ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Breyer, Tina; Gutschmidt, Anne; Nerdinger, Friedemann W.

Research Report

Expertenfeedback im Notfall-Management-Training: Eine experimentelle Studie

Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, No. 11

Provided in Cooperation with:
Seniorprofessur für Wirtschafts- und Organisationspsychologie, Universität Rostock

Suggested Citation: Breyer, Tina; Gutschmidt, Anne; Nerdinger, Friedemann W. (2013) : Expertenfeedback im Notfall-Management-Training: Eine experimentelle Studie, Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, No. 11, Universität Rostock, Wirtschafts- und Sozialwissenschaftliche Fakultät, Lehrstuhl für ABWL: Wirtschafts- und Organisationspsychologie, Rostock,
https://doi.org/10.18453/rosdok_id00002251

This Version is available at:
https://hdl.handle.net/10419/232104

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

UNIVERSITÄT ROSTOCK
Lehrstuhl für Wirtschafts- und Organisationspsychologie

Tina Breyer, Anne Gutschmidt & Friedemann W. Nerdinger

Expertenfeedback im Notfall-Management-Training. Eine experimentelle Studie.

Rostocker Beiträge
zur Wirtschafts- und Organisationspsychologie
Nr. 11

Tina Breyer, Anne Gutschmidt & Friedemann W. Nerdinger

Expertenfeedback im Notfall-Management-Training. Eine experimentelle Studie.

KNoTen 2.0

Herausgeber: Lehrstuhl für ABWL: Wirtschafts- und Organisationspsychologie der Universität Rostock

Kurztitel: Breyer, T., Gutschmidt, A. & Nerdinger, F. W. (2013) Expertenfeedback im Notfall-Management-Training. Eine experimentelle Studie. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 11*. Universität Rostock

Druck: Druckerei Hahn GmbH

Autoren: Tina Breyer (tina.breyer@uni-rostock.de)
Anne Gutschmidt (anne.gutschmidt@uni-rostock.de)
Friedemann W. Nerdinger

Universität Rostock
Lehrstuhl für ABWL: Wirtschafts- und Organisationspsychologie
Ulmenstr. 69
18057 Rostock

© Universität Rostock, Wirtschafts- und Sozialwissenschaftliche Fakultät, Lehrstuhl für ABWL: Wirtschafts- und Organisationspsychologie, 2013.

Hinweis zum Projekt KNoTen 2.0:

„KNoTen 2.0“ ist ein Verbundprojekt, das aus Mitteln des Bundesministeriums für Bildung und Forschung im Rahmen des Programms „Förderung von Vorhaben zur Weiterentwicklung und zum Einsatz von Web 2.0 Technologien in der beruflichen Qualifizierung“ sowie durch den Europäischen Sozialfonds (ESF) gefördert wird. Ziel des Verbundprojektes ist es, unter Einsatz moderner Kommunikationstechnologien insbesondere die Aus- und Weiterbildung im Bereich der Schiffssicherheit und des Notfall-Managements in der Schifffahrt zu verbessern. So sollen im Rahmen des Projektes auf der Basis vorhandener Lehrgänge neuartige Qualifikationsmöglichkeiten entwickelt werden, die den in den letzten Jahren stark gestiegenen Anforderungen der maritimen Branche Rechnung tragen.

Weitere Informationen zum Projekt und Ansprechpartner unter www.knoten-2null.de.

Das dieser Publikation zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) und durch den Europäischen Sozialfonds (ESF) unter dem Förderkennzeichen 01PF08021C gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
ESF Europäischer Sozialfonds für Deutschland
EUROPÄISCHE UNION
DLR Deutsches Zentrum für Luft- und Raumfahrt e.V. Projektträger im DLR

https://doi.org/10.18453/rosdok_id00002251

Inhalt

Abbildungsverzeichnis IV
Tabellenverzeichnis V
Abkürzungs- und Symbolverzeichnis VI
1 Einleitung 1
1.1 Die Ausgangslage in der Seeschifffahrt 1
1.2 Einordnung des Forschungsprojektes KNoTen 2.0 3
1.3 Anliegen und Gang der Arbeit 4
2 Zum Projekt KNoTen 2.0 – Komplexes Notfall-Management-Training gekoppelt mit Web 2.0 Technologien 5
2.1 Zum Verbundprojekt KNoTen 2.0 5
2.2 Die KNoTen-Station 6
3 Theoretischer und empirischer Hintergrund zur Ableitung der Forschungsfragen 9
3.1 Feedback im Trainingsprozess 9
3.1.1 Bedeutung von Feedback in Trainings- und Lernprozessen 9
3.1.2 Expertenfeedback im Rahmen des KNoTen 2.0-Projektes 10
3.1.3 Zur allgemeinen Wirkung von Feedback 11
3.2 Wahrnehmungs- und Konzentrationsfähigkeit als spezifische Anforderungen der KNoTen-Station 13
3.3 Forschungsfragen - Mögliche Auswirkungen zusätzlicher Funktionalitäten 16
3.3.1 Die Sichtbarkeit von Feedback-Kommentaren anderer Experten 17
3.3.2 Erhöhung der Wahrnehmungswahrscheinlichkeit anderer Expertenkommentare 23
3.3.3 Die Möglichkeit des Pausierens während der Beobachtung 23
3.3.4 Die Möglichkeit, einen auditiven Fokus zu setzen 27
3.4 Ergänzende Forschungsfragen 30
4 Methodik 31
4.1 Forschungsdesign 31
4.2 Grundaufbau der Experimental-Software 32
4.3 Besonderheiten der Varianten des Grundaufbaus 36
4.4 Indikatoren zur Leistungsmessung und zur Erhebung subjektiver Eindrücke 38
4.4.1 Indikatoren der Leistung 38
4.4.2 Erhebung der subjektiven Eindrücke 42
4.5 Untersuchungsablauf und weitere Erhebungsinstrumente 43

4.6 Methoden der statistischen Auswertung 47
4.7 Qualitative Auswertung 48
5 Ergebnisse 48
5.1 Zur Stichprobe 48
5.1.1 Soziodemografie in den Versuchsgruppen 48
5.1.2 Vergleichbarkeit der Versuchsgruppen 50
5.2 Ergebnisse der statistischen Auswertung 53
5.2.1 Anzahl der geschriebenen Kommentare 53
5.2.2 Anzahl der einbezogenen Videos 54
5.2.3 Ausführlichkeit der Kommentare 55
5.2.4 Tippgeschwindigkeit 56
5.2.5 Gesamtdauer 57
5.2.6 Richtigkeit der Kommentare 58
5.2.7 Verständlichkeit 58
5.2.8 Zusammenfassung 59
5.3 Ergebnisse aus den Interviewdaten 60
5.3.1 Subjektive Schwierigkeit 60
5.3.2 Gefühlslage während der Aufgabenbearbeitung 61
5.3.3 Verbesserungsvorschläge der Probanden 71
5.3.4 Zusammenfassung 78
6 Diskussion 79
6.1 Zusammenfassung 79
6.2 Bedeutung der Ergebnisse für das Forschungsprojekt KNoTen 2.0 81
6.2.1 „Expertenkommentare“ 81
6.2.2 „Expertenkommentare mit Signal“ 82
6.2.3 „Pausieren“ 84
6.2.4 „Ton variierbar“ 85
6.2.5 Weitere Handlungsempfehlungen 86
6.3 Limitationen der Untersuchung 87
Literaturverzeichnis 90
Anhang 97

Abbildungsverzeichnis

Abbildung 1: Grundprinzip der KNoTen-Station. 8
Abbildung 2: Der Prozess der Feedback-Wirkung 11
Abbildung 3: Informationsverarbeitung bei der Arbeit mit der KNoTen-Station..... 14
Abbildung 4: Dialogfeld für die Grundeinstellungen der Experimentalsoftware. 33
Abbildung 5: Anleitung für die Probanden der Kontrollgruppe. 34
Abbildung 6: Screenshot des Bearbeitungsbildschirms in der Grundvariante. 35
Abbildung 7: Screenshot vom suggerierten Hinzuschalten weiterer Kommentatoren. 37
Abbildung 8: Screenshot der Tipptest- Software. 44
Abbildung 9: Ausschnitt der deutschen Version des Konzentrationstestmaterials 45
Abbildung 10: Beispiel für eine Multiple Choice-Frage im Konzentrationstest. 46
Abbildung 11: Geschlechterverteilung in den Experimentalgruppen. 49
Abbildung 12: Bildungsabschlüsse in den Versuchsgruppen 50
Abbildung 13: Führerscheinbesitz in den Versuchsgruppen..... 51
Abbildung 14: Anzahl der Kommentare - Mittelwerte der Gruppen. 54
Abbildung 15: Anzahl einbezogener Videos - Mittelwerte der Gruppen. 55
Abbildung 16: Ausführlichkeit der Kommentare - Mittelwerte der Gruppen..... 56
Abbildung 17: Tippgeschwindigkeit - Mittelwert der Gruppen..... 56
Abbildung 18: Versuchsdauer in Sekunden - Mittelwerte der Gruppen. 57
Abbildung 19: Anzahl richtig erkannter Verstöße - Mittelwerte der Gruppen. 58
Abbildung 20: Verständlichkeit der Kommentare - Mittelwerte der Gruppen 59
Abbildung 21: Auftretende Gefühle in den Versuchsgruppen. 70

Tabellenverzeichnis

Tabelle 1: Übersicht zur Manipulation der unabhängigen Variablen 36
Tabelle 2: Anzahl der Probanden in den der Kontrollgruppe und in den Versuchsgruppen.... 49
Tabelle 3: Alter der Probanden in Jahren nach Experimentalgruppen..... 50
Tabelle 4: Mittelwerte und Standardabweichungen der Variablen zu Erfahrung mit Computern und dem Wissen über Straßenverkehr. 52
Tabelle 5: Deskriptive Statistiken für die Variable Aufgabenschwierigkeit..... 61
Tabelle 6: Anzahl der Statements zur Gefühlslage während der Nutzung Experimentalsoftware. 61
Tabelle 7: Subkategorien zur Kategorie positives / neutrales Gefühl 62
Tabelle 8: Subkategorien zur Kategorie negatives Gefühl..... 63
Tabelle 9: Subkategorien zur Kategorie gestresst / überfordert 64
Tabelle 10: Weitere Gefühlskategorien im Überblick 65
Tabelle 11: Prozentualer Anteil (%) und absolute Häufigkeiten (H) des Vorkommens der Gefühlskategorien in den Versuchsgruppen..... 66
Tabelle 12: Anzahl der genannten Verbesserungsvorschläge nach Varianten..... 71
Tabelle 13: Verbesserungsvorschläge und ihre Häufigkeiten 72
Tabelle 14: Verbesserungsvorschläge zur Videoanzeige 73
Tabelle 15: Verbesserungsvorschläge zum Kommentieren 75
Tabelle 16: Weitere Verbesserungsvorschläge 77

Anhang

Anhang 1: Interviewleitfaden und Fragebogen 97
Anhang 2: Signifikanzwerte des Shapiro-Wilk-Tests auf Normalverteilung der Variablen zu Erfahrungen mit PC und Straßenverkehr 98
Anhang 3: Signifikanztests zu Unterschieden hinsichtlich Erfahrungen mit PCs und Straßenverkehr..... 98
Anhang 4: Deskriptive Statistiken der abhängigen Variablen je Versuchsgruppe 100
Anhang 5: Tests auf Normalverteilung für die abhängigen Variablen 101
Anhang 6: Ergebnisse der Signifikanztests der abhängigen Variablen..... 102
Anhang 7: Test auf Normalverteilung der Variable Aufgabenschwierigkeit..... 103
Anhang 8: U-Tests zur erlebten Aufgabenschwierigkeit 103

Abkürzungs- und Symbolverzeichnis

AV ……	abhängige Variable
BMBF …….	Bundesministerium für Bildung und Forschung
CBT …….	Computer based Training
df …….	Freiheitsgrade
ESF …….	Europäische Sozialfonds
ISV …….	Institut für Schiffssicherheit und Sicherheitstechnik e.V.
H …….	Absolute Häufigkeit
M …….	Mittelwert in den Versuchsgruppen
MARSIG ……	Ingenieurgesellschaft für Maritime Sicherheitstechnik und Management mbH
N …….	Stichprobengröße
p …….	Probability (Irrtumswahrscheinlichkeit)
s …….	Standardabweichung
STCW95 ……..	Standards of Training, Certification and Watchkeeping for Seafarers 1995
UKZ …….	Ultrakurzzeitspeicher

1 Einleitung

1.1 Die Ausgangslage in der Seeschifffahrt

Etwa 90 Prozent der weltweit transportierten Güter werden über Seewege befördert (Oldenburg & Jensen, 2012). Durch die hohe Bedeutung der Seeschifffahrt im Zusammenspiel mit dem stetig wachsenden ökonomischen Druck auf die Reeder haben sich auch die Arbeits- und Lebensbedingungen für Seeleute verändert. Obwohl durch neue technische Entwicklungen starke, rein körperliche Belastungen reduziert werden konnten, stellt der normale Alltag an Bord von Schiffen eine immense Belastungs- und Stressquelle dar. Immer kleiner werdende Besatzungen, kürzere Verweildauern in Häfen und die zunehmende Gefahr durch Piraterie und Entführungen prägen das Leben an Bord. Zu den physischen und psychischen Belastungen zählen unnatürliche und unregelmäßige Schlaf- und Wachzeiten durch die einzuhaltenden Wachpläne, lange Arbeitstage, die Trennung vom familiären Umfeld, die Exposition gegenüber extremen Temperaturen und Luftfeuchtigkeiten sowie gegenüber Lärm und der Schiffsvibration (Oldenburg & Jensen, 2012; Oldenburg, Jensen, Latza & Baur, 2008). Diese Bedingungen wirken sich auf die physische und psychische Leistungsfähigkeit der Besatzungen aus und werden als Ursache für eine Vielzahl der Unfälle gesehen, die dem menschlichen Versagen zugeschrieben werden (Emad & Roth, 2008). Allein in deutschen Gewässern wurden im Berichtszeitraum 2011 115 Seeunfälle in registriert. Davon waren zwar 89 als „weniger schwer“ einzustufen, acht der Unfälle zählten jedoch zu den „sehr schweren“, bei denen Menschen zu Tode kamen oder die Vorfälle zu einem Totalverlust der Technik führten. Darüber hinaus wurden 18 „schwere“ Unfälle registriert, bei denen immenser Sachschaden entstand (BSU, 2012). Auch wenn die Bundesstelle für Seeunfalluntersuchungen (BSU) im Jahr 2011 keine Erhöhung der Unfallzahlen auf See verzeichnete machen diese Zahlen deutlich, wie groß die Gefahr von Seeunfällen ist.

Zwei weitere Faktoren, neben den bereits genannten, beeinträchtigen die Sicherheit in der Seeschifffahrt zusätzlich. Zum einen führt die Multinationalität der Besatzungen sowohl zu sozialkulturellen als auch zu rein kommunikativen Barrieren (Oldenburg & Jensen, 2012; Gerstenberger, 2002). So führt eine erschwerte Verständigung der Crewmitglieder untereinander zum einen zu erhöhtem Stress durch die erlebte Isolation der einzelnen Besatzungsmitglieder (Oldenburg & Jensen, 2012) und verhindert zum anderen insbesondere in Notsituationen eine schnelle und effektive Koordination (Hahne, Tober & Brühe, 1997).

Des Weiteren ist seit geraumer Zeit der Trend erkennbar, dass die Verweildauer der Seeleute in ihrem Beruf stark abnimmt. So hat sich die Anzahl der Jahre, in denen bspw. Schiffsoffiziere

tatsächlich zur See fahren, von zuvor durchschnittlich sieben Jahren auf drei bis vier Jahre beinahe halbiert (Gerstenberger, 2002). Auch gerade die Crewmitglieder unterer Hierarchieebenen, die aus Schwellen- oder Entwicklungsländern stammen, sind hiervon betroffen, wenn auch – im Vergleich zu den höheren Offizieren – nicht freiwillig. Die Heuerpolitik ist in diesem Personalbereich so weit, dass die Dauer einer Anstellung zum Teil gerade einmal 12 oder sogar weniger Wochen beträgt (Oldenburg & Jensen, 2012). Nach dieser Beschäftigungszeit kehrt nur ein Teil der Seeleute wieder zurück in ein neues Anstellungsverhältnis auf See und die Schiffsbetreiber verlieren mit jedem Seemann immer wieder einen Teil an Erfahrung und Know-how. Dies hat zur Folge, dass sich die Schifffahrt permanent auf der Suche nach Personal befindet und sich der Aufgabe gegenüber sieht, für neue und unerfahrene Berufseinsteiger sowohl wirkungsvolle aber auch effiziente Wege der Aus- und Weiterbildung vorzusehen, um den Verlust an Fach- und Erfahrungswissen zu kompensieren und so den regulären Betrieb sowie die Sicherheit an Bord gewährleisten zu können.

Die Förderung der Sicherheit mittels Schulungen und hierbei die Vorbereitung auf Notfälle stellt eine besondere Herausforderung sowohl inhaltlicher als auch organisatorischer Art dar. Insbesondere Seenotlagen, unabhängig von ihrer Entstehungsursache, sind häufig geprägt von komplexen, sich schnell verändernden Bedingungen sowie von unvollständigen und nur schwierig zu bewertenden Informationen. Sie sind daher in der Regel nur sehr schwer beherrschbar (Hahne et al., 1997).

In Notfallsituationen, also in „zeitlich eng begrenzten Extremsituationen verdrängen Angst und Stress […] rationales Handeln“ (Frey, Savage, Schmidt & Torgler, 2011, S. 237). Obwohl selbst betroffen, müssen die Entscheidungsträger und Akteure in der Notfallsituation, nämlich die verantwortlichen Schiffsoffiziere und das Schiffspersonal, sowohl für die eigene Sicherheit aber auch für die Sicherheit von Personen ohne aktive Rolle im Notfallgeschehen – v.a. für Passagiere – sorgen. Hierfür ist Wissen über die einzusetzende Technik sowie über die Abfolge und Koordination bestimmter Handlungen notwendig. Darüber hinaus spielt aber auch psychische Leistungsfähigkeit eine wichtige Rolle für das Treffen von Entscheidungen, die Handlungsplanung und Handlungsausführung (Hahne et al., 1997). Um demnach auch in Notfallsituationen handlungsfähig zu bleiben, kommt es auf das Zusammenspiel verschiedener Einzelfertigkeiten und -fähigkeiten (Können) sowie auf die notwendigen Kenntnisse an. Das erforderliches Wissen und Können stammen aus vielen unterschiedlichen Bereichen (z.B. Technik, Nautik, Psychologie, Kommunikation, Führung usw.), die alle Inhalt der Aus- und Weiterbildung sein sollten, um die Besatzungen umfassend zu befähigen Notfallsituationen zu bewältigen (Hahne, Baaske, Sedlaček & Schubert, 2002).

Einen der wichtigsten Teile der Ausbildung von Seeleuten stellen die Übungen und Trainings in realitätsnahen Umgebungen, z.B. an Bord von Schiffen oder in Simulatoren, dar (Emad & Roth, 2008). Diese haben den Vorteil nicht nur Wissen zu vermitteln, sondern dieses Wissen auch mit entsprechenden Handlungen zu verknüpfen, was den Anforderungen an die Aus- und Weiterbildung gemäß den international geforderten Standards entspricht (STCW95 - Standards of Training, Certification and Watchkeeping for Seafarers 1995), eher kompetenzbasierte Schulungen anzubieten (Emad & Roth, 2008). Dennoch wird an dieser Art der Ausbildung kritisiert, dass häufig nur kurze Handlungssequenzen trainiert werden und meist die entsprechende Überwachung und Kontrolle der Übungsdurchläufe durch Supervisoren fehlt (Emad & Roth, 2008).

Darüber hinaus erfordern Trainings in Simulatoren oder an Bord die physische Präsenz der Auszubildenden am Übungsort. Dies führt zu personellen Ausfällen im Bordbetrieb und häufig auch zu hohen Kosten für die Anreise zu den Trainingsorten. Schiffsbetreiber sind häufig nicht bereit, diese Kosten in Kauf zu nehmen, da sie zum einen schlichtweg finanzielle Einbußen bedeuten, zum anderen aber auch gefürchtet wird, dass gut ausgebildete Seeleute schnell durch Konkurrenten abgeworben werden könnten (Sampson, 2004).

Dennoch kann man auf diesen praktischen Teil der Ausbildung kaum verzichten und es müssen Möglichkeiten entwickelt werden, insbesondere das Training effizienter zu gestalten.

1.2 Einordnung des Forschungsprojektes KNoTen 2.0

Vor dem dargelegten Hintergrund versucht das vom Bundesministerium für Bildung und Forschung und dem europäischen Sozialfonds geförderte Forschungsprojekt „KNoTen 2.0 – Komplexes Notfall-Management-Training gekoppelt mit Web 2.0 Technologien“, Trainings in realitätsnahen Umgebungen zu verbessern. Der Teil der Projektarbeiten, über den im vorliegenden Arbeitspapier berichtet wird, beschäftigte sich mit der Entwicklung, Implementierung und Evaluation der sogenannten „KNoTen-Station“. Hierbei handelt es sich um eine Kombination aus Hard- und Software, die bei der Durchführung von komplexen Sicherheitstrainings in der Schifffahrt eingesetzt werden soll. Die KNoTen-Station soll es ermöglichen, die geforderte Supervision von Trainingsprogrammen durch Experten, die den Trainierenden Feedback zu ihren Leistungen im Training geben, zu realisieren. Um hierbei keine zusätzlichen Kosten zu verursachen, sind die Experten jedoch nicht physisch präsent, sondern verfolgen das Trainingsgeschehen in Echtzeit über das Internet und verfassen schriftliche Kommentare sobald ein Verhalten der Teilnehmer zu sehen ist, das Feedback erfordert. Die Kommentare werden zusammen mit

den Videos auf einem Server gespeichert und stehen direkt nach dem Training zur Nachbesprechung mit den Trainierenden zur Verfügung. Auf diese Weise kann zumindest die Ortsunabhängigkeit der erforderlichen Experten gewährleistet werden, was Kosten reduziert. Zudem erhalten die Trainierenden auf diesem Wege wertvolles Feedback, was neben der erforderlichen Feedback-Information quasi eine weitere Lerneinheit im Anschluss an das eigentliche Training beinhaltet (Musch, 1999).

1.3 Anliegen und Gang der Arbeit

Das Ziel der vorliegenden Arbeit besteht darin, durch qualitative und quantitative empirische Untersuchungen Anhaltspunkte darüber zu gewinnen, wie die im Forschungsprojekt KNoTen 2.0 entwickelte Hard- und Softwareeinheit „KNoTen-Station“ weiter optimiert werden könnte. Hierfür werden verschiedene zusätzliche und schon heute technisch realisierbare Funktionalitäten vorgestellt und aus psychologischer Sicht hinsichtlich ihrer möglichen positiven und negativen Auswirkungen auf die Usability durch Experten untersucht. Dabei geht es vorrangig um die Fragestellung, ob die Funktionalitäten insgesamt zu einem quantitativ und qualitativ besseren Trainingsfeedback führen, d.h. ob die zusätzlichen Features einen Gewinn für die KNoTen-Station darstellen und daher in Zukunft mit implementiert werden sollten.

Hierfür ist es zunächst notwendig, die Projektidee eingehend zu beschreiben, was in Kapitel 2 erfolgt. Im Anschluss daran werden die relevanten Bereiche der psychologischen Forschung sowie deren Befunde dargelegt, um auf diesem Weg ein Verständnis für die Bedeutung von Feedback im Trainingsprozess und die grundsätzlich möglichen Auswirkungen der zusätzlichen Funktionalitäten herzustellen (Kapitel 3). Hierbei wird deutlich, dass die aus praktischen Erwägungen und hauptsächlich wegen ihrer umgehenden technischen Realisierbarkeit entstandenen Ideen für zusätzliche Funktionalitäten zwar Vorteile haben, aber auch einige Nachteile daraus erwachsen können. Aus diesem Grund werden ausgehend von der Befundlage zunächst Forschungsfragen abgeleitet, die sich direkt auf die denkbaren Varianten des Grundaufbaus der KNoTen-Station beziehen. Diese werden im weiteren Verlauf der Arbeit systematisch mit Hilfe einer experimentellen Studie untersucht, um Aussagen darüber treffen zu können, ob eine Implementierung der Funktionalitäten sinnvoll insbesondere im Hinblick auf die Qualität des Feedbacks durch die Experten wäre.

Darüber hinaus soll aber auch untersucht werden, wie die Anwender der KNoTen-Station, in diesem Fall die Experten, die Handhabung der KNoTen-Station erleben. Dies ist von besonderer Bedeutung, um Experten auch langfristig für den Einsatz der Station zu gewinnen. Um Auf-

schluss hierüber zu erlangen, wurden jeweils im Anschluss an die Experimente Interviews mit den Probanden geführt, um die Eindrücke der Versuchspersonen während der Nutzung der Station zu ermitteln. Zusätzlich ist denkbar, dass die Probanden auf Basis ihrer persönlichen Erfahrungen mit der KNoTen-Station zusätzliche Verbesserungsideen entwickeln. Diese wurden ebenfalls im Rahmen der Interviews abgefragt (siehe ergänzende Forschungsfragen in Abschnitt 3.4).

Im vierten Kapitel der Arbeit werden die qualitativen und quantitativen Methoden der Datenerhebung und –auswertung dargestellt. Hierbei erfolgt ebenfalls die Beschreibung des Grundaufbaus der KNoTen-Station sowie der zu testenden Funktionalitäten. Die ermittelten Ergebnisse sowie die Beschreibung der Stichprobe und die Überprüfung der Vergleichbarkeit der Experimentalgruppen werden im Kapitel 5 vorgestellt. Das abschließende Kapitel 6 enthält die Zusammenfassung der Ergebnisse sowie deren Diskussion hinsichtlich der Limitationen der Studie, der sich ergebenden Implikationen für das KNoTen 2.0-Projekt und für die weitere Forschung.

Bevor im folgenden Abschnitt die KNoTen-Station erläutert wird, soll zunächst das Projekt „KNoTen 2.0“ vorgestellt werden.

2 Zum Projekt KNoTen 2.0 – Komplexes Notfall-Management-Training gekoppelt mit Web 2.0 Technologien

2.1 Zum Verbundprojekt KNoTen 2.0

Im vom Bundesministerium für Bildung und Forschung (BMBF) und dem Europäischen Sozialfonds (ESF) geförderten Forschungsverbund KNoTen 2.0 waren drei Verbundpartner tätig: das Institut für Schiffssicherheit und Sicherheitstechnik e.V. (ISV), ein gemeinnütziger außeruniversitärer Verein mit dem Zweck der Förderung von Wissenschaft und Forschung, Aus- und Fortbildung auf den Gebieten Sicherheitstechnik und Schiffssicherheit (ISV, 2013); die Firma MARSIG (Ingenieurgesellschaft für Maritime Sicherheitstechnik und Management mbH) sowie die Universität Rostock, hier im Speziellen der Lehrstuhl für Wirtschafts- und Organisationspsychologie. Die Partner verfolgten das Ziel, auf Basis schon vorhandener Sicherheitslehrgänge neuartige Qualifikationsmöglichkeiten zu schaffen, indem moderne Computer- und Internettechnologien mit komplexen, realen Übungen von Handlungsabläufen in Notfallsituationen auf See kombiniert werden. Während der Projektlaufzeit von November 2010 bis April 2013 wurden hierbei mehrere ineinandergreifende Teilaufgaben bearbeitet. So wurde durch die Umstel-

lung bereits in der Sicherheitsausbildung eingesetzter Lehrgänge auf online verfügbare computerbasierte Trainingsprogramme, die auch moderne Web 2.0-Technologien wie bspw. Wikis und Chats beinhalten dazu beigetragen, die Flexibiltät und Ortsunabhängigkeit der Nutzer der Lehrangebote zu steigern. Hierfür sollte ein online verfügbares Lernmanagementsystem entwickelt werden, das neben dem für die Lehrgänge notwendigen Stoff auch aktuelle Informationen über Änderungen der für die Seefahrt wichtigen nationalen und internationalen Regularien beinhaltet.

Darüber hinaus umfasste die Projektarbeit die Entwicklung komplexer Notfallübungen, die mit Hilfe von Kameratechnik begleitet und in Echtzeit über das Internet zu Experten unterschiedlicher Fachbereiche übertragen werden können, die die so beobachteten Trainingsverläufe mit Feedbackkommentaren versehen. Auf diese Weise sollte mit den kommentierten Videos ein Instrument entstehen, das die Effektivität von Auswertungsgesprächen mit den Trainierenden, die sich an eine Trainingssequenz anschließen, erhöht. Beide Teilaufgaben tragen hierbei durch den Einsatz der modernen Technologien zu einer Flexibilisierung der Ausbildung bei.

Die im Folgenden beschriebene Studie fokussiert insbesondere die zweite Teilaufgabe und versucht zu klären, wie das entwickelte Auswertungsinstrument, die sogenannte KNoTen-Station, zur Effektivierung von Feedback im Trainingsprozess beitragen kann und welche weiterhin denkbaren Funktionalitäten bei der Implementierung der KNoTen-Station in Zukunft in Betracht gezogen werden sollten.

2.2 Die KNoTen-Station

Ziel des Projektes KNoTen 2.0 ist es, spezifische Probleme der Aus- und Weiterbildung in der maritimen Wirtschaft sukzessive abzubauen. Zu den größten Belastungen der Teilnehmer solcher Programme gehören die meist vorhandene Orts- sowie Zeitgebundenheit von Trainingsprogrammen und Übungen. Neben der Erstellung von computerbasierten Trainingsprogrammen (CBTs), die für das Selbststudium der Seeleute einen ersten Schritt in Richtung Ortsunabhängigkeit gehen, bleiben Realübungen[1] jedoch unabdingbar. Mit ihrer Hilfe werden bestimmte motorische Fertigkeiten, aber auch die Zusammenarbeit mit anderen Seeleuten erlernt und weiterentwickelt. Hierfür müssen die Aus- und Weiterzubildenden an einen festgelegten Übungsort reisen. Zusätzlich ist es häufig erforderlich, dass Experten für die jeweiligen Übungen am Trainingsort anwesend sind, die entweder das Training leiten oder aber, wie im KNoTen-Projekt

[1] Unter Realübungen werden im Projekt KNoTen 2.0 solche Übungen verstanden, in denen Lernende bestimmte Abläufe in einer möglichst realitätsnahen Umgebung trainieren können. Es handelt sich demnach um Vorortübungen, zu denen die Teilnehmer physisch anwesend sein müssen, aber dennoch nicht um kognitive Lehreinheiten in Unterrichtsform.

vorgesehen, das Trainingsgeschehen beobachten, um den Trainingsteilnehmern im Anschluss ein Feedback geben zu können. In diesem Feedback erfahren die Teilnehmer, in welchen Trainingsbereichen ihre Leistungen besonders gut waren und in welchen Punkten Verbesserungsbedarf besteht.

Bei Seenotfällen, die im Rahmen des Forschungsprojektes das zentrale Trainingsfeld darstellen, ist per se die Gefahr für Menschenleben, Natur, aber auch die Gefahr des Verlustes sehr teurer Technik besonders groß. Daraus ergeben sich Anforderungen an das Schiffspersonal, die im Rahmen der Realübungen trainiert werden sollen. Diese Anforderungen sind sowohl nautischer und technischer, aber in erster Linie auch psychologischer Art. In Realübungen, die genau auf solche Situationen vorbereiten sollen, erscheint es daher sinnvoll, Experten all dieser Fachrichtungen / Domänen gleichzeitig zu integrieren, um eine umfassende Vorbereitung auf Seenotfälle zu gewährleisten.

Stellt allein das Zusammenkommen der zu trainierenden Seeleute bereits ein großes organisatorisches Problem dar, das zudem mit hohen Kosten verbunden ist, so ist es umso schwieriger, zusätzlich Experten verschiedener Fachbereiche zur selben Zeit am selben Ort zu versammeln. Um diese Herausforderung zu meistern, entwickelten die Verbundpartner des Verbundprojektes KNoTen 2.0 das Konzept der sogenannten KNoTen-Station. Die KNoTen-Station ist eine Kombination aus Hard- und Software, die es erlauben soll, Realübungen mittels moderner Kameratechnik in Echtzeit über das Internet zu übertragen, so dass Experten das Trainingsgeschehen von jedem beliebigen Ort aus verfolgen und mittels schriftlicher Kommentare Feedback geben können.

Das Besondere der Realübungen, die mit Hilfe der KNoTen-Station begleitet werden sollen, ist, dass nicht mehr nur einzelne Notfallsequenzen (bspw. das Löschen eines Brandes im Maschinenraum oder das Besetzen der Rettungsboote) trainiert werden, sondern vielmehr komplexe Notfallsituationen, die vor allem durch das Zusammenwirken der verschiedenen Akteure, die in einer Notfallsituation miteinander interagieren müssen, gekennzeichnet sind. Aus diesem Grund wird während einer Realübung an verschiedenen Orten gleichzeitig geübt. So kann sich bspw. eine Übungsgruppe auf der Brücke befinden, während eine zweite Übungsgruppe mit der Brandbekämpfung beschäftigt und eine dritte Übungsgruppe für die Evakuierungseinrichtungen zuständig ist. Auf diesem Weg werden nicht nur die einzelnen, auszuführenden Tätigkeiten trainiert, sondern auch die Koordination und Kommunikation innerhalb einer Crew, die während eines Notfalls v.a. aufgrund des damit verbundenen Stress‘ besonders erschwert wird.

Mittels der Kameratechnik der KNoTen-Station werden also mehrere Schauplätze der Realübung gleichzeitig gefilmt und Experten über das Internet zugänglich gemacht, so dass diese mehrere Szenerien gleichzeitig beobachten und kommentieren können. Der vollständige Aufbau der Knotenstation ist der folgenden Abbildung 1 zu entnehmen.

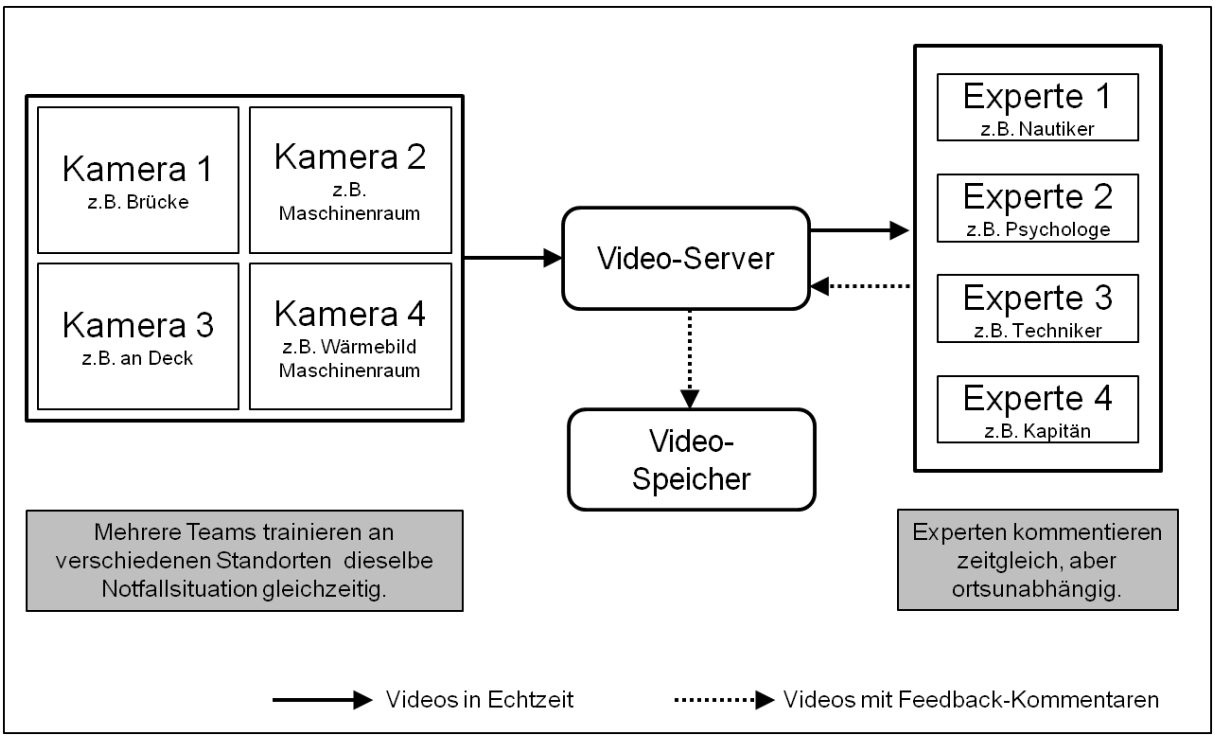


Abbildung 1: Grundprinzip der KNoTen-Station.

