



Fachbereich Informatik und Medien

MASTERARBEIT

**Das Usability-Labor im wissenschaftlichen Einsatz -
Aufbau und Bewertung eines Versuchs im Usability-Labor**

Vorgelegt von
Claudia Matthias B.Sc.

zum
Erlangen des akademischen Grades

MASTER OF SCIENCE
(M.Sc.)

Abgabedatum: Brandenburg a. d. Havel, den 26.10.2012
Erstbetreuer: Herr Prof. Alexander Urban
Zweitbetreuer: Herr Prof. Dr. sc. techn. Harald Loose

KURZREFERAT

Ziel der Arbeit ist die Entwicklung eines Forschungsdesigns zum vorgegebenen Thema „Kommunikation und Kommunikationsmittel im Arbeitsprozess“ unter Einbezug der Erhebungsmöglichkeiten des Usability-Labors (UL) der Fachhochschule Brandenburg (FHB). Angeregt wurde die Thematik durch Vorschläge der Betreuer, durch Recherche von Fallstudien und Berichten sowie durch Introspektion (Selbstbeobachtung). Im Vordergrund der Arbeit stehen die *Entwicklung* des Forschungsdesigns zur Hypothesenprüfung sowie die abschließende probeweise Versuchsdurchführung zur *Bewertung* und Anpassung des Designs.

Erweist sich das entwickelte Forschungsdesign als sinnvoll, kann es zur Grundlagenforschung beitragen und Ausgangspunkt für weiterführende hypothesenprüfende Untersuchungen mit repräsentativen Stichproben bilden. Lässt sich kein definitives Ergebnis herausarbeiten und/oder erweist sich das Design als unzweckmäßig, sind ggf. Optimierungen des Versuchsaufbaus, der Fragestellung, abgeleiteter Hypothesen und/oder verwendeter Erhebungsmethoden zu überdenken und kurz darzulegen.

ABSTRACT

The objective of this work is the development of an experimental design for the topic "communication and communicative devices in the work process" with the integration of methods from the University's (Brandenburg University of Applied Sciences) own Usability Lab. The topic was inspired by suggestions from professors, through research of case studies and reports, and through introspection. The focus of this work are the *development* of a research design to test hypotheses, and pilot tests to *evaluate* and adapt the design.

If the design proves useful, it may contribute to basic research and provide a starting point for further hypothesis-based research with representative samples. If no definitive results are found and/or the design proves impractical, possible optimization of the experimental design, main questions, hypotheses and / or survey methods need to be considered and briefly listed and explained.

VORWORT UND DANKSAGUNG

Der Versuch wurde in Zusammenarbeit mit Frau Tina Meißner (Masterstudiengang Medizininformatik) behandelt. Die Aufgabenbereiche wurden in zwei voneinander abgetrennte Schwerpunkte unterteilt, die in separaten Masterarbeiten dargestellt werden sollen - zum einen die Entwicklung des theoretischen Forschungsdesigns, inklusive aller benötigten Materialien sowie erste praktische Durchführungen durch die Autorin, zum anderen die technische Durchführung des Versuchs und Auswertung der Messdaten anhand der Werkzeuge und Techniken aus dem UL und ggf. Biosignalverarbeitungslabor (BS) durch Frau Tina Meißner.

In den nachfolgenden Ausführungen werden wichtige Inhalte von Frau Tina Meißner, die die Planungsarbeit beeinflussen, zur Übersicht und zum Verständnis zusätzlich kurz aufgeführt, mit konkretem Hinweis auf ihre Erstellung außerhalb dieser Arbeit. Nachfolgend wird Frau Tina Meißner in dieser Arbeit als „technischer Experimentator“ oder Versuchsleiter 1 (VL1) bezeichnet.

Verweise auf Abbildungen, Tabellen, Anhänge und Inhalte der beigefügten CD werden mit dem Symbol ► gekennzeichnet. Zusammenfassungen wichtiger Inhalte der einzelnen Abschnitte, relevante Hinweise oder Ergänzungen werden im laufenden Text in einem grau hinterlegten Kasten dargestellt.

Erster Dank geht an Herrn Prof. Dr. Harald Loose für die Überlassung des Themas und Herrn Prof. Alexander Urban für ihre gemeinsame Unterstützung und fachliche Betreuung. Besonderer Dank gilt Frau Tina Meißner für die produktive Zusammenarbeit am Thema. Weiterhin ist allen freiwilligen TeilnehmerInnen zu danken, die mit ihrer Bereitschaft zur Teilnahme sehr halfen, diese Arbeit zu erstellen sowie Familie und Freunden für ihre Geduld und liebevolle Unterstützung.

INHALTSVERZEICHNIS

1. Einleitung	1
1.1 EINFÜHRUNG	1
1.2 FOKUS UND ZIELE	1
1.3 ABGRENZUNG	2
1.4 AUFBAU UND VORGEHEN	2
2. Grundlagen	3
2.1 ALLGEMEINE PRINZIPIEN UND SCHRITTE SOZIALWISSENSCHAFTLICHER FORSCHUNG	3
2.1.1 Forschungsdesign	3
2.1.2 Quantitative Forschung	4
2.1.3 Hypothesenprüfende Untersuchungen	4
2.1.4 Versuchspläne	6
2.1.5 Faktor Mensch	8
2.1.6 Grundgesamtheit, Stichprobe und Auswahlentscheidung	9
2.2 METHODENREPERTOIRE	10
2.2.1 Experiment	11
2.2.2 Interview und Befragung	14
2.3 USABILITY-LABOR IN DER INFORMATIK	15
2.3.1 Allgemein	15
2.3.2 Methoden	15
2.3.3 Gütekriterien	17
2.3.4 Vor- und Nachteile	17
3. Forschungsdesign	18
3.1 FORSCHUNGSSTAND	18
3.2 THEMA UND FRAGESTELLUNG	21
3.3 (SACH-)HYPOTHESEN UND ERWARTUNGEN	22
3.4 VORGABEN UND AUSGANGSSITUATION	23
3.5 EXPERIMENTALVARIABLEN	24
3.5.1 Definition	24
3.5.2 Typen	24
3.5.3 Anwendung	27
3.5.4 Kontrolle von Störvariablen	28
3.6 METHODEN UND ERHEBUNGSVERFAHREN	31
3.6.1 Experiment	31
3.6.2 Fragebogen	33
3.7 TECHNISCHER RAHMEN	35
3.8 KOMMUNIKATIONSMITTEL	36

3.9 VERSUCHSZIEL, AUFGABENPLANUNG UND -AUSWAHL.....	37
3.9.1 <i>Versuchsziel und Bedingungen</i>	38
3.9.2 <i>Ansätze und Auswahl</i>	39
3.10 ORGANISATORISCHER RAHMEN.....	42
3.10.1 <i>Versuchsmaterialien</i>	42
3.10.2 <i>Versuchsleiter</i>	43
3.11 VORGEHEN	43
3.11.1 <i>Allgemein</i>	43
3.11.2 <i>Planung</i>	44
3.11.3 <i>Versuchsplan</i>	45
3.12 VERSUCHSPERSONEN.....	47
3.13 MESSGRÖßEN UND AUSWERTUNG.....	51
3.14 GÜTEKRITERIEN.....	52
3.15 ETHISCHE BEDENKEN	55
3.16 EMPIRISCHE VORHERSAGE UND STATISTISCHE HYPOTHESEN.....	55
4. Durchführung.....	57
4.1 EINLEITUNG.....	57
4.2 (SACH-)HYPOTHESEN UND OPERATIONALISIERUNG	57
4.3 METHODE UND VERSUCHSPLAN	57
4.3.1 <i>Methoden, Versuchsapparatur und –aufbau</i>	57
4.3.2 <i>Aufgaben</i>	58
4.3.3 <i>Versuchspersonen</i>	58
4.3.4 <i>Versuchsplan und –ablauf</i>	59
4.3.5 <i>Versuchsbedingungen</i>	60
4.3.6 <i>Empirische Vorhersage</i>	61
4.4 VOREXPERIMENTE	61
4.4.1 <i>Ergebnisse und Auswertung</i>	61
4.4.2 <i>Diskussion und Zusammenfassung</i>	62
4.5 HAUPTEXPERIMENTE	64
4.5.1 <i>Durchführung und Protokollierung</i>	64
4.5.2 <i>Ergebnisse und Auswertung</i>	65
4.5.3 <i>Diskussion und Zusammenfassung</i>	67
5. Auswertung und Zusammenfassung.....	70
5.1 PLANUNG UND DURCHFÜHRUNG	70
5.2 ERGEBNISSE UND GENERALISIERUNG	74
5.3 FAZIT UND AUSBLICK	75

Literaturverzeichnis	A-1
Tabellenverzeichnis	A-5
Abbildungsverzeichnis	A-6
Eidesstaatliche Erklärung.....	A-7
Anhang A. Details.....	A-8
A.1 Schritte der Forschung.....	A-8
A.2 Versuchspläne	A-9
A.3 Stichproben	A-13
Anhang B. Planung.....	A-16
B.1 Gesamtplanung	A-16
B.2 Versuchsaufbau.....	A-17
B.3 Zeitplan.....	A-19
B.4 MindMap zum Versuchsfragebogen.....	A-20
B.5 MindMap zu den Versuchspersonen-Aufgaben.....	A-21
Anhang C. Vorlagen.....	A-22
C.1 Anwerbung Versuchspersonen (Vorexperiment)	A-22
C.2 Anwerbung Versuchspersonen (Hauptexperiment).....	A-23
C.3 Teilnehmerinformation	A-24
C.4 Kommunikationsübersicht	A-25
C.5 Abläufe.....	A-26
C.6 Zeitablaufplan.....	A-27
C.7 Anleitung Versuchsperson.....	A-30
C.8 Probeaufgaben	A-31
C.9 Anleitung Versuchsleiter.....	A-33
C.10 Stimmungsfragebogen.....	A-34
C.11 Versuchsfragebogen	A-36
C.12 Debriefing.....	A-39
Anhang D. Ergebnisse.....	A-41
D.1 Versuchspersonen-Aufgaben.....	A-41
D.2 Technik und Material.....	A-42
D.3 Stimmungsfragebogen.....	A-43
D.4 Versuchsfragebogen	A-44
D.5 Versuchspersonen.....	A-48

Inhalte der beiliegenden CD: Masterarbeit als PDF-Datei

- Alle entwickelte Versuchsmaterialien (u.a. Vorlagen, Anweisungen, Textbausteine, siehe ►3.10.1)
- Abbildungen und Quellen (Internet, Paper, Buchauszüge)
- Unterlagen und Ergebnisse aller Versuchspersonen (codiertes, anonymisiertes Rohmaterial)

ABKÜRZUNGSVERZEICHNIS

aV	abhängige Variable
BS	Biosignalverarbeitungslabor
D1	Durchgang 1 (<i>Durchführung der ersten Messung</i>)
D2	Durchgang 2 (<i>Durchführung der zweiten Messung</i>)
EG	Experimentalgruppe
FHB	Fachhochschule Brandenburg
HE	Hauptexperiment(e)
HH	Haupthypothese
KG	Kontrollgruppe
MWT	Mehrfach-Wortschatz-Intelligenz-Test (MWT-B und Parallelförm MWT-A)
SV	Störvariable
TH	Teilhypothese
UL	Usability-Labor(e)
uV	unabhängige Variable
VE	Vorexperiment(e)
VL	Versuchsleiter
VL1	VersuchsleiterIn im UL, technischer ExperimentatorIn (<i>hier: Tina Meißner, Masterstudentin Medizininformatik an der FHB</i>)
VL2	VersuchsleiterIn im separaten Beobachtungsraum (<i>hier: Verfasserin dieser Arbeit</i>)
VP	Versuchsperson(en)

TEIL I

EINFÜHRUNG

1. EINLEITUNG

1.1 Einführung

In der heutigen „Informations- und Kommunikationsgesellschaft“ besitzt Kommunikation zunehmend Relevanz und erfolgt sowohl analog als auch digital, zeitversetzt und in Echtzeit und hebt Distanzen auf. Die Vielzahl an Kommunikationsmitteln (u.a. Email, Telefon, Instant Messaging) nimmt gehäuft Einzug in Privat- und Berufsalltag, besteht oft losgelöst voneinander und ermöglicht ein breites Spektrum an Kommunikation und Informationser-/vermittlung. Dies erzeugt jedoch gleichzeitig Unterbrechungen mit wichtigen *und* unwichtigen Informationen (cognitive overload, Stichwort: Flaschenhals menschlicher Informationsverarbeitung). Inwieweit beeinflusst diese Entwicklung nun den Menschen - was bewirken Kommunikation und die Integration gängiger Kommunikationsmittel in den Arbeitsprozess? Hemmen und beeinträchtigen sie die Leistung (Ablenkung, Aufmerksamkeits-/ Konzentrationsstörung, Kosten) oder fördern sie dynamischeres Arbeiten und schnellere effektivere Arbeitsabläufe (Teamarbeit, zeitnaher Informationsaustausch, informierte Entscheidung, rechtzeitige Intervention)? Untersuchungen zum Thema sind noch dünn gesät, fallen teils unwissenschaftlich aus oder begrenzen sich stark auf spezifische Kommunikationsmedien oder Personengruppen (In-House Studien). Empirische Ergebnisse können Anstoß geben, Arbeitsprozesse neu zu gestalten, um Kommunikation zu managen, Produktivität zu erhöhen und Kosten zu senken. Dazu ist Grundlagenforschung notwendig (Folgen von Kommunikation am Arbeitsplatz), auf deren Basis Kommunikationsmodelle konzipiert werden können (*aktives* Managen von Kommunikation). (s. [Br10][Bi98][Sc11])

1.2 Fokus und Ziele

Ziel der Arbeit ist die Entwicklung und Bewertung eines Forschungsdesigns zu einem vorgegebenen Thema mit festgelegtem Versuchsort im UL der FHB. Fokus liegt auf der Planung eines standardisierten Aufbaus inklusive Entwicklung aller benötigten Materialien und abschließender Bewertung via erster Versuchsdurchführungen. Alle behandelten Aspekte sollen konzeptionell beleuchtet werden, auch bzgl. der Wechselwirkung mit inhaltlichen Kriterien (d.h. praktische Umsetzbarkeit, Qualität, Validität). Die Arbeit soll an die soziologische Forschung und Studien aus der Wirtschaft anknüpfen und diese ggf. fortführen. Praxisrelevanz ergibt sich aus der tagesaktuellen Situation (Kommunikation/-smitteln im Arbeitsalltag). Dabei besitzen Effektivität, Effizienz (Zeit) und Umsetzbarkeit der Lösung zentralen Stellenwert. Zudem überschneiden sich wissenschaftliche und praxisnahe Bereiche, die bei Planung und Umsetzung zu berücksichtigen sind (interdisziplinärer Ansatz) - Aufbau von Forschungsdesigns in den Sozialwissenschaften, Methoden und Möglichkeiten des UL, Kommunikation/-smittel bei der Arbeit. Es handelt es sich um ein anwendungsorientiertes Forschungsvorhaben, das sich gegenüber rein theoretischen Arbeiten dadurch auszeichnet, dass es sich mit einer praxisrelevanten Problemstellung auseinandersetzt und Grundlage für weiterführende praktische Maßnahmen bilden soll. Auf dem Weg dahin sind der aktuelle Forschungsstand und Anknüpfungspunkte für die eigene Arbeit zu klären (►3.1). Das Projekt ist angesiedelt im Bereich Grundlagenforschung zur Lösung praktischer Probleme (Kommunikation im Arbeitsalltag) und besitzt funktionalen Wert.

1.3 Abgrenzung

Die technische Planung, Durchführung und Auswertung des Versuchs mit den Erhebungswerkzeugen des UL (und BS) gehören nicht zum Schwerpunkt dieser Arbeit und werden in einer separaten Arbeit durch Dritte behandelt (Datenerhebung, -auswertung, -interpretation). Die Vorgabe des Versuchsortes (im UL) und z.T. der Erhebungsmethoden und –werkzeuge sowie die Auslagerung technischer Aspekte müssen bei den theoretischen Überlegungen berücksichtigt werden und ziehen ggf. Einschränkungen nach sich. Weitere Grenzen ergeben sich durch den organisatorischen Rahmen u.a. verfügbare Technik, begrenzte Ressourcen / Zeit / Aufwand / Kosten im Rahmen der Masterarbeit. Die Arbeit erreicht nicht das finale Ziel, welches aus der empirischen Prüfung der (hier entwickelten) Hypothesen mit großen repräsentativen Stichproben und verallgemeinerbarer Ergebnisse besteht. Im Vordergrund der Arbeit steht der theoretische Aufbau eines geeigneten Forschungsdesigns mit anschließender Bewertung mittels exemplarischer Versuchsdurchführung. Die Datenaufbereitung /-auswertung aus den Durchführungen erfolgt an dieser Stelle wertungsfrei und dient primär dem Zweck der Bewertung des Forschungsdesigns. Ergebnisse sollen maximal erste Tendenzen aufzeigen und zusammen mit dem Forschungsdesign eine Grundlage für zukünftige empirische Weiterarbeit bilden.

1.4 Aufbau und Vorgehen

In den Grundlagen erfolgt die Ausarbeitung relevanter Aspekte zum Thema Forschungsdesigns im Bereich der Sozialwissenschaften, immer auch im Hinblick auf das geplante Forschungsvorhaben, um anschließend Techniken, Methoden und Möglichkeiten von ULs zusammenzufassen.

Der Hauptteil der Arbeit befasst sich mit der Entwicklung des theoretischen Forschungsdesigns, unter Berücksichtigung praktischer Möglichkeiten. Es erfolgt zunächst die Recherche und kritische Auseinandersetzung mit einschlägigen Arbeiten zum Versuchsthema (bisheriger Forschungsstand), um davon ausgehend sinnvoll Fragestellung und inhaltliche Untersuchungshypothesen abzuleiten. Anknüpfend werden Untersuchungsvariablen festgelegt (Operationalisierung) und geeignete Erhebungsverfahren ausgewählt. Zur Darstellung des methodischen Verfahrens zählt zudem ein sorgfältig ausgearbeiteter Untersuchungsplan, Beschreibung und Auswahl der Untersuchungsobjekte, Entwicklung des Untersuchungsmaterials, Aufführung eingesetzter Geräte, Durchführung und Auswertungsmethoden. Ziel ist es, ein begründetes, möglichst standardisiertes Forschungsdesign zu entwickeln, das auch durch andere Wissenschaftler theoretisch nachvollziehbar und ggf. replizierbar ist.

In der Durchführung wird das Design anhand erster exemplarischer Versuchsdurchführungen systematisch und methodisch kontrolliert erprobt, ggf. optimiert und bewertet. Dazu werden zunächst relevante Eckpunkte der Versuchsplanung aufgeführt, die Versuche durchgeführt und anschließend Durchführung und Ergebnisse beschrieben und dargestellt, kurz ausgewertet und zusammengefasst.

Zum Schluss erfolgt eine zusammenfassende Bewertung und Diskussion des Designs und der Ergebnisse, ggf. weitere Verbesserungsvorschläge sowie ein Ausblick für die zukünftige Anwendung des Designs.

TEIL II

GRUNDLAGEN

2. GRUNDLAGEN

Nachfolgend werden zentrale Begriffe und Konzepte für das eigene Vorhaben behandelt, beginnend mit Aspekten empirischer Forschung, an die Erhebungsmethoden des Versuchsorts im UL anknüpfen.

2.1 Allgemeine Prinzipien und Schritte sozialwissenschaftlicher Forschung

Empirische Untersuchungen lassen sich anhand zweier Kriterien klassifizieren¹ (s. [BN03, S.53ff]).

Hypothesenüberprüfung vs. -exploration: Nach der Literatursuche wird entschieden, ob sich explorative (Hypothesensuche) oder explanative Untersuchung (Hypothesenableitung/-prüfung) eignet. Explorative Untersuchungen sind erkundende Untersuchungen zur Hypothesenentwicklung oder Schaffung theoretischer Voraussetzungen zur Hypothesenformulierung (geringe Normierung, wenig standardisierte Datenerhebung, meist qualitative Methoden u.a. offene Befragung, Feldbeobachtung). Hypothesenprüfende Untersuchung meint begründete Hypothesenformulierung anhand des Standes der Theorienentwicklung, mit dem Ziel, vorhandene Annahmen zu prüfen (Hypothesenarten ►2.1.3).

Angestrebte Gültigkeit bzw. Eindeutigkeit der Aussagen in Hinsicht auf die Ergebnisse: Dies betrifft 1) den Gültigkeitsanspruch der Befunde, d.h. es wird die Untersuchungsart bzgl. interner (Eindeutigkeit der Ergebnisinterpretation und Aussagekraft) und externer Validität (Generalisierbarkeit und Gültigkeit) gewählt. Es wird zudem zwischen 2) experimentell und quasiexperimentell unterschieden, d.h. Versuche werden mit natürlichen (quasiexperimentell) oder zufällig zusammengestellten Gruppen (experimentell) durchgeführt, was primär interne Validität betrifft (►2.1.4). Zudem wird der Versuchsort auf 3) Feld oder Labor festgelegt. Ersteres nutzt eine möglichst unbeeinflusste, natürliche Umgebung (hohe externe Validität), störende Einflüsse sind aber schwer kontrollierbar und Ergebnisse uneindeutig (Alternativerklärungen, sinkende interne Validität). Dagegen sind untersuchungsbedingte Störgrößen im Labor gut kontrollierbar oder ausschaltbar, was interne Validität erhöht aber externe Validität durch die unnatürliche Versuchsumgebung mindert. (►2.2.1)

2.1.1 Forschungsdesign

Als Grundlage wissenschaftlicher Untersuchungen mit lebenden Subjekten, behandelt es das *Wie* der Untersuchung und ist entscheidend für Ergebnisaussagekraft. Im *experimentellen* Forschungsdesign, das sich auch für das eigene Vorhaben eignet (►2.1.4), werden Hypothesen via Manipulation der unabhängigen Variable (uV) geprüft und Störvariablen (SV) kontrolliert (*Kontrollstrategien* ►3.5.2).

Empirische Untersuchung: Sie fordert systematische Vorgehensdokumentation² (Wiederholbarkeit), präzise Terminologie und spezielle Auswertung (statistische Analyse), für allgemeingültige Aussagen (Generalisierung), um Gütekriterien zu prüfen (interne/externe Validität) und mit Theorien umzugehen (unterliegen ständiger Prüfung). Um notwendige Durchführungs- und Auswertungsmaßnahmen festzulegen, sind spezielle Schritte zu durchlaufen, an denen sich auch der Aufbau des eigenen Vorhabens orientieren soll ►Anhang A.1. (s.[el09][Ho97][Wp12][Zw98][Fa00][Wi12a][Ps12][BN03])

¹ Die Klassifikation soll hier auch zur Einordnung des eigenen Vorhabens und den Planungsschwerpunkten dienen.

² Standardisierung (methodische Regeln, Dokumentation) für Transparenz, Objektivität und intersubjektive Nachprüfbarkeit.

2.1.2 Quantitative Forschung

Grundsätzlich lassen sich in den Sozialwissenschaften zwei Methodenansätze unterscheiden, wobei in vielen Untersuchungen eine Kombination beider Ansätze oft sinnvoll ist.

Qualitative Forschung: Sie legt zu Explorationszwecken Gewicht auf „*Interpretationen von verbalem Material*“ [BN03, S.295] und verzichtet überwiegend auf Messungen. Datenmaterial kommt durch Verbalisierung der Erfahrungsrealität zustande (qualitative, nichtnumerische verbale Daten) und Verfahren sind teils auf spezielle Fragestellungen zugeschnitten und nicht immer übertragbar.

Quantitative Forschung: Hier erfolgen „*Quantifizierungen* der Beobachtungsrealität“ [BN03, S.295] mittels Experimentieren, Messen, Testen d.h. ein Ausschnitt der Beobachtungsrealität wird numerisch beschrieben und statistisch ausgewertet. Untersuchungsvorgänge sollten standardisiert sein. Vorteile der Erhebung und Auswertung quantitativer Daten liegen in der Zeitökonomie gegenüber qualitativen Daten sowie leichter vergleichbaren Ergebnissen. Die Vielfalt subjektiver Antworten sinkt bei standardisierten Verfahren dagegen. Speziell *quantitative Experimente* legen Fokus auf *Hypothesenprüfung*. Da das eigene Vorhaben auf (experimentelle) Messungen im UL abzielt, bietet sich hier *Hypothesenprüfung* mittels *quantitativer Datenerhebung* im *Experiment* an. (s. [el09][Ge03][BN06])

2.1.3 Hypothesenprüfende Untersuchungen

„Hypothesenprüfende Untersuchungen testen Annahmen über Zusammenhänge, Unterschiede und Veränderungen ausgewählter Merkmale bei bestimmten Populationen.“ [BN06, S.490] und sollen Ursache-Wirkungs-Beziehungen identifizieren oder Kausalannahmen testen.

Wissenschaftliche Hypothesen: Sie sind 1) empirisch untersuchbar d.h. Bezug auf Reales und Untersuchbares 2) generalisierbar³ als allgemeingültige Aussagen über den Einzelfall *hinaus* 3) als Konditionalsatz formuliert d.h. zumindest implizit als sinnvoller Wenn-Dann-/Je-Desto-Satz und 4) falsifizierbar d.h. widerlegbar, widerspruchsfrei, operationalisierbar. Hypothesen gelten als vorläufige Antwort auf Forschungsfragen, sind aber empirisch zu beweisen d.h. Bewährung als wahr (Verifizierung) oder falsch (Falsifizierung) durch Untersuchungen. (s. [As01][Ex12][el09][BN06][BN03])

Hypothesenentwicklung: Generell geht hypothesengeleitete Forschung von einer Fragestellung aus. Auf Grundlage wissenschaftlicher Literatur, Theorien, Beobachtung etc. können Hypothese(n) auf der Sachebene abgeleitet und Variablenidentifikation sowie die Auswahl der Versuchspersonen (VP) und Akquirierung der Stichprobe vorgenommen werden. Fragestellung und Hypothese(n) sind einfach, kurz und eindeutig zu formulieren und der untersuchte Bereich klar abzugrenzen. Hypothesen werden schließlich zur empirischen Vorhersage der Ergebnisse konkretisiert und in statistische Hypothesen⁴ transformiert. Um Allgemeingültigkeit zu gewährleisten, ist für statistische Hypothesen die gesamte interessierende Population wichtig statt nur Stichproben. Sie haben den Charakter von Wahrscheinlichkeitsaussagen mit probabilistischer Variablenbeziehung. Zuletzt erfolgen Durchführung und Messung via (nicht-)experimenteller Untersuchungen sowie Ergebnisauswertung und -darstellung gemäß der Sachhypothese.

³ Die behauptete Beziehung zwischen mind. 2 Variablen ist auf alle Populationen vergleichbarer Ereignisse generalisierbar.

⁴ Eine *statistische Hypothese* „sagt das Ergebnis einer empirischen Untersuchung vorher (Prognose)“ [BN03, S.27].

Hypothesenprüfung: Es wird angenommen, A sei Ursache von B, im Falle das „(1) A B vorausgeht (2) andere Ursachen für das Zustandekommen von B ausgeschlossen werden können, was (3) nur durch den Nachweis, wenn Non-A, dann Non-B stringent falsifiziert, nicht aber verifiziert werden kann“ [Fa00, S81] Vorausgesetzt wird Widerspruchsfreiheit, Kritisierbarkeit d.h. Ereignisse existieren, bei denen die Hypothese nicht zutrifft, Operationalisierbarkeit (Variablenfestlegung) sowie Hypothesen *vorab* d.h. unabhängige Hypothesenformulierung und –prüfung durch u.a. schriftliches Festhalten um nachträgliche Umdeutung/Immunsierung zu vermeiden. Einzelne Ergebnisschwankungen sind zulässig, human-/sozialwissenschaftliche Hypothesen beinhalten den Nachweis, dass die „behauptete Beziehung zwischen den Variablen ‚im Prinzip‘ besteht, und nicht, daß der Wenn-dann-Satz für *jedes einzelne* Untersuchungsobjekt *perfekt* zutrifft.“ [BN03, S.13] Widersprechende Ergebnisse sind nicht auszuschließen, sodass Verifikationsverfahren auf Induktionsschlüsse hinauslaufen, bei denen „von einer begrenzten Anzahl spezieller Ereignisse unzulässigerweise auf die Allgemeingültigkeit der Theorie geschlossen wird.“ [BN03, S.21], da Populationen oft nicht voll untersuchbar sind.

Hypothesenarten: Nicht alle Arten sind falsifizierbar/verifizierbar (nach [BN06][el09][Fa00]).

Unterteilung nach Aussagebereich der Hypothese

Universelle Hypothese: Aussagen gelten ausnahmslos für alle Fälle des Bereichs. Es sind strikt universelle (ausnahmslos) und quasiuniverselle Hypothesen (Einschränkung auf Teilbereich z.B. Frauen) zu trennen. Ist der Geltungsbereich zu groß, um alle Fälle zu untersuchen, kann die universelle Hypothese nicht verifiziert aber falsifiziert werden, 1 Fall genügt. Bei kleinem Geltungsbereich, d.h. alle betreffenden Objekte sind überprüfbar, ist die universelle Hypothese auch verifizierbar.

Existenzielle Hypothese: Aussagen gelten für mind. 1 Fall. Bei großem Geltungsbereich können sie nur verifiziert werden. Ist der Geltungsbereich klein, sind sie auch falsifizierbar.

Hypothese über Anteile: Es bestehen Vermutungen über prozentuale Anteile aller möglichen Fälle (statistische stochastische Sachhypothese) oder Wahrscheinlichkeit, mit der abhängige Variablen (aV) zutreffen (probabilistische stochastische Hypothese). Bei großem Geltungsbereich ist Falsifikation oder Verifikation nicht möglich, ist er kleiner, ist beides möglich.

Hypothesen sind auch am Grad der Konkretisierung unterscheidbar

Forschungshypothese (inhaltliche Hypothese): Sie ist eine aus der Theorie abgeleitete Vermutung (enthält Konstrukte).

Operationale Hypothese: Sie sagt den Ausgang der konkreten Untersuchung anhand der allgemeinen Forschungshypothese voraus und resultiert aus Untersuchungsplanung bzw. Operationalisierung der uV und aV (Konstrukte sind operationalisiert).

Statistische Hypothese: Sie besteht aus einem statistischen Hypothesenpaar, d.h. aus Alternativ- und Nullhypothese⁵.

Unterscheidung anhand des Untersuchungsziels (formale Klassifikation) - Übersicht formaler Forschungshypothesen

Zusammenhangshypothese: Sie postuliert einen Zusammenhang zwischen mind. 2 Merkmalen.

Unterschiedshypothese: Es werden mind. 2 Stichproben verglichen, die sich bzgl. mind. einer aV unterscheiden.

Veränderungshypothese: Es erfolgt ein Vergleich mind. einer Vorher- mit einer Nachhermessung.

Hypothesen in Einzelfalluntersuchungen: Annahmeprüfung über Merkmale / Verhaltensweisen einzelner Person/Objekts.

Unterscheidung (statistischer) Hypothesen durch Angabe der Richtung des hypothetischen Effekts (Gerichtetheit)

Ungerichtete (zweiseitige) Hypothese: Der Effekt kann in „beide Richtungen“ gehen (Richtung des Zusammenhangs / Unterschieds / Veränderung nicht vorgegeben). Der Hypothesentyp gilt in beiden Fällen als verifiziert.

Gerichtete (einseitige) Hypothese: Die Richtung des Effekts (des Unterschieds / Veränderung) wird vorhergesagt.

Zuletzt kann noch nach Spezifität für (statistische) Hypothesen unterteilt werden

Unspezifische Hypothese: Es gibt keine Größenangabe (Effektrichtung) über Zusammenhang / Unterschied / Veränderung.

Spezifische Hypothese: Es gibt eine Angabe über die Effektgröße des Zusammenhangs / Unterschieds / Veränderung.

⁵ Die Alternativhypothese ist die statistische Formalisierung der inhaltlichen Forschungsfrage und entspricht oft der statistischen Annahme. Die Nullhypothese ist die Verneinung der Alternativhypothese (komplementäre statistische Annahme), die nicht zu beweisen ist.

2.1.4 Versuchspläne

Versuchspläne sind in 3 Kategorien klassifizierbar: experimentell, quasi-experimentell und korrelativ. Die Bewertung dazugehöriger Designs erfolgt über ihre interne Validität, angeführt von experimentellen Designs (► Tab.1). Die Design-Zuordnung zu eine der 3 Formen liegt oft im Ermessen des Autors und hat erheblichen Einfluss darauf, ob Ergebnisinterpretationen zulässig sind (interne Validität).

(Strenges) Experiment: Der Ursache-Wirkungs-Zusammenhang wird durch systematische, willkürliche uV-Änderung und Erfassung der aV-Effekte via Experimental (EG)- und Kontrollgruppen (KG)-Vergleich erhoben und grenzt sich ab durch Randomisierung⁶ (Gruppenunterschiede minimieren durch Neutralisieren personenbezogener SV), uV-Manipulation und hohe interne Validität.

Quasi-Experiment⁷: Nicht alle relevanten Variablen sind kontrollier-/variierbar und VP den uV-Stufen *nicht* zufällig zugeordnet, ggf. bestehen *vorab* Gruppenunterschiede. Diese bewirken z.B. durch Selektion, dass aV-Änderungen nicht eindeutig auf die uV zurückzuführen und Ergebnisse uneindeutig sind, mit Alternativerklärungen und eingeschränkter interner und externer Validität.

Korrelationsstudie (Zusammenhang zwischen mind. 2 Merkmalen): Die Studie ist weder Experiment noch Quasi-Experiment (kein eigentliches Treatment). Variablen werden gleichzeitig erfasst, uV-Identifikation erfolgt nachträglich (ex post facto). Man spricht nur von begründeten Vermutungen.

Tab. 1 Gegenüberstellung der Designs und Mischformen bzgl. relevanter Kriterien ([Fa00, S.83])

Designtyp	Multiple Randomisierung	uV-Änderung	Hypothese vorab	Interne Validität
strenges Experiment	Ja	Ja	Ja	+++
Experiment	Ja	Ja	Ja	++
experimentell-korrelative Studie	Zum Teil	Zum Teil	Ja	+
Quasiexperiment	Nein	Ja	Ja	+
Mischpläne	Zum Teil	Zum Teil	Ja	-
Ex post facto-Studie	Zum Teil	Zum Teil	Ja	--
Korrelationsstudie	Nein	Nein	Nein	---

Versuchsplan: Als „standardisiertes, routinemäßig anwendbares Schema (Strukturschema), das dem Aufbau, der Kontrolle und der methodologischen Bewertung einer empirischen Untersuchung von unabhängigen (UV) und abhängigen (AV) Variablen sachlogisch zugrunde liegt.“([Sa92] zit. nach [el09]), umfasst er als Schnittstelle zwischen Problemstellung, Versuchsaufbau und -ablauf die Variablen- und VP-Auswahl, Operationalisierung, Erhebungsart, Auswertung und Interpretation. Via formalisierter Darstellung wird ermittelt, ob aV-Effekte der uV-Änderung zuzuschreiben und Ergebnisse verallgemeinerbar sind (Hypothesenprüfung, interne / externe Validität). (s. [Ps12][Fa00][BN06])

Vorteile standardisierter, abstrakter Versuchspläne: Vorteilhaft sind eine hohe Objektivität, routinemäßige Anwendung, Dokumentation von Studien, höhere Nachvollziehbarkeit, leichtere wissenschaftliche Kommunikation.

Information aus Versuchsplänen: Anzahl uV / aV / Messungen, uV-Stufen, VP-Zuteilung zu Stufen, Kontrolltechniken

Funktion: Sie enthalten formalisierte Planungsprinzipien/Handlungsanweisungen, erhöhen Objektivität und wissenschaftliche Kommunikation, dokumentieren Untersuchungen (nachvollziehbar, transparent), sind Entscheidungsgrundlage für statistische Auswertung und Abwägung zwischen Geltungs- (*externe Validität*) und Gültigkeitsanspruch (*interne Validität*).

⁶ Einfache Randomisierung ist eine Zufallsstichprobe, multiple Randomisierung durchläuft die Schritte 1) Zufallsstichprobe 2) zufällige VP-Zuweisung zu den Gruppen und 3) zufällige Zuweisung dieser Gruppen zu den Treatments (EG und KG).

⁷ “Werden mehrere Gruppen untersucht, die nicht durch Randomisierung hergestellt werden, sondern in ihrer ‚natürlichen‘ Zusammensetzung vorliegen, so spricht man von einer quasiexperimentellen Untersuchung.“[BN06, S.525]

Versuchspläne von Experimenten: Sie stellen schematisch Ursache-Wirkungs-Zusammenhänge dar, zum Hypothesenprüfen oder Explorieren von Zusammenhängen / möglichen Hypothesen (z.B. vorexperimentelle Designs) und erlauben *beschreibende* Versuchsanordnungen (korrelative Designs).

In gültigen Versuchsplänen sind 1) uV und aV abgrenzbar, 2) *folgt* die aV der uV und 3) werden mind. 2 Gruppen verglichen. Die Auswahl des Versuchsplans richtet sich nach Versuchssituation und Frage (es gibt nicht *den* Versuchsplan). Es wird nach wichtigen Gemeinsamkeiten unterteilt.

Die 4 wichtigsten Gruppen der Methodenlehre und ihre Untergruppen (zusätzliche Details ► Anhang A.2)

1) Vorexperimentelle Designs: Es findet eine 1malige Untersuchung an einer einzigen Versuchsgruppe statt, d.h. Vorher-Nachher-Messung an einer einzigen Versuchsgruppe oder ein statischer Gruppenvergleich.

2) Experimentelle Designs: Experimentelle Typen sind Randomisierungs-, Messwiederholungs- und Blockversuchspläne. Erweiterungen/Mischformen sind Solomon-, Mischversuchs-, multivariate, mehrfaktorielle oder unvollständige Pläne.

3) Quasiexperimentelle Designs und 4) Ex-Post-Facto-Designs und Korrelative Designs

Vorexperimentelles Design: Dies ist ein ungültiger (intern invalider) Plan ohne zulässige kausaltheoretische Rückschlüsse (keine Hypothesenprüfung), da präexperimentelle Ausprägungen der aV und uV nicht kontrolliert werden, suffiziente Vergleichs-/Kontrollmöglichkeiten von SV fehlen und Versuchsbedingungen oft unstandardisiert sind. Gültige experimentelle Studien sind teuer, daher bietet sich dieses Design für kleinere kostengünstige Voruntersuchungen an, um zu ermitteln, ob sich weitere Untersuchungen lohnen oder für explorative und hypothesengenerierende Fragestellungen.

Experimentelles Design: Dies sind Pläne für echte, strenge, intern valide Experimente zur (Kausal-) Hypothesenprüfung (► Tab.2). VP werden uV-Stufen oder Versuchsbedingungen zufällig zugeordnet (Randomisierung) und SV möglichst ausgeschlossen. Kontrollstrategien erlauben hohe interne Validität, u.a. *Randomisierung* (gleichmäßige Zufallsverteilung un-/bekannter SV für Gruppenvergleiche), *Wiederholungsmessung* (intraindividuelle Unterschiede konstanthalten durch Vergleich der Gruppe mit sich selbst) und *Blockbildung* (homogene, innerhalb vergleichbare Blöcke mit bekannten SV).

Tab. 2 Übersicht über die 3 wichtigsten experimentellen Designs (nach [el09])

	Randomisierungspläne (Between) ⁸	Wiederholungspläne (Within) ⁹	Blockversuchspläne
<i>Kontrolle</i>	Randomisierung	Wiederholungsmessung	Blockbildung
<i>Voraussetzung</i>	Gesetz der großen Zahl (viele VP)	Ausreichend Zeitabstand zwischen Messungen, ggf. Ausbalancierung ¹⁰	Vorkennntnis der SV, ggf. Voruntersuchung
<i>Vorteile</i>	auch unbekannte SV kontrollierbar, keine Übertragungseffekte	Auch unbekannte interindividuelle SV werden konstant gehalten, ökonomisch (geringe Stichprobe reicht), geringere Datenstreuung	bekannte SV kontrollierbar, ökonomisch (geringe Stichprobe nötig), wenig Datenstreuung, keine Übertragungseffekte
<i>Nachteile</i>	Große Stichprobe nötig	Reifung, Zeiteinfluss, Test-, Übertragungseffekt, Interferenz mehrfacher Behandlungen	SV muss explizit bekannt sein

⁸ Between-Designs sind Randomisierungspläne (hohe interne Validität), d.h. verschiedene Bedingungen werden verschiedenen Teilstichproben zugeordnet und uV-Wirkung durch Gruppenvergleiche ermittelt (d.h. interindividuelle Varianz).

⁹ Within-Design sind Messwiederholungspläne. Dieselbe VP wird unter verschiedenen uV-Ausprägungen untersucht und jeweils aV erhoben. Die mehrfachen Messungen erlauben Vergleiche innerhalb der Gruppe (intraindividuelle Varianz).

¹⁰ Ausbalancierung bedeutet, dass die Reihenfolge der experimentellen Bedingungen ausgeglichen variiert wird, möglichst ohne Wiederholung oder gleichmäßige Verteilung, ggf. per Zufall, um Reihenfolgeeffekten und Interferenz vorzubeugen.

Quasiexperimentelles Design: VP-Einteilung in KG und EG erfolgt nicht zufällig (EG liegt bereits vor). Interne Validität hängt aber bei Veränderungshypothesen stark von der KG ab und inwiefern EG und KG äquivalent sind. Diese Designs ergänzen experimentelle Untersuchungen, Versuchsbedingungen sind variierbar aber Sekundärfaktoren mindern interne Validität. (s. [BN06][Wi12c][Fa00])

Design für Entwicklungen: Zur Prüfung *entwicklungsbedingter* Veränderungshypothesen werden Quer- oder Längsschnittuntersuchungen (Mehrzeitpunktuntersuchungen) angewandt, die jedoch aufgrund konfundierter Alters-, Generations- und epochaler Effekte problematisch sind. (s. [BN06])

Zu **Ex-Post-Facto-** und **korrelativen Designs** zählen u.a. Korrelations- und Ex-post-facto-Studien.



Abb. 1 Differenziell-/sozialpsychologische Aspekte [Fa00,S.50]

stimmung, Informationsverarbeitung, Bewusstseinsprozesse, Interaktion zwischen VP und VL wie Kommunikation und Rollen und ethische Richtlinien (s. [el09]). Differenziell- und sozialpsychologische Aspekte eines Versuchs können auch als SV wirken¹¹ ►3.5.4.

Ethische Richtlinien: 1) VL sorgen für ethische Unbedenklichkeit der Forschung und 2) sind verantwortlich für Mitarbeiterverhalten (diese sind zudem für sich verantwortlich). 3) Sie müssen VP über den Versuch möglichst ganz aufklären, bevor VP zustimmen, 4) offen und ehrlich sein, Täuschungsmanöver vermeiden bzw. VP *nach* Experiment aufklären und Teilnahme nicht erzwingen. VP dürfen jederzeit ausscheiden und 5) müssen die Vereinbarungen kennen. VL müssen eingegangene Verpflichtungen erfüllen und 6) VP vor Unbequemlichkeiten, Schaden, Gefahr bewahren und über Risiken voll informieren. Experimente, die zu schweren, dauerhaften Schäden führen können, sind unzulässig. 7) Am Versuchende sind VP über Art und Bedeutung der Befunde vollständig aufzuklären, 8) unerwünschte Folgen festzuhalten / zu beseitigen und 9) gesammelte Informationen vertraulich zu behandeln. Falls dies nicht möglich ist, ist die VP *vorher* zu informieren.

Versuchsleiter: Sie sind für die Versuchsdurchführung zuständig, d.h. Versuchsplan korrekt und einheitlich und ethische Grundsätze umsetzen. Daher sind menschliche Fehler nicht auszuschließen (u.a. Registrierfehler, Ermüdung) und Strategien zu finden, Ergebnisse VL-unabhängiger (objektiver) und / oder VL als „Messinstrument“ zuverlässiger zu machen z.B. durch Schulung ►3.5.4.

Versuchspersonen: Sie sind in empirischen Untersuchungen die *Untersuchungstichprobe* und bestehen gemäß ethischen Grundsätzen meist aus Freiwilligen (ergibt ggf. Validitätsprobleme ►2.1.6). VP

¹¹ Dazu zählen u.a. Neugier, Gefälligkeit, Teilnahmeanhänger, Bereitschaft, Forschung zu unterstützen. Verhaltensmotive *während* des Experiments sind *positiv* (Anpassungsverhalten durch u.a. Bedürfnis nach sozialer Anerkennung, Bewertungsangst) oder *negativ* (Vermeidung, versteckter oder offener Widerstand, psychologische Abwehr, Reaktanz d.h. "Gegenhandeln" u.a. durch zu hohe Anforderungen, subjektiv als Nötigung erlebte Instruktion, als Folge von Täuschung). (s. [Fa00, S.51])

2.1.5 Faktor Mensch

Da Menschen aktiv auf Umstände und Situationen reagieren, lernen, Verhalten steuern und kontrollieren, Hypothesen bilden, interagieren und das Messinstrument selbst ein Mensch ist (VL), entstehen Wechselwirkungen (Subjekt-Subjekt-Relation), die zu berücksichtigen sind

►Abb.1. Humanpsychologische Experimente charakterisieren sich u.a. durch Teilnahme-Motivation (Freiwilligkeit, Verweigerung), Ein-

werden in Experimenten oft in KG und EG unterteilt. In der EG wird die uV manipuliert (Treatment). Bei mehreren uV-Stufen sind mehrere EG notwendig, sie sollten homogen und vergleichbar sein, möglichst durch Zufallsgruppenbildung. Die KG wird ohne experimentelles Treatment untersucht (unbeeinflusste Situation), um Störeinflüsse des zwischenzeitlichen Geschehens (Änderungen inner-/außerhalb der VP, die nicht durch uV ausgelöst wurden) zu kontrollieren d.h. Sekundär- und Fehlervarianz. EG und KG müssen bzgl. ihrer Ausgangswerte vergleichbar sein, um mittels Differenz zwischen EG und KG Störeinflüsse herauszufiltern. Zudem kann durch KGs der auf die uV zurückgehende Effekt in seiner Größenordnung erfasst werden. (s. [BN06][Fa00][el09])

2.1.6 Grundgesamtheit, Stichprobe und Auswahlentscheidung

Grundgesamtheit (Population): Sie meint „alle potenziell untersuchbaren Einheiten, die ein gemeinsames Merkmal aufweisen.“([SR05, S.203] zit. nach [el09]) Ihr Umfang (N) ist die „Gesamtheit aller N Beobachtungseinheiten, über die Aussagen getroffen werden sollen.“[BN03, S. 397] Sie besteht aus Einheiten (n) z.B. Personen und aus ihr wird die Stichprobe entnommen.

Stichprobe: Die gesamte interessierende Population (Vollerhebung)¹² ist oft nicht untersuchbar, so dass Schlussfolgerungen notwendig sind. Daher werden Teilmengen (Stichproben) untersucht, die sich nach verschiedenen Prinzipien und Umständen bilden (systematisch gewonnene Auswahl) und deren Ergebnisse auf die Population verallgemeinert werden. Sie sind weniger aufwendig und lassen sich schneller durchführen und auswerten. Zu beachten sind Zuverlässigkeit der Ergebnisverallgemeinerung, Stabilität bzw. Replizierbarkeit, interne/externe Validität, der praktisch mögliche Aufwand, verfügbare Kompetenzen, bereits vorhandener Instrumente sowie Zumutbarkeit und Ökonomie.

Größe: Um Schlüsse zuzulassen, muss die Stichprobe hinreichende Größe besitzen, Daumenregel ist $n \geq 30$. „Es ist ein weit verbreiteter Irrtum, daß mit wachsender Stichprobengröße die Repräsentativität der Stichprobe generell steigt. Dies trifft nur bei *unverzerrter Auswahl* zu.“[BN03, S.402] d.h. die *Auswahlart* ist ebenso entscheidend, um Ergebnisse zu verallgemeinern.

Vorgehen: VP werden aus der Population gezogen, müssen zugreifbar sein und aus der definierten Population stammen. Dazu ist eine Liste aller Elemente der Population anzufertigen als Erhebungsrahmen. Es gilt $n < N$. Befunde aus Stichprobenuntersuchungen bilden aber nicht immer die Population exakt ab, es können unvermeidbare Unterschiede zwischen Grundgesamtheit und Stichprobe bestehen, wodurch Stichprobenfehler¹³ entstehen.

Fehler: Weitere systematische Fehler entstehen, wenn Einheiten der Grundgesamtheit nicht in der Erhebungsgesamtheit (*Untererfassung*) oder mehrfach enthalten (*Übererfassung*) sind oder nicht zur Grundgesamtheit gehören (*Übererfassung*). Aus Stichproben gewonnene Aussagen gelten an sich nur für diese Stichprobe, Gültigkeit für die Population wird nur angenommen. Je größer der Stichprobenumfang und je geringer die Verteilungsvarianz, desto kleiner der Stichprobenfehler. Systematische

¹² „Werden alle Objekte einer Population untersucht, so spricht man von einer Vollerhebung [...]. Wird nur ein Ausschnitt der Population untersucht, so handelt es sich um eine Stichprobenerhebung...“[BN03, S.397] Vollerhebung ist problematisch bei unendlich großer oder nur teilweise bekannter Population oder wenn die Untersuchung diese stark beschädigen / zerstören würde oder zu aufwendig wäre (s.[BN06]).

¹³ Unzulässige Verallgemeinerung der Ergebnisse auf die Grundgesamtheit gefährdet externe Validität ► 3.14 (s. [BN06]).

Fehler können auch durch Antwortausfälle entstehen, indem über Einheiten der Stichprobe keine oder keine vollständigen Infos zu erlangen sind oder als Messfehler durch missverständliche Fragestellung oder ungeeichte Erhebungsinstrumente. Neben Täuschung und bewusster Fehlreaktion der VP, kann die Verweigerungsrate durch Teilnahmeverweigerung/-ausfall Ergebnisse verzerren, wenn davon auszugehen ist, dass sich Verweigerer systematisch bzgl. relevanter Merkmale von Freiwilligen unterscheiden und die ursprünglich vorgesehenen Stichprobenumfänge nicht realisiert werden (Missing-Data-Probleme). Potentiellen VP ist die Teilnahme daher zu erleichtern, u.a. durch motivierende Einwirkung, angenehme Untersuchungsbedingungen und Vorbeugen technischer Pannen.

Arten: „Stichproben, bei denen [...] die Elemente die gleiche (oder zumindest eine bekannte) Auswahlwahrscheinlichkeit haben, nennt man probabilistische Stichproben [Ziehung via Zufallsprinzip]. Sind die Auswahlwahrscheinlichkeiten dagegen unbekannt, spricht man von nichtprobabilistischen Stichproben [gezielte Auswahl].“ [BN03, S.402] Repräsentative Stichproben sind Voraussetzung, um Vollerhebungen durch Stichproben zu *ersetzen*. Zufallsstichproben bieten die beste Repräsentativität.

Verschiedene Arten der Stichprobenauswahl (Details zu den einzelnen Arten ► Anhang A.3)

1) *Probabilistische Stichproben* sind einfache Zufallsstichproben, zu ihren *Sonderformen* gehören die proportionale, disproportionale und mehrdimensionale Schichtung sowie die geschichtete, mehrstufige und Klumpenstichprobe.

2) *Nicht-probabilistische Stichproben* sind die Gelegenheits- (Ad-hoc-Stichprobe), Quoten- und theoretische Stichprobe.

Nicht-probabilistische Stichprobe: Auswahlwahrscheinlichkeiten sind unbekannt / unkontrollierbar, sodass VP-Selektion nach subjektiven Kriterien erfolgt, ohne objektiven Zufallsmechanismus und keinen Rückschluss auf die Population/Verallgemeinerung zulässt. Für nicht-probabilistische Stichproben empfehlen sich Studien mit explorativem Charakter.

Probabilistische Stichprobe: Sie fordert Zufallsauswahl, die oft unrealistisch oder mit großem Aufwand und Datenschutzproblemen verbunden ist, was Zugeständnisse erforderlich macht (u.a. Abwesenheit der laut Stichprobenplan ausgewählten VP, Verweigerung). „Um mit Hilfe einer Stichprobenerhebung (anstelle einer Vollerhebung) gültige Aussagen über eine Population treffen zu können, muß die Stichprobe repräsentativ sein, d. h. sie muß in ihrer Zusammensetzung der Population möglichst stark ähneln.“ [BN03, S.401] *Repräsentativität* gibt das Ausmaß der Übereinstimmung der Stichprobe mit der Population an: „Eine Stichprobe ist (merkmals-)spezifisch repräsentativ, wenn ihre Zusammensetzung hinsichtlich einiger relevanter Merkmale der Populationszusammensetzung entspricht. Sie ist global repräsentativ, wenn ihre Zusammensetzung in nahezu allen Merkmalen der Populationszusammensetzung entspricht“ [BN03, S.401] *Repräsentativität* ist also eher vom Auswahlverfahren als der Größe abhängig. Daher sollten objektive und einfach festzustellende Merkmale gewählt werden. Repräsentative Stichproben lassen sich nur probabilistisch bilden, sind aussagekräftiger als nicht-probabilistische und empfehlen sich für wissenschaftliche Versuche. (s. [e109][Fa00][Fa97][BN06])

2.2 Methodenrepertoire

Methoden (gr. *Metá hodós*: der Weg zu etwas hin) sind der Weg wissenschaftlichen „Vorgehens“, Methodenlehre (Methodik, Methodologie) bezieht sich ergo auf *Wege* wissenschaftlicher Erkenntnis und *Verfahrensweisen*, mit Anleitung für planmäßiges wissenschaftliches Vorgehen. „Die Methoden

der empirischen Datenerhebung haben die Funktion, Ausschnitte aus der Realität, die in einer Untersuchung interessieren, möglichst genau zu beschreiben oder abzubilden.“[BN03, S.137] Sozialwissenschaftliche Methoden sind weitgehend etabliert und erprobt und werden aufgeteilt in: 1) mehr aktives oder rezeptives (*passives*) Handeln des Forschers 2) drei Ebenen des Abstraktionsniveaus der Methoden 3) unmittelbarem oder unvermitteltem Kontakt des Forschers mit dem Gegenstand. Das *Abstraktionsniveau* (Festlegung) wird untergliedert in (s. [Kle86][Ge03][Fa00][Ho97][BN06]):

Alltagstechnik: Sie gilt nicht als Methode, ist also unwissenschaftlich und umfasst das tägliche Handeln (erlebe, tue).

Qualitative Methode: Sie abstrahiert von den Alltagstechniken (geringerer Abstraktionsgrad als quantitative Methoden).

Quantitative Methode: Sie bieten sich bei heuristischen Verfahren an.

(Sozial-)wissenschaftliche Grundmethoden sind Beobachtung (Rezipieren) und Experiment (Tun). Erfahrungswissenschaftliche Methoden der Psychologie trennen darüber hinaus noch:

Beobachtung: Gemeint ist die gezielte und geplante Wahrnehmung und Erfassung eines Teilbereichs der Wirklichkeit.

Experiment: Ein Vorgang wird zur gezielten Beobachtung absichtlich und planmäßig herbeigeführt.

Test: Dies ist ein Messverfahren zur Feststellung der individuellen Ausprägung eines psychischen Merkmals.

Befragung (Interview): Sie gilt als Technik zur Erfassung von Daten mit Hilfe der Beantwortung von Fragen.

„Im Vordergrund bei den sog. quantitativen Methoden steht die Frage, wie die zu erhebenden Merkmale operationalisiert bzw. quantifiziert werden sollen.“[BN03, S.137] Übliche human- und sozialwissenschaftliche Methoden sind Zählen, Urteilen, Testen, Befragen, Beobachten und physiologische Messungen. Empirische Untersuchungen kombinieren diese Arten oft. Folgend werden *Experiment* und *Befragung* beleuchtet als zentrale Methoden des eigenen Vorhabens (Auswahlbegründung ►3.6).

2.2.1 Experiment

„Will man Hypothesen über Ursache-Wirkungs-Relationen prüfen, so liefern experimentelle Untersuchungen die stringentesten Belege für oder gegen die behauptete Kausalität [...]“[BN06, S.491] Experimente (lat. Experimentum: Versuch, Beweis, Prüfung) sind methodisch angelegte Untersuchungen, um empirisch Informationen (Daten) zu gewinnen. Kennzeichnend sind *absichtliche Auslösung* („Willkürbarkeit“), *Bedingungsvariation* (aktive Bedingungsänderung), *Wiederholbarkeit* und gute *Kontrolle*. VP sind den EG und KG möglichst zufällig zuzuweisen (Randomisierung).

Experimente sind **systematische Beobachtung** und **planmäßige Bedingungsvariation** (aktiver Eingriff, uV-Isolation und -Änderung, Prüfung der Auswirkung auf aV) und lassen sich anhand spezieller Eigenschaften klassifizieren (►Tab.3).

In Experimenten wird mind. 1 Variable systematisch variiert, Effekte registriert und die Wirkung weiterer (Stör-)Variablen ausgeschaltet. Nicht-experimentelle Forschung beobachtet *ohne* Eingriff, d.h. nur Experimente erlauben Kausalitätsprüfung (Beobachterunabhängigkeit). Wichtige Variablen sind: *uV* (wird aktiv verändert), *aV* (wird vorhergesagt) und *SV* (beeinflussen aV, stören uV-Effekt, sind zu neutralisieren oder aktiv einzubeziehen). Es wird nur vom Experiment gesprochen, sind uV und aV *unterscheidbar*, geht die uV der aV stets *voraus* und werden Daten mind. 2er VP verglichen.

Vorgehen: VP werden oft in EG und KG eingeteilt und uV in der EG als Treatment manipuliert, um aV in KG und EG zu messen. Unterschiede zwischen KG und EG werden auf uV-Effekte zurückgeführt, es dürfen keine präexperimentellen Unterschiede zwischen KG und EG vorliegen. Alle Variab-

len außer uV sind konstant und Bedingungen in KG und EG identisch zu halten, um eindeutige Kausalschlüsse zu ziehen, Alternativerklärungen auszuschließen und interne Validität zu erhöhen. Zugrunde liegt das Forschungsdesign für planmäßige Durchführung. Erhobene Daten werden meist mit statistischen Methoden ausgewertet und Erkenntnisse via Schlussfolgerung gewonnen. Experimente können ethisch oder wegen unberücksichtigter Gefahren umstritten oder unzulässig sein u.a. Menschen-/Tierversuche in der Medizin. (s. [Wi12c][Ge03][Fa00][Ps12][Zw98][Wi12a])

Tab. 3 Gängige Klassifikationen von Experimenten (nach [Zw98][Fa00][Tr09, S.65])

Klassifikation	Arten
Versuchsziel	1) Prüfexperiment: "klassisches" Experiment zur kausalen Hypothesenprüfung durch Variierung der uV in verschiedenen Stufen und Erfassung der Effekte auf die aV 2) Erkundungsexperiment (exploratives Experiment): Datensammlung zur Hypothesenbildung durch uV-Variation ohne Hypothesenformulierung vorab, Erkundung des aV-Verhaltens (nicht hypothesengeleitet) 3) Vorexperiment: „kleines“ Experiment zur Planung von Erkundungs- oder Prüfexperimenten, Erprobung oder Verbesserung jeweiliger experimenteller Bedingungen 4) Entscheidungsexperiment: hat besonderen Stellenwert in der Bewertung verschiedener Theorien, meist wird so geplant, dass das Ergebnis eine von mehreren konkurrierenden Theorien begünstigt 5) Demonstrationsexperiment: Darstellung bekannter Ergebnisse durch experimentelle Demonstration
Zahl der uV	1) einfaktorielles Experiment: Auswirkung 1 uV auf 1 oder mehrere aV 2) mehrfaktorielles Experiment: Auswirkung mehrerer uV auf 1 oder mehrere aV
Zahl der aV	1) univariates Experiment: Wirkung 1 oder mehrerer uV auf 1 aV 2) multivariates Experiment: Wirkung 1 oder mehrerer uV auf mehrere aV
Ort	1) Labor- (spezieller Versuchsraum) oder 2) Feldexperiment (Durchführung in natürlicher VP-Umgebung)
Planung	1) Echtes Experiment: aktive Manipulation mind. 1 uV, SV-Kontrolle, aktive Zufallszuweisung der VP zu Bedingungen oder ex post facto, randomisierte VP-Verteilung auf EG und KG, zumeist Laborsetting 2) Quasi-Experiment: keine SV-Kontrolle, uV (teils un-)kontrollierbar, VP-Gruppe gegeben/VP-Eigenschaft bestimmt Zuteilung in EG / KG (ggf. präexperimentelle Gruppenunterschiede), keine Randomisierung bei VP-Zuteilung zu uV-Stufen, keine/begrenzte Kausalaussage, keine / schwere Wiederholbarkeit

Merkmal (echter) Experimente ist Randomisierung, d.h. zufällige Bedingungs- und VP-Zuweisung zu Gruppen, um Scheinerklärungen wie präexperimentelles Verhalten auszuschließen. Vollständige Randomisierung ist oft unmöglich, da VP-Gruppen meist aus organisatorischen Gründen vorgegeben sind. Am Randomisierungsgrad lassen sich die Experimenttypen Feld oder Labor unterscheiden.