Im Anschluss an die Realübung findet eine Auswertung mit den Trainingsteilnehmern statt. Hierbei werden die nun kommentierten Videos vollständig oder in Auszügen vorgeführt, um den Teilnehmern ihr Verhalten während der Übung vor Augen zu führen. In Verbindung mit den Feedbackkommentaren, die an den Stellen im Video eingeblendet werden, an denen die Experten sie zuvor platziert haben, entsteht somit eine Reflexions- und Diskussionsbasis[2], die als eine weitere Lehreinheit aufzufassen ist. So können die Trainierten nun zum ersten Mal sehen, wie ihre Kollegen und sie selbst während des Notfalls agiert haben, und können die Gelegenheit nutzen, das Zusammenspiel der Teams aus einer anderen Perspektive zu beurteilen.

[2] Das Vorgehen, fremde oder eigene Trainings- oder Verhaltensaufnahmen zur Unterstützung der Reflexion einzusetzen, wird seit geraumer Zeit in psychologischen Therapien erfolgreich eingesetzt (s. z.B. Griffiths, 1973; Galassi, Galassi & Litz, 1974; Roter, Larson, Shinitzky, Chernoff, Serwint, Adamo & Wissow, 2004).

3 Theoretischer und empirischer Hintergrund zur Ableitung der Forschungsfragen

Die vorliegende Studie sowie das beschriebene Forschungsprojekt basieren auf der Idee, dass insbesondere ein nach bestimmten Grundregeln gestaltetes Feedback besonders effektiv ist und bei der Aneignung benötigter und in Notsituationen erfolgversprechender Verhaltensweisen helfen kann. Daher wird das Thema Feedback im kommenden Abschnitt näher beleuchtet.

3.1 Feedback im Trainingsprozess

3.1.1 Bedeutung von Feedback in Trainings- und Lernprozessen

Die Wirkung von Feedback bzw. Rückmeldung an Lernende im Lern- oder Trainingsprozess wird seit geraumer Zeit in der psychologischen Literatur untersucht. Zahlreiche empirische Befunde zeigen, dass sich richtig gestaltete Rückmeldung positiv auf den Lernerfolg und die Leistung von Menschen (Bangert-Drowns, Kulik, Kulik & Morgan, 1991) sowie ihre Motivation zur Verhaltensänderung auswirken kann (Hattie & Timperley, 2007). Zudem finden sich Hinweise, dass Feedback Verhalten steuert, zukünftige Leistungsziele beeinflusst, die Fähigkeit, Fehler auch selbständig zu erkennen, erhöht sowie individuelle Lernprozesse und die persönliche Selbsteinsicht fördert (London, 2003; Klebl, 2006).

Unter Feedback ist im vorliegenden Zusammenhang der Austausch darüber zu verstehen, was eine Person an einer anderen Person wahrgenommen hat (Fengler, 2010). Im Training bzw. im Lernprozess sind meist die Ausbilder, Instrukteure oder andere Lernende diejenigen, die das Verhalten einer anderen Person während eines Trainings oder in einer Übung durch Beobachtung wahrnehmen. In einem Gespräch während oder nach der Lehreinheit melden sie den Beobachteten zurück, was ihnen an deren Verhalten aufgefallen ist. Gegebenenfalls kann die Rückmeldung Vorschläge darüber enthalten, wie das beobachtete Verhalten verbessert werden kann (Fengler, 1975). Die betroffenen Lernenden werden demnach mit ihrem eigenen Verhalten konfrontiert und erhalten im Gespräch die Gelegenheit, sich an der Reflexion über das eigene Verhalten zu beteiligen (Fengler, 2010).

Die Feedback-Forschung geht seit einiger Zeit von einem kognitiven und konstruktivistischen Lernkonzept aus (Musch, 1999). Nach dem kognitiven Lernkonzept wurde die Funktion von Feedback bis in die siebziger Jahre des 20. Jahrhunderts lediglich darin gesehen, im Sinne des operanten Konditionierens erwünschtes Verhalten zu belohnen und damit zu verstärken sowie negative Verhaltensweisen zu bestrafen und dadurch in der Zukunft zu reduzieren (für einen

Überblick zur operanten Konditionierung s. Zimbardo & Gerrig, 2004). Heute werden Konsequenzen des eigenen Verhaltens gemäß dem konstruktivistischen Lernkonzept eher als wichtige Informationsquelle gesehen, die erforderlich für die Überprüfung des eigenen Verhaltens, die Verfeinerung von Fähigkeiten und die Anpassungsfähigkeit ist. Demnach versorgt Feedback die Lernenden mit strategisch nützlichen Informationen und unterstützt damit die Selbstregulierung (Bangert-Drowns, Kulik, Kulik & Morgan, 1991). Insbesondere negatives Feedback im Sinne eines Hinweisens auf Fehler im Verhalten gibt einen Anstoß für Reflexionen und bewegt Lernende dazu, ihr eigenes Wissen, aber auch ihre grundsätzliche Herangehensweise an Probleme zu überprüfen (Musch, 1999).

3.1.2 Expertenfeedback im Rahmen des KNoTen 2.0-Projektes

Im Projekt KNoTen 2.0 wird Feedback eingesetzt, um bei Teilnehmenden an Sicherheitstrainings erforderliche Verhaltensänderungen herbeizuführen oder bereits gelernte Strategien zu festigen. Dabei sollen Internet- und insbesondere Web 2.0-Technologien den Teilnehmenden von Realübungen ein Expertenfeedback bezüglich ihrer Trainingsleistung ermöglichen. Hierbei müssen die rückmeldenden Experten nicht mehr am Trainingsort anwesend sein, sondern verfolgen und kommentieren das Trainingsgeschehen via Internet in Echtzeit unabhängig von ihrem aktuellen Aufenthaltsort. Durch die Experten zu kommentieren sind all jene Verhaltensweisen, die besonders erfolgreich waren oder die zu besonders schlechten Ergebnissen führten, was bedeutet, dass sowohl positives als auch negatives Feedback gegeben werden soll.

Nach Beendigung der Übung soll das Feedback möglichst zeitnah gegeben werden. Anhand der Videoaufzeichnungen, die zuvor von den Experten schriftlich kommentiert wurden, wird das Training von den Teilnehmern und den Instrukteuren der Übung gemeinsam ausgewertet. Hierbei sollen insbesondere die Verhaltensweisen fokussiert werden, die von den Experten mit einem Feedbackkommentar versehen wurden. Diese Kommentare werden an der Stelle des Videos eingeblendet, die den Zeitpunkt des Kommentierens wiedergibt. Auf diese Weise sind das kommentierte Verhalten im Video wie auch der Feedback-Kommentar für die Trainierenden gleichzeitig sichtbar. Dieses nachträgliche Feedback hat den Vorteil, dass neben der eigentlichen Lernsituation in der Realübung auch das Nachbereitungsgespräch zur Lernsituation wird, die zu verinnerlichenden Inhalte also wiederholt verarbeitet und hierdurch besser gespeichert werden (Musch, 1999). Damit Feedback diese Wirkungen entfaltet, müssen aber einige Bedingungen beachtet werden.

3.1.3 Zur allgemeinen Wirkung von Feedback

Die oben beschriebenen positiven Auswirkungen von Feedback entstehen in einem Reiz-Reaktions-Mechanismus, der vielen Einflussgrößen unterliegt. Feedback an sich ist lediglich eine bloße Information des Feedbackgebers an den Feedbacknehmer (Latham & Locke, 1991). Welche Konsequenzen sich aus dieser Nachricht letztlich für das zukünftige Verhalten des Feedbacknehmers ergeben, ob er beispielsweise in Zukunft positive Verhaltensweisen zeigt oder Fehler vermeidet, hängt von seiner Bewertung des Feedbacks sowie den daraus abgeleiteten Entscheidungen ab (Latham & Locke, 1991). Ilgen, Fischer und Taylor (1979) beschreiben den Prozess beginnend von der Aufnahme der Feedbackbotschaft bis zur möglichen Verhaltensreaktion in vier Schritten, die durch verschiedene Größen beeinflusst werden. Abbildung 2 stellt diesen Prozess dar.

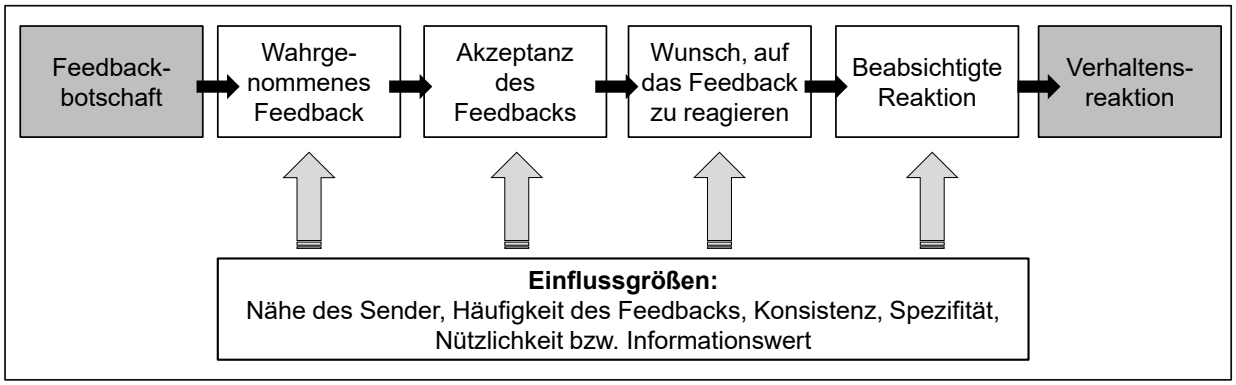


Abbildung 2: Der Prozess der Feedback-Wirkung
(Quelle: in Anlehnung an Ilgen et al., 1979)

Wahrgenommenes Feedback. Nachdem der Feedbackgeber seine Information mittels Kommunikation an den Empfänger gesendet hat, wird das Feedback vom Empfänger wahrgenommen. Wie das Feedback wahrgenommen wird, ergibt sich zum einen aus der objektiven Botschaft und zum anderen aus der Interpretation dieser Botschaft durch den Empfänger. Damit das Feedback möglichst so, wie es gemeint ist, ankommt, sollten einige Voraussetzungen erfüllt sein. Hierzu zählen bspw. die vom Empfänger eingeschätzte fachliche Nähe des Feedbackgebers und seine Glaubwürdigkeit. So wird ein Feedback, dessen Sender für fachlich nicht kompetent oder unglaubwürdig gehalten wird, wahrscheinlich gar nicht erst Gehör beim Feedbackempfänger finden. Darüber hinaus sind die grundlegende Tendenz – wird positives oder negatives Feedback gegeben – sowie die Häufigkeit der Übermittlung von Rückmeldungen von Bedeutung. Diese Faktoren sind grundlegend für die Akzeptanz der Rückmeldung, die Ilgen et al. (1979) als zweiten Schritt des Wirkprozesses anführen.

Akzeptanz des Feedbacks. Die Akzeptanz von Feedback ist vor allem davon abhängig, in welchem Maße der Feedbackempfänger die Rückmeldung als adäquate Beschreibung des eigenen Verhaltens auffasst. Nur wenn die Rückmeldung vom Feedbacknehmenden akzeptiert wird, entstehen der Wunsch und letztendlich die Absicht, das Feedback in entsprechendes Verhalten umzusetzen. Auch für die Akzeptanz des Feedbacks ist die Glaubwürdigkeit des Feedbackgebenden entscheidend. Diese ergibt sich vor allem aus der ihm zugeschriebenen Expertise und seiner Vertrauenswürdigkeit. Zusätzlich kommt in dieser Stufe des Wirkungsprozesses von Feedback aber hinzu, dass die Feedbackbotschaften zum einen konsistent sein sollten, d.h. einander nicht widersprechen und Spezifität besitzen sollten, d.h. konkret auf einen bestimmten Sachverhalt bezogen sind (Ilgen et al., 1979; Klebl, 2006). Hierzu kann das Einbringen von konkreten Verhaltensbeispielen beitragen (Greif & Breckwoldt, 2012).

Der Wunsch, auf das Feedback zu reagieren und die beabsichtigte Reaktion. Bei entsprechender Akzeptanz des Feedbacks entwickelt der Feedbacknehmer den Wunsch, entsprechend dem Feedback zu reagieren. Der Wunsch nach einer tatsächlichen Verhaltensreaktion auf das Feedback wird insbesondere von der Nützlichkeit bzw. dem Informationswert der Rückmeldung beeinflusst. Feedback mit einem hohen Informationswert unterstützt den Feedbacknehmer dabei, sich neue Leistungsziele zu setzen und liefert gegebenenfalls Anhaltspunkte für ein erfolgreiches Vorgehen. Je mehr nützliche Informationen der Feedbacknehmer also erhält, umso wahrscheinlicher ist es, dass er sich wünscht sein Verhalten in der vom Feedbackgeber vorgeschlagenen Weise zu verändern. Aus dem Wunsch, zu reagieren, kann sich im besten Fall eine Verhaltensabsicht entwickeln, wobei in der psychologischen Forschung davon ausgegangen wird, das diese Intention ein guter Prädiktor dafür ist, dass ein bestimmtes Verhalten dann auch wirklich in die Tat umgesetzt wird (Ajzen, 1991). Auch das Entstehen der Verhaltensabsicht ist in hohem Maße von der Spezifität des Feedbacks beeinflusst. Sehr konkretes Feedback kann in dieser Phase dem Feedbacknehmer helfen, seine neuen Ziele ebenfalls sehr spezifisch zu formulieren. Hierdurch kann mit einer positiven Wirkung auf die durch das Verhalten entstehende Leistung gerechnet werden (Latham & Locke, 1991).

Ilgen et al. (1979) nennen in ihrer Konzeption des Wirkprozesses noch weitere Einflussfaktoren, insbesondere solche, die in der Persönlichkeit des Feedbackempfängers ihren Ursprung finden. Hierzu zählen das Alter, die Kontrollüberzeugung (internal / external), das Selbstkonzept sowie die Motivstruktur. Hierbei handelt es sich jedoch um größtenteils nicht beeinflussbare Persönlichkeitsvariablen, die daher im KNoTen 2.0-Projekt nicht näher betrachtet werden.

Bei der Auswahl eines Feedbackgebers ist zu bedenken, ob die Person Glaubwürdigkeit, also Expertise und Vertrauenswürdigkeit, ausstrahlt. Gegebenenfalls sollte der feedbackgebende Experte beispielsweise durch die Veranstalter der Realübung vorgestellt und auf seine Qualifikation hingewiesen werden. Bei der Feedbackbotschaft sollte insbesondere auf die Häufigkeit, die Konsistenz, den Informationsgehalt und die daraus folgende Nützlichkeit sowie die Spezifität des Feedbacks Wert gelegt werden. Für die Formulierung der Feedbackbotschaft ist gegebenenfalls ein Training der Experten selbst in Erwägung zu ziehen, um so eine positive Wirkung durch das Feedback zusätzlich zu unterstützen.

3.2 Wahrnehmungs- und Konzentrationsfähigkeit als spezifische Anforderungen der KNoTen-Station

Durch den Einsatz der KNoTen-Station können Experten den Trainierenden Rückmeldung geben, ohne physisch beim Training präsent sein zu müssen. Vielmehr verfolgen die Experten an einem Bildschirm den Trainingsverlauf, der in Echtzeit via Internet übertragen wird. Hierbei beobachten sie mehrere Gruppen von Trainierenden gleichzeitig. Die Gruppen agieren dabei jeweils an verschiedenen Übungsorten, aber im gleichen simulierten Notfall. Sobald den Experten ein Verhalten auffällt, das sie als kommentierungswürdig einschätzen, ist es ihre Aufgabe, einen Feedbackkommentar über die Tastatur des verwendeten Computers einzugeben.

Die erfolgreiche Bewältigung dieser Aufgabe erfordert unterschiedliche Wahrnehmungs- und Konzentrationsprozesse. Dabei ist unter Wahrnehmung der Prozess zu verstehen, durch den die Informationen, die dem Individuum durch seine Sinnesorgane bereitgestellt werden, organisiert und interpretiert werden (Hagendorf, Krummenacher, Müller & Schubert, 2011). Durch Wahrnehmung erhält das Individuum demnach Kenntnis über das Geschehen in seiner Umwelt und auch in seinem Körper und kreiert ein mentales Modell seiner inneren und äußeren Umwelt (Nerdinger, 2012). Unter Konzentration ist ein „bewusstes Lenken und Kontrollieren der Aufmerksamkeit auf Reize und Reizkonstellationen (Sachverhalte der Umgebung und des eigenen Körpers)“ zu verstehen (Brunner & Zeltner, 1980, S. 122). Demnach ist Konzentration ein willentlicher Teil von Wahrnehmungsprozessen, der dafür sorgt, dass eine Person bestimmte, für ihn relevante Informationen aufnehmen kann. Welche Wahrnehmungs- und Konzentrationsfähigkeiten im Feedbackerstellungsprozess notwendig sind, und damit als Anforderungen an die Arbeit mit der KNoTen-Station zu sehen sind, wird im Folgenden beschrieben.

Zur Beschreibung der dabei ablaufenden Prozesse eignet sich das sogenannte Drei-Speichermodell der Informationsverarbeitung von Atkinson und Shiffrin (1968). Das Modell

verdeutlicht, welche Schritte der Filterung und Verarbeitung eingehende Informationen bzw. Reize durchlaufen, bis sie zu einer Verhaltensreaktion oder zur langfristigen Abspeicherung im Gedächtnis führen. Da diese Vorgänge bedeutsam für die Arbeit mit der KNoTen-Station sind und sich daraus Anforderungen an die Wahrnehmungs- und Konzentrationsfähigkeiten der Experten ableiten lassen, wird das Modell hier als theoretischer Rahmen für die weiteren Ausführungen herangezogen. Die folgende Abbildung 3 zeigt das Modell in Anlehnung an die ursprüngliche Version von Atkinson und Shiffrin (1968) sowie seine Erweiterung um die wichtigsten Aspekte der Feedbackerstellung im Rahmen des Einsatzes der KNoTen-Station (grau unterlegte Felder).

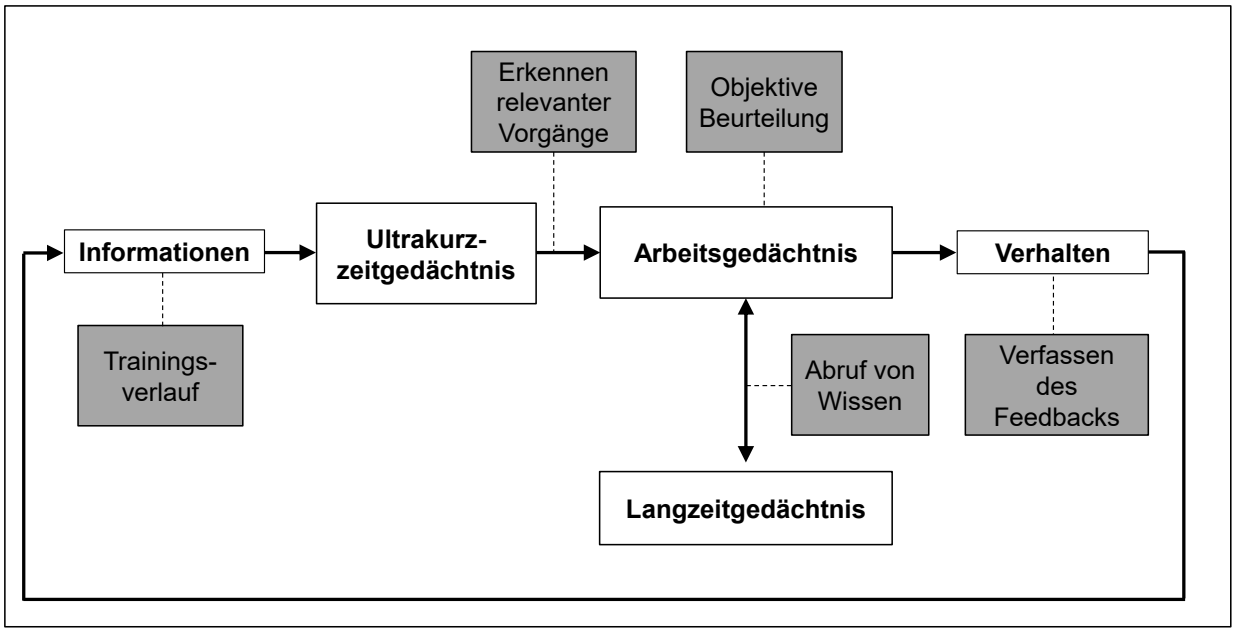


Abbildung 3: Informationsverarbeitung bei der Arbeit mit der KNoTen-Station (Quelle: In Anlehnung an Atkinson & Shiffrin, 1968).

Das Modell von Atkinson und Shiffrin (1968) geht davon aus, dass Informationen aus der Umwelt oder aus dem eigenen Körper über die Sinnesorgane aufgenommen werden. Die eingehenden Informationen werden nahezu vollständig im sogenannten Ultrakurzzeitgedächtnis (UKZ) für kurze Zeit abgespeichert. Im UKZ, das aus einer Vielzahl sensorischer Speichereinheiten besteht, wird die eingehende Information aber nicht verarbeitet (Hasebrook, 1995). Im nächsten Schritt wird eine geringe Zahl der im UKZ gespeicherten Information in das Arbeitsgedächtnis weitergeleitet. Das Arbeitsgedächtnis, in dem die Informationen bewusst verarbeitet werden, verfügt über Selektionsmechanismen, die aus der Flut an Informationen die jeweils relevanten herausfiltern. Diese Auswahl erfolgt sowohl auf rein physischer Ebene, z.B. werden besonders laute Geräusche oder grelle Farben wahrgenommen, als auch auf semantischer Ebene (Heuer, 2010). Letzteres bedeutet, dass für das Individuum bedeutsame Informationen eher ausgewählt

werden als unwichtige. Informationen, die nicht ausgewählt werden gehen, werden aus dem Ultrakurzzeitgedächtnis gelöscht. Die ins Arbeitsgedächtnis übertragenen Informationen werden dort verarbeitet, indem sie mit bestehendem Wissen verglichen und dadurch identifiziert werden. Anschließend werden sie langfristig im Langzeitgedächtnis abgespeichert bzw. führen unmittelbar zu Verhalten. Durch die Speicherung im Langzeitgedächtnis werden die verarbeiteten Informationen entweder in bestehende Wissensstrukturen integriert oder es werden neue Wissensstrukturen aufgebaut (Atkinson & Shiffrin, 1968).

In ähnlicher Weise läuft der Feedbackerstellungsprozess durch die Experten mit Hilfe der KNoTen-Station ab. Neben den Reizen aus der Arbeitsumgebung erreichen die Informationen in Form von Bildern und Geräuschen aus dem Trainingsverlauf die Sinnesorgane der Experten. Die Filterung relevanter Informationen aus der Reizflut erfordert von den Experten hohe Konzentration, d.h. ein Experte muss seinen Aufmerksamkeitsfokus auf jene Informationen richten, die für die auszuführende Aufgabe relevant sein könnten (Matthews, Davies, Westerman & Stammers, 2000). Als erste Anforderung an die Experten ergibt sich demnach eine hohe Konzentrationsfähigkeit, um die Aufmerksamkeit auf das relevante zu bewertende Trainingsgeschehen zu richten und sich nicht von den zusätzlich vorhandenen Umgebungs- oder Körperreizen ablenken zu lassen. Darüber hinaus müssen die Experten potenzielle Störgrößen während der Arbeit mit der KNoTen-Station ausschalten. So wird die Chance erhöht, dass die relevanten Reize trotz der beschränkten Verarbeitungskapazität des Arbeitsgedächtnisses zur weiteren Verarbeitung ausgewählt werden.

Unter relevanten Reizen sind bei der Arbeit mit der KNoTen-Station beobachtbare Verhaltensweisen der Trainierenden zu verstehen, die nach Auffassung der Experten einer Rückmeldung bedürfen. Um festzustellen, welche Verhaltensweisen dieses Kriterium erfüllen und mit welcher Rückmeldung diese versehen werden sollen, greift der Experte auf sein fachspezifisches Wissen und seine langjährige Erfahrung zurück. Nur im Vergleich der aktuellen Beobachtung mit abgespeichertem Wissen und Erfahrungen bezüglich richtiger und falscher Verhaltensweisen kann das gerade beobachtete Verhalten richtig eingeordnet werden (vgl. Christmann, 1999). Wird festgestellt, dass das Verhalten in Konflikt zum Wissen über richtiges Verhalten steht, muss ein Feedbackkommentar geschrieben werden. Hieraus leitet sich die zweite Anforderung an die feedbackgebenden Experten ab. Um fehlerhaftes oder im entgegengesetzten Fall besonders effizientes Verhalten der Trainierenden zu erkennen, muss der Experte über ein umfangreiches Wissen in seinem Fachgebiet verfügen, das einen Abgleich mit dem Beobachteten ermöglicht (Matthews et al., 2000). Diese Anforderung sollte allerdings bereits durch den Expertenstatus der vorgesehenen Nutzer der KNoTen-Station erfüllt sein.

Ist der Experte im Rahmen der Informationsverarbeitung zu einer Einschätzung des Verhaltens gekommen und hat sich für eine Feedbackformulierung entschieden, erfordert es die KNoTen-Station, dass er dieses Feedback in möglichst kurzer Zeit über die Tastatur in den verwendeten Rechner eingibt, um sich anschließend wieder der Beobachtung der Videos zu widmen. Hierbei vollzieht sich ein Aufgabenwechsel, bei dem die Tätigkeit des Beobachtens und Einschätzens beendet und die Tätigkeit des Schreibens am Computer gestartet werden muss (vgl. Grajewski, 2005). Dieser Vorgang geht in der Regel mit „Wechselkosten“ einher, die sich in einem Zeitverlust und auftretenden Fehlern zwischen dem Ende der einen und dem effizienten Ausführen der neuen Aufgabe niederschlagen können (Heuer, 2010). Ein erneuter Aufgabenwechsel vollzieht sich nach dem Fertigstellen eines Kommentars, wenn der Experte sich sofort wieder der Beobachtung der Videos widmen muss, um möglichst wenig zu verpassen. Eine weitere Anforderung ist also in der Fähigkeit zum möglichst fehlerfreien, schnellen Wechseln zwischen den beiden Aufgaben „Beurteilen“ und „Texteingabe“ sowie den Aufgaben „Texteingabe“ und „Beobachtung“ zu sehen.

Zusammenfassend kann festgehalten werden, dass die Arbeit mit der KNoTen-Station mehrere Anforderungen an die Experten stellt. So müssen die Experten zunächst in der Lage sein, sich weitgehend auf das Trainingsgeschehen zu konzentrieren. Des Weiteren sollten sie über entsprechendes Wissen und Erfahrungen verfügen, um die zu kommentierenden Verhaltensweisen erkennen zu können. Dabei müssen die Experten in der Lage sein, auf ihren Beobachtungen basierend zu beurteilen, was durch entsprechende Schulungen erreicht werden kann. Letztlich sollten sich die Experten schnell anpassen können, um die „Kosten“, die durch den Wechsel zwischen den unterschiedlichen Teilaufgaben entstehen, so gering wie möglich zu halten.

3.3 Forschungsfragen - Mögliche Auswirkungen zusätzlicher Funktionalitäten

Ausgehend vom Grundaufbau der KNoTen-Station ist zu untersuchen, ob weitere Gestaltungsvarianten der Benutzeroberfläche die Wirksamkeit des Feedbacks unterstützen und sogar steigern können. Diese Gestaltungsalternativen könnten unterschiedlichen Einfluss auf jene Größen haben, die sich bereits in empirischen Studien als förderlich in den Phasen des Feedbackwirkprozesses erwiesen haben. Hierzu zählen die Genauigkeit, die Konsistenz, die Glaubwürdigkeit, die Häufigkeit sowie der Informationswert bzw. die Nützlichkeit des Feedbackgebers und seiner Botschaft (Ilgen et al., 1979). Im Folgenden werden mögliche Gestaltungsvarianten vorgestellt und hinsichtlich ihres zusätzlichen Nutzens, aber auch möglicher Nachteile, die sich aus Befunden der Wahrnehmungs- und Konzentrationsforschung sowie der Sozialpsychologie ableiten, erläutert. Aus der Diskussion der Varianten werden Forschungsfragen abgeleitet, die im

weiteren Verlauf der Arbeit empirisch untersucht werden. Auf diesem Weg soll ermittelt werden, welche zusätzlichen Funktionalitäten für die endgültige Umsetzung der KNoTen-Station in Betracht gezogen werden sollten.

3.3.1 Die Sichtbarkeit von Feedback-Kommentaren anderer Experten

Das Konzept der KNoTen-Station besteht darin, durch die Ortsunabhängigkeit von feedbackgebenden Experten zu mehr Kosteneffizienz in der Aus- und Weiterbildung zu gelangen, ohne dabei Einbußen hinsichtlich der Wirksamkeit der Trainingsmaßnahme und insbesondere der Feedbackintervention hinnehmen zu müssen. Experten verschiedener, für das Training relevanter Themenbereiche können gleichzeitig und von verschiedenen Standorten Kommentare zum Trainingsverlauf abgeben. Da bei Notfalltrainings mehrere Domänen, wie die Nautik, Technik und die Psychologie eine wesentliche Rolle spielen, kann es sinnvoll sein, mehrere Experten aus unterschiedlichen Fachbereichen via Internet dem Training hinzuzuschalten, um ein umfassendes, informationsreiches Feedback zu erhalten. Wenn demnach mehrere Experten gleichzeitig online kommentieren, stellt sich die Frage, ob diese Experten die Kommentare der jeweils anderen gleichzeitig lesen können sollten und dies als zusätzliche Funktionalität in die KNoTen-Station integriert werden sollte.

Die empirische und theoretische Befundlage ist uneindeutig, es finden sich sowohl Belege für Vorteile als auch für Nachteile der gewählten Anordnung, die im Folgenden hinsichtlich der Güte möglicher Expertenfeedbackkommentare diskutiert werden. Diese Diskussion orientiert sich an vier möglichen theoretischen Erklärungen: Soziale Erleichterung, Wirkung von Heuristiken, Social Loafing und Informationsüberfrachtung.

Soziale Erleichterung (Social Facilitation)

Eine Wirkung, die die Sichtbarkeit anderer, gleichzeitig hinzugeschalteter Experten haben könnte, wird in der Sozialpsychologie als Social Facilitation (Soziale Erleichterung) (vgl. Zajonc, 1965) bezeichnet. Hierunter versteht man den Effekt, dass durch die bloße Anwesenheit von Zuschauern oder Personen, die an der gleichen Aufgabe selbständig arbeiten (Co-Akteure), die Leistung verbessert wird. Voraussetzung dabei ist, dass das Individuum die Aufgabe bereits gut beherrscht. Die Effekte, die Social Facilitation erzielt, sind in diversen Experimenten mit unterschiedlichen Versuchspersonen und zu lösenden Aufgaben - sowohl rein motorischen als auch intellektuellen und Vigilanz-Aufgaben[3] - ermittelt worden. So fanden Bond und Titus (1983) in ihrer Metaanalyse zur Sozialen Erleichterung heraus, dass sich durch die Anwesenheit anderer die Geschwindigkeit bei der Bearbeitung von Aufgaben und in gewissem Maße auch

[3] Unter Vigilanz-Aufgaben werden hier Aufgaben zur Ermittlung der Daueraufmerksamkeit verstanden.

die Genauigkeit bei der Bearbeitung erhöhte. Zajonc (1965) berichtete darüber hinaus von einer quantitativen Leistungssteigerung in dem Sinne, dass rein zahlenmäßig mehr Antworten gegeben wurden, als in Kontrollgruppen ohne Präsenz anderer.

Inzwischen wurde empirisch gezeigt, dass Social Facilitation auch durch die lediglich virtuelle Anwesenheit anderer Personen auftritt, z.B. bei Personen, die sich gleichzeitig auf einer Internetplattform befinden (Rafaeli & Noy, 2002; Christopherson, 2007). Daher kann davon ausgegangen werden, dass die empirischen Befunde zur Wirkung von Social Facilitation auch auf den vorliegenden Kontext der virtuellen Anwesenheit anderer Experten übertragbar sind.

Obwohl vielfach gezeigt werden konnte, dass Social Facilitation grundsätzlich zu einem höherren Output führt, d.h. bspw. mehr Antworten in Anwesenheit anderer Personen gegeben werden, so ist dieser Output hinsichtlich seiner Qualität aber davon abhängig, ob die Inhalte, die abgefragt werden, vom Individuum bereits beherrscht werden oder nicht. Das bedeutet, dass man nur bei gut gelernten Aufgaben von der Anwesenheit anderer profitiert. Bei Aufgaben hingegen, die erst neu erlernt werden müssen, die also zunächst zu falschen Ergebnissen führen können, bewirkt die Anwesenheit anderer eine Leistungsverschlechterung dahingehend, dass falsche Antworten vermehrt auftreten (Zajonc & Sales, 1966).

Bei den Kommentatoren, die die KNoTen-Station nutzen sollen, handelt es sich um Experten, „was bedeutet, dass er / sie über disziplinär strukturiertes Fachwissen verfügt, das im Laufe einer mehrjährigen Ausbildung erworben und durch einschlägige Berufserfahrung vertieft wurde," (Bromme, Jucks & Rambow, 2004, S.176). Somit beherrschen Experten die Aufgabe des Feedbackgebens, so dass – aufgrund der virtuellen Anwesenheit der Experten – eher von einer Verbesserung der Leistungen als von einer Verschlechterung ausgegangen werden kann.

Demnach ist zu vermuten, dass die virtuelle Anwesenheit der anderen Experten eine Erhöhung der Arbeitsgeschwindigkeit bewirkt, die sich sowohl in kürzeren Bearbeitungsdauern einzelner Kommentare als auch in der Dauer der Gesamtbearbeitung eines Trainingsdurchlaufes niederschlägt. Durch schnelleres Kommentieren bietet sich die Möglichkeit, sich zügiger wieder dem Geschehen auf den Bildschirmen zuzuwenden, wodurch das Risiko, wichtige Ereignisse zu verpassen, reduziert wird. Dies und der allein durch Social Facilitation gesteigerte Output können sich sowohl in einer höheren Anzahl an Kommentaren als auch in einer gesteigerten Länge der Einzelkommentare zeigen. Hierdurch sind positive Wirkungen hinsichtlich der wahrgenommenen Feedbackhäufigkeit, der wahrgenommenen Genauigkeit und Spezifität sowie ein

Zugewinn beim Informationsgehalt, also der Nützlichkeit der Feedbackkommentare, zu erwarten. All diese Größen sind wichtig für die Wirkung des Feedbacks.

Neue Perspektiven vs. Heuristiken

Durch das Lesen, Verarbeiten und Nachvollziehen der Kommentare der anderen Fachleute können den Experten auch neue Perspektiven im Hinblick auf das Trainingsgeschehen eröffnet werden (vgl. Zajonc, 1965). Situationen, die bspw. seltener wahrgenommen werden, weil sie ihren Ursprung eher in Domänen der anderen Experten haben, könnten daraufhin auch aus der eigenen fachlichen Sicht betrachtet werden. So ist z.B. denkbar, dass der Experte für Psychologie durch einen Kommentar des Experten für Technik auf einen bestimmten Vorgang im Trainingsgeschehen hingewiesen wird und daraufhin auch aus seiner psychologischen Perspektive einen Feedbackkommentar abgibt.

Insbesondere durch die Wahrnehmung der Sichtweise von Experten anderer Domänen könnte der Experte davon abgehalten werden, sonst zur kognitiven Erleichterung angewendete Verfügbarkeitsheuristiken (Tversky & Kahneman, 1973) bei seiner Beurteilung heranzuziehen. Heuristiken sind allgemeine, einfach anwendbare, aber zumeist unbewusste Regeln, die es erlauben, besonders schnell zu Urteilen und Entscheidungen zu gelangen (Raab, Unger & Unger, 2010). Verfügbarkeitsheuristiken im Speziellen sind auf den Grad der Leichtigkeit zurückzuführen, mit der Informationen aus dem Gedächtnis abgerufen können, d.h. „je einfacher wir uns an ein Ereignis erinnern und je mehr Beispiele uns dazu einfallen, desto häufiger und wahrscheinlicher erscheint uns dieses Ereignis," (Raab et al., 2012, S. 121). Der Rückgriff auf diese Art von Heuristiken führt häufig zu Fehleinschätzungen, da lediglich leicht aus dem Gedächtnis abrufbare Informationen, also bspw. Informationen über häufig beobachtete Fehlverhaltensweisen in Trainings, zur Beurteilung des aktuellen Geschehens herangezogen werden, da ihre Wahrscheinlichkeit höher eingeschätzt wird (vgl. Raab et al., 2010). Weniger präsentes Wissen über richtiges oder falsches Verhalten des Experten, das aber vielleicht dennoch hilfreich wäre, würde vernachlässigt. So würden bestimmte zu verändernde Verhaltensweisen der Trainierenden gar nicht wahrgenommen und folglich nicht mit Feedback versehen werden. Die Wahrnehmung der Perspektiven anderer Experten, durch das Einblenden ihrer Kommentare könnte hier Abhilfe schaffen und dazu führen, dass auch weniger gut im Gedächtnis verfügbare Informationen herangezogen werden.

Die Informationen, die der Experte durch das Lesen der anderen Kommentare erhält, können also als Anhaltspunkte für bestimmte, relevante Verhaltensweisen dienen, die ansonsten weniger präsent sind. Darüber hinaus können sie aber auch Hinweise auf Geschehnisse liefern, die

sich in den anderen Kameraperspektiven als denen abspielen, auf die der Experte gerade seinen Fokus gerichtet hat. Hat der Experte bspw. gerade auf Monitor 1 geachtet, könnte ein Kommentar zu Monitor 4 dazu führen, dass er seine Aufmerksamkeit auch zu jener Anzeige hinwendet. Demzufolge könnte durch die Sichtbarkeit zusätzlicher Kommentare anderer Experten die Aufmerksamkeit des Einzelnen einen breiteren Fokus erhalten. Er kann mehr beobachten und hat daraus resultierend mehr Gelegenheiten, Feedback zu geben. Ausgehend von dieser Überlegung ist anzunehmen, dass durch die Einblendung zusätzlicher Expertenkommentare mehr Feedbackkommentare zustande kommen und mehr Monitore in die Beobachtung einbezogen werden, sofern die anderen Experten Hinweise darauf liefern. Dies könnte das Feedback im Vergleich zur Ursprungsvariante der KNoTen-Station verbessern, sowohl in Bezug auf Häufigkeit als auch Informationsgehalt der Feedback-Kommentare.

Andererseits ist zu vermuten, dass Experten mehr Zeit zum Nachdenken über die Formulierung des Kommentars oder beim Aufschreiben selbst benötigen, wenn sie Sachverhalte im Kommentar schildern wollen, die im Gedächtnis nicht ganz so präsent sind als wenn sie über sehr präsente Inhalte schreiben würden. Ebenso ist davon auszugehen, dass die Kommentare ausführlicher sind, da die Experten gegebenenfalls Schwierigkeiten haben könnten, neue Sichtweisen sofort präzise zu formulieren und daher auf Be- und Umschreibungen zurückgreifen. Sowohl die längere Dauer beim Überlegen als auch die steigende Länge der Kommentare könnte sich demnach auf die Gesamtdauer des Beurteilungsprozesses auswirken, wenn der Experte auch nach Abschluss der Videoübertragung noch mit dem Schreiben des Kommentars beschäftigt ist. Hierbei sollte jedoch nicht vernachlässigt werden, dass diese Effizienzeinbußen einem Feedback gegenüberstehen, das hinsichtlich der Vollständigkeit und demnach des Informationsgehaltes überlegen sein könnte.

Social Loafing und Verantwortungsdiffusion

Neben diesen tendenziell als positiv zu bewertenden Auswirkungen der wahrgenommenen Präsenz anderer Experten können auch Effekte auftreten, die die Feedbackleistung behindern. Ein negativer Effekt, der durch die wahrgenommene Anwesenheit anderer entstehen kann, ist das sogenannte Social Loafing (Soziales Faulenzen; vgl. Latané, Williams & Harkins, 1979). Auch dieser Effekt ist auf virtuelle, soziale Situationen übertragbar, wie erste Studien untermauern (für einen Überblick s. Christopherson, 2007). Soziales Faulenzen zeigt sich darin, dass Mitglieder von Arbeitsgruppen durch ihr Wissen um weitere Verantwortungsträger die eigenen Anstrengungen herunterfahren und damit die Gruppenleistung reduzieren. Dieser Effekt tritt insbesondere dann auf, wenn den Gruppenmitgliedern nicht klar ist, wie wichtig ihre Arbeit zur Erreichung der Gesamtleistung ist. Daher ist ein Weg, Social Loafing einzudämmen, Personen,

die gemeinsam mit anderen arbeiten, ihren Beitrag und dessen Bedeutung zur Erreichung des gemeinsamen Ziels zu verdeutlichen (Williams, Harkins & Latané, 1981).

Neben dem Social Loafing kann es durch die wahrgenommene Präsenz anderer zu einer Verantwortungsdiffusion kommen (Darley & Latané, 1968; im virtuellen Kontext s. Christopherson, 2007). D.h., die wahrgenommene, wenn auch nur virtuelle Präsenz der anderen Experten, könnte dazu führen, dass der Einzelne sich nicht mehr selbst in der Verantwortung sieht und seine Aktivität einstellt, weil potenziell noch weitere Personen hierfür zur Verfügung stehen, die genau dieselbe Aufgabe bearbeiten.

Im Rahmen des KNoTen-Projektes könnten Social Loafing und Verantwortungsdiffusion dazu führen, dass der Experte denkt, bestimmte, eigentlich zu kommentierende Vorkommnisse während des Trainings könnten ebenso gut durch die anderen Fachleute bearbeitet werden. In der Folge schriebe der Experte selbst keinen Kommentar oder überließe die Beobachtung der Szenerie im Extremfall für eine Zeit sogar ganz den anderen Kommentatoren. In einem solchen Fall würde die Anzahl der Kommentare des entsprechenden Experten sinken, was sich negativ auf die Häufigkeit sowie die Konsistenz des Feedbacks auswirkt. Durch diese ausbleibenden Hinweise aus einer speziellen Domäne wird der Informationsgehalt und somit die Nützlichkeit des Feedbacks gesenkt. Dies sind jedoch Größen, die für die Wirksamkeit von Feedbackintervention von hoher Bedeutung sind.

Zusätzlicher Input und Informationsüberfrachtung

Durch das Einblenden fremder Kommentare werden die feedbackgebenden Experten einem zusätzlichen Stimulus ausgesetzt, der mit der eigentlichen Aufgabe konkurriert – der Formulierung von Kommentaren zu vier parallel gezeigten Videos. Damit ist schon der Grundaufbau der KNoTen-Station per se als sehr komplexe Reizvorlage anzusehen. So ist zu vermuten, dass bereits ohne die Einblendung der Fremdkommentare ein Überangebot an Informationen besteht und der Experte demzufolge nicht alles, was auf dem Bildschirm seines Computers eingeblendet wird, bewusst wahrnehmen kann (vgl. Kebeck, 1997).

Aufgrund der generell begrenzten Aufmerksamkeitskapazität des Menschen kann es durch Reize (Distraktoren), die zusätzlich zur eigentlichen Aufgabe präsentiert werden, zu einem Aufmerksamkeitskonflikt (Groff, Baron & Moore, 1982) kommen, indem die Distraktoren die Aufmerksamkeit von der eigentlichen Aufgabe abziehen. Im KNoTen-Projekt besteht solch ein Distraktor im Kommentieren der Trainingsaufnahmen, das parallel zur Beobachtung der Videos vorgenommen werden muss. Dies kann durch die verringerte kognitive Kapazität zu Leistungseinbußen bei der Erfüllung der Aufgabe führen (Jonas & Tanner, 2006). Im Anwendungsfall

der KNoTen-Station könnte dies bewirken, dass entweder die Kommentare gar nicht bewusst wahrgenommen werden und so nicht hilfreich sein können, oder dass vielleicht sogar die Fremdkommentare verfolgt werden und dadurch weniger oder keine Aufmerksamkeit mehr für das Geschehen in den Trainingsszenarien zur Verfügung steht. Im letzteren Fall liegt es nahe, dass weniger Vorkommnisse erkannt und kommentiert werden, so dass gegebenenfalls weniger Feedback-Kommentare verfasst werden. Die entsprechenden negativen Auswirkungen wären eine geringere Häufigkeit und Konsistenz des Feedbacks, wenn beispielsweise Fehlverhalten an einigen Stellen gesehen und kommentiert wird und an anderen nicht.

Demgegenüber kann aber auch die Nichtwahrnehmung der Fremdkommentare als Resultat des Wahrnehmungskonfliktes auftreten. In diesem Fall könnte nicht mit den positiven Effekten, die die Präsenz anderer herbeiführen kann, gerechnet werden.

Fazit

Hinsichtlich der Sichtbarkeit anderer Expertenfeedbacks bei der Beobachtung und Kommentierung von Realübungen liefern die sozialpsychologischen Theorien ein sehr ambivalentes Bild. Darüber hinaus kann zur inhaltlichen Qualität der Kommentare auf Basis der bisherigen Forschung noch keine Einschätzung möglicher Folgen der zusätzlichen Sichtbarkeit der anderen Expertenkommentare abgegeben werden. Folglich lassen sich zu diesem Zeitpunkt kaum eindeutige, gerichtete Hypothesen zur positiven oder negativen Wirkung der Sichtbarkeit der Kommentare anderer Experten formulieren. Demnach verbleibt hier für den weiteren Forschungsprozess nur die systematische Exploration, um einschätzen zu können, wie sich die angeführten psychologischen Effekte konkret in ihrem Zusammenspiel in der vorliegenden Projektidee auswirken. Daher wird für die Funktionalität, zusätzliche Expertenkommentare einzublenden (im Weiteren kurz „Expertenkommentare“ genannt), die folgende Forschungsfrage zugrunde gelegt:

Forschungsfrage 1:

Welche Auswirkungen hat die Sichtbarkeit anderer Expertenkommentare auf die Einzelleistung des Experten bei der Kommentierung des Trainingsgeschehens?

a) Welche Auswirkungen ergeben sich für die quantifizierbaren Merkmale der Leistung, also die Anzahl der Kommentare, die Anzahl der einbezogenen Videos, die Ausführlichkeit bzw. Länge der Kommentare, die Bearbeitungszeit der einzelnen Kommentare sowie die Gesamtdauer des Feedbackprozesses?

b) Welche Auswirkungen ergeben sich für die qualitativen Merkmale der Leistung, also die Richtigkeit der Kommentare und die Verständlichkeit der Kommentare?

3.3.2 Erhöhung der Wahrnehmungswahrscheinlichkeit anderer Expertenkommentare

Ausgehend von den bisherigen Ausführungen stellt sich außerdem die Frage, ob die positiven Wirkungen der Wahrnehmung der Kommentare der anderen Experten sich überhaupt entfalten können oder ob sie eventuell durch den entstehenden Wahrnehmungskonflikt gar nicht aufgenommen werden. Um Social Facilitation und das Ausschließen der Verwendung von Verfügbarkeitsheuristiken zu unterstützen, besteht neben dem bloßen Einblenden der Kommentare die zusätzliche Möglichkeit, den Experten mittels eines Hinweisgeräusches auf das Erscheinen der Kommentare vorzubereiten, wie es beispielsweise in Mail- und Chatprogrammen bei Eintreffen neuer Nachrichten eingesetzt wird. In diesem Fall würde jedes Erscheinen eines Kommentars eines anderen Experten von einem Geräusch begleitet.

Nach Sanders (1983) können akustische Signale bei Unsicherheit darüber, wann ein Reiz, auf den reagiert werden soll, eintrifft, dabei helfen, die Aufmerksamkeit auf den kommenden Reiz auszurichten und sogar die Reaktionszeit darauf zu steigern. Auch Posner (1994) weist darauf hin, dass durch Warn- oder Hinweistöne die Aufmerksamkeit gesteigert und auch eine Erhöhung der Arbeitsgeschwindigkeit herbeigeführt werden kann.

Um zu untersuchen, ob eine Forcierung der Wahrnehmung der anderen Expertenkommentare dazu führt, dass die Kommentare der anderen eher wahrgenommen werden und zu positiven Effekten führen, wird zusätzlich folgende Forschungsfrage aufgestellt:

Forschungsfrage 2:
Hilft ein zusätzlich eingefügter Hinweiston, sobald ein Kommentar eines anderen Experten eingeblendet wird, die positiven Auswirkungen der wahrgenommenen Anwesenheit zu steigern?

a) Welche Auswirkungen ergeben sich durch den zusätzlichen Hinweiston auf die quantifizierbaren Merkmale der Leistung, also die Anzahl der Kommentare, die Anzahl der einbezogenen Videos, die Ausführlichkeit bzw. Länge der Kommentare, die Bearbeitungszeit der einzelnen Kommentare sowie die Gesamtdauer des Feedbackprozesses?

b) Welche Auswirkungen ergeben sich durch den zusätzlichen Hinweiston auf die qualitativen Merkmale der Leistung, also die Richtigkeit der Kommentare und die Verständlichkeit / Nachvollziehbarkeit der Kommentare?

3.3.3 Die Möglichkeit des Pausierens während der Beobachtung

Neben der Einblendung der Kommentare anderer zeitgleich kommentierender Experten, erscheint auch die Möglichkeit, die Videos zu stoppen, während ein Feedback-Kommentar geschrieben wird, als eine wichtige Alternative. Je nach Trainingsstand der Trainierenden und der daraus folgenden Anzahl an Verhaltensweisen, die ein Experte kommentieren sollte, gerät der

schreibende Experte mehr oder weniger unter Zeitdruck. Im Grundaufbau der KNoTen-Station ist geplant, die Trainingsmitschnitte in Echtzeit[4] über das Internet zu übertragen. Dies bedeutet, dass, auch während ein Experte gerade einen Kommentar schreibt, die vier Videoübertragungen weiterlaufen. Um also nichts zu verpassen, während der Experte gerade einen Kommentar verfasst, muss sich er sich beeilen, um schnellstmöglich wieder seine Konzentration weg vom Eingabefeld und der Tastatur und hin zum Trainingsgeschehen zu lenken. Der Zeitdruck entsteht also dadurch, dass die Eingabe möglichst schnell erfolgen muss. Dieser Zeitdruck wirkt hierbei als Stressor auf den Experten (vgl. Bronner, 1974), was sich in einem subjektiv intensiven und unangenehmen Spannungsgefühl äußert (Greif, 1991).

Folgen des Zeitdrucks und Auswirkungen auf die Expertenleistung

Zeitdruck und der daraus resultierende Stress kann zu mehreren Folgen führen. Hierzu zählen eine eingeschränkte Kommunikation, die Ausführung einer geringeren Anzahl von Problemlöseschritten und es werden weniger Informationen in Betracht gezogen, wobei aber bei der Qualität der einbezogenen Informationen keine Abstriche zu verzeichnen sind. Es kann demnach davon ausgegangen werden, dass sich eine Person auf die wesentlichen und relevanten Informationen beschränkt (Bronner, 1974; Gilliland & Schmidt, 1993). Außerdem ist anzunehmen, dass Aussagen unter Zeitdruck kürzer formuliert werden (Wahlser, Blocker, Baus, Stopp & Speiser, 1998) und Menschen unter Zeitdruck generell versuchen, diesen zu kompensieren, indem sie ihre Arbeitsgeschwindigkeit erhöhen, worunter allerdings die Richtigkeit der Lösungen leidet (Bryan & Locke, 1967). Schließlich kann Zeitdruck als ein Auslöser für die Verwendung von Heuristiken gesehen werden (Raab, Unger & Unger, 2010). Wie im vorherigen Abschnitt erläutert, dienen Heuristiken der Vereinfachung und können dazu führen, dass lediglich sehr leicht aus dem Gedächtnis abrufbare Informationen in Entscheidungsprozesse einbezogen werden.

Nicht zu vergessen ist aber auch ein motivationaler Aspekt, der durch das Erleben von sehr hohem Zeitdruck entstehen kann. So könnte ein als extrem erlebter Zeitdruck dazu führen, dass der Betroffene die Ansicht entwickelt, dass die Zeit nicht ausreicht, um eine bestimmte Aufgabe zu lösen. Er verliert die Hoffnung, die Aufgabe bewältigen zu können, und gibt die Bearbeitung der Aufgabe vielleicht sogar ganz auf (Durham, Locke, Poon & McLeod, 2000).

Die angesprochenen Folgen von Zeitdruck könnten sich demnach stark auf die Leistung der Experten im KNoTen-Projekt auswirken, sollten sie bei ihrer Aufgabe des Kommentierens der Trainingsvideos Stress erleben. So ist anzunehmen, dass, wenn weniger Informationen gesam-

[4] Es ist damit zu rechnen, dass es bei der Übertragung der Inhalte zu minimalen Delays kommt, die aber im Sekundenbereich liegen.

melt werden, der Experte eventuell weniger Videos in seine Kommentierung einschließt. Vielmehr wird er sich auf ein oder zwei Videos konzentrieren, in denen er eher Anhaltspunkte für ein Feedback findet. Die Anzahl der einbezogenen Videos würde demnach sinken und damit auch die Möglichkeit, in den nun nicht betrachteten Aufnahmen Fehler zu entdecken und zu kommentieren, was die Anzahl der Kommentare insgesamt reduzieren würde. Letzteres könnte allerdings auch darauf zurückzuführen sein, dass, wie Bronner (1974) beschreibt, weniger kommuniziert wird, d.h. weniger Botschaften versendet werden. Sowohl die Anzahl der einbezogenen Videos als auch die Anzahl der verfassten Kommentare könnten das Gesamtfeedback nach Ilgen et al. (1979) weniger nützlich im Sinne eines geringeren Informationsgehaltes machen. Außerdem könnten die Häufigkeit des Feedbacks und vor allem die Konsistenz leiden, da möglicherweise nicht alle Vorkommnisse gleichmäßig mit Feedback versehen werden.

Des Weiteren könnte erlebter Zeitdruck dazu führen, dass die geschriebenen Kommentare kürzer werden, da sich der Experte auf die wesentlichen Aussagen konzentriert und keine zusätzlichen Beschreibungen oder Umschreibungen verwendet. Dies hätte gegebenenfalls Konsequenzen für die Ausführlichkeit der Kommentare, was sich wiederum auf die Spezifität des Feedbacks negativ niederschlägt. Nicht zu vernachlässigen ist darüber hinaus die steigende Fehlerhäufigkeit unter Stress. So könnte damit zu rechnen sein, dass auch falsche Einschätzungen durch den Experten getätigt werden.

Die einzig positive Auswirkung des Zeitdrucks kann hier in der Bearbeitungsgeschwindigkeit gesehen werden. Das adaptive Verhalten, das durch den Stressor Zeitdruck ausgelöst wird (Bronner, 1974), führt zu einer höheren Arbeitsgeschwindigkeit, also bspw. die Schnelligkeit, mit der ein Kommentar geschrieben wird. Ein schnellere Kommentierung kann hierbei den Vorteil bringen, dass sich der Experte nach dem Schreiben eines Kommentars schnell wieder dem Geschehen auf den Videoanzeigen zuwenden kann und somit die Gefahr reduziert wird weitere Vorkommnisse zu übersehen, die auftreten, während der Experte noch mit dem Schreiben beschäftigt ist.

Pausieren, um Zeitdruck zu reduzieren

Aufgrund dieser eher negativen Wirkungen eines als stark erlebten Zeitdrucks während der Beobachtung und Kommentierung des Trainingsgeschehens sollten Möglichkeiten in Erwägung gezogen werden, den Zeitdruck für die Experten zu reduzieren. Hierbei liegt es nahe, den Experten die Möglichkeit zu geben, die Beobachtung während des Schreibens eines Feedback-Kommentars zu stoppen, den Kommentar ohne Zeitdruck einzugeben und erst danach wieder mit der Beobachtung fortzufahren.

Diese Funktionalität könnte demnach sehr einfach die negativen Folgen durch Zeitdruck verhindern, würde aber zwangsläufig dazu führen, dass sich die Gesamtbearbeitungsdauer erhöht, was sich wiederum auf die Verzögerung des Feedbackgesprächs im Anschluss an das Training auswirkt, da der Experte insgesamt länger zum Beobachten und Kommentieren benötigt. Mit der „Pausieren"-Funktionalität würde der Experte seine Beobachtungen zwar in Echtzeit beginnen, aber mit jedem Kommentar einen weiteren Delay erzeugen, der auch die Trainingsauswertung weiter hinauszögert. Diese Verzögerung müsste in jedem Fall bei der zeitlichen und finanziellen Planung von Trainings dieser Art berücksichtigt werden.

Fazit

Angesichts der Ausführungen ergeben sich wie bei der ersten Variationsmöglichkeit, den zusätzlichen Expertenkommentaren, sowohl Vor- als auch Nachteile durch die zusätzliche Funktionalität. Eine zusätzliche Einbuße hinsichtlich der zeitlichen Abläufe könnte nur dann hingenommen werden, wenn die Reduktion des Zeitdrucks tatsächlich den beschriebenen negativen Auswirkungen von Stress vorbeugen kann. Es stellt sich damit folgende Forschungsfrage mit den entsprechenden Teilfragen:

Forschungsfrage 3:

Hilft eine zusätzlich eingefügte Funktionalität, die das Pausieren der Videos während des Kommentierens erlaubt, die Güte des Feedbacks im Vergleich zur Variante ohne diese Funktionalität zu erhöhen?

a) Welche Auswirkungen ergeben sich durch die Möglichkeit, zu pausieren, für die quantifizierbaren Merkmale der Leistung, also die Anzahl der Kommentare, die Anzahl der einbezogenen Videos, die Ausführlichkeit bzw. Länge der Kommentare, die Bearbeitungszeit der einzelnen Kommentare, sowie die Gesamtdauer des Feedbackprozesses?

b) Welche Auswirkungen ergeben sich durch die Möglichkeit, zu pausieren, für die qualitativen Merkmale der Leistung, also die Richtigkeit der Kommentare und die Verständlichkeit / Nachvollziehbarkeit der Kommentare?

c) Wie stark sind die zeitlichen Einbußen, die durch die zusätzliche Funktionalität des Pausierens entstehen?

3.3.4 Die Möglichkeit, einen auditiven Fokus zu setzen

Bedeutung der Tonübertragung während des Trainings

Während eines Notfalls ist es von essentieller Bedeutung, dass die in der Notlage befindlichen Personen effektiv miteinander kommunizieren können, um so in der Lage zu sein, arbeitsteilig lebensrettende Maßnahmen durchzuführen und sich einen zeitlichen Vorteil zu verschaffen. Aus diesem Grund sollte die Kommunikation während eines Notfalls ein fester Bestandteil eines umfassenden Notfall-Management-Trainings sein. Verwendete Trainingsinstrumente sollten die Möglichkeit einschließen, Kommunikation zu trainieren und Feedback darüber zu erhalten.

Daher beinhaltet der zuvor erläuterte Grundaufbau der KNoTen-Station neben den vier Kameraeinstellungen ebenfalls die Aufzeichnung und Übertragung der dazugehörigen Tonspuren, die es ermöglichen die Kommunikationsinhalte und Kommunikationsweisen der Trainierenden zu verfolgen. D.h., der feedbackgebende Experte sieht nicht nur, was die Teilnehmer während der komplexen Notfallübung tun, er hört es auch. Hierbei handelt es sich allerdings nicht wie bei der separaten Aufnahme und anschließenden Zusammenführung von Musikinstrumenten um ein harmonisches Zusammenklingen, sondern vielmehr um ein Durcheinander von Geräuschen, bei dem es zunächst schwerfällt, eine Tonspur der jeweiligen Aufnahme zuzuordnen. Der auditive Wahrnehmungsprozess, der von dem Experten auszuführen ist, beinhaltet demnach zunächst die Separation und Organisation der auditiven Informationen sowie das Herausfiltern der relevanten Informationen (Imhof, 2003) aus den vier Videos. Um diesen Vorgang der Zuordnung der Tonspuren zu den Videos abzukürzen, wäre es denkbar eine Funktionalität einzufügen, die es dem Experten erlaubt, die Vielzahl an präsentierten Geräuschen zu reduzieren und jeweils nur eine Tonspur, bspw. die, deren Video gerade betrachtet wird, zu aktivieren. Auf diese Weise würden die als Ablenkungen wirkenden anderen Geräusche wegfallen und der Experte könnte sich zugleich auf Bild und Ton aus einer Videoeinstellung konzentrieren.

Mit einer solchen Reduktion der in dem Moment nicht relevanten Geräusche könnten einige Vorteile verbunden sein. So konnte Kiefer (2002) zeigen, dass die Reduktion von Umgebungsgeräuschen (in seinem Fall Flughafendurchsagen) zu positiven Effekten in der Kommunikation von Versuchspersonen führte. So reduzierte sich die Zeit, die die Probanden für bestimmte Äußerungen brauchten, es gab weniger Sprechpausen und Fülllaute, die Probanden brauchten zum Versenden kommunikativer Botschaften weniger Worte, Silben und Sätze und benutzten weniger Wortfragmente (Teile von Wörtern, die nicht zu Ende gesprochen wurden). Darüber hinaus argumentieren Berti und Schröger (2003), dass jede Ablenkung an sich und die im Anschluss

daran erforderliche Reorientierung zurück zur eigentlichen Aufgabe mit Zeitkosten verbunden sind. Ohne die potentiellen Distraktoren (Ablenker) würden diese Kosten wegfallen. Lange (2005) postuliert des Weiteren, dass eine Unterdrückung irrelevanter Reize, also der Versuch, sich nicht ablenken zu lassen, zusätzliche Ressourcen beansprucht, die dann nicht mehr für andere kognitive Tätigkeiten zur Verfügung stehen. Dies würde zu einer Verschlechterung der Leistung führen, was in ihren Untersuchungen anhand der Richtigkeit der von Probanden gegebenen Antworten feststellbar war. Im Umkehrschluss könnte demnach eine Reduktion der Reize von vornherein eine solche Unterdrückungsarbeit überflüssig machen und es blieben mehr Ressourcen für die anderen kognitiven Aktivitäten.

Schließlich spricht für diese Erweiterung des Grundaufbaus der KNoTen-Station, dass die Kontrolle über die Nebengeräusche, die der Experte bspw. durch die eigenständige Auswahl der zu einer Zeit aktivierten Tonspur erhalten könnte, sich auf den Grad der gefühlten Belästigung und das Stresserleben durch die Vielzahl von Umgebungsgeräuschen auswirken kann (Kjellberg, Landström, Tesarz, Söderberg & Akerlund, 1996). Der durch diese Kontrolle gesenkte subjektive Level von Geräuschbelästigung kann sich auf das allgemeine Wohlbefinden bei der Aufgabenausführung auswirken (Nivison & Endresen, 1993).

Mögliche Effekte für die KNoTen-Station

Ausgehend von diesen Befunden könnten sich auch positive Effekte auf die Leistung der Experten im KNoTen-Projekt ergeben. So könnte es sein, wenn durch die Reduktion von auditiver Ablenkung schnelleres Kommunizieren möglich ist, dass auch die Experten schneller bei der Formulierung ihrer Feedbackkommentare sind und die Gesamtbearbeitungsdauer nicht nennenswert größer sein dürfte als die eigentliche Trainingsdauer. Letzeres bedeutet dann nichts anderes, als dass mit dieser Variante des Grundaufbaus keine zusätzlichen Kosten durch eine Verlängerung der Bearbeitungszeit entstehen sollten. Durch die zügigere Bearbeitung haben die Experten schneller wieder Zeit, sich auf die Beobachtung des Trainingsgeschehens zu konzentrieren, wohingegen bei einem langsameren Arbeiten eher die Gefahr bestünde, dass den Experten wichtige Vorkommnisse entgehen, während sie noch mit dem Kommentieren des letzten Vorfalls beschäftigt sind. Unter der Bedingung einer reduzierten Geräuschkulisse könnte mehr Zeit zu einer höheren Anzahl an Feedback-Kommentaren führen, da dem Experten weniger entgeht und er mehr Vorkommnisse beobachten kann. Die Variante könnte demnach Vorteile hinsichtlich der Häufigkeit und der Konsistenz des Feedbacks haben.

Der Argumentation Langes (2005) folgend sollte durch die Einsparung kognitiver Ressourcen durch überflüssig gewordene Unterdrückungsaktivitäten die Leistung der Experten auch hin-

sichtlich der Richtigkeit ihrer Einschätzungen steigen. Die Experten haben mehr Kapazitäten übrig, um alle relevanten Informationen einzubeziehen und sorgfältig abzuwägen. Zusätzlich wird die Fehlerwahrscheinlichkeit dadurch reduziert, dass die Möglichkeit ausgeschlossen wird, einen Kommunikationsstrang einem falschen Video zuzuordnen. Die vorgeschlagene Variation des Grundaufbaus brächte also auch Vorteile qualitativer Art.

Die bisherigen Befunde legen des Weiteren nahe, dass die Kommunikation ohne Störgeräusche präziser erfolgt. So fand Kiefer (2002) heraus, dass unter diesen Bedingungen – zumindest in der mündlichen Kommunikation – Aussagen weniger Wörter, Silben und Sätze beinhalten und weniger unbrauchbare Wortfragmente vorkommen. Es ist zu vermuten, dass auch die KNoTen-Experten ohne zusätzliche Ablenkung durch Geräusche ihr Feedback besser auf den Punkt bringen können. Hierin ist ein Vorteil hinsichtlich der Verständlichkeit bzw. eher hinsichtlich der Unmissverständlichkeit zu vermuten. Wenn die Experten ihre Eindrücke prägnant zusammenfassen können, sollte es während der Trainingsauswertung seltener zu Fehlinterpretationen kommen.

In der Kürze der Aussagen bzw. der fehlenden Ausführlichkeit kann allerdings auch ein Nachteil gesehen werden. So könnten wichtige zusätzliche Informationen, die sich beispielsweise in nun weggelassenen Nebensätzen oder zusätzlichen einzelnen Wörtern finden ließen, gegebenenfalls nicht aufgeschrieben werden, was einen geringeren Informationswert und damit eine geringere Nützlichkeit des Feedbacks für die Trainierenden zur Folge hätte. Es wird an dieser Stelle klar, dass nicht eindeutig gesagt werden kann, ob eine prägnante, kurze Ausdrucksweise oder ein ausführlicher Feedbackkommentar von größerem Wert für die Feedbacknehmer ist.

Schließlich besteht eine weitere Gefahr, die mit der Möglichkeit selbständig einen auditiven Fokus zu setzen, einhergeht. Durch das Einschalten nur einer Tonspur können den Experten wichtige Hinweise für relevante Vorkommnisse in den anderen Videomonitoren entgehen. So kann außerhalb des gerade auditiv fokussierten Videos nicht mehr gehört werden, ob es bspw. zu kommunikativen Auseinandersetzungen kommt. Es verbleibt lediglich ein tonloses Bild als einzige Informationsquelle, aus der Hinweise für solche rein auditiv feststellbaren Vorkommnisse nicht generiert werden können. Es besteht also das Risiko, dass weniger Vorfälle wahrgenommen werden können und demnach weniger Kommentare entstehen. Darüber hinaus fehlt durch die nun nicht mehr vorhandenen potenziellen Hinweisreize auf das Geschehen außerhalb des gerade Betrachteten die Gefahr, sich auch über eine längere Zeit – man wird ja nicht mehr abgelenkt – nur auf eine Anzeige, also einen Trainingsort zu fixieren. Demnach werden die anderen, potentiell ebenso wichtigen Trainingsgeschehen auf den anderen Anzeigen nicht mehr

in die Beobachtung und Kommentierung einbezogen. Hierunter kann erneut die Anzahl der geschriebenen Kommentare und damit die Feedbackhäufigkeit und Konsistenz, aber auch die Nützlichkeit des Feedbacks stark leiden, wenn bspw. einige Verhaltensweisen nur in einem Video, also bei einer Personengruppe kommentiert werden oder einige Vorkommnisse gar kein Feedback erhalten, weil sie sich in einem der nicht beachteten Videos abspielten.

Fazit

Wie bei den anderen Funktionalitäten, die zusätzlich zum Grundaufbau der KNoTen-Station denkbar sind, ergibt sich auch für die Möglichkeit, einen auditiven Fokus zu setzen, ein differenziertes Bild der Vor- und Nachteile. Für die eher praktisch angelegte Problemstellung, der Suche nach sinnvollen Zusatzfunktionalitäten für den bisher feststehenden Aufbau der KNoTen-Station, soll im Weiteren wieder auf explorativem Weg geklärt werden, welche Auswirkungen auf die Leistung der Experten sich tatsächlich ergeben bzw. wie sich die Variante, einen auditiven Fokus setzen zu können, auf einzelne Aspekte der Leistung der Experten auswirkt. Das führt zu folgenden Forschungsfragen:

Forschungsfrage 4:

Hilft eine zusätzlich eingefügte Funktionalität, die es erlaubt einen auditiven Fokus zu setzen, die Güte des Feedbacks im Vergleich zur Variante ohne diese Funktionalität zu erhöhen?

a) Welche Auswirkungen ergeben sich durch diese Möglichkeit für die quantifizierbaren Merkmale der Leistung, also die Anzahl der Kommentare, die Anzahl der einbezogenen Videos, die Ausführlichkeit bzw. Länge der Kommentare, die Bearbeitungszeit der einzelnen Kommentare sowie die Gesamtdauer des Feedbackprozesses?

b) Welche Auswirkungen ergeben sich durch diese Möglichkeit für die qualitativen Merkmale der Leistung, also die Richtigkeit der Kommentare und die Verständlichkeit / Nachvollziehbarkeit der Kommentare?

3.4 Ergänzende Forschungsfragen

Neben den zuvor diskutierten Forschungsfragen sollen in der vorliegenden Arbeit aber auch noch weitere Fragen geklärt werden. Im Hinblick auf die dauerhafte Gewinnung von Experten, die sich für das Feedbackgeben zur Verfügung stellen, sollte darauf geachtet werden, dass die Experten nicht überfordert werden und sich zumindest einigermaßen wohl fühlen bei Ihrer Tätigkeit. Daher soll diesen Fragen im Weiteren ebenfalls nachgegangen werden. Es ergeben sich folgende ergänzende Forschungsfragen:

Forschungsfrage 5:

Für wie schwierig halten diese Personen die Aufgabe, Feedbackkommentare mit Hilfe der KNoTen-Station zu erstellen?

Forschungsfrage 6:

Wie fühlen sich Personen, die mit der KNoTen-Station bzw. mit ihren möglichen Varianten arbeiten? Was sind die konkreten Auslöser für dieses Gefühl?

Forschungsfrage 7:

Welche zusätzlichen Funktionalitäten oder Veränderungen sehen die Nutzer der KNoTen-Station auf Basis erster Erfahrungen als nützlich an, um die Leistung der Experten weiter zu verbessern?

4 Methodik

4.1 Forschungsdesign

Da aufgrund der widersprüchlichen Forschungslage keine gerichteten Hypothesen ableitbar waren, werden die Forschungsfragen auf explorativem Wege überprüft (Bortz & Döring, 2006). Dabei ist eine Kombination verschiedener Erhebungsmethoden notwendig. Zum einen muss ein Weg gefunden werden, mit dem ermittelt werden kann, ob die beschriebenen Zusatzfunktionalitäten einen Einfluss auf die Leistung der Experten haben. Hierzu ist es zunächst notwendig, die Funktionalitäten softwaretechnisch umzusetzen. Beim Einsatz der Funktionalitäten müssen dann Einflüsse anderer Faktoren ausgeschaltet werden, um Veränderungen in der Leistung eindeutig auf die Manipulation der Funktionalitäten zurückführen zu können (Mittag, 2011). Aus diesem Grund wurde ein experimentelles Design zur Untersuchung der Forschungsfragen 1 bis 4 gewählt.