Feldstudie: Sie finden in der natürlichen VP-Umgebung statt. Bedingungen und VP können nicht zufällig auf Gruppen verteilt werden. Forscher finden Bedingungen vor und agieren als nicht eingreifende Beobachter u.a. „natürliches Experiment“, was kausaltheoretischen Wert begrenzt (maximal empirische Evidenz). Kritisiert werden multiple konfundierte Effekte, schwierige bis unmögliche Randomisierung der Ausgangsbedingungen d.h. uneindeutige Folgerung, schlechte SV-Kontrolle was Hypothesenprüfung erschwert sowie unzuverlässige Datenerhebung. Auch höhere bürokratische Hürden z.B. Datenschutz, Kooperationszwang z.B. mit Unternehmen, Kosten und Risiken unerwünschter Effekte z.B. Beschwerden, wirken sich nachteilig aus. Feldexperimente verfügen aber über höhere externe Validität durch Praxisbezug der naturalistischen Beobachtung, repräsentative Versuchsplanung, d.h. Generalisierbarkeit mit Gültigkeit für Situation vor Ort, Übertragbarkeit in die An-

wendungssituation und leichter gewährleistete Repräsentativität, da besonders sozialpsychologische Aspekte wirken (natürliche Versuchsbedingung, geringere Reaktivität, realistischere Effektstärken, ergo größere praktische Gültigkeit) und Versuche auch ohne Kenntnis der VP möglich sind.

Laborversuche: Sie finden in künstlicher Umgebung unter kontrollierten Bedingungen statt (uV manipulieren, aV messen). Validität (interne/externe¹⁴) ist zentrales Bewertungskriterium. Laborexperimente besitzen höchste interne Validität dank strenger Variablenisolation, (aktiver) Kontrolle von SV

◀ Natürlichkeit (Feld)	(Labor) Künstlichkeit ▶
<ul style="list-style-type: none"> • höhere externe Validität • Repräsentativität ist leichter zu gewährleisten • niedrigere Reaktivität • mehr bürokratische Hürden • mehr Kooperationszwang • Risiko von unerwünschten Auswirkungen • höhere Kosten 	<ul style="list-style-type: none"> • höhere interne Validität • leichte Veränderbarkeit der UV • leichtere Kontrolle von Störvariablen • Versuchspersonen sind den Bedingungen frei zuordenbar • leichteres Messen von Variablen • weniger Erfolgszwang • leichter wiederholbar • Selektionseffekte bei Teilnehmern

Abb. 2 Vor- und Nachteile Labor und Feld [Wp12]

und Versuchsbedingungen, insbesondere Randomisierung der Ausgangsbedingungen und gehören zu den experimentellen Studien (experimentelles Design), was die Suche eindeutiger Kausalbeziehungen zwischen isolierten Reizbedingungen und Verhaltensparametern und vergleichbare, repräsentative Ergebnisse zulässt. Die ceteris-paribus-Annahme¹⁵ ist nur unter Laborbedingungen realisierbar, sie besitzen hohen kausaltheoretischen Anspruch mit zuverlässigen Folgerungen. Zudem können VP frei und willkürlich Bedingungen zugewiesen und Variablen erhoben werden, die im Feld kaum/gar nicht möglich sind u.a. Blickverlauf. Kritisiert werden Künstlichkeit, geringe externe Validität d.h. wenig praktische Relevanz, Selektionseffekte¹⁶, höhere Reaktivität (VP wissen sich beobachtet) und Artefakte durch Erwartungen/Einstellung der VP und VL. Effektstärken können aus ethischen und praktischen Gründen beschränkt sein. Dennoch erlaubt die inhärente Künstlichkeit optimale *SV-Kontrolle* und gewährleistet am besten *Wiederholbarkeit* (definierte Versuchsbedingungen). Labor- und Feldversuche bilden zusammenfassend Extreme „lebensnaher“ bzw. „biotischer“ Versuche (►Abb.2).

Tab. 4 Validität der primären Experimenttypen (nach [BN03, S.61])

	Experimentell	Quasiexperimentell
Feld	interne Validität + externe Validität +	interne Validität – externe Validität +
Labor	interne Validität + externe Validität –	interne Validität – externe Validität –

Störeinflüsse/Gütekriterien: Stör- oder Sekundärvarianz ist die systematische Fehlerstreuung d.h. Varianz, die unkontrolliert neben der Primärvarianz der uV wirkt. Speziell sind das situationale und Personenvarianz. Hierneben tritt Fehlervarianz auf als unsystematische Fehlerstreuung z.B. Messfehler. Kontrolltechniken nach dem Max-Min-Kon-Prinzip¹⁷ sind Eliminieren, Konstanthalten, Parallelisieren und Wiederholungsmessung (►3.5) zum Neutralisieren oder Umwandeln weiterer Variablen in uV. Sozialwissenschaftliche Experimente sind oft weniger steuerbar, fordern keine strenge Reproduzierbarkeit aber Reliabilität und Validität (Trade-off ►Tab.4). Viele Forderungen sind „höchstens

¹⁴ Interne Validität meint, dass aV-Variation einzig auf die uV-Manipulation zurückführbar ist (Präzisierung der Bedingungskontrolle). Externe Validität befasst sich mit der Generalisierbarkeit bzgl. der Grundgesamtheit, Bedingungen, uV und aV.

¹⁵ „Unter der Annahme, dass alle außer den (vorher) genannten Rahmenbedingungen (Prämissen) gleich bleiben.“

¹⁶ Selektionseffekt treten auf, wenn (zufällige) VP-Gruppen sich in einer Eigenschaft signifikant von einander unterscheiden (s. [Re03]).

¹⁷ Primärvarianz (zu messende Varianz) ist zu *maximieren*, Sekundärvarianz zu *kontrollieren* und Fehlervarianz zu *minimieren*.

näherungsweise oder überhaupt nicht zu erfüllen [...]: Randomisierung, Standardisierung der Untersuchung, Kontrollen, außerdem im Prinzip auch Zufallsstichproben aus der Population.“[Fa00,S.8] Zudem dürfen Kausalschluss und Korrelation nicht gleichgesetzt werden mit dem vermuteten Variablenzusammenhang, der durch dritte (unabhängige) Variable entsteht und andere (abhängige) beeinflusst. (s. [Wi12d][Fa00][Re03][Ps12][Ge03])

2.2.2 Interview und Befragung

Befragung ist *die* Methode der Sozialwissenschaften mit ca. 90% aller Daten s. [BN03, S.237]. Gemäß der Standardisierung sind zu trennen: *freies nichtstandardisiertes Interview* ohne zwingende Vorgabe, *strukturiertes halb-/teilstandardisiertes Interview* mit Interviewleitfaden und *standardisierte Befragung* u.a. Fragebögen. Festzulegen sind zudem Kontakt (persönlich, telefonisch, schriftlich), Befragte (einzeln, Gruppe, Survey) und Befrager (einzeln, Tandem, Kommission). (s.[Hi12][BN03]) Störfaktoren entstehen u.a. durch Interviewer (persönliche Eigenschaften / Verhalten), differentielle Erreichbarkeit und Bereitschaft der Befragten, Missverständnisse zwischen Interviewer und Befragten, Antworttendenzen wie Bejahung oder soziale Erwünschtheit und die Situation selbst (Vertrautheit, Störungen, Atmosphäre). Gegenmaßnahmen sind u.a. Interviewer-Schulung, zufällige Zuordnung von Terminen und Interviewern, keine doppelte Verneinung, keine Suggestivfragen oder Verlegenheiten und keine unklaren Antwortkategorien („teilweise“).

Fragebögen: Sie gehören zur *schriftlichen Befragung*, sind höchst standardisiert, kostengünstig und werden anonym erlebt, „was sich günstig auf die Bereitschaft zu ehrlichen Angaben und gründlicher Auseinandersetzung mit der erfragten Problematik auswirken kann.“[BN03, S.237] Zur Präzisierung der Fragen helfen Strukturierungshilfen wie „MindMaps“.

Gestaltung: Eine klare Einleitung (Instruktion) sollte vorangestellt und „harte Daten“ d.h. in Zahlen ausgedrückte Ergebnisse, durch einzelne offene Fragen ergänzt werden. Die Situation sollte für alle VP gleich und angenehm sein, u.a. durch gute Organisation (Zeit-Puffer, kein Druck), Wertschätzung und inhaltliche Freiheit für VP (allein arbeiten lassen, nicht beobachten) und Anonymität. Schwachstellen der eigenen Erhebung sind in die Interpretation einzubeziehen (Offenlegung).

Fragenformate¹⁸: Dazu gehören *offene Fragen* für selbst Formuliertes ohne Vorgabe (ggf. Verbalisierungsprobleme, schwere Auswertung), *geschlossene Fragen* mit Faktfragen, Antwortvorgabe, Bezug auf Konkretes und einfache Beantwortung / Auswertung sowie *Mischformen* mit vorgegebenen Antwortmöglichkeiten und offenen Kategorien, sodass VP ihre „wahre“ Antwort geben können.

Skalierung: Sie erfolgt via Ratingskalen (numerisch, grafisch, verbal, symbolisch). „Likert-Skalen“, nach dem Sozialforscher Likert, messen das Ausmaß der Zustimmung oder Ablehnung zu vorgegebenen Gründen. Antwortvorgaben sind abgestuft zu gestalten, mit möglichst gleichen Abständen.

Formulierung: Fragen sollten einfach, kurz und genau sein, für leichtere Auswertung und weniger Informationsflut, mit geläufigen Begriffen und Fragen, deren Beantwortung nicht an Voraussetzungen

¹⁸ Offene Beantwortung (freie Gestaltung, Assoziation), halboffene Beantwortung (Einfachantwort, Mehrfachantwort, Reihenantwort, Sammelantwort), Antwortvorgaben (Alternativ-, Auswahl-, Umordnungs-, Zuordnungs-, Ergänzungsantwort). (s. [BN03])

gebunden sind, die nicht alle VP erfüllen. Verzweigungen und doppelte Verneinungen sind zu vermeiden und Fragen nicht suggestiv oder stereotyp zu formulieren, mit möglichst nur einem Inhalt. Antwortmöglichkeiten sollten erschöpfend und eindeutig sein ohne Überlappung.

Layout: Der Aufbau sollte logischen Regeln folgen und übersichtlich sein u.a. Tabellen, Schattierung. Da Aufmerksamkeit im mittleren Drittel oft am höchsten ist, sind hier die wichtigen Fragen zu platzieren, gemäß des Trichterprinzips „vom Allgemeinen zum Besonderen“ d.h. jede Fragengruppe beginnt mit unverfänglichen und steigert sich mit persönlichen/schwerer zu beantwortenden Fragen. Soziodemographische Daten befinden sich am Ende, sie werden hier besser toleriert oder direkt nach der Instruktion. Ansprechende Optik erleichtert das Erfassen der Fragen und zeigt Bemühung auf.

Beantwortungsdauer: 90 Minuten sind nicht zu überschreiten. Lange Fragebogen erhöhen zwar Mess-, „Genauigkeit“, Konzentration sinkt jedoch (VP-Alter, organisatorische Grenzen, Datenmenge).

Auswertungsmöglichkeiten: Antworten können in Zahlen übersetzt werden d.h. verrechenbare Datensätze u.a. für Häufigkeitsanalysen. Offene Fragen lassen sich nach inhaltlichen Aussagen zu Kategorien zusammenfassen. Antworthäufigkeiten sind als Tabelle oder Diagramm deskriptiv visualisierbar via Kreis-, Balken-, Linien-, Streu-, Histo-, Säulendiagramm oder „Piktogramm“, um Zusammenhänge und Unterschiede aufzuzeigen (schlussfolgernde Statistik). Mögliche Fehlerquellen umfassen unzulässige Diagrammtypen oder Darstellungswahl, die das Eigentliche nicht zeigt. Es existieren keine festen, wissenschaftlich ausgearbeiteten Prinzipien sondern nur Erfahrungswerte, daher bieten sich Pretests an für Fehlersuchläufe und um den Nutzen der Daten zu ermitteln. (s. [As01][Ki06][BN03])

2.3 Usability-Labor in der Informatik

Im Folgenden wird das UL beleuchtet, als der vorgegebene Untersuchungsort des eigenen Vorhabens.

2.3.1 Allgemein

UL werden für Usability¹⁹-Tests verwendet, um Benutzerfreundlichkeit / Benutzbarkeit, Nutzererleben- und -verhalten zu prüfen. Primär für Websites, Software-Produkte / Prototypen, Online-Angebote oder psychophysiologische Untersuchungen von Software-Anwendern z.B. Computerspieler, eingesetzt, werden sie sowohl an Hochschulen für Forschungsprojekte, Studiengänge / Module zur Benutzerfreundlichkeit als auch in der Wirtschaft (u.a. Informationsdesigner) genutzt. (s. [Li11][FA03])

2.3.2 Methoden

UL setzen sich aus VP-Computer und Beobachtungsarbeitsplatz mit zugehörigem Überwachungssystem zusammen. Ausstattung und Testmethoden variieren stark, das Minimum besteht aus einem Laptop mit Webcam in einem Raum. Günstig ist die Erweiterung der Laborumgebung auf weitere Räume zum Testen der VP und Aufteilung in Test- und Beobachtungsraum, mit Einsicht durch Einwegspiegel oder Kameras, inkl. Mikrofon, Aufzeichnungsgeräten sowie die Simulation einer natürlichen Umgebung im Nutzungskontext z.B. Büro. Dies erleichtert objektive, kontrollierte Beobachtung ohne Störung der VP durch VL oder Protokollant(en) und begünstigt natürliches VP-Verhalten. Messungen

¹⁹ Usability meint hier zielgruppengerechte Informationsgestaltung (u.a. Handbücher, Websites). Die Gestaltung orientiert sich an Nutzerzielen/-bedürfnissen, beeinflusst von Erkenntnissen über Wahrnehmung, Gedächtnis, Lernen und Emotionen.

können in Schnelltests mit 6 VP und für präzisere Messungen mit ca. 20 VP durchgeführt werden. Zur Durchführung gehören u.a. Beobachtung, Protokollierung, evtl. Logfile-Auswertung, Fragebögen, Interviews und Überprüfung eines Kriterienkatalogs. (s. [Li11][Wi11b][Pe11])

Usability-Testmethoden lassen sich unterteilen in:

Nutzerbefragung: Produkthanforderungen und Wünsche werden durch geschultes Personal ermittelt.

Nutzertests: Dies umfasst u.a. VP-Beobachtung, lautes Denken oder Blickbewegung.

Zudem sind Usability-Tests durch zwei (zeitliche) Kategorien unterscheidbar (s. [Ge03, S.21ff]):

Formative Evaluation (Gestaltender Test): Tests finden während der Produktentwicklung statt, sollen primär Benutzungsschwächen finden und sind durch ihre fehlersuchende Natur oft qualitativ.

Summative Evaluation (Kontrolltest): Dies sind oft quantitative Tests am Entwicklungsende, um das System als Ganzes zu prüfen und quantitative Daten zu erheben (Vergleich mit Zielen und Konkurrenzprodukt). Usability-Testmethoden sind qualitativ, quantitativ oder existieren als Mischform.

Eine weitere Aufteilung untergliedert die drei folgenden Kategorien:

Diagnostic Evaluation: Als klassische Usability-Tests im Labor müssen VP verschiedene Aufgaben am fertigen System (oder Prototyp) bewältigen. Möglich sind qualitative²⁰ (Thinking Aloud) und quantitative Techniken (Performance Testing²¹) mit entsprechend qualitativen und quantitativen Ergebnissen. Der Ablauf orientiert sich am quantitativen Experiment mit laborähnlicher Umgebung. Am Ende kann von VP ein Post Test Questionnaire ausgefüllt werden.

Participatory Evaluation: Dies sind meist qualitative Methoden (u.a. Interviewtechniken) mit der Möglichkeit, auch quantitative Daten zu erheben. Folglich sind Untersuchungen nicht laborgebunden und oft zu Entwicklungsbeginn (*vor* Prototyp) oder später durchzuführen (Gruppendiskussion, Interviewtechnik) mit aktivem Kontakt zwischen VP und VL, um Wünsche und Erwartungen zu ermitteln.

Subjective Evaluation: Messinstrumente werden durch Menschen ersetzt, die Schlüsse ziehen, in der Regel numerisch anhand von Fragebögen o. ä. (meist Online-Umfragen oder Post Test Fragebögen).

Hierneben existieren noch weitere Werkzeuge und Test-Methoden (s. [et07][Wi11a][Ge03]):

Interaktionsaufzeichnung: Umfasst Eingaben (Tastatur, Maus) und Aufzeichnung der Interfaceansicht.

Videobeobachtung: Bildschirmgeschehen und ggf. VP-Verhalten werden aufgezeichnet (aufwändige Methode, empfohlene Durchführung im UL mit mehreren Kameras).

Eyetracking (Blickerfassung): Die Blickbewegung der VP wird aufgezeichnet, d.h. Bewegungsrichtung und -intervalle (Aufmerksamkeitsverteilung), mit Aufzeichnungsgeräten wie mobile (Befestigung am Kopf) und externe Eyetracker (Remote-Geräte oder mit Kopffixierung).

Thinking Aloud (lautes Denken): Eindrücke werden wiedergegeben, im Labor (Ton- und Videoaufzeichnung) oder via Protokollierung der Aussagen. Eine Variation ist das *Write-along*, bei dem VP während der Lösung der Aufgaben schriftlich Notizen machen und Usability-Fragen mit auf dem Monitor angezeigt werden. Das durchgehende Kommentieren der eigenen Handlung ist aber unge-

²⁰ Qualitative Tests befassen sich mit Interaktion und dem System selbst, um gewählte Bereiche voll abzudecken. (s. [Ge03])

²¹ Performance Testing ist die Erhebung quantitativer Daten zur statistischen Auswertung (u.a. arithmetisches Mittel) mit Fokus auf Effizienz (Task Time, Completion Rate/Mean Time on Task) und Effektivität (Completion Rate, Errors, Assists).

wohnt, VP müssen öfter erinnert werden, was verunsichert und Leistungsfähigkeit negativ beeinflussen kann, die unnatürliche Testsituation besitzt nur bedingt aussagekräftige Ergebnisse.

Cognitive Walkthrough: Ein Experte löst anhand von Szenarien Aufgaben, andere notieren Probleme und Missverständnisse. Vorteile liegen hier in der raschen Durchführbarkeit und geringeren Kosten.

Moderierter Gruppentest: Eine audioaufgezeichnete Gruppe beantwortet Fragen und Aufgaben, Äußerungen werden in der Gruppe nachgefragt. Vorteile sind Einfachheit und Kostengünstigkeit.

2.3.3 Gütekriterien

Kriterien valider Testsettings ähneln stark denen der Sozialwissenschaften. Dabei ist jedoch zu beachten, dass Usability-Tests andere Ziele verfolgen: „Usability Tests sind kein Forschungsgebiet, sie dienen der Qualitätssicherung eines Produktes, welches an den Mann gebracht werden muss. Dementsprechend effektiv müssen sie sein was verwertbare Ergebnisse angeht, ansonsten sind sie den Aufwand ganz einfach nicht wert.“ [Ge03, S.32] Ähnlich quantitativen Experimenten der Sozialwissenschaften stehen *interne* und *externe Validität* in Konflikt miteinander. *Dritt- und SV* hingegen spielen beim Usability-Testing eine geringere Rolle, da Versuche meist aus Einzeltests bestehen. Die kleine VP-Anzahl erschwert jedoch das Fehler-Abfangen. Es sind identische Bedingungen für alle Tests zu gewährleisten. Für *Repräsentativität* und *Ergebnisverallgemeinerung* ist eingangs die Grundgesamtheit (Zielgruppe) genau zu definieren. Im Usability-Testing erzeugt dies Probleme, da Marketingausagen sehr allgemein formuliert sind, absolute Zahlen verlangen und absolute Usability-Ziele definieren. Quantitative Experimente können dies nicht leisten, zudem fehlt die KG. Störeinflüsse wie *Demand Characteristics* treten zurück, da VP das Versuchsziel kennen. Die Situation wirkt jedoch genauso unnatürlich. Um *Forced Exposure* zu vermeiden, ist einleitend zu erklären, dass Aufgaben oder der gesamte Test jederzeit angebrochen werden können (Details zu diesen und weiteren SV ► 3.5.4).

2.3.4 Vor- und Nachteile

Vorteile: Die Laborausstattung mit Aufzeichnungssoftware und -hardware liegt fertig vor, sodass große Stichproben untersucht und exakte Videoaufzeichnung gut durchgeführt werden kann. UL eignen sich hervorragend zur Erhebung quantitativer Daten. Zusätzlich unterstützt die Künstlichkeit im UL die Kontrolle von Störeinflüssen und begünstigt die Ziehung eindeutiger Kausalschlüsse. Dies erhöht interne Validität und fördert Wiederholbarkeit und Generalisierung. (s. [Wi11b][Ge03])

Nachteile: Die Künstlichkeit der Versuchsanordnung unterbindet andererseits in vielen Fällen natürliche Situationen (z.B. Unterbrechung durch Kollegen oder Anrufe), die Büroeinrichtung ist genormt und auf die Testsituation angepasst. Zudem bestehen VP aus Freiwilligen, die sich oft allgemein von der Gesamtpopulation unterscheiden. Aus Kosten-Nutzen-Gründen sind oft nur kleine VP-Zahlen untersuchbar, d.h. Fehler in Ergebnissen lassen sich schlechter abfangen.

TEIL III

VERSUCHSPLANUNG

3. FORSCHUNGSDESIGN

Im Hauptteil erfolgt die Entwicklung des Forschungsdesigns. Auf Basis der zentralen Fragestellung und Forschungsstands sollen Hypothese(n) abgeleitet und der Versuch im UL konzipiert werden.

3.1 Forschungsstand

Eingangs soll der eigene Frageansatz in den übergreifenden Zusammenhang eingebettet und nach Maßgabe vorhandener Ergebnisse und Theorien eingegrenzt werden. Dies umfasst interessante Untersuchungen zum Themenbereich, Ergebnisse, zentrale Schlussfolgerungen und eingesetzte Methoden, um nachfolgend die eigene Fragestellung einzuordnen, ans Publierte anzuschließen (Anknüpfungspunkt, Erweiterungsmöglichkeit), geeignete Hypothesen abzuleiten und ein sinnvolles Design zu erarbeiten. Im Vordergrund stehen Kommunikation/-smittel und ihre Auswirkungen auf den Arbeitsprozess, mit Augenmerk auf praktische Relevanz, d.h. Nutzen / Folgen für die Wirtschaft.

„Interruption is a major workplace concern today, especially for people engaged in information work, and computer technologies are widely viewed as exacerbating the problem.“[GD07, S.23]

Unterbrechung: Sie ist das „Pausieren einer Tätigkeit, um sich möglicherweise etwas Anderem zuzuwenden, mit der wahrscheinlichen Absicht, die pausierte Tätigkeit zu einem späteren Zeitpunkt wieder aufzunehmen.“ und trennt *innere* (selbst geplant u.a gleiche Folgen wie Multitasking) und *externe Unterbrechung* (von außen, oft ungeplant u.a. Kollegen). Letztere verursachen größere Folgen (Hauptkostentreiber), einige Studien [Gö08] berichten von externen Unterbrechungen mind. 4mal pro Stunde. Ferner lassen sich Unterbrechungen aufteilen nach ihren *unterschiedlichen* Auswirkungen. Negative Effekte liegen im stärkeren Frustrations- und Stressempfinden, hauptsächlich wirken sich aber *zeitliche* Aspekte auf Unterbrechungskosten aus. Die *Zeit zur Wiederaufnahme* unterbrochener Tätigkeiten steigt mit *Länge der Unterbrechungszeit*. Auch erhöhen sich *gefühlte* Arbeitsbelastung und *Zeitdruck*. Im Gegensatz dazu werden Unterbrechungen als hilfreich empfunden (positive Effekte, niedrigere empfundene Arbeitsbelastung), je stärker unterbrochene Tätigkeit und Tätigkeit der Unterbrechung (inhaltlich) im Zusammenhang stehen. Hier verkürzt sich auch die *Wiederaufnahmezeit* der ursprünglichen Tätigkeit [Gö08]. Aus Studien geht hervor [GD07][Sc11], dass Mitarbeiter ca. 11 Minuten unterbrechungslos an einer Aufgabe verbringen, bevor sie unterbrochen werden oder zu einer neuen Aufgabe wechseln. 57% der Unterbrechungen hatten keinen Bezug zur anfallenden Aufgabe. Menschen benötigen ca. 8 Minuten, um volle Konzentration wiederzuerlangen [Sc11]. Allgemein werden Unterbrechungen als signifikante Behinderung der Produktivität angesehen (längere Bearbeitungszeit, Fehleranfälligkeit²²), die Denkprozesse/Abläufe stören. Zwischenmenschliche Kommunikation gilt als eine Hauptursache für Arbeitsunterbrechung (Telefon, Gespräche) [GD07].

Wilson-Studie: Ein In-House Experiment unter Aufsicht von Dr. Wilson beschäftigte sich 2005 direkt mit Auswirkungen der „always-on“ Technologie. 8 VP (4 männlich, 4 weiblich) wurden 2mal untersucht – sie sollten IQ-Tests einmal unter ruhigen und einmal unter störenden Bedingungen (*klin-*

²² „People have cognitive limitations that make them sensitive to interruption. These limitations can cause people to make serious mistakes when they are interrupted.“[Mc99, S.1]

gelnde Handys, eintreffende Emails) lösen. VP unter Ablenkung erzielten durchschnittlich 10 IQ-Punkte weniger. Neben Verwendung paralleler Ausgaben eines matrizenbasierten IQ-Tests, wurden im Experiment Messungen von Hautleitwiderstand (Stressmessung), Herzfrequenz und Blutdruck sowie Selbsteinschätzung des Stresses durch VP vorgenommen. Ergebnisse der Studie zeigen deutlich, dass technologische Ablenkung die Leistung in IQ-Tests minderte (Mittelwert 143.38 unter ruhigen sank auf 132.75 unter störenden Bedingungen), sich Ablenkung stärker auf männliche (von 145.50 auf 127) als auf weibliche VP (von 141.25 auf 138.50) auswirkte und Selbsteinschätzung des Stresslevels klar unter störenden Bedingungen erhöhte (auf 0-10 Skala „erlebter Stress während des Tests“ erhöhte sich Stress für männliche VP von 2.75 auf 5.5 und für weibliche von 4.75 auf 6.75). Der Stressindikator Hautleitwiderstand stieg leicht an (Männer von 23.00 auf 25.75, Frauen von 16.25 auf 18.25). Herzfrequenz und Blutdruck zeigten keine konsistente Beziehung [Wi10].

In anderen Studien [HA03][CHW04] werden explizit **Unterbrechungsart, Multitasking und Kosten** untersucht, die wirtschaftliche Relevanz und Notwendigkeit weiterer Studien zum Thema hervorgehoben, um Kostenschätzungen vorzunehmen, Schwachstellen zu ermitteln und geeignete Gegenstrategien zu entwickeln [CHW04]. Häufige Wechsel / Unterbrechungen verschiedener Aufgaben gelten als primärer Kostenfaktor. Hinzu kommt „Multitasking“ (Wechsel zwischen nebenläufigen Arbeiten), mit hohem Mehraufwand an Zeit und Aufmerksamkeit (Termin-/Aufgabenkoordination, ständige Zeitplanaktualisierung, korrekte Priorisierung, Termineinhaltung) und erhöhter Komplexität (Folge: verzögerte Aufgabenerfüllung, Auswirkungen auf Zeitplan, Stressempfinden und Produktivität) [Gö08].

Nutzen: Untersuchungen zu *Multitasking (Task Switching) und Unterbrechungen* [CHW04] zielen u.a. darauf ab, Kosten, Dichte und Arten von Unterbrechung zu charakterisieren, um Richtlinien für sinnvolle Benutzerschnittstellen aufzustellen, die Tätigkeitswiederaufnahme nach Unterbrechung unterstützen, da Task-Switching und Unterbrechungen in komplexen langen Projekten bisher wenig von Software unterstützt wird [CHW04, S.1]. Es werden Modelle und Methoden entwickelt, um Unterbrechungskosten vorherzusagen, Aufmerksamkeitsfokus und Workload des Benutzers im Benutzerinterface zu finden und Multitasking unterstützende Software-Werkzeuge zu entwerfen [HA03][CHW04]. Grundlagenforschung zum *Umgang mit Kommunikation und Unterbrechungen bei der Arbeit* nutzt speziell der praktischen Entwicklung geeigneter Software und Projektmanagement-Werkzeuge, um Produktivität und Mitarbeiterbefinden zu erhöhen. In den letzten Jahren wuchs das Interesse an der Methodenentwicklung [Ha10] für Computersysteme mit Verständnis über Benutzeraufmerksamkeit, Arbeitspensums (MWL²³) und „Unterbrechbarkeit“ (ökonomische, kosteneffiziente Unterbrechungspunkte im Arbeitsfluss d.h. geplante Unterbrechungen). Basis ist u.a. psychologische Forschung zu Themen wie Unterbrechung und geteilte Aufmerksamkeit [HA03]. Weitere Maßnahmen umfassen die Entwicklung von Echtzeitüberwachung, Scheduling von Unterbrechungen und Entwicklung von

²³ Mental Workload (mentale Beanspruchung) ist der Beanspruchungsgrad, der von Aufgabenkomplexität (Anforderung) und eigener Leistungsvoraussetzung abhängt (Fähigkeit, Fertigkeit, motivationale Einstellung), d.h. „our cognitive activities share common mental resources, mental resources are limited in their capacities, time-sharing mental activities can, at times, not have access to enough mental resource, which can cause a sub-optimal level of cognitive performance or behavior.“ [To09,S.8]

„Physiologically Attentive User Interface (PAUI) that measures mental load using Heart Rate Variability (HRV) signals, and motor activity using electroencephalogram (EEG) analysis.“[CV04, S.1]

Zudem wurden Nutzerstudien durchgeführt zu Effekten von **Unterbrechungsarten in verschiedenen Situationen** und Tests zur Arbeitsbelastung und Verfügbarkeit von Menschen in Büroumgebungen [HA03]. Menschliche Aufmerksamkeit wird als endliche Ressource definiert: “Cognitive demands and limitations will ebb and flow in situations of divided attention, due to an interruption of a primary task, or engaging in dual- (or multi)-tasking [...] in the context of an interruption, attention switches from one task to another, whether the interruption is relevant or a distraction.” [Ha10, S.301] Mentale Überlastung durch zu viele Quellen können Leistungsfähigkeit ändern (ggf. mindern), da Ressourcen geteilt werden müssen und nicht genügend zur Verfügung stehen: “MWL, cognitive resource scarcity, and a tendency to make cognitive errors due to limited available cognitive resources are three aspects of the mental status. They all link to or lead to poor cognitive task performance.”[To09, S.8]

Andere Studien prüfen **Unterbrechungen zu verschiedenen Zeiten** und Auswirkungen auf „performance“, „emotional state“ sowie „social attribution“ [AB04]. Ansatz ist die Auffassung, dass Nutzer für „Unterbrechungsüberlastung“ (*interruption overload*) anfällig sind und Anwendungen wie Email, Instant Messaging, Web-Assistenten „do not consider the impact an interruption has on a user, and even the most well-meaning application has the potential to cause interruption overload. Poorly timed interruptions can adversely affect task performance [...] and emotional state...”[AB04, S.1] Messungen erfolgen mittels Event-Wahrnehmung, um anhand prognostizierter kognitiver Belastung des Anwenders Vorhersagen über bessere und schlechtere Unterbrechungszeitpunkte zu machen. Ergebnisse belegen u.a. dass vorhergesagte „beste“ Unterbrechungszeitpunkte weniger Ärger, Frust, Zeitdruck verursachten und weniger geistige Anstrengung forderten, d.h. *verschiedene* Unterbrechungszeitpunkte bewirken andere Effekte, woraufhin die Idee eines „attention managers“ entstand [AB04].

Weitere Untersuchungen zu **Unterbrechungen durch Kommunikation** (*Emails, Telefon, Personen*) erheben die Dauer der Unterbrechung und des Zurückfindens zum alten Arbeitsstand²⁴. Hochrechnungen belegen, dass diese Unterbrechungen einen nicht zu vernachlässigenden Anteil der Arbeitszeit fordern und entstehende Kosten weiterer Evaluierung, Grundlagenforschung und Gegenmaßnahmen bedürfen: “So far, there has been no reported empirical research into how long it takes to recover from an email interrupt.“ [JDW02, S.2-3] Studien speziell zu (textbasiertem) Instant Messaging schlussfolgerten [GD07], dieses trage dazu bei, Unterbrechungen effektiver zu bewältigen (u.a. schneller Bezug aufgabenrelevanter Information, Verhandeln von Gesprächsverfügbarkeit). Das hohe Potenzial digitaler Kommunikationsmittel, aus Unterbrechung eine Störung zu machen, wird primär auf 3 Eigenschaften untergliedert: Nachrichten-Notifikation (sofortige Bildschirmanzeige), Verfügbarkeitsan-

²⁴ „Research carried out by Solingen into communication interrupts showed 15-20 percent of an employee’s effort is spent dealing with interrupts and in real terms 15-20 minutes per interrupt [...] Personal visits and telephone calls caused 90 percent of all interrupts and email caused the rest [...] the average developer receives three to five interrupts per day [...] DeMarco reported that the recovery time after a phone call interruption is at least 15 minutes [...] However, DeMarco’s research was carried out using software developers as the subjects [...] they are likely to require extra time to recover from an interrupt compared to other job roles [...] because the nature of a developer’s work is likely to be quite complex...” [JDW02, S.2-3]

zeige des Nutzers (wird selten erneuert, lädt zur unkontrollierten Unterbrechung ein) und Möglichkeit/Ermutigung polychronischer Kommunikation (multiple simultane Konversationen). Unterbrechung muss nicht immer nachteilig sein und kann eine wertvolle Arbeitskomponente bilden (zeitnahe Lieferung kritischer Informationen für informierte Entscheidungen, rechtzeitiges Intervenieren), Interaktionsmöglichkeit für Informationsaustausch, Koordination, lange Zusammenarbeit bei der Verbindungen uneingeschränkt offen gehalten werden können, ohne ständige Konversation zur erfordern (schnelle Information mit minimaler Unterbrechung, Kontrolle über Kommunikationszeit) [Gah07].

3.2 Thema und Fragestellung

„Eine gute wissenschaftliche Theorie ist durch eine präzise Terminologie, einen logischen konsistenten Informationsgehalt (Widerspruchslosigkeit), durch eine möglichst breite inhaltliche Tragweite sowie durch eine begrenzte Anzahl von Annahmen (Sparsamkeit) gekennzeichnet.“ [BN03, S.105] Unter Berücksichtigung dieser Aspekte sind Fragestellung, Hypothesen und Erwartungen zu entwickeln. Es gilt zunächst das Problem vom Allgemeinen in einen konkreten messbaren Frageansatz zu überführen (Präzisierung). Daraus sind empirisch untersuchbare Hypothesen abzuleiten (► 3.3) und unklare mehrdeutige Begriffe für den Kontext genau zu definieren (begriffliche Klarheit).

Problembereich / Vermutungen: Der breite Frageansatz thematisiert den Einfluss von Kommunikation bei der Arbeit und soll via experimenteller UL-Untersuchung geprüft werden. Heute existiert eine Vielzahl an Kommunikationsmitteln, die in Berufs- und Privatalltag während der Arbeit allgegenwärtig sind. Viele Menschen sind ständig erreichbar, Informationsaustausch findet zeitversetzt und in Echtzeit statt. Basierend auf den Forschungsansätzen ► 3.1 lassen sich konkrete Fragen ableiten: Welche Auswirkungen verursachen Kommunikation(-smittel) während der Arbeit? Hemmen und beeinträchtigen sie Leistung und Effizienz (Zeit) durch Ablenken und Störung der Konzentration (Kosten) oder unterstützen / fördern sie ein dynamischeres Arbeiten und schnellere effektivere Arbeitsabläufe (wertvolle Arbeitskomponente)? Es wird vermutet, dass Kommunikation/-smittel während der Arbeit zwangsläufig (physische und/oder gefühlte) Ablenkung erzeugen. Zusätzlich werden Auswirkungen auf die Arbeitsleistung durch Kommunikation und gängige Kommunikationsmittel prognostiziert. Es kann innere (selbst initiiert, durch eigene Motivation) und äußere Unterbrechung (von außen erzeugt/ fremd initiiert, ggf. ungewollt) differenziert werden. Angenommene Auswirkungen beinhalten positive und negative Effekte auf Leistung: Förderung der Leistung²⁵ (Zusammenarbeit, Hilfe u.a.) oder ihr Beeinträchtigung²⁶ (Störung u.a.). Die Richtung der Auswirkung wird nicht festgelegt, da ohne weitere Spezifikation und Isolation der Vielzahl an Kommunikation/-smittel (Medienvielfalt, aktiv/

²⁵ Populäres Beispiel ist die Fernsehshow „Wer wird Millionär“, in der Kandidaten u.a. durch Telefon-Joker oder Publikum unterstützt werden. Zudem wird bei (aktiver/beidseitiger/selbst initiiertes) Kommunikation ein unterstützender Effekt vermutet durch Teamarbeit, höhere Motivation (Effizienz, soziale Komponente), dynamische, schnelle, effektive Abläufe (► 3.1).

²⁶ z.B. Überlastung durch häufige Unterbrechung, unterbrochene Konzentration, Gedankengänge sind schwer zu Ende formulierbar und ggf. neu zu beginnen (Rückkehr zum Ausgangspunkt/ Zustand vor Unterbrechung), was Zeit und Energie kostet, ggf. Zeitdruck erhöht, Stimmung senkt (Gereiztheit, Ungeduld, Frust) und Ergebnisqualität senkt. Probleme können ggf. nicht intensiv durchdrungen werden und Arbeitsprozesse sind fragmentiert, was in oberflächlicher Bearbeitung resultieren und sich negativ auf Arbeit und Person auswirken kann (Frust, Unzufriedenheit, sinkende Motivation).

passiv, ein-/beidseitig, fremd/selbst initiiert, Inhalt, Intervall) beide Effekte denkbar sind. Zuerst soll aber überprüft werden, ob sich Arbeitsleistung *überhaupt* ändert, bevor die Arten isoliert prüfbar sind.

Daraus ergibt sich die präzisierte Fragestellung unter Berücksichtigung des praktischen Nutzens: Erzeugen Kommunikation und gängige Kommunikationsmittel in einer Arbeitssituation zwangsläufig Ablenkung? Und inwiefern wirken sie sich auf die Arbeitsleistung aus?

Definition der Termini: *Arbeitssituation* meint hier einen Zustand, der 1) Konzentration fordert und kennzeichnet sich durch 2) Abarbeitung vorgegebener Aufgaben als konzentrierter zielgerichteter Arbeitsprozess unter 3) Begrenzung auf *Arbeit am Computer* mit Schwerpunkt digitale Kommunikation, Anlehnung an Büro, Messungen im UL u.a. Eyetracking und in der 4) Leistung gefordert ist d.h. „natürlicher“ Druck/Stress ohne *außerordentlichen* Leistungsdruck (keine Prüfungssituation oder Auswertung im VP-Beisein oder „Bestrafung“ negativer Leistung), die Situation ist keine Ausnahme (unnatürliche Angst/Stress) sondern möglichst neutrale Arbeitssituation mit Basis *Arbeitssituation im Alltag*. Aufgrund technischer Grenzen d.h. Einzelarbeitsplatz im UL der FHB wird zudem 5) Einzelarbeit gefordert. *Kommunikation und gängige Kommunikationsmittel* werden in ►3.8 definiert. Operationalisierung d.h. Messbarmachung von *Ablenkung* und *Arbeitsleistung* erfolgt in ►3.5.3.

3.3 (Sach-)Hypothesen und Erwartungen

Nachfolgend werden sachliche Hypothesen abgeleitet, ausgehend von theoretischen Überlegungen, empirischen Beobachtungen und Forschungsergebnissen und die Messbarmachung überdacht.

Schlüsselannahmen und Indikatoren: Es werden zwei Sätze (Behauptungen) aus ►3.2 abgeleitet.

1) Kommunikation und gängige Kommunikationsmittel erzeugen zwangsläufig Ablenkung in einer Arbeitssituation.

Das Konstrukt *Ablenkung* kann durch physiologische Indikatoren und (Selbst-)beobachtung messbar gemacht werden:

- *Physiologischer Ablenkungszustand:* durch physiologische Indikatoren mithilfe der Techniken des UL und BS
- *Subjektiver Ablenkungszustand:* durch Selbsteinschätzung (Fragebogen)

2) Kommunikation und gängige Kommunikationsmittel wirken sich auf die Arbeitsleistung aus.

Eine angenommene Folge ist die Änderung der Arbeitsleistung (*sprich* Minderung oder Erhöhung). Diese Veränderung der *Arbeitsleistung* kann anhand der Indikatoren *Qualität* der Arbeit und *Arbeitsmenge* erhoben werden (Vergleich KG mit EG).

- *Qualität:* Fehleranzahl bei bearbeiteten Aufgaben
- *Arbeitsmenge:* Anzahl gelöster Aufgaben in vorgegebener Zeit
- *IQ:* Es bieten sich unterstützende IQ-Messungen an, um durch Vergleiche von KG und EG mit evtl. veränderten IQ-Werten Rückschlüsse auf Leistungsänderungen zu ziehen und Ergebnisse zu stärken (s. Wilson-Studie ►3.1)

Daraus lässt sich eine (universelle ungerichtete) Kausalhypothese als Forschungshypothese ableiten:

Haupthypothese (HH): Durch Kommunikation und gängige Kommunikationsmittel wird in einer Arbeitssituation Ablenkung erzeugt und eine Veränderung der Arbeitsleistung bewirkt.

Es ergeben sich die beiden (universellen kausalen) Teilhypothesen, die es im Versuch zu prüfen gilt²⁷:

Teilhypothese 1 (TH1): Kommunikation und gängige Kommunikationsmittel erzeugen in einer Arbeitssituation Ablenkung. („Wenn Kommunikation/-smittel angewendet werden, dann erzeugt dies Ablenkung.“)

Teilhypothese 2 (TH2): Kommunikation und gängige Kommunikationsmittel verändern die Arbeitsleistung. („Wenn Kommunikation/-smittel angewendet werden, dann verändert dies die Arbeitsleistung.“)

²⁷ Bei Kausalhypothesen sind Wenn-(Ursache) und Dann-Teil (Wirkung) sprachlich und inhaltlich nicht austauschbar (s.[BN03]).

Daraus leiten sich Vorgehen (Hypothesenprüfung und Datenerhebung im UL) und Untersuchungsziele (Verifikation²⁸ der Haupthypothese durch Verifikation der 2 Teilhypothesen) ab.

Erwartungen: Vorab sind denkbare Ausgänge (und Ursachen) zu erfassen, um später zu entscheiden, welche Ergebnisse eindeutig für und welche gegen die inhaltlichen Hypothesen sprechen (► Tab. 5).

Tab. 5 Mögliche Ergebnisse der Untersuchung (Erwartungen)

Verifizierung	Falsifizierung	Kein Ergebnis
TH1: Ablenkung findet statt	TH1: Keine Ablenkung	TH1: Ablenkung wird nicht eindeutig gemessen - signifikante Ergebnisschwankungen
TH2: Veränderte Arbeitsmenge oder Qualität oder beides (IQ-Werte zur Ergebnisfestigung)	TH2: Keine Änderung der Arbeitsmenge und Qualität (IQ-Werte zur Ergebnisfestigung)	- Ergebnisverzerrung durch SV bei Durchführung - Ungeeignete Methoden, Werkzeuge, Variablen, Fragebögen, Planung - Fehlerhafte Datenauswertung
HH: Ablenkung und veränderte Arbeitsleistung finden statt	HH: Keine Ablenkung und Leistungsänderung - Ablenkung aber keine veränderte Arbeitsleistung (Arbeitsmenge und Qualität) - keine Ablenkung aber Arbeitsleistung ändert sich (Arbeitsmenge oder Qualität)	TH2: Keine aussagekräftigen Ergebnisse über Arbeitsleistung (Qualität und Arbeitsmenge) - signifikante Ergebnisschwankungen (uneindeutig) - Ergebnisverzerrung durch SV bei Durchführung - Fehlerhafte Datenauswertung - Ungeeignete VP-Aufgaben und / oder Planung - Ungeeignete aV (Qualität, Arbeitsmenge, ggf. IQ)
		HH: uneindeutige, nicht aussagekräftige Ergebnisse der TH1 und TH2

Falsifikation: Kann HH sich nicht bewähren z.B. aufgrund falscher Hypothesen, Zusatzannahmen oder Mischform, gilt zu überdenken, inwieweit TH1 und TH2 durch UND-Komponente²⁹ in Zusammenhang miteinander gebracht werden sollten oder ob sich getrennte Hypothesenprüfung lohnt.

Zusammenfassung: Die Hypothesen beziehen sich auf reale, empirisch untersuchbare Sachverhalte in Formalstruktur eines Konditionalsatzes, sind empirisch falsifizierbar und allgemeingültig, d.h. erfüllen Kriterien *wissenschaftlicher Hypothesen*. Im Vordergrund steht in der Arbeit die TH2 ►3.5.

Aussagebereich: universell (Geltungsbereich ist zu groß, sodass Hypothese nicht verifizierbar aber falsifizierbar ist)

Grad der Konkretisierung: Forschungshypothese (inhaltliche Hypothese) d.h. abgeleitete Vermutung (mit Konstrukten)

Untersuchungsziel: Veränderungshypothese (Vergleich EG und KG, Treatment hat verändernde Wirkung auf aV)

Gerichtetheit: ungerichtet (Veränderungsrichtung der Arbeitsleistung wird nicht vorgegeben)

Spezifität: unspezifisch (keine Angabe über Größe der erwarteten Leistungsänderung)

3.4 Vorgaben und Ausgangssituation

Neben dem übergeordneten Thema sind weitere Bedingungen für das Forschungsvorhaben fix vorgegeben und sollen kurz benannt werden, um diese später im Planungsprozess zu berücksichtigen und sinnvoll in den Versuchsaufbau zu integrieren. Insgesamt unterliegt die Untersuchung im Rahmen dieser Arbeit einigen zeitlichen, finanziellen, technischen, räumlichen und personellen Grenzen. Das

²⁸ Finales Ziel ist Hypothesenprüfung. Diese Arbeit realisiert nur ein Teilziel (Entwicklung des Forschungsdesigns, probeweise nicht-repräsentative Durchführung). Hypothese(n) sind nicht verifiziert, Ergebnisse zeigen maximal Tendenzen auf.

²⁹ Und-Komponenten im Dann-Satz erhöhen den Informationsgehalt (d.h. Zahl potenziell falsifizierender Ereignisse wächst) (s.[BN06]).

Vorgehen darf daher nicht zu rigide gestaltet sein, um Versuche nicht an kleinen Pannen scheitern zu lassen (Flexibilitätsanforderung) (s. [BN03]). So ist bei Entwicklung des Forschungsdesigns der technische und organisatorische Rahmen zu beachten: eine Vorgabe ist Umsetzung im UL, was den Versuchsort auf das *Labor* festlegt und die Methode *Experiment* nahe legt. Daher soll der Ansatz „*experimentelle Studie*“ verfolgt werden, d.h. Untersuchung der Beziehung zwischen uV und aV. Zudem können nur Geräte und Techniken einbezogen werden, die im UL der FHB vorrätig sind. Nach den begrenzten Ressourcen (Anzahl, Auswahl) richtet sich auch die Festlegung der aV (physiologische Indikatoren für Ablenkung ►3.5.3). Zudem begrenzt der technische Rahmen die Versuchszeit für die zwei Durchgänge (D1, D2) pro VP auf je ca.10 Minuten. *Technische* Planung, Geräteanwendung, Datenauswertung und -präsentation erfolgen außerhalb dieser Arbeit, sodass hier keine Aussagen zu TH1 möglich sind und sich die Auswertung auf TH2 beschränkt. Aufgrund zeitlicher und organisatorischer Grenzen ist zudem nur eine limitierte Anzahl an Geräten, Szenarien und VP prüfbar.

3.5 Experimentalvariablen

Für empirisch-quantitative Forschung ist Operationalisierung die Basis der Variablenmessung. Nach analytischer Definition zentraler Begriffe (Bedeutungsanalyse) folgt die *operationale* Definition: sie „standardisiert einen Begriff durch die Angabe der Operationen, die zur Erfassung des durch den Begriff bezeichneten Sachverhaltes notwendig sind, oder durch Angabe von meßbaren Ereignissen, die das Vorliegen dieses Sachverhaltes anzeigen (Indikatoren).“[BN03, S.67] Folgend wird ein kurzer Abriss über wichtige Variablentypen gegeben, um diese in ►3.5.3 auf die Hypothesen anzuwenden.

3.5.1 Definition

Variablen sind Größen in wissenschaftlichen Hypothesen und Versuchsplänen (s. [el09]). Sie sind 1) qualitative oder quantitative veränderliche Größen (qualitative Variablen mit 2 Ausprägungen, quantitative Variablen mit mehreren), 2) beobachtbare (konkret/manifest) oder indirekt beobachtbare (abstrakt/latent) Merkmale („Konstrukte“), 3) bilden im Experiment wesentliche Bestandteile von Hypothesen, 4) können bei der Planung abstrakte Platzhalter sein, die in konkrete Kontexte übersetzbar sind, 5) besitzen mind. 2 Abstufungen, von denen jeweils nur eine realisierbar ist wobei 6) viele Abstufungen auf wenige reduzierbar sind, z.B. Alter auf „unter 20“/„über 20“, was aber Informationsverlust nach sich zieht. Bei Reduktion werden stetige in diskrete Variablen umgewandelt.

3.5.2 Typen

Wichtige Variablentypen sind in ►Tab. 6 aufgeführt (s. [BN03][BN06][el09][Zw98][Ge03][Fa00]):

Tab. 6 Variablentypen, gruppiert nach Stellenwert, Merkmalsausprägung und empirischer Zugänglichkeit

Stellenwert für die Untersuchung	Art der Merkmalsausprägung bzw. Skalenniveaus	Empirische Zugänglichkeit
unabhängige und abhängige Variable, Störvariable, Kovariable (Kontrollvariable), Moderator- und Mediatorvariable	diskrete und stetige Variable, dichotome und polytome Variable	manifeste und latente Variable

Die Abgrenzung von uV und aV gilt als wichtigstes Gruppierungskriterium sozialwissenschaftlicher Variablen. Änderungen der aV werden mit dem Einfluss der uV erklärt.

Unabhängige Variable (Treatment, Behandlung, Bedingung, Faktor): Sie ist die experimentelle, zu manipulierende Bedingung. Ihre Ausprägung wird durch Manipulation (z.B. Dosierung) direkt oder

Selektion (z.B. Alter) indirekt festgelegt. Mehrere uV zugleich sind möglich (mehrfaktorieller Plan). Die uV muss in der Hypothese variierbar sein und ist in *Reiz- und Organismusvariable* unterteilbar.

- *Reizvariable*: Sie umfasst alle Umweltreize, die auf einen Organismus einwirken können (z.B. Temperatur, Lichtverhältnisse) und werden als uV *manipuliert* und als Bedingung eingeführt.

- *Organismusvariable*: Dies betrifft Eigenschaften von Organismen und wird als uV *selektiert*, d.h. Selektion/Sortierung statt Manipulation, z.B. Geschlecht. Sie bedingt im Versuchsplan, dass keine Erklärung konstruiert wird. Sortiervariablen können aber mit Wirkvariablen konfundiert sein.

Abhängige Variable (Wirk-/Zielvariable, Kriterium): Indem sie im Experiment gemessen wird, zeigt sich in ihr der Effekt der Maßnahme (uV) auf die aV-Ausprägung. Die aV muss operationalisiert werden, um messbar zu sein. Mehrere aV zugleich sind möglich (multivariates Design).

Störvariable (Drittvariable): SV beeinflussen aV zusätzlich und können Eindeutigkeit von Schlussfolgerungen behindern (alternative/konkurrierende Erklärungen, interne Validität sinkt). Sie gehören nicht zur Hypothese, d.h. alle Variablen *neben* uV sind potentiell SV. Drittvariablen sind oft auf Fehler in der Hypothesenaufstellung zurückzuführen (Verwechslung vermuteter Kausalität mit Korrelation). *Konfundierung* entsteht durch Mischen von Störfaktor und Stimulus und verzerrt Ergebnisse. SV sind fast nie auszuschließen und sollten zumindest erkannt, *neutralisiert* oder *aktiv* (als uV) integriert werden. *Ausschluss* erhöht interne Validität, *Integration* externe, oft bieten sich Kompromisse an.

Kontrollvariable (Kovariable): SV können in uV umgewandelt werden, um bekannte SV zu kontrollieren, d.h. potentielle SV werden gemessen und herausgerechnet (Ergebnisbereinigung). Die SV wird zur Kontrollvariable (Dritt-/ Zusatzvariable) und der Versuchsplan mehrfaktoriell. Sind Störeinflüsse nicht quantifizierbar, sind sie (un)kontrollierte SV. Um mit SV umzugehen, gibt es *Kontrollstrategien*.

Moderatorvariable: Sie ist eine Kovariable und beeinflusst die uV-Wirkung auf die aV (z.B. Straßenlärm). Ihr Einfluss muss durch zusätzliche Erhebung berücksichtigt werden.

Mediatorvariable: Sie ist eine Kovariable und wirkt nicht direkt auf die aV sondern über eine dritte Variable, z.B. löst statt des Klingelns schon die *Erwartung* die Speichelsekretion beim Pawlowschen Hund aus. Ein Wegfall dieser Variable würde den Zusammenhang zwischen uV und aV aufheben.

Kontrollstrategien: Zur SV-Kontrolle (Ausschaltung der Wirkung) existieren die Techniken Randomisierung, Parallelisierung (Matching), Konstanthalten, Elimination (s. [Fa00][BN06]).

- *Randomisierung*: Sie ist *die* Kontrolltechnik und berücksichtigt als einzige (theoretisch) *alle* Fehlerfaktoren simultan. SV müssen vorab nicht bekannt sein. Randomisierung ist korrekt durchzuführen (mit Zufallszahlen) und benötigt ausreichend große N (Gesetz der großen Zahlen).

- *Parallelisierung*: Es wird u.a. eine VP-Rangordnung bzgl. bestimmter Merkmale festgelegt, diese in Blöcke zerlegt und VP zufällig den uV-Stufen zugeordnet. Sie senkt oft die externe Validität³⁰.

- *Konstanthalten* (von SV in allen Durchführungen): Bei personengebundenen und untersuchungsbedingten SV wird ihre Variabilität, bei quantitativen Variablen der Wertebereich eingeschränkt (z.B. Alterseinschränkung), bei qualitativen erfolgt Ausschluss ganzer Kategorien (z.B. Geschlecht). Kon-

³⁰ Parallelisierung erfolgt oft bzgl. individueller Merkmale (u.a. Intelligenz), Haushaltsmerkmalen (u.a. Wohngegend), Organisationsmerkmalen (u.a. Anzahl der Abteilungen) oder kommunaler Merkmale (u.a. Bewohneranzahl) (s. [BN03]).

stanthaltung wirkt vor allem bei untersuchungsbedingten SV bzw. situativen Bedingungen (Tageszeit, Instruktionen etc.) die nicht zur uV gehören. Sie verringert jedoch externe Validität.

- *Elimination* (Ausschluss potentieller SV): Sie eignet sich für untersuchungsbedingte SV (z.B. Standardisierung, Abschirmung, Automatisierung), erleichtert Erkennung der Effektstärke der uV und Versuchsreplikation, senkt aber externe Validität. Personengebundene SV sind selten eliminierbar.

- *EG und KG*: Um Effekte von VP-Merkmalen auszuschalten, werden Gruppen per 1) Randomisierung gebildet, um VP-Unterschiede gleichstark zu verteilen und systematische Ergebnisverzerrung zu vermeiden oder per 2) Matching/Parallelisierung d.h. existieren wenige VP, die sich relevant unterscheiden, werden sie so aufgeteilt, dass relevante Merkmale in allen Gruppen gleichstark sind. Dabei sind Situationsmerkmale konstant zu halten u.a. Instruktion und EG/KG gemischt zu erheben, um uV und Situationsmerkmale nicht zu konfundieren d.h. *nicht* Tag 1 Bedingung A, Tag 2 Bedingung B.

- *Interne Validität*: Sie wird durch Störeinflüsse gesenkt u.a. Reifungsprozess (z.B. Kleinkindentwicklung), zwischenzeitliches Geschehen (z.B. Schwarzer Freitag), Messeffekt, Hilfsmittel (Änderung im Messinstrument), Gestikänderung des VL oder verzerrte VP-Auswahl (ungleiche KG und EG).

- *Externe Validität*: Sie wird u.a. beeinträchtigt durch Reaktivität (Messeffekt z.B. soziale Erwünschtheit), reaktive Effekte (methodenbedingter Effekt durch Datenerhebungsverfahren), Demand Characteristics³¹/Hawthorne Effekte³² (VP weiß, dass sie an Experiment teilnimmt, was ggf. Verhalten beeinflusst). Zur Kontrolle bieten sich postexperimentelle Befragung (VP-Wahrnehmung prüfen), nicht-Experimente (Schilderung) oder Rollenspiele/Simulationen an.

Diskrete/diskontinuierliche Variable: Diese hat endlich viele Ausprägungen (abzählbar) und kann vereinfachte, stetige Variable sein (z.B. Kinderanzahl). Diskrete Merkmale werden nach ihren Abstufungen unterschieden in *dichotom* (binär, 2fach gestuft z.B. Geschlecht m/w) oder *polytom* (mehrfach gestuft) und nach Entstehung der Abstufungen in *natürlich* (z.B. Augenfarbe) oder *künstlich* (Einteilung eines stetigen Merkmals z.B. Alter in jung-mittel-alt).

Stetige/kontinuierliche Variable: Sie besitzt unendlich viele Ausprägungen (unendlich abzählbar) in einem beliebigen Intervall und ist in eine diskrete Variable umwandelbar (z.B. Körpergewicht in kg).

Manifeste Variablen sind beobachtbar, **latente Variablen** liegen manifesten Variablen als hypothetisches Konstrukt zugrunde. Ein Konstrukt (Konzept) ist „psychisches oder soziales Phänomen, das nicht direkt beobachtbar (manifest) ist, sondern aus manifesten Indikatoren erschlossen wird“ [BN06, S.731] und wird über beobachtbare Indikatoren messbar gemacht (z.B. Stress). Das Beobachtbare ist der *Indikator* (stellvertretend für Konstrukt), der Erschließungsprozess die *Operationalisierung*.

- *Operationalisierung*: Als Oberbegriff von Messung, Skalierung und Indexbildung, legt sie fest, mit welchen Indikatoren theoretische Konstrukte (oder abstrakte Variablen) zu messen (beobachten, ermitteln, zählen) sind. Sie beschreibt Messgröße, Erhebungsmethode und -instrument. Zur

³¹ Demand Characteristic (Aufforderungscharakter) meint die Eigenschaft der Situation, Handlungen nahe zu legen / auszulösen aufgrund der Wechselwirkung von Feldsituation und Befindlichkeiten. (s. [Fa00])

³² Hawthorne-Effekt meint, dass die VP natürliches Verhalten zu „künstlichem“ ändert, weil sie weiß, dass sie an Versuchen teilnimmt und beobachtet wird, d.h. Ergebnisverfälschung, Effekte durch Versuchssituation selbst, was externe Validität senkt. (s. [Ot06][BN03])

operationalen Definition existieren 2 Arten: 1) *messorientiert* d.h. wie gemessen werden soll, mit Festlegung der aV und Erhebungsinstrumente (Messvorschriften) und 2) *untersuchungsorientiert* d.h. wie untersucht werden soll, mit Festlegung der uV und Auswahl des Untersuchungsplans. Fehler treten auf bei unangemessener/unzulässiger operationaler Definition (Konstruktvalidität), unzureichender Referenz (nicht alle Bedeutungskomponenten erfasst), Missspezifikation (Variablen gehören nicht zum Zielkonstrukt) und Überschussbedeutung (Variablen sind fälschlich dem Zielkonstrukt zugeschrieben). Zur Vermeidung bieten sich multivariate und differentiell orientierte Vorgehen an d.h. multiple Operationalisierung, mehrere aV und Methodenkombination.

3.5.3 Anwendung

Es werden experimentelle Variablen festgelegt und SV identifiziert mit geeigneten Kontrollstrategien, um Ergebnisverzerrung zu vermeiden. Zum Wenn-Teil (Bedingung) gehörige Variablen sind uV, zum Dann-Teil (Folge) aV. Die Hypothesenprüfung soll mittels *mehrerer* aV erfolgen (multivariat), die durch das Treatment ggf. verschieden beeinflusst werden für zuverlässigere Ergebnisse, eindeutigere Schlüsse, unabhängige Beobachtung, geringe Fehlerwahrscheinlichkeit und Vergleiche.

Variablenverteilung für TH1: Es soll die Wirkung der implementierten Maßnahme (*Kommunikation, gängige Kommunikationsmittel*) auf die Ausprägung der aV (Konstrukt *Ablenkung*) gemessen werden. Die uV wird vom VL direkt manipuliert und als Bedingung eingeführt, d.h. es handelt sich um eine Reizvariable. Sie soll in der Untersuchung 2fach gestuft auftreten (dichotome Variable), d.h. Ablauf *ohne* und *mit* Einsatz gängiger Kommunikation/-smittel. Dies erlaubt einen aussagekräftigen Einsatz der uV bei Aufteilung in KG und EG (KG: ohne | EG: mit), um mittels Vergleichen nachzuweisen, ob bei uV-Auftreten signifikante Ablenkung vorliegt. *Ablenkung* äußert sich u.a. anhand physiologischer Faktoren und soll daher über physiologische Indikatoren³³ messbar gemacht werden. Werkzeuge und Methode des UL (und BS) lassen sich hier sinnvoll einsetzen³⁴. Vor der Durchführung bieten sich zudem Ruheaufnahmen an („Normalwerte“ der VP), um diese mit Ergebnissen der KG und EG zu vergleichen und sicherzustellen, dass keine weiteren SV die uV-Wirkung verzerren. Zusätzlich erscheint die Erhebung subjektiver VP-Empfindung via Fragebogen sinnvoll, um Daten zu ergänzen, zu untermauern oder zu hinterfragen.