Zum anderen soll in der vorliegenden Studie geklärt werden, wie stark sich die Probanden während der Arbeit mit den unterschiedlichen Funktionalitäten beansprucht gefühlt haben. Hieraus sollen Rückschlüsse auf die Auswirkungen der verschiedenen Funktionalitäten auf die Probanden gezogen werden. Möglicherweise ergeben sich daraus Effekte für die Motivation und die Gewinnung von Experten für die Mitarbeit bei KNoTen-gestützten Trainingsprogrammen. Solche Daten, bei denen das Spektrum an möglichen Antworten aufgrund der Neuartigkeit des Untersuchungskontextes nicht absehbar ist, lassen sich am besten auf dem sehr offenen Weg der qualitativen Befragung erheben (Möhring & Schlütz, 2010).

Das experimentelle Design mit einem abschließenden Interview wurde schließlich um einige weitere Tests zur Erhebung von Kontrollvariablen ergänzt, die in Abschnitt 4.5 beschrieben werden.

4.2 Grundaufbau der Experimental-Software

Für den experimentellen Teil der Studie, in dem die im Forschungsprojekt entwickelte KNoTen-Station mit den anderen denkbaren Varianten verglichen werden soll, wurde die KNoTen-Station mit Hilfe einer eigens hierfür entwickelten Software vereinfacht nachgebildet. Diese Nachbildung simuliert die Anzeige der Videos auf dem Bildschirm eines Experten sowie die Kommentarfunktion. Auf den Einsatz spezieller Hardware, wie sie für die tatsächliche KNoTen-Station benötigt wird, wurde im Rahmen dieser Studie verzichtet, da dies angesichts einer Vielzahl von Versuchsleitern und Probanden sowohl mit einem erheblichen logistischen Aufwand als auch mit einem hohen Risiko für die Hardwarekomponenten verbunden gewesen wäre. Im Folgenden soll zunächst die Grundvariante der Software erläutert werden, die mit dem aktuellen Entwicklungsstand der KNoTen-Station weitgehend übereinstimmt und die in der vorliegenden Studie die Softwarevariante für die Kontrollgruppe darstellt. In der Studie wurde hierfür die Bezeichnung „Grundvariante“ verwendet. Im Anschluss daran wird die Umsetzung der in Kapitel 3 vorgestellten zusätzlichen Varianten erläutert, die aus systematischen Abwandlungen der Grundvariante hinsichtlich jeweils eines spezifischen Merkmals bestehen und die Software-Varianten für die vier Experimentalgruppen bilden. Dabei handelt es sich um eine Variante, in der Kommentare anderer Experten sichtbar sind (Bezeichnung des Szenarios: „Expertenkommentare“), eine Variante, in der ebenfalls Kommentare anderer sichtbar sind, die aber von einem Signalton begleitet werden (Bezeichnung: „Expertenkommentare mit Signal“), eine Variante, in der die Nutzer die Videos während der Bearbeitung anhalten dürfen (Bezeichnung: „Pausieren“) und zuletzt eine Variante, bei der die Nutzer aktiv steuern können, für welches Video der Ton eingeschaltet wird (Bezeichnung: „Ton variierbar“).

In der Nachbildung der KNoTen-Station wurden keine Aufzeichnungen von Notfalltrainingsprogrammen verwendet, da für deren Bearbeitung nur Experten in Frage gekommen wären. Daher wurde auf Videos von Straßenverkehrssituationen zurückgegriffen, für deren Verwendung mehrere Gründe sprachen. Zum einen waren die Videos unproblematisch erstellbar und sie enthielten Verhaltensweisen einer Vielzahl von Personen, die sich eindeutig als angemessen oder unangemessen einstufen lassen. Zum anderen kommt für diese Einstufung eine große Anzahl an Personen in Frage, weil die Teilnahme am Straßenverkehr für fast alle Menschen eine Alltagssituation darstellt. Es ist zu erwarten, dass ein großer Teil der Bevölkerung zumindest

über ein Minimum an Wissen und Erfahrung in diesem Bereich verfügt und in der Lage ist, Verhalten anderer Personen im Straßenverkehr hinsichtlich seiner Angemessenheit einzustufen.

Um bei der Durchführung der Versuche möglichst flexibel bleiben zu können, wurde eine offline arbeitende Software zur Nachbildung der KNoTen-Station gewählt, so dass die Versuchsleiter die Experimente ortsunabhängig mit der ihnen zur Verfügung stehenden Rechentechnik durchführen konnten. Die Software wurde auf CD-Rom oder USB-Sticks übergeben.

Um ein Mindestmaß an stabilen Versuchsbedingungen sicherzustellen, wurden die Versuchsleiter instruiert, die Experimente nicht mit Tablet-PCs oder Netbooks durchzuführen, da diese nur über sehr kleine Displays verfügen und im Vergleich zum Umgang mit herkömmlichen PCs oder normalgroßen Laptops über andere Eingabesysteme (Touchscreen oder verkleinerte, anders angeordnete Tastensätze) verfügen. Darüber hinaus wurden die Versuchsleiter dazu angehalten, eine Computer-Maus anzuschließen und den Browser Mozilla Firefox© in einer möglichst aktuellen Version zur Durchführung der Versuche zu verwenden.

Nach der Installation der Software auf den jeweiligen Endgeräten kann das Versuchsprogramm in einem Fenster des genannten Browsers gestartet werden. Hierbei erscheint zunächst auf dem Bildschirm ein Dialogfeld, in dem die Lautstärke der Videos und des in einer Variante verwendeten Signaltons sowie die Softwarevariante („Grundvariante“, „Expertenkommentare“, „Expertenkommentare mit Signal“, „Pausieren“ oder „Ton variierbar“) durch die Versuchsleiter festgelegt wird (s. Abbildung 4).

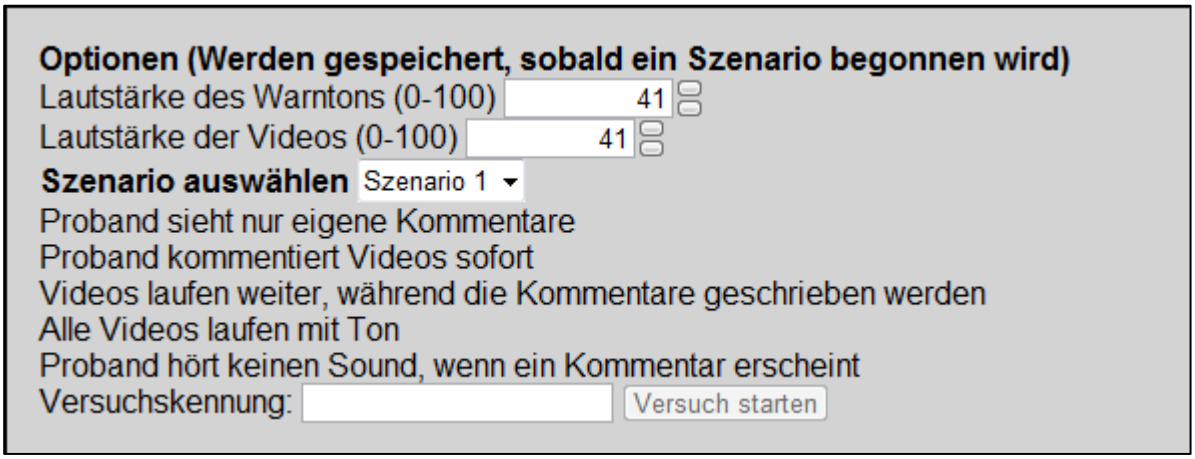


Abbildung 4: Dialogfeld für die Grundeinstellungen der Experimentalsoftware.

Darüber hinaus ist an dieser Stelle eine Versuchskennung einzugeben, die der eindeutigen Identifizierung der im Weiteren aufgezeichneten Daten dient. Nach der Vornahme dieser Einstellungen folgen Nutzerdialoge, die sich nun nur noch an die Probanden wenden.

Der erste Dialog, der für den Probanden eingeblendet wird, enthält die Anleitung zum Umgang mit der Software. Hierbei handelt es sich um einen Text, der sowohl die Aufgabenstellung für die Versuche als auch Hinweise zur Bedienung der Software beinhaltet und auf die jeweils ausgewählte Variante zugeschnitten ist. Abbildung 5 zeigt beispielhaft die Aufgabenstellung[5] der Grundvariante (Kontrollgruppe).

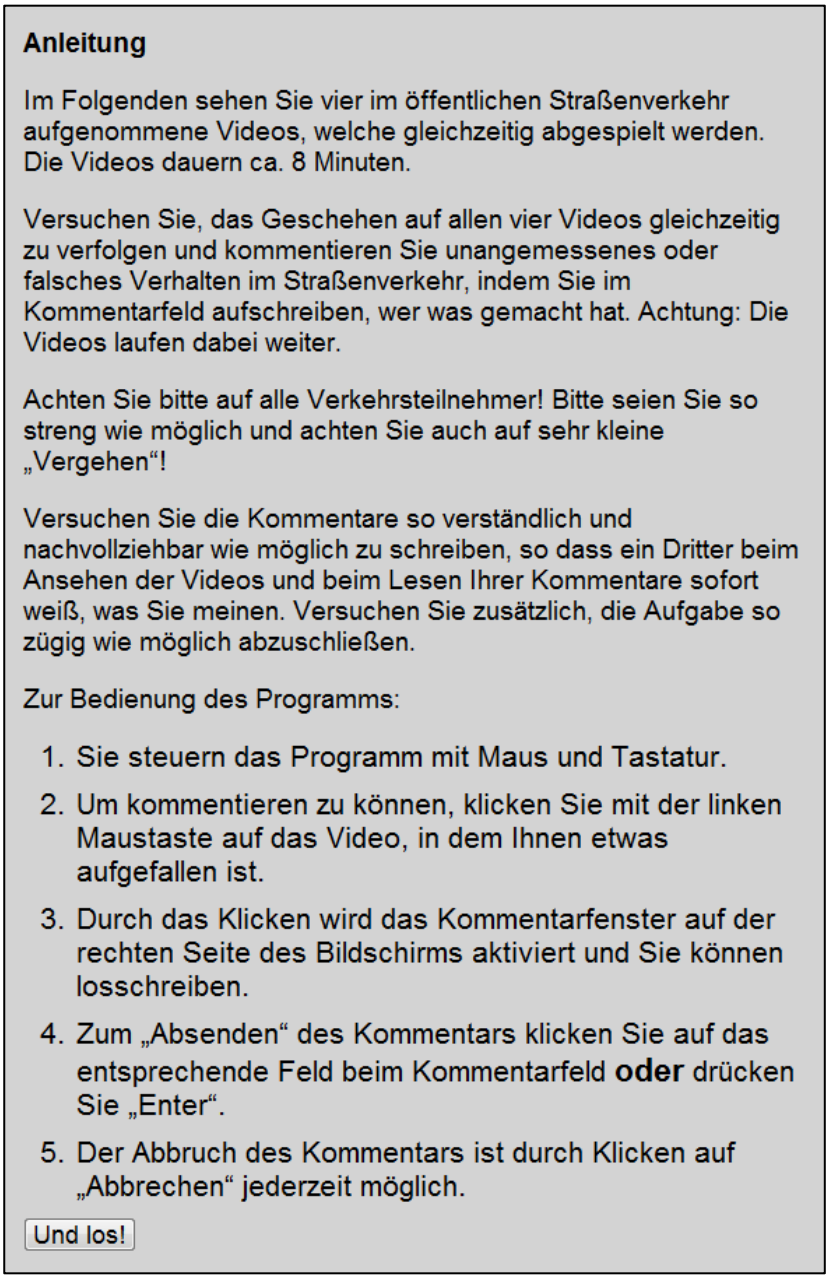


Abbildung 5: Anleitung für die Probanden der Kontrollgruppe.

Bei einem Mausklick auf den Button „Und los!“ wechselt der Bildschirm zur nächsten Anzeige, in der der Proband gebeten wird, seinen Namen einzugeben.

[5] Die konkreten Aufgabenstellungen für alle Varianten der nachgebildeten KNoTen-Station können bei den Autorinnen nachgefragt werden (Kontakt: tina.breyer@uni-rostock.de; anne.gutschmidt@uni-rostock.de).

Nach der Namenseingabe werden die Probanden gebeten mit einem Klick auf den „OK“-Button den Versuch zu starten. Mit dem Klick erscheint der in Abbildung 6 dargestellte Bearbeitungsbildschirm und die achtminütigen Videos inklusive der dazugehörigen Tonkanäle startet automatisch.

Abbildung 6: Screenshot des Bearbeitungsbildschirms in der Grundvariante.

Um einen Feedbackkommentar zu verfassen, muss das Video angeklickt werden, in dem die zu kommentierende Beobachtung gemacht wurde. Daraufhin wird das Eingabefeld am rechten oberen Bildschirmrand aktiviert[6] und der Nutzer kann seine Beobachtung schildern. Währenddessen laufen die Videos weiter (nur in der später näher zu erläuternden Variante „Pausieren“ ist dies nicht der Fall). Abgesendet wird ein fertig geschriebener Kommentar entweder durch Anklicken des Buttons „Absenden“ oder durch Drücken der Enter-Taste. Der Kommentar erscheint dann unterhalb des Eingabefeldes. Spätere Kommentare werden hierbei immer an oberster Stelle eingeblendet, so dass dort jeweils der aktuellste Kommentar sichtbar ist. Das Schreiben eines Kommentars kann bei Bedarf auch durch das Betätigen des entsprechenden Buttons unterhalb des Eingabefeldes abgebrochen werden. Nach Ende der vier Verkehrsvideos bzw. nach dem Absenden eines letzten, über das Ende der Videos hinausgehenden Kommentars wechselt die Bildschirmanzeige erneut und es werden die Worte „Vielen Dank“ eingeblendet.

Während der Bearbeitung werden durch die Software die eingegebenen Kommentare inklusive Zeitstempel, der jeweiligen Videonummer sowie der Gesamtversuchsdauer für jeden Probanden aufgezeichnet und als CSV-Dateien gespeichert.

[6] Um den Probanden anzuzeigen, dass sie nun ihren Kommentar eingeben können, wurde das Eingabefeld in diesem Moment farblich gelb hinterlegt.

4.3 Besonderheiten der Varianten des Grundaufbaus

Wie in Kapitel 3.3 argumentiert wurde, sind neben dem bereits beschriebenen Grundaufbau der KNoTen-Station und ihrer Softwarenachbildung weitere Varianten mit zusätzlichen Eigenschaften bzw. Funktionalitäten denkbar. Da keine eindeutige, theoretisch fundierte Prognose der Auswirkungen dieser Varianten auf die Leistung der mit der Station arbeitenden Personen ableitbar ist, wurden die vier Varianten ebenfalls jeweils in einem Teil der Software umgesetzt. Hierbei wurde darauf geachtet, dass in jeder Variante nur der interessierende Aspekt im Vergleich zum Grundaufbau verändert wurde, um sicher zu gehen, dass auftretende Veränderungen in den Leistungen bei der Aufgabenbearbeitung auch auf die jeweils veränderten Merkmale des Versuchs zurückzuführen sind. Die Besonderheiten der vier Teilprogramme, die in der vorliegenden Studie die Manipulationen der unabhängigen Variablen darstellen, sind in Tabelle 1 im Überblick dargestellt.

Tabelle 1: Übersicht zur Manipulation der unabhängigen Variablen

Eigenschaft / **Gruppe**	Kommentare anderer sichtbar	Signalton	Pausieren	Videoton
Grundvariante (Kontrollgruppe)	nein	ohne	nicht möglich	nicht variierbar
Expertenkommentare	**ja**	ohne	nicht möglich	nicht variierbar
Expertenkommentare mit Signal	**ja**	**mit**	nicht möglich	nicht variierbar
Pausieren	nein	ohne	**möglich**	nicht variierbar
Ton variierbar	nein	ohne	nicht möglich	**selbst steuerbar**

Die erste Erweiterung der Grundvariante der KNoTen-Station, bezeichnet als „Expertenkommentare“, besteht darin, den mit der KNoTen-Station arbeitenden Experten die Kommentare anderer Experten, die demselben Trainingsverlauf folgen, zugänglich zu machen. Für die Simulation dieser Funktionalität im Rahmen dieser Studie wurden zwei Veränderungen der Grundvariante vorgenommen. Zum einen werden zuvor festgelegte Kommentare zu bestimmten Zeiten im Verlauf der Bearbeitung eingeblendet. Diese enthalten Aussagen zu tatsächlich beobachtbaren Verhaltensweisen der Personen in den verwendeten Videos. Es werden insgesamt drei weitere Kommentatoren simuliert, deren Kommentare gemeinsam mit den Kommentaren der Probanden unter dem Eingabefeld angezeigt werden. Um das Bewusstsein für die angebli-

che Präsenz der anderen Kommentatoren zu erhöhen, wurde außerdem vor Beginn der Aufgabe eine zusätzliche Bildschirmanzeige in die Software integriert, die den Probanden suggerieren soll, dass in diesem Moment noch andere Kommentatoren hinzu geschaltet werden. Abbildung 7 zeigt einen Screenshot der verwendeten Anzeige. Die Namen der virtuellen Mitkommentatoren werden sukzessive eingeblendet. Anschließend erscheint der Schriftzug: „Achtung, Versuch beginnt!“, dem ein eingeblendeter Countdown von drei bis null folgt. Danach wird der normale Bearbeitungsbildschirm automatisch eingeblendet und die Videos werden abgespielt.

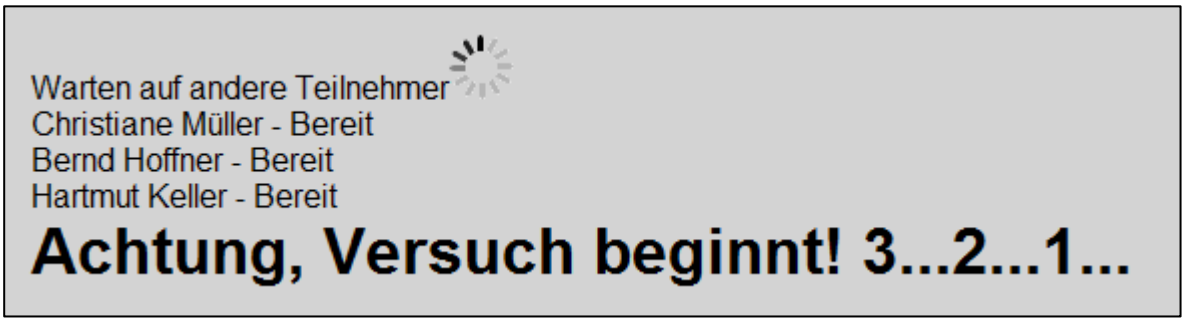


Abbildung 7: Screenshot vom suggerierten Hinzuschalten weiterer Kommentatoren.

Um die Probanden auf diese Variante angemessen vorzubereiten, wurde die Aufgabenstellung im Vergleich zur Grundvariante entsprechend verändert. So wurde den Probanden mitgeteilt, dass bei der Bearbeitung der Aufgabe weitere Kommentatoren durch das Programm simuliert werden. Sie sollten sich hierüber nicht wundern, sondern die Aufgabe weiterhin so konzentriert wie möglich erledigen.

Bei der zweiten Variante, bezeichnet als „Expertenkommentare mit Signal“, erklingt zusätzlich ein Signalton, sobald ein Kommentar eines der simulierten Kommentatoren eingeblendet wird. Das Signal ähnelt hierbei den akustischen Hinweisen, die beim Erhalt einer Nachricht in Chat- oder Messenger-Programmen wie z.B. Skype ertönen. Alle anderen Gegebenheiten und Instruktionen entsprechen der Variante „Expertenkommentare“.

Bei der Variante „Pausieren“ werden die vier Videos gestoppt, solange das Eingabefeld für die Kommentare aktiv ist. Durch den Mausklick auf eines der Videos kann somit die Beobachtung vollständig unterbrochen werden und die Nutzer können bis zum Absenden ihres Kommentars ohne zeitliche Begrenzung schreiben. Diese Möglichkeit wurde den Probanden in der Anleitung zur Aufgabenstellung und zur Bedienung des Programms beschrieben.

In der Variante „Ton variierbar“ war es möglich, durch die Bewegung der Maus über ein Video den Ton genau dieses Videos zu aktivieren. Die übrigen drei Videos hingegen wurden durch

diese Aktivierung ausgeschaltet. In der Anleitung zu dieser Variante wurde auf die Funktionalität, den Ton selbst steuern zu können hingewiesen, und die Probanden wurden gebeten, während der gesamten Bearbeitung mindesten einen Ton aktiviert zu lassen. Von den beschriebenen Abwandlungen abgesehen, stimmten jeweils alle übrigen Eigenschaften mit der Grundvariante überein.

4.4 Indikatoren zur Leistungsmessung und zur Erhebung subjektiver Eindrücke

Die vorliegende Studie basiert auf einer Kombination aus mehreren Erhebungsmethoden. Durch den Einsatz der nachgebildeten KNoTen-Station sowie der vier weiteren Varianten und durch die sich anschließende Befragung der Probanden wurden sowohl (weitgehend) objektive als auch subjektive Daten erhoben, die dabei helfen sollten die in Kapitel 3.3 und 3.4 aufgeworfenen Forschungsfragen zu beantworten. Im Folgenden wird zunächst beschrieben, anhand welcher objektiven Indikatoren die Leistung der Probanden eingeschätzt wird. Im darauf folgenden Abschnitt wird erläutert, wie die subjektiven Einschätzungen der Probanden bezüglich der erlebten Aufgabenschwierigkeit sowie des emotionalen Zustands nach der Arbeit mit der Experimentalsoftware in der post-experimentellen Befragung erhoben wurden.

4.4.1 Indikatoren der Leistung

Im Rahmen der Experimentalstudie wurden insgesamt sieben Indikatoren zum Vergleich der zuvor beschriebenen Varianten mit der Kontrollgruppe herangezogen. Hierbei ist zwischen Maßen zu unterscheiden, die rein mengenmäßig (quantitative Leistungsmaße) erfasst wurden und solchen, die sich auf die Güte der Kommentarinhalte (qualitative Leistungsmaße) beziehen. Hierzu zählen als rein quantitative Größen die Anzahl der Kommentare, die Anzahl der einbezogenen Videos, die Ausführlichkeit der Kommentare, die Tippgeschwindigkeit sowie die Gesamtdauer des Versuchs. Als qualitative Leistungsmaße wurden die Richtigkeit der geschriebenen Kommentare und deren Verständlichkeit für Dritte erhoben.

Anzahl der geschriebenen Kommentare

Als erstes rein quantitatives Leistungsmaß wurde die Anzahl der von einem Probanden geschriebenen Kommentare erhoben. Diese Größe ist für die Untersuchung von Bedeutung, da es im Sinne einer möglichst hohen Konsistenz ist, sämtliche kommentarbedürftigen Verhaltensweisen mit einem Feedback zu versehen. Da es sich bei dem in den Versuchen verwendeten Videomaterial um Aufnahmen von Straßenverkehrssituationen handelte, in denen insgesamt 92 gefährliche Verhaltensweisen der Verkehrsteilnehmer zu beobachten waren, hätten dementsprechend im Idealfall 92 Kommentare geschrieben werden müssen, um ein konsistentes Feed-

back zu geben. Für die statistische Auswertung wurde die Anzahl aller durch die Software registrierter Kommentare für jeden einzelnen Probanden betrachtet.

Anzahl der einbezogenen Videos

Um ein konsistentes Feedback abzugeben ist es wichtig, ähnliche Verhaltensweisen auch in unterschiedlichen Kontexten mit einem ähnlichen Feedback zu versehen. Damit dies möglich ist, stehen die zukünftigen Anwender der KNoTen-Station vor der Herausforderung, vier Videos gleichzeitig zu beobachten, in denen ähnliche Verhaltensweisen auftreten können. Um demnach die Konsistenz des Feedbacks gewährleisten zu können, ist es von Bedeutung, dass möglichst alle Videos bei der Kommentierung einbezogen werden. Aus diesem Grund wurde die Anzahl der Videos ausgezählt, zu denen jeder Proband Kommentare verfasst hatte.

Ausführlichkeit der Kommentare

Für ein informationsreiches und demnach möglichst nützliches Feedback ist es von Bedeutung, dass die Kommentare tatsächlich möglichst viele Informationen beinhalten und dementsprechend ausführlich sind. Um diese Größe zunächst rein quantitativ zu erheben, wurde die Anzahl der eingegebenen Zeichen (inklusive Leerzeichen) pro Kommentar erfasst und hieraus ein Durchschnittswert für jeden Probanden berechnet, der in die statistische Auswertung einging.

Tippgeschwindigkeit der Kommentare

Bei der Erstellung des Feedbacks unter Einsatz der KNoTen-Station ist es – gerade, wenn es keine Möglichkeit des Pausierens gibt – von besonderer Bedeutung, dass die Kommentare möglichst schnell verfasst werden, damit die Experten während der Eingabe möglichst wenig von den weiterlaufenden Videos verpassen. Um zu erheben, wie sich die unterschiedlichen Varianten auf die Tippgeschwindigkeit auswirken, wurde daher die durchschnittliche Tippgeschwindigkeit pro Proband erhoben.

Gesamtdauer der Bearbeitung

Als letztes quantitatives Maß wurde die Gesamtdauer der Versuche gemessen. Diese ist von Bedeutung, da erst nach vollständigem Abschluss der Kommentierung das Auswertungsgespräch mit den Trainingsteilnehmern erfolgen kann. Hierbei ist allerdings zu erwähnen, dass die Gesamtversuchsdauer bei der Grundvariante und in den Varianten „Expertenkommentare“, „Expertenkommentare mit Signal“ und „Ton variierbar“ kaum Unterschiede aufweisen dürfte, da diese Versionen automatisch, nach Ende der Videosequenzen bzw. nach dem Absenden des letzten über das Videoende hinausgehenden Kommentars, beendet werden. Die Variante „Pausieren“ hingegen könnte je nach Tippgeschwindigkeit und Anzahl der geschriebenen Kommentare zeitlich weit darüber hinausgehen. Um also zu ermitteln, um wie viel sich die Bearbei-

tungsdauer für die künftigen Anwender erhöht und damit gegebenenfalls zu höheren Honorarausgaben für die Experten oder einer stark verzögerten Auswertung nach Ende der Realübung führt, wurde die Dauer vom Start des Programms durch die Probanden bis zur Absendung des letzten Kommentars bzw. bis zu automatischen Abschaltung erhoben.

Richtigkeit

Zur Messung der Richtigkeit der Kommentare wurde im Vorfeld durch den Versuchsleiter in Zusammenarbeit mit sechs Studierenden der Universität Rostock eine Musterlösung erstellt, mit der die späteren Kommentare abgeglichen werden konnten. Hierfür sahen sich die insgesamt sieben Personen gemeinsam alle verwendeten Videos an und diskutierten beobachtete Verkehrsverstöße bis sie zu einer einstimmigen Meinung darüber gekommen sind, ob das Beobachtete einen Verkehrsverstoß darstellte oder nicht. Wenn Ersteres der Fall war, wurde die Zeit im Video gemeinsam mit einer Kurzbeschreibung des Verstoßes dokumentiert.

Beim Abgleich der Kommentare der Versuchspersonen mit der Musterlösung wurden die Kommentare für jedes der vier verwendeten Videos nach ihrem Zeitstempel sortiert und durch die Versuchsleiter hinsichtlich ihrer Übereinstimmung mit der Musterlösung beurteilt. Für einen Kommentar, der der Vorlage zeitlich ungefähr entsprach und aufgrund seines Inhalts eindeutig einem Vorkommnis in der Musterlösung zugeordnet werden konnte, wurde ein Punkt vergeben. In den Fällen, in denen eine Zuordnung nicht möglich war, wurden keine Punkte vergeben. Für die statistische Auswertung hinsichtlich der Richtigkeit der Kommentare wurde schließlich die Summe der richtigen Kommentare je Proband herangezogen. Die finale Punktvergabe wurde durch die Versuchsleiterinnen vorgenommen und beinhaltete bei der Wertung einen gewissen Ermessensspielraum, so dass einschränkend zu erwähnen ist, dass es sich bei der Einstufung der Richtigkeit der Kommentare lediglich um kein vollständig objektives Verfahren handelte.

Verständlichkeit der Kommentare

Unter Verständlichkeit wird hier die Nachvollziehbarkeit bzw. der Wiedererkennungswert eines Kommentars beim Betrachten der Videos für außenstehende Dritte verstanden. Ist ein Text verständlich, ist es für Personen, die einen Kommentar lesen, möglich, mit Hilfe des entsprechenden Videos eindeutig nachzuvollziehen, was der Autor des Kommentars ausdrücken wollte. Dies ist von besonderer Bedeutung für das im KNoTen-Projekt konzipierte Feedback-Gespräch nach Beendigung der Realübung. Hierfür stehen dem Gesprächsleiter und den Trainierten nicht mehr die Experten selbst, sondern nur noch die von ihnen verfassten Kommentare sowie die dazugehörigen Videosequenzen zur Verfügung. Wenn ein Kommentar unverständlich ist – es

kann also nicht nachvollzogen werden, was der Experte beim Verfassen aussagen wollte – besitzt der Kommentar keinen Informationswert für die Trainierenden.

Um die Verständlichkeit der von den Probanden verfassten Kommentare zu messen, wurde ein eigenes Scoring-System entwickelt, da der Rückgriff auf bestehende Verfahren wie bspw. das Hamburger Verständlichkeitskonzept nach Langer, Schulz von Thun und Tausch (1981) ungeeignet erschien. Letzteres kann zur Bewertung von längeren Texten eingesetzt werden, wohingegen es sich bei der KNoTen-Station überwiegend um sehr kurze Texte in Form der Kommentare handelt. Zur Entwicklung des hier genutzten Bewertungsschemas wurden per Zufall aus jeder Untersuchungsgruppe zwei Probanden ausgewählt. Die Kommentare dieser Probanden wurden unabhängig voneinander von zwei Auswertenden hinsichtlich auftretender Auffälligkeiten in Bezug auf ihre Verständlichkeit analysiert. So wurde induktiv nach Komponenten gesucht, die dazu beitragen, dass der Leser Anhaltspunkte über die Sichtweise des Kommentierenden erhält, diese nachvollziehen kann, sie also versteht. Gesucht wurde nach Merkmalen der verwendeten Sprache oder der Funktion von Satzbausteinen. Bei der anschließenden gemeinsamen Auswertung dieser ersten Sichtung wurde das final genutzte Bewertungsschema festgelegt.

Das Bewertungsschema basierte auf der Vergabe von Plus- und Minuspunkten für bestimmte Inhalte, die die Nachvollziehbarkeit des Geschriebenen förderten oder reduzierten. Pluspunkte wurden immer dann erteilt, wenn:

- genau nachvollzogen werden konnte, welche Personen im Video gemeint waren;
- klar wurde, welches Verhalten als Verstoß gegen das richtige Verhalten im Straßenverkehr gewertet wurde (z.B. „gehen bei rot“);
- zusätzliche Informationen zur gemeinten Person im Video gegeben wurden (z.B. „person in weißer Jacke“);
- zusätzliche Informationen zum im Video gezeigten Verhalten gegeben wurden (z.B. „überqueren eilig“);
- Textpassagen vorhanden waren, die zusätzliche, subjektive Einschätzungen und Beschreibungen des Gesehenen durch den Experten enthielten (z.B. „ungeeignete Stelle“ oder „nutzt als Abkürzung“);
- Textpassagen enthalten waren, die Anhaltspunkte zur räumlichen Orientierung im Video beinhalteten (z.B. „im Hintergrund“).

Beinhaltete ein Kommentar mehrere dieser positiven Merkmale, wurden dem Auftreten entsprechend viele Pluspunkte erteilt. Ähnlich wurde bei der Vergabe der Minuspunkte verfahren. Diese wurden für einen Kommentar erteilt, wenn die Statements Folgendes beinhalteten:

- unbekannte bzw. nicht gängige Abkürzungen (z.B. „ff“ für Fahrradfahrer);
- lediglich Verweise auf zuvor geschriebene Kommentare (z.B. „dito“ oder „siehe oben“);
- nicht eindeutig interpretierbare Aussagen (z.B. ironische Kommentare wie „professioneller Wendevorgang“ bei einer unerlaubten Autowendung im Video).

Nach Entwicklung der Scoring-Regeln wurde das gesamte Textmaterial von zwei unabhängigen Ratern, die das Material zuvor nicht gesichtet hatten, mit den Plus- und Minuspunkten versehen. Zur statistischen Auswertung wurde die Differenz aus Plus- und Minuspunkten für jeden Probanden gebildet.

4.4.2 Erhebung der subjektiven Eindrücke

Zur Beantwortung der Forschungsfragen 4 bis 7 wurde nach der Arbeit mit der Experimentalsoftware jeweils ein kurzes Interview mit den Probanden geführt. Der zugrundeliegende Interviewleitfaden befindet sich im Anhang der Arbeit. Die mündlichen Befragungen wurden als Tondateien aufgezeichnet, die anschließend transkribiert wurden.

Um die erlebte Aufgabenschwierigkeit (Fragestellung 5) zu ermitteln, wurden die Probanden gebeten auf einer fünfstufigen Likert-Skala einzuschätzen, wie schwierig sie die Aufgabe des Kommentierens fanden. Die Verwendung einer geschlossenen Fragestellung hatte hierbei den Vorteil, durch einen einheitlichen Bezugsrahmen das Antwortverhalten der Probanden im Sinne der Fragestellung einzugrenzen, was zum einen die Validität erhöht und zum anderen zu einer besseren Vergleichbarkeit der Daten führt (Möhring & Schlütz, 2010).

Die übrigen zu untersuchenden Fragen hingegen wurden offen gestellt, was sich anbietet, wenn das Spektrum der möglichen Antworten zum Zeitpunkt der Befragung noch nicht bekannt ist und eine Beeinflussung der Interviewten vermieden werden soll (Möhring & Schlütz, 2010). Die Gefühlslage der Bearbeitung wurde demnach durch die Frage: Wie haben Sie sich insbesondere bei der Bearbeitung der Kommentaraufgabe gefühlt? Hierbei wurden die Interviewer im Rahmen der Interviewerschulung angewiesen auch die folgenden Nachfragen zu stellen: „Wieso haben Sie sich so gefühlt?“ bzw. „Was war der konkrete Auslöser?“, um möglichst tiefgründige Antworten zu erhalten.

Um Verbesserungsvorschläge zu ermitteln, wurden den Probanden die Fragen gestellt: „Wie müsste das Programm der letzten Aufgabe verändert werden, damit Sie die gestellte Aufgabe *noch besser* lösen könnten?“ und „Wie müsste das Programm der letzten Aufgabe verändert werden, damit Sie die gestellte Aufgabe *noch schneller* lösen könnten?“

4.5 Untersuchungsablauf und weitere Erhebungsinstrumente

Die experimentelle Studie wurde im Rahmen einer Lehrveranstaltung mit Studierenden der wirtschafts- und sozialwissenschaftlichen Fakultät der Universität Rostock im Zeitraum vom 29. November 2012 bis zum 8. Januar 2013 durchgeführt. Die involvierten Studierenden rekrutierten Probanden für die Versuche aus ihrem Bekannten- und Verwandtenkreis, wobei jeder der freiwilligen Teilnehmer der Kontrollgruppe (Grundvariante) bzw. den vier weiteren Versuchsgruppen zufällig zugewiesen wurden (Bortz & Döring, 2006; Mittag, 2011). Hierdurch sollte sichergestellt werden, dass gegebenenfalls vorhandene, personengebundene Störvariablen in der Kontrollgruppe und in den Experimentalgruppen mittels des statistischen Fehlerausgleichs kompensiert werden. Aus diesem Grund wurde zusätzlich darauf geachtet, dass jeder Untersuchungsgruppe mindestens 20 Versuchspersonen zugewiesen wurden (Bortz & Döring, 2006).

Um während der Untersuchung die Bedingungen für alle Probanden konstant zu halten (Bortz & Döring, 2006), wurde im Vorfeld der Untersuchung ein Handbuch[7] zum Versuchsablauf ausgehändigt, in dem die Versuchsleiter sämtliche Informationen zur Durchführung der Datenerhebung nachlesen konnten. Darüber hinaus wurde eine Schulung aller beteiligten Versuchsleiter durchgeführt, in der sie den Umgang mit der Experimentalsoftware erlernten.

Der Ablauf jeder Untersuchungssitzung mit den Probanden enthielt folgende Schritte:

1. Durchführung eines sogenannten „Tipptests“
2. Durchführung eines Konzentrationstests nach Connelly, Hasher und Zacks (1991)
3. Durchführung der Hauptuntersuchung unter Einsatz einer Variante der Experimentalsoftware.
4. Durchführung eines Interviews sowie abschließendes Ausfüllen eines Fragebogens.

[7] Das vollständige Handbuch kann bei den Untersuchungsleiterinnen unter tina.breyer@uni-rostock.de und anne.gutschmidt@uni-rostock.de angefordert werden.

Der Tipptest

Für die Durchführung des Tipptests wurde eine eigens für diesen Zweck erstellte Software eingesetzt, bei der die Probanden einen vorgegebenen Text fehlerfrei abtippen mussten. Abbildung 8 zeigt einen Screenshot der verwendeten Software.

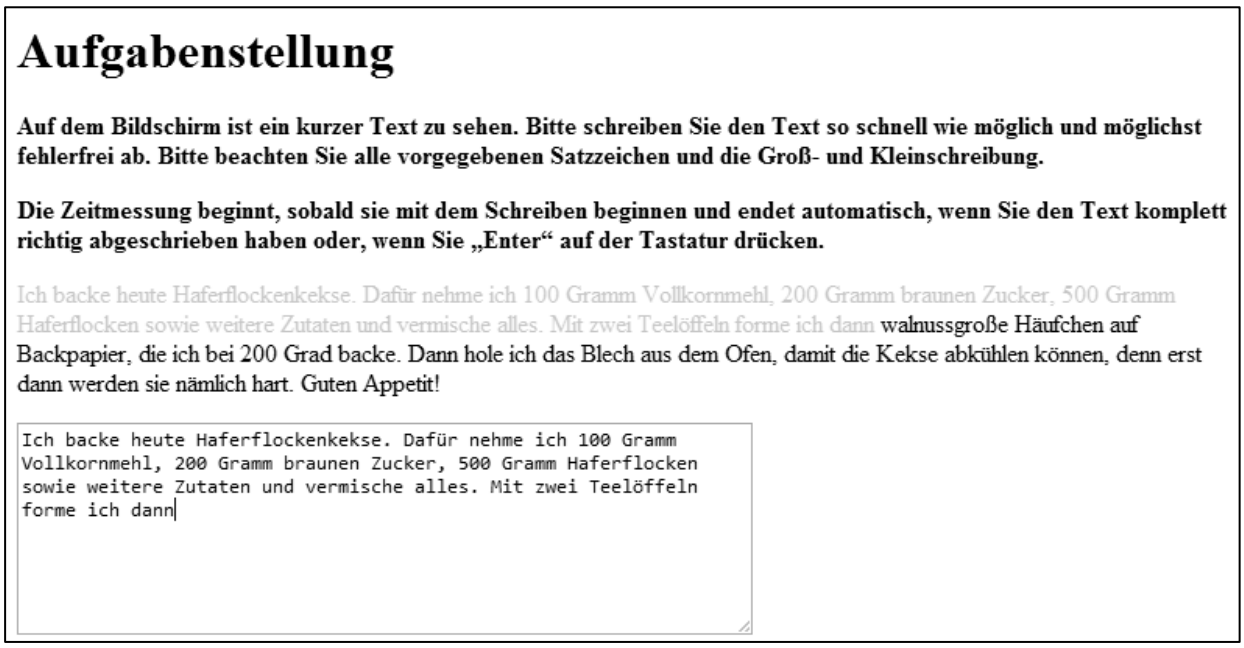


Abbildung 8: Screenshot der Tipptest- Software.

Der verwendete Text enthielt große und kleine Buchstaben, Zahlen, Satz- sowie Leerzeichen, wobei aber darauf geachtet wurde, dass nicht zu schwierige Worte benutzt wurden. Der Test wurde durch den Probanden mit der Eingabe des ersten Buchstabens im dafür vorgesehenen Eingabefeld begonnen. Sobald Zeichen eingegeben waren, wurde den Versuchspersonen durch eine grüne Färbung des Ausgangstextes signalisiert, dass sie bis zu diesem Punkt fehlerfrei abgeschrieben hatten. Eine rote Einfärbung sollte die Probanden darauf aufmerksam machen, dass ihnen ein Fehler unterlaufen war. Um das Programm abzuschließen, musste der gesamte Text vollständig und fehlerfrei im Eingabefeld eingeben werden. Mit der Eingabe des letzten richtigen Zeichens bzw. mit Korrektur des letzten vorhandenen Fehlers stoppte das Programm automatisch und gab die Dauer der Bearbeitung in Sekunden, die Anzahl der vorgenommenen Korrekturen sowie die durchschnittliche Tippgeschwindigkeit in Sekunden und Minuten an. Sobald diese Werte auf dem Bildschirm angezeigt wurden, wurden sie vom Versuchsleiter auf dem Fragebogen notiert. Mit dem Fragebogen wurden im weiteren Verlauf u. a. auch die soziodemografischen Daten aufgenommen.

Der Konzentrationstest

Um die allgemeine Konzentrationsfähigkeit der Probanden zu ermitteln, wurde ein zweiteiliger Test nach Connelly et al. (1991) durch geführt. Dieser bestand zum einen aus einer Vorleseaufgabe und zum anderen aus der Beantwortung von inhaltlichen Multiple Choice-Fragen zum vorgelesenen Text. Der von Connelly et al. (1991) verwendete Text über eine archäologische Ausgrabung wurde hierfür in Deutsche übersetzt. Abbildung 9 zeigt einen Ausschnitt des verwendeten Materials.

Die Ausgrabung
Die Autofahrt Fluss *wurde holprig* Jeep *jetzt wo* der religiöse *Georg* religiös *die* Grabwerkzeuge *Hauptstraße verließ, um den* religiösen *Feldweg zu benutzen. Er* Grabwerkzeuge Fluss *musste nicht in die* Jeep *Universität, weil er* Grabwerkzeuge Fluss *während der* religiösen *Sommerpause* Jeep *nicht vor Ort sein muss. Er war* Jeep Fluss *froh* Grabwerkzeuge *aus den* Jeep *stickigen* religiöse *Büros* Flüsse *der Archäologieabteilung* religiös *raus zu kommen* Fluss Jeep *und im* Grabwerkzeuge *Feld forschen zu können. Er* Fluss *mochte es schon immer* religiöse *an archäologischen* Grabwerkzeuge *Ausgrabungen* religiösen *teilzunehmen* Jeep.

Abbildung 9: Ausschnitt der deutschen Version des Konzentrationstestmaterials.

Die Probanden wurden aufgefordert, nur die kursiv gedruckten Worte laut vorzulesen. Dies sollte möglichst fehlerfrei und so schnell wie möglich erfolgen. Wie in Abbildung 9 ersichtlich ist, kamen im Text aber auch andere, nicht kursiv gedruckte Wörter vor, die aufgrund ihrer semantischen Bedeutung eine starke Verbindung zum vorzulesenden Text aufwiesen und deshalb als Distraktoren wirkten. Das Vorlesen wurde aufgezeichnet, so dass sowohl die Richtigkeit als auch die Dauer der Bearbeitung im Nachhinein feststellbar waren.

Im Anschluss an das Vorlesen wurde den Probanden das Textblatt wieder entzogen und sie bekamen stattdessen ein Blatt mit vier Multiple Choice-Fragen, die sich auf den zuvor vorgelesenen Text bezogen. Abbildung 10 gibt ein Beispiel für diese Fragen. Die Probanden wurden gebeten die jeweils richtige Antwort anzukreuzen, wobei die Schwierigkeit darin bestand, dass einige der Antwortalternativen den immer wieder im Text aufgetretenen Distraktoren entsprachen. Bei der Auswertung wurde die Anzahl der richtig beantworteten Multiple Choice-Fragen herangezogen.

Sowohl die erhobenen Daten des Tipptests als auch die Daten aus dem Konzentrationstest gingen in die Überprüfung der Vergleichbarkeit der Teilstichproben ein und dienten somit als Kontrollvariablen.

Womit fuhr Georg?	
❒	Auto
❒	Jeep
❒	Zug
❒	Bus
❒	Motorrad
❒	Traktor

Abbildung 10: Beispiel für eine Multiple Choice-Frage im Konzentrationstest.

Die Kommentaraufgabe

Nachdem eingangs erste Kontrollvariablen erhoben wurden, begann die eigentliche experimentelle Untersuchung, in der die Probanden mit der ihnen zugelosten Variante der Experimentalsoftware arbeiten sollten.

Um die Versuchsbedingungen bis auf die gezielte Variation der Softwarevarianten für alle Probanden einheitlich zu gestalten, wurden den Versuchspersonen unabhängig von den jeweiligen Versuchsleitern die Instruktionen zur Bearbeitung der Aufgabe in Textform durch die Software vermittelt (Bortz & Döring, 2006). Die grundlegende und für alle Versuchsgruppen einheitliche Anweisung bestand darin, dass die Probanden versuchen sollten, das Geschehen auf den vier gezeigten Videos gleichzeitig zu verfolgen und unangemessenes oder falsches Verhalten im Straßenverkehr zu kommentieren, indem sie aufschrieben, welche Person welches Verhalten gezeigt hatte. Die Instruktionen wurden je nach vorliegender Variante um Hinweise zur Steuerung der Software ergänzt (s. Kap. 4.2 und 4.3).

Nach Beendigung der Bearbeitung und einer anschließenden Aktualisierung des Browser-Fensters konnten die während der Bearbeitung aufgezeichneten Kommentare und andere Messgrößen schließlich abgespeichert werden.

Interview mit abschließendem Fragebogen

Im Anschluss an den Hauptteil der Untersuchung wurde ein kurzes Interview durchgeführt, in dem die Fragen zur Beantwortung der Forschungsfragen 5 bis 7 gestellt wurden. Die Interviews wurden aufgezeichnet, transkribiert und im Fall der Frage zur Aufgabenschwierigkeit mittels statistischer Testverfahren analysiert. Die Fragen zur Gefühlslage während der Bearbeitung und zu den Verbesserungsvorschlägen wurden mittels qualitativer Inhaltsanalyse (s. Kapitel 4.7) ausgewertet. Ein Fragebogen, der anschließend mit den Probanden bearbeitet wurde, diente der

Erhebung der soziodemografischen Daten und weiterer subjektiver Einschätzungen für die Stichprobenbeschreibung sowie zur Überprüfung der Vergleichbarkeit der Teilstichproben hinsichtlich wesentlicher Merkmale wie z.B. der Altersstruktur oder allgemeinen Computerkenntnissen. Der Interviewleitfaden sowie der verwendete Fragebogen findet sich in Anhang 1.

Im Anschluss an die Untersuchung wurde den Probanden für ihre Mitwirkung gedankt und gegebenenfalls wurden weitere Fragen zur durchgeführten Studie beantwortet.

4.6 Methoden der statistischen Auswertung

Um festzustellen, ob es hinsichtlich der abhängigen Variablen – im vorliegenden Fall verschiedene Facetten der Leistung der Probanden beim Erstellen von Feedbackkommentaren – signifikante Unterschiede zwischen der Kontrollgruppe und einer der Gruppen mit alternativer Softwareversion gibt, wurden Mittelwertvergleiche mittels t-Tests durchgeführt, sofern die dafür notwendigen Vorbedingungen der Normalverteilung und der Varianzhomogenität erfüllt waren (Bortz, 2005). Hierfür wurden die Daten, die durch die Experimentalsoftware aufgezeichnet wurden, verwendet. Die in Kapitel 4.4.1 beschriebenen Variablen wurden für die insgesamt fünf zu untersuchenden Gruppen mit dem Shapiro-Wilk-Test auf Normalverteilung getestet, eine Alternative zum Kolmogorov-Smirnov-Test bei Stichprobengrößen kleiner als 30 (Bühner & Ziegler, 2009). Bei einem Signifikanzwert von ≤ .05 wurde davon ausgegangen, dass keine Normalverteilung vorliegt. Auf Varianzhomogenität wurde mit dem Levene-Test geprüft (Bühner & Ziegler, 2009), wobei ein Signifikanzwert von .05 als Grenze gewählt wurde, deren Unterschreitung gegen Varianzhomogenität spricht (≤ .05). Waren diese Vorbedingungen erfüllt, so wurde der t-Test angewendet, wobei ein nicht auf Zufall basierender Unterschied zwischen den jeweils getesteten Gruppen nur dann angenommen wurde, wenn die Irrtumswahrscheinlichkeit unter oder bei genau fünf Prozent lag. Lag Normalverteilung vor, aber keine Varianzhomogenität, wurde der t-Test unter Anwendung einer Freiheitsgradkorrektur durchgeführt (Rasch, Friese, Hofmann, Naumann, 2010).

Alternativ zum t-Test wurde der U-Test von Mann-Whitney angewandt, wenn alle Vorbedingungen für einen t-Test nicht erfüllt waren (Bortz & Lienert, 2008). Dieser verteilungsfreie Test wird bei rangskalierten Daten angewandt. Bortz (2005) empfiehlt diesen Test aber auch für intervallskalierte Daten, wenn die Vorbedingungen des t-Test nicht erfüllt sind und es sich vor allem um kleinere Stichproben handelt. Auch hier wurde für die Signifikanz ein Grenzwert von .05 festgelegt, so dass nur Werte ≤ .05 einen signifikanten Unterschied anzeigen.

4.7 Qualitative Auswertung

Um die Forschungsfragen 6 und 7 (Gefühlslage während der Arbeit mit der Experimentalsoftware und Verbesserungsvorschläge) zu beantworten, wurde im Anschluss an die experimentelle Studie mit jedem Probanden ein Interview geführt. Diese Gespräche wurden aufgezeichnet und transkribiert. Um aus den nun vorliegenden schriftlichen Daten Erkenntnisse gewinnen zu können, wurde sich der Methode der qualitativen Inhaltsanalyse (vgl. dazu Früh, 2007; Mayring 2008) bedient.

Hierfür wurde zunächst in den Antworten auf die gestellten Fragen nach relevanten Kodiereinheiten gesucht. Im Falle der Frage nach dem Gefühl während der Bearbeitung dienten hierbei von den Probanden genannte Adjektive als zentrale Orientierungspunkte. Bei der Frage nach Verbesserungsvorschlägen wurden solche Teilaussagen zu einer Kodiereinheit zusammengefasst, die sich genau auf eine mögliche zusätzliche Verbesserung oder bspw. eine zusätzlich denkbare Funktionalität bezogen. So kam es vor, dass ein Proband auch mehrere Verbesserungsvorschläge äußerte, die im Weiteren unterschiedlichen Kategorien zugeordnet wurden.

Im nächsten Schritt wurden für beide Fragen diejenigen Kodiereinheiten, die eine inhaltliche Ähnlichkeit aufwiesen, zu einer Kategorie zusammengefasst. Aufgrund der Vielzahl der so identifizierten Kategorien sowohl zur sechsten als auch zur siebten Forschungsfrage, wurde mit Hilfe einer anschließenden induktiven Oberkategorienbildung durch eine Zusammenfassung inhaltlich noch sehr ähnlicher Kategorien eine besser handhabbare Kategorienzahl erreicht. Die Ergebnisse der qualitativen Auswertung finden sich in Kapitel 5.3.2 und 5.3.3.

5 Ergebnisse

5.1 Zur Stichprobe

5.1.1 Soziodemografie in den Versuchsgruppen

An der Studie nahmen insgesamt 129 Personen teil, die per Losverfahren den fünf Versuchsgruppen zufällig zugeordnet wurden. Die Kontrollgruppe, in der die Grundvariante der KNo-Ten-Station untersucht wurde, umfasste 26 Personen, der Gruppe, in der Kommentare anderer Teilnehmer sichtbar waren („Expertenkommentare“), wurden 28 Personen zugeteilt und die Gruppen, in denen Kommentare anderer begleitet von einem Signalton eingeblendet wurden („Expertenkommentare mit Signal“), umfasste 23 Personen. Die Gruppe „Pausieren“ umfasste insgesamt 27 Personen, und in der Gruppe mit manueller Steuerung des Tons („Ton variierbar“) wurden 25 Personen untersucht. Damit waren die Gruppen nahezu gleich besetzt (s. Tabelle 2).

Tabelle 2: Anzahl der Probanden in den der Kontrollgruppe und in den Versuchsgruppen

Variante	Grund-variante	Experten-kommentare	Experten-kommentare mit Signal	Pausieren	Ton variierbar	Gesamt
Anzahl der Probanden	26	28	23	27	25	129

Insgesamt nahmen 57 Frauen und 72 Männer an der Studie teil. Abbildung 11 zeigt, dass die Anteile von Männern und Frauen in den einzelnen Gruppen ähnlich verteilt waren.

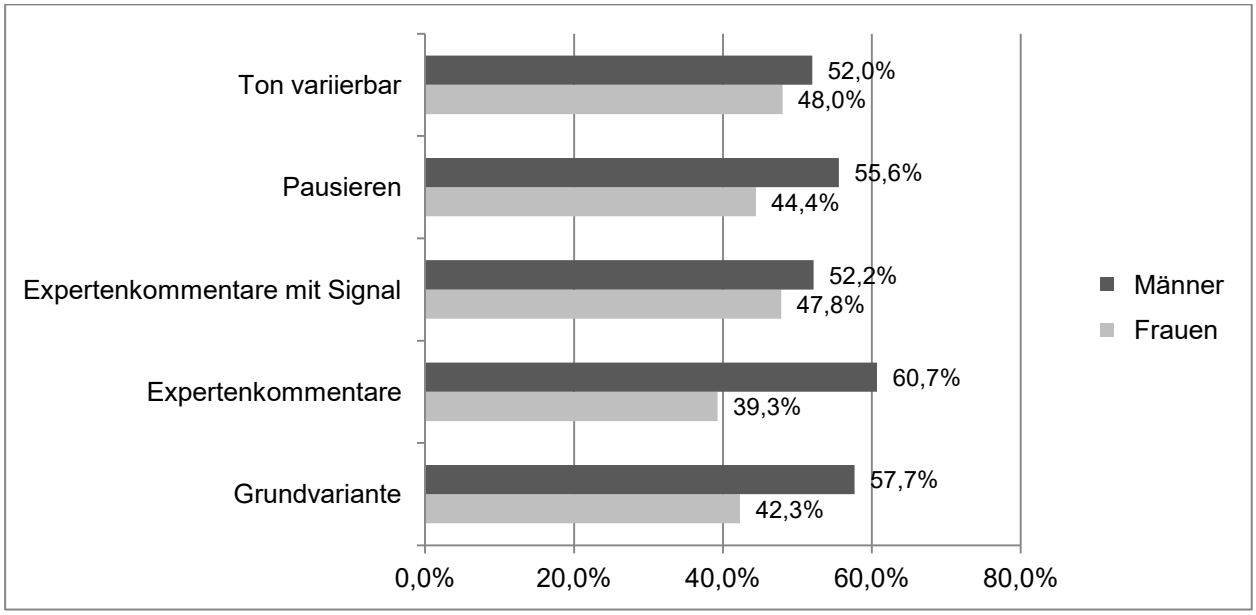


Abbildung 11: Geschlechterverteilung in den Experimentalgruppen.

Die Probanden waren in der Gesamtheit zwischen 18 und 62 Jahren alt mit einem Durchschnittsalter von 30,1 Jahren und einer Standardabweichung von 11,2 Jahren. Die Probanden der Gruppe mit der Grundvariante der KNoTen-Station waren im Durchschnitt 33,1 Jahre alt (s = 13,5 Jahre), die Probanden in der Gruppe der Variante „Expertenkommentare“ waren durchschnittlich 28,9 Jahre alt (s= 8,9 Jahre) und die Probanden in der Gruppe der Variante „Expertenkommentare mit Signal“ im Durchschnitt 31,8 Jahre alt (s = 12,9 Jahre). Das durchschnittliche Alter für die Gruppe „Pausieren“ betrug 26,8 Jahre (s = 8,6 Jahre) und für die Gruppe „Ton variierbar“ 30,1 Jahre (s = 11,2 Jahre). Tabelle 3 zeigt die deskriptiven Statistiken für das Alter, differenziert nach Versuchsgruppen (in den letzten beiden Gruppen hat jeweils eine Person das Alter im Fragebogen nicht angegeben).

Tabelle 3: Alter der Probanden in Jahren nach Experimentalgruppen

	Grundvariante	Experten-kommentare	Experten-kommentare mit Signal	Pausieren	Ton variierbar
N	26	28	23	26	24
Mittelwert (Standardabweichung)	33,1 (13,5)	28,9 (8,9)	31,8 (12,9)	26,8 (8,6)	30,1 (11,2)

Abbildung 12 zeigt eine Übersicht über die Bildungsabschlüsse in den einzelnen Gruppen, wobei die Gruppe mit der Variante mit Fremdkommentaren und Signalton einen höheren Anteil an Hochschulabschlüssen aufweist.

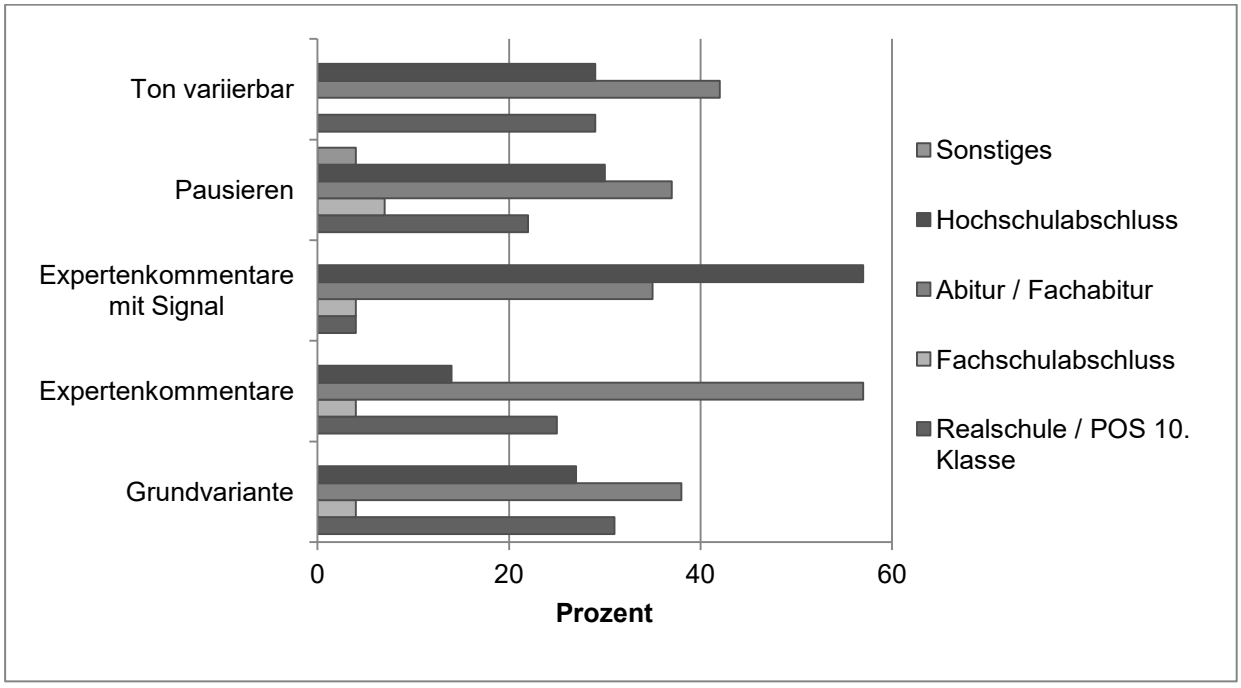


Abbildung 12: Bildungsabschlüsse in den Versuchsgruppen

Es wird deutlich, dass die Probanden aller Versuchsgruppen hinsichtlich ihrer soziodemografischen Hintergründe hohe Ähnlichkeiten aufweisen. So findet sich in jeder Gruppe eine ähnliche Verteilung der Geschlechter. Auch die deskriptiven Statistiken bezüglich des Alters sind vergleichbar und in allen Gruppen zeigt sich ein ähnlich hohes Niveau an Bildungsabschlüssen. Demnach lassen sich aus der Verteilung der Probanden auf die Versuchsgruppen keine Verzerrungen der Ergebnisse ableiten.

5.1.2 Vergleichbarkeit der Versuchsgruppen

Neben der Ähnlichkeit der soziodemografischen Hintergründe der Probanden ist es für die Aussagekraft der Studie von besonderer Bedeutung, dass sich die Probanden auch hinsichtlich ihrer

Erfahrung im Umgang mit dem Medium „Computer“ und ihres Vorwissens in Bezug auf den verwendeten Untersuchungskontext ähneln. Sollte dies nicht der Fall sein, könnten sich Verzerrungen dadurch ergeben, dass Probanden bspw. generell langsamer den Text eingeben können oder aufgrund mangelnden Wissens über die Straßenverkehrsordnung falsche Kommentare eingeben. Um solche Verfälschungen kontrollieren zu können, wurden im Verlauf der Untersuchung mehrere Kontrollvariablen erhoben. Hierzu gehörten die allgemeinen Fähigkeiten beim Umgang mit dem Computer (erhoben im Tipptest) sowie die allgemeine Konzentrationsfähigkeit (erhoben im Konzentrationstest nach Conelly, Hasher & Zacks (1991), bei dem zwei Variablen Vorlesedauer und die Anzahl der korrekten Antworten auf die gestellten inhaltlichen Fragen gemessen wurden).

Um auch Aussagen über die allgemeine Erfahrung mit Computern sowie das subjektiv eingeschätzte Wissen und die Erfahrung der Probanden im Straßenverkehr zu ermitteln, beantworteten die Probanden im Anschluss an die Durchführung der Experimente einen schriftlichen Fragebogen. Der verwendete Fragebogen findet sich im Anhang 1.

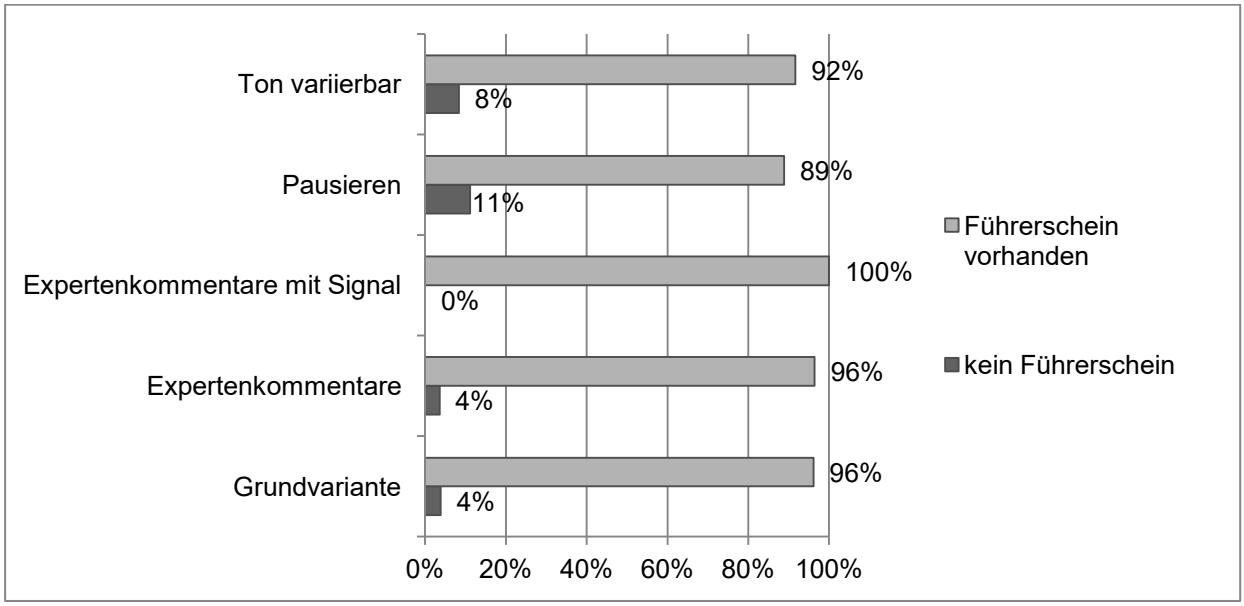


Abbildung 13: Führerscheinbesitz in den Versuchsgruppen

In allen Gruppen gab es fast ausschließlich Probanden, die einen Führerschein besitzen. Abbildung 13 zeigt den Führerscheinbesitz differenziert nach den fünf zu untersuchenden Gruppen[8]. Da in allen Gruppen sehr wenige Probanden keinen Führerschein besitzen, wird davon ausgegangen, dass sich durch dieses Merkmal keine Verzerrungen hinsichtlich der Ergebnisse zeigen werden.

[8] Jeweils ein Proband aus der Gruppe Expertenkommentare mit Signal und Ton machte keine Angaben zum Führerschein.

Die übrigen angesprochenen Variablen besaßen ein metrisches bzw. ein ordinales Skalenniveau, daher wurden statistische Testverfahren angewendet, um Unterschiede auf Signifikanz zu prüfen. Tabelle 4 zeigt die arithmetischen Mittel sowie die Standardabweichungen (in Klammern) der erhobenen Variablen.

Tabelle 4: Mittelwerte und Standardabweichungen der Variablen zu Erfahrung mit Computern und dem Wissen über Straßenverkehr.

		Grund-variante	Experten-kommentare	Experten-kommentare mit Signal	Pausieren	Ton variierbar
Tipptest	Tippgeschwindigkeit (Zeichen pro Sekunde)	2,6 (1,1)	2,4 (0,9)	2,7 (0,9)	2,5 (0,7)	2,6 (0,9)
Konzentrationstest	Dauer des Vorlesens (Sek.)	78,3 (13,1)	81,4 (28,2)	75,5 (15,0)	77,9 (19,1)	75,8 (12,7)
	richtige Antworten	2,3 (1,2)	1,9 (1,1)	2,2 (1,0)	2,3 (0,7)	2,0 (1,0)
Computerkenntnisse	Jahre der Erfahrung mit PCs	14,2 (6,5)	11,9 (5,9)	14,0 (5,9)	12,9 (3,8)	11,0 (5,6)
	Computertage pro Woche	6,0 (1,1)	5,4 (1,7)	6,0 (1,2)	6,1 (0,9)	5,8 (1,4)
	Subj. Einschätzung PC-Fähigkeiten	2,5 (0,9)	2,4 (1,2)	1,9 (0,8)	2,6 (1,1)	2,7 (0,9)
Subjektive Einschätzung	Wissen über Straßenverkehr	2,5 (1,0)	2,7 (0,8)	2,6 (1,2)	2,5 (1,0)	2,6 (0,7)
	Erfahrung im Straßenverkehr	3,8 (1,3)	3,5 (1,3)	4,0 (1,0)	3,7 (1,5)	3,6 (1,2)

Um signifikante Unterschiede zwischen der Kontrollgruppe (Grundvariante) und den einzelnen Experimentalgruppen, in denen die Zusatzfunktionalitäten erprobt wurden, auszuschließen, wurden t-Tests berechnet. Falls die Vorbedingung der Normalverteilung für den t-Test nicht erfüllt war, wurde der parameterfreie U-Test angewendet (siehe Anhang 2). Dem t-Test vorausgehend wurde der Levene-Test auf Varianzhomogenität durchgeführt und bei einer Signifikanz kleiner als 0,05 für den t-Test eine entsprechende Korrekturformel angewandt. Durch t-Test und U-Test kann zwar nicht gezeigt werden, dass zwei Gruppen hinsichtlich einer Eigenschaft gleich sind, aber ein Signifikanzwert, der größer als 0,05 ist, zeigt zumindest an, dass nicht von

einem signifikanten Unterschied auszugehen ist. In Anhang 3 sind die detaillierten Ergebnisse der t-Tests und U-Tests dargestellt.

Die Vergleiche zwischen der Grundvariante und den verschiedenen Experimentalgruppen führen zu keinen signifikanten Unterschieden. Nur für die Variable „Jahre der Erfahrung mit PCs“ gab es einen signifikanten Unterschied zwischen der Kontrollgruppe und der Experimentalgruppe „Ton variierbar“ (Grau hinterlegt Tabelle 4). Demzufolge besitzen die Probanden dieser Experimentalgruppe statistisch gesehen signifikant weniger Erfahrung im Umgang mit Computern. Da sich jedoch keinerlei Unterschiede hinsichtlich der übrigen Variablen zur Messung der Erfahrung im Umgang mit Computertechnik finden und die Probanden dieser Gruppe auch durchschnittlich mehr als zehn Jahre Erfahrung mit Computern aufweisen, ist nicht davon auszugehen, dass sich aus diesem Umstand Verzerrungen hinsichtlich der Ergebnisse ergeben.

Somit ist festzuhalten, dass die zufällig zugelosten Versuchsgruppen hinsichtlich der soziodemografischen Merkmale sowie der Konzentrationsfähigkeit, der Erfahrung im Umgang mit der in den Versuchen verwendeten Technik sowie hinsichtlich des Wissens über und der Erfahrung im Straßenverkehr keine nennenswerten Unterschiede aufweisen. Sie können daher im Folgenden hinsichtlich der vorgestellten quantitativen und qualitativen Leistungsmaße miteinander verglichen werden.

5.2 Ergebnisse der statistischen Auswertung

Nach der Eingabe der Daten wurden Signifikanztests (U-Tests und t-Tests) durchgeführt, um zu ermitteln, ob die Varianten des Grundaufbaus der KNoTen-Station signifikante Vor- oder Nachteile im Vergleich zur Grundvariante hinsichtlich der Leistung der Probanden mit sich bringen. Die Ergebnisse der Tests werden im Folgenden getrennt nach der jeweils untersuchten Variablen, dargestellt. Die vollständige deskriptive Statistik zu den untersuchten Variablen, die Tests auf Normalverteilung sowie die Statistiken zu den Signifikanztests finden sich im Anhang 4, Anhang 5 und Anhang 6.

5.2.1 Anzahl der geschriebenen Kommentare

Die Variable Anzahl der geschriebenen Kommentare konnte für alle 129 Teilnehmer der Studie ausgewertet werden. Die Verteilung der Probanden auf die Versuchsgruppen ist den jeweiligen N-Werten in der nachstehenden Abbildung 14 zu entnehmen. Zur besseren Veranschaulichung wurde in Höhe des Mittelwertes der Grundvariante eine gestrichelte Linie eingefügt, um die Abweichungen der übrigen Gruppen von diesem Vergleichsniveau zu verdeutlichen.

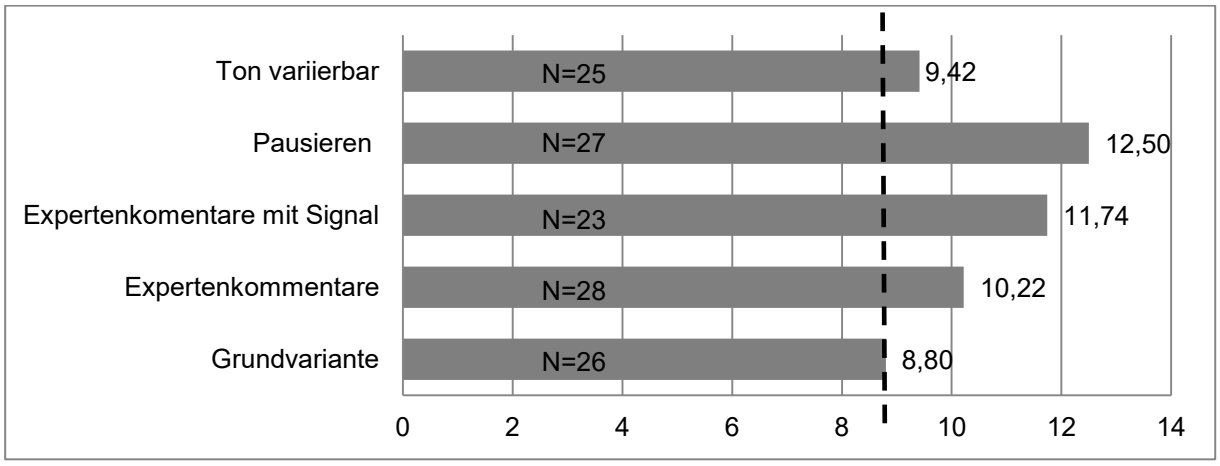


Abbildung 14: Anzahl der Kommentare - Mittelwerte der Gruppen.