2fach gestufte uV (Reizvariable): X1 = Kommunikation, gängige Kommunikationsmittel (*ohne/ mit*)

aV: Z = Konstrukt „Ablenkung“, soll anhand folgender Indikatoren gemessen werden:

2fach gestufte aV: Z1 = Eyetracking (Blickpunkte/Fixationen, Pupillenradius, Lidschläge) *signifikante Ablenkung ja/ nein*

2fach gestufte aV: Z2 = EOG-System (Sakkaden, Lidschläge) *signifikante Ablenkung ja/ nein*

2fach gestufte aV: Z3 = Thermographie (Hauttemperatur) *signifikante Ablenkung ja/ nein*

2fach gestufte aV: Z4 = ProComp (EKG, EMG, BVP, Atmung, Hautleitfähigkeit, -temperatur) *signifikante Ablenkung ja/ nein*

aV: Z5 = Fragebogen (*subjektives Empfinden*)

³³ „Physiologische Messungen gehören zu den objektiven Meßmethoden, d.h. der Proband hat nicht die Möglichkeit, in direkter Weise auf die Messergebnisse verfälschend Einfluß zu nehmen [...]sog. Biosignalen [...], z.B. der Hauptleitfähigkeit als indirekt elektrisches Signal oder der Herzaktivität und den Gehirnströmen als direkte elektrische Signale.“[BN03, S.278]

³⁴ Festlegung geeigneter physiologischer Indikatoren (z.B. Pupillengröße) erfolgt außerhalb der Arbeit durch den Experimentator des UL. Gewählte Indikatoren und Erhebungswerkzeuge werden hier nur der Vollständigkeit halber und zur weiteren Planung genannt. Daten- und Hypothesenauswertung (TH1) erfolgt zeitversetzt und wird daher hier nicht aufgeführt.

Variablenverteilung für TH2: Es soll die Wirkung der implementierten Maßnahme (*Kommunikation, gängige Kommunikationsmittel*) auf die aV-Ausprägung (Konstrukt *Arbeitsleistung*) gemessen werden. Als deskriptives Maß zur Beschreibung der Arbeitsleistung eignen sich Fehleranzahl und Anzahl gelöster Aufgaben in der vorgegebenen Zeit, die gleichzeitig Nutzenfaktor in der Praxis darstellen (Produktivität am Arbeitsplatz, Kosten). Die aV wird also durch die Indikatoren *Qualität* und *Arbeitsmenge* messbar gemacht, die je endlich viele Ausprägungen besitzen (diskrete Variablen). Dadurch kann sie präziser erhoben werden (differenzierte Ergebnisse), z.B. besteht die Möglichkeit, dass Aufgaben in der EG genauso schnell aber fehlerbehafteter als in der KG oder Aufgaben in der EG langsamer aber genauso fehlerfrei wie in der KG bearbeitet werden. Wie angedeutet, soll die Durchführung 2malig als EG und KG erfolgen und die aV in beiden Gruppen erhoben werden (Vergleiche).

2fach gestufte uV (Reizvariable): X1 (s. o.)

aV: Y = Konstrukt „Arbeitsleistung“, soll anhand folgender Indikatoren gemessen werden:

(diskrete Variable) aV: Y1 = Qualität (Anzahl der Fehler bei bearbeiteten Aufgaben)

(diskrete Variable) aV: Y2 = Arbeitsmenge (Anzahl gelöster Aufgaben in vorgegebener Zeit)

aV: Y3 = Fragebogen (*subjektives Empfinden*)

Da sich Rückschlüsse auf Leistungsänderungen gut über IQ-Werte ziehen lassen, werden sie zusätzlich ermittelt und bei der Auswertung einbezogen. Es soll zwischen den 3 Werten „IQ gestiegen“, „IQ gleich bleibend“, „IQ gesunken“ unterschieden werden. Da VP durch uV nicht ungestört arbeiten und auch selbst kommunizieren dürfen (evtl. Hilfestellung von außen), ist mit „verzerrter“ IQ-Messung zu rechnen. Erhobene Werte sollen hier nicht den „wahren“ IQ erheben, sondern nur Aufschluss über mögliche IQ-Änderungen geben, um Leistungsunterschiede zwischen KG und EG nachzuweisen.

3fach gestufte aV: Y4 = Messung IQ-Wert (IQ gestiegen / IQ gleich bleibend / IQ gesunken)

Zudem bietet sich die Erhebung subjektiver VP-Empfindung an, um Ergebnisse zu unterstützen/zurück hinterfragen und ggf. SV aufzudecken. HH gilt als vorläufig bestätigt bzw. *nicht falsifiziert*, liefern TH1 und TH2 eindeutige Ergebnisse und werden nicht falsifiziert, d.h. es liegt Ablenkung vor (TH1) und veränderte Arbeitsleistung (TH2) (Arbeitsmenge und/oder Qualität). Ergebnisverzerrung durch unberücksichtigte Variablen und SV ist auszuschließen bzw. durch Kontrollstrategien zu verhindern (►3.5.4). Die aV werden in möglichst wenige Stufen eingeteilt, da die zu untersuchende Hypothese im ersten Schritt nur unspezifizierte Leistungsänderung (ungerichtet) und Ablenkung überhaupt (ungeachtet der Ausprägung und Art) prüft. In weiteren Untersuchungen bieten sich auf dieser Basis komplexere Fragestellungen an, z.B. Richtung der Leistungsänderung unter speziellen Bedingungen.

3.5.4 Kontrolle von Störvariablen

Standardisierung der Bedingungen: Mit klaren Handlungsanweisungen, standardisierten Abläufen, Instruktionen und (verbalen) Vorgaben soll sichergestellt werden, dass Ergebnisse objektiv und Versuche wiederholbar sind und alle Variablen neben der uV für alle VP konstant sind³⁵. Generell ist es

³⁵ Ansatz ist, dass störende Bedingungen „für die Ergebnisse weniger erheblich sind, wenn alle Untersuchungsteilnehmer ihrem Einfluß in gleicher Weise ausgesetzt sind. Konstante störende Bedingungen mindern zwar die Generalisierbarkeit (externe Validität), aber nicht

fast unmöglich, SV ganz zu verhindern, sie sollen daher konstant gehalten und ins Testsetting passiv oder aktiv integriert werden. „Ein relativ kleiner, unaufgeklärter Rest an Störeinflüssen und Meßfehlern ist tolerierbar. Schließlich wird in der empirischen Forschung mit Hypothesen in der Regel nicht das Ziel verfolgt, den Einzelfall detailliert zu erfassen, sondern zusammenfassend und übergreifend Tendenzen und Trends aufzuzeigen.“[BN03, S.16-17] Durch Erfassung in Fragebögen, Vorexperimente (VE), Protokolle und Nachgespräche sollen SV aufgedeckt werden. Nachfolgend sind für den Versuch relevante SV und Kontrollstrategien aufgeführt (s. [BN03][Fa00][Ge03][e109][Re03]):

Störeinflüsse der VP (personengebundene SV): u.a. ungewohnter Umgang mit den Kommunikationsmitteln, Misstrauen, unnatürliches Verhalten der VP, Stimmung, Teilnahmemotivation, Bewertungsangst, Bedürfnis nach sozialer Anerkennung
Kontrolle: Eliminierung interindividueller Variation, indem VP mehrfach untersucht wird (in KG und EG) und als ihre eigene Kontrolle fungiert (Messwiederholungsplan), d.h. jede VP absolviert *beide* Bedingungen, sodass personengebundene SV für beide Bedingungen *konstant* bleiben (*Vorteil:* Ökonomie, Parallelisierung der SV; *Nachteil:* ggf. eingeschränkte Generalisierbarkeit), Vergleich mehrerer verschiedener VP miteinander problematisch (nur eingeschränkt möglich)

Soziale Situation: u.a. durch soziale Interaktion zwischen VP und VL (Kommunikation, Rollenübernahme), psychologische Abwehr/Reaktanz („Sabotage“) per Vermeidung, verstecktem/offenem Widerstand (Gründe: zu hohe Anforderungen, subjektiv als Nötigung erlebte Instruktionen, Argwohn gegenüber Täuschung, erzwungene Verhaltens- und Reaktionsweisen)
Kontrolle: neutrales freundliches Klimas, kein unnatürlicher Leistungsdruck, plausibles Versuchsziel, höfliche Formulierung der Instruktion, entspannte Anwerbungsituation und Durchführung (persönliche Freiheit, Handlungsspielraum)

VP-Effekte, -Motivation, -Rolle³⁶: 1) Erwartungen: Auswirkungen experimenteller Bedingung (z.B. Placeboeffekt), soziale Erwünschtheit von Verhalten 2) Motive: Freiwilligkeit (Interesse an Wissenschaft, Belohnung), Test-/Bewertungsangst
Kontrolle: Versuchsziel tarnen, möglichst keine Interaktion mit VP während Durchführung, SV im Nachgespräch aufdecken

Maturation Effect (s. [Re03]): Veränderung der VP im Experimentverlauf (Müdigkeit, Hunger) mindert interne Validität
Kontrolle: kurze Versuchszeit (2x ca. 10 Minuten), Beobachtung der VP und Abfrage im Nachgespräch

Positions- und Carry-over-Effekte³⁷ (bei Messwiederholung): Position der Bedingung in immer gleicher Reihenfolge und Einfluss früherer Bedingungen auf spätere, wirken sich auf uV-Effekte aus (Ermüdung, Übung, sinkende Angst)
Kontrolle: Kommunikationsmedien /-arten und Reihenfolge in allen Versuchen unvollständig ausbalancieren, Vertauschen der EG-KG-Reihenfolge bei verschiedenen VP, variierte Zeitintervalle für uV-Einsatz hier unmöglich (zu kurze Versuche)

(unbeabsichtigte) VL-effekte und VL-Erwartungseffekte³⁸: z.B. unwillentliche Änderung der Gestik beim VL
Kontrolle: nicht vollständig vermeidbar, möglichst einheitliches neutrales VP-unabhängiges Verhalten, reduzierte direkte Kommunikation und Interaktion zwischen VP und VL (*Nachteil:* gesteigerte Künstlichkeit), Standardisierung des Ablaufs um Bedingungen konstant zu halten (schriftliche Instruktion, Instruktionen per Email), gegenseitiges und Selbstbeobachtung der VL (Kontrolle), Prüfen der Videoaufzeichnung, Anonymität zwischen VL und VP (gleiches Verhalten allen VP gegenüber), VP-Beobachtung per Live-Videoübertragung in *separaten* Raum (d.h. keine direkte „störende“ Beobachtung), Protokollieren aller unerwarteten Vorkommnisse, in späteren Untersuchungen u.a. VL ausschalten, VL- Schulung, Manipulation

zwangsläufig die Eindeutigkeit der mit der Untersuchung gewonnenen Erkenntnisse (interne Validität).“[BN03,S.88] Fokus liegt hier auf hoher interner Validität, noch keine Generalisierung.

³⁶ Rollen (s.[Fa00]) sind „gute“ VP (motiviert, folgt Instruktion, versucht Hypothese zu bestätigen), „um ihre Bewertung besorgte“ VP (will guten Eindruck machen, beeinflusst Reaktionsverhalten), „ehrliche“ VP (bemüht Instruktion zu folgen, keine eigene Hypothesenbildung) und „negativistische“ VP (Widerstand, versucht Hypothese zu widerlegen / Versuch zu sabotieren).

³⁷ Carry-over-Effekt meint u.a. Sensibilisierung, Übungs-, Übertragungs-, Ermüdungs-, Erinnerungs-, Sättigungseffekte (VP verliert Lust an Teilnahme), zwischenzeitliches Geschehen, allgemeine Testerfahrung, zwischenzeitliches Training (s. [BN06]).

³⁸ VL-Erwartungseffekt und VL-Effekt (experimenter bias) meint (un)bewusste Beeinflussung durch VL und systematische Ergebnisverzerrung. Sie treten nur in Versuchen auf, wo VP und VL kommunizieren, Ergebnisse sind nicht VL-unabhängig oder objektiv. Der VL-Effekt (experimenter effect) ist ein erwartungsunabhängiger Effekt des VL, sprich 1) bio-soziale Effekte (Geschlecht etc.), 2) psycho-soziale Effekte (Persönlichkeitsmerkmale, Verhaltensstil) und 3) situative Effekte (aufdringlich, verunsichernd). Der VL-Erwartungseffekt (experimenter expectancy effect) ist die 1) (unbewusste) VP-Beeinflussung (u.a. paraverbale Signale wie Betonung, non-verbale Signale wie Mimik) oder 2) hypothesenkonforme Fehler (Datenbeobachtung, -aufzeichnung, -verarbeitung).

der VL-Erwartungen, Blind- oder Doppelblindversuch³⁹, VL nicht vollständig über Versuchsziel informieren

Situative SV, verschiedene Testzeitpunkte, andere Störquellen (äußere Faktoren neben uV): VL, Personen/Lärm außerhalb des UL, Lichtverhältnis, Temperatur (u.a. Einfluss auf Konzentration), unangenehmes Arbeitsklima, Getränke, Speisen
Kontrolle: Minimierung externer Geräuschkulisse durch Durchführung in Semesterferien in abgelegenen ruhigen Untersuchungsraum und geeignete Warnhinweise am Labor (Schild), neutrales angenehmes Arbeitsklima mit verdeckter Platzierung der VL (externer Beobachtungsraum, nur ein VL im Labor, indirekte Beobachtung), UL abschirmen (geschlossene Fenster, Jalousien, Türen), Ergebnisse nachträglich auf externe SV prüfen (Videos, Nachgespräch), UL gut belüftet, beleuchtet, wohltemperiert, ausgeglichen, konstante Licht- und Temperaturverhältnisse bei allen Durchführungen, keine Ablenkung durch VL, Getränke/Erfrischung erst nach Versuch, insgesamt Situationsmerkmale möglichst über alle Versuche hinweg konstant halten (festgelegter Instruktionswortlaut, identische Beschreibung des Versuchs/Geräte/Vorwand, gleicher Raum/ Beleuchtung/ Temperatur und VL), soweit organisatorisch möglich alle Durchführung auf konstante Zeitpunkte festlegen

Untersuchungsbedingte SV (u.a. Instruktionsfehler, unerwartete Zwischenfragen, Stromausfall)

Kontrolle: Kontrolltechnik „Registrieren“ d.h. sind Versuchsbedingungen nicht vergleichbar, sind Störungsart und -intensität zu registrieren und protokollieren (*hier:* Video, Protokoll) für nachträgliche statistische Korrektur (s. [BN06, S.528])

Kohorteneffekte (Generationseffekte) und Altereffekte

Kontrolle: Mischung der Generationen/Altersgruppen (später randomisierte Stichproben oder selektive Gruppen)

Demand Characteristics und Hawthorne-Effekt: „künstliche“ Verhaltensänderung (unnatürliches Verhalten) der VP aufgrund des Wissens über Experimentsituation bzw. Beobachtung durch Dritte, VP versucht z.B., ‚wahres‘ Versuchsziel zu erkennen („Hypothesis Guessing“) und Antworten zu geben, die VL hören möchten (sinkende interne und externe Validität)

Kontrolle: glaubhaftes vorgetäushtes Versuchsziel, Situationskontrolle per postexperimenteller Befragung (Debriefing)

Forced Exposure: VP können während des Versuchs nur Anweisungen befolgen (unnatürlich, senkt externe Validität)

Kontrolle: VP vorab erklären, dass sie Aufgaben und den Test jederzeit abbrechen kann (Lockerung des Versuchsablaufs)

Versuchsziel: VP deckt „wahres“ Versuchsziel vorzeitig auf oder besitzt Vorwissen über Experiment (Verhaltensänderung)

Kontrolle: räumliche Trennung der VP, die Experiment hinter sich haben, von prospektiven VP, Schweigepflichterklärung, glaubhaftes Versuchsziel formulieren und stringent in allen Materialien vermitteln (Instruktionen, verbale Anweisung etc.)

SV zwischen Messungen, Instrumentänderung, Geräte-/Hilfsmittelausfall, weitere Mess- und Auswertungsfehler

Kontrolle: Ablauf, Skalen und Instrumente für alle Durchführungen im Hauptexperiment (HE) konstant halten, klar definierte standardisierte Bedingungen (schriftlich), Möglichkeit des Versagens berücksichtigen d.h. Auswahl zuverlässiger Instrument, VE zum Proben, Biosignale prüfen (Stresslevel der VP, vorab Ruheaufnahme), Vorgehen und Auswertung vorher genau planen, Ergebnisse und Auswertung nachträglich prüfen (mögliche Mess- und Auswertungsfehler)

Mangelnde instrumentelle Reliabilität (ungenau/fehlerhafte Erfassung des zu Messenden durch Messinstrument, senkt interne Validität) **Mangelnde instrumentelle Validität** (Messinstrument misst nicht Gewünschtes, senkt externe Validität)

Kontrolle: prüfen, ob VP-Aufgaben und Methoden das zu Überprüfende messen und Messgenauigkeit (VE), ob Parallelformen der VP-Aufgaben gleichartig und auf gewählte Stichprobe anwendbar sind (z.B. Altersgrenze bei genormten IQ-Tests)

Laborexperiment: Künstlichkeit der Anordnung kann VP-Verhalten beeinflussen (geringere externe Validität als im Feld)

Kontrolle: möglichst neutrale Arbeitsumgebung (Computerarbeit, Ruhe, wenige Leute, keine offene Beobachtung)

Gleichsetzung des „Kausalschlusses“ mit „Korrelation“: vermuteter Zusammenhang zwischen 2 Variablen, der nicht gegeben ist sondern durch dritte (unabhängige) Variable entsteht, die auf die beiden anderen (abhängigen) Einfluss nimmt

Kontrolle: Zusammenhang zwischen Variablen prüfen, u. U. weitere Variable identifizieren, die Einfluss nimmt

Auftreten von Dritt- und Störvariablen

Kontrolle: Laborversuch (gewährleistet beste SV-Kontrolle), weitestgehende Isolation oder Integration der Variablen

Gefahr deutlich ungleicher Gruppe an VP: relevanter Unterschiede (u.a. Geschlechterverteilung)

Kontrolle: Ausgeglichenheit der VP bzgl. relevanter Merkmale (in späteren repräsentativen Untersuchungen) (► 3.12)

³⁹ Blindstudie heißt, VP besitzen keine Kenntnis darüber, ob sie der EG oder KG angehören, um Einfluss von Erwartungen zu eliminieren. Doppelblindversuch sind Versuche, in denen VL und VP nicht die VP-Zugehörigkeit zur Gruppe (KG bzw. EG) kennen.

Konfundierung von Treatment und Situationsmerkmalen oder durch Vermischung von Störfaktoren und Stimulus

Kontrolle: EG und KG in gemischter Form erheben (d.h. *nicht* Tag 1 Bedingung A, Tag 2 Bedingung B) und prüfen (währenddessen und nachträglich durch Videoaufzeichnung und Nachgespräch), inwieweit Effekte auf aV von uV ausgehen

SV durch zwischenzeitliches Geschehen (ändert ggf. VP-Verhalten)

Kontrolle: Stimmungsfragebogen zu Beginn (VP-Befinden erheben), Tagesgeschehen prüfen, Nachgespräch (Debriefing)

Reaktivität (reaktive Effekte des Messens z.B. soziale Erwünschtheit, Sensibilisierung) und **reaktive Effekte** (durch die Situation, Messeffekte auf VP, Messvorgang beeinflusst das zu Messende, Konfundierung mit aV, gefährdet Eindeutigkeit der Schlussfolgerungen und interne Validität): Ergebnisse kaum generalisierbar, senken externe Validität

Kontrolle: ständige Kontrolle (Beobachtung, Protokoll, Nachgespräch / Debriefing), Berücksichtigung bei der Auswertung (Messeffekte treten bei Messwiederholungsdesigns auf und beeinflussen interne Validität negativ)

Lern- und Reifungseffekte: kann bei VP *und* VL auftreten bei Versuchen über längere Zeiträume, Routine (VL passen Verhalten an, was sich auf spätere Tests auswirkt), Hunger, Unaufmerksamkeit, ähnliche VP-Aufgaben in KG und EG etc.

Kontrolle: gegenseitige Kontrolle der VL, Ablauf nach jedem Versuch prüfen, Ergebniskontrolle bzgl. Gewöhnung, Vertrautheit mit Aufgabe, Lernen (speziell D2), standardisierte Abläufe (schriftliche Instruktion, Handlungsvorgabe für VL)

Veränderte experimentelle Bedingungen

Kontrolle: konstante Bedingungen in KG / EG für alle VP (Parallelfarm der Aufgaben, gleiche Instrumente und Materialien)

Kompensatorischer Wettstreit: VP versucht auch unter schlechteren Bedingungen (hier: EG) gleiche (leistungsstarke) Ergebnisse zu liefern **und Treatmentdiffusion:** VP kennt EG-Treatment und versucht, Reaktion zu antizipieren/imitieren

Kontrolle: Aufmerksame Beobachtung durch VL (ggf. Aufdecken der Effekte), Nachgespräch (Debriefing)

Stichprobe: systematisch verzerrte Stichprobe, verzerrte Ergebnisse, Auswahlfehler, nur Freiwillige als VP (► 3.12)

Zufällige und systematische Fehler: SV unentdeckt oder nicht zu verhindern (meist im Einzelversuch), beeinflussen Ergebnisgenauigkeit aber verändern sie nicht in eine Richtung, hohe Fehlerrate kann Unterschiede in KG und EG verdecken

Kontrolle: in VE aufdecken, Konstanthalten / Integration der SV (passiv oder aktiv), Kontrollstrategien anwenden (► 3.5.3)

3.6 Methoden und Erhebungsverfahren

Zusammenfassend sollen im eigenen Vorhaben experimentell durch uV-Manipulation die aV-Effekte untersucht und SV kontrolliert/integriert werden. Die Konstrukte „Ablenkung“ und „Arbeitsleistung“ werden als aV durch Variablen konkretisiert. Anknüpfend sind geeignete Methoden zu wählen und extern festgelegte UL-Werkzeuge zu nennen ► 3.7. Fragestellungen lassen sich oft nicht mit nur einer Methode suffizient beantworten, es empfehlen sich *Methodenkombinationen* (verringert zudem Risiko von Operationalisierungsfehlern). Für die angestrebte Untersuchung bieten sich *Experiment* und *Befragung* an. Experimente kennzeichnen sich durch „absichtliches und planmäßiges Herbeiführen eines Vorganges zur gezielten Beobachtung“ [Ho97, S.63] und eignen sich speziell für Hypothesenüberprüfung und Laborversuche. Als sekundäres Verfahren soll Befragung eingesetzt werden als „Technik zur Erfassung von Daten mit Hilfe der Beantwortung von Fragen“ [Ho97, S.63]. Die Fragebögen sollen Ergebnisse aus dem Experiment bestätigen/hinterfragen, die Stichprobe prüfen, SV aufdecken und werden hier speziell zur Bewertung des Forschungsdesigns eingesetzt (Schwachstellen, Optimierung).

3.6.1 Experiment

Aus obigen Kapiteln geht hervor, dass das (strenge) *quantitative Experiment* eine gute Methode für den geplanten Versuch im UL darstellt. Hier sind 4 Merkmale bedeutsam: 1) *Beschreibbarkeit der Versuchsbedingungen* als zentrales (Qualitäts-)Merkmal und Grundlage, um Ergebnisse sinnvoll zu interpretieren und Wiederholbarkeit durch andere Forscher zu gewährleisten, 2) *Willkürlichkeit* d.h. freie willkürliche uV-Änderung für sichere Ursache-Wirkungs-Zusammenhänge, 3) *SV-Kontrolle* als

zentrales Qualitätsmerkmal, bei dem angestrebt wird, SV auszuschalten oder zu erheben (Integration), um Effekte abzuschätzen und aus Ergebnissen herauszurechnen und 4) *Wiederholbarkeit*, die erst Ergebnisprüfung, Experimentiererweiterung und Prüfen neuer Fragestellungen erlaubt und repräsentative Ergebnisse unterstützt. Wiederholung unter exakt gleichen Bedingungen ist jedoch schwierig z.B. durch Trainingseffekte sind nicht gleiche VP verwendbar. In der Praxis ist daher eine möglichst große Annäherung an gleiche Bedingungen gängig.

Labor: Dieses eignet sich am besten, echte Experimente durchzuführen. Zudem können im UL Variablen erhoben werden, die im Feld kaum/unmöglich sind, d.h. leichtere umfangreichere Messung.

Gruppe: Statt getrennter Einteilung in KG und EG, soll *dieselbe* VP *vor* und *nach* uV-Änderung in

Wissen der Versuchsperson	Benennung der Untersuchung			
	offen	getarnt	quasi-biotisch	(voll-)biotisch
Kenntnis des Versuchsziels	✓	✗	✗	✗
Kenntnis eines (vorgeschobenen) Zieles	✓	✓	✗	✗
Wissen, dass ein Versuch stattfindet	✓	✓	✓	✗

Abb. 3 Durchschaubarkeit der Untersuchung [Wp12]

KG und EG untersucht werden. Dies bietet sich bei 2-stufigen uV an, erlaubt Datenvergleich der VP ohne individuelle SV, senkt Kosten und eignet sich für Einzelversuche mit wenigen VP.

Durchschaubarkeit der Versuchssituation:

Für Experimente gilt aus Forschersicht - *je weniger durchschaubar, desto besser* - sowohl bei

VP als auch bei VL mit Kontakt zu VP. Es lässt sich unterteilen in vollbiotische, quasibiotische, getarnte und offene Versuchssituationen (►Abb. 3). Allgemein ist es sinnvoll, wenn VP und VL die Situation möglichst wenig durchschauen, um SV wie VL- und VP-Effekte zu vermeiden (Reduktion der Durchschaubarkeit: „Verblindung“). Eine vollbiotische Versuchssituation lässt sich im Labor schwer umsetzen, eine offene Untersuchungssituation erhöht hingegen die Wahrscheinlichkeit von Reaktivität (unnatürliches Verhalten durch Versuchssituation und -ziel) und verzerrt Ergebnisse. Es bietet sich hier die getarnte Versuchssituation an, in der VP wissen, dass sie an Versuchen teilnehmen aber das eigentliche Versuchsziel nicht kennen (Vorwand für realistischere uV-Integration und Effekte). Es gilt dabei, ethische Regeln zu beachten, d.h. Selbstbestimmung und persönliche Freiheit nie einzuschränken und VP zum Ende in einem Debriefing das „wahre“ Ziel zu enthüllen (s. [Ge03]).

Aus diesem und vorherigen Kapiteln ergibt sich folgende zusammenfassende Übersicht:

Art: quantitatives „echtes“ Experiment d.h. beschreibbare Versuchsbedingungen, Willkür, SV-Kontrolle, Wiederholbarkeit

Vorgehen: quantitative Datenerhebung auf experimenteller Ebene zur Überprüfung von Ursache-Wirkungs-Hypothesen

Klassifikation nach: 1) *Untersuchungsziel* als Hypothesenüberprüfung, „klassisches“ (Prüf-)Experiment, kausale Prüfung vorhandener Annahmen (Hypothesen) durch aktive Manipulation der uV und Erfassung der Effekte auf aV, 2) *Zahl der uV*: einfaktorielles Experiment (eine uV), 3) *Zahl der aV*: multivariates Experiment (mehrere aV) und 4) *Versuchsort* im Labor (UL, BS) mit optimaler Dritt- und SV-Kontrolle, leichter direkter uV-Änderung (klare aV-Effekte), hoher interner Validität, freier willkürlicher VP-Zuordnung zu Bedingungen, leichter umfangreicherer Messung gegenüber Feld (u.a. physiologische Indikatoren), weniger Erfolgszwang und Kosten, leichter wiederholbar durch definierte Bedingungen. *Nachteile* sind Selektionseffekte, inhärente Künstlichkeit der Anordnung, eingeschränkte externe Validität (Realitätsmangel, geringere Generalisierbarkeit), Verhaltensänderung bei VP durch Wissen um beobachtete Situation.

Versuchsplanung: VP wissen um Versuch, Ziel ist aber *getarnt* (abschließendes standardisiertes Debriefing notwendig)

- KG und EG mit jeweils gleicher VP und identischer Laborsituation, jede VP absolviert beide Bedingungen mit Vorteil der Ökonomie, Parallelisieren der SV der VP, vergleichbaren Daten selber VP d.h. konstante interindividuell, personelle SV

- VE als kleines Experimente vorab zur Planung des HE und Erprobung und Verbesserung experimenteller Bedingungen
- HE als finales Prüfexperiment zur Bewertung des Forschungsdesigns (und Hypothesenprüfung)

3.6.2 Fragebogen

Für das eigene Vorhaben werden Fragebögen mit verbalen Marken für die Rating-Skalen verwendet (verbale Charakterisierung der numerischen Abstufung), für bessere Anschaulichkeit und annähernd äquidistanter Ausprägung des Merkmalskontinuums der Begriffe sowie ungeradzahlige 5-stufige Skalen⁴⁰. Zur Ermittlung des Zustimmungsgrades werden Häufigkeits- (nie-selten-gelegentlich-oft-immer) und Intensitätsskalen (gar nicht-wenig-mittelmäßig-ziemlich-völlig) eingesetzt ([Ro87] zit. nach [BN06, S.177]). Sie enthalten je eine neutrale Mittelkategorie, um unsicheren VP das Ausweichen auf die Neutralkategorie zu erleichtern und Antwortpolarisierung zu vermeiden (VP nicht zur Aussage drängen). Es wird hier nicht mit übermäßig *zentralen Tendenzen* als SV gerechnet, das „Ambivalenz-Indifferenz-Problem“ ist aber trotzdem zu beachten (neutrale Antwort kann besagen, dass die VP keine dezidierte Meinung vertritt oder ambivalent ist). So sind Ergebnisse zu hinterfragen, messtheoretische Probleme und Urteilsfehler beim Einsatz der Rating-Skalen⁴¹ zu beachten sowie Verzerrungen durch stereotypes Ankreuzen, Akquieszenz oder willkürliche, kalkulierte Antwortveränderung. Es folgen wichtige SV und Kontrollen für den eigenen Versuch (s. [BN03][Fa00]):

Mangelnde Wiederholbarkeit, keine Verallgemeinerung möglich (keine Generalisierung, niedrige externe Validität):

Kontrolle: standardisierte Befragung d.h. schriftliche Fragebögen (weniger fehleranfällig, erlauben objektivere Auswertung, kostengünstiger, wiederholbar durch andere Forscher d.h. bessere Verallgemeinerung / Generalisierung / Prüfung)

Auswertungsfehler

Kontrolle: schriftliche Fragebögen mit geschlossenen Fragen und genormten Skalen (statistische Auswertung möglich, einfach, objektiv, deutlich), Schwachstellen bei Auswertung einbeziehen (und offen legen), Prüfung der Eignung durch VE

Unehrliche Antwort/Falschantwort/soziale Erwünschtheit/Antworttendenz (Unentschiedenheit, Nein-/Ja-sage-Tendenz)

Kontrolle: Anonymisierung der Fragebögen (anonyme Antworten, keine direkte VP-Beobachtung), individuelle Antworttendenz bei Auswertung beachten, keine sozial erwünschten Antworten herausfordern, Versuchsziel undurchschaubar machen

Beeinflussung der VP durch Gruppendynamik und differentielle Bereitschaft der VP

Kontrolle: Beantwortung erfolgt einzeln (direkt zum Versuchszeitpunkt), Fragebogen kurz halten um vollständiges Ausfüllen zu ermutigen und Sinken der Konzentration/Ermüdung vorzubeugen (Nicht-Antworten in Auswertung berücksichtigen)

Formulierung und Missverständnisse

Kontrolle: neutrale eindeutige einfache Formulierung, *keine:* Suggestivfragen, stereotype Formulierung, (doppelte) Verneinung, Verlegenheiten, VL stehen für Nachfragen zur Verfügung, keine unklaren Antwortkategorien, Testen in VE, mehrere Aspekte in getrennte Fragen erheben (Eindeutigkeit), neutrales übersichtliches Layout (Tabellen, Schattierung)

Beeinflussung der Antworten durch VL (persönliche Eigenschaften, Verhaltensweisen) und Situation

Kontrolle: Kontakt mit VP auf das Schriftliche reduzieren (nur Fragebogenübergabe), Kommunikation nur bei Fragen, Stimmungsfragebogen *vorab* um Befinden zu erheben, Fragen zur Versuchssituation um SV aufzudecken, externe Störung vermeiden, Ausfüllsituation für alle VP gleich halten (konstante SV), neutral und angenehm: gute Organisation (Zeit-Puffer, nicht hetzen), inhaltliche Freiheit (unbeobachtet, allein arbeiten), Anonymität sichern durch u.a. VP-Codes

⁴⁰ Laut Faustregel sind maximal 7 Antwortkategorien zu verwenden, um noch zu differenzieren. Optimal gelten annähernd gleichabständige 5-stufige Skalen mit erschöpfenden, eindeutigen Antwortmöglichkeiten ohne Überlappung ([Ro87] zit.nach [BN06,S.181]).

⁴¹ Dies sind u.a. Halo-Effekte (Pauschalurteil, Objekt wird bzgl. verschiedener Merkmale gleich eingestuft), Milde-Härte-Fehler (systematisch zu positiv/negativ), Tendenz zur Mitte (extreme Ausprägungen vermeiden, Antwortverweigerer, „weiß-nicht“-Antworten, nicht-entscheidungsfreudige VP) und Self-Serving-Bias (Selbstbeurteilung gemäß Selbstkonzept, oft ‚selbstwertschützend‘). (s.[BN03])

Vorgehen: Zunächst wird die Fragestellung präzisiert, indem der *Zweck* erfasst („Auswertung des Versuchs“) und in relevante Unterthemen unterteilt wird⁴² (Strukturierungshilfen ► Anhang B.4):

Zum Versuchsthema:

Thema 1: Befragung zur Person u.a. soziodemographische Daten, um Repräsentativität der Stichprobe prüfen

Thema 2: Befragung zum Verhalten/Gewohnheit, um Repräsentativität der Stichprobe, Messungen und Daten zu prüfen

Zum Forschungsdesign:

Thema 3: „Bewertung des Versuchs (Eindruck)“, u.a. offene Fragen für freie Meinung der VP, zur Optimierung des Versuchs (hier wird evtl. das „wahre“ Versuchsziel entlarvt daher Positionierung der Fragen am Ende, Störeinflüsse aufdecken)

Es sind primär geschlossene Fragen enthalten, aufgelockert durch einzelne offene, um freie Ergänzung zu gestatten. Geschlossene Fragen erleichtern die Auswertung, bieten größere Objektivität, Standardisierung (Wiederholbarkeit) und bedürfen keiner zeitaufwendigen, kostspieligen Kodierungs- und Kategorisierungsarbeit. Besonders in den letzten beiden Teilen soll aufgedeckt werden, inwieweit das „wahre“ Versuchsziel vorab erkannt wurde, ob das Vorgehen unnatürlich wirkte, inwieweit VP vom VL beeinflusst wurden (Ergebnisverzerrung), ob verwendete Kommunikationsmedien bekannt oder ihre Nutzung „erzwungen“ und VP-Aufgaben angemessen waren (Optimierung, Ergebnisprüfung).

Einsatz: Der Fragebogen ist am Versuchende zu beantworten, um kein Misstrauen zu erregen, das „wahre“ Versuchsziel nicht vorab preiszugeben oder VP-Verhalten zu beeinflussen. Er erhebt nicht die Forderung, den zentralen Kriterien der Testgüte (Objektivität, Reliabilität, Validität) zu genügen, da die ermittelten quantitativen Daten weniger der Hypothesenprüfung dienen als primär der Bewertung des entwickelten Forschungsdesigns und der VP. Hierneben wird *vor* Versuchsstart ein Stimmungsfragebogen eingesetzt (übernommen aus [Kl83]), inkl. Erhebung wichtiger technischer Angaben für die physiologische Messung. Vor allem in den VE sind zusätzlicher Aufwand und Testkosten (Fragebogenanfertigung-, ausfüllen, -auswertung) mit dem Nutzen abzuwägen und Inhalte zu testen.

Aus diesem und vorherigen Kapiteln ergibt sich folgende zusammenfassende Übersicht:

Standardisierte quantitative Befragung mittels Fragebögen für leichtere objektive Auswertung und Wiederholbarkeit

Kontakt: schriftlich (zur Standardisierung und Wiederholbarkeit, weniger Störeinflüsse u.a. durch VL)

Befragte: VP-Anzahl analog zum Experiment, einzelne Beantwortung (Einzelversuche), alle VP sind erreichbar da Befragung direkt vor und nach Versuch stattfindet, Befragung ist anonymisiert (Datenschutz, ermutigt ehrlichere Beantwortung)

Interviewer: Übergabe schriftlicher Fragebögen durch VL, Kommunikation nur bei Fragen (keine direkte VP-Beobachtung)

Fragenformate: Mischform (primär geschlossene Fragen, vereinzelte offene Fragen für Ergänzungen)

Offene Fragen: Versuchsbewertung (selbst formuliert) für interne Optimierung / Prüfung des Versuchsaufbaus

Geschlossene Fragen: Angaben zur Person (soziodemographische Daten) und Versuch(thema), zur Auswertung und Prüfung der Versuchsergebnisse und Stichprobe, gleichabständige 5-stufige Rating-Skalen mit Mittelkategorie und verbalen Marken

Anrede: „Siezen“ (offiziell, neutral, höflich, angebracht für gemischte Stichprobe aus Unbekannten, konstant im Versuch)

Anordnung: soziodemographische Daten vorab, Wichtiges ins mittlere Drittel (hier hohe Aufmerksamkeit), Trichterprinzip

Bearbeitungszeit: nicht länger als Versuchszeit selbst (gilt es in VE zu prüfen und ggf. zu kürzen)

⁴² Zusätzliche Fragen zum technischen Teil („Physiologische Signale“) wurden von VL1 eingefügt und extern ausgewertet.

3.7 Technischer Rahmen

Weitere wichtige Planungsschritte umfassen technische Aspekte wie die Wahl der Geräte (►3.7) und „gängigen Kommunikationsmittel“ (►3.8) sowie Planung der VP-Aufgaben (►3.9), Ausarbeitung der Versuchsunterlagen und weitere organisatorische Faktoren (Räume, Zeitplan, Finanzen) (►3.10).

Geräteauswahl: Diese hängt auch von der Laborausstattung der FHB ab und sollte den zeitlichen Rahmen und Ressourcen berücksichtigen und VP nicht ergebnisverzerrend hemmen/stören. Zu verwendende Werkzeuge des UL (und BS) werden hier lediglich aufgelistet (►Tab.7), um darauf weitere theoretische und praktische Überlegungen aufzubauen⁴³. Im VE (►4.4) sind Geräte, VP-Aufgaben, Materialien und Vorgehen auf ihre Eignung hin zu prüfen, um jene einzusetzen, die notwendige Daten liefern und in der gegebenen Bearbeitungszeit zu bewältigen sind.

Tab. 7 Versuchsapparatur im UL (s. [Ne12][In12][De12][BP02][Sc06][Wi11a][et07][Dp10][Br12])

Geräte	Messung	Beschreibung
ProComp Infiniti	EKG	Elektrokardiografie („Herzstromkurve“, „Herzschrift“): Ableitung elektrischer Aktivitäten des Herzens und Aufzeichnung in Form von Kurven
	EMG	Elektromyografie: Messung der natürlichen, elektrischen Muskelaktivität über Nadel-EMG oder Oberflächen-EMG (<i>im geplanten Versuch wird Letzteres verwendet</i>)
	BVP (Puls)	Blutvolumenpuls: relative Menge Blut, die aktuell durch das Gefäß fließt (gesteuert durch Herzschlagvolumen und Gefäßstatus), gilt als Maß für die Durchblutung
	Atmung	Respiration
	GSR (Hautleitfähigkeit, elektrodermale Aktivität)	Galvanic Skin Response: Messung des elektrischen Hautwiderstandes ("Sprache der Haut"), sehr präziser zuverlässiger Parameter im Rahmen von Biofeedback-Messungen - emotionaler Zustand / allgemeines Entspannungsniveau beeinflusst durch das vegetative Nervensystem die Schweißbildung und damit die elektrische Leitfähigkeit der Haut
	Hauttemperatur	
Morae	Live-Überwachung und Aufzeichnung von VP-Monitor und VP (von vorne)	Morae besteht aus den 3 Komponenten Recorder, Morae Observer und Manager <i>Morae Recorder:</i> zeichnet Nutzer-Video, Audio und Systemdaten auf (Event-Recording) und synchronisiert diese in Echtzeit; <i>Morae Observer:</i> Livebeobachtung, Aufzeichnung; <i>Morae Manager:</i> zur Bearbeitung, Suche, Analyse, Videoproduktion der Aufzeichnung
EOG	Elektrookulographie	- Registrierung der Augenbewegung (<i>im Versuch hier:</i> Ableitung über Napfelektroden)
Eyetracking (Remotesystem)	Blickerfassung (Eye-gaze Analysis System, Nyan Analyse Software, stationär)	- Aufzeichnung der Blickbewegung (<i>Bewegungsrichtung, -intervalle, Aufmerksamkeitsverteilung</i>) durch mobile Eyetracker (Befestigung auf Kopf) oder externe (Remote-Geräte oder mit Kopf-Fixierung), Kalibrierung der Eyetracking-Kameras auf VP individuell vor Versuchsbeginn, da Stellung und Größe der Augen je Person variieren
Thermokamera	Thermokamera, (Infrarot-, Wärmebildkamera)	Aufnahme der Intensitätsverteilung elektromagnetischer Wellen (durch integrierte Infrarotsensoren) u.a. zur Visualisierung der Infrarotstrahlung, Umsetzung in Thermografien (Wärmebilder auf Bildschirm anzeigen oder als Thermografien entwickeln)

Der Anschluss des BS an das UL der FHB erlaubt die gemeinsame Nutzung der Methoden und Geräte beider Labore im Versuch⁴⁴. Das UL besteht aus VP⁴⁵- und Beobachtungsarbeitsplatz⁴⁶ (mit u.a. Moni-

⁴³ Geräteauswahl, technische Durchführung, Datenerhebung/-auswertung werden extern durch VL1 behandelt (s. [Me13]).

⁴⁴ Untersuchung im Biosignallabor I (010) am Usability Arbeitsplatz, mit zusätzlichen Geräten des Biosignallabor II (003)

⁴⁵ Probandenarbeitsplatz besitzt u.a. 1 Windows-Rechner mit Webcam und unterhalb des Monitors montierte Eyetracking-Kamera.

toren, Web-, Beobachtungs-, Eyetracking-Kameras) (s. [Me11], ► Anhang B.2). Die Versuche werden zudem videoaufgezeichnet, um Messungen und Protokolle nachträglich zu prüfen.

Versuchszeit: Diese wird hier primär von (externen) technischen Überlegungen bzgl. Biosignal-Messung begrenzt. Es bietet sich eine Versuchszeit von ca. 11 Minuten an, um noch überschaubare Datenmengen zu erheben, die im zeitlichen Rahmen auswertbar sind. Nachteilig wirkt sich die verkürzte Versuchszeit ggf. auf VP und Arbeitssituation aus, da die Eingewöhnungszeit nur knapp ist und sich schwer eine „realistische Arbeitssituation“ simulieren lässt. Durch die kurzen Abläufe tritt hier das Zeitlimit in den Vordergrund, d.h. die VP weiß, dass sie die Aufgaben in kurzer Zeit zu absolvieren hat und wird vermutlich fokussiert und leistungsorientiert arbeiten. Infolgedessen ist es eher unwahrscheinlich, dass sie eigenständig angebotene oder andere Kommunikationsmittel wie Handy oder soziale Netzwerke nutzen wird, sodass sich Kommunikation hier eher auf die VL-initiierte beschränkt⁴⁷.

Anzahl der VP: Aus praktischen Gründen verringert sich die Anzahl, da in der verfügbaren Zeit eine limitierte Anzahl an Versuchen realisierbar und nur begrenzt große Datenmengen verarbeitbar sind.

Technik: Es wird vermutet, dass die Fülle technischer Geräte und damit verbundene umfassende Verkabelung an Gesicht, Hand und Oberkörper nur geringe Bewegungsfreiheit zulässt, die sich negativ auf die Natürlichkeit der Arbeitssituation und das VP-Verhalten auswirkt und Ergebnisse verzerrt.

3.8 Kommunikationsmittel

Hinsichtlich der Fragestellung ist zunächst zu definieren, welche „Kommunikation und gängige Kommunikationsmittel“ in den Versuch einzubeziehen sind, um Ergebnisse später einzuordnen. Kommunikation lässt sich in selbst gemachte (innere) und „erzwungene“ (äußere) unterteilen:

Kategorie I (äußere): Kommunikation wird vom VL als uV in den Versuch integriert. Sie tritt für VP willkürlich auf, ihr Auftreten ist von VP nicht kontrollier- oder planbar, z.B. Emails, Anrufe, Eintreten einer Person (*hier*: VL2).

Kategorie II (innere): Kommunikation wird durch VP initiiert und kontrolliert. Der VL nimmt hier maximal indirekten Einfluss bzw. partizipiert, z.B. Benutzung sozialer Netzwerke oder Handys im Versuch.

Äußere Ablenkung lässt sich durch VL kontrollieren und in den Ablauf integrieren (uV-Variation).

Innere Ablenkung wird durch VP erzeugt und ist daher schlecht kontrollierbar oder aktiv einplanbar.

Die Unterteilung der uV nach Kategorie I und II ist aus zeitlichen und organisatorischen Gründen im Rahmen dieser Arbeit nicht möglich und für die Fragestellung nicht zweckmäßig. Beide Kategorien sollen im ersten Schritt unter dem Oberbegriff „Kommunikation/gängige Kommunikationsmittel“ zusammengefasst werden⁴⁸. Nach der Durchführung sind Ergebnisse aber auch bzgl. des Auftretens der jeweiligen Kategorie hin zu überprüfen und ggf. nachträglich eine Differenzierung vorzunehmen.

⁴⁶ Beobachtungsarbeitsplatz besitzt u.a. 2 Rechner (für Eyetracking, Beobachtungscomputer zur Steuerung der Kameraanlage oder Durchführung der Blickpunktprüfung) und Überwachungssystem (Beobachtungscomputer, Kameraanlage mit 3 Monitoren und Rekorder, Kontrollmonitore für Augenposition und Echtzeit-Aufnahmen der 3 Überwachungskameras die VP und Umgebung aufzeichnet) (s.[Me13]).

⁴⁷ Bei der kurzen Versuchszeit wirkt sich zeit- (und alters-)bedingtes Konzentrationsnachlassen wenig aus, zudem werden diese SV durch gleiche VP in EG und KG konstant gehalten. In späteren Versuchen bieten sich längere Abläufe an, um VP an die Situation zu gewöhnen (natürlicheres Verhalten/Situation), für Aufgabenlösung mit weniger Zeitdruck (ermutigt ggf. selbst initiierte Kommunikation) und größere Datenmengen (ggf. Einflüsse durch Konzentrationsnachlassen / Ermüdung).

⁴⁸ Deuten weiterführende repräsentative Untersuchungen auf Bewährung der Hypothese hin, wäre eine Spezifizierung interessant z.B. „Unterschiedliche Medien wirken sich verschieden auf Leistung und Ablenkung aus“, „Innere und äußere Kommunikation wirkt sich verschieden auf Leistung und Ablenkung aus.“).

In jedem Fall sind beide im Protokoll zu vermerken und in der Auswertung zu beachten. **Planung** (►Tab.8): Es sollen Personen ins Testsetting integrieren werden, die Kommunikation initiieren (uV). Inhaltlich soll Kommunikation mit den VP-Aufgaben (ursprüngliche Tätigkeit) oder dem Versuch korrelieren (►3.10), um Zusammenhang zwischen Kommunikation und Tätigkeit herzustellen und uV plausibler einzubinden (Versuchsvorwand⁴⁹ unterstützen ohne Misstrauen, „natürlicher“).

Tab. 8 Liste der eingesetzten „Kommunikation und gängigen Kommunikationsmittel“ (Kategorie I)

Medium	Anwendung
Person VL 2 (Beobachtungsraum)	Tritt in UL ein, direkte verbale und nonverbale Kommunikation
Telefon (Internet-Telefonie via Skype ⁵⁰)	direkte verbale Kommunikation
Email (Web-Oberfläche oder installiertes Email-Programm: thunderbird, web.de, gmx.de)	schriftliche Kommunikation: Eintreffende Emails sollten nicht zu übersehen sein, u.a. blinkende Anzeige, akustisches Signal und/oder Anzeigen der Nachricht als Vorschau und Instruktionen per Email eingehen, um VP mit der Anwendung vertraut zu machen, Scheu zu nehmen und uV plausibel einführen.
Chat (via Skype, Kombination mit Telefonie, sodass VP nicht so viele Anwendungen gleichzeitig öffnen und „lernen“ muss)	schriftliche Kommunikation

Es existieren 4 bekannte Methoden, um Unterbrechung in Mensch-Computer-Interaktion zu koordinieren: 1) „immediate“ 2) „negotiated“ 3) „mediated“ und 4) „scheduled“⁵¹. Reaktionen bei Mensch-Mensch-Interaktion fallen darauf wie folgt aus: 1) „take-up with full compliance“ 2) „take-up with alteration“ 3) „decline“ oder 4) „withdraw“⁵². (s.[Mc99]) Die uV in diesem Versuch soll teils „negotiated“ eingeführt werden, d.h. die VP hat Kontrolle darüber, *wann* sie Unterbrechung bearbeitet (Email, Chat) und teils „immediate“, d.h. Unterbrechung tritt in den Vordergrund und *muss* beachtet werden (Eintreten des VL, z.T. Anruf). „(1) people perform very well when they can negotiate for the onset of interruptions, however giving people this kind of control also means that they may not handle interruptions in a timely way; (2) when people are forced to handle interruptions immediately, they get the interruption tasks done promptly but they make more mistakes and are less effective overall.“[Mc99, S.8] So sollen Unterbrechungsarten variiert werden (vgl. natürliche Arbeitssituation, in der diverse Varianten auftreten, die VP-abhängig gehandhabt werden), um Auswirkung nicht vorab zu stark in eine Richtung zu verzerren (vgl. o.g. Effekte s. [Mc99]).

3.9 Versuchsziel, Aufgabenplanung und -auswahl

Zunächst wird ein (vorgeschobenes) Versuchsziel entwickelt, anschließend Anforderungen an die VP-Aufgaben aufgeführt, um darauf aufbauend die konkrete Planung der VP-Aufgaben vorzunehmen.

⁴⁹ Vorwand ist, dass Anweisungen „von oben“ kommen (externer Professor), der nicht-technische VL befindet sich während des Versuchs nicht im Labor „um Biosignale nicht zu verfälschen“ und ist alleinig für die Aufgaben zuständig und Ansprechpartner für VP.

⁵⁰ Bietet sich hier an, da Telefonat am Rechner stattfindet und Messung nicht unterbricht bzw. einbezieht (u.a. Eyetracking)

⁵¹ 1) unmittelbare 2) ausgehandelt (verhandelt, abgesprochen) 3) vermittelt 4) geplant (freie Übersetzung der Autorin)

⁵² 1) Annahme mit voller Zustimmung 2) verzögerte Annahme 3) Ablehnung 4) (sich) zurückziehen (freie Übersetzung der Autorin)

3.9.1 Versuchsziel und Bedingungen

Vorgeschobenes Ziel: Um Störeinflüsse wie VP-Effekte und an das Ziel angepasstes VP-Verhalten möglichst zu eliminieren und uV sinnvoll einzuführen, wird die *getarnte Versuchssituation* gewählt. Das „wahre“ Versuchsziel soll der VP während des gesamten Versuchs verborgen bleiben, damit die uV maximal wirken kann. Entsprechend ist ein Vorwand zu wählen, der angesichts des Aufbaus (u.a. Geräte, VP-Aufgaben, Vorgehen) schlüssig erscheint und kein Misstrauen erregt. Da die Geräteanbringung an der VP sehr aufwendig und auffällig ist, bietet sich ein daran angelehnter Vorwand an:

Vorwand: „Der Versuch beinhaltet die Aufzeichnung physiologischer Signale (Herzrate, Puls, Atmung, Augenbewegung etc.) bei normaler Arbeitstätigkeit am Computer. Die Datensammlung dient der Erforschung physiologischer Vorgänge beim Menschen während der Arbeit am Computer, um in nachfolgenden wissenschaftlichen Untersuchungen das Arbeitsverhalten von Arbeitnehmern in organisierte, gesunde Strukturen zu überführen.“⁵³

Damit VP nicht versuchen, das „wahre“ Versuchsziel aufzudecken bzw. ihr Verhalten (un)bewusst zu ändern, ist besondere Vorsicht bei der Formulierung schriftlicher und mündlicher Anweisungen geboten und spezielle Schlüsselworte des echten Versuchs nicht zu erwähnen (u.a. Neutralität wahren, Zweifel am vorgeschobenen Versuchsziel ausräumen, VP nicht durch Wortwahl beeinflussen).

Zu vermeidende Schlüsselworte: Ablenkung, Konzentration, Kommunikation(-smittel), Leistung

Bedingungen für VP-Aufgaben: VP-Aufgaben dürfen nicht zu leicht sein (Misstrauen über Sinn und Ziel des Versuchs, weniger aussagekräftige Ergebnisse) aber auch nicht zu schwer (u.a. unnatürlicher Leistungsdruck, Stress und Angst über „normale“ Arbeitssituation hinaus, Überforderung, Entmutigung, Datenausfall) und zudem kein fachspezifisches Wissen/Vorwissen verlangen, dass nicht alle VP leisten können, um Ergebnisse nicht zu verzerren und Aufgaben auf die definierte Grundgesamtheit (►3.12) anwenden zu können. Sie sollten zielgerichtete Leistung und Konzentration fordern ohne aber VP in abnormalen Stress- und Angstzustand (Ausnahmestand) zu versetzen. Die Bearbeitung findet am Computer statt (►3.2), d.h. alle Aufgaben müssen digital umsetzbar sein. Um Störung durch ablenkende Bildschirminhalte zu vermeiden, sollten sie zudem in Vollbildanzeige darstellbar sein. Eine klare Zeitvorgabe an VP ermutigt zielgerichtetes und konzentriertes Arbeiten. Zudem eignen sich sprechende Bezeichner für die VP-Aufgaben (u.a. in Anweisungen), um Aufgabenart und gestellte Erwartungen zu verdeutlichen. Die Aufgabenfülle sollte möglichst schnell erfassbar sein z.B. durch Seitenzahlangabe oder mehrere Aufgaben auf einer Seite, damit VP ähnlich „normaler“ Arbeitssituation selbst priorisieren und ihre Zeit einteilen können (Zeitmanagement). Zusätzlich wird damit Druck ausgelöst, der auch „normalen“ Arbeitssituationen innewohnt. Da die Messgeräte teils die Interaktionsmenge beeinflussen (Verkabelung an Händen erschwert Tippen) ist dies bei der Aufgabenwahl/-aufbereitung zu beachten (weniger Schreibarbeit, einfache Klick-Aktionen).

Vorgehen: VP bekommen Aufgaben gestellt, die sie am Computer abarbeiten. Jeder Versuch hat die gleiche Versuchszeit, die der VP mitgeteilt wird, d.h. normale Arbeitssituation mit natürlichem Zeitdruck. Nach Ablauf der Zeit wird der Versuch beendet. Die VP darf nicht das Gefühl haben, sie hätte

⁵³ Ausarbeitung der Formulierung in Zusammenarbeit mit VL1

versagt (keine Schädigung der VP). Der Fokus sollte ergo nicht zu stark auf den Aufgaben liegen. Jede VP absolviert 2 Durchläufe (KG, EG), daher müssen Parallelformen der Aufgaben vorliegen, um diese sinnvoll vergleichen zu können. Hier sind auch SV wie Übungs- und Lerneffekte zu beachten:

Vorwand (Erweiterung): „Der erste Durchlauf ist beendet. Zur Verifizierung (Sicherstellung) der ersten Messung würden wir gerne noch einen zweiten Durchlauf machen.“

3.9.2 Ansätze und Auswahl

Nachfolgend werden relevante Ansätze vorgestellt und darauf aufbauend VP-Aufgaben gewählt. Eine Übersicht über die erste Ideenfindung liegt im ►Anhang B.5 vor.

Ansatz Leistungstest (Erhebung der Arbeitsleistung via Leistungstests): Neben der Ermittlung von Qualität und Arbeitsmenge (TH1), lassen sich über IQ-Tests zusätzlich IQ-Werte erheben, um unterstützende Vergleiche anzustellen (s. auch Wilson-Experiment ►3.1). Die Nutzung genormter, wissenschaftlich validierter Aufgaben gewährleistet u.a. Wiederholbarkeit und Ergebnisunabhängigkeit (Reliabilität, Validität, Objektivität). Zudem lassen sich kurze, voneinander unabhängige Einzelaufgaben (statt langer Knobelaufgaben oder aufeinander aufbauender Aufgaben) besser in den beschränkten zeitlichen Rahmen integrieren und problemlos erweitern/austauschen. (IQ-)Tests existieren in großer Fülle und erheben unterschiedliche (IQ-)Werte. Daher lassen sie sich schnell und einfach gemäß jeweiliger Grundgesamtheit (u.a. Alter, Nationalität, Bildungsstand) und Versuchsziel austauschen. Nachteilig wirken bei „echten“ Leistungstests der hoher Anschaffungsaufwand (Zugang nur für wissenschaftliches Personal) und Kosten. Um passende VP-Aufgaben zu wählen, werden im ersten Schritt Intelligenz, Intelligenztests und die Arten aufgeführt (s. [BN03][Re05][Wi12b][An08][Le05][Re05]).

Intelligenz: Sie umfasst geistige Fähigkeiten wie Denkvermögen, Auffassungsgabe, Logik, Urteilsvermögen, Kreativität. *Die* Intelligenz ist nicht definierbar, nur die Sonderformen intelligenten Denkens. Der Begriff „Intelligenz“ ist doppeldeutig, er meint zum einen die biologisch vorgegebene Kapazität (*flüssige Intelligenz*), die befähigt, neue Probleme zu lösen und nahezu unabhängig von Einflüssen durch Erziehungsträger (z.B. Schule) ist. Zum anderen meint er *kristallisierte Intelligenz*, d.h. erfahrungsabhängige Intelligenz als Kumulation von Fertigkeiten und Wissen. Intelligenz ist ein nicht direkt messbares soziales Konstrukt, IQ ist nie 100%ig feststellbar (zufällige Messfehler) und ein Intelligenztest allein umfasst nicht alle Teilbereiche der Intelligenz. Wesentliche Persönlichkeitsdimensionen werden auch gar nicht erfasst (z.B. emotionale Intelligenz). Nach moderner Ansicht setzt sich Intelligenz aus mehreren, relativ unabhängigen Faktoren zusammen. Intelligenztests bestehen daher oft aus einer Kombination von Einzeltests, die verschiedene Merkmalsbereiche untersuchen (u.a. Auffassungsgabe, Sprachbeherrschung, analytisches Denken, logisches Kombinieren, Merk- und Rechenfähigkeit, räumliches Vorstellungsvermögen).

Intelligenztests: Sie sind Leistungstests, für deren Bearbeitung meist eine Zeitbegrenzung vorliegt. Sie sind *kulturgebunden* (fordern sprachliche Kompetenz und kulturspezifisches Hintergrundwissen) oder *kulturfrei* (kulturfair, Denkleistung unabhängig von schulischer Vorbildung). „Die Qualität eines Tests [...] lässt sich an drei zentralen *Kriterien der Testgüte* festmachen: Objektivität, Reliabilität und

Validität [...]“ [BN03, S.193]⁵⁴ Ein Ergebnis (und das bekannteste) ist der IQ (*Intelligenzquotient*)⁵⁵. Der Binet-Simon-Test (1905) als erster brauchbarer Intelligenztests (Grundidee ist Basis moderner Tests), besteht aus diversen Einzelaufgaben (Subtests, u.a. Alltagsfragen/-probleme, logische/ mathematische Aufgaben). Man fasst bei Gesunden die Intelligenz oft erst als Generalfaktor auf (= g-Faktor). Danach kann Spezifizierung erfolgen, bei der zumeist Allgemeinfaktoren flüssiger und kristalliner Intelligenz unterschieden werden. Gebräuchliche Intelligenztests sind oft zeitraubend und liefern zu viele, besonders den Nichtfachmann verwirrende Informationen. Es gibt Tests, die eine (einzige) allgemeine Intelligenz betrachten und solche für Komponenten der Intelligenz (abhängig von der jeweiligen zugrunde liegenden Intelligenztheorie) u.a.

- *allgemeine Intelligenz*, eine ältere Theorie, die auf dem Generalfaktormodell⁵⁶ von C. Spearman basiert, z.B. Zahlenverbindungstests oder Progressive Matrizen-Tests,
- *kristallisierte (kristalline) und fluide Intelligenz* (Problemlösefähigkeit, Wissen),
- *Ravens Progressive Matrizen*, ein kulturunabhängiges, sprachfreies Verfahren, um Verzerrungen für VP anderer Kulturen auszuschließen,
- *verbale und praktische (Handlungs-)Intelligenz*, d.h. Intelligenz als „zusammengesetzte oder globale Fähigkeit des Individuums zielgerichtet zu handeln, rational zu denken und sich wirkungsvoll mit seiner Umwelt auseinander zu setzen“ [We56] S.13 zit. nach [Wi12b], Leistungsfähigkeit u.a. als Sprachverständnis, Arbeitsgedächtnis,
- *Mehrfaktorenkonzepte* wie z.B. Intelligenz-Struktur-Tests und
- *Kurztests* u.a. Mehrfachwahl-Wortschatz-Intelligenz-Test (MWT) von S.Lehrl.

Im angestrebten Versuch bedarf es eines einfachen, schnell durchführbaren, möglichst objektiven Verfahrens zur Messung des allgemeinen Intelligenzniveaus mittels Kurztests.

Auswahl für den Versuch: Die VP-Aufgaben, insbesondere zur Intelligenzmessung, sollen obigen Gütekriterien genügen (zuverlässige objektive Werte). Daher werden standardisierte, genormte und wissenschaftlich erprobte Tests bevorzugt. Hinsichtlich der Fragestellung bietet sich vor allem die Erfassung der allgemeinen Intelligenz an. Die technisch stark begrenzte Versuchszeit legt zudem die Verwendung von Kurztests nahe (z.B. MWT), unter Beachtung der Testökonomie (Praktikabilität, Zeitaufwand, Kosten). Um die Grundgesamtheit wenig zu begrenzen, sollten die Tests bildungsunabhängig und milieuneutral sein und eine große Altersgruppe umfassen (weitestgehende Unabhängigkeit von Alter, Bildung, Nationalität, körperlicher Verfassung). Da bereits einfache (elementare kognitive) Aufgaben ein Maß der Intelligenz ergeben, erscheinen simplere Testformen/-aufgaben in Anbetracht des beschränkten zeitlichen und organisatorischen Rahmens hier sinnvoll und erleichtern ggf. die Auswertung. Interessante allgemeine Intelligenztests sind u.a. der Grundintelligenztest, Raven-Matri-

⁵⁴ Objektivität (Anwenderunabhängigkeit) ist das Ausmaß, in dem Testergebnisse vom Testanwender unabhängig sind (s. [BN03]).

Reliabilität (Zuverlässigkeit) ist der „Grad der Genauigkeit, mit dem das geprüfte Merkmal gemessen wird.“[BN03, S.195].

Validität (Gültigkeit) gibt an, wie genau der Test misst, was er zu messen vorgibt (wichtigstes Testgütekriterium) (s.[BN03]).

⁵⁵ Aufgrund der Gefahr der Verabsolutierung des IQ als Etikett, wird diese Größe in der Fachsprache nicht mehr verwendet.

⁵⁶ Der allgemeiner Faktor der Intelligenz (Generalfaktor, g-Faktor) geht davon aus, Intelligenz basiere auf einer allgemeinen geistigen Fähigkeit (sog. Generalfaktor), die alle Intelligenzleistungen beeinflusst und etwa dem IQ entspricht. Er ist aber nur Orientierungswert für Intelligenzniveaus, es ist weiter die Ausprägung spezifischer Intelligenzfaktoren zu erfassen.

zentest, Snijder-Oomen non-verbaler Intelligenztest, Zahlenverbindungstest, MWT, dreidimensionaler Würfeltest. Sie zeichnen sich speziell aus durch geringe Abhängigkeit von Bildung und Kultur, kurze Testdauer, sehr ökonomische Bedingungen, geringes Testmaterial sowie zügige Durchführung und schnelle Erwerbung der Kompetenzen zur Durchführung. Nachteilig wirkt hier ggf., dass nur ein Ausschnitt der intellektuellen Fähigkeiten gemessen wird und nicht immer Bezug zur Leistungsfähigkeit im Alltag besteht. Unter Berücksichtigung genannter Faktoren, technischer, zeitlicher, organisatorischer Bedingungen (u.a. Verfügbarkeit, finanzieller Aufwand, Beschaffungsaufwand), bietet sich hier der MWT an, um Arbeitsleistung und IQ-Werte zu erheben (s. [Te12b][Te12a]).

MWT: Er bietet Eigenschaften, die für das eigene Vorhaben hohe Relevanz besitzen (s.[Le05,S.8 ff]).

Der MWT misst das allgemeine Intelligenzniveau, ist durch seelisch-geistige Störungen kaum beeinflusst, sehr objektiv, schnell und einfach abnehm- und auswertbar (Ökonomie), erfordert geringen Materialaufwand (Abarbeitung am Computer möglich), kann durch Hilfspersonal abgenommen werden (ermöglicht spätere Doppelblindversuche), ist sowohl bei Einzelpersonen als auch Gruppen (Einzelarbeitsplatz im UL) und mehrmals an der gleichen Person abnehmbar, d.h. wiederholbar in EG und KG ohne Lerneffekte. Zudem ist er robust, wurde an repräsentativen Zufallsstichproben von 1952 Erwachsenen der BRD geeicht und langjährig erprobt. Hinsichtlich Testgütekriterien gilt er als zuverlässig, objektiv, stabil, wiederholbar (Reliabilität) und verfügt über standardisierte Anweisung, Durchführung, Auswertung, Interpretation (Objektivität). Auch die kurze Bearbeitungsdauer von 4-6 Minuten und geringe Auswertungszeit sind ideal für die eigene begrenzte Versuchszeit. Zudem hält er sich finanziell im Rahmen. Der ursprüngliche MWT-B wurde später um den MWT-A erweitert, der als Paralleltest dient und den MWT damit ideal für die zweimalige Durchführung an derselben VP macht.

Die Anwendung des MWT unterliegt Einschränkungen (Ausschlusskriterien für Stichprobe ►3.12): Eichnorm bzgl. des Alters liegt bei 20-64 Jahren, VP sollten Muttersprache deutsch besitzen und möglichst keine Legastheniker sein (Lese-Rechtschreib-Störung). Sehschwäche, pharmakogene Akkomodationsstörung, Motivationsmangel und schwere seelisch-geistige Störung mindern Ergebnisse⁵⁷, fallen aber hier weniger ins Gewicht, da für EG und KG *dieselbe* VP genutzt wird (konstante individuelle Bedingungen). Neben dem MWT als primäre VP-Aufgabe, sollen noch weitere Leistungsausgaben verwendet werden. Da sich kein genormter IQ-Test für die verbleibenden ca. 5 Minuten anbietet, werden fragmentarisch Aufgabenteile aus anderen IQ-Tests [Re05] genutzt, die als Parallelform vorliegen, um Qualität und Arbeitsmenge zu erheben (aufgrund der unvollständigen Bearbeitung dieser Tests ist IQ-Ermittlung hier nicht möglich). Es heißt, Intelligenzleistung sei „weitestgehend unabhängig von Schwankungen der Stimmungsleistung“ [K183, S.18] aber zu verbissenes oder ängstliches Herangehen an Testaufgaben ist dennoch zu vermeiden. Daher wird *vor* dem Versuch ein Stimmungsfragebogen eingesetzt, um SV zu erkennen (►3.6.2).

Anmerkungen: Die MWT-Beschaffung erwies sich als aufwendig⁵⁸. Die schwierige Erhältlichkeit gewährleistet aber, dass wenige VP den Test vorab kennen, er wird nicht in Bildungseinrichtungen durchgeführt. In späteren Versuchen könnten sich auch sprachfreie Tests als sinnvoll erweisen, um VP mit sprachlichen Defiziten nicht auszuschließen bzw. Ergebnisse nicht durch diese unberücksich-

⁵⁷ Das Auftreten dieser SV wird z.T. via Fragebögen und Debriefing erhoben und ist in die Auswertung mit einzubeziehen.

⁵⁸ Wissenschaftliche IQ-Tests sind der Masse nicht zugänglich (Verteilung zentral und exklusiv über www.testzentrale.de). Der Erwerb des MWT-A und B erfolgte in Papierform (online-Bestellung). Die Aufgaben waren zu digitalisieren und bearbeitbar zu machen (interaktive Felder) für Bearbeitung am Computer.

tigte SV zu verzerren (z.B. Bochumer Matrizenest). Es boten sich hier aber keine sprachfreien Tests für die kurze Versuchszeit und unter Berücksichtigung finanzieller Möglichkeiten an. Zusammenfassend wurden folgende VP-Aufgaben für den Versuch ausgewählt ►Tab.9:

Tab. 9: Übersicht über die ausgewählten Aufgaben mit ihren jeweiligen Parallelförmigen, Kürzeln und Dauer

Aufgabe: genormte Leistungstests für KG und EG (Parallelförmigen), inkl. Probeaufgaben für MWT- und IQ-Tests mit Lösungsweg zur Eingewöhnung	Kürzel	Zeit
MWT von S. Lehl (Messung der allgemeinen Intelligenz) für VE und HE	MWT-A, MWT-B	4-5 Minuten
logisches Denken + rechnerische Intelligenz als Teil eines IQ-Tests für HE und VE	IQ-1, IQ-2	10,5 Minuten
Zusatz: Sprachliche Intelligenz + Technik und räumliche Vorstellung als Teil eines IQ-Tests für HE	IQ-Zusatz	9-10 Minuten

3.10 Organisatorischer Rahmen

3.10.1 Versuchsmaterialien

Die Versuchsdurchführung erfordert einige Materialien und organisatorische Überlegungen (u.a. finanzielle, personelle, räumliche, zeitliche, ggf. rechtliche, ethische Erwägungen ►Anhang B.1). Im Rahmen der Arbeit wurden folgende Unterlagen entwickelt ►Tab.10 (s. ►Anhang C, beigelegt CD).

Tab. 10: Übersicht über die entwickelten Versuchsmaterialien

Material	Inhalt
<i>Zeitplan (VE, HE)</i>	Datum/Uhrzeit der Versuche, anonymisierte VP-Spezifika, Abläufe, Reihenfolge EG-KG
<i>Zeitablaufplan</i>	Detaillierte Arbeitsschritte des Versuchs, d.h. Vorbereitung, Durchführung, Nachbereitung, inkl. Zuordnung von Verantwortlichkeiten und Vorgabe verbaler Anweisungen bei der Durchführung
<i>Anleitung für VL</i>	Zweck, Aufbau und Funktion, Auftreten, Aufgaben
<i>Anleitung für VP</i>	Instruktionen an VP per Email im UL kurz vor Versuchsbeginn
<i>Kürzelübersicht</i>	Übersicht über alle hier entwickelten Codes und Kürzeln und ihre Verwendung
<i>Protokoll-Vorlage</i>	Vorlage für VL im Beobachtungsraum
<i>VP-Akquirierung</i>	Separate Textvorlagen zur Akquirierung von VP per Email für VE und HE
<i>Stimmungsfragebogen</i>	Fragebogen zur Beantwortung vor der Durchführung zur Erfassung der Stimmung der VP und Angaben hinsichtlich physiologischer Signale ⁵⁹
<i>Versuchsfragebogen</i>	Fragebogen zur Beantwortung nach der Durchführung mit Angaben zur Person, über Umgang mit Computern und Kommunikationsmitteln, VP-Aufgaben, Versuchsablauf und Arbeitssituation
<i>Debriefing-Vorlage für VL</i>	Angaben zu Muttersprache, physiologische Signale ►Fußnote 59, Kommunikationsverhalten im Versuch, Versuchsziel, Feedback der VP, verbaler Hinweis zu Rechten und Pflichten der VP
<i>Probeaufgaben</i>	Probeaufgabe je Aufgabenkomplex mit Lösung zur Bearbeitung vor Versuchsbeginn
<i>Interaktive PDFs mit VP-Aufgaben</i>	MWT-Aufgaben (Parallelförmigen: MWT-A und -B), IQ-Aufgaben (Parallelförmigen: IQ-1 und IQ-2) für EG und KG und IQ-Zusatzaufgaben für die EG
<i>Kommunikationsübersicht</i>	Vorwand, Übersicht über Kommunikation/-smittel, vorgegebene (verbale, schriftliche) Texte für standardisierte uV, Übersicht aller vordefinierten Abläufe zur Manipulation der uV ►3.11.3

Weitere extern erstellte Unterlagen sind das Messprotokoll (Geräte, Material, Einverständnis-, Datenschutz-, Schweigepflichterklärung, Teilnahmefreiwilligkeit, Recht auf Versuchsabbruch, Dateneinsicht, -löschung, -korrektur, Anonymität, Widerrufsrecht) und Teilnehmerinformation, mit Angaben

⁵⁹ Der Frageteil zu physiologischen Signalen wurde außerhalb dieser Arbeit durch den technischen Experimentator erstellt.

zur Kleiderordnung, Geräten, Zweck, Dauer etc. für die Versendung an die VP vorab per Email. Es wurden zudem Codes eingeführt, um VP, Versuche und Materialien eindeutig zuzuordnen (► Abb.4).

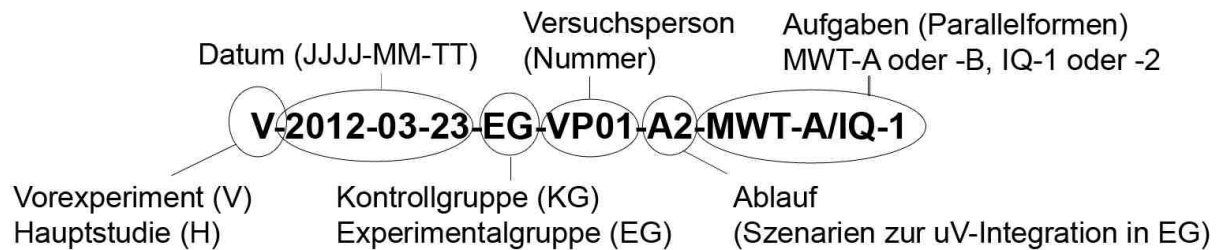


Abb. 4 Erklärung des verwendeten Versuchscode anhand eines Beispiels

3.10.2 Versuchsleiter

Idealerweise sind VL nicht vollständig über das Versuchsziel informiert (z.B. externe Experten), damit Ergebnisse weniger beeinflusst werden können (► 3.5.4). Im geplanten Versuch ist dies im Rahmen der Arbeit nicht möglich, sollte aber für nachfolgende Durchführungen überlegt werden. Davon abgesehen haben VL im geplanten Versuch Aufgaben und Rollen: „Je nach Art des Experiments hat der VL die Aufgabe den Versuchspersonen den Test zu erklären, die Messungen durchzuführen, SV auszugrenzen, den Versuchspersonen die Aufgabe zu präsentieren und für eventuelle Hilfestellungen oder Rückfragen zur Verfügung zu stehen.“ [Ge03, S.11-12] Neben diesen primären Aufgaben sollen sich VL hier vor der VP als (nur) die „Durchführenden“ ausgeben und eine dritte externe Person (Professor, Autoritätsperson) vortäuschen, die von außen Anweisungen gibt zum glaubhaften Einbinden der uV. Aufgaben der VL sind genau zu dokumentieren, um spätere Standardisierung und Prüfung des Ablaufs zu ermöglichen, SV zu vermeiden, Wiederholbarkeit zu gewährleisten und Aufgaben eindeutig für den jeweiligen VL vorab festzulegen (u.a. was wird gesagt, wer sagt es, Verhalten bei Fragen, Geräteeinstellung- und Bedienung, Protokollierung, Debriefing). Das VL-Auftreten kann in psychologischen Experimenten das VP-Verhalten beeinflussen (VL als Respekt einflößend oder Gegenteil). Da VL hier (nur) „Durchführende“ sind und möglichst angenehm und „ansprechbar“ wirken sollen, wird gepflegte aber „normale“ Tageskleidung empfohlen (kompetent aber nicht autoritär), genaue Vorgaben für VL ► Anhang C.9.

3.11 Vorgehen

Als nächstes sind das konkrete Vorgehen und ein geeigneter Versuchsplan zu entwickeln, unter Berücksichtigung theoretischer Überlegungen und praktischer Bedingungen vorangegangener Kapitel.

3.11.1 Allgemein

Ein Forschungsprozess beginnt mit einer klar und eindeutig formulierten Idee, die empirisch untersuchbar ist und von theoretischen Vermutungen bzw. Theorien ausgeht: *Erzeugen Kommunikation und gängige Kommunikationsmittel in einer Arbeitssituation zwangsläufig Ablenkung? Und inwiefern wirken sie sich auf die Arbeitsleistung aus?* (► 3.1, 3.2) Anschließend werden Vorhersagen entwickelt bzgl. möglicher Beobachtungen: *Es wird angenommen, dass durch Kommunikation und gängige Kommunikationsmittel in einer Arbeitssituation Ablenkung erzeugt und eine Veränderung der Arbeitsleistung bewirkt wird* (► 3.3). Nachfolgend wird die Beobachtung durchgeführt (hier: *quantitati-*

ves *Experiment im UL*) und mit den formulierten Vorhersagen (► 3.16) verglichen (► 4). Bei den Versuchen handelt es sich um klassische, szenariobasierte Tests mit festgelegten Aufgaben, die die VP zu lösen hat. Bei Übereinstimmung von Beobachtung und Hypothese kann die Theorie durch Ergebnisse erweitert und ausgebaut werden (► 6). Bei keiner Übereinstimmung, muss sie modifiziert werden (Hypothese gilt als nicht bewährt).

Zusammenfassung wichtigster Entscheidungen zum Vorhaben: Geplant ist eine experimentelle Untersuchung als Stichprobenuntersuchung in Form eines Laborexperiments im UL. Die Daten werden quantitativ durch Messgeräte des UL (und BS), VP-Aufgaben und schriftliche Fragebögen (subjektive Einschätzung der VP) erhoben. Zudem erfolgt indirekte Beobachtung der VP durch VL (Live-Videoübertragung in separaten Beobachtungsraum).

3.11.2 Planung

Die *Planung* beschäftigt sich mit dem konkreten Versuchsaufbau (► Tab.11), immer in Hinblick auf Versuchsthema und Umsetzung im technischen und organisatorischen Rahmen der FHB. Das Vorgehen in dieser Arbeit untergliedert sich in VE (► 4.4) und finale HE (► 4.5). Mithilfe der VE sollen Aufbau, Ablauf, Erhebungsmethoden, Geräte sowie Versuchsmaterialien mit VP des Hochschulumfeldes der FHB⁶⁰ erprobt werden, um mögliche Fehlerquellen aufzudecken/zu eliminieren und zu kontrollieren, inwiefern erhobene Daten zweckmäßig für die Hypothesenprüfung sind. Anschließend kann das Design bzgl. aufgetretener Fehler, SV, Geräte- und Aufgabenauswahl optimiert werden, um die HE mit VP aus der Wirtschaft durchzuführen. Zum Ende ist das Design ggf. nochmals zu optimieren, um repräsentative Untersuchungen (außerhalb der Arbeit) zu ermöglichen (Generalisierung).

Tab. 11 Eckdaten zur Versuchsplanung

Zusammenfassender Versuchsaufbau	
<i>Anwesende bei Durchführung</i>	<i>Im UL:</i> eine VP (Einzelversuche) und ein (technischer) VL (<i>hier:</i> VL1) <i>Im getrennten Beobachtungsraum:</i> ein VL (Beobachtung, Protokollierung, uV-Manipulation) (<i>hier:</i> VL2)
<i>Durchläufe je Versuch</i>	2 Durchläufe mit <i>gleicher</i> VP (1mal KG, 1mal EG) mit identischer Laborsituation und Ablauf, gleichen Skalen/Instrumenten, ruhigen konstanten Untersuchungsbedingungen, Parallelformen der Aufgaben, uV-Manipulation in EG (Kommunikation/-smitteln) → <i>Vorteil: Ökonomie, Parallelisierung der SV der VP</i>
<i>Versuchsziel</i>	Getarnte Versuchssituation (VP weiß um Experiment aber kennt Versuchsziel nicht)
<i>Schlüsselworte</i>	<i>Nicht zu verwendende Begriffe:</i> Ablenkung, Konzentration, Kommunikation(-smittel), Leistung → damit VP nicht (versuchen), „wahres“ Versuchsziel aufzudecken, ihr Verhalten (un)bewusst zu ändern, um Neutralität zu wahren und VP nicht durch Wortwahl (unbewusst) zu beeinflussen
<i>Laborumgebung</i>	Möglichst „normale“ Arbeitssituation (natürlicheres VP-Verhalten), d.h. keine Ausnahmesituation (z.B. Prüfungssituation, unnatürlicher zusätzlicher Stress/Druck), keine offene Beobachtung der VP
<i>Instruktionen</i>	Schriftlich (per Email): Zweck (Vorwand), Ablauf, Eckdaten (einfache, verständliche Formulierung)
<i>Aufgaben</i>	- Leistungstests (Parallelform für EG und KG d.h. gleiche Art und Schwierigkeit, Lerneffekte beachten) - Bearbeitung durch VP am Monitor, Zeitvorgabe (identisch in D1 & D2), ohne zusätzliches Schreibzeug
<i>Versuchszeit pro Durchlauf</i>	11 Minuten (identisch für D1 und D2 und alle VP in VE und HE), VP ist Versuchszeit bekannt (Zeitdruck, ermutigt zielgerichtete konzentrierte Bearbeitung, ggf. Zeitmanagement und Priorisierung)
<i>Durchführung (Ablauf)</i>	<i>Vorexperimente:</i> Erprobung, Verbesserung und Auswahl der experimentellen Bedingungen (<i>aufgrund der geringen VP-Anzahl im VE besteht die Möglichkeit, dass SV und Fehler trotzdem nicht aufgedeckt werden</i>) <i>Hauptexperiment:</i> Prüfexperiment zur Hypothesenprüfung

⁶⁰ Für VE bietet sich eine leicht erreichbare Grundgesamtheit an, für HE sind VP gemäß der Hypothese zu wählen (► 3.12).

VP und Anzahl	<ul style="list-style-type: none"> - im VE: aus dem Hochschul Umfeld der FHB (ökonomische Gründe, leicht erreichbare Gesamtheit), N = 4 - im HE: aus der definierten Grundgesamtheit (Akquirierung aus der Wirtschaft), N = 12 (<i>angestrebt</i>) - VP-„Pfleger“: faire Vorab-Information, Begrüßung, angenehme Situation, „wahres“ Ziel aufdecken (Debriefing), Verabschiedung, Anwerben <i>Freiwilliger</i>, unangenehme Aspekte nicht verschweigen, Anonymität und Datenschutz gewährleisten, Zumutbarkeit der Untersuchung (u.a. Dauer, Aufgaben)
---------------	---

Anmerkung: Das Experiment ist abzubrechen, wenn die VP es wünscht oder die Aufgaben vollständig bearbeitet hat. Letzteres sollte nicht eintreten, damit die Arbeitszeit bei allen VP identisch und vergleichbar ist. Die Bearbeitung der VP-Aufgaben wird im VE erprobt und Aufgaben ggf. erweitert. Zudem sollte vermerkt werden, ob VP das „wahre“ Versuchsziel bereits während des Versuchs (er)kannten (Ergebnisverzerrung). Dies soll mittels Debriefing (Nachgespräch) überprüft werden (►Anhang C.12). Der Versuch ist vorab genau zu beschreiben (inkl. aller Arbeitsschritte) und während der Durchführung zu protokollieren (inkl. Fotos/Skizzen des Aufbaus), um Wiederholbarkeit (und Gültigkeit) zu gewährleisten und SV aufzudecken. Anschließend erfolgt die Auswertung erhobener Daten. In den VE ist u.a. zu ermitteln, wie viel Zeit Vorbereitung, Durchführung und Nachbereitung beanspruchten (Details zu Schritten ►Anhang C.6). Die uV-Manipulation muss vorab genau ausgearbeitet und bei der Durchführung protokolliert werden (►3.11.3), um kontrollierte wiederholbare Abläufe zu gewährleisten. Es sind auch getrennte Versuchs- und Beobachtungsräume angedacht, um objektive kontrollierte Beobachtung zu erleichtern, natürliches VP-Verhalten zu fördern und störende VL-Einflüsse zu mindern. Im Rahmen der Arbeit muss aber VL1 durchgehend im UL zugegen sein, um Geräte zu bedienen und zu kontrollieren.

3.11.3 Versuchsplan

Für die Veränderungshypothesen, die hier zu prüfen sind, kommen folgende Pläne in Frage: *experimentelle* oder *quasiexperimentelle Versuche*, *Veränderungshypothesen für Entwicklungen* und für *Zeitreihen*. Aufgrund des vorgegebenen Versuchsortes im Labor, besonderer Eignung zur Hypothesenprüfung und Sicherung der Gütekriterien sowie weiterer o.g. Gründe, bietet sich die *experimentelle* Behandlung an. Gemäß ►3.5.3 eignen sich *elementare* Versuchspläne mit *einer* uV, die in *zwei* Stufen variiert wird. Da die Stichprobe zu klein sein wird, um vergleichbare Ausgangsbedingungen der VP zu gewährleisten und Randomisierung sicherzustellen (►3.12), sollen alle VP beide Bedingungen (mit/ohne uV) als EG und KG absolvieren (Versuchs-Kontrollgruppenplan, Kontrolle der VP-Unterschiede, echte Vergleichsmöglichkeit). So sind auch keine Pretests⁶¹ notwendig. Die KG dient der SV-Kontrolle und Vergleich mit der EG. EG und KG finden direkt nacheinander statt, um VP-Ausfälle zu vermeiden (unvollständige Datensätze). (s. [BN06, S.547])

Fazit: Prüfung der Veränderungshypothesen via Laborexperiment mit EG und KG als elementarer Versuchsplan.

⁶¹ Pretests sind bei zu kleinen Stichproben (<30) durchzuführen, die keinen zufälligen Ausgleich personenbedingter SV in EG und KG gewährleisten können oder bei zweifelhafter Randomisierung. Pretest bedeutet, dass sowohl EG als auch KG in Vortests mit dem Treatment (uV) behandelt werden, um ggf. Unterschiede der Ausgangssituation aufzudecken. (s. [BN06])

Versuchsplan: Um intern valide Untersuchungen mit kausaltheoretischer Erklärung zu ermöglichen, sind experimentelle Designs zu nutzen. Es bietet sich die Untergruppe *Within-Design* an, d.h. dieselbe VP wird unter verschiedenen uV-Ausprägungen untersucht und jeweils aV erhoben. Sie gestattet Vergleiche der EG und KG einer VP (intraindividuelle Varianz). Within-Designs sind Messwiederholungspläne, hier eignet sich das einfachste Design, der 2-Stichprobenversuchsplan mit Messwiederholung (1 Versuchsgruppe wird zu 2 Messzeitpunkten untersucht, d.h. einmal ohne und einmal mit uV). Messwiederholung ermöglicht, Sekundärvarianz zu kontrollieren, personelle SV konstant zu halten (Vergleich verschiedener Messwerte *einer* VP), situative SV sind jedoch zusätzlich zu kontrollieren. Vorteile dieser Pläne sind kleine VP-Anzahl, geringe Datenstreuung und klarer nachzuweisende experimentelle Effekte durch Wegfall der Between-Varianz (interindividuelle SV bei verschiedener VP in KG und EG). Nachteile von Messwiederholung sind speziell SV, die bei gleicher Bedingungsreihenfolge für alle VP auftreten (primär Positions-Effekt, Carry-Over-Effekt ▶3.5.4), sowie steigender Zeitaufwand und Übertragungsfehler (vorausgehende Bedingungen beeinflussen nachfolgende). Um Letzteren vorzubeugen, sollten Zeitabstände zwischen experimentellen Bedingungen groß gehalten und ihre Reihenfolge ausbalanciert sein. Große zeitliche Abstände sind hier praktisch nicht möglich, jedoch eine variierte Bedingungsreihenfolge. Zusätzlich bieten sich *multivariate Versuchspläne* an (Untergruppe des Experimentellen Designs), die sich durch mehrere aV kennzeichnen (▶3.5.3), d.h. eine globale aV (Konstrukt) wird anhand diverser Einzelmaße operationalisiert, um das Konstrukt facettenreicher zu erfassen und wechselseitige Beziehungen der aV untereinander zu berücksichtigen und aufzudecken. So lassen sich Ergebnisse besser verallgemeinern (hebt externe Validität). Zugleich erhöht sich aber der Aufwand und Ergebnisse sind nicht immer besser oder aufschlussreicher als bei univariaten Designs. Die hier gewählten aV sind jedoch komplex und lassen sich durch mehrere operationale Indikatoren sinnvoller erfassen. Es wird daher eine Mischform aus Messwiederholungsplan und multivariatem Versuchsplan gewählt⁶².

<p>Versuchsplan: Experimentelles Design (gestattet intern valide Untersuchung mit kausaltheoretischer Erklärung)</p> <p><i>Untergruppen:</i> Messwiederholungsplan (Within-Design) und multivariater Versuchsplan (mehrere aV) als Mischform</p> <p><i>Umsetzung:</i> Unifaktorieller multivariater 2-Stichprobenversuchsplan mit Messwiederholung (vgl. Vorher-Nachher-Messung)</p> <p><i>Vorteile:</i> Geringe VP-Anzahl notwendig, geringe Datenstreuung, Vergleiche innerhalb EG-KG (Erfassung intraindividuelle Varianz, Personenvarianz-Kontrolle), Wegfall der Between-Varianz (personelle SV konstant), Konstrukt facettenreicher erfassbar d.h. Ergebnisse besser verallgemeinerbar (steigende externe Validität)</p> <p><i>Nachteile:</i> Größerer Aufwand (speziell physiologische Messung), Übertragungsfehler, Positions-Effekt, Carry-Over-Effekt</p>
--

Ausbalancierung: Die Reihenfolge der experimentellen Bedingungen wird variiert (Abfolge der uV-Arten) und soll sich möglichst nicht wiederholen oder gleichmäßig verteilt sein, um Effektabhängigkeit von einer bestimmten Variante zu vermeiden (höhere Ergebnisrepräsentativität). Im 2-Stichprobenplan mit Wiederholungsmessung bietet sich ferner an, EG-KG-Reihenfolge für VP zu variieren,

⁶² Entgegen „normaler“ Messwiederholungspläne wird eine VP aus zeitlichen Gründen hier nicht zu zwei verschiedenen Messzeitpunkten untersucht. Für weiterführende Versuche sind auch mehrfaktorielle Pläne in Betracht zu ziehen (mehrere uV z.B. Organismusvariablen wie VP-Alter). Bei großer VP-Anzahl bieten sich auch 2-Stufen-Randomisierungspläne an (verschiedene EG und KG), die *unbekannte* SV kontrollieren (ohne Übertragungseffekte) und intern valide sind.

um SV durch Gruppenreihenfolge aufzudecken/auszubalancieren. Es wird vermutet, dass KG in D1 der VP Sicherheit gibt, sie an die Situation gewöhnt, Misstrauen, Stress und Angst abbaut (ungestörtes Arbeiten ohne negative Folgen) und so natürlicheres Verhalten begünstigt. Andererseits werden verzerrte Biosignale befürchtet (Normalwertmessung in KG), da VP in D1 ggf. aufgeregt sind („ungewohnte“ Experimentsituation, u.a. Stress, Aufregung, Angst vor Unbekanntem). Bei uV-Einführung in D1 (EG) wird befürchtet, dass VP verfrüht auf das Versuchsziel aufmerksam oder/und in erhöhten Stress-/Frustrationszustand versetzt werden und dies Ergebnisse in D2 beeinflusst. Daher wird das Tauschen der Gruppenreihenfolge in VE erprobt. Bei großer VP-Anzahl beugt zufällige Bedingungs-zuteilung zusätzlichen SV vor. Hier ist jedoch nur *unvollständiges Ausbalancieren* mit zufällig gewählten Abfolgen möglich (wenige VP, nicht alle möglichen Abfolgen umsetzbar). Aufeinanderfolgende VP sollten nicht exakt gleichen Bedingungen ausgesetzt sein.

Möglichst ausgeglichenes Design hinsichtlich folgender Bedingungen: Geschlecht, Alter und Position(Tätigkeit) der VP, Art und Schwierigkeitsgrad der Aufgaben in KG und EG (Parallelformen), Reihenfolge der Kommunikation/-smittel (zufällige Zuteilung, unvollständiges Ausbalancieren), Reihenfolge der EG und KG bei unterschiedlichen VP

Unterschiedliche Zeitpunkte und Anzahl für uV-Auftreten wäre wünschenswert (falls aV-Effekte von bestimmter uV-Abfolge abhängen), aber aufgrund der kurzen Versuchszeit hier nicht möglich. Die uV sollte nicht zu selten auftreten, da sonst die Wirkung nicht eindeutig zuzuweisen ist aber auch nicht zu oft, um nicht unrealistisch⁶³ zu wirken, ggf. Misstrauen zu erregen und die Arbeitszeit zu stark zu beeinträchtigen (Vorgaben, inhaltliche Aspekte und Abläufe s. ► Anhang C.4, C.5).

Ablauf: Es wurden 3 Abläufe definiert, die sich in der Reihenfolge der Kommunikation (Inhalt) und Kommunikationsart (Medium) unterscheiden, um alle VP gleichen Bedingungen auszusetzen aber Reihenfolge- und Positioneffekte zu vermeiden. Der Abfolgefaktor ist als „Random Factor“ konzipiert, d.h. aus allen möglichen uV-Abfolgen sind einige zufällig gewählt (s. [BN06, S.550]). Neben geplanter Kommunikation von außen, darf die VP auch selbst initiiert kommunizieren (u.a. Telefon, Chat, Email). Dies kann nicht direkt von VL manipuliert werden, ist jedoch zu protokollieren (► 3.8).

Zeitpunkt 1: Aufgabenbearbeitung am Rechner als KG, d.h. ohne uV-Einsatz (keine Kommunikation/-smittel)
Ausgangswerte messen („Normalwerte“): Biosignale (UL, BS), Arbeitsmenge, Qualität, ggf. IQ, Fragebögen
Zeitpunkt 2: Aufgabenbearbeitung am Computer als EG, d.h. mit uV-Manipulation (mit Kommunikation/-smittel)
Beobachtung und Messung von VP in Situation mit Kommunikation (Erwartung: Biosignale verzeichnen erhöhte Ablenkung, Arbeitsmenge und/oder Qualität und ggf. IQ-Werte unterscheiden sich von Zeitpunkt 1)

3.12 Versuchspersonen

Quantitative Experimente erfordern die Definition der Grundgesamtheit: Art, Anzahl, Merkmale (Stichprobenbeschreibung, Repräsentativität), sonstige Merkmale sind summarisch zu nennen, inkl. Rekrutierungsart, Verweigerungsbegründung und Überprüfung von Machbarkeit und Praktikabilität.

Auswahl: Die Untersuchung der gesamten interessierende Population und vollständige Randomisierung sind hier organisatorisch nicht möglich, unökonomisch und entsprechen nicht dem gesetzten Ziel

⁶³ Es ist glaubhaft zu machen, dass neben VL weitere (Autoritäts-)Personen beteiligt sind, die sich extern zuschalten und ggf. kurzfristige Änderungen veranlassen (rechtfertigt Kommunikation per Email, Telefon etc.). Zusätzlicher kurzer Smalltalk (Frage nach Wohlbefinden) soll Ängste/Hemmungen abbauen und Natürlichkeit unterstützen (Misstrauen vermeiden).

(►1.2). Es wird nur eine kleine Stichprobe untersucht und die Grundgesamtheit aus praktischen Gründen auf eine Submenge eingegrenzt, sodass Ergebnisse maximal für diese gelten können. Zudem sind Auswahlfehler zu beachten (kein Zufälligkeitsprinzip, nur anwesende Personen) sowie Freiwilligkeit der Teilnahme als SV⁶⁴ (nur Freiwillige), was Stichproben systematisch verzerren kann⁶⁵.

Nicht-probabilistische Stichprobe: Sie eignet sich bei unbekanntem oder unkontrollierbarem Auswahlwahrscheinlichkeiten, Elemente werden nicht über objektive Zufallsmechanismen entnommen sondern mittels bewusster Selektion nach festgelegten subjektiven Kriterien (unzulässige Verallgemeinerung auf Population). Sie sind weniger aufwändig als echte Zufallsstichproben (geringere Kosten). Da die Untersuchung im Rahmen dieser Arbeit eher „probierenden“, experimentellen Charakter besitzt und gemäß der Zielsetzung noch keine repräsentativen Ergebnisse fordert (►1.2), ist dieses Verfahren hier praktikabel. Aufgrund mangelnder Ressourcen und des leichteren Zugangs wird die Gelegenheitsstichprobe präferiert, die sich aus Freiwilligen zusammensetzt (nicht-repräsentative Stichprobe). Ihr Fokus liegt nicht auf Verallgemeinerung sondern Erfassen der aV-Veränderung nach uV-Manipulation, was mit dem Anliegen der Arbeit einhergeht. Die praktikable ökonomische VP-Rekrutierung dieser Ad-hoc-Stichprobe ist günstig. Einschränkungen bei der Generalisierung lassen keine allgemeinen Aussagen zu (geringe externe Validität, VP möglicherweise atypisch für Gesamtpopulation). Da Ziel dieser Arbeit nicht die Erhebung repräsentativer Daten ist, sondern Entwicklung des Forschungsdesigns inklusive Bewertung und Optimierung durch erste Versuche, ist im Rahmen dieser (experimentellen) Studie eine kleine, leicht zu erreichende Gesamtheit aus Freiwilligen der definierten Grundgesamtheit zulässig, fälschliche Generalisierung aber zu vermeiden. In späteren Replikationsstudien sind repräsentative Untersuchungen mit größeren Zufallsstichproben⁶⁶ aber unerlässlich, um verallgemeinerbare Aussagen zu erlauben. Für Untersuchungen mit kleiner VP-Anzahl bieten sich Extreme und typische Fälle an, um Unterschiede oder Gemeinsamkeiten zu ermitteln, z.B. verschiedene Branchen, die sich im Idealfall bereits in Kommunikationsverhalten und ihrer Struktur unterscheiden. Für die HE wurden Medien-/IT-Firmen⁶⁷ (freie Wirtschaft, primäre Ausrichtung IT/neue Medien) und Ämter (Behörden/staatliche Einrichtung, Computer und neue Medien sekundär) gewählt. VP der VE setzen sich aus ökonomischen Gründen aus verfügbaren VP der FHB zusammen (Mitarbeiter, Studenten). In VE sind u.a. Geräte, Ablauf, Reihenfolgen (EG/KG, uV), Materialien,

⁶⁴ Freiwillige entsprechen nicht Populationsdurchschnitt u.a. bessere schulische Ausbildung, bessere Schätzung des eigenen sozialen Status, höhere Intelligenz/Bedürfnis nach sozialer Anerkennung/Geselligkeit, unkonventionellere Einstellung, meist weiblich, weniger autoritär oder konformistisch ([RR75, S.1955ff] ergänzt durch [EB77], zit. nach [BN06, S.73]).

⁶⁵ Solche Fehler lassen sich vermeiden, indem VP-Akquirierung im Idealfall per Post erfolgt, via Adresskartei und Namen, die zufällig gezogen werden. Dies beansprucht jedoch höheren zeitlichen, organisatorischen und finanziellen Aufwand.

⁶⁶ z.B. Klumpenstichproben (probabilistische Stichprobe), u.a. Liste aller Firmen erstellen (via Branchenbuch) und zufällig Firmen ziehen. Durch geringe Kontrolle über Mitarbeiterzahl in gezogenen Firmen (Stichprobengröße ist Zufallsvariable) können angetroffene Klumpen ggf. zu groß für vollständige Erhebung sein (ggf. mehrstufige Stichprobe d.h. Klumpen nicht vollständig erfassen). Mind. 30 VP sind erforderlich, d.h. „Für praktische Zwecke können wir davon ausgehen, dass die Mittelwertverteilung auch für extrem von der Normalität abweichende Grundgesamtheiten hinreichend normal ist, wenn $n \geq 30$ ist.“ [BN06, S.411]

⁶⁷ Es „liegt der Verdacht nahe, daß Untersuchungen über entwicklungsbedingte, sozialisationsbedingte und durch das Altern bedingte Prozesse im kognitiven und intellektuellen Bereich, die vorwiegend mit Studenten durchgeführt werden, zu falschen Schlüssen führen[...] Die Fragwürdigkeit humanwissenschaftlicher Forschungsergebnisse [...] erhöht sich um ein Weiteres, wenn man in Rechnung stellt, daß an diesen Untersuchungen nur ‚freiwillige‘ Studenten teilnehmen.“ [BN03, S.78-79]

Aufgaben, Erhebungsmethoden, Daten (korrekte Messung, sinnvolle Ergebnisse) zu prüfen und Problemen und SV aufzudecken (u.a. Synchronisierung aller Geräte). VP sind bzgl. beeinflussender Faktoren wie Geschlecht (2 männlich, 2 weiblich), Position (2 Mitarbeiter, 2 Studenten) auszugleichen. Zur Aufnahme in die Studie für die HE müssen VP einige wichtige Kriterien erfüllen.

Einschlusskriterien⁶⁸:

- VP sind in der Lage und willigen ein, an der Untersuchung im UL teilzunehmen (Verkabelung, Video- und Audioaufzeichnung, Teilnahmebedingungen: freiwillig, unbezahlt) und unterschreiben Einverständnis- sowie Datenschutzerklärung
- VP arbeiten primär am Computerarbeitsplatz⁶⁹ s. Schwerpunktsetzung ► 3.2 („Die tägliche Verwendung des Computers als Arbeitsmittel wird bei allen Teilnehmern als Bedingung vorausgesetzt.“ ► Anhang C.1, C.2)
- Alter liegt zwischen 18 – 65 Jahre (offizielles Renteneintrittsalter) und umfasst alle erwachsenen berufsfähigen Menschen, keine Kinder oder Rentner (*hier*: Stichprobenentnahme direkt aus Firma, Erhebung der Arbeitszeit mittels Fragebogen)
- Mischung der Berufsposition ist erwünscht aber nicht zwingend (Erhebung des Angestelltenverhältnisses via Fragebogen)
- Geschlecht umfasst männliche und weibliche VP (möglichst ausgeglichenes Verhältnis in der Stichprobe)
- VP stammen möglichst ausgeglichen aus IT/ Medien-Branche (freie Wirtschaft) oder Amt (Behörde ohne IT-Fokus)
- VL sollten den VP nicht vorher bekannt sein, um Verzerrung durch verändertes, voreingenommenes Verhalten zu meiden
- VP besitzen formale Bildung (Erhebung des höchsten bisherigen Bildungsabschlusses mittels Fragebogen)
- VP sind vertraut mit Email und PDF-Bearbeitung, nach Möglichkeit auch mit den genutzten Medien (*hier*: Skype, Email)⁷⁰
- VP haben keine Lese-Rechtschreibschwäche und die kognitiven Fähigkeiten, um Aufgaben und Instruktionen zu verstehen
- VP geben Einverständnis, keine Details nach Versuchende preiszugeben und unterschreiben Schweigepflichterklärung
- VP sollten sich vor der Untersuchung nicht in einem extremen Gemütszustand befinden sondern in gutem Allgemeinzustand (Erhebung der Stimmung via Fragebogen vor Versuchsstart, ggf. nachträgliches Aussortieren „extremer“ VP)

Soziodemographische, psychographische Faktoren sowie die Nutzung von Computern und den definierter Kommunikationsmittels werden im Fragebogen erhoben, um Verzerrungen⁷¹ aufzudecken. Ein (nachträglicher) VP-Ausschluss erfolgt, liegt eines der folgenden Kriterien vor.

Ausschlusskriterien:

- Lese-Rechtschreibschwäche, keine ausreichenden kognitiven Fähigkeiten zur Aufgabebearbeitung, „extreme“ Stimmung
- unfreiwillige Teilnahme, Verweigerung der Unterschrift für Einwilligung-, Datenschutz- oder Schweigepflichterklärung
- VP sind nicht vertraut mit Arbeit am Computer (keine Erfahrung mit Computerarbeit, kein Büroarbeitsplatz am Computer)
- VP liegen außerhalb der definierten Altersbegrenzung z.B. Kinder, Rentner
- VP sind berufsunfähige oder stammen nicht aus den definierten Branchen
- VP benötigen sehr starke Kontaktlinsen oder Brillen, was sich nicht für das Eyetracking und die MWT-Aufgaben eignet
- Muttersprache ist nicht deutsch (Erhebung im Debriefing), da VP-Aufgaben, Fragebögen, Instruktionen und speziell MWT dies erforderlich machen und genormte Texte hier ausschließlich in Deutsch vorliegen
- VP sind mit den Versuchsdurchführenden vertraut (möglichst Anonymität gefordert)

VP-Akquirierung: Um die Anzahl an Verweigerern zu reduzieren⁷² (s.[BN03]), sind VP individuell und persönlich anzusprechen (mündlich oder schriftlich) und das Vorhaben, soweit bei getarntem Versuchsziel möglich, inhaltlich zu erläutern, mit Angaben darüber, wem es zugute kommt, speziell

⁶⁸ Ein-/Ausschlusskriterien wurden via Akquirierungsmail, Fragebogen, Einführungs- und Nachgespräch gefiltert / erhoben.

⁶⁹ Begrenzung auf Büromitarbeiter (Untersuchung *aller* interessierenden Arbeitssituationen *aller* Menschen hier unmöglich)

⁷⁰ Kenntnisse werden hier im Gespräch vor Untersuchung erhoben und ggf. Programme nachträglich erklärt und vorgeführt

⁷¹ Befürchtete Gefahren sind Verzögerungen bei Aufgabebearbeitung durch „Lernen des Mediums“ oder Ignorieren der uV (Unkenntnis).

⁷² „Bereits bei der Anwerbung werden Erwartungshaltungen erzeugt, die die Reaktion der Untersuchungsteilnehmer auf die spätere Untersuchungssituation nachhaltig beeinflussen. [...] Teilnehmer [müssen den Experimentatoren] persönliches Vertrauen entgegenbringen [um] absichtliche Täuschungen und bewußte Fehlreaktionen [zu] verhindern [...]“[BN03, S.75]

wenn VP selbst profitieren. Zusätzlich ist mitzuteilen, wann und wie VP eigene Ergebnisse erfahren. Günstig wirkt sich auch Anwerbung durch Personen mit hohem sozialem Status aus. Für den Versuch eignet sich die direkte Anfrage per Email an Firmen als Erstkontakt, bevorzugt über direkte Ansprechpartner und persönlich, unter Mitteilung wichtiger Teilnahmebedingungen (Ein-/ Ausschlusskriterien, Zweck, Vorgehen, Bedingung, Mitarbeiter). Um VL-Fehler und andere SV vorzubeugen und den Versuch zu standardisieren (Wiederholbarkeit), erfolgt die Anwerbung für alle VP identisch.

Untersuchungsmerkmale (Verweigererrate reduzieren, Stichprobenverzerrung entgegenwirken): Folgende Merkmale erhöhen die Teilnahme (s. [BN06]) - Interesse der VP für den Versuchsgegenstand, größer eingeschätzte Geltung der Untersuchung, Entlohnung in Form von persönlichen kleinen Geschenken und Aufmerksamkeiten statt Geld *vor* Entscheidung zur Teilnahme, persönliche Bekanntschaft mit der anwerbenden Person und öffentliche Unterstützung der Untersuchung, d.h. Teilnahme gehört „zum guten Ton“. Das eigene Vorhaben ist vorausgehend zu erklären, um Interesse zu wecken (*hier*: Anwerbungsmail vorab), die Relevanz hervorzuheben (*hier*: „um in nachfolgenden wissenschaftlichen Untersuchungen das Arbeitsverhalten von Arbeitnehmern in organisierte, gesunde Strukturen zu überführen“ ► Anhang C.1, C.2), Belohnung nach Versuchende in Aussicht zu stellen (*hier*: Erfrischung⁷³) und die Unterstützung durch die FHB zu unterstreichen. Ein Stimmungsfragebogen soll über die subjektive Befindlichkeit der VP Aufschluss geben (Untersuchungsbereitschaft).

Versuchsbedingte Verweigerung: Die hohe zeitliche Belastung mit ca. 1,5h pro VP, Versuchsart mit Verkabelung, teils Entkleidung und Rasur, ohne finanzielle Entschädigung sowie die Durchführung in den Sommerferien (Urlaubssaison, Semesterferien) erhöhen hier die Ausfall- und Verweigerungsrate.

Anmerkung: Die Stichprobe ist begrenzt und VP-Anzahl gering aufgrund u.a. technischer, praktischer und zeitlicher Grenzen (Datenmenge, wenige Freiwillige). VP sind nicht nach u.a. Alter oder Geschlecht selektiert (Kovariablen, beeinflussen ggf. Leistung oder Konzentration). In repräsentativen Untersuchungen sollte Randomisierung erfolgen, um Drittvariablen auszubalancieren. Weiterführend ließen sich VP auch nach Kovariablen gruppieren, um ihre Effekte einzeln zu erheben / prüfen. Insgesamt erlaubt das hier verwendete nicht-probabilistische Verfahren (► 3.14) keine Hypothesenprüfung⁷⁴ und Ergebnisse sind ohne nachfolgende Versuche mit (Zufalls-)Stichproben nicht verallgemeinerbar.

Aus diesem und vorherigen Kapiteln ergibt sich folgende zusammenfassende Übersicht:

Vorgehen: (Experimentelle) Studie an einer kleinen, leicht zu erreichenden Gesamtheit (Stichprobenuntersuchung)

Grundgesamtheit/Population (Submenge): N = Büromitarbeiter (primär Computerarbeitsplatz)

Stichprobe (HE): n = Medien/IT-Firma und Amt, *nicht-repräsentativ* **Stichprobengröße:** (Angestrebt) N = 12

Stichprobe (VE): Aus ökonomischen Gründen nur Freiwillige aus FHB-Umfeld (für Testzwecke) **Stichprobengröße:** N = 4

Auswahl: Nicht-probabilistische „anfallende“ *Gelegenheitsstichprobe*, Freiwillige, ökonomisch, nicht generalisierbar, praktikabel, Replikationsstudien mit repräsentativer größerer Zufallsstichprobe erforderlich für zulässige Hypothesenprüfung

Akquirierung: Direktkontakt (Ansprechpartner, Firmen-Email) **Quelle:** Städte Brandenburg, Potsdam (regional)

⁷³ Buffet nach Versuchsbeendigung (räumliche Trennung von UL, damit „fertige“ VP sich nicht mit „neuen“ austauschen).

⁷⁴ „Repräsentationsschlüsse ohne Zufallsstichproben sind grundsätzlich fragwürdig und unzuverlässig.“ [Fa00, S.29]

3.13 Messgrößen und Auswertung

Nachfolgend werden kurz Messgrößen und Auswertungsmethoden für die hier durchgeführte Erhebung zu TH2 dargestellt⁷⁵ und auswertungsrelevante Aspekte beschrieben. Allgemein bieten sich zur Hypothesenprüfung statistische Grundlagen an (Experiment, Fragebogen), dazugehörige statistische Auswertungsverfahren sind aber erst bei repräsentativen Stichproben sinnvoll (u.a. *Signifikanztests* s. [BN06, S.491-501]). Ergebnisse hier durchgeführter Versuche sollen wertneutral beschrieben und dargestellt werden (z.B. Zusammenfassung erhobener Daten als Grafik, Darstellung des Datenmaterials durch Tabellen) und dienen vorrangig der Bewertung und Optimierung des Forschungsdesigns und nur sekundär der Hypothesenprüfung (Tendenzen). Es werden maximal Vorschläge zur möglichen Quantifizierung der Merkmale, ihrer Messung und Darstellung aufgezeigt. VP sind nicht repräsentativ für die zugrunde liegende Population.

Satz für TH1 (Erwartung): Bei VP in der EG liegt signifikant höhere (physiologische) Ablenkung vor als in der KG.

Vorgehen: Messung der Ablenkung (aV: Z_n) bei allen n VP (Vergleich EG und KG)

Satz für TH2 (Erwartung): Bei VP in der EG ist die Arbeitsleistung (Qualität, Bearbeitungszeit, ggf. IQ) ungleich der KG.

Vorgehen: Messung der Arbeitsleistung (aV: Y_n) bei allen n VP (Vergleich EG und KG)

Versuchsfragebogen (► Anhang C.11): Er spielt hier neben dem Experiment eine untergeordnete Rolle und soll hauptsächlich erhobene Messwerte und die Stichprobe prüfen und SV aufdecken. Es werden subjektive Empfindungen, d.h. „innerer Realität“ der VP erhoben, mittels offener und geschlossener Fragen (feste Antwortmöglichkeiten auf 5-stufiger verbal verankerter Ratingskala). Allgemein werden zur Auswertung die Antworten in Zahlen übersetzt, um verrechenbare Datensätze zu erhalten (z.B. Häufigkeitsanalyse). Variablen (Merkmale) werden in Kategorien eingeteilt (Abstufungen). „Die Häufigkeit in diesen Kategorien sind dann die Grundlage einer tabellarischen oder graphischen Darstellung des Datenmaterials.“ [BN03, S.142] *Offene Fragen* können nach ähnlichen inhaltlichen

Fragebogenauswertung allgemein:

1. Fragen und Antworten sind verschiedene Zeichen / Zahlen / Abkürzungen zuzuordnen, z.B. in unausgefülltem Fragebogen der Frage eine „Variable“ (z.B. A, B, ...) und der Antwort eine Nummer (z.B. A1, A2, ...) zuweisen. Bei offenen Fragen sind Tendenzen zu sammeln und kodieren (für unbeantwortete Fragen kein / extra Symbol).

2. Danach erfolgt Dateneingabe z.B. mit Excel und Fehlerbereinigung (Plausibilitätsprüfung u.a. Suche nach ausschließenden Antworten, Qualität).

Aussagen in Kategorien (Motivtypen) zusammengefasst/operationalisiert werden (Antwort-Auflistung zu jeder Frage). *Geschlossene Fragen* besitzen meist Rating-Skalen (Zuordnung von Zahlen zu Skalenstufen), die z.B. *Mittelwertberechnung* ermöglichen. Zur *grafischen Ergebnisdarstellung* wird oft die deskriptive Darstellung der Antworthäufigkeiten verwendet, z.B. Kreis-, Balken-, Säulen-, Linien- (Trends d.h. zeitliche Entwicklung) oder Streudiagramme (veranschaulichen Zusammenhänge), Histogramm (Klassifikation und

Darstellung stetiger metrischer Merkmale z.B. Alter) oder „Piktogramme“ (Darstellung mittels Symbolen). Es lassen sich Zusammenhänge und Unterschiede aufzeigen (*schlussfolgernde Statistik*) oder Tabellen mit Prozentangaben nutzen. *Fehlerquellen* liegen in der Wahl eines unzulässigen Dia-

⁷⁵ Physiologische Messungen und Auswertung für TH1 erfolgen außerhalb dieser Arbeit (s. [Me13]).

grammtyps oder ungeschickter Darstellung, die nicht das Eigentliche abbildet. Auswertung kann sowohl durch Erfassungsbüros (*Nachteil*: höhere Kosten, ungenügender Einblick in Vorgang d.h. „Qualität“) oder selbstständig durchgeführt werden (*Schrittfolge*: 1.Fragebogen verkodieren 2.Datenmaske erstellen 3.Dateneingabe, s.[As01]). Fragebögen hier werden selbstständig ausgewertet.

Stimmungsfragebogen: Dieser wurde samt Auswertung aus [Kl83, S.18] übernommen. Das Ergebnis bringt die momentane Stimmung zum Ausdruck und sollte mind. 4,0 erreichen (► Anhang C.10).

Experiment: Im Versuch werden Daten mittels Techniken des UL und BS (Biosignale) und anhand von VP-Aufgaben (Qualität, Arbeitsmenge, IQ) erhoben. Techniken und Planung sind in ►3.5.3, 3.7, 3.11 nachzuschlagen. Es wird via Messwiederholung dieselbe VP in EG und KG untersucht, um beide Durchgänge zu vergleichen. Die Relevanz der Ergebnisse für die Hypothese wird durch VE erprobt.

VP-Aufgaben: Zur Auswertung der TH2 werden Leistungstests eingesetzt, um Leistungs- und IQ-Werte von KG und EG zu vergleichen (Gemeinsamkeit, Veränderung). In der KG wird der „Normalwert“ (Richtwert) erhoben, in der EG der Wert unter Einfluss der uV. Für den MWT liegt eine genormte Auswertungsanweisung mit *Normskalen für Intelligenzmessung* vor. Die weiteren hier verwendeten Leistungstests können aufgrund ihres Umfangs nur fragmentarisch eingesetzt werden (einzelne Aufgabenkomplexe), sodass keine Intelligenzmessung möglich ist. Alle VP-Aufgaben eignen sich aber gut für die Erhebungen der primären aV „Qualität“ und „Arbeitsmenge“.

Auswertung: Es gilt zu prüfen, inwieweit Gefordertes gemessen wurde, wie Daten aufzubereiten und ob sie verwertbar sind und der Versuch der Hypothese angemessen ist, u.a. Aufbau, VP-Aufgaben, Methoden. Mess-, Auswertungsmethoden und der Einfluss des Versuchaufbaus sind zu hinterfragen, u.a. Künstlichkeit, Experimentsituation, SV. Zu prüfen gilt, ob das Versuchsziel vorab erkannt wurde (*hier*: Debriefing) und inwieweit störende Einflüsse wie Fremdartigkeit der Kommunikationsmittel oder menschliche/technische Einwirkung auftraten (*hier*: mittels Fragebogen und Nachgespräch). Es sollten nur notwendige Daten erhoben werden, um „Informationsflut“ vorzubeugen (*hier*: VE vorab).

3.14 Gütekriterien

Um in der wissenschaftlichen Forschung gültige, aussagefähige Ergebnisse zur Hypothesenprüfung und Verallgemeinerung zu erheben, sind Gütekriterien einzuhalten: u.a. Objektivität, Reliabilität und Validität. **Objektiv gültige Messungen** werden in den Sozialwissenschaften erschwert, da Beobachter und Beobachteter ggf. in sozialer Interaktion aufeinander einwirken (vgl. Experiment). Hauptproblem für die **Repräsentativität** des geplanten Experiments ist die VP-Auswahl (►3.12), sie sollte optimal den verkleinerten Ausschnitt der jeweiligen Grundgesamtheit darstellen. In dieser Arbeit wird eine Submenge der Grundgesamtheit genutzt (geringe VP-Anzahl, Freiwillige), um Durchführbarkeit zu gewährleisten. Entsprechend besitzen Ergebnisse maximal für diese Stichprobe Bedeutung (keine „perfekte Auswahl“, Abstriche, niedrige externe Validität). „Grundsätzlich sind ohne weitere Untersuchungen die Ergebnisse eines Experiments [...] meistens nicht repräsentativ für die entsprechende Grundgesamtheit, dürfen also nicht verallgemeinert werden...“[Ge03, S.8] Zudem findet die Untersuchung im Labor statt (inhärente Künstlichkeit), unter technisch begrenzter Versuchszeit (ca. 11 Minuten), sodass das Treatment ggf. unnatürlich wirkt und Ergebnisübertragung auf die Realität behin-

dert. Unter diesem Gesichtspunkt wären Feldstudien ggf. vorteilhafter oder könnten später zusätzlich durchgeführt werden. (s. [Ge03][Bor06])

Gültigkeit (Validität): Es gilt zu kontrollieren, ob das Gewünschte gemessen wird (Eignung der Ergebnisse zur Hypothesenprüfung) und soll hier besonders in VE erfolgen. Hypothesenüberprüfende Versuche fordern eindeutig interpretierbare (interne valide) und über die Bedingungen der Untersuchungssituation und VP hinausgehende, generalisierbare Ergebnisse (externe valide). Gefährdet wird Gültigkeit durch zu strenge oder ungenaue Definitions-, Operationalisierungs- oder Messvorschriften, sodass der Realitätsausschnitt verkürzt, verzerrt, unvollständig oder mehrdeutig erfasst wird.

Variablenvalidität: Dies bezieht sich auf die Übersetzung theoretischer Begriffe in empirisch beobachtbare Variablen (Realisierung) bzw. mögliche Operationalisierungsfehler (Güte).

Statistische Validität (Überprüfung statistischer Auswertungsmethoden): Sie gilt als Ergänzung der internen Validität und verweist auf u.a. typische Fehler und Anwendungsfehler bei statistischen Verfahren, ungenaue Messinstrumente, Effektgrößen und (zu kleine) Stichprobenumfänge.

Interne Validität: Verlangt wird SV-Kontrolle/Ausschaltung (Elimination, Randomisierung, Konstanthaltung, Parallelisierung), sodass ΔV -Änderungen eindeutig auf uV -Manipulation rückführbar sind (keine Alternativerklärung). Um interne Validität in Experimenten (mit Kausalschlüssen) zu sichern, sollte das Max-Kon-Min-Prinzip⁷⁶ beachtet werden (SV-Kontrolle im eigenen Vorhaben ►3.5.4). Zudem erhöht die Künstlichkeit des Versuchsaufbaus im UL (Laborsituation) die interne Validität. Zu beachten ist „Variation personaler und situationaler Merkmale, Störfaktoren bei Meßwiederholung sowie bei der interindividuellen und bei der intraindividuellen Bedingungsvariation“ [Fa00, S.55]. Zur strengen Hypothesenprüfung in Replikationsstudien muss Randomisierung von VP und Treatment vorliegen (hohe interne Validität), Bedingungen und Zuordnungen werden im eigenen Versuch variiert (►3.11.3) (s.BN03,S.57]).