In der Gruppe, die mit der Grundvariante der KNoTen-Station arbeitete, wurden im Durchschnitt 8,8 Kommentare pro Proband geschrieben (s = 5,71). In allen übrigen Versuchsgruppen lag der Mittelwert höher. So wurden in der Variante mit Expertenkommentaren im Durchschnitt 10,22 Kommentare (s=5,37), in der Variante „Expertenkommentare mit Signal“ 11,74 Kommentare (s=6,83), in der Variante „Pausieren“ 12,5 Kommentare (s=8,82) und in der Variante „Ton variierbar“ 9,42 Kommentare (s=5,96) pro Proband geschrieben. Um zu ermitteln, ob es sich hierbei um signifikante Unterschiede handelt, wurden Mann-Whitney-U-Tests durchgeführt, die auch bei nicht vorliegender Normalverteilung verwendet werden können. Es zeigt sich, dass lediglich der Unterschied zwischen der „Grundvariante“ und der Variante „Expertenkommentare mit Signal“ signifikant ist (p = 0.038). Demnach ist davon auszugehen, dass der Unterschied zwischen diesen beiden Gruppen auch auf die Grundgesamtheit übertragen werden kann, so dass bei der Implementierung der KNoTen-Station inklusive der Einblendung von Expertenkommentaren mit Signal damit gerechnet werden kann, dass mehr Kommentare geschrieben werden, als ohne diese zusätzliche Funktionalität. Über die Qualität der Kommentare ist damit aber nichts ausgesagt.

5.2.2 Anzahl der einbezogenen Videos

Auch die Variable Anzahl der einbezogenen Videos konnte für alle Studienteilnehmer ausgewertet werden. Die Mittelwerte sind in Abbildung 15 dargestellt.

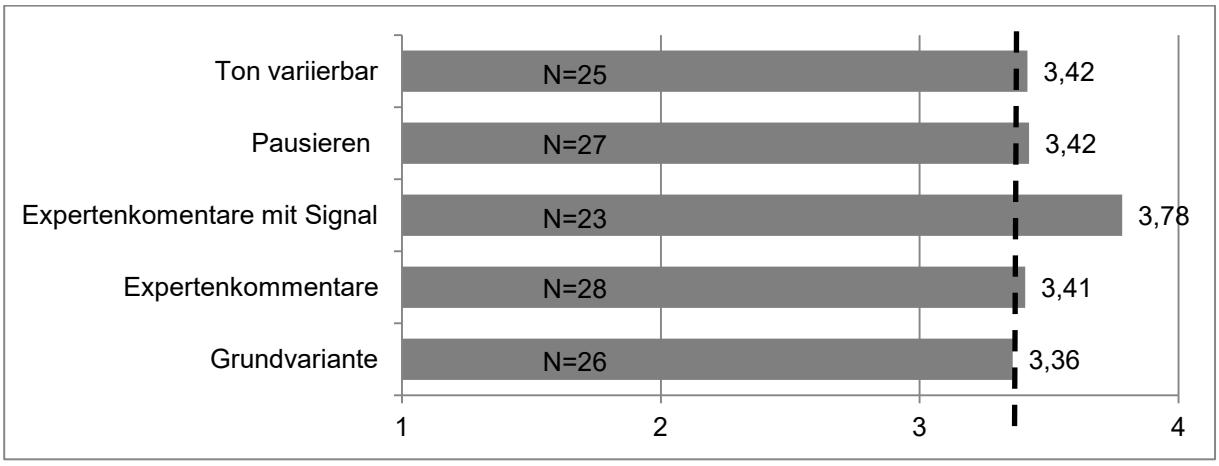


Abbildung 15: Anzahl einbezogener Videos - Mittelwerte der Gruppen.

Bei der Betrachtung der Abbildung 15 wird deutlich, dass sich alle Gruppen hinsichtlich der Variable Anzahl einbezogener Videos sehr ähneln. Die Grundvariante wies einen Mittelwert von 3,36 Videos (s= 0,91); die Variante „Expertenkommentare“ einen Mittelwert von 3,41 (s=0,57), die Varianten „Pausieren“ und „Ton variierbar“ jeweils einen Mittelwert von 3,42 auf ($s_{_Pausieren}$ = 0,81; $s_{_Ton\ variierbar}$ = 0,83). Lediglich die Gruppe, in der die Expertenkommentare mit Signalton eingeblendet wurden wich etwas stärker ab mit einer durchschnittlichen Anzahl von 3,78 einbezogenen Videos (s=1,09). Bei der Überprüfung mit dem U-Test stellte sich allerdings kein statistisch signifikanter Unterschied heraus.

5.2.3 Ausführlichkeit der Kommentare

Die Variable Ausführlichkeit der Kommentare wurde gemessen an der Anzahl der Zeichen, die ein eingegebener Kommentar durchschnittlich enthielt. Die Abbildung 16 enthält die Darstellung der Mittelwerte für diese Variable.

In der „Grundvariante“ wurden im Durchschnitt 31,25 Zeichen (inklusive Leerzeichen) (s= 8,21) pro Kommentar geschrieben. Die Varianten „Expertenkommentare“, „Expertenkommentare mit Signal“ und „Ton variierbar“ fallen hierhinter etwas zurück. So betrugen die Mittelwerte bei der Variante „Expertenkommentare“ 30,13 Zeichen (s=10,77), bei der Variante „Expertenkommentare mit Signal“ 28,65 Zeichen (s=7,95) und bei der Variante „Ton variierbar“ 30,37 Zeichen (s=9,93). Einzig in der Variante „Pausieren“ wurden im Durchschnitt mehr Zeichen je Kommentar eingegeben. Hier lag der Mittelwert bei 43,10 Zeichen (s=11,59). Dieser Wert unterscheidet sich hoch signifikant von der Grundvariante (Mann-Whitney-U-Wert=138,5; Z= -3,781; p=0,000).

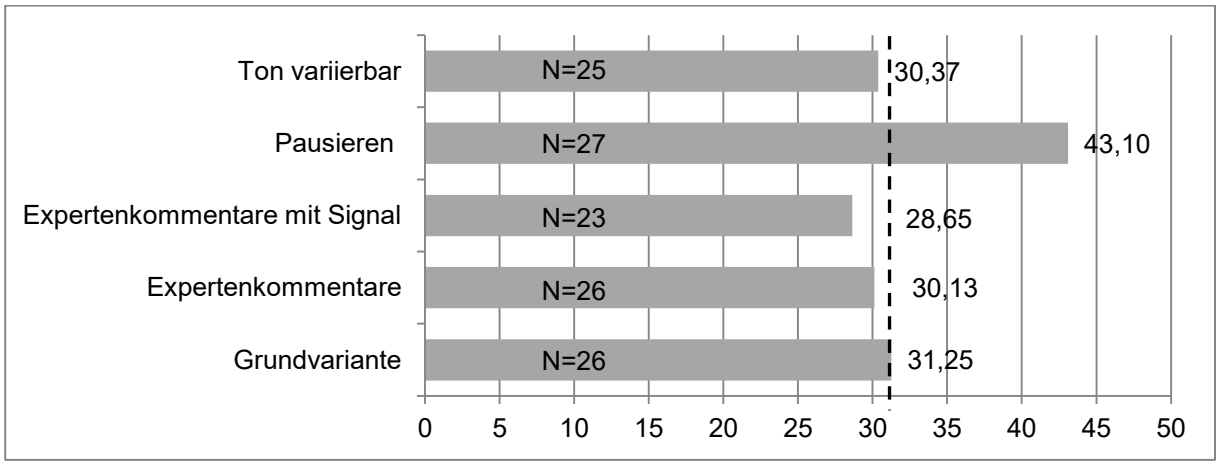


Abbildung 16: Ausführlichkeit der Kommentare - Mittelwerte der Gruppen.

5.2.4 Tippgeschwindigkeit

Die Variable Tippgeschwindigkeit i.S. der Schnelligkeit der Texteingabe während der Versuche wurde mit Hilfe der eingegebenen Zeichen pro Sekunde erfasst. Dafür konnten insgesamt 128 Datensätze ausgewertet werden. Die Mittelwerte sind in Abbildung 17 dargestellt.

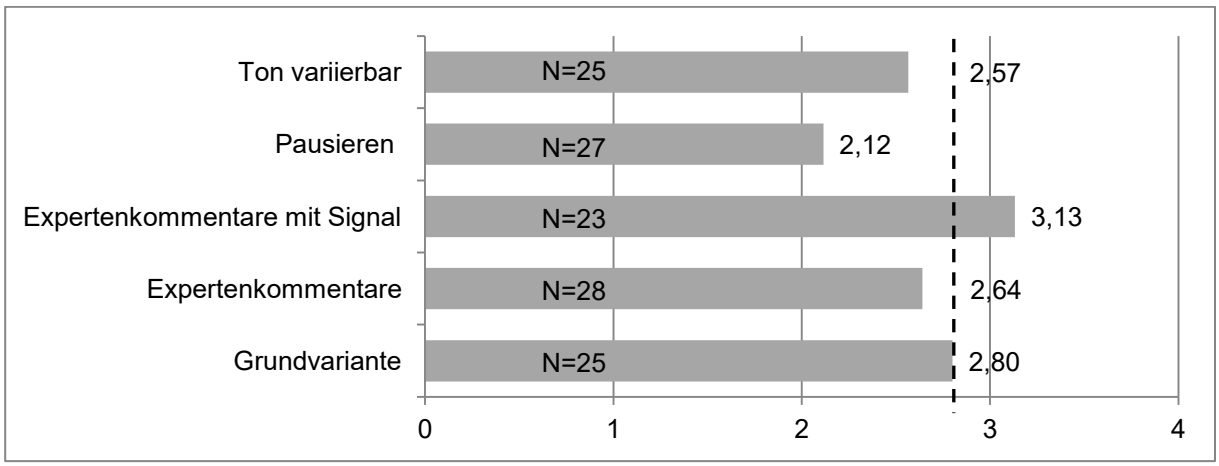


Abbildung 17: Tippgeschwindigkeit - Mittelwert der Gruppen.

Bei dieser Variable wies die Grundvariante einen Mittelwert von 2,8 Zeichen pro Sekunde auf (s=1,01). Lediglich für die Variante „Expertenkommentare mit Signal“ zeigte sich eine höhere Geschwindigkeit (M=3,13 Zeichen / Sekunde, s= 1,00). Die Varianten „Expertenkommentare“ (M=2,64 Zeichen / Sekunde, s= 0,90), „Pausieren“ (M=2,12 Zeichen / Sekunde, s= 0,88) und „Ton variierbar“ (M=2,57 Zeichen / Sekunde, s= 0,90) hingegen schnitten schlechter als die Grundvariante ab.

Der Unterschied zwischen der Grundvariante und der Variante „Pausieren“ ist hoch signifikant (T=2,757, df=50, p = 0.008). Experten, die die Möglichkeit des Pausierens haben, sollten demnach generell langsamer schreiben als ohne diese Option. Damit verlieren sie unter dieser Be-

dingung mehr Zeit und können sich folglich weniger auf das Beobachten der Videos konzentrieren.

5.2.5 Gesamtdauer

Neben der Variable Tippgeschwindigkeit wurde auch die Gesamtdauer[9] der Versuche erhoben. Da diese stark an die Dauer der Videos gebunden ist – sie dauerten in den Versuchen acht Minuten – waren bei der Untersuchung, mit Ausnahme der Variante „Pausieren“, kaum Unterschiede zur Grundvariante abzusehen. Die Abbildung 18 zeigt die Mittelwerte der Gruppen für diese Variablen.

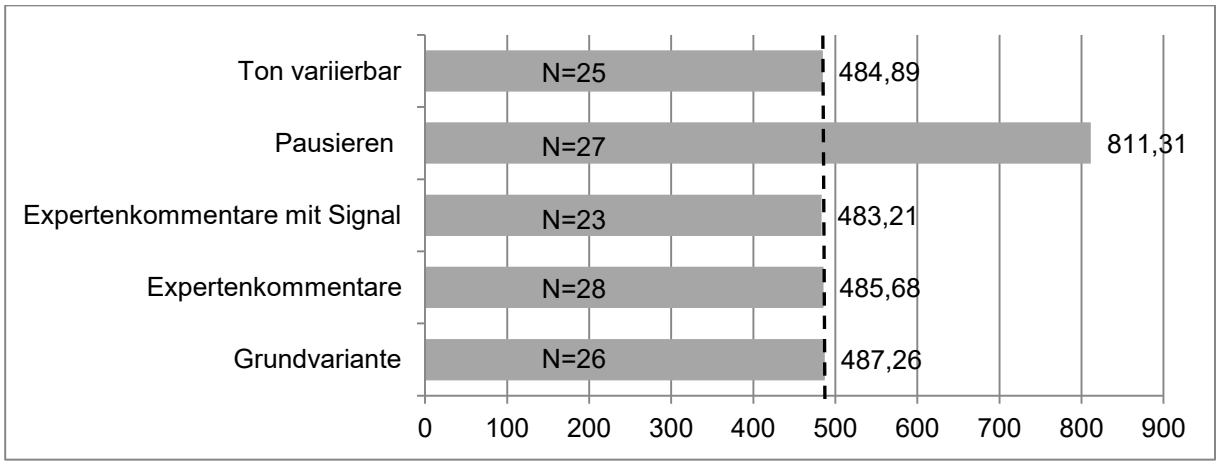


Abbildung 18: Versuchsdauer in Sekunden - Mittelwerte der Gruppen.

Wie zu vermuten war, ergeben sich kaum Unterschiede zwischen den Versuchsgruppen mit Ausnahme der Variante „Pausieren“. Die kleineren Abweichungen der Mittelwerte der anderen Varianten sind wahrscheinlich auf minimale technische Verzögerungen zurückzuführen, die sich beim Abspielen der Software mit unterschiedlichen Endgeräten ergeben. Die Mittelwerte und Standardabweichungen sowie weitere deskriptive Statistiken können Anhang 4 entnommen werden.

Die Variante „Pausieren“ mit einem Mittelwert von 811,31 Sekunden (s=224,70) hingegen benötigt deutlich mehr Zeit als die „Grundvariante“. Der Unterschied erwies sich als hoch signifikant (Mann-Whitney-U-Wert = 2,0, Z = - 6,09, p = 0.000).

[9] Die Gesamtdauer entspricht der Zeit in Sekunden zwischen manuellem oder automatischem Start bzw. Erscheinen der Benutzeroberfläche, in der die vier Videos angezeigt wurden und der automatischen Beendigung, wenn entweder alle Videos zu Ende waren oder, falls ein Proband über dieses Ende hinaus mit der Eingabe eines Kommentars befasst war, bis zum Absenden dieses Kommentars.

5.2.6 Richtigkeit der Kommentare

Die Richtigkeit der geschriebenen Kommentare bildet eines der eher inhaltlich-qualitativen Maße, das sich zur Bewertung der Leistung der Probanden heranziehen lässt. Hierbei musste zunächst ermittelt werden, inwieweit die Kommentare der Probanden mit der zuvor erstellten Musterlösung übereinstimmten. Jeder Proband erhielt je einen Punkt für jeden richtigen Kommentar, die Anzahl der Punkte wurde abschließend für jeden Probanden aufsummiert. Die nachstehende Abbildung zeigt die Mittelwerte der Gruppen im Überblick.

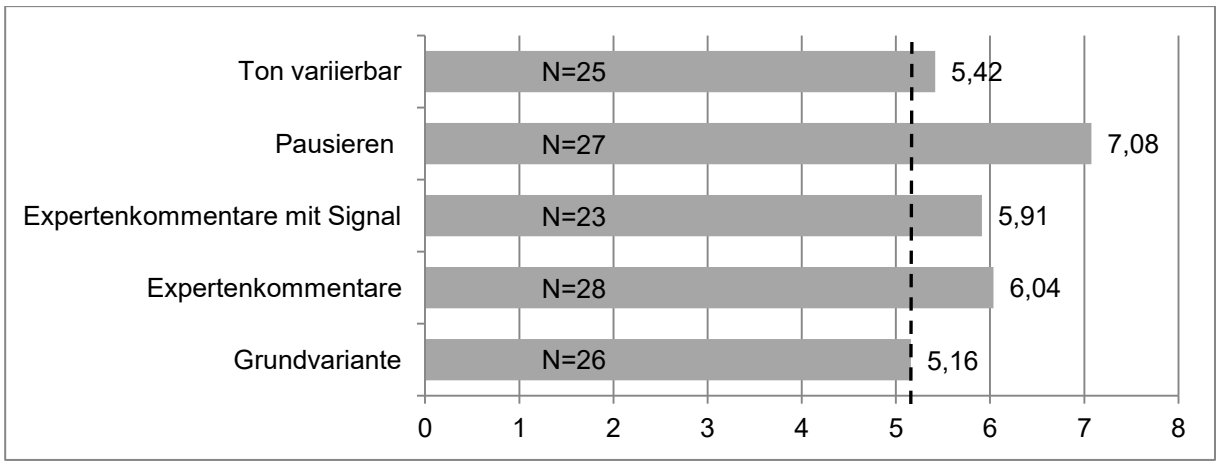


Abbildung 19: Anzahl richtig erkannter Verstöße - Mittelwerte der Gruppen.

Den kleinsten Mittelwert hinsichtlich der Richtigkeit der Kommentare bezogen auf die Musterlösung weist die Grundvariante mit 5,16 richtigen Kommentaren (s=3,09) auf. In allen übrigen Versuchsgruppen lag der Mittelwert höher. So wurden in der Variante mit Expertenkommentaren im Durchschnitt 6,04 richtige Kommentare (s=3,09), in der Variante Expertenkommentare mit Signal 5,91 richtige Kommentare (s=2,89), in der Variante „Pausieren“ 7,08 richtige Kommentare (s=6,49) und in der Variante „Ton variierbar“ 5,42 richtige Kommentare (s=3,19) von den Probanden geschrieben. Die Überprüfung der Daten mittels U-Test zeigte allerdings keine signifikanten Unterschiede.

5.2.7 Verständlichkeit

Für die Ermittlung der Verständlichkeit wurde ein Scoring-Verfahren entwickelt, bei dem alle erfassten Kommentare Plus- und Minuspunkte erhielten. Die im Folgenden verwendeten Mittelwerte beziehen sich auf die Verständlichkeitspunkte, die jeder Proband aufsummiert über alle von ihm geschriebenen Kommentare erhalten hat, abzüglich der Minuspunkte. Der Verständlichkeits-Score konnte für 126 Probanden ermittelt werden. Die Abbildung 20 zeigt die Mittelwerte im Vergleich der untersuchten Varianten der KNoTen-Station.

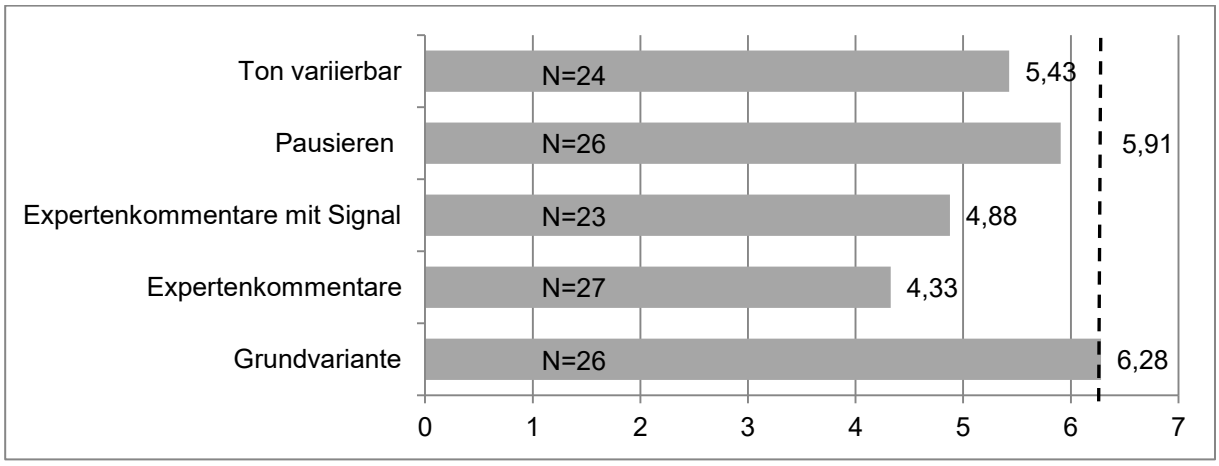


Abbildung 20: Verständlichkeit der Kommentare - Mittelwerte der Gruppen

Die „Grundvariante“ lieferte bei der Variable Verständlichkeit mit 6,28 Punkten (s=5,38) den höchsten Mittelwert, gefolgt von der Variante „Pausieren“ mit einem Mittelwert von 5,91 Punkten (s=4,61). Am geringsten fiel der Mittelwert in der Variante „Expertenkommentare“ aus (M = 4,33 Punkte, s=2,96). In der Variante „Expertenkommentare mit Signal“ zeigte sich ein Mittelwert von 4,88 Punkten (s=4,29) und in der Variante „Ton variierbar“ ein Mittelwert von 5,43 Punkten (s=3,6). Nach einer Überprüfung der Unterschiede zwischen der „Grundvariante“ und den experimentellen Varianten mittels Mann-Whitney-U-Test zeigten sich jedoch keine signifikanten Differenzen.

5.2.8 Zusammenfassung

Bei der Untersuchung der erhobenen Daten ergaben sich insgesamt vier signifikante Unterschiede zwischen der Grundvariante der KNoTen-Station und den denkbaren Abwandlungen. In der Variante „Expertenkommentare mit Signal“ wurden signifikant mehr Kommentare geschrieben und in der Variante „Pausieren“ waren die Kommentare im Vergleich zur „Grundvariante“ signifikant ausführlicher, dafür waren aber die Leistungen in dieser Gruppe hinsichtlich der Tippgeschwindigkeit und der Gesamtdauer des Kommentierungsprozesses der „Grundvariante“ signifikant unterlegen.

Hinsichtlich der Anzahl der einbezogenen Videos und der beiden qualitativen Merkmale Richtigkeit der Kommentare und Verständlichkeit konnten keine signifikanten Unterschiede im Vergleich zur „Grundvariante“ festgestellt werden. Ob sich weitere Empfehlungen hinsichtlich der Implementierung der KNoTen-Station ableiten lassen, wird bei der Auswertung der Interviewdaten im folgenden Abschnitt dargelegt.

5.3 Ergebnisse aus den Interviewdaten

Im Anschluss an die Experimentaldatenerhebung wurde mit allen Probanden noch ein Interview durchgeführt, um über die quantitativen Daten hinaus Rückmeldung der Probanden über deren subjektives Erleben der jeweiligen Untersuchungssituation zu erhalten. Damit sollten qualitative, nicht nur auf die Leistung bezogene Anhaltspunkte zur Nutzerfreundlichkeit der KNoTen-Station und ihrer verschiedenen Varianten ermittelt werden. Hierfür wurden die Studienteilnehmer gefragt:

1. Wie schwierig fanden Sie die Aufgaben (erfasst auf einer Skala von 1=gar nicht schwierig bis 5 schwierig)?
2. Wie haben Sie sich insbesondere bei der Bearbeitung der Kommentaraufgabe gefühlt? Wieso haben Sie sich so gefühlt? Was war der konkrete Auslöser?
3. Wie müsste das Programm der letzten Aufgabe verändert werden, damit Sie die gestellte Aufgabe *noch besser* lösen könnten? Wie müsste das Programm der letzten Aufgabe verändert werden, damit Sie die gestellte Aufgabe *noch schneller* lösen könnten?

5.3.1 Subjektive Schwierigkeit

Die erlebte Schwierigkeit der Aufgabe ist deshalb von Bedeutung, weil deren Fortsetzung bzw. das Abbrechen der Bearbeitung davon abhängen kann (Orvis, Horn & Belanich, 2008). Mit Schwierigkeit ist der Grad gemeint, zu dem eine Aktivität für das Individuum eine beanspruchende Situation darstellt, welche einer bestimmten Menge an kognitiver und physischer Anstrengung bedarf. Die erlebte Aufgabenschwierigkeit wirkt sich auf die Motivation des Individuums aus, welche am höchsten ist, wenn es bei der Bearbeitung einer Aufgabe gleich wahrscheinlich ist, dass man sie mit Erfolg oder Misserfolg beendet. Insbesondere, wenn eine Aufgabe als zu schwierig erlebt wird, kann es dazu kommen, dass die Aufgabe abgebrochen wird (Orvis et al., 2008).

Für das KNoTen-Projekt lässt sich daraus schlussfolgern, dass die Aufgabe, die Trainingsvideos parallel zu verfolgen und gleichzeitig Feedback-Kommentare zu verfassen, eine mittlere Schwierigkeit für die Experten darstellen sollte. Ein solcher Schwierigkeitsgrad beeinflusst die Motivation positiv (Nerdinger, 2013) und reduziert die Wahrscheinlichkeit, dass der Feedbackprozess vorzeitig abgebrochen wird (Orvis et al., 2008).

Im Rahmen der vorliegenden Studie wurde die subjektiv erlebte Aufgabenschwierigkeit mittels einer fünfstufigen Skala erhoben (1 = gar nicht schwierig bis 5 = sehr schwierig). Den Proban-

den wurde im Anschluss an die Kommentierungsaufgabe im Interview die Frage gestellt, wie schwierig sie diese Aufgabe fanden. 119 der 129 Probanden beantworteten diese Frage. Tabelle 5 zeigt die deskriptiven Statistiken für die Variable Aufgabenschwierigkeit im Überblick.

Tabelle 5: Deskriptive Statistiken für die Variable Aufgabenschwierigkeit

	Grundvariante	Experten-kommentare	Experten-kommentare mit Signal	Pausieren	Ton variierbar
N	25	27	19	26	22
Fehlende Werte	1	1	4	1	3
Mittelwert	4,00	4,11	4,00	4,04	4,02
Standardabweichung	0,87	0,80	0,82	0,87	0,91

Bei der Betrachtung der Mittelwerte in allen Versuchsgruppen wird deutlich, dass diese mit Werten über 4 in dem Teil der verwendeten Skala angesiedelt sind, der für eine eher hohe Schwierigkeit steht. Darüber hinaus gab kein einziger Proband an, dass er die Aufgabe für gar nicht schwierig hielt. Im Vergleich zur Kontrollgruppe wies keine der übrigen Varianten einen statistisch signifikanten Unterschied in der erlebten Schwierigkeit auf (Statistiken der durchgeführten U-Tests sind dem Anhang 8 zu entnehmen). Es bleibt demnach festzuhalten, dass alle untersuchten Varianten eine ähnliche, eher hoch erlebte Schwierigkeit aufweisen.

5.3.2 Gefühlslage während der Aufgabenbearbeitung

Die Frage nach der gefühlsmäßigen Verfassung während der Arbeit mit der KNoTen-Station oder einer ihrer Varianten beantworteten 127 der insgesamt 129 Probanden. Eine Übersicht über die Verteilung dieser Aussagen über die fünf Versuchsgruppen findet sich in Tabelle 6.

Tabelle 6: Anzahl der Statements zur Gefühlslage während der Nutzung Experimentalsoftware.

Variante	Grund-variante	Experten-kommentare	Experten-kommentare mit Signal	Pausieren	Ton variierbar
Anzahl der Proban-den insgesamt	26	28	23	27	25
Anzahl der verwert-baren Statements	26	28	24	22	25

In einer Antwort ließen sich mehrere Sinneinheiten nachweisen, die einzeln kategorisiert wurden. Außerdem haben zwei Probanden die Aufnahme der Interviews nicht erlaubt, ihre Antworten wurden daher nicht ausgewertet. In 13 Antworten fanden sich keine Bezüge zur eigentlichen Fragestellung, daher wurden auch diese Aussagen vernachlässigt.

5.3.2.1 Beschreibung der Kategorien

Im Rahmen der qualitativen Inhaltsanalyse kristallisierten sich insgesamt acht Kategorien heraus, die zum Teil weitere Unterkategorien umfassen. Im Folgenden sollen diese Kategorien zunächst benannt und beschrieben werden, bevor ihre Zuordnung zu den Versuchsgruppen vorgestellt wird.

Die erste Oberkategorie wurde mit *positives / neutrales Gefühl* bezeichnet. Hierunter sind drei Arten von Aussagen zusammengefasst, die entweder ein „gutes“ oder ein „neutrales“ Gefühl bzw. eine als positiv erlebte, „spielerische Herausforderung“ bei der Bearbeitung der Aufgabe beinhaltet. Jeweils eine kurze Beschreibung sowie wörtliche Beispiele für die drei Subkategorien finden sich in Tabelle 7.

Tabelle 7: Subkategorien zur Kategorie positives / neutrales Gefühl

Subkategorie	Beschreibung	Beispiele
gut	Probanden äußern ein wenig spezifiziertes, positives Gefühl	"Also gefühlt hab ich mich gut,…" (cs895.005-Sz.1) "Hab ich mich bei gut gefühlt." (mj.172.08-Sz.2)
neutral	Probanden äußern kein spezifisches Gefühl, beschreiben, dass sie sich nicht ungewöhnlich fühlen, einfach die Aufgabe als solche versuchen zu meistern	"Normal." (mg383.03-Sz.1) "Eigentlich nicht anders als sonst. Man hat sich das halt angeguckt…" (jm33.01-Sz.1)
spielerisch herausgefordert	Proband berichtet von der Aufgabe, wie von einem Wettkampf gegen die anderen Experten, fand es lustig und spaßig	"Ich fand es interessant. Es war lustig aber auch ein bisschen Nervenkitzel, weil man ja schon an sich den Anspruch hat die anderen drei, was heißt zu schlagen, aber mehr herauszufinden oder mehr Kommentare [aufzuzählen].“ (tb190.02-Sz.6)

Zur Begründung für das *gute* Gefühl führten die Probanden bspw. an: "Weil's ne interessante Aufgabe war" (cs895.005-Sz.1) oder "Es waren Aufgaben, die mir Spaß gemacht haben" (ms1222.03-SZ.1). Für die neutrale Gefühlslage und die spielerisch aufgefasste Herausforderung wurden keine spezifischen Gründe genannt.

Die drei Unterkategorien *gut*, *neutral* und *spielerisch herausgefordert* wurden zusammengefasst, da sie alle keine negative Grundtendenz aufweisen. Dies ist bei allen übrigen Kategorien der Fall, wie im Folgenden deutlich wird. Hinter der zweiten Kategorie *negatives Gefühl* verbergen sich alle Aussagen, die zum Inhalt haben, dass Probanden entweder *frustriert* waren oder sich *unwohl* gefühlt haben. Die Beschreibung und entsprechende Beispiele für die Subkategorien finden sich in der folgenden Tabelle 8.

Tabelle 8: Subkategorien zur Kategorie negatives Gefühl

Subkategorie	Beschreibung	Beispiele
frustriert	Probanden empfanden die Aufgabe als schwierig zu bewältigen und waren enttäuscht über die eigene Leistung	"Genervt. Weil ich ganz genau wusste, […] dass ich nicht alles erfassen kann […] egal wie konzentriert ich mich dieser Aufgabe widme." (jm337.02-Sz.1) "Frustrierend, weil man sich ärgert, dass man so wenig herausgefunden hat…" (sh594.01-Sz.1)
unwohl	weniger spezifiziertes, negatives Gefühl	"...deswegen hab ich mich sehr unsicher gefühlt und sehr unwohl." (mg383.10-Sz.1) "Nicht gut." (lk146.04-Sz.1)

Als Gründe für das *Unwohlsein* wurden bspw. die Angst etwas zu verpassen, während man geschrieben hat (mg383.10-Sz.1) und die eigene mangelnde Erfahrung mit dem Untersuchungsgegenstand (Straßenverkehr) (lk146.04-Sz.1) genannt, wobei Letzteres nur für die Teilnehmer an der vorliegenden Studie relevant sein dürfte.

Etwas spezifischer als in der Kategorie negatives Gefühl waren die Aussagen, die der Kategorie *gestresst / überfordert* zugeordnet werden konnten. Diese werden im Folgenden aufgrund ihrer Spezifität und ihres Zusammenhangs mit dem Gefühl der Überlastung und des empfundenen Stress‘ separat betrachtet. In der Kategorie *gestresst / überfordert* finden sich vier Unterkategorien. Die Subkategorie *überfordert* beinhaltet alle Aussagen, in denen die Probanden explizit das Wort „überfordert“ benutzten. Die Subkategorie *gestresst* beinhaltet sämtliche Aussagen, in denen entweder wieder die Worte Stress, stressig oder gestresst vorkamen oder aber von Angespanntheit, Druck oder Anstrengung die Rede war. *Hilflos* beinhaltet Aussagen darüber, dass die Probanden den Eindruck hatten, dass sie die Aufgabe nicht kontrollieren konnten. Schließlich zählt auch ein Gefühl der Unsicherheit dazu. Die hier gemeinte Unsicherheit bezieht sich ganz konkret auf das fehlende Bewusstsein, die Aufgabe (Beobachten und Kommentieren) bewältigen zu können, weil zu viele Dinge gleichzeitig erledigt werden müssen. Es handelt sich also bei dieser Art der Unsicherheit (hier benannt als *unsicher, Aufgaben gleichzeitig zu schaf-*

fen) um ein fehlendes, aufgabenspezifisches Selbstvertrauen (Bandura, 1997) ausgelöst durch die Vielzahl der gleichzeitig zu erledigenden Aufgaben. Die Beschreibungen und Beispiele der vier Subkategorien finden sich in Tabelle 9.

Tabelle 9: Subkategorien zur Kategorie gestresst / überfordert

Sub-kategorie	Beschreibung	Beispiele
überfordert	Probanden sehen sich der Aufgabe nicht gewachsen.	"überfordert" (ap019.02-Sz.1)
gestresst	Probanden fühlen sich unter Druck und angespannt, erleben Stress, empfinden die Aufgabe als sehr anstrengend	"Zuerst gestresst" (lh066f-Sz.1) "Man ist unter Druck." (sh594.02-SZ.1); "Es war sehr anstrengend." (lk146.02-SZ1)
hilflos	Probanden fühlen sich ratlos, unbeholfen bei der Bearbeitung der Aufgabe, empfinden einen Kontrollverlust	"Ich habe mich als erstes etwas hilflos gefühlt, weil ich nicht wirklich wusste, wo ich gucken sollte […]." (nm157.2-Sz.5) "Ja und irgendwo so ein bisschen hilflos, weil man kann nicht überall gleichzeitig sein […] man kann nicht gucken und schreiben und alles, also man verpasst viel." (lk146.01-Sz.5)
unsicher, Aufgaben gleichzeitig zu schaffen	Fehlende Selbstwirksamkeit aufgrund kognitiver Überlastung: Probanden sind sich nicht sicher, ob sie die gestellte Aufgabe lösen können, weil es zu viele Aufgaben gleichzeitig sind	"Mmh, unsicher auf jeden Fall. Durch die vielen Aufgaben gleichzeitig." (sk725.02-Sz.4) "[…] eigentlich hab ich mich dabei eher unsicher gefühlt. Weil ich mehr… weil ich viel wissen müsste, aber man gar nicht so schnell alles erkennen kann, das ist das Problem." (jm337.06-Sz.4)

Alle vier Subkategorien beschreiben das Gefühl, der Aufgabe nicht oder nur unter Einsatz sehr großer Anstrengung gewachsen zu sein und betrachten die gestellte Aufgabe als einen Stressor, der Auslöser für die genannten negativen Empfindungen ist.

Daneben fanden sich Kategorien, die sich nicht weiter zu einer übergeordneten Kategorie zusammenfassen ließen. Hierunter fallen die Kategorien

- *gelangweilt*, bei der die Probanden die zu erledigende Aufgabe als reiz- und spannungslos empfanden,
- *abgelenkt*, bei der sich die Probanden durch bestimmte Gegebenheiten gestört fühlten,
- *unsicher (Wissen)*, bei der die Probanden angaben, sich bezüglich des ihren Kommentaren zugrunde liegenden Wissens unsicher zu sein,

- *unsicher (Aufgabenstellung)*, bei der den Probanden nicht genau klar war, was bei der Erledigung der Aufgabe exakt von ihnen verlangt wird und schließlich
- *hektisch / nervös*, die dann geäußert wurde, wenn die Probanden sich während der Bearbeitung der Aufgabe unruhig und aufgeregt fühlten.

Die Beschreibungen der Kategorien sowie entsprechende Beispiele finden sich in Tabelle 10.

Tabelle 10: Weitere Gefühlskategorien im Überblick

Kategorie	Beschreibung	Beispiele
gelang-weilt	Probanden empfanden das Kommentieren als reiz- bzw. spannungslos	"…nee, es war nachher langweilig" (ap019.06-Sz.1) "Am Anfang noch recht konzentriert […], am Ende ziemlich gelangweilt" (lb119.05-Sz.1)
abgelenkt	Probanden fühlten sich durch etwas in ihrer Aufmerksamkeit für ihre eigentliche Aufgabe gestört.	"Ich hab mich abgelenkt gefühlt durch die Kommentare an der Seite,…" (nv027.9-Sz.2) "Ich habe Versucht nicht hinzugucken, aber man wird natürlich dazu geleitet." (jm377.03-Sz.6)
unsicher (Wissen)	Bei den Probanden herrscht Ungewissheit über die inhaltliche Richtigkeit ihrer Kommentare aufgrund fehlenden oder unsicheren Wissens	"Ich habe mich ein bisschen blöd gefühlt, also blöd im Sinne von oh Gott, ich hab keine Ahnung." (mg383.09-Sz.2) "Ich hab da keine Ahnung von den Verkehrsregeln hier, was die Straßenbahn angeht. Ob man da rüber laufen darf oder nicht […]. " (lb092.04-Sz.4)
unsicher (Aufgaben-stellung)	Bei den Probanden herrschte Unklarheit über die Anforderungen, die an sie gestellt wurden, Aufgabenstellung war nicht ausreichend	"…und dann wusste ich nicht genau, was ich nun alles kommentieren sollte."(nv027.8-Sz.2) "…ist das jetzt oder ist das jetzt kein Fehler…" (mg383.06-Sz.2)
hektisch / nervös	Probanden fühlten sich unruhig, flatterig, fahrig	"Dadurch bin ich stark nervös geworden." (cs895.006-Sz. 6) "Eher hektisch"(ms1222.05-Sz.6)

5.3.2.2 Gefühlskategorien in den verschiedenen Versuchsbedingungen

Im Folgenden wird das Auftreten der vorgestellten Kategorien in den befragten Experimentalgruppen dargestellt. Hierbei wird die Reihenfolge entsprechend der nachfolgenden Tabelle 11 gewählt, die sich aus der Gesamthäufigkeit des Vorkommens der Kategorien in allen Gruppen ergibt. Angegeben werden sowohl der prozentuale Anteil der einbezogenen Aussagen bezogen auf die Gesamtzahl der Aussagen in einer Experimentalgruppe sowie die absolute Anzahl der Aussagen, die der jeweiligen Kategorie zugeordnet wurden.

Die Kategorie *gestresst / überfordert* tritt über alle Interviews hinweg betrachtet am häufigsten auf. 59 Aussagen wurden dieser Kategorie zugeordnet (s. letzte Spalte der Tabelle 11). In allen Szenarien beinhalteten über ein Drittel der Aussagen Hinweise darauf, dass sich die Probanden

gestresst oder überfordert fühlten. Hierbei liegen die Anteile der Aussagen in den Varianten „Expertenkommentare“, „Expertenkommentare mit Signal“ und „Pausieren“ mit ca. 40% sehr nah beieinander. Etwas höher liegt der Anteil der Aussagen mit 46% in der Grundvariante der KNoTen-Station. Im Szenario „Ton variierbar“ weisen dagegen mehr als zwei Drittel (68 %) der Aussagen auf Überforderung oder Stress bei der Bearbeitung der Aufgabe durch die Probanden hin. Möglicherweise haben die Probanden die selbstständige Steuerung des Tons als weitere Aufgabe neben dem Beobachten und Kommentieren angesehen, was zu einer zusätzlichen Arbeitsbelastung führt.

Tabelle 11: Prozentualer Anteil (%) und absolute Häufigkeiten (H) des Vorkommens der Gefühlskategorien in den Versuchsgruppen

Gefühl während der Bearbeitung	Grundvariante		Experten-kommentare		Experten-kommentare mit Signal		Pausieren		Ton variierbar		
	%	H	%	H	%	H	%	H.	%	H	Summe
gestresst / überfordert	46,15	12	39,29	11	41,67	10	40,91	9	68,00	17	59
positives / neutrales Gefühl	23,08	6	35,71	10	20,83	5	22,73	5	20,00	5	31
unsicher (Aufgabenstellung)	-	-	7,14	2	8,33	2	13,64	3	8,00	2	9
negatives Gefühl	15,38	4	3,57	1	4,17	1	4,55	1	4,00	1	8
unsicher (Wissen)	7,69	2	3,57	1	4,17	1	13,64	3	-	-	7
gelangweilt	7,69	2	7,14	2	4,17	1	-	-	-	-	5
hektisch / nervös	-	-	-	-	12,50	3	4,55	1	-	-	4
abgelenkt	-	-	3,57	1	4,17	1	-	-	-	-	2

Die am zweithäufigsten genannte Kategorie ist das *positive / neutrale* Gefühl. In allen Varianten der KNoTen-Station finden sich in mindestens 20 % aller Aussagen Hinweise auf dieses Gefühl, d.h., dass mindestens jede fünfte Aussage einen positiven oder zumindest neutralen Grundtenor besitzt. Das Szenario „Expertenkommentare“ (ohne Signal) sticht hierbei etwas hervor, da es hier wesentlich mehr (fast 40 %) positive oder neutrale Rückmeldungen von den Probanden gab. Hier ist denkbar, dass die Simulation anderer anwesender Experten zu einem angenehmen Gefühl des Nicht-Alleinseins geführt hat oder die Probanden Spaß dabei empfanden, sich mit den anderen Experten zu messen, wie einer der Probanden angab (s. nochmals Tabelle 7).

Unsicher (Aufgabenstellung) ist die am dritthäufigsten angesprochene Kategorie. Diese kam jedoch nur in vier der fünf Versuchsgruppen vor. So äußerte diese Art der Unsicherheit nie-

mand in der Grundvariante, jedoch mindesten jeweils zwei Personen in den übrigen mit zusätzlichen Funktionen versehenen Varianten der KNoTen-Station. Zur Erklärung des unterschiedlichen Vorkommens dieser Kategorie in den Versuchsgruppen könnte die jeweils den vorgenommenen Manipulationen der Softwarevarianten entsprechende, unterschiedlich formulierte Aufgabenstellung herangezogen werden. In der Kontrollgruppe (Grundvariante) mussten im Gegensatz zu den anderen Szenarien keine zusätzlichen Funktionalitäten erläutert werden, so dass die Aufgabenstellung vermutlich übersichtlicher für die Probanden war, wohingegen sich die Probanden der anderen Szenarien mit weiteren Informationen zur Bedienung des Programms auseinandersetzen mussten. Dies könnte ein Hinweis darauf sein, dass die Erläuterung der Funktionalitäten die Aufnahme der eigentlichen Aufgabenstellung stört, was während der Bearbeitung zu Unsicherheiten führt.

Ebenfalls auffällig bei der Kategorie *unsicher (Aufgabenstellung)* ist die Variante „Pausieren". Im Vergleich zu den anderen betroffenen Varianten ist hier die Unsicherheit in fast doppelt so vielen Aussagen zu finden. Die könnte damit erklärt werden, dass die Probanden dieses Szenarios mehr Zeit haben darüber nachzudenken, ob ihre Kommentare wirklich auch der Aufgabenstellung entsprechen. Sie können mehrere Interpretationsweisen ihrer Beobachtungen in Erwägung ziehen, was zu Unsicherheiten bezüglich einer Festlegung auf eine Interpretationsweise und ihrer Passung zur Aufgabenstellung geführt haben könnte.

Ein eher unspezifisches, *negatives Gefühl* kam in insgesamt acht Aussagen zum Ausdruck. Davon betroffen waren alle Versuchsgruppen. Auffällig ist hierbei wieder die Grundvariante, in der Aussagen mit diesem Inhalt deutlich häufiger vorkommen, als in den anderen Varianten. Das Gefühl könnte durch die generell hohen Anforderungen an die Konzentrationsleistung bei der Nutzung der KNoTen-Station entstehen. Fraglich bleibt allerdings, warum die anderen, nicht weniger anspruchsvollen Varianten in dieser Kategorie besser abschneiden. Eine Begründung könnte darin gesehen werden, dass die zusätzlichen Funktionalitäten in den anderen Varianten eher die Aufmerksamkeit der Probanden auf sich gezogen haben und ihnen die Ursache für ein eher unspezifisches, negatives Gefühl zugeschrieben wurde.

Die Kategorie *unsicher (Wissen)* kam in vier der fünf Kategorien vor, nicht jedoch in der Variante „Ton variierbar". Hierbei ist die Variante „Pausieren" übermäßig betroffen. Der Grund könnte auch hier sein, dass in der Variante „Pausieren" die Probanden mehr Zeit zum Nachdenken haben und sich aufgrund unterschiedlicher Interpretationsweisen der eigenen Beobachtungen Unsicherheit in Bezug auf das eigene Wissen entsteht. Im Szenario „Ton variierbar" könn-

ten die Probanden die Steuerung des Tons eventuell als zusätzliche Belastung erlebt haben, die sie noch zusätzlich unter Druck setzt und noch weniger Zeit zum Überlegen und Abwägen lässt.

In der Grundvariante sowie in den Varianten mit anderen Expertenkommentaren haben Probanden angegeben, dass sie sich *gelangweilt* fühlten. Als Begründung hierfür wurde jeweils angeführt, dass im Verlauf der Bearbeitung klar wurde, dass immer wieder die gleichen Vergehen zu kommentieren waren und es demnach nach einiger Zeit nur noch wenig Abwechslung gab. Dies ist insofern von Bedeutung, als dass erlebte Monotonie sich auf die Motivation der Kommentatoren (vgl. Nerdinger, 2012) und auch auf ihre Konzentrationsleistung auswirken kann (Pierre, Hofinger, & Buerschaper, 2011). Dass die Szenarien „Ton variierbar“ und „Pausieren“ nicht betroffen waren könnte daran liegen, dass durch die vollständige Konzentration auf jeweils nur ein Video mit Bild und Ton die gleichen Vergehen weniger häufig wahrgenommen werden. Die Probanden des Szenarios „Pausieren“ könnten durch das Abwägen ihrer Beobachtungen und die Vielzahl der einbezogenen Gedanken weniger Wiederholungen wahrgenommen haben, weil ihre Überlegungen die einzelnen Vergehen in jeweils unterschiedlichen Kontexten einbezogen haben.

Die Gefühlskategorie *hektisch / nervös* kam nur in den Varianten „Expertenkommentare mit Signal“ (drei Aussagen) und „Pausieren“ (eine Aussage) vor, wobei die erste Variante stärker betroffen ist. Das Ergebnis für die „Pausieren“-Variante überrascht, da aufgrund des fehlenden Zeitdrucks die Probanden dieser Versuchsgruppe eigentlich entspannter sein sollten als in den anderen. Der Auslöser der Hektik in der Variante der „Expertenkommentare mit Signal“ könnte auf den Signalton zurückzuführen sein. Dieser stellt die einzige Manipulation gegenüber der Variante „Expertenkommentare“ dar. Außerdem berichtete einer der Probanden: „Und ja, auf jeden Fall war es so, dass ich nervös wurde, weil die anderen Kommentatoren, dass die, die haben ja auch immer die ganze Zeit geschrieben. Das hat man ja auch immer gehört durch dieses Dip-Dip-Dip und ja, wenn plötzlich alle schreiben, dann hat man ja auch dieses Gefühl, dass man etwas übersehen hat und da steigt dann natürlich die Nervosität,“ (cs895.006-Sz.6). Da Nervosität ein Ausdruck erhöhter Informationsbereitschaft ist (Schneider, 2011), wird in dieser Aussage deutlich, dass der Signalton als Hinweis für die Aktivität der anderen Experten wirkt, was den Probanden offensichtlich unruhiger hat werden lassen in Vorbereitung für die eigene Aktivität. Die Frage, ob demnach Hektik und Nervosität als negativ bewertet werden sollte, weil sie zwar als unangenehm erlebt werden aber zu erhöhter Wachsamkeit führen, bleibt an dieser Stelle offen.

Die letzte ermittelte Kategorie *abgelenkt* kommt nur in den Varianten „Expertenkommentar“ und „Expertenkommentar mit Signal“ jeweils einmal vor. In der erstgenannten Variante wurde explizit gesagt, dass sich der Proband durch das bloße Erscheinen der anderen Kommentare abgelenkt fühlte. In der Variante mit Signalton wurde die Ablenkung explizit dem Signalton zugeschrieben. Hier stellt sich die Frage, ob sich die Störung tatsächlich auf das Geräusch an sich bezieht oder eher das Bewusstwerden neuer Kommentare die Ablenkung darstellt. Da an dieser Stelle im Interview aufgrund der für alle Versuchsleiter gleichen Leitfäden nicht nachgefragt wurde, kann die Frage nach dem Ursprung der Ablenkung nicht eindeutig identifiziert werden. Da Kommentare an sich, und der Signalton aber zusätzliche Reize darstellen, muss angenommen werden, dass zusätzlich dargebotene Stimuli tatsächlich, wenn auch nur für einen kleinen Teil der Nutzer der KNoTen-Station, zu Ablenkungen von der eigentlichen Aufgaben führen.

5.3.2.3 Zusammenfassung und Diskussion

Auf Basis der Auswertung des Interviewmaterials hinsichtlich der Gefühlslage während der Arbeit mit der KNoTen-Station und ihrer denkbaren Varianten sollten einige Ergebnisse bei der endgültigen Implementierung nicht außer Acht gelassen werden. Zur besseren Veranschaulichung wird in der folgenden Abbildung 21 das Vorkommen der Aussagen in den verschiedenen Gefühlskategorien nochmals zusammengefasst. Um die Vergleichbarkeit zu ermöglichen, sind die relativen Häufigkeiten als Grundlage für die Erstellung des Diagramms herangezogen worden.

Zum einen ist festzustellen, dass das erlebte Stressgefühl und der Eindruck der Überforderung in allen Varianten sehr hoch sind. Die Arbeit mit der KNoTen-Station ist demnach generell als sehr anspruchsvoll einzustufen. Als besonders stressend / überfordernd wird die Variante „Ton variierbar“ erlebt. Obwohl zu vermuten gewesen wäre, dass die Möglichkeit der Reduktion der Nebengeräusche durch das Setzen eines auditiven Fokus sich positiv auf das Stresserleben der Experten auswirkt, scheint die selbstständige Steuerung des Tons als zusätzliche Belastung erlebt zu werden. Angesichts dieser Tatsache sollte von dieser Variante Abstand genommen werden.

Die Integration anderer Expertenkommentare (ohne Signalton) wird am häufigsten neutral oder sogar positiv beschrieben. Die virtuelle Zusammenarbeit könnte demnach einen wichtigen motivationalen Vorteil darstellen. Dabei ist aber zu beachten, dass die Expertenkommentare auch ablenkend wirken können.

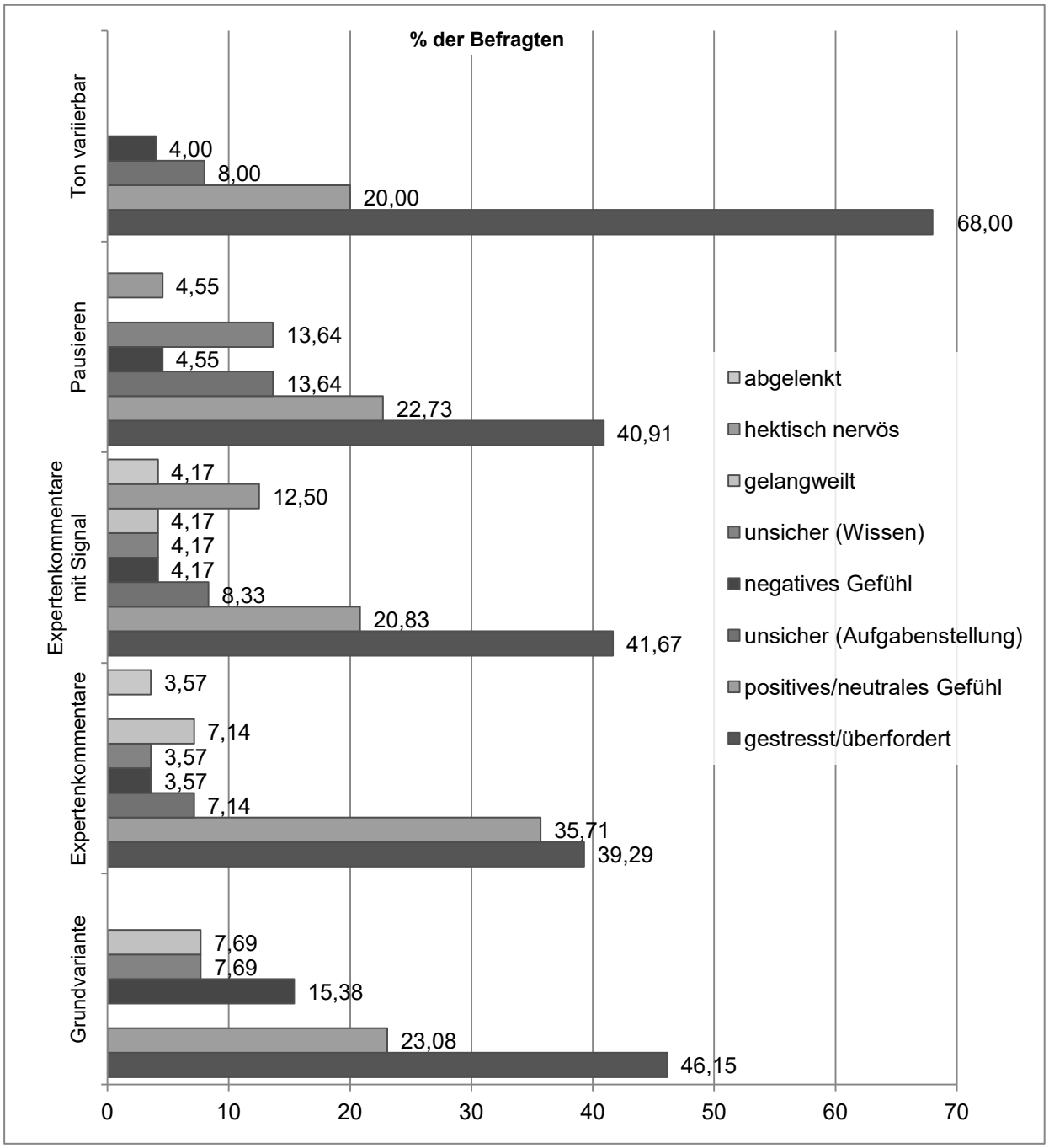


Abbildung 21: Auftretende Gefühle in den Versuchsgruppen.

Schließlich fällt auf, dass das Szenario „Pausieren“, das zuvor insbesondere wegen der Reduktion des Zeitdrucks bei der Bearbeitung eher mit positiven Ausgängen assoziiert wurde, offensichtlich dazu führte, dass die Probanden Gefühle der Unsicherheit sowohl in Bezug auf die Passung der eigenen Arbeit im Hinblick auf die Aufgabenstellung als auch im Hinblick auf das eigene Wissen entwickeln. Offensichtlich führt die zur Verfügung stehende Zeit eher zum Infragestellen der eigenen Einschätzungen, wohingegen sich diese Möglichkeit in den übrigen

Varianten nicht ergibt und die Probanden zur Zeitersparnis kommende Gedanken eher einfach aufschreiben und absenden um sich wieder der Beobachtung widmen zu können.

Es bleibt einschränkend festzuhalten, dass die Interviews erst nach der Bearbeitung der Kommentaraufgabe geführt wurden. Es handelt sich demnach um eine retrospektive Betrachtung, die daher fehlerbehaftet sein kann (Pohl & Haracic, 2005). Dennoch liefert die Auswertung erste Anhaltspunkte darüber, wie sich auch die späteren Experten der KNoTen-Station während der Bearbeitung fühlen könnten, was sowohl Einfluss auf ihre Leistung als auch auf ihre Motivation während der Bearbeitung und die Absicht, sich langfristig als Experte für mehrere Trainingsdurchläufe zur Verfügung zu stellen, haben könnte.

5.3.3 Verbesserungsvorschläge der Probanden

In den 129 Interviews beantworteten 127 Probanden die Frage nach Verbesserungsvorschlägen. Insgesamt wurden 256 Vorschläge unterbreitet, die sich wie folgt auf die Versuchsgruppen verteilten (s. Tabelle 12).

Tabelle 12: Anzahl der genannten Verbesserungsvorschläge nach Varianten.

Variante	Grund-variante	Experten-kommentare	Experten-kommentare mit Signal	Pausieren	Ton variierbar
Anzahl der Probanden insgesamt	26	28	23	27	25
Anzahl der Verbesserungsvorschläge	59	60	47	46	44

Im Rahmen der qualitativen Inhaltsanalyse wurden elf Kategorien ermittelt, in die sich die genannten Vorschläge einordnen lassen. Ihre Bezeichnung sowie die Häufigkeit ihres Vorkommens sind der nachstehenden Tabelle zu entnehmen. Die Kategorien werden im Folgenden vorgestellt (s. Tabelle 13) und hinsichtlich ihrer Umsetzbarkeit im Rahmen des KNoTen-Projektes diskutiert.

Tabelle 13: Verbesserungsvorschläge und ihre Häufigkeiten

Bezeichnung	Häufigkeit des Vorkommens
Videoanzeige	120
Kommentieren	46
Pausieren	35
Dauer der Videos	11
Vorbereitung für die Bearbeitung	9
Geschwindigkeit	9
Aufgabenstellung	7
Andere Kommentare	5
Ton	4
Signalton	3
Sonstige	7

5.3.3.1 Verbesserungsvorschläge zur Videoanzeige

Die am häufigsten vertretene Kategorie *Videoanzeige* umfasst sämtliche Vorschläge, die sich darauf beziehen, was der Nutzer auf dem Monitor seines Rechners sieht. Hierunter finden sich insgesamt sechs Unterkategorien: *die Anzahl der integrierten Videos, die Qualität der Videoaufnahmen, die Größe der Videoanzeigen*, *die Anordnung der Kommentarfenster* sowie *die Anordnung der Videoanzeigen*. In der Tabelle 14 sind die genannten Subkategorien zusammenfassend beschrieben. Zur Komprimierung der Inhalte wurde auf eine Anführung wörtlicher Zitate verzichtet werden, da die Aussagen eindeutig verständlich waren.

Hier handelt es sich prinzipiell um umsetzbare Verbesserungsvorschläge, da es sich weitgehend um allgemeine Gestaltungsgrundsätze handelt. Besonders wichtig ist die Qualität der Videoaufnahmen und hinsichtlich der Anordnung der Kommentar- und Videofenster müssten vor Inbetriebnahme der KNoTen-Station weitere Tests mit den Nutzern durchgeführt werden. Dabei sollte ermittelt werden, welche Anordnung zu den besten Leistungen führt bzw. von den Anwendern akzeptiert wird.

Tabelle 14: Verbesserungsvorschläge zur Videoanzeige

Subkategorie	Beschreibung	Anzahl der Aussagen
Anzahl der integrierten Videos	Reduktion der Anzahl der integrierten Videos oder die Beobachtung immer nur eines Videos zu einem Zeitpunkt wäre sinnvoll.	68
Qualität der Videoaufnahmen	Bildqualität der verwendeten Videos müsste verbessert werden. Angesprochen wurden die Erhöhung der Bildauflösung, Unterbinden von Kamerawackeln und die günstigere Positionierung der Kameras zur Aufzeichnung der Videos, so dass die Nutzer einen besseren Überblick haben.	21
Größe der Videoanzeigen	Videoanzeigen sollten vergrößert werden (z.B. durch eine Lupenfunktion, bei der durch die Bewegung der Maus über das jeweilige Video dieses vorübergehend größer dargestellt wird).	14
Inhalt der Videos	Inhalt der Videos sollte verändert werden. Es wird u.a. Abwechslungsreichtum innerhalb des Videomaterials, die Verwendung von den Probanden eher unbekannten Drehorten, die Festlegung auf entweder nur Stand- oder nur bewegte Bilder etc. gewünscht.	13
Anordnung der Kommentarfenster	Position des Eingabefensters, in das die Kommentare eingetragen werden, sollte verändert werden; für jede Videoanzeige sollte es ein eigenes Eingabefeld geben, das sich jeweils unter oder im Video selbst befindet	3
Anordnung der Videoanzeige	Anordnung der Videos untereinander	1
	Summe	120

Einzig die Subkategorien *Inhalt der Videos* und die *Anzahl der integrierten Videos* erscheinen zumindest zu einem Teil nicht beliebig anpassbar. Da der *Inhalt der Videos* aus dem angedachten, komplexen Notfall-Management-Training bestehen wird, scheidet eine Gestaltung des Abwechslungsreichtums oder der Bekanntheit der Drehorte für die Experten aus. Die angesprochene Variabilität der Videoinhalte ergibt sich aus den geplanten Trainingsinhalten oder bspw. aus den simulierten Notfällen. Die Vielzahl von zu kommentierenden Verhaltensweisen jedoch ist kaum steuerbar, da sie vom Trainingsstand der Trainierenden abhängt. So werden weniger Trainierte wahrscheinlich unterschiedlichere und insgesamt mehr Fehler begehen, die kommentiert werden müssten, als Teilnehmer mit einem sehr guten Trainingsstand. Die auf den Videos ersichtlichen Drehorte, als ein genannter Aspekt der Subkategorie *Inhalt der Videos*, sind ebenfalls nicht beliebig wählbar und richten sich nach den Erfordernissen des Trainings an sich. Für eine Übung eines Brandnotfalls ist bspw. das Training an einem dafür geeigneten Ort durchzuführen, der dann auch auf den Videos erscheint. Darauf, ob ein Experte diesen speziellen Ort kennt oder nicht, kann jedoch nicht geachtet werden.

Auch hinsichtlich der Variation der *Anzahl der integrierten Videos* besteht wenig Spielraum. Die KNoTen-Station versucht es zu ermöglichen, dass sich ein komplexes Training an mehreren Übungsorten gleichzeitig durchführen lässt und das Zusammenspiel mehrerer Teams in ein und derselben Notfallsituation geübt werden kann. Darüber hinaus soll allen Teilnehmern gleichermaßen ein umfassendes Feedback gegeben werden. Dazu ist es notwendig, alle Teams zu filmen und von den Experten parallel kommentieren zu lassen. Die Anzahl der gezeigten Videos hängt demnach von der Anzahl der Trainingsteams ab. Eine einfache Reduktion der Videos wäre daher nicht sinnvoll, weil in diesem Fall zum einen nicht das komplette Zusammenspiel der Teams ersichtlich wäre und zum anderen nicht jede Gruppe nach Ende des Trainings Feedback erhalten würde. Eine Möglichkeit zur Verringerung der Anzahl könnte durch die sequentielle Kommentierung der einzelnen Videos geschaffen werden. So könnten sich die Experten jeweils auf nur ein Video konzentrieren. Nachteile wären aber hierbei, dass ebenfalls das Zusammenspiel der beteiligten Teams nicht so deutlich ersichtlich wäre und eine sequenzielle Bearbeitung je nach Anzahl der Videos ein Vielfaches an Zeit beanspruchen würde. Dadurch würde sich die Feedback- und Auswertungsrunde für die Trainierenden entsprechend verzögern. Hinsichtlich dieses Aspektes der Gestaltung der Videoanzeige ist demnach damit zu rechnen, dass ein Trade-off zwischen Zeit und Entlastung durch weniger Videos in Kauf genommen werden müsste. Hinsichtlich dieses Aspektes sollten weitere Forschungsarbeiten durchgeführt werden, die sich mit der Frage auseinandersetzen, welche Anzahl von eingebundenen Videos für die Experten optimal ist.

5.3.3.2 Verbesserungsvorschläge zum Kommentieren

In der Kategorie *Kommentieren* wurden alle Verbesserungsvorschläge zusammengefasst, die sich damit beschäftigten, auf welche Weise das laut Aufgabenstellung abzugebende Feedback in den Computer eingegeben werden sollte. Hierbei konnten vier Subkategorien identifiziert werden (s. Tabelle 15).

Die Vielzahl der Vorschläge, die sich auf die Umstellung des Kommentierens auf eine automatisierte *Spracheingabe* beziehen, zeigt deutlich, dass das Eingeben des Feedbacks mittels Tastatur nicht als eine optimale Lösung angesehen wird. Die Eingabe auditiver Kommentare, die vom Computer in Schrift umgewandelt werden, könnte dadurch erfolgen, dass durch einen Mouse-Klick in das jeweilige Video diesem eine Tonspur zugewiesen und mit Hilfe von Sprache-zu-Text-Programmen zu einem schriftlichen Kommentar umgewandelt wird. Eine solche Variante ist für die Weiterentwicklung der KNoTen-Station durchaus denkbar, zumal die heute

vorhandenen technischen Möglichkeiten mittlerweile eine hohe Zuverlässigkeit besitzen und auch auf unterschiedliche Dialekte der Sprecher robust reagieren können (Römer, 2009).

Tabelle 15: Verbesserungsvorschläge zum Kommentieren

Sub-kategorie	Beschreibung	Anzahl der Aussagen
Spracheingabe	Integration einer Funktionalität in die KNoTen-Station, die gesprochene Kommentare automatisch verschriftlicht.	21
standardisierte Eingabehilfen	Einführung einiger standardisierter Eingabehilfen, z.B. Buttons, für wiederkehrende Ereignisse oder erneut betroffene Personen. Der zu erstellende Kommentar kann durch die Betätigung des Buttons ganz oder zumindest zum Teil ersetzt werden.	11
Nachträglich Kommentieren	Formulierung der Kommentare nach der Betrachtung der Videos. Während der Beobachtung sollen handschriftliche Notizen gemacht werden.	8
Sichtbare Markierung im Video	Markieren derjenigen Personen, die ein Fehlverhalten zeigen. Mit dem Markieren wird angezeigt, wo im Bild bzw. bei welcher Person sich das Fehlverhalten gezeigt hat.	6
	Summe	46

Standardisierte Eingabehilfen könnten die Experten dahin gehend entlasten, als dass für immer wieder auftretende Verhaltensweisen oder Personen die entsprechenden Beschreibungen per Klick automatisch in die Kommentare einfügt werden. Dies setzt voraus, dass das zu beobachtende Verhalten oder die zu beobachtenden Personen standardisiert beschrieben werden können. In den Versuchen, die in dieser Studie durchgeführt wurden, war dies möglich. Es handelte sich um Verkehrsverstöße von Fußgängern, Auto- oder Fahrradfahrern, die immer wieder in allen vier Videos auftraten. In diesen Fällen würden Buttons für die betroffenen Personen (z.B. ein Button für Fußgänger) oder das gezeigte Verhalten (z.B. ein Button für Überqueren der Straße bei auf rot geschalteter Ampel) durchaus Sinn machen. Eine Übertragbarkeit auf die KNoTen-Station erscheint allerdings schwierig: Die Trainingsteilnehmer mit Hilfe von standardisierten Buttons werden kaum voneinander zu unterscheiden sein. Zwar ist vorstellbar, dass man für einige Trainingsteilnehmer (z.B. Kapitän sowie die höherrangigen Offiziere) Eingabehilfen erstellen kann, eine Differenzierung von Übungsteilnehmern ohne sichtbaren Status erscheint jedoch kaum möglich. Auch die zu kommentierenden Verhaltensweisen bieten aufgrund ihrer möglichen Vielfalt wenige Anhaltspunkte zur Standardisierung, v.a. da die Experten aus unterschiedlichen Fachgebieten kommen und deshalb auf ihre Domäne abgestimmte Schaltflächen zu entwickeln wären.

Das vorgeschlagene *nachträgliche Kommentieren* der Videos scheint generell weniger geeignet zu sein, da in diesem Fall sogenannte Rückschaufehler eintreten können. Dieser denkpsychologische Effekt besagt, dass durch neue Informationen Einschätzungen von Sachverhalten nachträglich verzerrt werden (Pohl & Haracic, 2005). Im Fall der KNoTen-Station würde dies bedeuten, dass nachträglich geschriebene Feedbackkommentare zu einem anderen Ergebnis führen würden als sofort geschriebene. Die Verzerrung resultiert aus der Hinzuziehung von Informationen, die eine Person im Laufe der Zeit gewinnt, also bspw. durch Informationen, die der Experte im weiteren Verlauf seiner Beobachtungen aufnimmt. Zeigt ein Trainierender bspw. zu Beginn des Trainings positiv zu bewertende Verhaltensweisen und verschlechtert sich im Verlauf des Trainings, so könnte das Verhalten aufgrund dieser Beobachtungen insgesamt negativer bewertet werden. Zudem würde eine Kommentierung erst im Anschluss an das Beobachten mehr Zeit in Anspruch nehmen als die parallele Durchführung beider Aufgaben. Demnach könnte diese Variante zu einer Verzögerung des Feedbackgesprächs nach dem Training führen. Aus beiden Gründen scheint der Vorschlag, die Beobachtungen erst später mit Kommentaren zu versehen, weniger geeignet.

Die Markierung einer Person zum Zeitpunkt ihres besonders guten oder besonders negativen Verhaltens im Video (*sichtbare Markierung im Video,* in der Fachsprache als „Taggen“ bezeichnet) könnte eine sinnvolle Ergänzung der Funktionalitäten der KNoTen-Station bilden, insbesondere in unübersichtlichen Trainingsmomenten. Da der Vorschlag aber gänzlich von Feedbackkommentaren absieht, die das eigentliche Ziel im KNoTen-Projekt darstellen, kann es sich bestenfalls um eine Ergänzung der Funktionalitäten handeln (würde ganz auf schriftliche Kommentare verzichtet, müssten bei der Auswertung der Videos wieder Experten zugegen sein, die die markierten Videosequenzen dann mit den Trainierten auswerten, womit der Vorteil der Ortsunabhängigkeit der Experten verloren ginge).

5.3.3.3 Weitere Verbesserungsvorschläge

Die qualitative Auswertung der Interviews ergab neben den genannten Kategorien weitere Ansatzpunkte zur Verbesserung. Diese werden im Folgenden gemeinsam betrachtet, da sie weniger ausdifferenziert sind. Unter den Kategorien finden sich vier, die sich direkt auf Varianten beziehen, die im Rahmen der vorliegenden Studie getestet wurden. Um Redundanzen zu vermeiden, werden diese Kategorien erst in den Implikationen für das KNoTen-Projekt (KAP. 6.2) erläutert. Es handelt sich um die Kategorien *Pausieren*, *andere Kommentare sichtbar*, *Ton* und *Signalton*, die weitgehend mit den getesteten Varianten, „Pausieren“, „Expertenkommentare“,

„Ton variierbar“ und Expertenkommentare mit Signal“ übereinstimmen. Tabelle 16 gibt einen Überblick über die weiteren Kategorien der genannten Verbesserungsvorschläge.