Externe Validität (ökologische Validität): Strenge Hypothesenprüfung setzt voraus, dass keine Beschränkung besteht auf Untermengen der Population oder zeitlich-räumlich-situationale Umstände d.h. Stichprobenergebnisse sich (auf andere Personen, Situationen, Zeitpunkte, ggf. andere Konstrukt-Operationalisierung, andere VL) verallgemeinern lassen. Im Mittelpunkt steht die Frage, inwieweit Ergebnisse durch den Versuchsaufbau beeinflusst sind (Behandlung unter ►4.4, 4.5, 5.2). Bei fälschlicher Generalisierung (oft in explorativen Studien mit kleiner leicht zu erreichender Gesamtheit) wird eine illusorische Allgemeingültigkeit beansprucht. Für den eigenen (eher explorativen) Versuch sind daher spätere repräsentative Studien erforderlich. Es böte sich an, den Versuch hier mit möglichst hoher *interner* Validität (im Labor) zu planen: „externe Validität sinkt mit wachsender Unnatürlichkeit der Untersuchungsbedingungen bzw. mit abnehmender Repräsentativität der unterschiedlichen

⁷⁶ "Maximiere die Primärvarianz, kontrolliere die Sekundärvarianz und minimiere die [unsystematische] Fehlervarianz!" [Sa92, S.213] Primärvarianz entsteht durch uV -Wirkung (Maximierung u.a. mit Extremgruppen als Stichprobe, uV -Extremwerte), Sekundärvarianz durch kontrollierte Drittvariablen (SV-Kontrolle und/oder Kontrollvariablen erforderlich) und unsystematische Fehlervarianz ist kontrollierbar d.h. eine unsystematische ΔV -Änderung, die nicht auf SV oder uV -Manipulation zurückführbar ist wie Mess- oder Auswertungsfehler (Minimierung u.a. durch Konstanthalten, SV-Umwandlung in uV oder Kovariable, Wiederholungsmessung, Standardisierung).

Stichproben.“[BN03, S.57] und danach Feldversuche (hohe externe Validität⁷⁷) durchzuführen (VP in natürlicher Umgebung mit natürlichen Einflüssen, Bearbeitung alltäglicher Aufgaben, ggf. andere Operationalisierung). Der Versuch sollte ferner Bedingungen für Experimente einhalten (beschreibbare Versuchsbedingung, Willkür, SV-Kontrolle, wiederholbar).

Fragebogen: Er erfasst „selbstbezogene“ VP-Auskünfte, abhängig von Erinnerungsvermögen, Aufmerksamkeit, Selbsterkenntnis etc. und ist für unwillkürliche Fehler, Verzerrungen und absichtliche Verfälschungen anfälliger (s. [BN03, S.190]) u.a. mangelnde Konzentration, Übermüdung, ungeeignete Items, Motivation, „Tagesform“, Missverständnisse, situative Störungen, schlechte Versuchsbedingungen (vgl. auch **VP-Aufgaben**). Dies gilt es während der Versuche und späteren Analyse des Videomaterials zu beachten (Validitätseinbuße durch Antwortverzerrung). Die Rücklaufquote wird hoch eingeschätzt (wenige Verweigerer), da Fragebögen direkt vor Ort im Versuch auszufüllen sind.

Konstruktvalidität (Bestandteil externer Validität): Sie befasst sich mit ungenauer Operationalisierung abgeleiteter Variablen und unzureichender Konstrukt-Explikation (Güte). Im VE ist zu prüfen, inwiefern Messinstrumente das theoretische Konstrukt erfassen (►2.4, 3.1). (s. [BN06])

Stichprobenbias: Die Stichprobe weicht vom Ideal der streng zufälligen Auswahl ab und verzerrt ggf. Ergebnisse (sinkende externe Validität). Probleme externer Validität werden hier eingehend in ►3.11.3, 3.12 behandelt. Effekte wie VP-Effekte und fehlerhafte Stichprobe, in denen VP sich systematisch vom Populationsdurchschnitt unterscheiden (keine randomisierte Stichprobe), gefährden zusätzlich externe Validität. Allgemein ist es fast unmöglich, SV gänzlich zu verhindern, sie sollten daher zumindest konstant gehalten und ins Testsetting (passiv/aktiv) integriert werden (Integration mindert Eindeutigkeit der Kausalschlüsse). „Durch standardisierte Testabläufe und kleinere Testgruppen können Störvariablen beziehungsweise deren Einfluss weiter reduziert werden.“[Ge03, S.9] Dies wird hier durch Entwicklung eines festgelegten Versuchsplans, Materialien und Texten angestrebt. Um wissenschaftlichen Ansprüchen zu genügen, sind noch weitere Bedingungen zu erfüllen [Wi12c]:
 1) *Planmäßigkeit:* experimentelle Versuchsbedingungen werden ausführlich vorausgeplant und beschrieben (►4) (nachvollziehbar, wiederholbar) 2) *Replikation:* Experiment ist bei Einhaltung der Bedingungen durch Andere wiederholbar und prüfbar 3) *Kontrollverlauf:* Es gilt zu prüfen, ob sich im Versuch psychische Veränderungen bei VP ereignen. VP dürfen den Versuch jederzeit abbrechen. Via Biosignale wird erfasst, inwieweit sie sich in einem Stresszustand befinden (Atmung, Puls etc.) und mit Fragebögen ihr subjektives Empfinden erfragt. Es gilt auch zu prüfen, dass D1 keine Auswirkungen auf D2 hat (Testeffekte), speziell in VE sollen hier EG-KG-Reihenfolge und VP-Aufgaben variiert und evaluiert werden, ob Lern-, Übungs- oder Sensibilisierungseffekte auftreten (**Testing**). Zwischen D1 und D2 dürfen keine prägnanten Ereignisse auftreten (**History**), die D2 beeinflussen (hier: beide Messungen direkt nacheinander) und keine Beschaffenheitsveränderung der Instrumente oder

⁷⁷ „Liegen zu einem weit fortgeschrittenen Forschungsgebiet vorwiegend Laboruntersuchungen vor, so daß an der internen Validität der Erkenntnisse kaum noch Zweifel bestehen, sollten die Resultate vordringlich mit Felduntersuchungen auf ihre externe Validität hin überprüft werden. Dominieren in einem gut elaborierten Forschungsgebiet hingegen lebensnahe Feldstudien, deren interne Validität nicht genügend dokumentiert erscheint, sollten vorrangig Überlegungen zur Umsetzung der Fragestellung in Laboruntersuchungen angestellt werden.“[BN03, S.61]

Kriterienänderung bei Beobachtung und Bewertung auftreten (**Instrumentation**, identische Bedingungen in allen HE). Durch kurze Versuchszeiten entsteht hier wenig „Reifung“ (**Maturation**: psychologische / physiologische Veränderung) wie u.a. Müdigkeit oder Hunger (s. [Fa00]).

3.15 Ethische Bedenken

Ethische Bedenken sind bei Planung und Durchführung (s. [Ge03] [Zw98]) von Experimenten einzu- beziehen (*besonders bei getarntem Versuchsziel*⁷⁸). Es gilt immer zwischen wissenschaftlichem Fortschritt und Menschenwürde abzuwiegen (Güterabwägung). VP sollen im geplanten Versuch zwar „normalem“ Leistungsdruck ausgesetzt (keine Prüfungssituation) aber VP psychisch und physisch nicht zu stark belastet werden mit ggf. sogar Aversionen gegenüber zukünftiger Experimentteilnahme. Die VP wird zuvor vollständig über alle Messungen, Geräteanbringung und alle möglichen negativen Aspekte informiert und kann danach selbst über die Teilnahme entscheiden (informierte Zustimmung, Einverständniserklärung) und hat zu jedem Zeitpunkt das Recht, den Versuch zu verweigern/abzubrechen (klar kommunizieren). Negative Aspekte des Versuchs (dauert ca.1,5h) sind möglichst durch positive aufzuheben (anschließendes Buffet). Zudem ist die VP über Video- und Audioaufzeichnung zu informieren und Datenschutz sicherzustellen. Es gilt zu prüfen, ob das Experiment (psychische) Änderungen bei der VP hervorgerufen hat⁷⁹ (Erfassung Biosignale, subjektive Einschätzung durch Fragebogen, Nachgespräch, neutrale freundliche Versuchssituation, sorgfältige schonende Durchführung). Die VP-Aufgaben dürfen nicht das Selbstwertgefühl der VP verletzen⁸⁰ (keine Bloßstellung, „Bestrafung“, starke Überforderung). VP-Eigenschaften dürfen und sollen im Versuch nicht manipuliert werden, Verletzungen der Vertraulichkeit oder des Datenschutzes sind streng untersagt. Es werden nur *Freiwillige* als VP akquiriert. Um Wiederholbarkeit des Versuchs zu gewährleisten und ggf. Störeinflüsse aufzudecken/zu vermeiden (u.a. VL-Effekt), sollten Fragebögen und Nachgespräch standardisiert vorliegen (schriftlich, vorgegebene Fragen).

3.16 Empirische Vorhersage und statistische Hypothesen

Die empirischen Prognosen für die beiden zu prüfenden Teilhypothesen des Versuchs lauten:

TH1: Kommunikation und gängige Kommunikationsmittel erzeugen in einer Arbeitsituation Ablenkung.

(„**Wenn** Kommunikation/-smittel angewendet werden, **dann** erzeugt dies Ablenkung.“)

→ Empirische Prognose: Ablenkung in EG > KG

TH2: Kommunikation und gängige Kommunikationsmittel verändern die Arbeitsleistung.

(„**Wenn** Kommunikation/-smittel angewendet werden, **dann** verändert dies die Arbeitsleistung.“)

→ Empirische Prognose: Arbeitsleistung in EG ≠ KG

Aus der allgemeinen Forschungshypothese wird die Vorhersage für ein konkretes Ergebnis (operationale Hypothese) abgeleitet, passende statistische Alternativhypothesen (H_1) formuliert und durch

⁷⁸ Es soll ein abschließendes „Debriefing“ mit Enthüllung des Ziels stattfinden und VP über ihr Recht aufgeklärt werden, die Auswertung ihrer Daten zu verweigern oder diese (schriftlich, mündlich oder als Präsentation) einzusehen (s. *Messprotokoll*).

⁷⁹ Physische und psychische Risiken vorher bedenken und kommunizieren, bei unbeabsichtigten Beeinträchtigungen hat der VL unverzüglich einzugreifen, persönliche Freiheit und Selbstbestimmung sind nie einzuschränken.

⁸⁰ Der MWT gibt in der Instruktion an, es existiere *maximal* eine richtige Antwort, um VP nicht zu verunsichern bzw. zu demoralisieren, wenn sie keines der Worte kennen. Tatsächlich existiert immer *exakt* eine richtige Lösung.

komplementäre statistische Nullhypothese (H_0) zum Hypothesenpaar ergänzt. Statistische Hypothesen sind gerichtet oder ungerichtet, spezifisch oder unspezifisch und decken alle möglichen Ausgänge ab (s.[BN06,S.494]). „Die Alternativhypothese postuliert dabei einen bestimmten Effekt, den die Nullhypothese negiert. Das komplementäre Verhältnis von H_0 und H_1 stellt sicher, daß bei einer Zurückweisung der H_0 ‚automatisch‘ auf die Gültigkeit der H_1 geschlossen werden kann“ [BN03, S.28-29]

Operationale Hypothese: Bei den ausgewählten Büroangestellten (verschiedene Branchen und Firmen, Geschlecht möglichst ausgeglichen) bestehen zwischen Durchführung ohne Einsatz der uV (KG) und Durchführung mit Einsatz der uV (EG) Unterschiede bzgl. der Messung (physiologischer) Ablenkung ($EG > KG$) und Arbeitsleistung (gemessen an VP-Aufgaben) ($EG \neq KG$).

Statistische Hypothese: Es wird jeweils im Sinne der Alternativhypothese getestet (►Tab.12).

Tab. 12 unspezifische statistische Hypothesenpaare für TH1 und TH2

TH1	<p>Nullhypothese: Es gibt entweder keine Unterschiede zwischen der (physiologischen) Ablenkung der VP in der KG („Normalwert“) und EG oder die (physiologische) Ablenkung in der KG ist höher als in der EG.</p> <p>(Gerichtete) Alternativhypothese: In der EG ist die (physiologische) Ablenkung der VP höher als in der KG.</p>
TH2	<p>Nullhypothese: Es gibt keine Unterschiede zwischen der Arbeitsleistung der VP in der KG („Normalwert“) und EG.</p> <p>(Ungerichtete) Alternativhypothese: In der EG ist die Arbeitsleistung der VP ungleich der Arbeitsleistung der KG.</p>

In späteren repräsentativen Versuchen ist ein kritischer Bereich festzulegen, für welche Maßzahlen die Nullhypothese verworfen wird. Fällt das Ergebnis jedes Mal uneindeutig aus, ist zu vermutet, dass SV das Experiment zu sehr beeinflussen oder eine uV-Manipulation nicht möglich ist.

Anmerkungen: Einschränkungen bei Operationalisierung und Ort („unnatürliches“ Labor), zeitlich und kapazitär, erfordern eine Eingrenzung der Ergebnisse auf vermutlich keine/geringe aktive, VP-initiierte Kommunikation. Die Arbeitsleistung der EG wird hier wohl primär durch fremd initiierte, „erzwungene“, vermehrt einseitige Kommunikation beeinflusst und entsprechend sinkende Leistung prognostiziert. Um Auswirkungen durch speziell VP-initiierte, beidseitige Kommunikation (Dialog) zu prüfen, könnte aktive Kommunikation in nachfolgenden (Replikations-)Studien durch längere Versuchszeiten mit Eingewöhnungsphase und Möglichkeit für VP, ihren „Arbeitsplatz“ selbst einzurichten, geplant werden (natürlicher, eigene Kommunikationsmedien/Accounts etc.). Zudem bietet sich an, VP selbstständig arbeiten zu lassen, ohne Anwesenheit von VL (SV vermeiden wie u.a. direkte Beobachtung und dadurch bedingt Hemmungen, VL-Effekte), in Gruppen und/oder im Feld zu untersuchen. Durch technische Bedingungen (kurze Versuchszeit, zwingende Anwesenheit mind. eines VL zur Betätigung der Messgeräte, Einzelversuche durch Einzelarbeitsplatz im UL) sind diese Überlegungen im Rahmen dieser Arbeit aber nicht realisierbar.

TEIL IV

VERSUCHSDURCHFÜHRUNG UND -AUSWERTUNG

4. DURCHFÜHRUNG

4.1 Einleitung

Anhand quantitativer Datenerhebung im UL soll die Fragestellung untersucht werden, ob Kommunikation und gängige Kommunikationsmittel in einer Arbeitssituation zwangsläufig Ablenkung erzeugen und inwieweit Änderungen der Arbeitsleistung feststellbar sind. Letztere Annahme bildet den Fokus dieser Arbeit, Messung und Auswertung der Ablenkung erfolgen extern. Um Funktionsfähigkeit der Geräte, Eignung von Material, Aufgaben und den reibungslosen Versuchsablauf zu prüfen sowie mögliche Fehler und SV aufzudecken, werden zuvor VE mit leicht zu erreichenden VP absolviert.

4.2 (Sach-)Hypothesen und Operationalisierung

Die universelle ungerichtete Wirkungshypothese (Kausalhypothese) des Versuchs lautet:

Forschungshypothese HH: Durch Kommunikation und gängige Kommunikationsmittel wird in einer Arbeitssituation Ablenkung erzeugt und eine Veränderung der Arbeitsleistung bewirkt.

Aus ihr leiten sich TH1 und TH2 ab, die anschließend wie folgt operationalisiert werden (►Tab.13):

Teilhypothese 1 (TH1): Kommunikation und gängige Kommunikationsmittel erzeugen in einer Arbeitssituation Ablenkung. („Wenn Kommunikation/-smittel angewendet werden, dann erzeugt dies Ablenkung.“)

Teilhypothese 2 (TH2): Kommunikation und gängige Kommunikationsmittel verändern die Arbeitsleistung. („Wenn Kommunikation/-smittel angewendet werden, dann verändert dies die Arbeitsleistung.“)

Ziele der Versuche hier sind Bewertung und Verbesserung des Forschungsdesigns und maximal erste vorläufige Tendenzen bzgl. der Hypothese (keine allgemeingültigen, repräsentativen Aussagen).

Tab. 13: Übersicht über die Operationalisierung (Variablenzuordnung zu den Teilhypothesen)

TH1: aV ist Z = Konstrukt „Ablenkung“	TH2: aV ist Y = Konstrukt „Arbeitsleistung“
Z1 = Eyetracking (Blickpunkte, Pupillenradius, Lidschläge)	Y1 = Qualität (Fehleranzahl bei bearbeiteten Aufgaben)
Z2 = EOG-System (Sakkaden, Lidschläge)	Y2 = Arbeitsmenge (Anzahl gelöster Aufgaben in vorgegebener Zeit)
Z3 = Thermographie (Hauttemperatur)	Y3 = Fragebogen (subjektive Selbsteinschätzung)
Z4 = ProComp (EKG, EMG, BVP, Atmung, Hautleitfähigkeit, Hauttemperatur)	Y4 = Messung IQ-Wert (gestiegen, gleich, gesunken)
Z5 = Fragebogen (subjektive Selbsteinschätzung)	
uV ist X1 = Kommunikation, gängige Kommunikationsmittel (ohne / mit)	

4.3 Methode und Versuchsplan

4.3.1 Methoden, Versuchsausrüstung und -aufbau

Experiment: Die quantitative Datenerhebung erfolgt primär via „echtem“ klassischen (Prüf-)Experiment, klassifizierbar über 1) *Versuchsziel:* Hypothesenüberprüfung (*aktive uV-Manipulation* und Erfassung der aV-Effekte für eindeutige Kausalschlüsse) 2) *Zahl der uV:* einfaktorielles Experiment (*eine uV*) 3) *Zahl der aV:* multivariates Experiment (*mehrere aV*) und 4) *Versuchsort:* Labor (UL). Letzteres zeichnet sich aus durch optimale SV-Kontrolle, höhere interne Validität, direkte uV-Änderung, willkürliche freie VP-Zuordnung zu Bedingungen, umfangreicheres Messen (im Feld teils so nicht möglich), geringerem Erfolgswang/Kosten und leichterem Wiederholbarkeit (definierte Bedingungen). Abträglich wirken die inhärente Künstlichkeit der Versuchsanordnung, Selektionseffekte bei VP, u. U. eingeschränkte externe Validität (Realitätsmangel, geringere Generalisierbarkeit) und Verhaltensänderung bei VP durch Wissen um beobachtete Laborsituation (►Anhang B.2).

Fragebogen: Zusätzlich erfolgt quantitative, standardisierte Befragung via *schriftlichen*, anonymisierten Fragebögen, für leichtere objektive Auswertung (Standardisierung), Wiederholbarkeit, Fehlervermeidung (weniger SV wie VL-Effekte) und Generalisierung der Datenerhebung und -auswertung. Die VP-Anzahl ist analog zum Experiment (Einzelbefragung). Der Fragebogen erhebt subjektive VP-Eindrücke mit der primären Aufgabe, den Versuch zu bewerten und zu optimieren, SV aufzudecken, Angaben zur VP zu erheben (Stichprobenprüfung bzgl. soziodemographischer Daten, Fragen zu Verhalten und Gewohnheit) und im Experiment erhobene Daten zu bestätigen bzw. zu hinterfragen. Insgesamt werden folgende Methoden und Geräte zur Datenerhebung verwendet (► Tab.14):

Tab. 14 Datenerhebung: Apparate, Mittel, intendierte Messung (Details ► Anhang D.2, B.2)

Geräte/Mittel	Messung
<i>Eyetracking (Remote)</i>	Pupillengröße (-radius), Blickpunkte bzw. Fixationen, Lidschläge
<i>EOG-System</i>	Sakkaden, Lidschläge
<i>Thermographiekamera</i>	Hauttemperatur
<i>ProComp Infiniti</i>	EKG, EMG, BVP, Atmung, Hauttemperatur, GSR (Hautleitfähigkeit, elektrodermale Aktivität)
<i>2 Fragebögen</i>	1) Stimmungsfragebogen (& physiologische Signale ⁸¹) 2) Versuchsfragebogen: soziodemographische Daten (Ein-/Ausschlusskriterien), Verhalten/Gewohnheit, Aufbau, Durchführung, Aufgaben
<i>VP-Aufgaben</i>	Arbeitsleistung (Qualität, Arbeitsmenge, <i>IQ-Werte</i>)
<i>Debriefing</i>	Vorlage für standardisiertes Debriefing-Gespräch mit VP nach Versuchsbeendigung
<i>Kamera(s) (Video- und Audioaufzeichnung)</i> <i>Apparative Beobachtung</i>	Aufzeichnung des Versuchs (Nachkontrolle) und Live-Überwachung (Echtzeit-Übertragung in anderen Raum, Protokollierung ohne direkte „störende“ VP-Beobachtung), <i>Kameras ändern ggf. VP-Verhalten</i> 1) TeamViewer (Live-Überwachung des VP-Monitors, Übertragung in den Beobachtungsraum) 2) Videokamera (<i>nur in den VE</i>) 3) Morae (Live-Überwachung und Aufzeichnung des VP-Monitors und VP von vorne)
<i>uV: Kommunikation und Kommunikationsmittel</i>	- Skype (Telefon, Chat) und Person (Erzeugung von Unterbrechung/Kommunikation, <i>hier: via VL</i>) - VP Email: Webanwendung web.de (<i>VE 1 und 2, alle HE</i>), Programm thunderbird (<i>VE 3 und 4</i>) - VL Email (Beobachtungsraum): Webanwendung gmx.de für Kontakt mit VP und uV-Integration

4.3.2 Aufgaben

Tab. 15: ausgewählte VP-Aufgaben für VE und HE

Ansatz	Aufgaben in der KG und EG (Parallelförmigen) ⁸²
genormte Leistungstests (Intelligenztests)	1) MWT (IQ-Messung) (<i>HE, VE</i>)
für KG / EG (mit Probeaufgabe für MWT- und IQ-Test, inkl. Lösungsweg, zur Eingewöhnung)	2) Logisches Denken + Rechnerische Intelligenz (<i>HE, VE</i>) 3) Zusatz: Sprachliche Intelligenz, Technik & räumliche Vorstellung (<i>HE</i>)

4.3.3 Versuchspersonen

Vorgehen: (Experimentelle) Studie an einer kleinen, leicht zu erreichenden Gesamtheit (Stichprobenuntersuchung)

Grundgesamtheit/Population (Submenge): N = Büromitarbeiter (primär Computerarbeitsplatz)

Stichprobe (HE): n = Medien/IT-Firma und Amt, *nicht-repräsentativ* **Stichprobengröße:** (Angestrebt) N = 12

Stichprobe (VE): Aus ökonomischen Gründen nur Freiwillige aus FHB-Umfeld (für Testzwecke) **Stichprobengröße:** N = 4

Auswahl: Nicht-probabilistische „anfallende“ *Gelegenheitsstichprobe*, Freiwillige, ökonomisch, nicht generalisierbar, praktikabel, Replikationsstudien mit repräsentativer größerer Zufallsstichprobe erforderlich für zulässige Hypothesenprüfung

Akquirierung: Direktkontakt (Ansprechpartner oder Firmen Email) **Quelle:** Städte Brandenburg und Potsdam

⁸¹ Der technische Teil des Debriefing („Physiologische Signale“) wurde von Dritten eingebracht und extern ausgewertet.

⁸² Der MWT (4-5 Minuten) kann in der Versuchszeit vollständig bearbeitet werden, sodass IQ-Werte ermittelbar sind und er als Erstes eingesetzt wird. Der zweite IQ-Test (ca. 45 Minuten) ist nur unvollständig einsetzbar (daher keine IQ-Werte, *nach* dem MWT einzuordnen).

4.3.4 Versuchsplan und –ablauf

Die Veränderungshypothesen werden mit einem elementaren Versuchsplan (*eine* 2stufige uV) geprüft:

Versuchsplan: Experimentelles Design (gestattet intern valide Untersuchung mit kausaltheorietischer Erklärung)

Untergruppen: Messwiederholungsplan (Within-Design) und multivariater Versuchsplan (mehrere aV) als Mischform

Umsetzung: Unifaktorieller multivariater 2-Stichprobenversuchsplan mit Messwiederholung (vgl. Vorher-Nachher-Messung)

Vorteile: Geringe VP-Anzahl notwendig, geringe Datenstreuung, Vergleiche innerhalb EG-KG d.h. Personenvarianz-Kontrolle und Wegfall der Between-Varianz (personelle SV konstant), Konstrukt facettenreicher erfassbar d.h. Ergebnisse besser verallgemeinerbar (steigende externe Validität)

Nachteile: Größerer Aufwand (speziell physiologische Messung), Übertragungsfehler, Positions-Effekt, Carry-Over-Effekt

Möglichst ausgeglichenes Design hinsichtlich folgender Bedingungen: Geschlecht, möglichst Alter (ab 18 Jahre), Art und Schwierigkeit der Aufgaben in KG und EG (Parallelförmigkeit), Reihenfolge der Kommunikation/-mittel (unvollständiges Ausbalancieren) und EG und KG bei unterschiedlichen VP (nicht exakt gleiche Reihenfolge bei aufeinander folgenden VP)

Die Versuche (je ca. 1,5h) finden im UL der FHB auf mehrere Tage verteilt statt. Im VE werden vorab 4 VP geprüft (3 männlich, 1 weiblich), akquiriert via FHB Email-Verteiler (► Anhang C.1).

Versuchsablauf in HE⁸³ (► Anhang C.6): Zunächst werden VP schriftlich (per Email) über den geplanten Versuch informiert und erhalten nach Teilnahmebestätigung weitere Details per Email (► Anhang C.3). Am Versuchstag wird der jeweilige Versuch vorab vorbereitet, u.a. Einrichtung aller Kommunikationsmittel (alte Emails, Chat- und Anrufprotokoll löschen, Instruktion senden, Probe- und VP-Aufgaben auf VP-Rechner öffnen), Starten aller Programme (u.a. Teamviewer, Morae), Gerätevorbereitung, Bereitstellen der Materialien (u.a. Fragebögen, Debriefing, inkl. VP-Code) und Vorbereitung des Erfrischungsraums. Nach Einfinden der VP im UL (Einzelversuch) wird sie begrüßt und nimmt am VP-Rechner Platz (Stuhl ohne Rollen um für Biosignalmessung Bewegung zu reduzieren). Sitzposition der VP wird so gewählt, dass sie Tastatur und Maus ohne Veränderung der Position bequem erreicht und VP instruiert, sich möglichst wenig zu bewegen. Sie erhält eine mündliche Einführung (teils vorgegebene Texte: Zweck, Ablauf, Geräte/-anbringung, VL-Funktion⁸⁴, Handlungsrahmen der VP, Kommunikationsmittel) und wird in Kenntnis gesetzt über Videoüberwachung und Anonymität der Messungen. Anschließend ist ein Fragebogen auszufüllen (gegenwärtige Stimmung, persönliche Angaben für Biosignalwertmessung), danach erfolgt Geräteanbringung an der VP (inklusive verbaler Erklärung). Als nächstes erhält VP die Instruktions-Email mit Versuchsziel und -vorgehen, beschäftigt sich mit Probeaufgaben (zum Verständnis) und kann Fragen stellen. Nach Kalibrierung des Eyetracking erfolgt 1-minütige Ruheaufnahme der Biosignale (Beruhigung, soll Augen schließen), danach startet der Versuch. Ein Durchlauf dauert 11 Minuten, in dieser Zeit soll VP Leistungstests am Monitor bearbeiten (erst MWT zur IQ-Erhebung danach weitere Aufgaben, Parallelförmigkeit für D1 und D2). Der Ablauf unterteilt sich in eine Durchführung ohne Kommunikation (KG: Normalwertmessung) und eine mit Kommunikation (EG: uV-Auswirkungen) mit selber VP, es werden je physiologische Signale der VP aufgezeichnet („Ablenkung“) und Arbeitsleistung ermittelt (Qualität, Arbeitsmenge, IQ-Messung), um Vergleiche beider Messungen anzustellen (Vorteile: pa-

⁸³ Ablauf und Bedingungen wurden nach den VE für die HE angepasst (Änderungen sind in ► 4.4 aufgeführt). Zeitphasen pro Versuch unterteilen sich in Vorbereitung (½ Stunde), Versuchsdurchführung (½ Stunde) und Nachbereitung (½ Stunde).

⁸⁴ Instruierung, Kommunikation im Versuch ist auf den VL2 zu beschränken (VL1 als „bloßer“ Techniker ausgewiesen).

parallelisieren und Konstanthalten individueller SV der VP, Ökonomie)⁸⁵. Nach D1 erfolgt eine kurze Pause, um Daten abzuspeichern, neue Instruktionen und VP-Aufgaben zu öffnen und ggf. kurz Absprache mit VP zu halten⁸⁶, danach schließt sich D2 unter exakt gleichen experimentellen Bedingung an (ausgenommen der uV⁸⁷). Der technische Experimentator (VL1) befindet sich durchgängig im UL (Gerätebedienung), während VL2 im Beobachtungsraum protokolliert (unaufdringliche Beobachtung, Absprache via Skype mit VL1), als Ansprechpartner für VP fungierte und (in EG) die uV manipuliert. Nach Versuche füllt VP einen Fragebogen zum Versuch aus, wird entkabelt und von VL2 in separaten Erfrischungsraum weitergeleitet (VP erhalten keine Vergütung). Dort findet Debriefing statt, in dem neben Angaben zu physiologischen Messungen und Kommunikation das „wahre“ Versuchsziel aufgedeckt wird (getarnte Versuchssituation), inkl. Smalltalk und Unterschreiben des Messprotokolls (Schweigepflicht-, Datenschutz-, Einverständniserklärung). VP wird verabschiedet und der Versuch nachbereitet (u.a. Datenspeicherung, Beschriftung, Absprache, Geräte zurücksetzen).

4.3.5 Versuchsbedingungen

Tab. 16: Zusammenfassung des Versuchsaufbaus

Zusammenfassender Versuchsaufbau	
<i>Anwesende bei</i>	<i>Im UL:</i> eine VP (Einzerversuche) und ein (technischer) VL (<i>hier:</i> VL1)
<i>Durchführung</i>	<i>Im getrennten Beobachtungsraum:</i> ein VL (Beobachtung, Protokollierung, uV-Manipulation) (<i>hier:</i> VL2)
<i>Durchläufe je Versuch</i>	2 Durchläufe mit <i>gleicher</i> VP (1mal KG, 1mal EG) mit identischer Laborsituation und Ablauf, gleichen Skalen/Instrumenten, ruhigen konstanten Untersuchungsbedingungen, Parallelformen der Aufgaben, uV-Manipulation in EG (Kommunikation/-smitteln) → <i>Vorteil: Ökonomie, Parallelisierung der SV der VP</i>
<i>Versuchsziel</i>	Getarnte Versuchssituation (VP weiß um Experiment aber kennt Versuchsziel nicht)
<i>Schlüsselworte</i>	<i>Nicht zu verwendende Begriffe:</i> Ablenkung, Konzentration, Kommunikation(-smittel), Leistung → damit VP nicht (versuchen), „wahres“ Versuchsziel aufzudecken, Verhalten (un)bewusst zu ändern, um Neutralität zu wahren und VP nicht durch Wortwahl (unbewusst) zu beeinflussen
<i>Laborumgebung</i>	Möglichst „normale“ Arbeitssituation (natürlicheres VP-Verhalten) d.h. keine Ausnahmesituation (z.B. Prüfungssituation, unnatürlicher zusätzlicher Stress/Druck), keine offene Beobachtung der VP
<i>Instruktionen</i>	Schriftlich (per Email): Zweck (Vorwand), Ablauf, Eckdaten (einfache, verständliche Formulierung)
<i>Aufgaben</i>	- Leistungstests (Parallelform für EG und KG d.h. gleiche Art und Schwierigkeit, Lerneffekte beachten) - Bearbeitung durch VP am Monitor, Zeitvorgabe (identisch in D1 & D2), ohne zusätzliches Schreibzeug
<i>Versuchszeit pro Durchlauf</i>	11 Minuten (identisch für D1 und D2 und alle VP in VE und HE), VP ist Versuchszeit bekannt (Zeitdruck, ermutigt zielgerichtete konzentrierte Bearbeitung, ggf. Zeitmanagement und Priorisierung)
<i>Durchführung (Ablauf)</i>	<i>Vorexperimente:</i> Erprobung, Verbesserung und Auswahl der experimentellen Bedingungen (<i>aufgrund der geringen VP-Anzahl im VE besteht die Möglichkeit, dass SV und Fehler trotzdem nicht aufgedeckt werden</i>) <i>Hauptexperiment:</i> Prüfexperiment zur Hypothesenprüfung
<i>VP und Anzahl</i>	- <i>im VE:</i> aus dem Hochschul Umfeld der FHB (ökonomische Gründe), N = 4 - <i>im HE:</i> aus der definierten Grundgesamtheit (Akquirierung aus der Wirtschaft), N = 12 (angestrebt) - VP-„Pflege“: faire Vorab-Information, Begrüßung, angenehme Situation, „wahres“ Ziel aufdecken (Debriefing), Verabschiedung, Anwerben <i>Freiwilliger</i> , unangenehme Aspekte nicht verschweigen, Anonymität und Datenschutz gewährleisten, Zumutbarkeit der Untersuchung (u.a. Dauer, Aufgaben)

⁸⁵ EG & KG werden je VP vertauscht, um SV aufzudecken. Die uV tritt in EG 5mal auf, dazu sind 3 Abläufe definiert (1 Ablauf pro VP).

⁸⁶ Bei der EG in D2 trat in der Pause der VL2 ein, um sich (als Ansprechpartner) vorzustellen und Kommunikationsmittel zu erklären („Ich befinde mich während der Durchführung nicht im UL, um die Biosignale nicht zu verzerren.“).

⁸⁷ uV ist hier Kommunikation per Email, Chat, Telefon, persönliches Eintreten von VL 2 (► Anhang Kommunikationübersicht).

4.3.6 Empirische Vorhersage

Empirische Prognosen und unspezifische statistische Hypothesenpaare für die Teilhypothesen lauten:

TH1: Empirische Prognose: Ablenkung in EG > KG

Vorgehen: Messung der Ablenkung (aV: Z_n) bei allen n VP (Vergleich EG und KG)

Nullhypothese (H_0): Es gibt entweder keine Unterschiede zwischen der (physiologischen) Ablenkung der VP in der KG („Normalwert“) und EG oder die (physiologische) Ablenkung in der KG ist höher als in der EG.

(gerichtete) *Alternativhypothese* (H_1): In der EG ist die (physiologische) Ablenkung der VP höher als in der KG.

TH2: Empirische Prognose: Arbeitsleistung in EG \neq KG

Vorgehen: Messung der Arbeitsleistung (aV: Y_n) bei allen n VP (Vergleich EG und KG)

Nullhypothese (H_0): Es gibt keine Unterschiede zwischen der Arbeitsleistung der VP in der KG („Normalwert“) und EG.

(ungerichtete) *Alternativhypothese* (H_1): In der EG ist die Arbeitsleistung der VP ungleich der Arbeitsleistung der KG.

4.4 Vorexperimente

Es werden Probleme protokolliert (Protokolle ►CD anbei) und die Datenaufbereitung für das HE erprobt (Codierung, Tabellen, Diagramme), um Veränderungen / Zusammenhänge / Unterschiede aufzuzeigen. Vollständige Ergebnisse und Versuchsangaben der VE und HE finden sich in ►Anhang D.

4.4.1 Ergebnisse und Auswertung

Aufgaben⁸⁸ (►Abb.5, 6): Auffällig ist, dass Werte aus D1 oft schlechter ausfielen, unabhängig von EG oder KG. Es werden Lern- und Übungseffekte vermutet, d.h. langsames Lernen der Bearbeitung in D1 und Abrufen/Anwenden in D2. Um dem abzuhelpfen, wurde VP ab VP3 vorab eine Probeaufga-

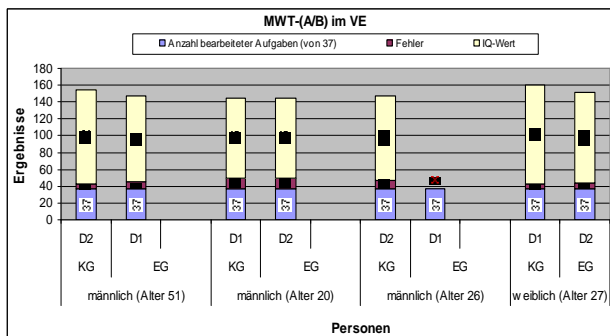


Abb. 6 VE-Ergebnisse des MWT (D1, D2) in KG & EG

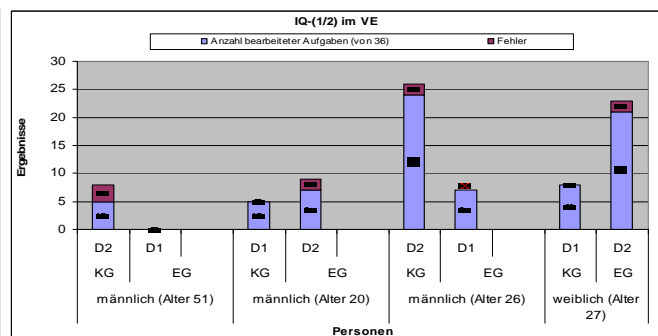


Abb. 5 VE-Ergebnisse der IQ-Tests (D1, D2) in KG & EG

be je Aufgabentyp mit Lösung ausgedruckt vorgelegt. Aufgrund von Datenausfällen (VP3) und wenigen VP ließ sich aber keine definitive Änderung feststellen (SV bestand scheinbar weiter). Dahinter wurde das passive Vorgehen vermutet („Überfliegen“ der Beispiele ohne Interaktion d.h. keine Verinnerlichung) und für HE digitalisierte interaktive Probeaufgaben genutzt (selbstständiges Lösen, Klick-Schreibaktion). VP4 arbeitete so schnell, dass die Zusatzaufgabe (uV) verfrüht zu senden war (verzerrter Ablauf). Daher liegen VP-Aufgaben im HE auf dem VP-Computer vor und müssen nicht erst von VP per Email geladen werden (Computer entlastet, keine verzerrte Arbeitszeit durch Laden).

Debriefing: Auch nach Ausfüllen des Fragebogens am Ende erkannte keine VP das Versuchsziel⁸⁹. Vorwand und Ablauf werden hier unter Vorbehalt als erfolgreich betrachtet. Eine VP empfand das Eintreten des VL2 (uV) als „auffällig“ aber wurde nicht misstrauisch, Versuch und Vorwand wurden

⁸⁸ IQ-Werte sind der MWT-B Normentabelle entnommen (der Testzentrale lag keine MWT-A Normentabelle vor). „IQ“ wird hier unter Vorbehalt zu Vergleichszwecken genutzt und erhebt keinen Anspruch, „echte“ IQ-Werte der VP abzubilden.

⁸⁹ Annahmen der VP rangierten von „Eyetracking“, „Aufnahme von Biosignalen“ bis hin zu „keine Ahnung“ (2 VP).

von einer anderen VP als „schlüssig“ und das Versuchsziel im Nachhinein als „plausibel“ beschrieben. 2 VP merkten an, dass sie mit web.de bzw. Skype nicht vertraut seien, im Versuch stellte sich heraus, dass beide VP das jeweilige Kommunikationsmittel nur unter Anweisung und teils Aufforderung des VL1 verwendeten (VL-Eingriff, verzerrt ggf. Ergebnisse, darf in HE nicht auftreten, d.h. VP sollten Kommunikationsmittel kennen oder sollten vor Durchführung erläutert / vorgeführt werden).

Technik/Material: Für VP3 und 4 fand der Versuch durch den langsamen VP-Computer unter konstant erschwerten Bedingungen statt (langes Laden der Aufgaben-PDFs, verzögertes Scrollen/Schreiben). So lassen sich maximal VP1 und 2 oder VP3 und 4 vergleichen (Bedingungsunterschied). EOG-Gerät (Gerätausfall), VP-Aufzeichnung und Email-Programm (thunderbird) verursachten die größten Probleme: VP1 und 2 wurden per Videokamera aufgezeichnet, Konvertierung und Aufbereitung der Daten hätte aber den zeitlichen Rahmen gesprengt, sodass danach die Aufzeichnung mit Morae getestet wurde. Das Email-Programm *thunderbird* stürzte im Versuch mehrfach ab, es war auf web.de zurückzugreifen. Da web.de normal keine Popups oder Töne für eingehende Nachrichten anbietet, musste VL1 die VP öfter darauf hinweisen, Emails zu prüfen (SV) ► Anhang D.2.

Stimmungsfragebogen (vor Versuch): Werte sollten nicht stark unter 4,0 liegen, da sonst die Gefahr

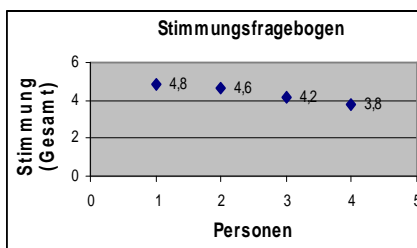


Abb. 7 Stimmung VE

besteht, dass VP zu ängstlich oder verbissen arbeiten. Im VE wurden keine extremen Stimmungen ermittelt (► Abb.7).

Versuchsfragebogen (nach Versuch): Alle VP nutzen Computer gelegentlich bis immer, auch in der Freizeit, die Arbeit damit fällt ihnen leicht. Die Kommunikationsmittel wurden nie bis immer verwendet (ggf. Einfluss auf Verhalten/Ergebnisse). Der Einführungstext der HE wurde daher um explizite Erklärung/Vorführung eingesetzter Medien erweitert. VP-

Aufgaben wurden als eindeutig und verständlich, ihre Schwierigkeit zwischen zu schwer bzw. gar nicht schwer eingeschätzt, einige Aufgaben waren bereits bekannt. Ab VP3 wurden MWT- und IQ-Aufgaben im Fragebogen getrennt erfragt, es wurden jedoch weiterhin ähnliche Einschätzungen vorgenommen. Der Versuchsablauf wurde als ziemlich bis völlig eindeutig befunden, VP fühlten sich wenig bis mittelmäßig aufgeregt, gar nicht bis wenig zeitlich unter Druck gesetzt und in EG wenig abgelenkt (gar nicht/mit leichten Abstufungen als störend/einschränkend empfunden, uV als nicht auffällig erlebt). Auch VL-Anwesenheit setzte nach eigener Aussage keine VP unter Druck, niemand fühlte sich beobachtet oder unwohl. Es ist unsicher, inwieweit die Angaben „wahr“ sind. Unter Vorbehalt wird abgeleitet, dass der Versuch keine Extremsituation darstellt, mit zumutbaren Aufgaben.

4.4.2 Diskussion und Zusammenfassung

Ethische Bedenken: VP-Monitor und VP (vorne) wurden aufgezeichnet und in den Beobachtungsraum übertragen, zur Koordination der VL, Integration der uV (durch VL2), Vermeidung von VL-Effekten (möglichst keine Interaktion zwischen VP und VL1 im UL, Kommunikation ausschließlich mit VL2) und um VP weniger störend zu beobachten (indirekt, Protokoll). Im Einführungstext wurde explizit die Aufklärung über die Videoaufzeichnung eingefügt. Dies beeinträchtigt ggf. „natürliches“ VP-Verhalten (u.a. reaktive Effekte), ethische Gesichtspunkte besitzen jedoch größere Relevanz.

Zeitplanung: Die Fragebogenbeantwortung überstieg nicht die Versuchszeit (Stimmungsfragebogen 2,5-3 Minuten, Versuchsfragebogen 4,5-5 Minuten). Vor- und Nachbereitungsphase des Versuchs betragen je 20-30 Minuten, die Gesamtversuchszeit pro VP zwischen 80-90 Minuten (► Anhang B.3).

EG und KG: Aufgrund weniger VP, unausgeglichener Stichprobe (1 weiblich/3 männlich, 1 Mitarbeiter/3 Studenten), SV (ungenauere Messung) und Datenausfällen sind keine Aussagen darüber möglich, inwieweit die Reihenfolge der KG und EG evtl. Ergebnisse verändert. Daher wird auch in den HE die Reihenfolge weiter ausbalanciert (Ausgleich von SV). Es besteht jedoch die Vermutung, dass die EG in D1 inhaltlich sinnvoller ist, da hier eher Fragen u.a. zu Ablauf und Aufgaben auftreten und die uV entsprechend natürlicher integriert werden kann (Kommunikation zwischen VL2 und VP).

Praktische Überlegung: Instruktion und VP-Aufgaben wurden ab VP3 in *einer* Email zusammengefasst, da VP trotz verbaler Anweisung und der Email-Betreffzeile die VP-Aufgaben öffneten ohne die Instruktionsmail zu lesen. Aufgrund langer Ladezeiten wurden VP-Aufgaben für die HE komprimiert (schlechtere Qualität) und auf dem VP-Computer vorab geöffnet, nur die Zusatzaufgabe wird als uV per Email versendet (keine verzerrte Bearbeitungszeit durch PDF-Probleme, weniger Frust/Stress der nicht explizit von uV ausgelöst wird). Da thunderbird (Vorteil: Popups/akustische Signale bei Nachrichten) wiederholt abstürzte, musste auf web.de (Webanwendung) zurückgegriffen werden. Dies zog leider nach sich, dass VL1 eingreifen und auf Emails hinweisen musste (SV)⁹⁰. Zudem waren nicht alle VP mit web.de vertraut, was Probleme bei der Benutzung (Verzögerungen, Frustration) nach sich zog (für HE wurde ein Add-On gefunden, um neue Nachrichten visuell und akustisch anzukündigen). In den Einführungstext wurde Erklärung und Vorführung der Kommunikationsmittel eingearbeitet⁹¹ und im Zuge der VE Instruktionstexte und Aufgaben-Anweisungen umformuliert und Materialien gemäß VP-Nachfragen und aufgetretener Schwierigkeiten angepasst⁹². Die Protokollvorlage wurde erweitert für Notizen zu den einzelnen Versuchsphasen (D1, D2, Vor-/Nachbereitung).

Versuchsfragebogen: Freitextfelder für offene Antworten wurden aufgrund hoher Antwortverweigerung, Doppelung, zur Vereinfachung und Verkürzung zusammengefasst und „Teil III: Angaben zur Organisation“ erweitert (Trennung von MWT- und IQ-Fragen, um Ergebnisse besser zuzuordnen).

VP: Wegen VP-Ausfällen (VP erschien nicht zu Termin) und geringer Beteiligung wurde nur eine ungleiche Anzahl männlicher (3) und weiblicher (1) VP getestet, zudem mehr Studenten (3) als Mitarbeiter (1) der FHB. VP-Termine mussten mehrfach verschoben werden, aufgrund technischer Ausfälle im UL und Krankheitsgründen und einige VP waren VL bekannt (ggf. anderes VP-Verhalten).

Abläufe: Der Ablauf verlief teils unruhig/hektisch/unsicher, was ggf. VP beeinflusste, zudem wurde von vorgegebenen Texten abgewichen d.h. ungenormter, unstandardisierter Ablauf, ohne identische Bedingungen für alle VP. VE waren nützlich, um das Vorgehen zu üben, Unregelmäßigkeiten zu korrigieren und Sicherheit zu erlangen (später ggf. VL-Schulung). Aus dem Protokoll ergab sich, dass VP fast nur mit VL1 kommunizierten. Um dies zu vermeiden, wurde die Funktion der VL in der ver-

⁹⁰ Ablauf wurde so umgearbeitet, dass per Chat oder Telefon (uV) vorab auf eingehende Emails (uV) hingewiesen wird.

⁹¹ VP1 kannte weder Skype noch web.de und ignorierte daher (nach eigener Aussage) jegliche Kommunikation dieser Art.

⁹² In den Instruktionen wurden VP-Aufgaben stärker hervorgehoben (Druck erhöhen, zielorientierte konzentrierte Arbeit).

balen Einführung deutlicher getrennt mit einem „technischen“ VL1 (zuständig für Geräte, kennt Versuch nicht) und VL2 (zuständig für uV, VP-Aufgaben, Fragebögen, Debriefing, Ansprechpartner für VP), damit VP aus eigenem Antrieb in der EG Kommunikation mit VL2 aufnehmen und keine Kommunikation mit VL1 auftritt. VP lasen Instruktionen und Aufgabentexte oft oberflächlich (speziell MWT) und stellten VL1 Fragen. Es gilt zu prüfen, ob dies durch Testsituation (Zeitdruck), uV (Ablenkung) oder andere Einflüsse (oberflächliche Arbeitsweise, Flüchtigkeitsfehler) bedingt wird.

Technische Schwierigkeiten: Ab VP3 ließ der überlastete VP-Computer nur stark verzögerte Bearbeitung der Aufgaben-PDFs zu und beeinflusste die Bearbeitungszeit negativ (keine einheitliche Auswertung von VP1-2 und VP3-4), sorgte neben der uV für störende Effekte (erhöhter Frust/Stress) und beeinflusste den Ablauf (Programm-Ausfall beim Teamviewer wodurch VL2 „blind und taub“ war, d.h. eingeschränkte/lückenhafte Mitverfolgung und Protokollierung, uV schwer zu synchronisieren).

Kommunikation (uV): 2 VP antworteten auf Chat-Nachrichten und initiierten sogar selbst Kontakt mit VL2 via Skype (Fragen zum Ablauf). Generell erfolgte Kommunikation im Versuch aber eher einseitig und von außen „erzwungen“ (VL2). Die uV wurde nach den VE umgestaltet, sodass VP nicht mehr gezwungen sind, auf diese einzugehen um weiterzuarbeiten (VP-Aufgaben nicht mehr per Email versendet) und VP sich „natürlich“ entscheiden können, wie sie Kommunikation handhaben (wie im Alltag) und zusätzliche (störende) Kommunikation zwischen VL1 und VP reduziert wird.

4.5 Hauptexperimente

4.5.1 Durchführung und Protokollierung

VP erschien im UL, füllte den 1. Fragebogen aus und wurde von VL1 verkabelt (inkl. Erklärung zu Geräten, Messungen, Ablauf). Danach las sie die Instruktionsmail, bearbeitete Probeaufgaben und wurde in Kenntnis gesetzt über Gesichts- und Monitoraufzeichnung („für interne Auswertung“, anonym). VL1 stellte das Eyetracking ein und es folgte eine 1-minütige Ruheaufnahme. Wurden Inhalte der Einführung oder Arbeitsschritte zwischendurch vergessen, ergaben sich Unregelmäßigkeiten im Ablauf oder wurden Besonderheiten bei der VP beobachtet, kommunizierten die VL per Skype-Chat unauffällig miteinander. Beide Durchläufe dauerten je 11 Minuten und starteten mit dem MWT (teils Hinweis durch VL1). Zwischen beiden Durchläufen fand eine kurze Pause statt (3-8 Minuten), hier trat VL2 ein: erfolgte die EG in D2, erklärte und führte VL2 die Kommunikationsmittel vor, erfolgte die KG in D2, kündigte VL2 seine Abwesenheit an („*Ich wollte nur Bescheid geben, dass ich ab jetzt nicht mehr per Skype erreichbar bin*“). Einige VP reagierten in der EG sofort auf die uV und lasen und antworten auf Nachrichten („*Alles gut*.“). Teils wurden Instruktionen aus den uV erst nach vollständiger Abarbeitung einer Aufgaben-PDF befolgt, Chat- und Emailnachrichten erst nach Beendigung des Durchlaufs geöffnet oder der VL1 um Hilfe gebeten („*[Zusatzaufgabe] muss ich aber erst danach machen?*“)⁹³. Andere bemerkten die uV (Chat, Email), ignorierten sie aber oder fragten beim VL1 nach („*Soll ich da jetzt irgendwas machen?*“) oder die uV wurde gar nicht bemerkt. Kommunikation zwischen VL1 und VP konnte auch in den HE nicht gänzlich unterbunden werden. VP stellten

⁹³ Im Text (Chat, Telefon, Email) wurde ab VP3 betont, dass Zusatzaufgabe *nach* anderen Aufgaben zu bearbeiten sei.

Fragen, die z.T. bereits in der verbalen Einführung und schriftlichen Instruktionen erklärt wurden (VP nach MWT: „Dann mach ich jetzt IQ-2 auf, ja?“), 1 VP legte sogar eine Pause zwischen MWT- und IQ-Aufgaben ein, weil sie auf Signal zur Weiterarbeit wartete (verzerrte Ergebnisse). VL1 bemühte sich, Antworten knapp und kurz zu halten, bei Unsicherheiten erfolgte unauffällige Absprache mit VL2 per Skype-Chat. Zudem musste VL1 einige VP wiederholt darauf hinweisen, sich nicht zu bewegen oder gerade hinzusetzen (Störung). Einige VP seufzten während des Versuchs oft oder erschienen frustriert bzw. beendeten Aufgabenteile nicht und wendeten sich anderen Teilen zu (beim MWT-B: „Keine Ahnung, weiß ich nicht, fertig!“). Alle VP waren stark auf die Aufgaben konzentriert und reagierten nur verzögert, wenig oder gar nicht auf uV Chat und Email (dies kann u.a. auch mit der kurzen Versuchszeit in Kombination mit der Aufgaben-Fülle zusammenhängen). Beim Eintreten von VL2 zur Übergabe des Fragebogens (uV), vermieden VP völlig den Blickkontakt und reagierten minimal oder stellten kurz Blickkontakt mit VL2 her und antworten zum Teil auch (VL2 „Ist sonst alles okay?“, VP „Ja, alles okay.“) oder unterbrachen ihre Arbeit komplett, um sich VL2 zuzuwenden. Generell hielten alle VP verbale Antworten auf Unterbrechungen knapp („mmh“). Eine VP fragte zwischen D1 und D2, ob Surfen im Internet gestattet sei (VL: „Ja“)⁹⁴. Aufgrund des überlasteten VP-Computers war dies nur mit Verzögerung möglich, sodass sie nach kurzer Zeit wieder aufgab. Einige VP versuchten, die Zusatzaufgabe sofort zu öffnen und einzeln oder parallel neben dem IQ-Test zu bearbeiten. Da Laden und Speichern der Email-Anhänge aufgrund der Rechnerüberlastung langwierig war, gaben sie das Vorhaben auf („PC ist voll langsam, der unterstützt mein Multitasking gar nicht.“) oder wurden von VL1 hingewiesen, erst IQ-Aufgaben zu bearbeiten. Nach den Versuchen wurden VP entkabelt, füllten den 2.Fragebogen aus, unterschrieben das Messprotokoll und wurden von VL2 in den Erfrischungsraum weitergeleitet, wo ein entspanntes Debriefing stattfand (►Protokolle s. CD).

4.5.2 Ergebnisse und Auswertung

Erhobene Daten sollen wertfrei präsentiert und Ergebnisse neutral beschrieben werden⁹⁵. Deutende Analysen sind aufgrund der Zielsetzung dieser Arbeit, geringer VP-Anzahl, technischer Probleme, nicht repräsentativer Stichprobe u. ä. (►3.15) unzulässig (keine Aussagekraft, maximal Vermutung).

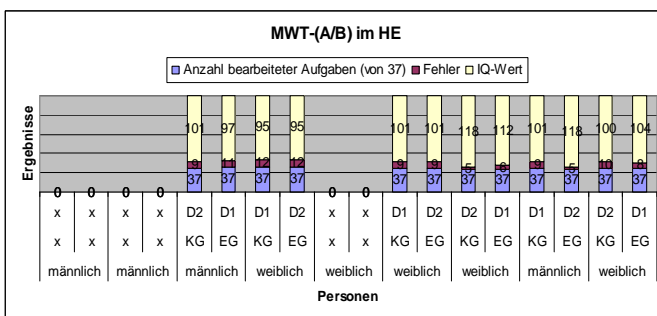


Abb. 8 HE-Ergebnisse des MWT (D1, D2) in KG & EG

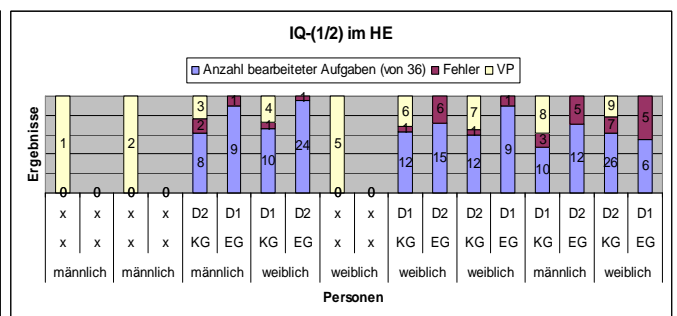


Abb. 9 HE-Ergebnisse der IQ-Test (D1, D2) in KG & EG

MWT-Aufgaben (►Abb.8): Es gilt, dass „bei kleineren Unterschieden der Gesamtpunktzahlen Zufälle eine große Rolle spielen. So kann es sehr gut allein vom Zufall abhängen, ob jemand 34 oder 35

⁹⁴ Dies verzerrt ggf. die Bearbeitungszeit (fand in KG statt), wurde aber im Sinne einer möglichst „natürlichen“ Arbeitssituation gestattet.

⁹⁵ Alle Rohdaten sind anonymisiert im Anhang und der beigefügten CD einsehbar.

oder auch 36 Gesamtpunkte erhält. Hingegen dürften so große Leistungsunterschiede wie fünf und mehr Rohpunkte nicht mehr bloß mit Zufallseinflüssen erklärt werden.“[Le05, S.36] Ergebnisse sind nicht aussagekräftig, Fehleranzahl und IQ liegen in der EG 2x über, 2x unter und 2x gleich auf mit KG-Werten. Aufgrund kleiner VP-Anzahl kann dies viel mit VP selbst zu tun haben (keine Ausbalancierung individueller Merkmale), sowie „Zufall“, Übungs-/ Lerneffekten und weiteren SV.

IQ-Aufgaben (►Abb.9): Ähnliches gilt für Ergebnisse hier. Es lässt sich aber die leichte Tendenz ablesen, dass VP in D2 generell mehr Aufgaben abarbeiteten, was auf Übungs- und Lerneffekt hinweist.

Stimmungsfragebogen (►Abb.10): Werte liegen zwischen 3,5 und 5,7, keine der VP befand sich

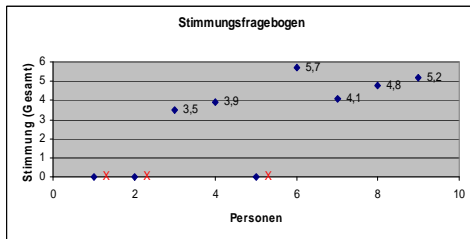


Abb. 10 Stimmung HE

nach eigener Angabe in schlechter Stimmung, die Mehrzahl lag sogar etwas über dem Richtwert 4,0 und deckte sich mit der Beobachtung, dass die VP entspannt und positiv gelaunt waren.

Versuchsfragebogen: Alle VP nutzten den Computer immer oder oft zum Arbeiten, die Nutzung fällt ihnen nach eigenen Angaben leicht und sie verwenden ihn mehrheitlich auch in der

Freizeit. Die VP gaben an, Telefon, Handy oder Skype auch im Alltag immer, oft oder zumindest gelegentlich zu nutzen, sowie SMS zu versenden und auch Emails außerhalb der Arbeit einzusetzen (nur 1 VP gab „selten“ an). Bei sozialen Medien hielt sich die Verwendung die Waage (3VP immer bis oft, 3VP selten bis nie). Keine der VP chattete in ihrer Freizeit gewohnheitsmäßig (maximal gelegentlich, 4VP nie). Alle VP gaben aber an, via o.g. Kommunikationsmittel privat während der Arbeit am Computer gelegentlich bis immer zu kommunizieren. Die MWT- und IQ-Aufgaben wurden überwiegend als verständlich und eindeutig eingeschätzt, der MWT schnitt hier ein wenig besser ab. Die Schwierigkeit wurden zwischen ziemlich bis mittelmäßig bei beiden Aufgabentypen eingeschätzt, der MWT hierbei von 2 VP als wenig schwierig und der IQ von 1 VP als gar nicht schwer. Der MWT wurde meist mittelmäßig und vereinzelt wenig bis gar nicht leicht empfunden, die IQ-Aufgaben dagegen hauptsächlich als wenig leicht und vereinzelt gar nicht oder mittelmäßig leicht. Nur 1VP gab an, den MWT bereits zu kennen, allen anderen war er unbekannt. Die IQ-Aufgaben waren hingegen nur 3 VP gar nicht bekannt, die restlichen 3 verteilten sich auf wenig, mittelmäßig und ziemlich bekannt. Generell empfanden alle VP Anweisungen im Versuch ziemlich bis völlig eindeutig und fühlten sich nur mittelmäßig, wenig bis gar nicht aufgeregt. Nur 1 VP fühlte sich zeitlich unter Druck gesetzt, die anderen gaben hier mittelmäßig, wenig bis gar nicht an. Die Anwesenheit Anderer schien keine VP unter Druck zu setzen und nur 1 VP fühlte sich ziemlich beobachtet, die anderen mittelmäßig, weniger oder gar nicht. Insgesamt fühlte sich keine VP unwohl. In der EG fühlten sich VP gar nicht (4 VP) oder wenig abgelenkt, nur 1 VP gab „ziemlich“ an. In der KG fühlten sich VP ebenso gar nicht (4 VP) oder wenig abgelenkt (1 VP „keine Antwort“). Größtenteils fühlt sich in der EG auch niemand gestört (1 VP „mittelmäßig“) und wenig bis gar nicht gezwungen oder genötigt. In der KG fühlten sich die VP noch minimal weniger gestört und genauso wenig gezwungen oder genötigt wie in der EG. 3VP gaben in der EG und KG an, sich gar nicht anders als bei alltäglicher Arbeit zu verhalten. 1VP gab hier in der EG „mittelmäßig“ und 2 VP „ziemlich“ an, in der KG dagegen 1 VP „wenig“, 1 VP „mit-

telmäßig“ und 1 VP „ziemlich“. In der EG existiert eine leichte Tendenz, dass hier Aufgaben schlechter und Aufgaben in der KG besser zu lösen waren. Aufgrund der geringen VP-Anzahl lassen sich hier aber keine zulässigen Aussagen machen. Insgesamt sind die Angaben eher uneindeutig.

Debriefing: Keine VP erkannte vorab das „wahre“ Versuchsziel, Vermutungen variierten hier zwischen „Usability-Tests (Verhalten am PC)“, „Stressfaktor unter ungewohnter Arbeit (Aufregung)“, „technische Messungen und VP-Aufgaben auswerten sowie Messung des Blick bei MWT“ und „Körperreaktion beim Eintreten des VL2“⁹⁶. 3 VP gaben an, die Kommunikation im Versuch (uV) bewusst ignoriert oder bewusst später gelesen/reagiert zu haben. Begründungen lauteten hier, dass VP nicht wussten, ob sie die Nachricht in der Versuchszeit öffnen sollten oder dachten, die Kommunikation bezöge sich auf die Zusatzaufgabe und sie erst alle weiteren bearbeiten wollten (Priorisierung, Zeitmanagement), der Umgang mit Skype oder die Bildschirmaufteilung (Programme) ungewohnt sei (Probleme bei Benutzung) oder sie zu stark auf die Aufgaben konzentriert waren. Einige VP merkten an, dass sie sonst nicht so unbeweglich und gerade saßen (unnatürliche steife verkrampfte Haltung), versucht hätten, bewusst Signale zu beeinflussen (u.a. ruhiger sein, Blick kontrollieren) oder bewusst ihr Handy ignoriert hätten, obwohl sie in einer „normalen“ Arbeitssituation auf SMS sofort reagiert hätten. Eine VP fügte an, dass sie die „unnatürlich“ empfundene Arbeitssituation aufgrund schwieriger Aufgaben zeitweilig ausblenden konnte. Andere bemerkten, dass sie in D2 motivierter gewesen wären oder sich selbst unter Druck gesetzt hätten.

4.5.3 Diskussion und Zusammenfassung

VP (►Anhang D.5): VP-Ergebnisse sind nur beschränkt untereinander vergleichbar (VP-Merkmale, soziodemographische Einflüsse), sinnvoller ist der EG-KG Vergleich derselben VP. Zudem bestanden alle VP aus Freiwilligen, die nicht dem Durchschnitt der Population entsprechen müssen⁹⁷. Generell wurde für die Auswertung hier keine VP verworfen, weil sie das Versuchsziel zu früh erkannte, Geräte oder Ergebnisse ausfielen oder sie Anforderung nicht genügte. Am Experiment nahmen 6 VP teil (3 weitere fielen kurzfristig aus), Geschlecht (2 Männer, 4 Frauen), Alter 26-47 Jahre (Durchschnittsalter: 33,8), alle sprachen deutsch und waren Angestellte mit Wochenarbeitszeit von 40 (5 VP) oder 31 Stunden (1 VP), von denen der Großteil im Büro verbracht wird⁹⁸. VP besaßen Abitur (1), Ausbildung (2) oder Studium (3). Alle VP nutzen im Arbeitsalltag primär den Computer (Vorerfahrung) und erhielten für die Versuchsteilnahme kostenlos Snacks und Getränke. Der Versuch dauerte pro VP ca. 1,5 h. Wegen niedriger Beteiligung⁹⁹, VP-Ausfällen infolge technischer Ausfälle an der FHB¹⁰⁰ (2-tägige Wartungsarbeiten) und kurzfristiger VP-Absage, waren Geschlecht und Branche (Amt vs. Fir-

⁹⁶ Eine VP wurde in der Pause zwischen KG-EG misstrauisch („Ihr seid putzig“, „Jetzt hat sie [VL2] mich verwirrt, das war doch Absicht.“) und verhielt sich anders (ignorierte u.a. Skype-Anruf „Kannste aber voll vergessen!“ und grinste in D2).

⁹⁷ Nach [RR75, S.1955ff] ergänzt durch [EB77], zit. nach [BN06, S.73], unterscheiden sich Freiwillige durch bessere schulische Ausbildung, bessere Schätzung des eigenen sozialen Status, höhere Intelligenz, höheres Bedürfnis nach sozialer Anerkennung und Geselligkeit, unkonventioneller, eher weiblich, weniger autoritär oder konformistisch.

⁹⁸ Retrospektiv wäre es ggf. hilfreicher, statt „Anzahl im Büro verbrachter Stunden“ direkt Computerarbeitszeit zu erfragen.

⁹⁹ Es konnten nur 9 der geplanten 12 VP angeworben werden.

¹⁰⁰ Versuchsausfall für der ersten beiden Tage, 4 VP konnten auf neue Termine umorganisiert werden, 2 (männliche) VP fielen aus.

ma¹⁰¹) nicht ausbalanciert (1 VP aus Amt, 5 VP aus 3 unterschiedlichen Firmen). Insgesamt wurden schriftlich per Email 14 Firmen und 2 Behörden angefragt. Rückmeldungen auf anonyme Emails fielen gering aus, mehr Freiwillige wurden akquiriert, wenn zuvor mündlich oder schriftlich ein direkter Ansprechpartner (persönlich) kontaktiert wurde (interne Weiterleitung der Email).

VL: VL überwachten sich gegenseitig und gaben sich Hinweise per Skype-Chat (vergessene Textteile, Fehler, Hinweise, Beobachtungen für das Protokoll). Die VL waren in den HE schneller und routinierter (Übung im VE, reibungsloserer Ablauf bzgl. uV/Verkabelung/Einführung, weniger Fehler). Dennoch waren VP nicht exakt gleichen Bedingungen ausgesetzt (variierende Inhalte, unterschiedliche Reihenfolge einzelner Schritte speziell bei der Einführung). Hier sollte später ein strengere Ablaufplan vorliegen sowie mehrmaliges Üben ohne VP und unter Beobachtung Dritter.

Abläufe: Die für die uV definierten 3 Abläufe wurden ausbalanciert den VP zugewiesen (jeder Ablauf kam einmal in D1 als EG und KG vor), um Abhängigkeit der Ergebnisse mit den Abläufen zu vermeiden (Verzerrung, Positions-, Reihenfolgeeffekte). Dieses Vorgehen lohnt sich aber erst bei signifikant höherer VP-Anzahl. Daher sind an dieser Stelle keine brauchbaren Aussagen möglich und es war nicht nachweisbar, ob sich die Reihenfolge KG-EG in D1 bzw. D2 auf das Ergebnis auswirkte.

VP-Aufgaben: Sinnvoller wäre vielleicht ein *kompletter* Probedurchlauf vorab, damit VP sich an Aufgaben und Ablauf gewöhnen können und anschließend mit voller Leistungsfähigkeit starten¹⁰². Zudem sollten vielleicht in der Auswertung zusätzlich die Bearbeitungszeit der MWT- und IQ-Aufgaben in Minuten erhoben werden, um hier noch dezidierter auszuwerten.

Kommunikation (uV): Ab VP 3 wurde der Skype-Name des VL2 von „FH Brandenburg“ auf „Claudia Matthias“ geändert, um den Kontakt für die VP persönlicher zu gestalten und Kommunikation zu ermutigen. Einige VP waren mit Skype vertraut (VP3) oder kannten das Programm, nutzten es aber privat nicht (3 VP). Eine VP kannten Skype gar nicht.

Emails: Wegen technischer Schwierigkeiten konnte die Email-Anwendung *Thunderbird* im HE nicht genutzt werden (stattdessen Web-Anwendung *web.de*), sodass Emails nicht wie geplant via Popup auf dem VP-Monitor angezeigt wurden und die uV so zu verändern war, dass Emails (uV) vorab via Chat oder Telefon (uV) angekündigt wurden¹⁰³.

Arbeitsleistung (aV): Die aktive Arbeitszeit wird in der EG durch das Auftreten der uV reduziert (Telefon, Eintreten des VL)¹⁰⁴, sodass insignifikante Unterschiede der Arbeitsmenge (Y2) in EG und KG vernachlässigbar sind (keine aussagekräftigen Ergebnisse, Wertung unter Vorbehalt). Es wird befürchtet, dass die kurze Versuchszeit hier keine deutlichen Unterschiede zwischen EG und KG zulässt. Ggf. lässt sich mittels der Biosignale wenigstens später ermitteln, wie lange der uV-Effekt nachwirkte (Ablenkung/Stress, Dauer bis Aufgabenbearbeitung fortgeführt wird). Daher spielen in

¹⁰¹ Branchen sind Liegenschaften / Bauen, Webprogrammierung und -design, Geodaten und Satellitenbilder, Layout und Druck.

¹⁰² Es bestanden einige Probleme im Ablauf, u.a. Lern- und Übungseffekte, Ungewissheit der VP welche Aufgabe wann und wie zu erledigen sei. Die Probeaufgaben konnten hier nicht viel ausrichten.

¹⁰³ Ab VP4 konnte „Mail-Check“ installiert werden, was Popups und Audiosignale bei neuen Emails für *web.de* bereitstellte.

¹⁰⁴ Es wurde bei der Entwicklung der Abläufe angestrebt, alle VP gleichen Bedingungen auszusetzen, d.h. die tatsächliche „Kommunikationszeit“ (Auftreten der uV) in allen Abläufen etwa konstant zu halten.

diesem Versuchsaufbau Fehleranzahl (Y1), sowie Auswirkungen auf physiologische Signale (Z) eine größere Rolle. In nachfolgenden Studien bietet sich ggf. bei längerer Versuchszeit eine Differenzierung in „Totale Versuchszeit“, „Dauer der Unterbrechung“ („Zurückfindung in Aufgabenbearbeitung“ inklusive oder als Extra-Wert) und „Zeit, die mit der tatsächlichen Aufgabenbearbeitung verbracht wurde“ an, um exaktere Ergebnisse zu erhalten.

Probleme: Zwischendurch traten praktische Schwierigkeiten auf u.a. war das Eyetracking mehrmals einzustellen, weil Aufnahmen nicht gespeichert wurden, Morae verursachte nach Eyetracking-Kalibrierung Probleme, die uV (Email) konnte nicht zeitgenau versendet werden, da Probleme mit der Synchronisation im Ablauf und dem Postfach bestanden (Abrufintervalle des Mailservers), der Versuch musste erneut gestartet werden, da Morae nach 10 Sekunden abbrach oder ganz ausfiel, ProComp-Aufnahmen wurden nicht gespeichert (Datenausfall), teils ergaben sich Probleme mit web.de (eingehende Emails akustisch oder via Popup nicht angekündigt, sodass VP sie übersahen), in einer EG stürzte der Teamviewer in der 5.Minute ab, sodass der VL2 „taub und blind“ war und nur schwer die uV mit dem VL1 synchronisieren und Protokoll führen konnte, Messungen wurden aufgrund starken VP-Bewegung gestört, EKG-Sensor fiel mehrmals ab und verzögerte den Ablauf, Sitzhaltung der VP musste mehrmals justiert werden, Skype stürzte nach Anruf ab und war neu zu starten (Aufblinken auf VP-Monitor).

Unregelmäßigkeiten: Der Versuch wurde teils durch externe Störungen beeinflusst: u.a. 30-sekündiges lautes Hundegebell, ständiges Vibrieren des VP-Handys, Aufstehen des VL1, Reinigungsfrauen vor dem UL in einem Durchlauf (laute Gespräche, Staubsauger, Poltern etc.), zudem wurde hier VL2 gestört (Eintreten der Reinigungsfrauen) und anschließend die uV verzögert abgeschickt. Aufgrund unzureichender VP-Anzahl und SV ist nicht ermittelbar, inwieweit die Normalwertmessung (KG) tatsächlich korrekte Werte erhebt oder bereits verzerrt vorliegt.

Vermutung: VP waren zu vielen Informationen/zu viel Neuem auf einmal ausgesetzt (speziell, wenn Kommunikationsmittel vorher nicht bekannt waren), sodass Instruktion und Ablauf nicht verinnerlicht wurden und VP z.T. überfordert wirkten, was sich ggf. auch in der geringen Reaktion auf uV äußerte.

Hypothese: Verifikation allgemeingültiger Aussagen ist nur durch Untersuchung der Gesamtpopulation möglich, d.h. anhand von Stichprobendaten getroffene Schlussfolgerungen sind logisch nicht zulässig. Das Ziel wissenschaftlichen Arbeitens besteht darin, Theorien auf ihren Bewährungsgrad hin zu überprüfen (Falsifikationsprinzip d.h. Eliminierung falscher/schlecht bewährter Aussagen, konträre *Einzelfälle* sind zulässig) (►3.11.1). Veränderungshypothesen mit zweifaktoriellem Messwiederholungsplan gelten u.a. als bestätigt, wenn der Haupteffekt EG vs. KG signifikant ist (gleichförmige Unterscheidung über alle Messungen hinweg) (s. [BN06, S.550]). Aufgrund o.g. Schwierigkeiten und Verzerrungen (►3.15 ff), eingeschränkter Validität (►3.14) und den uneindeutigen verzerrten Ergebnisse können keine validen Aussagen zur Hypothese gemacht werden. Ziel dieser Arbeit ist jedoch ein erster Versuchsaufbau und probeweise Durchführung. Gemessene Daten erheben noch keinen repräsentativen Anspruch, die Hypothese soll noch nicht als tatsächlich bewährt nachgewiesen oder verallgemeinert werden. Daher ist die Unsicherheit der Schlussfolgerungen akzeptabel.

TEIL V

DISKUSSION

5. AUSWERTUNG UND ZUSAMMENFASSUNG

Ziel war die Entwicklung eines Forschungsdesigns für einen Versuch im Usability-Labor zu einem vorgegebenen Thema. Ermittelte Daten erhoben keinen Anspruch auf Repräsentativität und dienten primär der Bewertung von Design und Durchführung. Nachfolgend sind Planung und Vorgehen zu diskutieren, um Ergebnisse und Relevanz des Versuchs zu bestimmen und mit einem Fazit und Ausblick zu enden.

5.1 Planung und Durchführung

Allgemein: Finanzielle, personelle, zeitliche und technische Grenzen sowie ethische Bedenken erschwerten die Konzeption eines „optimalen“ Untersuchungsplans. Besonders die kurze Versuchszeit wirkte sich negativ auf die Manipulation der unabhängigen Variable aus. So ließ sich hauptsächlich „erzwungene“ einseitige Kommunikation erreichen, es kam kaum aktiver, von der Versuchsperson initiiertes Dialog mit Versuchsleiter 2 zustande. Entsprechend waren Auswirkungen der unabhängigen Variable nur begrenzt prüfbar und die Arbeitssituation wirkte noch unnatürlicher (geringere Immersion in Aufgaben, keine Eingewöhnungsphase). Replikationsstudien sollten längere Versuchszeiten einschließen, um eine natürlichere Arbeitssituation zu schaffen und unabhängige Variablen sinnvoller zu variieren / einzusetzen und ggf. Geräte verwenden, die die Versuchsperson körperlich weniger stark einschränken (mehr Handlungsfreiheit, ggf. natürlicheres Verhalten).

Akquirierung der Versuchspersonen: Diese war zeitintensiv mit hohem organisatorischem Aufwand. Die gewünschten 12 Versuchspersonen im Hauptexperiment kamen nicht zustande (wenige Freiwillige, Ausfälle). In nachfolgenden Studien sind Versuchspersonen ggf. finanziell zu entschädigen. Ausfälle technischer Geräte und Wartungsarbeiten an der Fachhochschule (unplanmäßig über 2 Tage), zogen mehrmaliges Umorganisieren von Versuchspersonen, Versuchspersonenausfälle, Missverhältnisse (u.a. Alter, Geschlecht) und höhere Catering-Kosten nach sich. Stichproben von Replikationsstudien sollten repräsentativ und ausbalanciert als Zufallsauswahl vorliegen (z.B. Quotenstichprobe) mit mind. 30-40 Versuchspersonen (►3.12).

Beobachtung: Es wurde nur das Monitor-Geschehen in den Beobachtungsraum übertragen (aufgrund technischer Grenzen). Versuchsleiter 2 konnte daneben noch die Gespräche im Usability-Labor (Audio) mitverfolgen, band die unabhängige Variable ein und tauschte sich per Chat mit Versuchsleiter 1 aus. Zukünftig sollte die Versuchsperson selbst mit aufgezeichnet werden und ggf. die Integration der unabhängigen Variable und Protokollierung auf mehrere Versuchsleiter verteilt werden (Doppelbelastung und Fehleranfälligkeit senken).

Versuchsplan: Großer Vorteil des Messwiederholungsplans (gleiche Versuchsperson in Kontroll- und Experimentalgruppe) lag im Vergleich der Experimental - und Kontrollgruppenwerte ohne personelle Störvariablen¹⁰⁵. Nachteilig wirkten sich Lern- und Übungseffekte in Durchgang 2 aus, daher

¹⁰⁵ „In repeated measures designs [...] variability among the subjects due to individual differences is completely removed from the error term. This makes these designs much more powerful than completely randomized designs, where different subjects are randomly assigned to the different treatments“[St96, S.402] zit. nach [Fa00, S108]

wurde die Reihenfolge von Experimental- und Kontrollgruppe in Durchgang 1 und 2 je Versuchsperson variiert. Störvariablen waren aber trotzdem nicht auszuschließen (Ergebnisse der ersten Messung oft schlechter). Zudem wurden Versuchspersonen zu vielen Informationen auf einmal ausgesetzt, besonders bei der Experimentalgruppe im 1. Durchgang (Informationen zu Ablauf, Geräten, Aufgaben, Kommunikationsmitteln, Versuchsleitern). Anweisungen wurden vermutlich nicht verinnerlicht oder vergessen, Texte nicht vollständig oder sorgfältig gelesen und Versuchspersonen überlastet. Um dies zu vermeiden, Lern- und Übungseffekte im 1. und 2. Durchgang konstant zu halten (Ablauf und Aufgaben üben) und Versuchspersonen empfänglicher für die unabhängige Variable zu machen, erweist sich ggf. ein Probendurchlauf vor jedem Versuch als praktisch oder ein Versuchsplan mit 3 Durchläufen der Form Kontrollgruppe-Experimentalgruppe-Kontrollgruppe, ggf. Solomonpläne. Dies erhöht zwar die Versuchszeit (Ermüdungsgefahr, sinkende Konzentration, weniger Freiwillige) lässt aber eindeutiger Ergebnisse zu und eliminiert Störvariablen. Daneben sind Einführungstexte zu kürzen, um den Ablauf einfacher zu gestalten. Um von der Versuchsperson initiierte Kommunikation zu ermutigen, ist die Arbeitssituation natürlicher zu gestalten. Um ggf. Interaktion zu erproben (u.a. gegenseitige Hilfe), wären Gruppenversuche (mehrere Arbeitsplätze) denkbar. Versuchspersonen verhalten sich hier ggf. natürlicher, da sie nicht allein beobachtet werden (höhere Interaktion, technische Möglichkeiten und Gruppendynamik bedenken). Insgesamt ist das Design in dieser Arbeit eher vor-experimentell (Prinzip Vorher-Nachher-Messung) mit zu vielen Störvariablen (u.a. Zeiteinflüsse, Reaktivität von Versuchsperson/Versuchsleitern, Reifung, Testeffekte). Es bedarf nachfolgend sowohl einer repräsentativen Stichprobe als auch weiterer Kontrollstrategien (u.a. Randomisierung), um aussagekräftige Ergebnisse mit kausalthoretischem Anspruch zu gewährleisten. Im Rahmen einer ersten Pilotstudie mit einfacher Auswertung ist dieses Vorgehen jedoch durchaus zulässig.

Design: Versuchsplan, Durchführung und wichtige Eckpunkte sind in ►4 zusammengefasst, um Replikation durch Andere zu ermöglichen und die Güte der Untersuchung prüfbar zu machen. Angestrebt wird hohe Objektivität, routinemäßige Anwendung und höhere Nachvollziehbarkeit. Neben dem genutzten Messwiederholungsplan, eignen sich später auch „2-Gruppen-Randomisierungspläne“.

Geräte und Kommunikation: Die Durchführung im Usability-Labor unter Anbringung von Sensoren an der Versuchsperson und Verwendung von Remote Eyetracking erforderte eine starre (unnatürliche) Sitzhaltung, mit wenig Bewegungsfreiheit (speziell Kopf und Augen). Das Bewusstsein über die Empfindlichkeit der Geräte wurde durch Hinweise an die Versuchsperson verstärkt und hatte zur Folge, dass Versuchspersonen sich fast gar nicht bewegten und z.B. beim Eintreten des Versuchsleiters 2 (unabhängige Variable „Störperson“) nur minimal reagierten. Dies erschwerte die Simulation einer „natürlichen“ Arbeitssituation, Versuchspersonen waren sich durchgängig bewusst, dass sie sich in einem Versuch befanden, was ggf. natürliche, beidseitige Kommunikation hemmte (kein oder wenig Blickkontakt, „gezwungene“ Atmosphäre, „fremde“ Versuchsleiter, kurze Versuchszeit d.h. sehr fokussiertes Arbeiten). Kommunikation beschränkte sich im Versuch meist auf einseitige „erzwungene“ Kommunikation durch den Versuchsleiter.

Ergebnisdarstellung: Es wurde *Häufigkeitsverteilung* genutzt, um Versuchspersonen-Antworten und Messungen (Kontrollgruppe, Experimentalgruppe) derselben und verschiedener Versuchspersonen gegeneinander darzustellen (Streuung, Nutzung des Wertebereichs, Konzentration). Ergebnisse wurden zudem neutral in Tabellenform aufgeführt (► Anhang D).

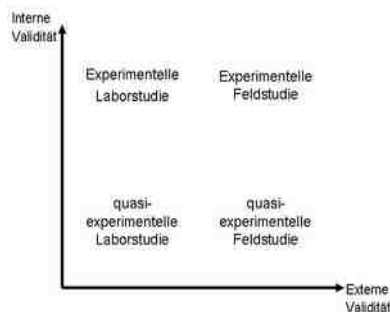


Abb. 11 Validität [Wi12a]

Erhebungsmethoden, inhaltliche und methodische Implikationen für weitere Forschung: Laborversuche erfordern weniger Zeit und Kosten (als im Feld) und gewährleisten optimale Störvariablen-Kontrolle d.h. bessere eindeutige kausale Schlüsse (systematische Manipulation der unabhängigen Variablen, Wiederholbarkeit) und entsprechend höhere *interne Validität*. Es deutet sich aber an, dass Laborversuche zur Hypothesenprüfung hier nicht ausreichen, um (intern *und* extern) ausreichend valide Ergebnisse zu liefern und

Replikationsstudien sowie Studien im Feld (oder eine Kombination) erforderlich sind¹⁰⁶ (► Abb.11). Ergebnisse aus Laborversuchen sind ohne weitere Zusatzinformationen nur auf vergleichbare Labor-situationen generalisierbar, wenn Störungsfreiheit durch restriktive Maßnahmen bzw. unnatürliche Rahmenbedingungen erkauft wurde. Die Anwendbarkeit der Ergebnisse auf praktische, reale Probleme ist zweifelhaft (künstlich, Versuchspersonen wissen um Experiment, befinden sich nur *kurz* im Experiment), Ergebnisse sind nicht einfach auf andere Personen, Settings, Zeitpunkte oder Orte verallgemeinerbar. Versuchspersonen wissen sich beobachtet, wodurch mit Beeinflussung des Verhaltens zu rechnen ist und „Konstanthalten“ erlaubt zwar, dass Ergebnisse im gesamten Experiment denselben Wert haben, die Störvariablen bestehen aber weiter und schränken die Generalisierbarkeit der Befunde auf die Realität ein (geringere externe Validität). Zudem sind die Versuchspersonen hier nicht repräsentativ (fehlende Zufallsstichprobe) und zahlenmäßig nicht ausreichend für sichere empirische Befunde. Validität lässt sich aber durch Replikation aufzeigen (weitere Bewährung für Anwendungsfeld, Bedingungsprüfung). Ideal sind *unabhängige* Replikationen (cross-laboratoy replication) mit anderen Laboren und/oder Versuchspersonen-Gruppen. Es bieten sich Untersuchungen in Labor-situationen mit variierenden Rahmenbedingungen an (u.a. Versuchszeit, Stichprobe, Geräte, Kommunikationsmittel), sowie Feldstudien, um externe Validität zu gewährleisten. Hier durchgeführte Versuche sind nur erste Studien an kleinen, leicht zu erreichenden Gesamtheiten (praktikable ökonomische Rekrutierung), die auf fälschliche Generalisierung verzichten und keinen Anspruch auf Allgemeingültigkeit der Ergebnisse erheben (nicht-probabilistisches Verfahren zur Stichprobenauswahl).

Aussagegehalt der empirischen Untersuchung: Insgesamt wurde hier interne Validität begünstigt. Operationalisierung der Konstrukte erfolgte anhand *mehrerer* abhängiger Variablen, um Ergebnisse besser zu kontrollieren und zu bestätigen (höhere Reliabilität). Nach weiteren Laborversuchen sollten Feldstudien durchgeführt werden, mit natürlicheren Untersuchungsbedingungen und größerer Ähn-

¹⁰⁶ Bzgl. interner und externer Validität ist „experimentelle Felduntersuchung“ allen anderen Kombinationen überlegen [...]. Dies gilt zumindest für die hypothesenüberprüfende Forschung [...]“[BN03, S.61] z.B. Feldversuche mit Methoden des Usability-Labors (u.a. Eyetracking). Es gilt zu prüfen, ob Geräte mobil verfügbar sind und sich für lange Messungen eignen (Datenmenge).

lichkeit von u.a. Versuchspersonen, Treatment und Wirkung zu Population und Situation, für die die Untersuchung gültig sein soll, um externe Validität zu erhöhen (natürliche Arbeitssituation und Kommunikation, repräsentative Zufallsstichprobe) und Generalisierung der Ergebnisse zuzulassen (Nachteil: Einschränkung der Messmöglichkeiten, schwierigere Auswertung). Die hier gewählte Extremsituation (viel variierendes Kommunikationsauftreten in kurzem Zeitraum¹⁰⁷), um in der begrenzten Versuchszeit die unabhängige Variable sinnvoll einzuarbeiten, sollte nachfolgend durch eine ausgeglichene, natürliche Situation balanciert werden (längere Versuchszeit, besser verteilte „natürliche“ Kommunikation/-smittel für höhere externe Validität). Es bestehen zudem Schwierigkeiten und Störungen (►3.15, 3.5.4, 4.5), die noch nicht eliminiert oder kontrolliert werden konnten und Ergebnisse verfälschten (u.a. Unerfahrenheit der Versuchsleiter, begrenzte technische Mittel und Zeit, Versuchspersonen-Probleme d.h. Anzahl, Repräsentativität, Auswahl). Störvariablen, alternierende Erklärungen und mehrdeutige Interpretierbarkeit lassen sich hier also nicht ausschließen.

Gültigkeit der Hilfstheorien: Operationalisierung und Erhebung des Konstrukts *Ablenkung* sowie die Analyse physiologischer Messungen erfolgte außerhalb dieser Arbeit. Die Operationalisierung der *Arbeitsleistung* mittels „Qualität“, „Arbeitsmenge“ (und „IQ-Werte“) erscheint aussagekräftig und angemessen, wobei die kurze Versuchszeit hier „Qualität“ und „IQ-Werte“ in den Vordergrund rückt¹⁰⁸. Die zusätzliche IQ-Erhebung soll Ergebnisse unterstützen (Konstruktvalidität), der Einsatz mehrerer abhängiger Variablen die Ergebnisse sichern (Bestätigung) und die Nutzung genormter, wissenschaftlich anerkannter Tests (speziell MWT) reliable, intern valide, objektive und wiederholbare Aufgaben(ergebnisse) gewährleisten. Leistungstests lassen sich entsprechend austauschen, um unterschiedliche Bereiche der Intelligenz zu untersuchen z.B. Problemlöseverhalten. Die Fragebögen ergänzten den Versuch sinnvoll, zur Bewertung von Planung und Durchführung (u.a. Versuchspersonen-Aufgaben, Ablauf), Beurteilung der Zulässigkeit der Versuchspersonen (soziodemographische Daten, Angaben über Arbeitsweise, Verhalten, Versuchspersonen-Kriterien) und Ergebnisprüfung (Stimmungsfragebogen, subjektives Empfinden bzgl. Ablenkung, Druck, Arbeitssituation). Das Debriefing erwies sich als hilfreich, um aufzudecken, inwieweit das Versuchsziel vorzeitig aufgedeckt wurde und festzustellen, inwieweit Geräte das Versuchspersonen-Verhalten einschränkten (subjektive Aussagen) und um Organisatorisches zu klären (Datenschutz, Schweigepflicht, Einsicht in eigene Daten). Technische Anforderung erwiesen sich als umfangreich und (zu zweit) kaum zu bewältigen (Geräteausfall, langsamer Versuchspersonen-Rechner durch Überlastung, problematische Synchronisation aller Geräte etc.).

Weitere Vorschläge: Es empfiehlt sich eine Versuchsleiter-Schulung, um Störvariablen aufzudecken, Versuchsleiter-Fehlern vorzubeugen und gleiche Versuchsbedingungen für alle Versuchspersonen zu schaffen (geübte, souveräne und ruhige Durchführung, immer gleiches Verhalten und Wortlaut). Zudem sollte die Betreuung der Technik (Vor-/Nachbereitung) mithilfe eines oder mehrerer Assistenten

¹⁰⁷ Auch fehlende Vertrautheit mit Versuchsleitern und Aufgaben, ungewohnte Situation, „erzwungene“, unnatürliche Kommunikation

¹⁰⁸ Zusätzlich böte sich ggf. noch die Erhebung der Bearbeitungszeit des MWT in Minuten an (eindeutigere Daten).

erfolgen¹⁰⁹, ein leistungsstärkerer Versuchspersonen-Rechner verwendet und ggf. Blind- oder Doppelblindversuche mit dem Versuchsleiter im Labor angestrebt werden, um Störvariablen zu vermeiden. Insgesamt initiierte die Versuchspersonen zu viel Kommunikation mit dem Versuchsleiter 1 (Zwischenfragen) und wendete sich selten an den Versuchsleiter 2 (eigentlicher Ansprechpartner). Falls technisch machbar¹¹⁰, sollte die Versuchspersonen zukünftig gänzlich allein im Labor arbeiten, ohne direkte hemmende Beobachtung und mit Versuchsleiter 2 als alleinigem Ansprechpartner. Dies könnte die Kommunikation zwischen Versuchsleiter 2 und der Versuchsperson wie gewünscht erhöhen, natürlicheres Versuchspersonen-Verhalten ermutigen und von der Versuchsperson initiierte Kommunikation ermöglichen (Nutzung Handy etc.). Da auffällig war, dass Versuchspersonen, die mit den Kommunikationsmitteln nicht vertraut waren, diese im Versuch selten bis gar nicht nutzten (und die unabhängige Variable ignorierten), ist zu überlegen, Versuche auf Versuchspersonen zu beschränken, die diese kennen (Begrenzung der Hypothese) oder vorab mit der Versuchsperson die Verwendung zu üben (z.B. Probedurchlauf).

5.2 Ergebnisse und Generalisierung

Generalisierung: Wie oben angegeben, empfehlen sich nach weiterer Verbesserung des Designs Replikationsstudien durch unterschiedliche Experimentatoren (cross-laboratory replication) und ggf. unter variierenden Bedingungen, um Befunde zu sichern und systematische Fehler, Labor- oder Forscherbezogene Verzerrung frühzeitig aufzudecken: „Man sagt, dass die Ergebnisse von Experimenten dann als gesichert und generalisierbar gelten, wenn mind. drei erfolgreiche Replikationen vorliegen.“[el09] Trotz der kritisierten Unnatürlichkeit von Laborversuchen (geringe externe Validität), kann es ausreichen „ein Design zu wählen, welches der Proband als realistisch empfindet (experimental realism) und sich demnach normal verhält, auch wenn der Versuch künstlich konstruiert wurde.“[Re03, S.15]. Fokus liegt hier auf einer Arbeitssituation, die Versuchspersonen ermutigt, sich möglichst wie bei „normaler (unbeobachteter) Arbeit am Computer“ zu verhalten. Generalisierbarkeit hängt zudem von der Ergebnisanwendung ab: es werden Effects- vs. Theory Application unterschieden (s.[Re03]): *Effects Application* strebt auf die reale Welt verallgemeinerbare Ergebnisse an, d.h. hohe Übereinstimmung mit experimentellen Bedingungen (u.a. Feldexperiment, für die Population repräsentative Versuchspersonen) während *Theory Application* zur Identifikation wissenschaftlicher Theorien genutzt wird (Theorieprüfung statt Gewinnung allgemeingültiger Regeln, Technik der Falsifizierung). Um kausale Zusammenhänge zu untersuchen (basic research), eignen sich z.B. auch Studenten, nicht aber für Vorhersage von Verhalten (applied research). Zusammenfassend ist die Untersuchung im Usability-Labor zur Grundlagenforschung ein sinnvoller Ansatz, trotz geringer Repräsentativität der Versuchspersonen, solange sie der Überprüfung wissenschaftlicher Theorien dient. Es sollte aber später auch im Feld getestet werden, um Ergebnisse auf die Praxis anzuwenden.

¹⁰⁹ Da speziell Geräteanbringung und -bedienung Zeit kosteten, wäre ein Assistent sinnig, um Fehler zu vermeiden (u.a. unsynchrone Geräte, Datenausfall) und Versuchszeit zu kürzen und den Versuch somit für potentielle Versuchspersonen attraktiver machen.

¹¹⁰ Ggf. muss hier über reduzierte, weniger aufwendige und fehlertolerantere Messmethodik nachgedacht werden.

5.3 Fazit und Ausblick

Leistungstests zur Darstellung von Leistungsänderung stellen ein wiederholbares, eindeutiges Verfahren dar. Ähnlich der Wilson-Studie wurden auch Biosignale gemessen (physiologische Indikatoren) und Befragungen durchgeführt. Demgegenüber wurde die Versuchsdauer hier aufgrund technischer Einschränkungen stark gekürzt und entsprechende Aufgaben gewählt (MWT Kurztests, unvollständige IQ-Tests). Im Vordergrund standen Ablenkung und Arbeitsleistung (Arbeitsmenge, Qualität, inkl. IQ-Werte). Der Versuchsaufbau wurde beschrieben und möglichst genormt (Text- und Dokumentvorlagen, definierte Abläufe, Debriefing, schriftliche Fragebögen), um Wiederholbarkeit und Überprüfbarkeit zu gewährleisten. Insgesamt knüpft das entwickelte Forschungsdesign an Wilsons und weitere Studien an (►3.1), konzentriert sich aber auf Grundlagenforschung (interne Validität) mit einigen Erweiterungen (Messungen im Usability-Labor, komplexere Messungen u.a. Verfahren wie Eyetracking). Die Erhebung konkreter Auswirkungen von Kommunikation auf die Leistung besitzt für die Wirtschaft große Relevanz (zukünftige Arbeitsplatzgestaltung). Viele Studien zum Thema basieren aber auf Feldstudien, Beobachtung und Befragung, oft mit niedriger Versuchspersonen-Anzahl, unausgeglichenen oder wenig repräsentativen Stichproben (u.a. Alter, Geschlecht, In-House Studien), teils unwissenschaftlich, ohne Angaben über Versuchsaufbau und Bedingungen (nicht nachvollziehbar oder durch andere replizierbar / überprüfbar). Weiterführende streng empirische Forschung auf dem Gebiet erscheint also durchaus sinnvoll, um konkrete Auswirkungen und Auslöser zu erfassen und Kommunikation in Unternehmen zu strukturieren (praktische Bedeutung z.B. Benutzer-Aufmerksamkeits-Management). Das eigene Vorhaben soll einen Beitrag zur Grundlagenforschung leisten und allgemeine Effekte von Kommunikation und gängigen Kommunikationsmitteln während der Arbeit prüfen, um den Bestand gesicherten Wissens im Untersuchungsbereich zu erweitern und ggf. unwissenschaftliche Studien bzw. Aussagen wissenschaftlich zu bestätigen oder zu falsifizieren. Stellt sich die Hypothese in nachfolgenden Replikationsstudien als ausreichend bewährt heraus, lässt sich z.B. testen, ob die Anzahl der „Unterbrechungen“ (zeitliche Intervalle), Kommunikationsart (einseitig vs. Dialog, selbst- vs. fremdinitiiert) oder Kommunikationsmittel die Arbeitsleistung zusätzlich beeinflussen und ob die abhängige Variable von weiteren (Organismus-)Variablen wie Altersgruppe (Generation) oder Medienaffinität beeinflusst wird.

TEIL VI

VERZEICHNISSE UND ANHÄNGE

LITERATURVERZEICHNIS

- [AB04] **Adamczyk P. D.; Bailey B. P.:** “If Not Now, When?: The Effects of Interruption at Different Moments Within Task Execution“(Paper) Graduate School of Library and Information Science and Department of Computer Science, University of Illinois at Urbana-Champaign Champaign, 2004
- [An08] **Antoniw, K.:** „Intelligenz und Intelligenztests“ (Seminar "Intelligenz- und Leistungsdiagnostik" (Pädagogische Psychologie II) Stand 07/2008
URL: http://wulv.uni-greifswald.de/2008_ka_iq_leistung/userdata/IntelligenzundIntelligenztests.ppt (22.03.2012, 18:03Uhr)
- [As01] **Aschemann-Pilshofer, B.:** „Wie erstelle ich einen Fragebogen“ (2.Aufl.) Jänner 2001, URL: <http://www.aschemann.at/Downloads/Fragebogen.pdf> (19.03.2012, 11:46 Uhr)
- [Bi98] **Bienert, P.:** „Information & Kommunikation – Technik und Anwendung in Wirtschaft und Medien“ Springer-Verlag Berlin Heidelberg, 1998
- [BN03] **Bortz J.; Döring N.:** „Forschungsmethoden und Evaluation“ (3. Aufl.) Springer-Verlag Berlin Heidelberg New York, 2003
- [BN06] **Bortz J.; Döring N.:** „Forschungsmethoden und Evaluation“ (4. Aufl.) Springer Medizin Verlag Heidelberg, 2006
- [BP02] **Bruns T.; Praun N.:** “Biofeedback – Ein Handbuch für die therapeutische Praxis” Vandenhoeck & Ruprecht, Göttingen, 2002
- [Br10] **Bry, F.:** „Dagstuhl Manifesto - Digital Social Media“ (Paper) HAW Hamburg, 2010
URL: <http://users.informatik.haw-hamburg.de/~ubicomp/papers.html> (01.03.2012, 10:07Uhr)
- [Br12] **brain mental systems:** „GSR“ (Stand 2012)
URL: http://www.brain-mentalsystems.de/stichwortverzeichnis/gsr.php?gb_dateiname=gsr (26.08.2012, 08:20Uhr)
- [CHW04] **Czerwinski M.; Horvitz E.; Wilhite, S.:** “A Diary Study of Task Switching and Interruptions” Artikel in CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM New York, NY, USA 2004; S. 175-182
- [CV04] **Chen, D.; Vertegaal, R.:** „Using Mental Load for Managing Interruptions in Physiologically Attentive User Interfaces“(Paper) Human Media Lab, Queen’s University, Kingston, Ontario, 2004
- [De12] **Deutsche Kaufberatung:** „Thermokamera“
URL: <http://www.deutsche-kaufberatung.de/waermebildkamera/thermokamera-51> (21.03.2012, 21:26Uhr)
- [Dp10] **news.de/dpa:** “Im ‚Usability‘-Labor werden Nutzer zu Versuchskaninchen“ (Stand 2010)
URL: <http://www.news.de/technik/855090522/im-usability-labor-werden-nutzer-zu-versuchskaninchen/1/> (23.12.2011, 08:52Uhr)
- [Dr97] **Drews, D.:** „Die Onlinegesellschaft“ Wirtschaftsverlag Langen Müller/Herbig, 1997

- [EB77] **Effler, M.; Böhmecke, W.:** "Eine Analyse des Verweigererproblems mit Beinahe-Verweigerern." Zeitschrift für experimentelle und angewandte Psychologie, 24, 1977; S.35-48
- [el09] **elearning (Petzoldt, J.):** „Versuchsplanung“ TU Dresden (Stand 01/2009)
URL: <http://elearning.tu-dresden.de/versuchsplanung> (17.03.2012, 7:40Uhr)
- [et07] **e-teaching.org:** "Usability Test" (Stand 2007)
URL: <http://www.e-teaching.org/didaktik/qualitaet/usability/> (22.12.2011, 15:55Uhr)
- [Ex12] **Experimentalpsychologie.de:** "Welche Experimente gibt es?"
URL: <http://www.experimentalpsychologie.de/page38.html> (31.03.2012, 19:05 Uhr)
- [Fa00] **Fahrenberg, J.; Klein, C.; Peper, M.; Zimmermann, P.:** „Versuchsplanung“ (Paper)
Psychologisches Institut der Universität Freiburg (Redaktion und Layout: U. Probst), 2000
- [FA03] **Fachhochschule Stuttgart, Hochschule der Medien:** „Usability“ (Stand 2003) (Auftritt zur Sonderbeilage in Stuttgarter Zeitung & Stuttgarter Nachrichten)
URL: <http://www.hdm-stuttgart.de/zeitungsbeilage/usability.htm> (22.12.2011, 15:32Uhr)
- [Fa97] **Fahrmeir, L.; Künstler, R.; Pigeot, L.; Tutz, G.:** „Statistik – Der Weg zur Datenanalyse“
Springer, Berlin, 1997
- [GD07] **Garrett, R. K.; Danziger, J. N.:** „IM = Interruption Management? Instant Messaging and Disruption in the Workplace“(Paper) Journal of Computer-Mediated Communication, 2007
- [Ge03] **Gerken, J.:** „Validität und Aussagekraft von Usability Test Methoden“ (Studienarbeit)
WiSe 2002/2003 (05.12.2011, 07:10Uhr)
- [Gö08] **Göddel, J.:** „Herausforderung der Wissensarbeit: Bewältigung von Unterbrechungen bei nebenläufigen Arbeiten“(Paper) Technische Universität Kaiserslautern/Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI GmbH) Kaiserslautern, Deutschland, 08/2008
- [HA03] **Horvitz, E.; Apacible, J.:** "Learning and Reasoning about Interruption" (Paper) Microsoft Research Redmond, Washington (USA), 2003
- [Ha10] **Haapalainen, E.; Kim, S. J.; Forlizzi, J. F.; Dey, A. K.:** „Psycho-Physiological Measures for Assessing Cognitive Load“(Paper) University of Oulo, Finnland, Carnegie Mellon University, USA, 2010
- [Hi12] **Hiemisch, A.:** „Sozialwissenschaftliche Methoden“, Universität Greifswald
URL: http://www.phil.uni-greifswald.de/fileadmin/mediapool/general_studies/Skript_Hiemisch_Methoden.pdf (19.03.2012, 11:12Uhr)
- [Ho97] **Hobmair, H. (Hg); Altenthann, S.; Betscher-Ott, S.; Dirrigl, W.; Gotthardt, W.; Ott, W.:** "Psychologie" (2. Aufl.) Köln: Stam Verlag, 1997
- [In12] **InfraTec:** „Eine Thermokamera unterstützt hochpräzise Fertigungsabläufe“
URL: <http://www.infratec.de/de/thermografie/thermografie-wissen/thermografie-schlagworte/thermokamera.html> (21.03.2012, 21:25Uhr)
- [JDW02] **Jackson, T.; Dawson, R.; Wilson, D.:** "Evaluating the Effect of Email Interruptions within the Workplace"(Paper) Danwood Group, Lincoln, Loughborough University, Leics., UK, 2002

- [**Ki06**] **Kirchhoff, S.; Kuhnt, S.; Lipp, P.; Schlawin, S.:** „Der Fragebogen – Datenbasis, Konstruktion und Auswertung“ (2. überarbeitete Aufl. 2003) VS Verlag für Sozialwissenschaften | GWV Fachverlag GmbH, Wiesbaden 2006
- [**Kl83**] **Klausnitzer, J. E.:** „Der IQ Selbsttest“ (9.Aufl.) Wilhelm Heyne Verlag München, 1983
- [**Le05**] **Lehrl, S.:** „Mehrfachwahl-Wortschatz-Intelligenztest MWT-B“ (5. Aufl. 2005) Balingen: Spitta Verlag GmbH & Co. KG, 2005
- [**Li11**] **Linek, S. B.:** „Drei Fragen (und Antworten) zum Usability-Lab“(Stand 2011)
URL: <http://www.zbw-mediataalk.eu/2011/05/drei-fragen-und-antworten-zumusability-lab/> (22.12.2011, 14:53Uhr)
- [**Mc99**] **McFarlane, D. C.:** “Coordinating the Interruption of People in Human-Computer Interaction”(Paper) Naval Research Laboratory, Code 5513. Washington, D.C., 1999
- [**Me11**] **Meißner, T.:** „Tutorial Usability-Labor – Eyetracking-System –“ (Paper) Fachhochschule Brandenburg, Stand 04.03.2011
- [**Me13**] **Meißner, T.:** „Untersuchung physiologischer Signale kognitiv-psychologischer Belastungen hinsichtlich berechenbarer Merkmale“ (Masterarbeit) Fachhochschule Brandenburg, 2013
(*noch nicht veröffentlicht*)
- [**Ne12**] **Netdokter.de:** „Gesundheit und Medizin“
URL: <http://www.netdokter.de> (21.03.2012, 20:34Uhr)
- [**Ot06**] **Otto, M.:** „Untersuchung der Validität des Erhebungsinstruments SaMon – Über die Validität eines Fragebogen“(Diplomarbeit) (1. Aufl. 2006) GRIN Verlag 2006
- [**Pe11**] **perceptionenhancement:** „Subjective Evaluation“
URL: http://www.perceptionenhancement.com/docs/human_factors_course/human_factors_course_lec12_subjective_evaluation.pdf (23.12.2011, 08:38Uhr)
- [**Ps12**] **Psychologie-Lexikon:** „Versuchsplanung“
URL: <http://www.psychology48.com/deu/d/versuchsplanung/versuchsplanung.htm>
(15.03.2012, 9:03Uhr)
- [**Re03**] **Reuß, T.:** „Die interne und externe Validität von Laborexperimenten“ (Seminararbeit) Johann Wolfgang Goethe-Universität, Frankfurt am Main, 2003
- [**Re05**] **Reichel, W.:** „Der große Intelligenztests“ (2. Aufl.) Klett-Cotta, Stuttgart 2005
- [**Ro87**] **Rohrmann, B.:** „Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung“ Zeitschrift für Sozialpsychologie, 9, 1987; S. 222-245
- [**RR75**] **Rosenthal, R.; Rosnow, R.L.:** “The Volunteer Subject” Wiley & Sons Inc, New York, 1975
- [**Sa92**] **Sarris, V.:** ”Methodologische Grundlagen der Experimentalpsychologie 2: Versuchsplanung und Stadien.” Ernst Reinhardt Verlag, München, 1992
- [**Sc06**] **Scoreberlin:** „Usability-Tests und Eyetracking“ (Stand 05/2006)
URL: <http://www.scoreberlin.de/usability-artikel/usability-tests-eyetracking>
(26.08.2012, 07:52Uhr)

- [Sc11] **Schmitt, I.:** „change@office: Virtuelle Chefentlastung – Die neue Form des Management Supports“ (1. Aufl.) (Paper) Gabler Verlag, Wiesbaden, 2011
- [SR05] **Sarris, V.; Reiß, S.:** „Kurzer Leitfaden der Experimentalpsychologie“ München: Pearson, 2005
- [St96] **Stevens, J.:** “Applied multivariate statistics for the social sciences” (3. Aufl.) Lawrence Erlbaum Associates, Inc., 1996
- [Te12a] **Testzentrale (Lehrl, S.):** „MWT-A: Mehrfach-Wortschatz-Intelligenztest“ (Stand 2012)
URL: <http://www.testzentrale.de/programm/mehrfachwahl-wortschatz-intelligenztest.html>
(12.04.2012, 11:53Uhr)
- [Te12b] **Testzentrale (Lehrl, S.):** „MWT-B: Mehrfach-Wortschatz-Intelligenztest“ (Stand 2012)
URL: <http://www.testzentrale.de/programm/mehrfachwahl-wortschatz-intelligenztest-1.html>
(12.04.2012, 11:56Uhr)
- [To09] **Tokuda, S.:** „An investigation into relationships between saccadic intrusions and mental workload“ (Dissertation) Wichita State University, 12/2009
- [Tr09] **Trimmel, M.:** „Wissenschaftliches Arbeiten in Psychologie und Medizin“ (1.Aufl.) UTB, Stuttgart, 2009
- [We56] **Wechsler, D.:** „Die Messung der Intelligenz Erwachsener“ Bern: Verlag Hans Huber, 1956
- [Wi10] **Wilson, G.:** „THE ‘INFOMANIA’ STUDY“ 2010
URL: <http://www.drglennwilson.com/links.html> (28.12.2011, 13:20Uhr)
- [Wi11a] **Wikipedia:** „Blickerfassung“ (Stand 2011)
URL: <http://de.wikipedia.org/wiki/Blickerfassung> (22.12.2011, 16:17Uhr)
- [Wi11b] **Wikipedia:** „Usability-Labor“ (Stand 09/2011)
URL: <http://de.wikipedia.org/wiki/Usability-Labor> (22.12.2011, 14:36Uhr)
- [Wi12a] **Wikipedia:** „Forschungsdesign“ (Stand 02/2012)
URL: <http://de.wikipedia.org/wiki/Forschungsdesign> (14.03.2012, 7:51Uhr)
- [Wi12b] **Wikipedia:** „Intelligenztest“ (Stand 03/2012)
URL: <http://de.wikipedia.org/wiki/Intelligenztest> (22.03.2012, 15:22Uhr)
- [Wi12c] **Wikipedia:** „Psychologisches Experiment“ (Stand 02/2012)
URL: http://de.wikipedia.org/wiki/Psychologisches_Experiment (14.03.2012, 11:00Uhr)
- [Wi12d] **Wikipedia:** „Sekundärvarianz“ (Stand 03/2012)
URL: <http://de.wikipedia.org/wiki/Sekund%C3%A4rvarianz> (14.04.2012, 11:29Uhr)
- [Wp12] **Wirtschaftspsychologische Gesellschaft:** „Forschungsdesigns“
URL: <http://www.wpgs.de/content/blogcategory/82/347> (15.03.2012, 11:16Uhr)
- [Zw98] **Zwiler, R.:** „Versuchsplanung“ (Stand 1998)
URL: <http://www.zwiler.de/scripts/methoden/node3.html> (16.03.2012, 8:19Uhr)

TABELLENVERZEICHNIS

TAB. 1	GEGENÜBERSTELLUNG DER DESIGNS UND MISCHFORMEN BZGL. RELEVANTER KRITERIEN ([FA00, S.83])	6
TAB. 2	ÜBERSICHT ÜBER DIE 3 WICHTIGSTEN EXPERIMENTELLEN DESIGNS (NACH [EL09])	7
TAB. 3	GÄNGIGE KLASSIFIKATIONEN VON EXPERIMENTEN (NACH [ZW98] [FA00] [TR09, S.65])	12
TAB. 4	VALIDITÄT DER PRIMÄREN EXPERIMENTTYPEN (NACH [BN03, S.61])	13
TAB. 5	MÖGLICHE ERGEBNISSE DER UNTERSUCHUNG (ERWARTUNGEN)	23
TAB. 6	VARIABLENTYPEN, GRUPPIERT NACH STELLENWERT, MERKMALS AUSPRÄGUNG UND EMPIRISCHER ZUGÄNGLICHKEIT	24
TAB. 7	VERSUCHSAPPARATUR IM UL (S. [NE12][IN12][DE12][BP02][SC06][W11A][ET07][DP10][BR12])	35
TAB. 8	LISTE DER EINGESETZTEN „KOMMUNIKATION UND GÄNGIGEN KOMMUNIKATIONSMITTEL“ (KATEGORIE I)	37
TAB. 9:	ÜBERSICHT ÜBER DIE AUSGEWÄHLTEN AUFGABEN MIT IHREN JEWEILIGEN PARALLELFORMEN, KÜRZELN UND DAUER	42
TAB. 10:	ÜBERSICHT ÜBER DIE ENTWICKELTEN VERSUCHSMATERIALIEN	42
TAB. 11	ECKDATEN ZUR VERSUCHSPLANUNG	44
TAB. 12	UNSPECIFISCHE STATISTISCHE HYPOTHESENPAARE FÜR TH1 UND TH2	56
TAB. 13:	ÜBERSICHT ÜBER DIE OPERATIONALISIERUNG (VARIABLENZUORDNUNG ZU DEN TEILHYPOTHESEN)	57
TAB. 14	DATENERHEBUNG: APPARATE, MITTEL, INTENDIERTE MESSUNG (DETAILS ► ANHANG D.2, B.2)	58
TAB. 15:	AUSGEWÄHLTE VP-AUFGABEN FÜR VE UND HE	58
TAB. 16:	ZUSAMMENFASSUNG DES VERSUCHSAUFBAUS	60
TAB. 17	VERSCHIEDENE ARTEN DER STICHPROBEAUSWAHL (ÜBERSICHT VGL. [EL09] [BN06])	A-13
TAB. 18	ÜBERSICHT ÜBER DIE RAUM-ORGANISATION FÜR DEN VERSUCH IM RAHMEN DIESER ARBEIT	A-16
TAB. 19	FINANZPLAN FÜR DIE VOR- UND HAUPTEXPERIMENTE	A-16
TAB. 20	ZEITPLAN FÜR VOREXPERIMENTE	A-19
TAB. 21	ZEITPLAN FÜR HAUPTEXPERIMENTE	A-19
TAB. 22	DEFINITION DER KOMMUNIKATION ZUR AKTIVEN MANIPULATION DER UNABHÄNGIGEN VARIABLE	A-25
TAB. 23	VORDEFINIERTER ABLÄUFE (VARIERT) FÜR DIE EXPERIMENTALGRUPPE: (1 MINUTE – 2 MINUTEN – 2 MINUTEN – 2 MINUTEN)	A-26
TAB. 24	ZEITABLAUF MIT ALLEN SCHRITTEN DES VERSUCHSABLAUFS FÜR DIE VERSUCHSLEITER	A-27
TAB. 25	AUFBAU UND FUNKTION DER VERSUCHSBETEILIGTEN	A-33
TAB. 26	ERGEBNISSE DER AUFGABEN (MWT-A/B, IQ-1/2) AUS DEN VOR- UND HAUPTEXPERIMENTEN FÜR ALLE DURCHGÄNGE 1 UND 2	A-41
TAB. 27	ÜBERSICHT ÜBER MATERIALIEN UND GERÄTE IN ALLEN VERSUCHEN (DURCHGANG 1 UND 2) DER VOR- UND HAUPTEXPERIMENTE	A-42
TAB. 28	ERGEBNISSE DES STIMMUNGSFRAGEBOGENS FÜR ALLE VOR- UND HAUPTEXPERIMENTE	A-43
TAB. 29	DEMOGRAPHISCHE MERKMALE DER VERSUCHSPERSONEN IM VOR- UND HAUPTEXPERIMENT	A-48

ABBILDUNGSVERZEICHNIS

ABB. 1 DIFFERENZIELL-/SOZIALPSYCHOLOGISCHE ASPEKTE [FA00,S.50]	8
ABB. 2 VOR- UND NACHTEILE LABOR UND FELD [WP12].....	13
ABB. 3 DURCHSCHAUBARKEIT DER UNTERSUCHUNG [WP12]	32
ABB. 4 ERKLÄRUNG DES VERWENDETEN VERSUCHSCODES ANHAND EINES BEISPIELS	43
ABB. 5 VE-ERGEBNISSE DER IQ-TESTS (D1, D2) IN KG & EG.....	61
ABB. 6 VE-ERGEBNISSE DES MWT (D1, D2) IN KG & EG.....	61
ABB. 7 STIMMUNG VE	62
ABB. 8 HE-ERGEBNISSE DES MWT (D1, D2) IN KG & EG.....	65
ABB. 9 HE-ERGEBNISSE DER IQ-TEST (D1, D2) IN KG & EG.....	65
ABB. 10 STIMMUNG HE	66
ABB. 11 VALIDITÄT [W112A].....	72
ABB. 12 ÜBERSICHT ZUM AUFBAU UND ABLAUF EINES EMPIRISCHEN FORSCHUNGSDESIGNS (SCHRITTFOLGE) .	A-8
ABB. 13 SKIZZE ZUM VERSUCHSAUFBAU	A-17
ABB. 14 ARBEITSPLATZ IM USABILITY-LABOR DES TECHNISCHEN VERSUCHSLEITERS (VL 1).....	A-18
ABB. 15 ARBEITSPLATZ IM USABILITY-LABOR DER VERSUCHSPERSON.....	A-18
ABB. 16 SEPARATER RAUM FÜR DAS CATERING UND DEBRIEFING-GESPRÄCH.....	A-18
ABB. 17 MINDMAP ZUR ENTWICKLUNG DES VERSUCHFRAGEBOGENS	A-20
ABB. 18 MINDMAP ZUR ENTWICKLUNG DER VERSUCHSPERSONEN-AUFGABEN (ERSTE IDEENFINDUNG).....	A-21
ABB. 19 VERSUCHSFRAGEBOGEN (VOREXPERIMENT) - PERSÖNLICHER EINDRUCK ZUM VERSUCHSABLAUF ...	A-44
ABB. 20 VERSUCHSFRAGEBOGEN (VOREXPERIMENT) – VERSUCHSAUFGABEN (VP3, VP4)	A-44
ABB. 21 VERSUCHSFRAGEBOGEN (VOREXPERIMENT) - VERHALTEN IM ALLTAG	A-44
ABB. 22 VERSUCHSFRAGEBOGEN (VOREXPERIMENT) - VERSUCHSAUFGABEN (VP1, VP2).....	A-45
ABB. 23 VERSUCHSAUFGABEN (VOREXPERIMENT) - PERSÖNLICHER EINDRUCK ZUR ARBEITSSITUATION	A-45
ABB. 24 VERSUCHSAUFGABEN (HAUPTEXPERIMENT) - VERHALTEN IM ALLTAG.....	A-46
ABB. 25 VERSUCHSAUFGABEN (HAUPTEXPERIMENT) - PERSÖNLICHER EINDRUCK ZUM VERSUCHSABLAUF....	A-46
ABB. 26 VERSUCHSAUFGABEN (HAUPTEXPERIMENT) - VERSUCHSAUFGABEN.....	A-47
ABB. 27 VERSUCHSAUFGABEN (HAUPTEXPERIMENT) - PERSÖNLICHER EINDRUCK ZUR ARBEITSSITUATION (NACH DURCHLAUF GEORDNET).....	A-47
ABB. 28 VERSUCHSAUFGABEN (HAUPTEXPERIMENT) - PERSÖNLICHER EINDRUCK ZUR ARBEITSSITUATION (NACH GRUPPE GEORDNET).....	A-47

EIDESSTAATLICHE ERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit zum Thema

**„Das Usability-Labor im wissenschaftlichen Einsatz –
Aufbau und Bewertung eines Versuchs im Usability- Labor“**

vollkommen selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe. Die Arbeit wurde bisher in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Brandenburg a. d. Havel, den 26. Oktober 2012

(Claudia Matthias)

ANHANG A. DETAILS

Dieser Anhang enthält weiterführende Informationen zu den einzelnen Schritten empirischer (sozial-wissenschaftlicher) Forschung, sowie Details zu den einzelnen Untergruppen von Versuchsplänen und Stichprobenarten.

A.1 SCHRITTE DER FORSCHUNG

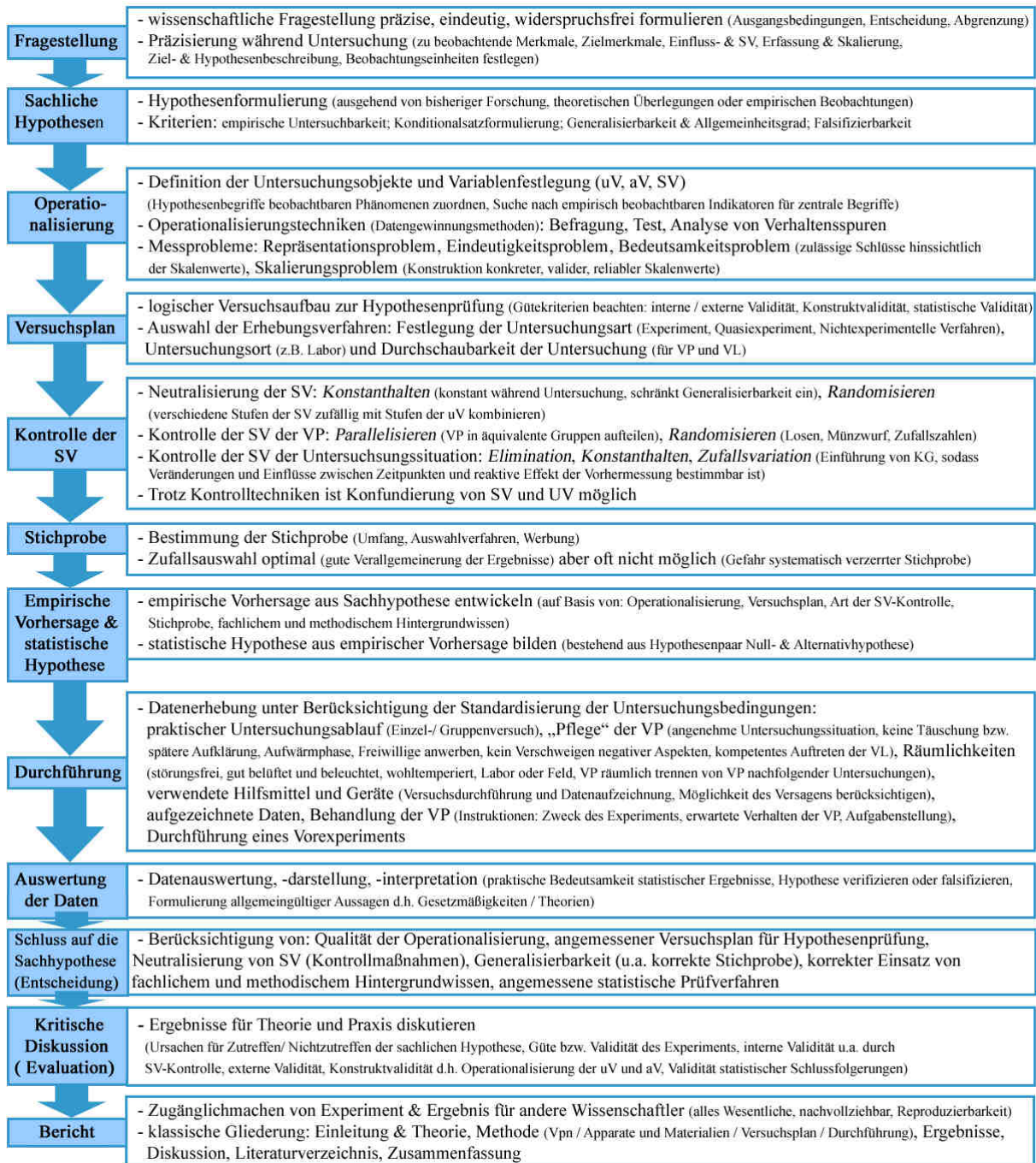


Abb. 12 Übersicht zum Aufbau und Ablauf eines empirischen Forschungsdesigns (Schrittfolge)

A.2 VERSUCHSPÄNE

Vorexperimentelle Designs:

Einmalige Untersuchung an einer Versuchsgruppe (Schrotschuss-Design, one-shot case studies): Dies ist „eine Versuchsanordnung, bei der eine einzige Versuchsgruppe nur einmal einer einzigen Behandlung X unterzogen wird, um danach den Effekt auf die abhängige Variable zu messen.“[Sa92, S.31] Ähneln eher unkontrolliertem Probieren und kann maximal als Erkundungsexperiment verwendet werden (meist in Alltagspsychologie eingesetzt, führt oft zu Pseudonachweisen). Es bietet sich an für Statusmessung oder Bestandsaufnahme, nicht aber für Ursache-Wirkungs-Zusammenhänge und besitzt keine Vergleichsmöglichkeiten oder Kontrolltechniken (Ergebnisse oft mehrdeutig, keine interne und externe Validität). Erneute Untersuchungen mit besseren (ideal: experimentellen) Designs sind zwingend, um valide Ergebnisse zu erhalten.

Vorher-Nachher-Messung an einer Versuchsgruppe (Eingruppen-Pretest-Posttest-Pläne): Eine Gruppe (repräsentative Stichprobe) wird vor und nach Treatment untersucht (Kontrolle interindividueller Unterschiede). *Vorteil*: echte Vergleichsmöglichkeit (Differenz zwischen End- und Ausgangswert). Dieser Plan besitzt jedoch erhebliche Mängel bei der internen Validität und keinerlei Kontrollstrategien, sodass von Manipulation der unabhängigen Variablen nicht auf Werte der abhängigen Variablen zu schlussfolgern ist (invalides Plan, Alternativerklärungen möglich). Grundsätzlich könnten Störvariablen die (Nicht-)Veränderung bewirken haben (u.a. Zeiteinflüsse, Reifung, Testeffekte, Reaktivität von Versuchsperson und Versuchsleiter). Der Versuchsplan sollte nicht in kausaltheoretischen Untersuchungen angewendet werden, kann aber bei einfachen statistischen Auswertungen im Rahmen von Pilotstudien genutzt werden.

Statischer Gruppenvergleich: Mind. 2 (statische) Gruppen werden nach Einführung der unabhängigen Variable einmalig untersucht. Der Plan ist invalide (geringe interne Validität), besitzt aber die Kontrollvorteile, dass interindividuell variable Ausgangsmesswerte berücksichtigt werden und Vergleichsmöglichkeiten durch mind. eine weitere statische Gruppe existieren (s. [Sa92]). Oft werden 2 Gruppen (Kontroll- und Experimentalgruppe) eingesetzt, mit 2fach gestufter unabhängiger Variable und Nachhermessung. Es werden oft bestehende (natürliche), nicht durch Zufall gewonnene Gruppen untersucht (wie Schulklassen), die sich schon vor Versuchsbeginn unterscheiden können und folglich nicht vergleichbar sind (*Auswahlverzerrung*, mindert interne Validität). Ist Zufallsgruppenbildung nicht möglich (Randomisierung), kann der Versuchsplan bei großer Versuchspersonen-Anzahl ($N > 30$) trotzdem eingesetzt werden (z.B. Schulversuche), das Design ist aber ungeeignet für kausaltheoretische Untersuchungen.

Experimentelle Designs

Randomisierungspläne: Als Between-Designs bezeichnet, da sie Mittelwertvergleiche zwischen Gruppen ermöglichen (Untersuchung interindividueller Varianzen) und eignen sich zur Untersuchung von Ursache-Wirkungs-Zusammenhängen (intern valide). Randomisierung als Kontrollstrategie ist Hauptmerkmal dieser Pläne – dazu müssen Versuchspersonen und Bedingungen dreifach randomisiert (3xR) werden: 1) Gesamtstichprobe durch einfache Zufallsziehung 2) daraus 2 oder mehr Untersuchungsgruppen durch Zufallsgruppenbildung erzeugen 3) erzeugte Gruppen per Zufall den Stufen der unabhängigen Variable zuordnen. *Vorteil*: systematische und unsystematische Kontrolle von Fehlervarianzen (vgl. Max-Kon-Min-Prinzip). Es besteht die Annahme, dass bei genügend großer Versuchspersonen-Anzahl sowohl bekannte als auch *unbekannte* Störvariablen in beiden Gruppen gleich verteilt sind (30 – 40 Versuchspersonen empfohlen). Durch die Gleichverteilung der Störvariablen gelten Gruppen, Ausgangsbedingungen und –messwerte als vergleichbar. *Nachteil*: hohe notwendige Versuchspersonen-Anzahl (nicht immer möglich, ökonomisch, umso problematischer je mehr Bedingungen und Faktoren bestehen).

Faktorielle Versuchspläne: „Bei einem faktoriellen Plan arbeitet man mit mehr als einer unabhängigen Variablen und einer abhängigen Variablen. Enthält ein faktorieller Plan zwei unabhängige Variablen, spricht man von einem zweifaktoriellen Plan; enthält er drei unabhängige Variablen, spricht man von einem dreifaktoriellen Plan usw.“[BN06, S.531] Entsprechend lassen sich faktorielle Pläne unterteilen in zwei- und dreifaktorielle Pläne (s. mehrfaktorielle Pläne).

- *Mehrfaktorielle Versuchspläne*: zwei oder mehr unabhängige Variablen werden zwei- oder mehrfach gestuft untersucht. Für alle unabhängigen Variablen (Faktoren) sind Kontrollstrategien festzulegen. Bei unterschiedlichen Kontrollstrategien für die Faktoren spricht man von Mischversuchsplänen. Bei vollständigen mehrfaktoriellen Plänen müssen alle Stufen eines jeden

Faktors (unabhängige Variable) mit den Stufen aller anderen Faktoren kombiniert werden und unter jeder Faktorstufenkombination muss eine Zufallsstichprobe des Umfangs n untersucht werden. *Beispiele:* 2×2 -faktorielles Design (einfachste Design, Untersuchung 2er jeweils 2fach-gestufteter Faktoren, benötigt 4 Gruppen), $2 \times 3 \times 2$ -faktorielles Design (benötigt 12 Gruppen. *Bezeichnung:* z.B. $3 \times 4 \times 5$ Design meint 3-faktorielles Design mit Faktor à 3 Stufen, à 4 Stufen, à 5 Stufen. Alle Faktorstufen müssen miteinander kombiniert werden (z.B. Kreuztabelle). Nach [Sa92] sind für mehrfaktorielle Randomisierungsdesigns pro Zelle ca. 5-15 Versuchspersonen notwendig. Es lassen sich neben der Wirkung der einzelnen Variablen (Haupteffekte) auch ihre kombinierten Wirkungen (Interaktionseffekte oder Wechselwirkung) ermitteln. Entsprechend sind mehrere Hypothesen zu formulieren. Bei steigender Zahl unabhängiger Variablen erhöhen sich jedoch Komplexität und Aufwand und die Interpretation der Interaktionen wird zunehmend kollektiver. Die Versuchspersonen-Anzahl nimmt exponentiell zu, die Kosten steigen (Kosten-Nutzen-Abwägung). (s. [BN06, S.531ff])

- *Einfaktorielle Pläne: Zweigruppenplan:* „Bei einem Zweigruppenplan arbeitet man mit einer zweifach gestuften unabhängigen Variablen und einer abhängigen Variablen.“[BN06, S.528] Bei kleineren Stichproben ist hier ein Testen der Vergleichbarkeit der Stichproben durch Pretests (Vortest) notwendig. Nachteilig wirkt sich aus, dass diese Kontrollmaßnahme die Versuchspersonen für das Treatment bzw. die 2. Messung (Posttestmessung) „sensibilisiert“ (instrumentelle Reaktivität/Testübung, mindert interne Validität). Zur Kontrolle der Pretesteffekte wurde der Solomon-Viergruppenplan entwickelt (s.u.). *Mehrgruppenpläne:* „Bei einem Mehrgruppenplan arbeitet man mit einer mehrfach gestuften unabhängigen Variable und einer abhängigen Variablen.“[BN06, S.530]

Solomonpläne: Diese kontrollieren durch zusätzliche Kontrollgruppe störende Effekte einer Vorhermessung, da davon ausgegangen wird, dass die Vorhermessung die Werte der Nachhermessung beeinflusst. Durch die Kontrollgruppe (zusätzliche Vergleichsmöglichkeit), wird der Einfluss dieser Störgröße eingeschätzt. Die Versuchspersonen-Zuteilung zu den Gruppen erfolgt zufällig (für „Gleichheit“ der Gruppen vor Untersuchungsbeginn). Solomonpläne existieren in verschiedenen Varianten: u.a. *Solomon-3-Gruppen-Plan* (einfachster Fall, führt ein Treatment mit 3 Gruppen durch, Einfluss der Vorhermessung wird durch Vergleich der Nachher-Werte bestimmt), *4-Gruppen-Solomon-Plan* (erlaubt explizitere Bestimmung des Vortestwerts und der Wechselwirkung durch weitere Kontrollgruppen, d.h. Gruppe 1 als „klassische“ Experimentalgruppe mit Pretest, Treatment, Posttest, Gruppe 2 als „klassische“ Kontrollgruppe mit Pretest und Posttest, Gruppe 3 als One-Shot-Case-Design mit Treatment und Posttest), Gruppe 4 mit Posttest, durch Differenzbildung sind Treatment, Vortest und Zeiteinflüsse bestimmbar und es lassen sich die „reinen“ Treatmenteffekte isolieren), *Solomon-6-Gruppen-Plan* besitzt entweder 3. Faktorstufe (vgl. Trendanalyse) oder irrelevantes Treatment (Überprüfung möglicher. Reaktivitätseffekte).

Messwiederholungsdesigns (Within-Subject Design): Sie vergleichen unterschiedliche Mittelwerte einer Gruppe (Untersuchung der intraindividuellen Varianzen, d.h. mehrfache Untersuchung derselben Stichprobe). Versuchsperson oder eine Gruppe wird zu verschiedenen Zeitpunkten unter allen Stufen der unabhängigen Variable(n) untersucht und mit sich selbst verglichen. Die einzelnen Messwerte pro Versuchsperson werden zu Mittelwerten zusammengefasst. Dieses Design erlaubt die Kontrolle der Sekundärvarianz und macht sich Unterschiede innerhalb einer Gruppe zunutze, d.h. personelle Störvariablen werden durch Vergleich der Messwerte einer Versuchsperson oder Gruppe konstant gehalten. Situative Störvariablen (Licht, Raumtemperatur etc.) sind zusätzlich zu kontrollieren. Einfachste Variante ist der Zweistichprobenversuchsplan mit Messwiederholung (1 Gruppe zu 2 Messzeitpunkten). Der Mehrstichprobenplan mit Messwiederholungen ist ein mehrfaktorielles Design, mit entweder mehreren Stufen der unabhängigen Variable d.h. Trendanalyse oder mehreren unabhängigen Variablen. Es wird in beiden Fällen nur eine Stichprobe untersucht, diese jedoch zwei- oder mehrfach. Vorteil: geringe benötigte Versuchspersonen-Anzahl, geringe Datenstreuung, klarerer Nachweis der experimentellen Effekte durch Wegfall der Between-Varianz (interindividuell). Nachteil: Zeitaufwand durch wiederholte Untersuchungen, Gefahr von Übertragungseffekten (vorausgehende Untersuchungsbedingungen können nachfolgende beeinflussen d.h. „carry over“-Effekt). Zur Vorbeugung Letzterer können die Zeitabstände zwischen experimentellen Bedingungen größer gewählt und die Reihenfolge der experimentellen Bedingungen ausbalanciert werden.

Blockversuchspläne: Versuchspersonen werden anhand eines blockbildenden Merkmals in homogene Blöcke eingeteilt (Kompromiss aus Randomisierungs- und Wiederholungsdesigns). Es werden verschiedene experimentelle Gruppen untersucht und die Between-Varianz kontrolliert. Die Blockbildung soll Ausgangswert-Unterschiede bzgl. des blockbildenden Merkmals minimieren, um die bekannte Störvariable zu kontrollieren. Zusätzliche (bekannte/unbekannte) Störvariablen sind

durch Kontrollstrategien zu kontrollieren. Es existieren *2-Stichprobenpläne mit Blockbildung* (Sonderfall, als Parallelisierungsdesign bezeichnet, da Blöcke aus genau 2 Versuchspersonen bestehen und *Paare* heißen) und *Mehrstichprobenpläne mit Blockbildung* (mehr als 2 Stufen der unabhängigen Variable). Blockversuchspläne sollen die geringere Präzision des Randomisierungsdesigns durch Kontrolle der Between-Varianz vermeiden (geringere Datenstreuung) und streben geringere Anfälligkeit für Reihenfolge- und Übertragungseffekte gegenüber Wiederholungsmessungen an, indem jede Versuchsperson nur unter einer Faktorstufe untersucht wird. Sie bieten sich an, wenn aufgrund kleiner Stichproben die Zufallsgruppenbildung unterschiedliche (inhomogene) Gruppen hervorbringen würde. Stattdessen werden Blöcke (statistisch) homogenisiert, um Fehlervarianzen zu reduzieren (parallelisierte Gruppen, Konstanthaltung bestimmter Ausgangsbedingungen, Kontrolle von Vortestvariablen die mit abhängigen Variablen korrelieren). *Nachteil*: Schwierigkeit, ein geeignetes blockbildendes Merkmal zu finden. Zudem sind Blöcke u. U. nicht ideal homogen bzgl. anderer Merkmale und der Aufwand erhöht sich durch Voruntersuchungen.

Mischversuchspläne: Dies sind zwei- oder mehrfaktorielle Pläne (mind. 2 unabhängige Variablen), wobei einzelne unabhängige Variablen unterschiedlichen Design-Haupttypen entsprechen. Zu den Planungsprinzipien / Haupttypen zählen: Randomisierung, Blockbildung, Wiederholungsmessung, Organismusvariable (korrelativer Faktor) oder quasiexperimentelle Faktoren sind zusätzlich möglich, mindern jedoch Qualität und interne Validität. *Vorteil*: gute Sekundär- und Fehlervarianzkontrolle, leichtere Verbindung experimenteller und korrelativer Ansätze. Es sollte bei der Anwendung beachtet werden, dass hier Vorteile (z.B. kleinere Stichproben) und Nachteile (z.B. Übertragungseffekte) einzelner Design-Haupttypen vereint werden.

Unvollständige Versuchspläne: Dies sind (unvollständige) mehrfaktorielle Designs, in denen nicht alle Stufen jedes Faktors miteinander kombiniert werden. Sie sind ökonomischer als mehrfaktorielle Pläne (weniger Versuchspersonen), überprüfen jedoch nicht immer alle Haupt- und Interaktionseffekt (Interpretationseinschränkungen, vor allem für Interaktionseffekte). Strukturbezogen lassen sich zwei Arten unterscheiden: *hierarchische Pläne* (hierarchische Verschachtelung der Faktoren, prüfen nur Haupteffekte und sollten nur angewandt werden, wenn Interaktionen unwahrscheinlich sind) und *quadratische Pläne* (drei oder vier Faktoren, die auf vollständigen zweifaktoriellen Designs basieren d.h. Darstellung als Quadrat, sie sind eine Sonderform der unvollständigen Pläne bei der alle Faktoren die gleiche Stufenanzahl aufweisen) s. [BN06, S.540ff]).

Multivariate Versuchspläne (*multi = viel; variat = Variablen*): Eine oder mehrere unabhängige Variablen beeinflussen mind. 2 abhängige Variablen. Es wird eine globale abhängige Variable gewählt (Konstrukt) und anhand verschiedener (sachrepräsentativer, aussagekräftiger) Einzelmaße operationalisiert, d.h. ein Konstrukt wird durch verschiedene abhängige Variablen facettenreicher erfasst (z.B. Konstrukt „Angst“ über Gesichtsausdruck, Puls etc.). Wechselseitige Beziehungen der abhängigen Variablen untereinander werden berücksichtigt und aufgedeckt. Dadurch sind Ergebnisse besser verallgemeinerbar und externe Validität steigt. Durch die verschiedenen Messungen und umfangreichere statistische Analyse erhöht sich aber der Aufwand und Ergebnisse sind nicht in jedem Fall besser oder aufschlussreicher als bei univariaten Designs. Angewandt sollte der Plan nur werden, wenn die abhängige Variable komplex ist und sich nur durch mehrere operationale Indikatoren sinnvoll erfassen lässt. (s. [BN06, S.545ff])

Quasiexperimentelle Designs

Designs für Entwicklungen: zur Überprüfung von entwicklungsbedingter Veränderungshypothesen

Querschnittuntersuchung: Die Stichprobe wird bzgl. bestimmter Merkmale in unterschiedlichen Altersgruppen (oder variable Generationen) zum *gleichen* Zeitpunkt untersucht. Dies dient der Veränderungsmessung (Prognosen über Entwicklungsverläufe durch direkten Altersgruppenvergleich bzgl. bestimmter Merkmale) und als Ausweg, da Alter, Epoche und Generation als Einflussgrößen bei Fragestellungen stets konfundiert und nicht vollständig voneinander isolierbar sind. Sie sind jedoch anfällig für Konfundierung von Alters- und Generationseffekten (z.B. Kohorteneffekte d.h. Messwertbeeinflussung aufgrund verschiedener Generationen), weswegen Rückführungen der Ergebnisse auf das Alter nur bei vernachlässigbaren Generationsunterschieden erlaubt sind und Messunterschiede nur bei vernachlässigbaren Altersunterschieden auf Generationsunterschiede zurückführbar sind. Weiterhin kann mit zunehmendem Alter eine systematische Veränderung der Stichprobe bzgl. bestimmter Merkmale stattfinden und Ergebnisse beeinflussen (sinkende interne Validität durch selektive Populations-

veränderung). Weiterhin hängt auch die Validität der Messinstrumente vom Versuchspersonen-Alter ab, sodass Messinstrumente nicht unbedingt vergleichbar sind (z.B. Testaufgaben, die von Jüngeren durch kreative Denkleistung und von Älteren durch Erfahrung erfüllt werden). Die zeitliche Entwicklung ist nicht ablesbar, im Mittelpunkt steht die aktuelle Struktur. Als Merkmal sind hier die heterogenen Skalen und Messinstrumente, sowie die geringe Wahrscheinlichkeit von Alternativerklärungen zu nennen. Es existiert ein klarer theoretischer Bezug und ein Vergleich zwischen verschiedenen Personen z.B. Versuchs-Kontrollgruppe ist möglich (Between Subject Design). Nachteilig wirkt sich z.B. die hohe Gefahr von Störvariablen zwischen den Messungen aus.

Längsschnittuntersuchung (*Mehrzeitpunktuntersuchungen, Longitudinalstudie*): Dies sind Untersuchungen zu mehreren Messzeitpunkten um zeitliche Entwicklung abzubilden, mit gleichen Versuchspersonen oder unterschiedlichen. Als vorteilhaft wirkt sich die geringe Gefahr für Störvariablen zwischen den Messungen aus. Diese Studien kennzeichnen sich u.a. durch homogene Skalen und Messinstrumente, klare Erhebungszeitpunkte, hohe Wahrscheinlichkeit von Alternativerklärungen, einem nicht so klar entwickelten theoretischen Bezug sowie der Möglichkeit, Vergleiche derselben Versuchsperson anzustellen (z.B. Vorher-Nachher-Messung, Within Subject Design). Längsschnittstudien sind Querschnittstudien überlegen, da alles was in Letzteren erhoben wird auch in Ersteren untersucht werden kann aber nicht umgekehrt. Zudem gelten die Ergebnisse als zuverlässiger (reduzierte Varianz). Als Nachteil der Längsschnittstudie sind Kosten und höherer organisatorischer und zeitlicher Aufwand (sehr große Stichproben kaum möglich) zu nennen sowie die zunehmende Schwierigkeit, alle Versuchspersonen des ersten Erhebungszeitpunktes für weitere zu gewinnen. Insgesamt sind in Längsschnittstudien Alters- und Epochenefekte konfundiert, mit oft selektiven Ausgangsstichproben (Freiwillige als Versuchspersonen, wenig repräsentative Stichprobe) und selektiven Stichprobenveränderungen (einseitige systematische Stichprobenausfälle, da nicht alle Versuchspersonen bereit sind, mehrere Messungen über längere Zeit durchzuführen¹¹¹). Auch Vergleichbarkeit der Messinstrumente (Validität der Messinstrumente ggf. altersabhängig), Testübung und Testeffekte sowie der Umstand, dass Ergebnisse nur für die untersuchte Generation gelten (generationsspezifische Aussagen) und der erhöhte Untersuchungsaufwand (Zeitaufwand), gehören zu den Schwächen der Längsschnittuntersuchungen. (s. [el09] [BN06])

Ex-Post-Facto-Designs und Korrelative Designs

Korrelationsstudie: Relevante Variablen werden nicht in abhängige und unabhängige Variablen unterteilt. Stattdessen sollen komplexe multivariate Variablenzusammenhänge aufgedeckt und strukturiert werden. Kausalaussagen auf Basis korrelativer Zusammenhänge sind aber unzulässig, da Kausalität zwar hinreichende aber nicht notwendige Bedingung für Korrelationen ist. Im Vordergrund steht hier die Prüfung von Zusammenhangshypothesen zwischen zwei oder mehreren Variablen ohne Manipulation experimenteller Variablen oder kausaltheoretischen Anspruch.

Ex-post-facto-Studien: Daten werden nur zu einem Zeitpunkt erhoben, ohne Manipulation der unabhängigen Variablen (z.B. Befragung), sodass Sequenzkenntnisse von unabhängigen und abhängigen Variablen fehlen und nur korrelative Aussagen möglich sind (beruht auf korrelativem Versuchsplan), d.h. Ursache-Wirkung ist nicht festzustellen. Es wird aus vorgefundenen Fakten im Nachhinein (*ex post facto*) auf Verursachungsbedingungen geschlossen. Es liegt jedoch relatives Wissen zur zeitlichen Abfolge der korrelierenden Variablen vor (in sog. Pfadanalysen), mit denen Kausalitätsannahmen fundiert werden. Der kausaltheoretische Anspruch kann jedoch nicht begründet werden.

¹¹¹ Positive Selektion (VP mit negativen Merkmalen fallen heraus) und negative Selektion (VP mit positiven Merkmalen fallen heraus) führen beide zu Ergebnisverzerrungen.

A.3 STICHPROBEN

Tab. 17 Verschiedene Arten der Stichprobeauswahl (Übersicht vgl. [eI09] [BN06])

Zufällig vs. nicht zufällig	Arten	Vorgehen (Schrittfolge)
Probabilistische Stichprobe (zufällig)	Einfache Zufallsstichprobe	1) Zielpopulation bestimmen 2) Vollständige Liste aller Populationsobjekte erstellen 3) Zufallszahlen/Lottoprinzip/systematische Auswahl 4) einfache Zufallsstichprobe
Sonderformen	Geschichtete Stichprobe	Reduktion / Ausschaltung der Zufallsstreuung für bestimmte Merkmale: 1) Zielpopulation in Schichten einteilen 2) pro Schicht vollständige Liste aller Objekte erstellen 3) Zufallszahlen/Lottoprinzip/systematische Auswahl 4) geschichtete Stichprobe Anwendung: Merkmale, die mit Variablen zusammenhängen, sind bekannt
	Klumpenstichprobe	1) Zielpopulation (aus Klumpen) festlegen 2) vollständige Liste aller Klumpen 3) Zufallszahlen/Lottoprinzip/systematische Auswahl 4) Klumpenstichprobe Anwendung: Population setzt sich aus homogenen Teilgesamtheiten zusammen, von denen jede zufällig ausgewählte Teilgesamtheit vollständig untersucht wird
	Mehrstufige Stichprobe	1) Zielpopulation (aus Klumpen) festlegen 2) vollständige Liste aller Klumpen 3) Zufallszahlen/Lottoprinzip/systematische Auswahl 4) 1. Ziehungsstufe (Zufallsauswahl von Klumpen) 5) pro Klumpen vollständige Liste aller Objekte erstellen 6) Zufallszahlen/Lottoprinzip/systematische Auswahl 7) 2. Ziehungsstufe (Zufallsauswahl von Objekten pro Klumpen) 8) zweistufige Stichprobe Anwendung: Population setzt sich aus großen Teilgesamtheiten zusammen und jede zufällig ausgewählte Teilgesamtheit wird stichprobenartig untersucht
	Proportionale Schichtung	1) Schichtung nach Merkmal, dessen Parameter gesucht werden (oder Ersatzmerkmal) 2) Größe der Grundgesamtheit und Schichten muss bekannt sein 3) Zufallsstichprobe aus jeder Schicht ziehen 4) resultierende Gesamtstichprobe muss bzgl. des Schichtenmerkmals völlig repräsentativ für Grundgesamtheit sein
	Disproportionale Schichtung	Bei starken Größenunterschieden der Schichten durch das Schichtungsmerkmal, werden Teilstichproben gleich groß gemacht
	Mehrdimensionale Schichtung	Schichtung nach mehreren Merkmalen, wenn interne Streuung einzelner Schichten (im Vergleich zur Streuung zwischen Schichten) deutlich abnimmt und folglich das Vertrauensintervall der Gesamtstichprobe einengt
Nicht-probabilistische Stichproben (nicht zufällig)	Gelegenheitsstichprobe (Ad-hoc-Stichprobe)	Unrepräsentative Gruppen: 1) freiwillige Personen 2) Aufteilung der Personen per Zufall in Experimental- und Kontrollgruppe 3) Gelegenheitsstichprobe
	Quotenstichprobe	1) selbst gewählte Zielpopulation (keine Zufallsauswahl) 2) Quotenbildung anhand untersuchungsrelevanter Merkmale (systematische Repräsentation wesentlicher Merkmale) 3) Stichprobenentnahme unter Beachtung definierter Quoten 4) Quotenstichprobe 5) ggf. Überprüfung der Quotenstichprobe (Überprüfungsmöglichkeit anhand nicht quotierter aber in Verteilung bekannter Merkmale)
	Theoretische Stichprobe	

Nicht-probabilistische Stichproben

Quotenstichprobe: Sie setzt kein Zufallsprinzip voraus und kennzeichnet sich durch Quotenbildung bzgl. untersuchungsrelevanter Merkmale innerhalb der Grundgesamtheit (z.B. Geschlecht, Alter), wobei die Merkmale Einfluss auf die abhängige Variable haben sollten. „Bei der Quotenstichprobe wird versucht, die Zusammensetzung der Stichprobe hinsichtlich ausgewählter Merkmale den Populationsverhältnissen durch bewußte Auswahl ‚passender‘ Objekte anzugleichen, also quasi ‚Quoten‘ für bestimmte Merkmale zu erfüllen“([BN03, S.405]). Negativ wirkt sich hier vor allem die mangelnde Repräsentativität aus. Ein Vergleich der Verteilung nicht quotierter Merkmale in der Stichprobe mit Verteilung dieser Merkmale in der amtlichen Statistik erlaubt die Prüfung der Stichprobeneildung. Fehlerquellen sind „Interviewer-Bias, Auffüllen bei Nicht-Antreffen oder Verweigerung, u. U. zu indirekte Beziehung zwischen Quoten- und Untersuchungsmerkmalen“[Fa00, S.29].

Gelegenheitsstichprobe (Ad-hoc-Stichprobe): Dies ist eine Untersuchung von Personen/Objekten, die gerade zur Verfügung stehen oder leicht zugänglich und Freiwillige sind (meist Einteilung, z.T. nach Zufallsprinzip, in Experimental- und Kontrollgruppe). Wenn repräsentative Stichproben nicht benötigt werden, d.h. Verallgemeinerungen nicht das Anliegen sind, kann dieses Verfahren zur Erfassung der Veränderung der abhängigen Variablen nach Manipulation der unabhängigen Variablen genutzt werden (Auswahlprinzip ist unbekannt). *Vorteile:* praktikable, ökonomische Rekrutierung. *Nachteile:* Einschränkungen bei Generalisierung der Ergebnisse auf die Grundgesamtheit. Sollen Ergebnisse dennoch verallgemeinert werden, sind Untersuchungen an einer weiteren Stichprobe vorzunehmen (d.h. Replikationsstudie). Stimmen Ergebnisse beider Studien überein, lässt sich Gültigkeit über die ursprüngliche Personengruppe hinaus annehmen. Bei Misserfolg der Replikationsstudie muss Generalisierbarkeit der Befunde angezweifelt werden. Insgesamt entsprechen Freiwillige nicht dem Durchschnitt der Population und unterscheiden sich ggf. von Verweigerern.

Theoretische Stichprobe: Nach theoretischen Überlegungen werden typische oder untypische Fälle bewusst ausgewählt. Sie findet gerade in qualitativer Forschung Anwendung. (s. [BN06, S.402ff])

Probabilistische Stichproben

Einfache Zufallsstichprobe (random sample, simple random sample): Welche Elemente der Population in die einfache Zufallsstichprobe einbezogen werden, ist zufallsabhängig (gleiche Auswahlwahrscheinlichkeit). Alle Untersuchungsobjekte der zu untersuchenden Population müssen identifizierbar und in einer Liste *vollständig* aufgezählt sein, damit aus dieser Liste die Stichprobe nach *Zufallsprinzip* ausgewählt werden kann. Auswahlstrategien unterteilen sich in: *Auswahl nach Lottoprinzip* (für jedes Element ein Los, aus durchgemischten Losen erfolgt Stichprobenbildung per Ziehung), *Auswahl mittels Zufallszahlen* (Elemente werden durchnummeriert, Stichprobenziehung mittels Tafeln mit Zufallszahlen) und *systematische Auswahl* (jedes K-te Element aus zufällig geordneter Reihenfolge der Objekte). Zufallsstichproben sind Voraussetzung für die Generalisierung auf die Population (Repräsentationsschluss). Zufallsstichproben gewähren *globale* Repräsentativität statt (merkmals-)spezifischer, d.h. Zusammensetzung entspricht in nahezu *allen* Merkmalen der Population statt in nur relevanten Merkmalen. Einfache Zufallsstichproben erweisen sich praktisch jedoch als problematisch, da oft nicht *jedes* Untersuchungsobjekt der Population erfassbar und *zufällig* auswählbar ist bzw. der Aufwand unzumutbar wäre. (s. [BN03, S.400ff] [Fa00])

Geschichtete Stichprobe (stratifizierte Stichprobe): Sie wird mit geschichteter Wahrscheinlichkeitsauswahl gebildet, wobei sämtliche Elemente der Population in mehrere Schichten aufgeteilt werden. Schichtenbildung erfolgt anhand eines gemeinsamen, für die Fragestellung relevanten und theoretisch begründeten Merkmals (u.a. Beruf, zumeist biografische und soziodemographische Merkmale), das mit der zu untersuchenden Variable im Zusammenhang steht. Stichproben werden so zusammengesetzt, dass „prozentuale Verteilung von Schichtungsmerkmalen (Alter, Geschlecht, Beruf etc.) in der Stichprobe der prozentualen Verteilung dieser Merkmale in der Population entspricht“[BN03, S.55] Aus den Schichten werden Teilstichproben nach Zufallsprinzip entnommen (einzelne Schichten sind in sich homogen aber untereinander verschieden). Nachteil ist der höhere organisatorische und rechnerische Aufwand. Andererseits wird von erhöhter Schätzgenauigkeit der Gesamtstichprobe durch geschichtete Wahrscheinlichkeitsauswahl ausgegangen (Teilstichproben streuen bzgl. der untersuchten Variable weniger als in Gesamtstichprobe d.h. homogene Teilstichproben). *Kleinere* (sinnvoll) geschichtete Stichproben können *gleiche* Schätzgenauigkeit erreichen wie Zufallsstichproben (mit weniger Versuchspersonen). Zusammenfassend soll für bestimmte Merkmale die Zufallsstreuung reduziert oder ausgeschaltet werden. ([BN06, S.425ff])

Klumpenstichprobe („Cluster Sample“): Hier geschieht die zufällige Auswahl einer einzelnen Versuchsperson aus der Population, wobei die Population aus vielen Teilpopulationen/Gruppen bestehen muss (z.B. Schulklassen) und aus diesen Gruppen die zufällige Auswahl getroffen wird¹¹². Klumpenstichproben sind Stichproben, „in denen mehrere zufällig ausgewählte ‚Klumpen‘ (z.B. Krankenhäuser, Wohnblocks, Schulklassen o. ä.) vollständig erhoben werden“. [BN03, S.55] und bieten sich bei einer nicht vollständig zugänglichen Grundgesamtheit an. Vorteil ist der geringe zeitliche und organisatorische Aufwand gegenüber einfachen Zufallsstichproben, da eine Liste sämtlicher Elemente der zu untersuchenden Population nicht nötig ist und Liste sämtlicher, für die Untersuchung relevanter Klumpen in der Population genügt. Sie bereiten den geringsten untersuchungstechnischen Aufwand. Nachteil ist die große Ähnlichkeit der Elemente eines Klumpens, was zu einem vergrößerten Stichprobenfehler führen kann. Jeder Klumpen sollte die Population annähernd gleich gut repräsentieren (Gleichheit der Klumpen aber Verschiedenartigkeit der Untersuchungsobjekte in den Klumpen). (s. [BN06, S.435ff, 481])

Mehrstufige Stichprobe („Multistage Sampling“): Statt Klumpen vollständig zu untersuchen, wird eine 2- oder mehrstufige Stichprobe gezogen durch. Hierbei erfolgt schrittweise die Auswahl aus einer schwer erfassbaren, nicht vollständig zugänglichen Grundgesamtheit, wenn es in der Praxis schwierig ist, die Stichprobe direkt aus der Grundgesamtheit zu ziehen (z.B. Fehlen von Listen über sämtliche Elemente der Grundgesamtheit, Verstreuung der Grundgesamtheit über großes Territorium oder zu große „Klumpen“). Die schrittweise Auswahl untergliedert sich in Ziehungsstufen (Auswahl nach mehreren Schichtungs- und Klumpenmerkmalen):

“Man zieht eine mehrstufige Stichprobe, indem man zunächst zufällig eine Klumpenstichprobe mit großen Klumpen zieht (1.Ziehungsstufe). Diese Klumpen werden nicht vollständig untersucht, sondern aus ihnen wird eine Zufallsstichprobe der Untersuchungsobjekte gezogen (2.Ziehungsstufe). Zieht man auf der zweiten Stufe wieder eine Klumpenstichprobe, ergibt sich durch Ziehung einer Zufallsstichprobe aus diesen Klumpen eine 3. Ziehungsstufe usw.“[BN06, S.441]

Diese Stichprobenart ist organisatorisch nicht so aufwendig wie die einfache Zufallsstichprobe und erfordert keine vollständige Liste aller Untersuchungsobjekte der Population sondern nur aller Klumpen und Untersuchungsobjekte in den ausgewählten Klumpen.

¹¹² “Man zieht eine Klumpenstichprobe, indem man aus einer in natürlichen Gruppen (Klumpen) gegliederten Population nach dem Zufallsprinzip eine Anzahl von Klumpen auswählt und diese Klumpen dann vollständig untersucht.“[BN06, S.436]

ANHANG B. PLANUNG

Dieser Anhang zeigt Einzelheiten zur Gesamtplanung des Versuchs dieser Arbeit (Organisation), sowie vorausgehende Überlegungen und Vorarbeiten hinsichtlich einiger relevanter Planungsschritte (Versuchsaufbau, Zeitplan für Vor- und Hauptexperiment mit Versuchspersonen, MindMaps zur Entwicklung des Versuchsfragebogens und der Versuchspersonen-Aufgaben).

B.1 GESAMTPLANUNG

Organisation vorab:

- Entwicklung und Beschaffung der Materialien (IQ-Aufgaben, MWT-A/B, Materialien s. ► 3.10.1)
- Anwerbung und Auswahl der Versuchspersonen, Akquirierung und zeitliches Eintakten für Vor- und Hauptexperimente
- Zeitplan für eigentliche Durchführung der Untersuchung (inklusive Pufferzeiten für evtl. Pannen)

Raum-Organisation:

Tab. 18 Übersicht über die Raum-Organisation für den Versuch im Rahmen dieser Arbeit

Zweck	Raum
Versuchsdurchführung mit Versuchspersonen	Usability-Labor (und angeschlossenes Biosignalverarbeitungslabor), <i>Zeitplan für Usability-Labor</i>
Separater Raum für die getrennte Beobachtung im Versuch	<i>R014 Informatik-Gebäude (Büro Tina Meißner)</i>
Separater Raum für das Catering in den Hauptexperimenten	<i>R011 Informatik-Gebäude</i>

Alphabetisches Personenregister:

Tina Meißner – Versuchsleiter 1

Claudia Matthias – Versuchsleiter 2

Finanzen:

Tab. 19 Finanzplan für die Vor- und Hauptexperimente

Versuch	Anschaffung	Kosten
Vorexperiment	Catering für die Versuchspersonen (<i>hier</i> : durch die Versuchsleiter organisiert und finanziert): Süßigkeiten	7,18 €
Hauptexperiment	Catering für die Versuchspersonen (<i>hier</i> : durch die Versuchsleiter organisiert und finanziert): Obst (Erdbeeren, Trauben), Brötchen, Aufschnitt (Käse, Wurst), Kekse & Süßes, Servietten, Getränke	36,23 €
Vor- und Hauptexperiment	MWT-A und MWT-B	31,55 €
Gesamt		74,96

B.2 VERSUCHSAUFBAU

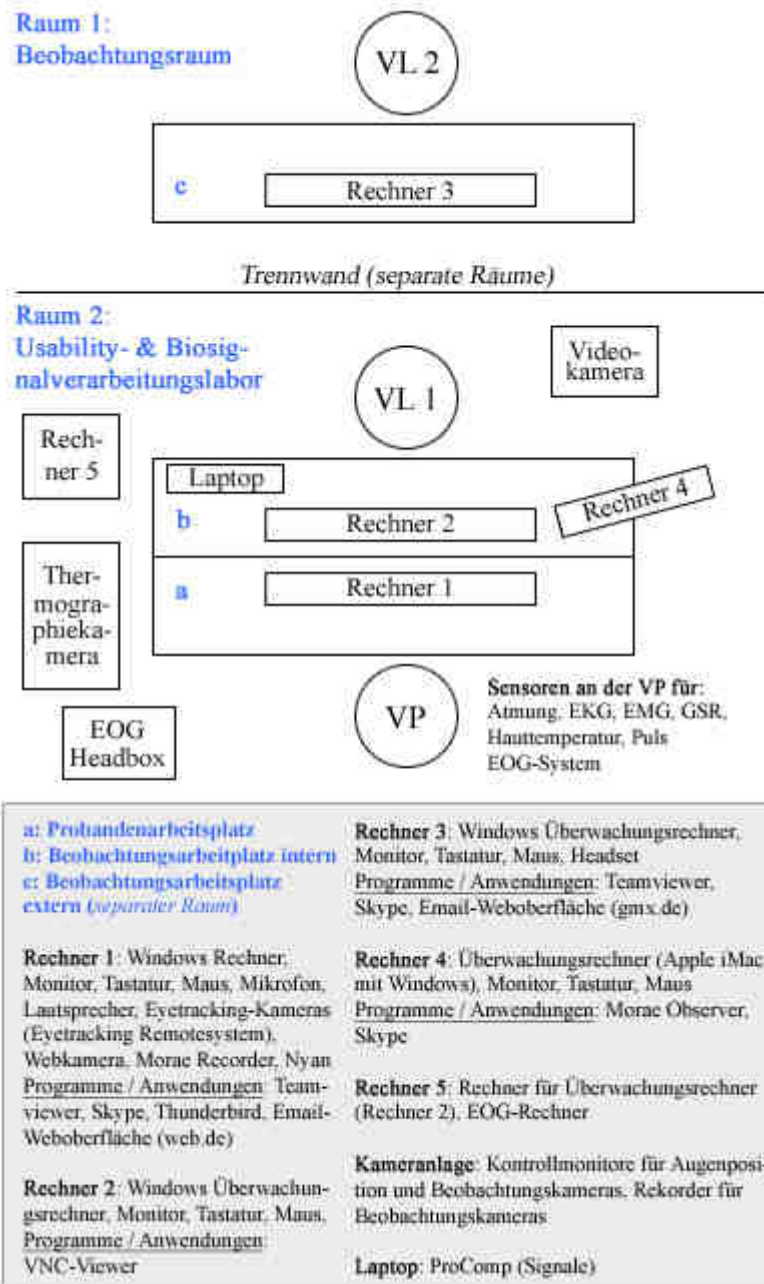


Abb. 13 Skizze zum Versuchsaufbau



Abb. 14 Arbeitsplatz im Usability-Labor des technischen Versuchsleiters (VL 1)



Abb. 15 Arbeitsplatz im Usability-Labor der Versuchsperson



Abb. 16 Separater Raum für das Catering und Debriefing-Gespräch

B.3 ZEITPLAN

Ort: Usability-Labor (Versuchsperson, Versuchsleiter 1), Beobachtungsraum (Versuchsleiter 2)

Dauer: ca. 1,5 Stunden pro Versuchsperson in Vor- und Hauptexperiment

Tab. 20 Zeitplan für Vorexperimente

Datum	Versuchsperson	Versuchspersonen-Code (Ablauf)	Zeit
18.06.2012	Tina Meißner, Claudia Matthias	Testen der Planung und Geräte durch Versuchsleiter	14:00Uhr
24.06.2012	Tina Meißner, Claudia Matthias	Testen der Planung und Geräte durch Versuchsleiter	17:00Uhr
25.06.2012	VP01 (Mitarbeiter, männlich)	V-2012-06-25-EG-VP01-A1-MWT-A/IQ-1 V-2012-06-25-KG-VP01-A1-MWT-B/IQ-2	16:00 Uhr (1,5 Stunden)
	VP02 (Student, männlich)	V-2012-06-25-KG-VP02-A2-MWT-B/IQ-1 V-2012-06-25-EG-VP02-A2-MWT-A/IQ-2	17:30Uhr (1,5 Stunden)
	VP03 (Student, männlich)	V-2012-07-12-EG-VP03-A3-MWT-B/IQ-1 V-2012-07-12-KG-VP03-A3-MWT-A/IQ-2	14:30Uhr (1,5 Stunden)
11.07.2012	VP04 (Studentin, weiblich)	V-2012-06-29-KG-VP04-A1-MWT-A/IQ-1 V-2012-06-29-EG-VP04-A1-MWT-B/IQ-2	16:30Uhr (1,25 Stunden)

Tab. 21 Zeitplan für Hauptexperimente

Datum	Versuchsperson	Versuchspersonen-Code (Ablauf)	Zeit
05.08.2012 (Sonntag)	Testen der Planung und Geräte, Versuchsvorbereitung	Tina Meißner, Claudia Matthias (Versuchsleiter 1 und 2)	16:30Uhr
06.08.2012 (Montag) VERSUCHSAUSFALL: kurzfristige Wartungsarbeiten an der Fachhochschule	TERMINVERLEGUNG VP 09 (neu nummeriert) VP 04 (neu nummeriert) VP 07 (neu nummeriert)		14:30Uhr 16:30Uhr 18:30Uhr
	VERSUCHSPERSON-AUSFALL VP01	H-2012-08-07-EG-VP01-A1-MWT-A/IQ-1 H-2012-08-07-KG-VP01-A1-MWT-B/IQ-2	13:00Uhr
	VERSUCHSPERSON-AUSFALL VP02	H-2012-08-07-KG-VP02-A2-MWT-A/IQ-1 H-2012-08-07-EG-VP02-A2-MWT-B/IQ-2	15:00Uhr
07.08.2012 (Dienstag) VERSUCHSAUSFALL: kurzfristige Wartungsarbeiten an der Fachhochschule	TERMINVERLEGUNG VP08 (neu nummeriert)	H-2012-08-07-KG-VP02-A2-MWT-A/IQ-1 H-2012-08-07-EG-VP02-A2-MWT-B/IQ-2	17:00Uhr
	VP03	H-2012-08-08-EG-VP03-A3-MWT-A/IQ-1 H-2012-08-08-KG-VP03-A3-MWT-B/IQ-2	14:30Uhr (1,25 Stunden)
	VP04	H-2012-08-08-KG-VP04-A1-MWT-A/IQ-1 H-2012-08-08-EG-VP04-A1-MWT-B/IQ-2	16:30Uhr (1,5 Stunden)
08.08.2012 (Mittwoch)	VERSUCHSPERSON-AUSFALL VP05	H-2012-08-08-EG-VP05-A2-MWT-A/IQ-1 H-2012-08-08-KG-VP05-A2-MWT-B/IQ-2	18:30Uhr
	VP06	H-2012-08-09-KG-VP06-A3-MWT-A/IQ-1 H-2012-08-09-EG-VP06-A3-MWT-B/IQ-2	14:30Uhr (1,25 Stunden)
	VP07	H-2012-08-09-EG-VP07-A1-MWT-A/IQ-1 H-2012-08-09-KG-VP07-A1-MWT-B/IQ-2	16:30Uhr (1,5 Stunden)
09.08.2012 (Donnerstag)	VP08	H-2012-08-09-KG-VP08-A2-MWT-A/IQ-1 H-2012-08-09-EG-VP08-A2-MWT-B/IQ-2	18:30Uhr (1,5 Stunden)
	VP09	H-2012-08-10-EG-VP09-A2-MWT-A/IQ-1 H-2012-08-10-KG-VP09-A2-MWT-B/IQ-2	10:00Uhr (1 Stunde)

B.4 MINDMAP ZUM VERSUCHSFRAGEBOGEN

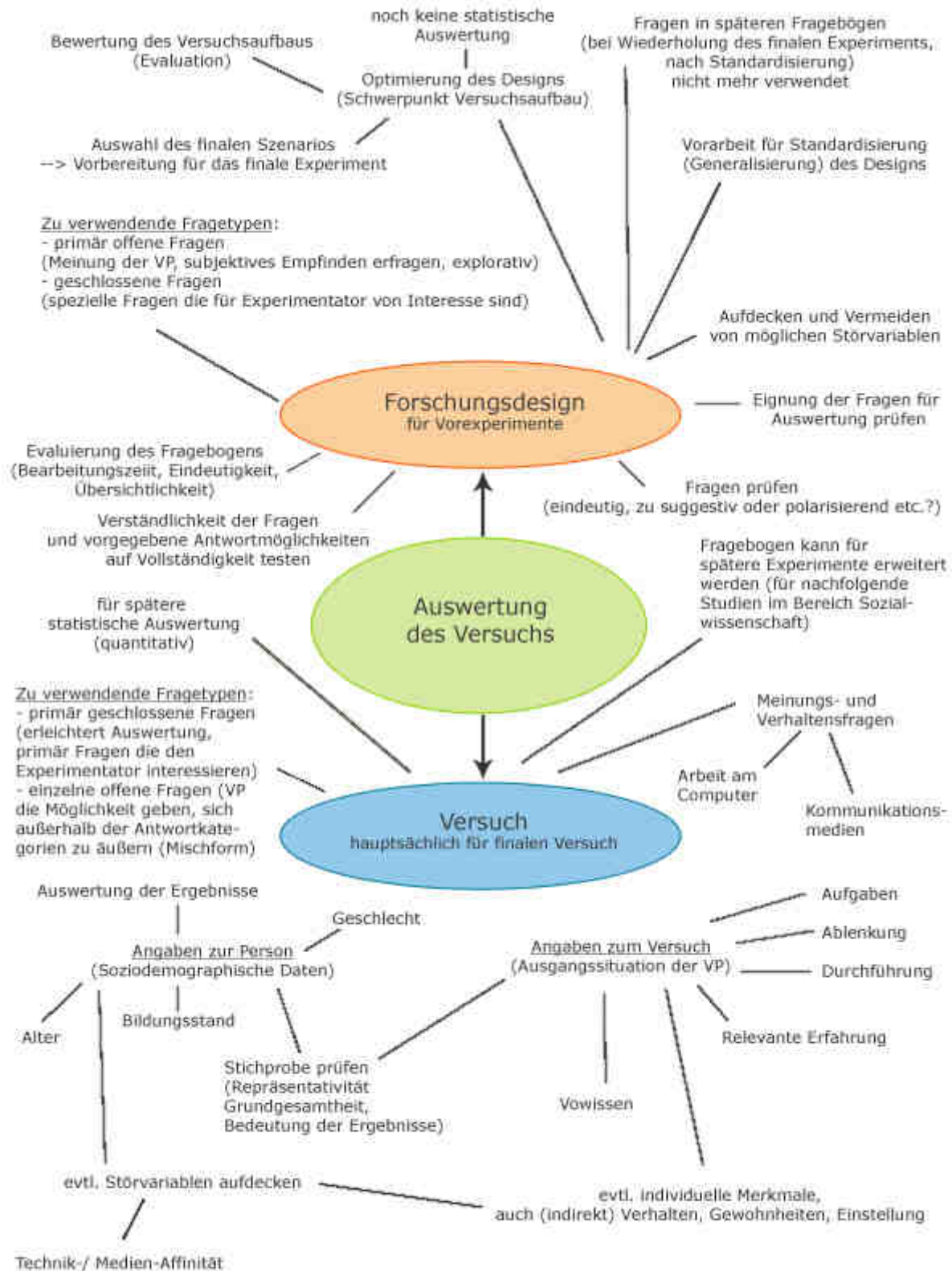


Abb. 17 MindMap zur Entwicklung des Versuchsfragebogens

B.5 MINDMAP ZU DEN VERSUCHSPERSONEN-AUFGABEN

Puzzle

Vorteile: erfordert Konzentration

- schnelles Weiterarbeiten nach uV ohne Gedanken nochmal vollständig zu durchlaufen (visuelle Unterstützung, keine intensive Problemlösung)
 - Parallelaufgaben für zwei Durchläufe mit wenig Aufwand möglich
 - VL-Effekte fast gänzlich auszuschließen
 - durch Einfachheit kein überdurchschnittlicher Leistungsdruck, Stress, Angst
 - Bearbeitung am Computer (Online-Puzzle), u.a. Eye-Tracking einsetzbar
- Nachteile: schwere Auswertung ("Fehler" schwer auffindbar)
- schwierige Standardisierung
 - Aussagekraft fraglich (Zählen korrekt zusammengefügtter Teile?)
 - eher Erhebung der Bearbeitungszeit (Einzelteile schwer identifizierbar)
 - keine durchgängige Konzentration gefragt
 - geringe Denkleistung für Bearbeitung (keine intensive Problemlösung, ungenügend fokussierte Arbeitssituation, ggf. Hypothese unbeantwortet)
 - Sechstärke kann Leistung beeinflussen (schwerer Vergleich zwischen VP)

Sudoku:

Vorteile: höhere durchgängige Konzentration und Denkleistung erforderlich

- nach erfolgter uV ggf. mehr Zeit zum Weiterspielen erfordert
- Nachteil: Lerneffekte (erprobte Sudoku-Spieler versus Anfänger, Unterschiede zwischen mehreren Durchläufen mit gleicher VP)
- höhere Bearbeitungszeit (Zeitvorgabe ungünstig)
 - schlechte Auswertung und Standardisierung (Fehlerrate bei halbfertiger Lösung schwer ermittelbar, ggf. nur Bearbeitungszeit messbar)
 - Vergleich zwischen mehreren VP schwierig

I. Spielerischer Ansatz

Knobel- und Konzentrationsaufgaben

Ziel: Hemmungen nehmen, "Prüfungssituation" lockern, Konzentration bei Bearbeitung fördern (Denkaufgaben)

Beispiele: „Wassermüllaufgabe“,

„Missionar-Kannibalen-Problem“, „Turm-von-Hanoi“-Aufgabe

Bewertung: bei u.a. Mathematikern, Physikern und Informatikern bekannt, gern in der Lehre verwendet, d.h. Gefahr der Bekanntheit der Aufgabe größer (ggf. verzerrte Ergebnisse, schlechte Eignung)

- schlecht auswertbar (lange zusammenhängende Aufgabe), schwer standardisierbar d.h. als Ganzes zu bearbeiten, große Wahrscheinlichkeit der Nicht-Beendigung in der Versuchszeit (Ergebnis „keine Lösung“ stellt ungünstige Aussage dar)
- schwere Aufteilung in Einzelteile (nicht eindeutig falsch/richtig, Folgefehler möglich)
- keine Teilaufgaben, sodass Anzahl erledigter Einzel-Aufgaben nicht zähl-/vergleichbar
- vergleichbare Parallelaufgabe für zweiten Durchlauf (ohne Lerneffekt) benötigt¹
- nicht einfach und schnell zu erklären (nicht immer selbsterklärend), d.h. erfordert ggf. Eingriff des VL (VL-Effekte, schlechtere Wiederholbarkeit)
- Aufgabe-Fokus zu stark auf Fachgebiet u.a. Mathematik (keine Chancengleichheit für alle VP)
- Zeitdruck bzw. Zeitvorgabe hier ungünstig
- Vergleich zwischen VP ungünstig (persönliche Merkmale entscheidenden Einfluss) (u.a. mathematisches/logisches Denken, Vorkenntnisse, Erfahrungen mit gleichartigen Aufgaben)

¹Flussüberquerungsaufgaben:

Missionar-Kannibalen-Problem, Eifersüchtige-Ehemänner-Problem

Vorteile: etwa gleiche Art und Schwierigkeitsgrad

- Konzentration spielt große Rolle (Konzentrationsaufgabe)
- Nachweis (siehe [KI07, S. 140-149]), dass bei Durchführungen des Missionaren-Kannibalen-Problems mit anschließender Bearbeitung des Eifersüchtige-Ehemänner-Problems praktisch kein Transfer (des Wissens auf zweite Aufgabe) stattfindet
- beide Aufgaben bereits in Versuchen erprobt

Nachteile: Eifersüchtige-Ehemänner-Problem trotzdem etwas höheren Schwierigkeitsgrad

- größere Wahrscheinlichkeit, dass Aufgaben schon bekannt sind
- Wissenstransfer nicht gänzlich auszuschließen
- nicht sicher nachgeweisbar, dass evtl. Lösungs-Schwierigkeiten durch uV bedingt sind
- Bearbeitungszeit (ca. 30 Minuten) sprengt gesetzten zeitlichen Rahmen (ca. 10 Minuten)

II. Kommunikation fördern

Ziel: (aktive) Kommunikation fördern

Ansatz: UL in Vordergrund stellen (Vorwand)

1) VP vorgeblich Benutzbarkeit von Websites prüfen lassen (Usability-Test, evtl. "Thinking Aloud"),

Aufgaben: "Suchen Sie...", (schriftlich) Fragen beantworten

(Fragen müssen generisch sein, dürfen kein Vorwissen fordern)

Beispiele: Vergleich von Uni-Websites (Suchen von Studiengängen, Stundenplänen, Gebühren, Seminare, Formulare, Freizeitangebote), Nutzung von Online-Bibliotheken (Anmeldung, Buchverlängerung, Buchsuche, eBook-Nutzung), Online-Shops

Vorteile: Aufgaben-Nähe zur Erhebungsmethode und -ort (im UL) d.h. glaubwürdiger Vorwand (Usability-Tests)

- Übertragung auf "normale" Arbeitssituation möglich (Nutzung von Websites, Recherche, Suche, Internet-Surfen)

- "freies Surfen" ermutigt VP ggf., Social Media-Seiten zu besuchen

Nachteile: Vorwissen und Lerneffekte (bei Parallelaufgaben) nicht auszuschließen

- uneindeutige Auswertbarkeit (Dauer, "Fehler")
- zweifelhafte Ergebnisse
- eindeutige ja/nein-Fragen erforderlich, um "Fehleranzahl" zu erheben
- individuelle VP-Merkmale beeinflussen Ergebnisse (Versiertheit im Umgang mit Websites, Gewohnheit, Alter, Bildungsstand etc.)
- > Vergleiche zwischen VP nur bedingt möglich
- Ergebnisse nicht eindeutig auf uV zurückzuführen (Ablenkung auf Website, d.h. schwer kontrollierbare SV)
- Lerneffekte bei Parallelaufgaben im zweiten Durchlauf



III. Methodenbezogen (UL)

Ziel: UL im Vordergrund

Aufgabe: Software-Prototyp testen (Aufgaben, Evaluation)

Vorteile: glaubhafter Vorwand (Usability-Test)

Nachteile: (siehe II)

Software-Prototyp benötigt (Aufwand, Kosten)

- Vertrautheit mit Programmart

(schwere Vergleiche zwischen VP, Lerneffekte, Vorwissen)

IV. Genormte Tests (Leistungstests)

Ansatz: unabhängige, genormte Testaufgaben,

Arbeitsleistung mittels Leistungstest erheben

Vorteile: Vergleiche zwischen VP möglich

- leicht standardisier- und auswertbar
 - Parallelaufgaben möglich
 - gut in zeitlichen Rahmen einbettbar (Kurztests)
 - genormte Aufgaben (wiederholbar, unabhängig, reliabel, valide)
 - kurze, unabhängige Einzelaufgaben (keine Folgefehler, leichte Einzelauswertung, Zeitvorgabe möglich, Auswertung auch bei "unfertigen" Lösungen)
 - zusätzlich IQ-Messung möglich
 - allgemeine, unspezifische Aufgaben die kein Vorwissen erfordern, wenig fachspezifisch
 - fordern Konzentration
 - Abstimmung der Aufgaben auf Grundgesamtheit (Alter, Nationalität etc.)
- Nachteile: Anschaffungsaufwand und Kosten

Abb. 18 MindMap zur Entwicklung der Versuchspersonen-Aufgaben (erste Ideenfindung)

ANHANG C. VORLAGEN

Dieser Abschnitt präsentiert auszugsweise relevante Materialien, die zur Durchführung des Versuchs im Rahmen dieser Arbeit entwickelt wurden und sich in vollständiger Form als PDF auf der beige-fügten CD befinden.

C.1 ANWERBUNG VERSUCHSPERSONEN (VOREXPERIMENT)

Anwerbungs-Email für das Vorexperiment (Akquirierung)¹¹³

Betreff: Männliche und weibliche Probanden gesucht! (18.06. und 25.06.)

Liebe ProfessorInnen, MitarbeiterInnen und KommilitonInnen,

im Rahmen unserer Masterarbeit möchten wir im Usability- und Biosignallabor unserer Fachhochschule einen Versuch durchführen. Für dieses Vorhaben benötigen wir noch Personen, die sich für die Probedurchläufe zur Verfügung stellen.

Der Versuch beinhaltet die Aufzeichnung von physiologischen Signalen (Herzrate, Puls, Atmung, Augenbewegungen etc.) bei normaler Arbeitstätigkeit am Rechner. Die Datensammlung soll der Erforschung der physiologischen Vorgänge beim Menschen während der Arbeit dienen.

Während der Aufnahme der physiologischen Daten arbeiten Sie ein paar einfache Aufgaben am Computer ab. Der zeitliche Umfang des Versuchs pro Teilnehmer liegt bei rund 60 Minuten. Ihre Daten sind dabei zu jedem Zeitpunkt anonym, Sie können Ihre eigenen Daten aber gern auf Anfrage einsehen.

Wir bedanken uns schon jetzt für Ihr Bemühen und freuen uns auf Ihre Rückmeldung!

Mit freundlichen Grüßen

Claudia Matthias und Tina Meißner

¹¹³ Stellenweise Umformulierungen durch VL1 (Tina Meißner)

C.2 ANWERBUNG VERSUCHSPERSONEN (HAUPTEXPERIMENT)

Anwerbungs-Email für das Hauptexperiment (Akquirierung)¹¹⁴

Betreff: Masterarbeit: Weibliche und männliche TeilnehmerInnen gesucht! (Brandenburg/Havel)

Sehr geehrte/r Damen und Herren,

im Rahmen unserer Masterarbeit führen Frau Tina Meißner und ich, gemeinsam mit Mitarbeitern und Professoren der Fachhochschule Brandenburg, Versuche im hiesigen Usability- und Biosignallabor durch. Hierfür benötigen wir noch Personen, die sich für die Versuche zur Verfügung stellen und uns so ermöglichen, unsere Masterarbeit erfolgreich zu gestalten.

Der Versuch beinhaltet die Aufzeichnung physiologischer Signale (Herzrate, Puls, Atmung, Augenbewegung etc.) bei normaler Arbeitstätigkeit am Computer. Die Datensammlung dient der Erforschung physiologischer Vorgänge beim Menschen während der Arbeit am Computer, um in nachfolgenden wissenschaftlichen Untersuchungen das Arbeitsverhalten von Arbeitnehmern in organisierte, gesunde Strukturen zu überführen.

Während der Aufnahme der physiologischen Daten arbeiten Sie einige einfache Aufgaben am Computer ab. Der zeitliche Umfang des Versuchs **pro Teilnehmer** liegt bei **rund 60-80 Minuten**. Ihre Daten sind dabei zu jedem Zeitpunkt **anonym**, Sie können diese aber auf Anfrage gern einsehen. Eine monetäre Entschädigung für Ihren Aufwand ist leider nicht möglich. Als kleines Dankeschön für die Teilnahme stellen wir aber im Anschluss an den Versuch Snacks und Getränke bereit.

Insgesamt werden aus Ihrem Haus optimal **sechs Teilnehmer (drei männlich, drei weiblich)** benötigt. Ein möglichst großer Querschnitt Ihrer Belegschaft, sowohl Alter (ab 18) als auch Position betreffend, wäre dabei wünschenswert. Die tägliche Verwendung des Computers als Arbeitsmittel wird bei allen Teilnehmern als Bedingung vorausgesetzt. **Die Versuche sollen im Zeitraum vom 02.08. – 10.08.2012 stattfinden.** Genaue Termine würden wir mit den Mitarbeitern **persönlich vereinbaren.**

Wir bedanken uns schon jetzt für Ihr Bemühen und freuen uns auf Ihre Rückmeldung!

Mit freundlichen Grüßen

Claudia Matthias und Tina Meißner

¹¹⁴ Stellenweise Umformulierungen durch VL1 (Tina Meißner)

C.3 TEILNEHMERINFORMATION

Informations-Email vorab nach Teilnahmebestätigung der Versuchsperson (in Vor- und Hauptexperiment)¹¹⁵

Betreff: Informationen für den Teilnehmer

Versuch: Aufnahme von verschiedenen physiologischen Parametern

Dauer: 60-80 Minuten (und anschließende Erfrischung)

Verkabelung:

Die Verkabelung mit den verschiedenen Elektroden und Sensoren erfolgt nur oberflächlich und ist nicht schmerzhaft oder unangenehm.

- Elektroden im Gesicht (4)
- Elektroden auf Oberkörper (3)
- Elektrode am Unterarm (1)
- Elektrode im Nackenbereich (1)
- Brustgurt
- 3 Sensoren an einer Hand

Hinweise für alle Teilnehmer:

Bitte lockere, bequeme Kleidung tragen (weiteres T-Shirt o. ä.)

Hinweise für Männer:

Bitte Rasur des Oberkörpers/Nackens (bei *starker* Behaarung)

Hinweise für Frauen:

Bitte kein starkes Make-up (Puder, Rouge etc.)

¹¹⁵ Alle Angaben von VL1 (Tina Meißner)

C.4 KOMMUNIKATIONSÜBERSICHT

Vorwand: Es soll der Versuchsperson glaubhaft gemacht werden, dass neben den beiden ausführenden Versuchsleitern noch weitere höhergestellte Personen an dem Experimente beteiligt sind (Autoritätspersonen), die sich extern zuschalten und ggf. kurzfristige Änderungen veranlassen. Der Kontakt mit der erfundenen Persona (Dr. Thomas Schmidt-Weber) erfolgt immer über den Versuchsleiter 2 (Übermitteln von Anweisungen via Skype, Chat, Weiterleitung von Emails, Eintreten in Labor zur Übergabe von zusätzlichem Material/ Anweisungen), der sich während des Versuchs außerhalb des Usability-Labors befindet¹¹⁶ (externe Überwachung).

Tab. 22 Definition der Kommunikation zur aktiven Manipulation der unabhängigen Variable

Mittel	Vorwand
Chat (CH1), Telefon (TE1)	Ankündigung: neue Aufgabe zusätzlich (per Email)
EM (EM1)	Senden der neuen zusätzlichen Aufgabe per Email
Email (EM2), Chat (CH2)	Ankündigung: Fragebogen ausfüllen (nach Versuch), umgekehrt zur Seite legen
Chat (CH3), Störer (ST2)	Varianten: 1) Ansage der verbleibenden Zeit (durch Versuchsleiter 1) oder 2) Erkundigung nach Fortschritt/Problem/Hilfe anbieten (Versuchsleiter 2: Ansprechpartner)
Störer (ST1)	Fragebogenübergabe, Fragebogen ausfüllen (nach Versuch), umgekehrt zur Seite legen

Standardisierte Textvorlagen:

EM1: Sehr geehrte/r TeilnehmerIn, anbei erhalten Sie die noch fehlenden Aufgaben, die Sie bitte zum Schluss bearbeiten. Nach vollständiger Bearbeitung der Aufgaben **behalten Sie die PDF bitte geöffnet**. Mit freundlichen Grüßen, Dr. Thomas Schmidt-Weber

EM2: Sehr geehrte/r TeilnehmerIn, nach Beendigung des Versuchs bitten wir Sie, noch einen Fragebogen auszufüllen, den Sie in Kürze von Frau Matthias erhalten. Mit freundlichen Grüßen, Dr. Thomas Schmidt-Weber

CH2: Hallo, nach Beendigung des Versuchs bitten wir Sie, noch einen Fragebogen auszufüllen, den Sie in Kürze erhalten. Vielen Dank!

CH1, TE1: Es ist noch eine Aufgabe hinzugekommen, die Sie bitte am Ende ebenfalls bearbeiten. Sie erhalten diese direkt per Email. Bitte aktualisieren Sie gegebenenfalls Ihr Email-Postfach, um die neue Email zu empfangen. Vielen Dank!

CH3 (Dialog mit Versuchsperson): Hallo, sind die Aufgaben verständlich? Oder gibt es irgendwelche Fragen? Hallo, kommen Sie gut voran oder brauchen Sie Hilfe?, Hallo, ich wollte mich nur erkundigen, ob alles in Ordnung ist oder ob Sie irgendwo Hilfe brauchen?

ST1: Nach Beendigung des Versuchs würden wir Sie bitten, noch diesen Fragebogen auszufüllen. Er ist vollständig anonym. Wenn Sie fertig sind, legen Sie ihn einfach umgekehrt zur Seite. Vielen Dank!

ST2: „Sind die Aufgaben verständlich oder haben Sie Fragen?“, „Kommen Sie gut voran oder hängen Sie irgendwo fest? Sie können Frau Matthias gerne auch Fragen stellen oder um Hilfe bitten.“, „Es verbleiben noch 2 Minuten.“

¹¹⁶ Vorwand für VL2-Abwesenheit in der EG ist „Biosignale nicht durch Anwesende verfälschen“ und in der KG aufgrund von Terminen.

C.5 ABLÄUFE

Tab. 23 Vordefinierte Abläufe (variiert) für die Experimentalgruppe: (1 Minute – 2 Minuten – 2 Minuten – 2 Minuten – 2 Minuten)

Zeit hh:mm	Störer	Ablauf 1 (ST2 - CH1 - EM1 – CH2 – ST1)
00:01	ST 2	Sind die Aufgaben verständlich oder haben Sie Fragen? Sie können Frau Matthias gerne auch Fragen stellen oder um Hilfe bitten.
00:03	CH 1	Es ist noch eine Aufgabe hinzugekommen, die Sie bitte am Ende ebenfalls bearbeiten. Sie erhalten diese direkt per Email. Bitte aktualisieren Sie gegebenenfalls Ihr Email-Postfach, um die neue Email zu empfangen. Vielen Dank!
00:05	EM 1	Sehr geehrte/r TeilnehmerIn, anbei erhalten Sie die noch fehlenden Aufgaben, die Sie bitte zum Schluss bearbeiten. Nach vollständiger Bearbeitung der Aufgaben behalten Sie die PDF bitte geöffnet . Mit freundlichen Grüßen, Dr. Thomas Schmidt-Weber
00:07	CH 2	Hallo, nach Beendigung des Versuchs bitten wir Sie, noch einen Fragebogen auszufüllen, den Sie in Kürze erhalten. Vielen Dank!
00:09	ST 1	Nach Beendigung des Versuchs würden wir Sie bitten, noch diesen Fragebogen auszufüllen. Er ist vollständig anonym. Wenn Sie fertig sind, legen Sie ihn einfach umgekehrt zur Seite. Vielen Dank!

Zeit hh:mm	Störer	Ablauf 2 (TE1 – EM1 - CH3 - CH2 – ST1)
00:01	TE 1	Es ist noch eine Aufgabe hinzugekommen, die Sie bitte am Ende ebenfalls bearbeiten. Sie erhalten diese direkt per Email. Bitte aktualisieren Sie gegebenenfalls Ihr Email-Postfach, um die neue Email zu empfangen. Vielen Dank!
00:03	EM 1	Sehr geehrte/r TeilnehmerIn, anbei erhalten Sie die noch fehlenden Aufgaben, die Sie bitte zum Schluss bearbeiten. Nach vollständiger Bearbeitung der Aufgaben behalten Sie die PDF bitte geöffnet . Mit freundlichen Grüßen, Dr. Thomas Schmidt-Weber
00:05	CH 3	Hallo, ich wollte mich nur erkundigen, ob alles in Ordnung ist oder ob Sie irgendwo Hilfe brauchen?
00:07	CH 2	Hallo, nach Beendigung des Versuchs bitten wir Sie, noch einen Fragebogen auszufüllen, den Sie in Kürze erhalten. Vielen Dank!
00:09	ST 1	Nach Beendigung des Versuchs würden wir Sie bitten, noch diesen Fragebogen auszufüllen. Er ist vollständig anonym. Wenn Sie fertig sind, legen Sie ihn einfach umgekehrt zur Seite. Vielen Dank!

Zeit hh:mm	Störer	Ablauf 3 (CH1 – EM1– EM2 – ST1 - ST2)
00:01	CH 1	Es ist noch eine Aufgabe hinzugekommen, die Sie bitte am Ende ebenfalls bearbeiten. Sie erhalten diese direkt per Email. Bitte aktualisieren Sie gegebenenfalls Ihr Email-Postfach, um die neue Email zu empfangen. Vielen Dank!
00:03	EM 1	Sehr geehrte/r TeilnehmerIn, anbei erhalten Sie die noch fehlenden Aufgaben, die Sie bitte zum Schluss bearbeiten. Nach vollständiger Bearbeitung der Aufgaben behalten Sie die PDF bitte geöffnet . Mit freundlichen Grüßen, Dr. Thomas Schmidt-Weber
00:05	EM 2	Sehr geehrte/r TeilnehmerIn, nach Beendigung des Versuchs bitten wir Sie, noch einen Fragebogen auszufüllen, den Sie in Kürze von Frau Matthias erhalten. Mit freundlichen Grüßen, Dr. Thomas Schmidt-Weber
00:07	ST 1	Nach Beendigung des Versuchs würden wir Sie bitten, noch diesen Fragebogen auszufüllen. Er ist vollständig anonym. Wenn Sie fertig sind, legen Sie ihn einfach umgekehrt zur Seite. Vielen Dank!
00:09	ST 2	Es verbleiben noch zwei Minuten.

Email: t.schmidt-weber@gmx.de (Versuchsleiter 2), probandFHB@web.de (Versuchsperson)

Skype: claudiafhb (Versuchsleiter 2), Proband2012 (Versuchsperson)

C.6 ZEITABLAUPLAN

Tab. 24 Zeitablauf mit allen Schritten des Versuchsablaufs für die Versuchsleiter

Vorbereitung vor jeder Durchführung	
VL 1, VL 2	Alle Kommunikationsmittel einrichten, Computer präparieren (für Versuchsperson & Versuchsleiter) (Skype, Email-, Chat-Programm, Teamviewer, Bildschirmschoner & Stand-By ausschalten)
VL 1	Geräte und Software einrichten und testen
VL 2	Fotos/Skizzen/Beschreibung zur Dokumentation (Versuchsaufbau, s. <i>Vorgaben im Messprotokoll</i>), mit und ohne Versuchsperson
	Aufbau und Einrichten der IP- & Videokameras
	Materialien bereitstellen (Fragebögen, Instruktionen, Aufgaben), inkl. Versuchspersonen-Code
	<u>Protokoll vorbereiten:</u> 1) Äußere Bedingungen (Temperatur, Getränke etc.) 2) Szenario (Messgeräte, Aufgaben, Kontroll- oder Experimentalgruppe mit Störungsplan)
	Buffet- und Warteraum vorbereiten (voneinander trennen), Stühle ohne Rollen für Versuchspersonen
	Emails mit Instruktionen & Versuchspersonen-Aufgaben vorbereiten
	Ggf. Aktivierung des TeamViewer Mitschnitts (in den Vorexperimenten)
Durchführung (Ablaufplan)	
VL 1, VL 2	Vorbereitung des Labors (s. o.)
VL 1	Begrüßung der Versuchsperson
	<u>Mündliche Einführung der Versuchsperson:</u> * 1) Kurze Erklärung zum Experiment & Zweck 2) Geräte/Funktionsweise & Vorgehen (Aufklärung über Video- & Bildschirmaufzeichnung) 3) Funktion der Versuchsleiter & Handlungsrahmen für Versuchsperson 4) Anonymität & Dokumentunterzeichnung (Datenschutz, Schweigepflicht, Einverständnis)
	<u>Kurze Erklärung zu den Kommunikationsmitteln:</u> (im Experimentalgruppen-Durchlauf) 5) zeigen (Skype-Telefon, Email-Programm, Chat-Programm etc.) 6) „alles benutzen“, „bitte Telefon beantworten“ etc. 7) Erlaubnis erteilen, eigene Kommunikationsmittel zu nutzen (soz. Netzwerke, Handy ...)
	Stimmungsfragebogen ausfüllen lassen (<i>vor</i> Durchführung)
	Erklärung der Aufgaben, Absolvierung 1 Testaufgabe pro Typ (MWT, IQ) zur Probe
	Anbringung und Justierung der technischen Geräte an Versuchsperson
	Geräte/Software einrichten (Synchronisation aller Geräte, ggf. Assistent)
VL 2	Dokumentation, Protokoll
	Anweisungen an Versuchsperson per Email (schriftliche Instruktionen)
	Aufgaben an Versuchspersonen (im Vorexperiment per Email, im Hauptexperiment auf dem Desktop vorliegend)
VL 1	Nachfragen, ggf. Hilfestellung bei Fragen zu Instruktionen & Aufgaben, 1-minütige Ruheaufnahme der Biosignale (Normalwerte ermitteln) 1) Versuch starten, Zeit stoppen (benötigte Zeit für Vorbereitung notieren) 2) Bereitstehen der Versuchsleiter für evtl. Rückfragen (Ansprechpartner)
	VL 2 (unaufdringliche) Beobachtung und Protokollierung der Durchführung (ggf. Markierung interessanter Stellen)
VL1, VL2	kurze Pause zwischen Durchgang 1 und 2: Instruktionen für Durchgang 2 versenden (Versuchsleiter 2), Ausdrucken der Versuchspersonen-Aufgaben aus Durchgang 1 & Abspeicherung der Daten aus Durchgang1 (Versuchsleiter 1), ggf. kurzes Gespräch mit Versuchsperson (Versuchsleiter 2)
VL 1	Versuch beenden nach vorgegebener Zeit
	Aushändigung Fragebogen / Aufforderung zum Ausfüllen

	Bearbeitete Versuchspersonen-Aufgaben ausdrucken und benennen (Versuchspersonen-Code)
	Email-Postfach der Versuchsperson leeren (für Durchlauf mit neuer Versuchsperson)
	Löschen des Versuchsperson Skype-Protokolls (für Durchlauf mit neue Versuchsperson)
	Ggf. Speicherung von Mitschnitten (TeamViewer-Session, Skype-Verlauf) in den Vorexperimenten
	Ggf. Fotos, Skizzierung des Versuchsaufbaus
	Versuchsperson über „wahres“ Versuchsziel aufklären (standardisiertes „Debriefing“)
VL 1,	Verabschiedung der Versuchsperson (Weiterleitung in Buffet-Raum, Erfrischung anbieten als Belohnung)
VL 2	Nachbereitung (s. u.)
Nachbereitung nach jeder Durchführung (Vervollständigung und Nummerierung)	
VL 1	Speichern der Messdaten (Benennung, Versuchspersonen-Code, Verzeichnisse)
VL 2	Protokoll und Messprotokoll beenden und nummerieren (Versuchspersonen-Code)
VL 1, VL 2	Ggf. Auffälligkeiten/Störeinflüsse notieren, Unregelmäßigkeiten bei Ablauf / Durchführung
VL 2	Fragebögen nummerieren (Versuchspersonen-Code)
	Ggf. weitere Fotos (zusätzlich Anzahl notieren, Versuchspersonen-Code)
VL 1, VL 2	Alle Geräte und Materialien „zurücksetzen“

<p><u>Legende:</u> VL 1: Tina Meißner VL 2: Claudia Matthias</p>
--

Mündliche Einführung der Versuchsperson (Standardtext)

Erklärungen vor dem 1. Durchlauf:

Zweck: Gemeinsam mit Mitarbeitern und Professoren der Fachhochschule Brandenburg, möchten wir einen Versuch im Usability- und Biosignallabor der Fachhochschule durchführen.

Der Versuch beinhaltet die Aufzeichnung von physiologischen Signalen (Herzrate, Puls, Atmung, Augenbewegungen etc.) bei normaler Arbeitstätigkeit am Rechner. Die Datensammlung soll der Erforschung der physiologischen Vorgänge beim Menschen bei verschiedenen Arbeitstätigkeiten dienen, um eine Biosignal-Datenbank aufzubauen und in nachfolgenden wissenschaftlichen Untersuchungen das Arbeitsverhalten für Arbeitnehmer in organisierte, gesunde Strukturen zu überführen.

Ablauf: Während der Aufnahme der physiologischen Daten arbeiten Sie ein paar einfache Aufgaben am Computer ab, die parallel zu den Biosignalen als separate Biosignalkategorie ebenfalls ausgewertet werden. Wir werden zur Sicherstellung der Messungen zwei Durchgänge mit vergleichbaren Aufgaben und Ablauf durchführen. Beide Abläufe dauern je 10 Minuten. [Der Versuch und Ihre Bildschirmaktivitäten werden zur Auswertung zusätzlich per Video aufgezeichnet. Die Daten bleiben vertraulich und werden ohne Ihre Zustimmung nicht anderweitig verwendet. Bitte bearbeiten Sie beide PDFs, d.h. MWT und IQ und starten Sie mit dem MWT.](#)

Zeit und Organisation: Der zeitliche Umfang des gesamten Experiments liegt bei rund 60-80 Minuten. Ihre Daten sind dabei zu jedem Zeitpunkt anonym. Sie können Ihre eigenen Daten aber gerne auf Anfrage einsehen. Als kleines Dankeschön für Ihre Teilnahme stellen wir im Anschluss an den Versuch Snacks und Getränke im Nebenraum für Sie bereit.

Mitarbeiter / Versuchsleiter: Ich (Frau Meißner)¹¹⁷ bin für die Aufnahme der Biosignale und Technik zuständig. Frau Matthias, mit der Sie bereits per Email Kontakt hatten, hat die Aufgaben und Fragebögen für Sie zusammengestellt und führt das Nachgespräch. Sie ist bei Fragen Ihr Ansprechpartner. Leider kann unser Professor selbst bei der Durchführung diese Woche nicht anwesend sein. Er steht aber per Email in Kontakt. Sollten also noch irgendwelche kurzfristigen Änderungen bei den Versuchsaufgaben notwendig sein, bekommen Sie diese direkt per Email oder durch Frau Matthias mitgeteilt. Sie erhalten außerdem gleich per Email die Instruktionen und Versuchsaufgaben. [Versuchsleiter 1: Hinweis, dass Email-Postfach ggf. manuell aktualisiert werden muss, um neue Emails zu empfangen.](#)

Falls Kontrollgruppe im 1. Durchlauf: Leider hat Frau Matthias gerade noch einen Termin, wird aber im Laufe des Versuchs dazu stoßen und sich ggf. bei notwendigen Änderungen mit Ihnen in Kontakt setzen.

Falls Experimentalgruppe im 1. Durchlauf: Frau Matthias ist per Skype, Chat und Email erreichbar. Bei noch offenen Fragen können Sie sich also gerne an Sie wenden.

Kommunikationsmittel: Wir haben dazu für Sie ein Email-Programm eingerichtet und zusätzlich noch Telefon (Skype) und eine Chat-Funktion. [Versuchsleiter 1: Kommunikationsmittel am Bildschirm zeigen und ggf. kurz erklären / vorführen, falls Versuchsperson damit nicht vertraut ist.](#)

Erklärungen vor dem 2. Durchlauf:

Falls Experimentalgruppe im 2. Durchlauf:

Versuchsleiter 1: „Ablauf und Aufgabenart sind wie im ersten Durchlauf.“

Versuchsleiter 2 (tritt ein): „Ich bin Frau Matthias, wir standen bereits per Email in Kontakt. Ich werde während der Durchführung nicht im Raum sein, um die Biosignal-Aufnahmen nicht zu verfälschen. Ich bin aber ab jetzt per Skype jederzeit ansprechbar, falls Sie noch Fragen oder Rückmeldungen haben. Sollte sich noch etwas am Ablauf ändern, werde ich mich bei Ihnen melden.“

Falls Kontrollgruppe im 2. Durchlauf:

Versuchsleiter 2 (tritt ein): „Ich bin leider ab jetzt nicht mehr erreichbar, da ich noch einen Termin habe.“

Genormte Antworten auf Fragen zu den MWT-Aufgaben (s. [Le05]):

Frage: „Kann mehr als ein Wort je Zeile richtig sein?“

Antwort: „Es ist *höchstens* ein Wort pro Zeile richtig. Es dürfen nie zwei Wörter in einer Zeile angekreuzt werden.“

Frage: „Wie viel Zeit steht mir zur Verfügung?“

Antwort: „Es gibt keine Zeitvorgabe für die Beantwortung dieses Tests.“

¹¹⁷ VL1 gibt vor, VP-Aufgaben nicht zu kennen, um Kommunikation (insbesondere in KG) zwischen sich und VP vorzubeugen. In EG soll Kommunikation möglichst nur zwischen VL2 und VP über o. g. Kommunikationsmittel erfolgen. Hierzu wird VL2 als Verantwortliche/r für VP-Aufgaben genannt und Kontaktmöglichkeiten (Skype, Chat, Email) vorab kurz vorgestellt.

C.7 ANLEITUNG VERSUCHSPERSON

Instruktionen

Sehr geehrte/r TeilnehmerIn,

willkommen zu unserem Experiment. Bitte lesen Sie die folgenden Instruktionen sorgfältig durch, bevor Sie beginnen. Falls Sie etwas nicht verstehen oder Fragen haben, können Sie die Versuchsleiterin gerne um Hilfe bitten.

Zweck: In diesem Experiment interessieren wir uns für die Messung Ihrer Biosignale bei verschiedenen Arbeitstätigkeiten.

Sie können während der Durchführung Ihrem ganz normalen Arbeitsverhalten am Computer nachgehen. Für uns ist die Aufzeichnung Ihrer physiologischen Daten vordergründig, was leider einige Einschränkungen in der normalen Bewegung erforderlich macht. Diese werden Ihnen noch von der Versuchsleiterin erläutert.

Die Ergebnisse der Aufgaben, die Sie während der Aufzeichnung lösen, fließen ebenfalls in die Bewertung des Versuchs ein und bilden eine separate Biosignalkategorie.

Ablauf: Bitte beginnen Sie mit der Bearbeitung der **MWT-Aufgabe (PDF)**, wenn Ihnen die Versuchsleiterin den Start der Aufzeichnung signalisiert.

Nach vollständiger Bearbeitung der PDF **behalten Sie diese bitte geöffnet** und widmen Sie sich den weiteren Aufgaben (PDF).

Nach 10 Minuten wird der Versuch beendet. Zur Sicherstellung der ersten Messung erfolgt ein **zweiter Durchlauf mit vergleichbaren Aufgaben und Ablauf** (bitte wieder mit der **MWT-Aufgabe starten**).

Der gesamte Versuch (inklusive Vor- und Nachbereitung) sollte nicht länger als 60-80 Minuten dauern.

Vielen Dank für Ihre Teilnahme am Experiment!

Mit freundlichen Grüßen

Dr. Thomas Schmidt-Weber

C.8 PROBEAUFGABEN

Aufgaben Beispiele

1) MWT-Aufgabe

Anweisung: Sie sehen Reihen mit Wörtern. In jeder Reihe steht **höchstens ein Wort**, dass Ihnen vielleicht bekannt ist. Wenn Sie es gefunden haben, setzen Sie bitte ein Häkchen dahinter.

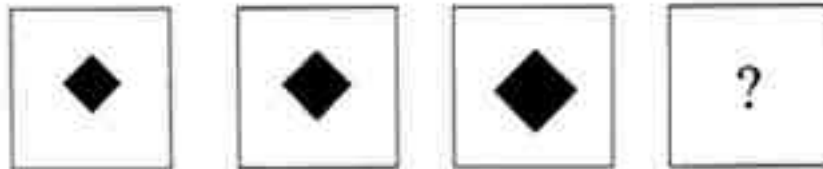
Beispiel: Mera Arme Rame Maer Raem

2) IQ-Aufgabe

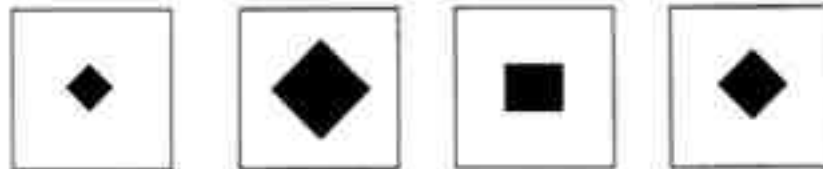
Thema: Logisches Denken

Anweisung: Die drei Figuren verändern sich nach bestimmten Regeln. Ergänzen Sie die vierte Figur aus den Lösungsvorschlägen der zweiten Reihe so, dass sie die Reihe sinnvoll fortsetzt.

Beispiel:



**Lösungs-
vorschläge:**



a

b

c

d

Lösung:

Thema: Zahlenreihen

Anweisung: Die Zahlenreihen sind nach einer bestimmten Regel aufgebaut. Ergänzen Sie die nächste Zahl, welche die Reihe richtig fortsetzt.

Beispiel:

7 5 8 6 9 7 ?

a) 11

b) 8

c) 10

d) 12

Lösung:

Lösungen

1) MWT-Aufgabe

Lösung: Mera Arme Rame Maer Raem

2) IQ-Aufgabe

Thema: Logisches Denken

Lösung:

Lösung: b
Die Rauten werden größer.

Thema: Zahlenreihen

Lösung:

Lösung: c)
Regel: $-2 +3 -2 +3 \dots$

C.9 ANLEITUNG VERSUCHSLEITER

Anleitung für Versuchsleiter: idealerweise sollte der Versuch mit 2 Versuchsleitern durchgeführt werden, die die (technische) Durchführung beherrschen aber nicht das „wahre“ Versuchsziel kennen d.h. schriftliche Instruktionen und größtmögliche Standardisierung des Versuchsaufbaus/-durchführung zur Vermeidung von Störvariablen (u.a. Versuchsleiter-Effekt), verbesserte Wiederholbarkeit, Objektivität und Generalisierung (*hier:* aus organisatorischen Gründen Doppelblindversuche oder Durchführung mit „externen“ Versuchsleitern nicht möglich).

Tab. 25 Aufbau und Funktion der Versuchsbeteiligten

Experimentator	Versuchsleiter 1	Versuchsleiter 2
„vorgetäuschter“ Experimentator (Autoritätsperson) außerhalb des Labors: tritt mit Versuchsperson per Email in Kontakt (u.a. Instruktionsübergabe)	- Technik-Verantwortliche - möglichst nicht während Durchführung mit Versuchsperson kommunizieren - kennt Fragebögen und Aufgabe nicht	- tritt in einigen Szenarien als „Störer“ auf - während Durchführung nicht anwesend (externe Beobachtung durch Live-Audio/Video-Übertragung) - soll alleinig mit Versuchsperson kommunizieren (über definierte Kommunikationsmittel) als Ansprechpartner - Verantwortliche für Fragebögen, Aufgaben, Debriefing
<p>2 Versuchsleiter (<i>neben Versuchsperson die einzigen Anwesenden während der Durchführung</i>):</p> <ul style="list-style-type: none"> - <i>Vorwand:</i> Versuchsleiter sind „nur Durchführende“ des Experiments und sorgen für reibungslosen Untersuchungsablauf - Vorwand (<i>getarntes Versuchsziel</i>) und Experiment so erklären, dass Aufbau eindeutig ist und kein Misstrauen erzeugt - Störvariablen-Kontrolle (Aufdeckung, Vermeidung): Versuchsleiter-Effekte (gegenseitiges Beobachten der Versuchsleiter, speziell im Vorexperiment), Versuchspersonen-Effekte (unaufdringliche Beobachtung, Auswertung), u.a. organisatorische Fehler, Geräte, Messung 		

Auftreten der Versuchsleiter: professionell, freundlich aber „studentisch“ („normale“ Alltagsbekleidung), unaufdringlich, unterstützend (Ansprechperson während des Versuchs)

Vermeiden: Einschüchterung, autoritäre Wirkung, offenes Beobachten der Versuchsperson

Aufgaben (klare Einteilung der Versuchsleiter *vor* Durchführung):

- Vorbereitung des Labors (Geräte, Materialien, Kommunikationsmittel)
- *Begrüßung und Einführung der Versuchsperson:* Erklärung zum Versuch (Zweck, Geräte, Vorgehen, Handlungsrahmen für Versuchsperson, Datenschutz-, Schweigepflicht-, Einverständniserklärung)
- Stimmungsfragebogen ausfüllen lassen (*vor* Durchführung)
- Anbringung und Einstellung der technischen Geräte
- Erklärung der Versuchsaufgaben, falls notwendig
- Versuch starten und Bereitstehen für Rückfragen der Versuchsperson (Ansprechpartner)
- (unaufdringliche) Beobachtung, Protokollierung (Interessantes, mögliche Störvariablen)
- Gegenseitige Beobachtung der Versuchsleiter, Versuchsperson beobachten (Störvariablen aufdecken)
- Beendigung des Versuchs (Zeit stoppen, Zeitvorgabe überwachen)
- Aushändigung Fragebogen (Versuch, Technik)
- „Debriefing“ (Versuchsziel aufdecken), Weiterleitung in Buffet-Raum, Verabschiedung
- Nachbereitung (Materialien vervollständigen, Versuchspersonen-Code einfügen, Daten speichern)

C.10 STIMMUNGSFRAGEBOGEN

Vielen Dank für Ihre Teilnahme am Experiment! Vor der Durchführung möchten wir Sie bitten, noch die folgenden Fragen zu beantworten.

Die Antworten können nicht auf Sie persönlich zurückgeführt werden.

Es ist wichtig, dass Sie die Fragen ehrlich beantworten, damit die Daten beim Experiment sinnvoll ausgewertet werden können.

Bitte kreuzen Sie immer nur eine Möglichkeit an.

Angaben zur momentanen Stimmung

- 0 – trifft momentan überhaupt nicht zu
- 3 – trifft halb und halb zu
- 6 – trifft völlig zu

Mit den Abstufungen 1, 2, 4 und 5 sind weitere Differenzierungen möglich.

<i>Ich fühle mich jetzt...</i>	0	1	2	3	4	5	6
Ausgeruht							
Zufrieden							
Gutgelaunt							
Entspannt							
Unvoreingenommen							
Zuversichtlich							
Stabil							
Aufnahmefähig							
Heiter							
Ungezwungen							

PHYSIOLOGISCHE SIGNALE¹¹⁸

Angaben zum Alltag: Ich bin...

Rechtshänder Linkshänder

Nichtraucher Raucher

Wie viele Zigaretten täglich: _____

Wann zuletzt geraucht: _____

Lese-Rechtschreib-Schwäche: ja nein

Regelmäßiger Kaffeetrinker:

nein ja

Wie viele Tassen pro Tag: _____

Wann zuletzt Kaffee getrunken: _____

¹¹⁸ Angaben zu Abschnitt „Physiologische Signale“ von VL1 (Tina Meißner)

Getränke, die ich hauptsächlich trinke: _____

Ich treibe regelmäßig Sport:

nein ja

Welchen: _____

Wie oft: _____

Wann zuletzt Sport getrieben: _____

Ich habe letzte Nacht geschlafen:

Mehr oder weniger als sonst. Stunden (circa): _____

Ich habe geträumt: Gutes Schlechtes Wirres Nichts

Ich habe entspannt durchgeschlafen.

Ich war vor dem Weckerklingeln wach.

Ich bin schnell aus dem Bett gekommen.

Ich habe mind. 1x auf Snooze gedrückt.

Ich war ständig wach , weil _____

Ich brauchte lange zum Einschlafen , weil _____

Ich bin eher ein „Morgenmuffel“ „Frühaufsteher“

Ich bin im Alltag auf der Arbeit Stress ausgesetzt:

nie selten gelegentlich oft immer

Vielen herzlichen Dank, dass Sie sich die Zeit nehmen, an dem Experiment teilzunehmen und den Fragebogen auszufüllen!

C.11 VERSUCHSFRAGEBOGEN

Vielen Dank für Ihre Teilnahme am Experiment! Wir möchten Sie bitten, noch den folgenden Fragebogen zu beantworten.

Die Antworten können nicht auf Sie persönlich zurückgeführt werden.

Es ist wichtig, dass Sie die Fragen ehrlich beantworten, damit die Daten sinnvoll ausgewertet werden können. **Bitte kreuzen Sie immer nur eine Möglichkeit an.**

TEIL I: Angaben zur Person

Geschlecht: weiblich männlich

Lebensalter: _____

Höchster bisheriger Bildungsabschluss:

- Kein Abschluss
- Hauptschulabschluss
- Realschulabschluss
- Abitur oder andere Hochschulzugangsberechtigung
- Abgeschlossene Berufsausbildung
- Meister
- Abgeschlossenes Studium
- Promotion
- Habilitation
- Sonstiges

Angestellte/r: ja nein , sondern _____

Wochenarbeitszeit: _____ Stunden

Anzahl im Büro verbrachter Stunden (pro Woche): ca. _____ Stunden

TEIL II: Angaben zum Versuch

Allgemeines Verhalten am Computer:

1 – nie 2 – selten 3 – gelegentlich 4 – oft 5 – immer

	1	2	3	4	5
Ich benutze den Computer zum Arbeiten					
Das Arbeiten mit dem Computer fällt mir leicht					
Ich benutze den Computer auch in meiner Freizeit.					

Im Alltag verwende ich folgende Kommunikationsmittel:

1 – nie 2 – selten 3 – gelegentlich 4 – oft 5 – immer

	1	2	3	4	5
Telefon, Handy, Internet-Telefonie (Skype o.ä.)					
SMS					
Emails					
Soziale Netzwerke (Facebook, StudiVZ, Xing o.ä.)					
Messenger, Kurznachrichtendienste (Twitter, ICQ, MSN, AIM, AOL, Yahoo Messenger o.ä.)					

	1	2	3	4	5
Während des Arbeitens am Computer kommuniziere ich mit anderen (via oben genannter Mittel)					

TEIL III: Angaben zur Organisation

Persönlicher Eindruck zu den gestellten Aufgaben:

1 –gar nicht 2 –wenig 3 – mittelmäßig 4 – ziemlich 5 - völlig

<i>Die MWT-Aufgaben (1. PDF) ...</i>	1	2	3	4	5
...waren für mich verständlich und eindeutig					
...empfand ich als zu schwer					
...empfand ich als zu leicht					
...waren mir bereits bekannt					
<i>Die IQ-Aufgaben (2. PDF) ...</i>	1	2	3	4	5
...waren für mich verständlich und eindeutig					
...empfand ich als zu schwer					
...empfand ich als zu leicht					
...waren mir bereits bekannt					

Anmerkungen zu den Versuchsaufgaben: _____

Persönlicher Eindruck zum Versuchsablauf:

1 –gar nicht 2 –wenig 3 – mittelmäßig 4 – ziemlich 5 - völlig

	1	2	3	4	5
Die Anweisungen waren für mich eindeutig					
Ich fühlte mich aufgeregt					
Ich fühlte mich zeitlich unter Druck gesetzt					
Ich fühlte mich durch die Anwesenheit anderer unter Druck gesetzt					
Ich fühlte mich beobachtet					
Ich fühlte mich unwohl					

Anmerkungen / Probleme: _____

Persönlicher Eindruck zur Arbeitssituation:

1 –gar nicht 2 –wenig 3 – mittelmäßig 4 – ziemlich 5 - völlig

<i>Im ersten Durchlauf...</i>	1	2	3	4	5
... fühlte ich mich abgelenkt					
... fühlte ich mich gestört					
... fühlte ich mich gezwungen oder genötigt					
... verhielt ich mich anders als bei der alltäglichen Arbeit					
... konnte ich die Aufgaben schlechter lösen					
... konnte ich die Aufgaben besser lösen					

<i>Im zweiten Durchlauf...</i>	1	2	3	4	5
... fühlte ich mich abgelenkt					
... fühlte ich mich gestört					
... fühlte ich mich gezwungen oder genötigt					
... verhielt ich mich anders als bei der alltäglichen Arbeit					
... konnte ich die Aufgaben schlechter lösen					
... konnte ich die Aufgaben besser lösen					

Ich habe mich gestört gefühlt durch: _____

Vielen herzlichen Dank, dass Sie sich die Zeit genommen haben, an dem Experiment teilzunehmen und den Fragebogen auszufüllen!

Weitere allgemeine Anmerkungen:

C.12 DEBRIEFING

Anleitung für „Debriefing“-Gespräch

Muttersprache: _____

„Physiologische Signale“¹¹⁹

Welche Geräte / Sensoren wurden als störend empfunden?

Welche Geräte / Sensoren ließen den Proband anders agieren als sonst?

Frage: „Glauben Sie, dass das Ableiten der Signale so bewusst vom Probanden wahrgenommen wird, dass dies die Signale verändert?“

Kommunikation

Haben Sie Kommunikation im Versuch ignoriert oder bewusst später gelesen / reagiert?

ja nein

Wenn „ja“, aus welchem Grund:

- Verwendete Kommunikationsmittel ungewohnt:
- Konzentration auf Bearbeitung der Aufgaben (Priorisierung):
- Anderer: _____

Versuchsziel

Versuchsperson nach dem vermuteten Versuchsziel befragen

(„Wir möchten gerne noch einmal nachfragen, ob Ihnen klar war, worum es in dem Experiment ging.“)

¹¹⁹ Angaben zu Abschnitt „Physiologische Signale“ von VL1 (Tina Meißner)

Aufdeckung des „wahren“ Versuchsziels

(„Das tatsächliche Ziel des Experiments bestand darin, zu erforschen, inwieweit Kommunikation und gängige Kommunikationsmittel in einer Arbeitssituation zwangsläufig Ablenkung erzeugen und inwieweit die Arbeitsleistung sich verändert.“)

Wurde das wahre Versuchsziel vorher erkannt: ja nein

Wenn „ja“, wurde es:

- vor dem Experiment durch andere erkannt: ja
- während der Durchführung erkannt: ja, in der KG
- ja, in der EG
- Anders erkannt: _____

Feedback

Weitere Bemerkungen / Kritik der Versuchsperson:

(„Haben Sie noch Anmerkungen oder Kritikpunkte für uns? Wir freuen uns über jedes Feedback.“)

Instruktionen an die Versuchsperson

- VP auf das Recht hinweisen, die Auswertung ihrer Daten nicht zu gestatten
- Schweigepflicht bis Herbst/Winter 2012
- Expliziten Dank aussprechen für Teilnahme am Experiment und Beantwortung der Fragebögen

Verabschiedung und Bedanken

„Vielen Dank, dass Sie sich die Zeit genommen haben.“

ANHANG D. ERGEBNISSE

An dieser Stelle werden die Ergebnisse aus den Vor- und Hauptexperimenten wertungsfrei dargestellt.

D.1 VERSUCHSPERSONEN-AUFGABEN

Tab. 26 Ergebnisse der Aufgaben (MWT-A/B, IQ-1/2) aus den Vor- und Hauptexperimenten für alle Durchgänge 1 und 2

Vorexperimente											
VP	Durchgang	MWT-A			MWT-B			IQ-1		IQ-2	
		AA (37)	AR	IQ-Werte	AA (37)	AR	IQ-Wert	AA (36)	AR	AA (36)	AR
Kontrollgruppe											
1	2				37	31	112			5	2
2	1				37	25	95	5	5		
3	2	37	27	100						24	22
4	1	37	32	118				8	8		
Experimentalgruppe											
1	1	37	28	101				0	0		
2	2	37	25	95						7	5
3	1				37	X	X	7	X		
4	2				37	30	107			21	19
Hauptexperimente											
Kontrollgruppe											
1	X	X	X	X	X	X	X	X	X	X	X
2	X	X	X	X	X	X	X	X	X	X	X
3	2				37	28	101			8	6
4	1	37	25	95				10	9		
5	X	X	X	X	X	X	X	X	X	X	X
6	1	37	28	101				12	11		
7	2				37	32	118			12	11
8	1	37	28	101				10	7		
9	2				37	27	100			26	19
Experimentalgruppe											
1	X	X	X	X	X	X	X	X	X	X	X
2	X	X	X	X	X	X	X	X	X	X	X
3	1	37	26	97				9	8		
4	2				37	25	95			24	23
5	X	X	X	X	X	X	X	X	X	X	X
6	2				37	28	101			15	9
7	1	37	31	112				9	8		
8	2				37	32	118			12	7
9	1	37	29	104				6	1		

Legende:

AA (x): Anzahl bearbeiteter Aufgaben (x - Gesamtanzahl)

AR: Anzahl richtiger Lösungen (von Gesamtanzahl aller bearbeiteten Aufgaben)

VP: Versuchsperson

X: Ausfall des Datensatzes (Ausfall der Versuchsperson, technischer Ausfall, menschliches Versagen)

D.2 TECHNIK UND MATERIAL

Tab. 27 Übersicht über Materialien und Geräte in allen Versuchen (Durchgang 1 und 2) der Vor- und Hauptexperimente

VP	Vorexperimente				Hauptexperimente								
	1	2	3	4	1	2	3	4	5	6	7	8	9
Eyetracking	D1, D2	D1, D2	D1, D2	D1, D2	X	X	D1, D2	D1, D2	X	D1, D2	D1, D2	D1, D2	D1, D2
EOG-System	nein	nein	D1, D2	D1, D2	X	X	D1, D2	D1, D2	X	D1, D2	D1, D2	D1, D2	D1, D2
Thermographie- kamera	D1, D2	D1, D2	D1, D2	D1, D2	X	X	D1, D2	D1, D2	X	D1, D2	D1, D2	D1, D2	D1, D2
ProComp	D1, D2	D1, D2	D1, D2	D1, D2	X	X	D1, D2	D1, D2	X	D2	D1, D2	D1, D2	D1, D2
Morae	nein	nein	D1, D2	nein	X	X	D1, D2	D1, D2	X	D1, D2	D2	D1, D2	D1, D2
Videokamera	D1, D2	D2	nein	nein	X	X	nein	nein	X	nein	nein	nein	nein
TeamViewer	D1, D2	D1, D2	D2	D1, D2	X	X	D1, D2	D1, D2	X	D1, D2	D1 (ersten 5 Minuten), D2	D1, D2	D1, D2
Skype (Telefon, Chat)	D1, D2	D1, D2	D1, D2	D1, D2	X	X	D1, D2	D1, D2	X	D1, D2	D1, D2	D1, D2	D1, D2
Email: thunderbird	nein	nein	D1	nein	X	X	nein	nein	X	nein	nein	nein	nein
Email: web.de	D1, D2	D1, D2	D2	D1, D2	X	X	D1, D2	D1, D2	X	D1, D2	D1, D2	D1, D2	D1, D2
Email: gmx.de	D1, D2	D1, D2	D1, D2	D1, D2	X	X	D1, D2	D1, D2	X	D1, D2	D1, D2	D1, D2	D1, D2
Fragebogen 1 (Stimmung)	D1	D1	D1	D1	X	X	D1	D1	X	D1	D1	D1	D1
Fragebogen 2 (Versuch)	D2	D2	D2	D2	X	X	D2	D2	X	D2	D2	D2	D2
VP-Aufgaben (MWT, IQ)	D1, D2	D1, D2	D1, D2	D1, D2	X	X	D1, D2	D1, D2	X	D1, D2	D1, D2	D1, D2	D1, D2

Legende:

D1: Durchgang 1

D2: Durchgang 2

VP: Versuchsperson

X: Ausfall der Versuchsperson

D.3 STIMMUNGSFRAGEBOGEN

Tab. 28 Ergebnisse des Stimmungsfragebogens für alle Vor- und Hauptexperimente

VP	Vorexperiment				Hauptexperiment									
	1	2	3	4	1	2	3	4	5	6	7	8	9	
1	4	3	2	4		X	X	3	3	X	3	3	3	4
2	4	4	3	2		X	X	2	6	X	6	3	5	5
3	5	5	4	3		X	X	2	6	X	6	3	5	6
4	5	4	4	3		X	X	3	4	X	6	3	4	5
5	5	6	6	6		X	X	4	0	X	6	6	6	6
6	5	4	6	6		X	X	4	5	X	6	5	6	6
7	5	5	3	5		X	X	5	5	X	6	5	6	5
8	5	4	4	4		X	X	4	4	X	6	5	2	5
9	5	5	4	3		X	X	4	6	X	6	4	5	5
10	5	6	6	2		X	X	4	0	X	6	4	6	5
Gesamt	4,8	4,6	4,2	3,8		X	X	3,5	3,9	X	5,7	4,1	4,8	5,2
					Lese-Rechtschreib-Schwäche	X	X	X	nein	X	nein	nein	nein	nein

Auswertung:

Addition aller Zahlenwerte, Division der Summe durch 10 (Mindestwert: 4,0)

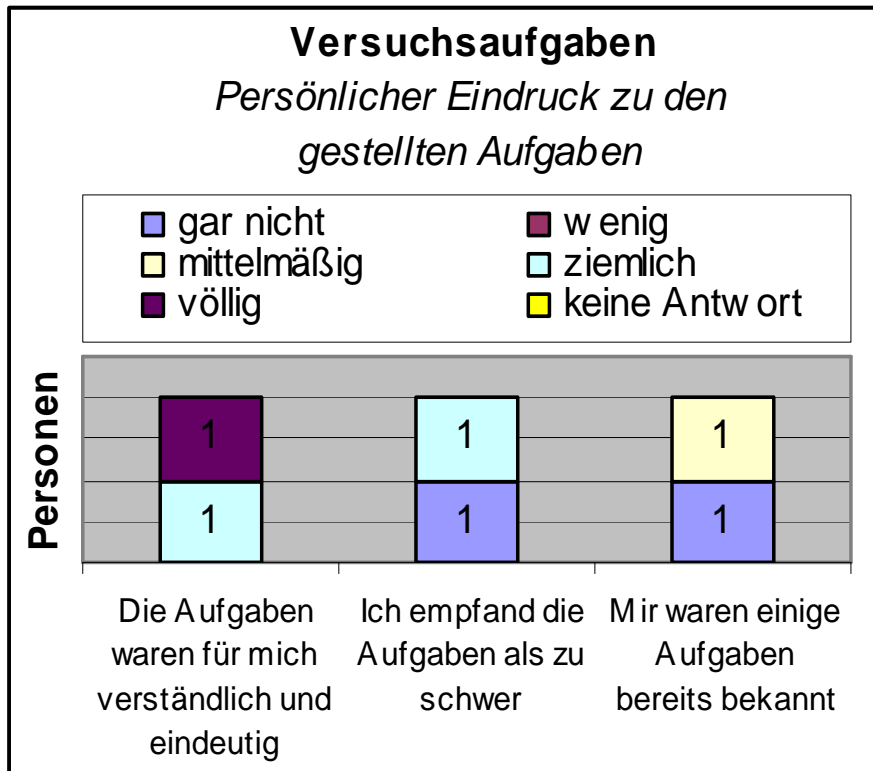


Abb. 22 Versuchsfragebogen (Vorexperiment) - Versuchsaufgaben (VP1, VP2)

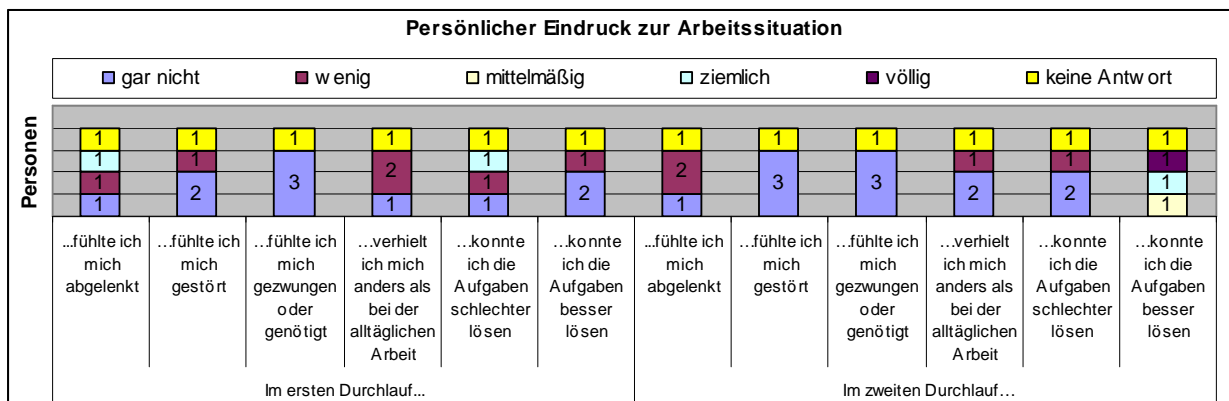


Abb. 23 Versuchsaufgaben (Vorexperiment) - Persönlicher Eindruck zur Arbeitssituation

Ergebnisse des Versuchsfragebogens im Hauptexperiment

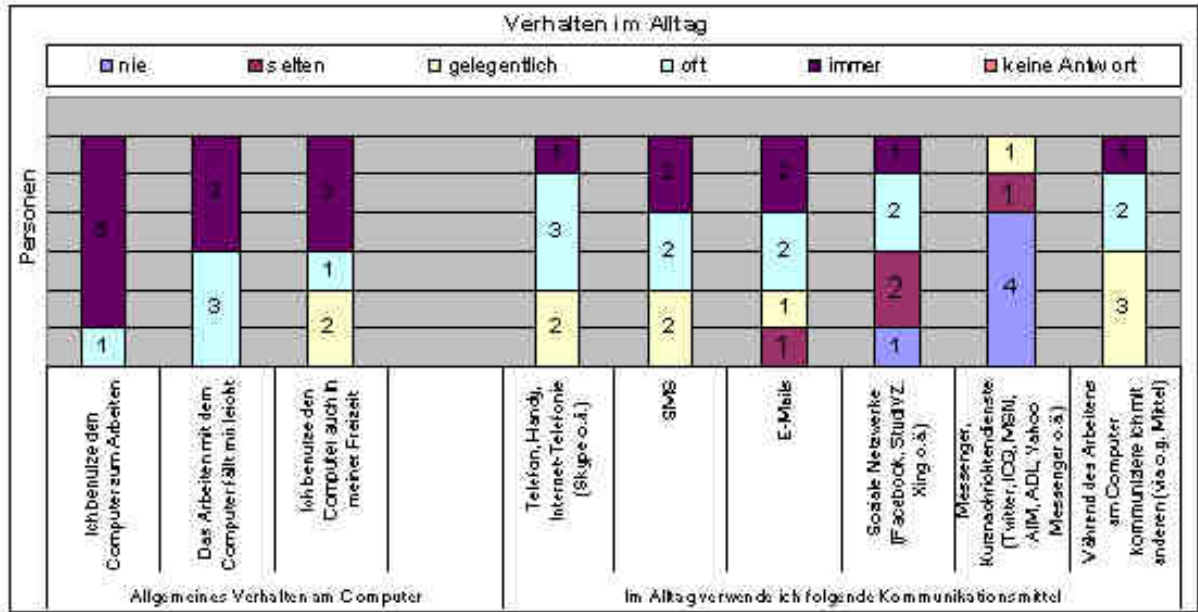


Abb. 24 Versuchsfragebogen (Hauptexperiment) - Verhalten im Alltag

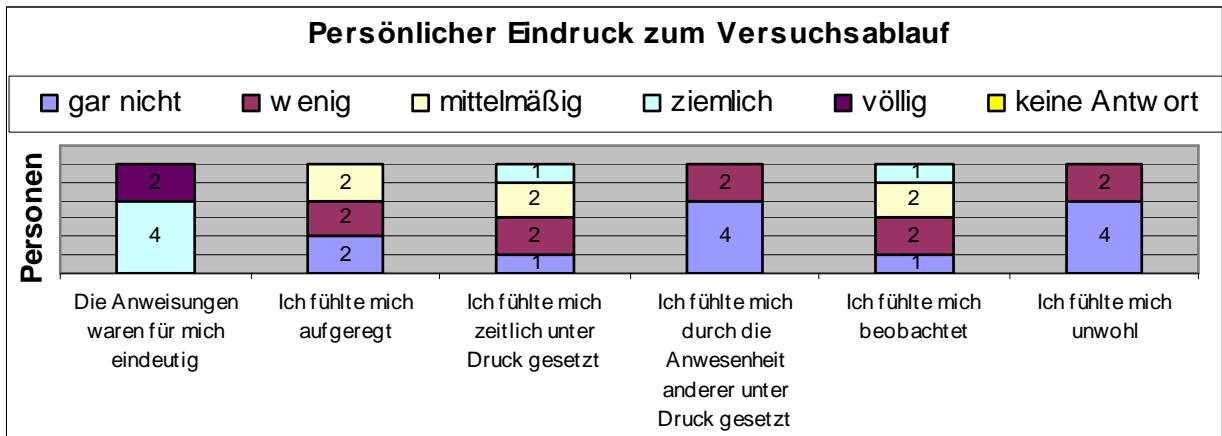


Abb. 25 Versuchsfragebogen (Hauptexperiment) - Persönlicher Eindruck zum Versuchsablauf

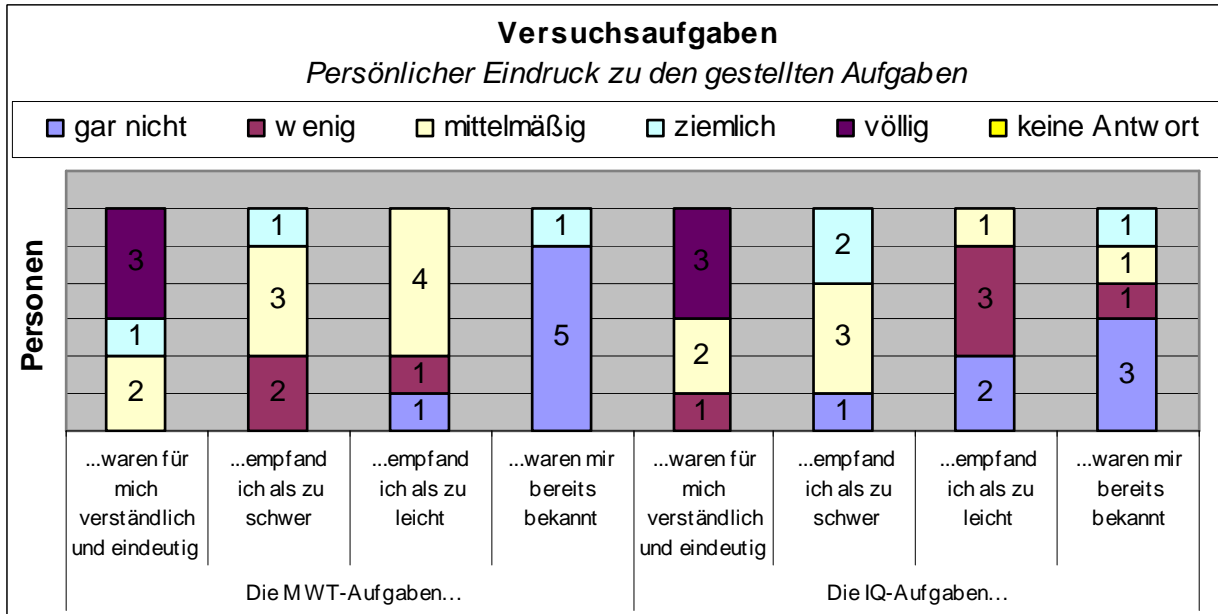


Abb. 26 Versuchsfragebogen (Hauptexperiment) - Versuchsaufgaben

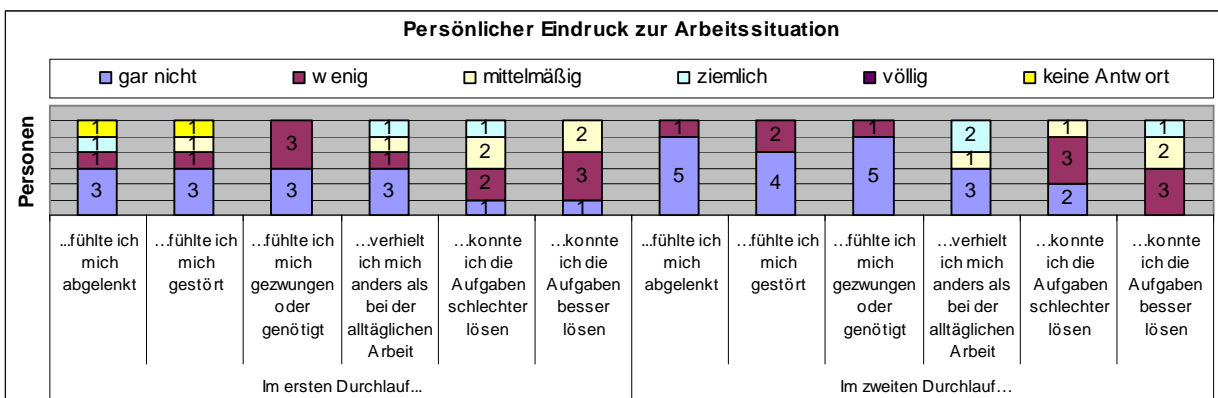


Abb. 27 Versuchsfragebogen (Hauptexperiment) - Persönlicher Eindruck zur Arbeitssituation (nach Durchlauf geordnet)

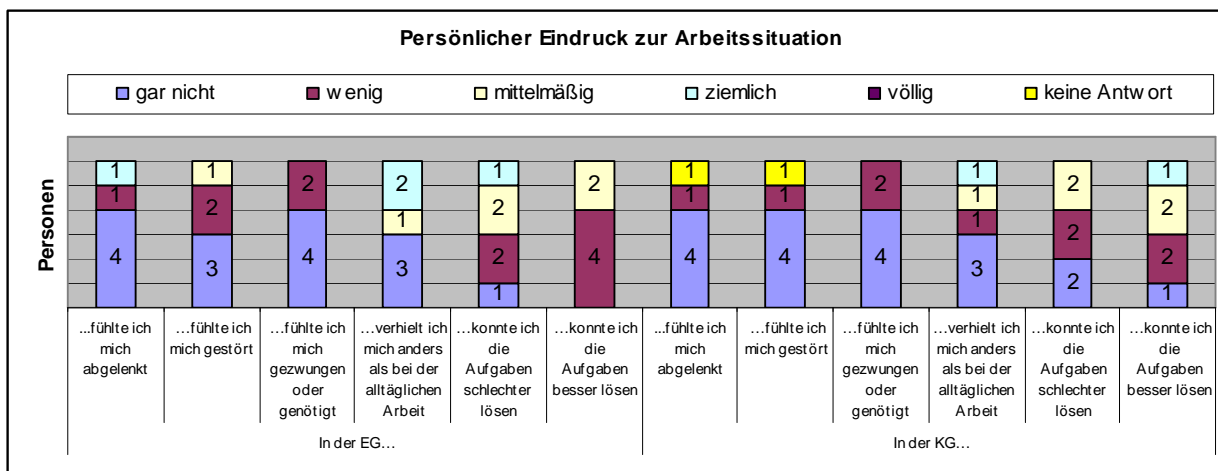


Abb. 28 Versuchsfragebogen (Hauptexperiment) - Persönlicher Eindruck zur Arbeitssituation (nach Gruppe geordnet)

D.5 VERSUCHSPERSONEN

Tab. 29 Demographische Merkmale der Versuchspersonen im Vor- und Hauptexperiment

VP	Vorexperimente				Hauptexperimente								
	1	2	3	4	1	2	3	4	5	6	7	8	9
Geschlecht	m	m	m	w	X	X	m	w	X	w	w	m	w
Alter	51	20	26	27	X	X	30	26	X	39	34	27	47
Tätigkeit	M	S	S	S	X	X	A	A	X	A	A	A	A
Muttersprache	d	d	d	d	X	X	d	d	X	d	d	d	d
Wochenarbeitszeit					X	X	40	40	X	31	40	40	40
Anzahl im Büro verbrachter Stunden (pro Woche)					X	X	40	40-50	X	6h 10 Min	45	40	40
Abschluss	1	2	2	1	X	X	1	3	X	4, 5	1	2	2, 1
Branche	FHB	FHB	FHB	FHB	X	X	IT, Web	Design, Druck	X	Design, Druck	Geodaten, Satellitenbilder	IT, Web	Behörde (Bauen, Liegenschaften)

Legende:

- 1: abgeschlossenes Studium
- 2: Abitur oder andere Hochschulzugangsberechtigung
- 3: Meister

- A: Angestellte/r
- d: deutsch
- FHB: Fachhochschule Brandenburg
- m: männlich
- M: MitarbeitIn (Fachhochschule Brandenburg)
- S: StudentIn
- VP: Versuchsperson
- w: weiblich
- X: Versuchspersonen-Ausfall