Tabelle 16: Weitere Verbesserungsvorschläge

Kategorie	Beschreibung	Anzahl der Aussagen
Pausieren	Integration einer Funktionalität, die es ermöglicht, die Videos während des Schreibens anzuhalten (einmal wird Möglichkeit eine Erholungspause einzulegen, gewünscht).	35
Vorbereitung	Vorbereitung auf den Inhalt (Verkehrsregeln) bzw. die Form der Aufgabe (Videos beobachten und kommentieren) wäre hilfreich.	10
Dauer der Videos	Die Videosequenzen sollen kürzer sein.	9
Geschwindigkeit	Möglichkeit, die Videos in langsamerer Geschwindigkeit abspielen zu lassen bzw. die Geschwindigkeit selbst steuern zu können.	9
Aufgabenstellung	Eine klarere Aufgabenstellung mit Hinweisen darauf, worauf in den Videos zu achten ist	7
Andere Kommentare sichtbar	Die Probanden aus den Varianten „Expertenkommentare“ und „Expertenkommentare mit Signal“ sprechen sich dafür aus, die Kommentare anderer Experten nicht einzublenden.	5
Ton	Ton abstellen.	4
Signalton	Im Szenario Expertenkommentare mit Signal den verwendeten Hinweiston weglassen.	3
Sonstige	Nachträgliche Bearbeitung bereits fertiger Kommentare, Verwendung größerer Bildschirme bzw. von vertrauten Tastaturen, Möglichkeit, zurückzuspulen.	4
	Summe	86

Zehn Probanden äußerten, dass sie eine *Vorbereitung* auf die Arbeit mit der KNoTen-Station sinnvoll fänden. Hierbei wurde differenziert in die inhaltliche Vorbereitung und die Vorbereitung auf die Handhabung der Software und die Erfüllung der Aufgabenstellung. Hinsichtlich der inhaltlichen Vorbereitung führten die Befragten an, dass sie sich zuvor mit den geltenden Verkehrsbestimmungen hätten auseinandersetzen wollen. Diese Art der Vorbereitung erscheint aufgrund des Expertenstatus der Kommentatoren im KNoTen-Projekt weniger sinnvoll. Die Vorbereitung auf die Handhabung der Nutzeroberfläche und die Gewährung einer Eingewöhnungszeit bspw. durch Testläufe für jeden Experten erscheint jedoch sinnvoll, da nicht davon ausgegangen werden kann, dass alle Experten im Umgang mit der Software oder mit Computertechnik im allgemeinen geübt sind.

Ein weiterer Vorschlag zur Verbesserung der Feedbackleistungen bezog sich auf die *Dauer der Videos*. Diese betrug in allen Experimentalgruppen acht Minuten. Alle Befragten, die sich hierzu äußerten, gaben an, sie hätten bessere Leistungen erbracht, wenn die Videos kürzer gewesen wären und schlagen deshalb eine Reduktion der Dauer vor. Auch dieser Vorschlag ist kaum auf die KNoTen-Station übertragbar, da sich die Dauer der Videos nach der Länge des Trainings bemisst. Werden umfangreichere Trainings durchgeführt, verlängert sich demnach die Videodauer. Darüber hinaus müssten auch weitere Forschungsarbeiten zeigen, ob sich tatsächlich eine Verbesserung der Leistungen durch eine Verkürzung der Versuchszeit ergibt.

Die *Geschwindigkeit* variieren zu können, und vor allem die Möglichkeit eines langsameren Abspielens nannten neun Befragte als Verbesserungsmöglichkeit. Sie führten an, dies würde sich positiv auf das Kommentieren auswirken, da ohne Zeitdruck während des Schreibens weniger Ereignisse verpasst würden. In diesem Fall würde aber die Übertragung der Videos nicht mehr in Echtzeit erfolgen und demnach wäre mit einer zeitlichen Verzögerung für das Auswertungsgespräch zu rechnen. Insofern weist dieser Vorschlag eine hohe Ähnlichkeit zur Variante „Pausieren“ auf, die im Diskussionsteil näher beleuchtet wird.

Sieben Probanden gaben an, dass ihnen eine klarer formulierte *Aufgabenstellung* zu besseren Leistungen verholfen hätte. Dies zeigt, dass die verwendeten Aufgabenstellungen anscheinend nicht eindeutig genug formuliert waren. Das ist als eine Limitation der vorliegenden Studie zu betrachten, schärft aber den Blick dafür, dass vor dem Einsatz der KNoTen-Station im Training den Experten die Anwendungshinweise eindeutig zu vermitteln sind. Dabei müssen auch die Grundregeln effektiven Feedbacks vermittelt werden, um die Experten bei ihrer Aufgabe zu unterstützen.

Hinter den *sonstigen* Vorschlägen verbergen sich seltener genannte Ideen der Probanden. Da sich diese alle nicht auf die KNoTen-Station übertragen lassen, werden sie hier nicht näher kommentiert.

5.3.4 Zusammenfassung

Insgesamt lieferten die Teilnehmer der Studie viele Hinweise, die sich bei der Implementierung und Inbetriebnahme der KNoTen-Station als wertvoll erweisen können. So wurden zahlreiche Hinweise gegeben, worauf bei der Gestaltung der Videoanzeige zu achten ist und welche zusätzlichen Verbesserungen für die Vereinfachung der Eingabe der Kommentare denkbar wären. Hiervon sind einige prinzipiell umsetzbar, erfordern allerdings weitere Forschungsarbeiten und Testläufe, um die zu erzielende Wirkung zu überprüfen. Einige Vorschläge wie nachträgliches

Kommentieren, standardisierte Eingabehilfen oder eine Verkürzung der Videosequenzen betreffen, sind jedoch nicht auf die KNoTen-Station übertragbar und machen nur im Kontext der in der Studie verwendeten Verkehrsvideos Sinn.

Weiterhin bleibt festzuhalten, dass sich vier der Vorschlagskategorien direkt auf die getesteten Varianten der KNoTen-Station bezogen. Eine detailliertere Erörterung dieser Kategorien wird daher in der Diskussion der Handlungsempfehlungen für die KNoTen-Station erfolgen.

6 Diskussion

6.1 Zusammenfassung

Das Ziel der vorliegenden Studie bestand darin, ausgehend vom im Forschungsprojekt KNoTen 2.0 entwickelten Prototypen der KNoTen-Station, vier weitere Varianten dieses Hard- und Softwareaufbaus mit zusätzlichen Funktionalitäten im Hinblick auf ihre Eignung zur Erstellung eines wirkungsvollen Feedbacks für die Auswertung von komplexen Realübungen im Bereich der Sicherheitsausbildung in der Seeschifffahrt sowie im Hinblick auf die Auswirkungen der Arbeit mit der Grundvariante bzw. der Variationen zu testen. Des Weiteren sollten zusätzliche Optimierungsmöglichkeiten für die KNoTen-Station ermittelt werden, die gegebenenfalls in weiteren Forschungsarbeiten eruiert werden können.

Auf Basis eines ausführlichen Literaturstudiums wurde hierfür zunächst versucht die Anforderungen herauszustellen, die erfüllt sein sollten, um ein Feedback, das Trainierende üblicherweise nach der Durchführung von Trainingsmaßnahmen erhalten, möglichst wirkungsvoll zu gestalten. Daraufhin wurde dargestellt, welchen Herausforderungen die Experten, die mit Hilfe der KNoTen-Station ein solches Feedback erstellen sollen, gegenüberstehen. Es zeigte sich, dass dies eine sehr anspruchsvolle Aufgabe ist, die nur mit Hilfe ausgeprägter Wahrnehmungs- und Konzentrationsfähigkeiten zu bewältigen ist.

Im darauffolgenden Schritt wurde darauf aufbauend der wahrnehmungs- und konzentrationspsychologische Forschungsstand zusammengetragen, der Hinweise darauf liefern sollte, wie sich bestimmte zusätzliche Funktionalitäten der KNoTen-Station auf die Leistung der Experten bei der Feedbackerstellung sowie auf ihr Befinden auswirken. Bei diesen zusätzlichen Funktionalitäten handelte es sich um 1) das Einblenden von Feedbackkommentaren anderer Experten, die gleichzeitig die selbe Aufgabe bearbeiten, 2) das Einblenden anderer Expertenkommentare begleitet von einem zusätzlichen Signalton, 3) die Möglichkeit während des Verfassens der Feedbackkommentare die zu beobachtenden Videosequenzen anzuhalten und 4) die Möglich-

keit, den Ton der Videosequenzen gezielt zu steuern. Da es aufgrund der widersprüchlichen Forschungsergebnisse nicht möglich war, gerichtete Hypothesen zur Wirkung der Funktionalitäten abzuleiten, wurden offene Forschungsfragen formuliert, die mit Hilfe einer explorativen, experimentellen Untersuchung mittels einer selbst erstellten Software-Nachbildung der noch in der Konstruktion befindlichen KNoTen-Station, gefolgt von einer offenen Befragung mit insgesamt 129 Probanden beantwortet wurden. Als Leistungsmerkmale wurden die Anzahl der geschriebenen Kommentare, die Anzahl der einbezogenen Videos, die Ausführlichkeit der geschriebenen Kommentare, die Tippgeschwindigkeit und die Gesamtdauer der Feedbackerstellung erhoben. Hierbei handelte es sich um rein quantitative Merkmale der Leistung der Experten. Um auch Aussagen darüber treffen zu können, welche Varianten qualitativ Vorteile oder Nachteile aufweisen, wurde darüber hinaus die Richtigkeit der Kommentare bezogen auf eine Musterlösung sowie die Verständlichkeit der Kommentare auf Basis eines Scoring-Verfahrens erhoben. Das Befinden der Probanden während der Arbeit mit der Software, die erlebte Schwierigkeit der bearbeiteten Aufgabe sowie weitere Verbesserungsmöglichkeiten wurden in einem Interview mit den Probanden erhoben. Die Daten wurden mit Hilfe von parametrischen und parameterfreien statistischen Tests sowie mittels qualitativer Inhaltsanalyse ausgewertet.

Die Ergebnisse der Studie sollen im Folgenden für jede der vier Varianten zusammengefasst und im Hinblick auf ihre Vor- und Nachteile diskutiert werden, um so zu Schlussfolgerungen zu gelangen, welche zusätzlichen Funktionalitäten eine Bereicherung für den aktuellen Prototypen der KNoTen-Station darstellen würden. Darüber hinaus werden Handlungsempfehlungen für die Vorbereitung der Arbeit mit der KNoTen-Station mit den Seefahrtsexperten gegeben und die Grenzen der Aussagekraft der Untersuchung dargestellt.

6.2 Bedeutung der Ergebnisse für das Forschungsprojekt KNoTen 2.0

6.2.1 „Expertenkommentare“

Die erste zu überprüfende Variante der KNoTen-Station wurde im Rahmen der vorliegenden Studie „Expertenkommentare“ genannt. Im Vergleich zum Grundaufbau der KNoTen-Station, die für alle Auswertungen die Kontrollgruppe darstellte, wurden bei dieser Variante während der Feedbackerstellung Kommentare fiktiver anderer Personen eingeblendet, wobei den Probanden suggeriert wurde, dass diese Personen parallel an der gleichen Aufgabe arbeiten. Bei der statistischen Auswertung bezüglich der Leistungsindikatoren und der erlebten Aufgabenschwierigkeit zeigten sich zwischen dieser und der Grundvariante keine signifikanten Unterschiede. Tendenziell, also bei der bloßen Betrachtung der Mittelwerte[10], wird jedoch deutlich, dass die Variante „Expertenkommentare“ hinsichtlich der Richtigkeit der Kommentare überlegen zu sein scheint. Hierbei wies die Variante den zweithöchsten Mittelwert aller untersuchten Gruppen auf. Eine mögliche Erklärung könnten die zusätzlichen Kommentare liefern, die den Probanden eventuell als Hilfestellung gedient haben. Die eingeblendeten Kommentare wurden auf Basis der Musterlösung erstellt und stellten damit korrekte Antworten dar, an denen sich die Probanden orientieren konnten.

Hinsichtlich der Verständlichkeit jedoch schnitt die Variante „Expertenkommentare“ – mit dem in allen Gruppen geringsten Mittelwert – am schlechtesten ab, so dass befürchtet werden muss, dass in dieser Variante entstandene Kommentare zwar vielleicht inhaltlich angebracht, sprachlich aber so formuliert sind, dass die gewünschte Botschaft eventuell nicht beim Feedbacknehmenden ankommt. Obwohl die deskriptiven Daten nur sehr geringe und nicht signifikante Unterschiede aufweisen, könnte vermutet werden, dass die geringere Verständlichkeit mit der vergleichsweise geringeren Ausführlichkeit in Verbindung steht. Bei diesem Indikator zeigte sich in der Gruppe „Expertenkommentare“ im Vergleich zu den anderen Gruppen der zweitniedrigste Mittelwert und der geringste Median. Dies könnte darin begründet sein, dass kürzere Kommentare weniger Anhaltspunkte enthalten können, die der Nachvollziehbarkeit der Inhalte der Kommentare dienen. Die geringere Ausprägung der Ausführlichkeit könnte hierbei ein Hinweis auf das Auftreten von Social Loafing gedeutet werden. Hierbei engagieren sich Individuen weniger wenn sie mit anderen gemeinsam arbeiten, als wenn sie allein mit einer Aufgabe betraut sind (Karau & Williams, 1993). So könnte die durch das Einblenden der anderen Kommentare induzierte Vorstellung, nicht allein an der Aufgabe zu arbeiten, dazu geführt haben, dass Kommentare kürzer gefasst werden, weil die Probanden eventuell davon ausgingen, dass die Kom-

[10] Anhang 4 zeigt die hier diskutierten deskriptiven Statistiken.

bination ihrer Statements mit denen der anderen ein ausreichend umfangreiches Bild der betrachteten Verkehrssituationen liefern. Möglicherweise nehmen sie an, dass eine schon von einem anderen verfasste Aussage nicht noch einmal von ihnen mit einem sehr ausführlichen Kommentar wiederholt werden müsste.

Obwohl demnach bei den statistisch auswertbaren Daten keine Vorteile bzw. ausgehend von der rein deskriptiven Betrachtung sogar eher Nachteile der Variante „Expertenkommentare“ gegenüber der „Grundvariante“ sichtbar wurden, zeigte die Auswertung der Interviewdaten hinsichtlich des Wohlbefindens während der Aufgabenbearbeitung ein anderes Bild. Hier erzielte die Variante „Expertenkommentare“ weit höhere Werte hinsichtlich positiver oder zumindest neutraler Gefühle. Mehr als 35 Prozent der Befragten gaben von sich aus an, sich bei der Bearbeitung nicht schlecht gefühlt zu haben. In den übrigen Gruppen lag dieser Wert nur zwischen 20 und gut 23 Prozent. Dies könnte darin begründet sein, dass unter anderem das Vorhandensein sozialer Kontakte Auswirkungen auf das Wohlbefinden von Personen hat. „So wird die Zeit, die man alleine verbringt, als weniger angenehm beurteilt als die Zeit, die man in Gesellschaft anderer Personen verbringt“ (Nestler, Back & Egloff, 2011, S. 57).

Auch hinsichtlich der Äußerung von erlebtem Stress oder Überforderung zeigten die Probanden der Variante „Expertenkommentare“ geringere und damit positivere Werte. Die Wahrnehmung anderer Personen könnte hierbei als Möglichkeit erlebt worden sein, zumindest Teile der Aufgabe zu delegieren, was zum Gefühl persönlicher Entlastung führen kann und dadurch das Stressgefühl reduziert (Hintz, 2011). Gerade in diesem Punkt stellt sich jedoch die Frage, ob diese Entlastungsempfindung auch bei dem Bewusstsein entsteht, dass die anderen Kommentatoren aus anderen Fachgebieten als dem eigenen stammen, wie es im KNoTen-Projekt angedacht ist. In diesem Fall sollten die anderen Kommentatoren als nicht geeignet einschätzt werden, die spezifische Aufgabe zu übernehmen, Feedback aus der eigenen Perspektive zu geben.

Insgesamt erscheint die Variante „Expertenkommentare“ hauptsächlich aus motivationaler Sicht vorteilhaft für die Arbeit mit der KNoTen-Station. Dennoch sollten gerade der Aspekt reduzierten Stressempfindens insbesondere in weiteren Tests mit Experten gleicher und unterschiedlicher Fachgebiete untersucht werden, um noch eindeutigere Aussagen treffen zu können.

6.2.2 „Expertenkommentare mit Signal“

Die Variante „Expertenkommentare mit Signal“ stellte eine Erweiterung der zuvor erörterten Variante „Expertenkommentare“ dar. Zusätzlich zur Einblendung der anderen Kommentare wurde mit Hilfe eines beim Erscheinen der Kommentare erklingenden Signaltons versucht, die

Wahrnehmung der zusätzlichen Kommentare sicherzustellen. Bei der statistischen Auswertung konnte festgestellt werden, dass diese Variante, anders als die Variante ohne Signal, einen signifikanten Unterschied zur Grundvariante bezüglich der Anzahl der verfassten Kommentare aufweist. Da der Signalton die einzige Abweichung zwischen dieser und der Expertenkommentar-Variante darstellte, muss der demnach für den zu verzeichnenden Unterschied verantwortlich sein. Die Wirkung könnte dabei auf verschiedene Weisen entstanden sein. Zum einen ist denkbar, dass der Signalton dazu geführt hat, dass die Expertenkommentare durch die vorgenommene Hervorhebung tatsächlich besser wahrgenommen werden und so die in Kapitel 3 dargestellten möglichen positiven Einflüsse der Expertenkommentare verstärkt haben. Andererseits könnte aber auch der Signalton allein dazu geführt haben, dass die Probanden immer wieder neu aktiviert wurden und so das allgemein höhere Erregungsniveau (Tomasi, Caparelli, Chang & Ernst, 2005) zu der höheren Quantität der Leistung geführt hat.

Neben diesem einzigen signifikanten Befund deuten sich bei der Betrachtung der deskriptiven Statistiken aller Gruppen im Vergleich weitere rein quantitative Vorteile dieser Variante an. So besaß die Gruppe „Expertenkommentare mit Signal“ den höchsten Mittelwert bei der Anzahl der einbezogenen Videos, die höchste Tippgeschwindigkeit und die geringste Gesamtversuchsdauer. Obwohl diese Unterschiede nicht signifikant waren, also nicht auf die Grundgesamtheit übertragbar sind, liefern sie doch weitere Hinweise darauf, dass die quantitativen Leistungsmerkmale durch Integration eines zusätzlichen Tons positiv beeinflusst werden.

Demgegenüber stehen allerdings die Werte der qualitativen Leistungsmaße. Obwohl die Probanden mehr Kommentare schrieben, erhöhte sich die Punktzahl für die Richtigkeit nicht. Das bedeutet, dass die zusätzlichen Kommentare also keinen zusätzlichen Nutzen brachten. Des Weiteren wiesen die Kommentare die zweitgeringste durchschnittliche Verständlichkeit auf, mit dem sogar über alle Gruppen geringsten Median. Dies könnte wie auch zuvor argumentiert mit der vergleichsweise geringen Ausführlichkeit der Kommentare zusammenhängen, die eine Folge der Verantwortungsdiffusion bzw. des Social Loafing durch die gefühlte Präsenz anderer Kommentatoren sein könnten.

Besonders auffällig bei der Betrachtung der Interviewdaten ist, dass allein durch das Hinzufügen des Signaltons eine erhebliche Reduzierung des Wohlbefindens erfolgte. So befand sich der Anteil der Personen, die sich bei der Aufgabenbeantwortung wohl oder zumindest nicht negativ fühlten, auf einem ähnlich hohen Niveau, wie die übrigen Gruppen und erreichte nicht, wie in der Variante ohne Signalton, höhere Werte. Der Signalton scheint demnach das Wohlbefinden negativ zu beeinflussen. Dies zeigt sich auch der hohen Zahl an Aussagen, die beinhalteten,

dass sich Probanden hektisch oder nervös fühlten, was wie oben beschrieben als ein Ausdruck erhöhter Erregung betrachtet werden kann.

Zusammenfassend kann demnach für die Variante „Expertenkommentare mit Signal“ festgehalten werden, dass sich die Vorteile dieser Variante lediglich auf die quantitative Leistung beziehen, nicht aber auf die Qualität der Kommentare, wobei das Wohlbefinden unter der Hinzufügung des zusätzlichen Tons leidet. Demnach ist die Integration eines zusätzlichen Hinweistons zusätzlich zur Einblendung der Kommentare anderer Experten eher nicht zu empfehlen.

6.2.3 „Pausieren“

In der dritten Versuchsgruppe, „Pausieren“,wurden keine Anmerkungen anderer Kommentatoren eingeblendet. Stattdessen wurde versucht, den empfundenen Zeitdruck der Probanden zu reduzieren, indem die gezeigten Videos während des Eingebens eines Kommentars angehalten wurden und erst weiterliefen, wenn der Kommentar abgeschlossen war. Wie in Kapitel 3 beschrieben, war unter anderem zu vermuten, dass die Versuchspersonen unter dieser Bedingung weniger Stress erleben und durch ausreichend zur Verfügung stehende Zeit eine intensivere und somit effektivere Informationsverarbeitung erfolgt. Zusätzlich wurde vermutet, dass die Probanden weniger Ereignisse verpassen würden, was sich sowohl auf die qualitativen als auch die quantitativen Leistungen bei der Feedbackerstellung auswirken kann. Dem gegenüber musste bei dieser Variante aber in Kauf genommen werden, dass durch das Anhalten der Videos die Bearbeitung der Aufgabe insgesamt dementsprechend mehr Zeit erfordert.

Die statistische Auswertung zeigte wie erwartet eine um 60 Prozent höhere Bearbeitungszeit als in der Kontrollgruppe (hochsignifikanter Unterschied, $p=0.000$). So benötigten die Probanden in der „Pausieren“-Gruppe fast 14 Minuten, wohingegen die Kontrollgruppe die Feedbackerstellung entsprechend der Dauer der Videos schon nach acht Minuten abschloss.

Weitere signifikante Unterschiede zur Grundvariante ergaben sich hinsichtlich der Ausführlichkeit der Kommentare, die in der Versuchsgruppe höher und der Tippgeschwindigkeit, die signifikant geringer war. Hinsichtlich der übrigen Leistungsmerkmale und der erlebten Aufgabenschwierigkeit ergaben sich jedoch keine signifikanten Unterschiede. Bei der Betrachtung der deskriptiven Statistiken wird aber zumindest tendenziell deutlich, dass die „Pausieren“-Gruppe den höchsten Mittelwert im Vergleich zu allen übrigen Gruppen hinsichtlich der Richtigkeit aufwies und den zweithöchsten Durchschnittswert bei der Verständlichkeit (wobei in der Grundvariante durchschnittlich aber immer noch die verständlichsten Kommentare verfasst

wurden). Dementsprechend besitzt die getestete Variante also hinsichtlich der Verständlichkeit keinen Vorteil gegenüber der Grundvariante.

Anders als vermutet, konnte durch die Möglichkeit zum Pausieren kein Vorteil hinsichtlich des Wohlbefindens generiert werden. Auch hinsichtlich des erlebten Stress‘ und der gefühlten Überforderung konnten im Vergleich zur Kontrollgruppe keine Verbesserungen erzielt werden. Vielmehr wurde bei der Auswertung der Interviewdaten deutlich, dass die zusätzlich zur Verfügung stehende Zeit auch negative Effekte haben kann. So berichteten die Probanden dieser Gruppe vermehrt vom Entstehen von Unsicherheit hinsichtlich ihres Fachwissens und hinsichtlich ihres Verständnisses der Aufgabenstellung bzw. bei der Steuerung der Software. Es ist zu vermuten, dass durch die zusätzliche Zeit die Probanden Gelegenheit hatten, ihre eigenen ersten Gedanken in Frage zu stellen. Interpretiert man diese Unsicherheit als eine geringere Selbstwirksamkeit bei der Aufgabenbewältigung, so ist gemäß der Theorie des geplanten Verhaltens (Ajzen, 1991) zu vermuten, dass so die Absicht sinkt, Kommentare zu verfassen, was dazu geführt haben kann, dass weniger und somit weniger richtige Kommentare geschrieben wurden, als grundsätzlich möglich gewesen wären. Hierbei stellt sich allerdings gerade im Hinblick auf die Unsicherheit bezüglich des eigenen Wissens die Frage, ob dieses Phänomen auch bei Experten mit langjähriger Erfahrung in ihrem Fachgebiet auftreten würde. Auch hinsichtlich der Unsicherheit bezüglich der Aufgabenstellung ist denkbar, dass weniger Bedenken bei den Experten entstehen, da ihnen mit großer Wahrscheinlichkeit sogenannte Drehbücher in Vorbereitung auf die durchgeführte Notfallübung zu Verfügung stehen werden, die den Experten Informationen über den Ablauf und so Hinweise auf potenzielle Verhaltensverstöße der Trainierenden liefern. Unsicherheiten bezüglich dessen, was in den Videos zu kommentieren ist, sollten sich so reduzieren lassen. Hier müssten weitere Tests mit Personen mit Expertenstatus und unter Einbeziehung von Drehbüchern durchgeführt werden, um sichere Aussagen treffen zu können.

Auf Basis der Ergebnisse der vorliegenden Studie bleibt es zum aktuellen Zeitpunkt in Anbetracht der positiven Tendenzen hinsichtlich der Qualität der Kommentare, denen eine signifikante Erhöhung der Bearbeitungszeit gegenübersteht, die mit einer Verzögerung des Auswertungsgespräches einhergeht sowie gegebenenfalls steigende Honorarkosten für die Experten verursacht, eine Kosten-Nutzen-Abwägung, ob diese Variante implementiert werden sollte.

6.2.4 „Ton variierbar“

In der letzten untersuchten Gruppe war es den Probanden möglich, statt der Geräusche aller angezeigten Videos per Mausbewegung auf eines der Videos nur einen Tonkanal zu aktivieren, während die anderen stumm geschaltet wurden. In dieser Gruppe wurden weder andere Exper-

tenkommentare eingeblendet noch war ein Pausieren möglich. Im Vergleich zur Kontrollgruppe ergaben sich keine statistisch signifikanten Unterschiede hinsichtlich der erhobenen Leistungsindikatoren oder der erlebten Aufgabenschwierigkeit. Auch bei einer rein deskriptiven Betrachtung der Mittelwerte und Mediane ergaben sich keine nennenswerten Unterschiede zu den anderen Gruppen, so dass angenommen werden kann, dass die Reduktion der Geräuschvielfalt weder einen positiven noch einen negativen Einfluss auf die Leistung und die erlebte Aufgabenschwierigkeit mit den damit verbundenen Abbruchtendenzen ausübt.

Diesem eher neutralen Ergebnis stehen die Erkenntnisse der Interviewstudie entegegen. Obwohl ähnlich wie in den übrigen Gruppen rund ein Fünftel der Probanden angaben, sich positiv oder neutral bei der Bearbeitung der Aufgabenstellung gefühlt zu haben, zeigt sich in der qualitativen Inhaltsanalyse, dass nahezu 70 Prozent der Versuchspersonen dieser Gruppe von sich aus angaben, dass sie sich gestresst oder überfordert gefühlt hatten. Dieser Wert variierte in den übrigen Gruppen nur zwischen knapp 40 und gut 46 Prozent.

Eine mögliche Erklärung hierfür könnte sein, dass die manuelle Steuerung des Tons von den Probanden als eine weitere Teilaufgabe wahrgenommen wurde neben den ohnehin anspruchsvollen Aufgaben, die vier Videos zu beobachten und sie gleichzeitig zu kommentieren, so dass zusätzlicher Stress bzw. das Gefühl der Überforderung entstand.

Angesichts fehlender Vorteile dieser Variante gegenüber der Grundvariante, sollte die Funktionalität „Ton variierbar“ in der getesteten Version nicht umgesetzt werden. In Anbetracht der Bedeutung des Tons für spätere Realübungen, um auch die Qualität der Kommunikation der Trainierenden einschätzen zu können, sollte demnach der Ton von allen Videos hörbar bleiben bzw. müssen andere, weniger anspruchsvolle Steuerungsmöglichkeiten gefunden werden, den Ton zu variieren.

6.2.5 Weitere Handlungsempfehlungen

Nachdem, ausgehend von den Ergebnissen der vorliegenden Studie, die Varianten „Expertenkommentare mit Signal“ und „Ton variierbar“ in den vorgestellten Versionen ausgeschlossen werden konnten, ergeben sich weitere Handlungsempfehlungen, die bei der Umsetzung der Projektkonzeption beachtet werden sollten.

Die Auswertung der erhobenen Daten weist darauf hin, dass die bearbeitete Aufgabe in allen getesteten Varianten gleichsam als schwierig eingeschätzt wurde. Die Mittelwerte aller Gruppen lagen hier bei ca. 4 von maximal 5 erreichbaren Punkten. Da die Motivation, eine Aufgabe zu Ende zu bringen, hiervon maßgeblich beeinflusst wird (Nerdinger, 2013; Orvis et al., 2008),

sollte vor Inbetriebnahme der KNoTen-Station darauf geachtet werden, dass die Experten im Umgang mit der einzusetzenden Technik ausreichend geschult sind. Neben der Handhabung der Technik sollte ein vorbereitendes Training die Vermittlung der Aufgabenstellung bzw. der Kriterien, die von den Experten zu beachten sind, und die Grundsätze wirkungsvollen Feedbacks beinhalten. Auf diese Weise könnten Unsicherheiten abgebaut und Kompetenzen bei den Experten zur Bewältigung des Feedbackauftrags aufgebaut werden. Dies könnte sich ebenfalls positiv auf das Gefühl der Überforderung auswirken. Da sich die Arbeit der KNoTen-Station aufgrund der zu bewältigenden Anforderungen (mehrere Videos, schneller Wechsel zwischen Beobachten und Schreiben) aber generell als sehr anspruchsvoll darstellt, sollten mit Hilfe weiterer Forschung Wege gefunden werden, um die beschriebenen negativen Auswirkungen auf das Wohlbefinden der Experten zu reduzieren. Ein Training könnte hier beispielsweise die Selbstwirksamkeitserwartung der Kommentatoren verbessern (Martins, Pundt, Breyer & Nerdinger, 2011) und so zu mehr Wohlbefinden während der Arbeit mit der KNoTen-Station beitragen (Halisch & Geppert, 2012).

Neben der Vorbereitung der Experten ist auch die Kompetenz der Trainingsleiter, die nach Beendigung der Realübung das Feedbackgespräch mit den Trainierten durchführen, von besonderer Bedeutung. Diese müssen in der Lage sein, die Kommentare der unterschiedlichen Experten mit Hilfe ihres umfangreichen, möglichst interdisziplinären Fachwissens zu verstehen und in das Feedbackgespräch zu integrieren. Hierbei müssen sie in der Lage sein, ad hoc die geschriebenen Kommentare aus unterschiedlichen Fachgebieten zu verstehen und gegebenenfalls mit eigenem Wissen zu ergänzen. Hierfür erscheint eine umfangreiche Vorbereitung der Trainingsleiter bspw. durch Schulungen erforderlich, wobei auch hier darauf zu achten ist, insbesondere die Grundregeln für effektives Feedback zu vermitteln.

6.3 Limitationen der Untersuchung

Der vorliegenden Untersuchung sind Grenzen gesetzt, von denen die wesentlichen hier kurz erläutert werden sollen. Einige dieser Limitationen geben Hinweise für zukünftige Weiterentwicklungen und weitere Untersuchungen der KNoTen-Station.

Die Studie wurde innerhalb eines kurzen Zeitraums mit insgesamt 129 Personen durch mehrere Versuchsleiter durchgeführt. Aus praktischen Gründen war es nicht möglich, die originale KNoTen-Station für diese Untersuchung zu verwenden, weshalb eine Software genutzt wurde, die die für die Studie im Vordergrund stehenden Funktionalitäten, die parallele Anzeige der Videos und die Kommentarfunktion, nachbildete. Darüber hinaus konnten mit Hilfe der Software kostengünstig die vorgeschlagenen zusätzlichen Funktionalitäten erprobt werden. Schwie-

rig war hierbei jedoch, immer exakt gleichen Versuchsbedingungen herzustellen, wie es für ein Experimentaldesign gefordert wird (Bortz & Döring, 2006; von Rosenstiel & Nerdinger, 2011), da die Versuchsleiter nicht über identische Computer mit gleicher Ausstattung wie Display und Lautsprecher verfügten, was z.B. im Szenario „Expertenkommentare mit Signal“ eine Rolle spielen könnte. Es ist jedoch davon auszugehen, dass auch unter realen Bedingungen die Experten, die die KNoTen-Station nutzen werden, über jeweils unterschiedliche technische Ausstattung verfügen werden.

Bei der vorliegenden Studie kann nicht mit Sicherheit davon ausgegangen werden, dass die Stichprobe in ihrer Gesamtheit repräsentativ ist für die eigentliche Zielgruppe, Experten aus Bereichen wie Nautik oder Psychologie. Es ist zu vermuten, dass die Zielgruppe tatsächlich homogener in ihrer Zusammensetzung ist. So ist aufgrund des Expertenstatus dieses Personenkreises beispielweise eher mit einem mittleren oder höheren Alter zu rechnen, damit die per Definition (Bromme, Jucks & Rambow, 2004, s. Abschnitt 3.3.1) geforderte Erfahrung gegeben ist. Eine Realisierung einer Studie dieses Umfangs mit Experten wäre allerdings aus praktischen Gründen kaum möglich gewesen.

Die untersuchte Stichprobe wurde für eine Aufgabenstellung gewählt, die für alle Probanden als lösbar erschien. Es ist jedoch damit zu rechnen, dass nur sehr wenige Probanden als Experten für Verhalten im Straßenverkehr gelten dürften, während die KNoTen-Station in Zukunft von Personen genutzt werden soll, die eben genau diesen Expertenstatus in Bezug auf die gestellte Aufgabe besitzen. Darüber hinaus wird den Experten vor dem eigentlichen Training ein Skript bereitgestellt, in dem der Ablauf des Trainings dargestellt ist. Folglich ist den Experten bereits vorher bekannt, was zu welchem ungefähren Zeitpunkt geschehen wird. Diesen Vorteil hatten die Probanden in der vorliegenden Studie wiederum nicht. Damit dürften die Experten bei der Beurteilung des beobachteten Verhaltens auch weniger unsicher sein, als es die Probanden dieser Studie waren. Auch die Dauer der Videos wird in der zukünftigen Anwendung sicherlich länger sein, da sie sich nach der Dauer eines Trainings richten muss. Allerdings zeigen die Aussagen der Probanden, dass es sich bereits bei achtminütigen Videos um eine sehr herausfordernde Aufgabe handelt.

Trotzdem liefert die Studie wichtige Hinweise, welche zusätzlichen Funktionalitäten für die KNoTen-Station mit größerer Wahrscheinlichkeit sinnvoll sind und welche Funktionalitäten keinen Mehrwert bringen, da psychologische Effekte wie z.B. das Social Loafing grundsätzliche Gültigkeit haben. Weitere Studien sollten zeigen, ob die gefundene positive Wirkung von Funktionalitäten, wie dem Pausieren, sich auch bei der Nutzung durch Experten und bei realen

Trainingsvideos bestätigen lässt. Insbesondere hier muss geprüft werden, ob die mit dieser Variante einhergehenden Kosten, in Form des erhöhten Zeitaufwandes und der Zeitverzögerung des Feedbacks, durch den gewonnenen Nutzen, z.B. in Form von mehr Ausführlichkeit und Richtigkeit, aufgewogen werden können.

Auch wenn Funktionalitäten, wie das manuelle Ein- und Ausschalten des Tons, aus den Ergebnissen als weniger nützlich hervorgehen, sollte trotzdem in Erwägung gezogen werden, ob eine solche Funktionalität in abgewandelter Form dennoch sinnvoll sein könnte. Das wiederholte Anwählen des Videos, dessen Ton angestellt werden sollte, wurde von den Probanden möglicherweise als zusätzliche Belastung wahrgenommen. Dabei konnte immer nur ein Video ausgewählt werden. Sinnvoller könnte es jedoch sein, wenn der Ton beliebig vieler Videos an- und ausgestellt werden könnte. Ist der Ton eines Videos für den Experten beispielsweise unbrauchbar aufgrund von alles übertönenden Alarm- oder Windgeräuschen, so könnte das Ausschalten des Tons hier die Aufgabe erleichtern.

Aufgrund des experimentellen Designs, in dem differenziert die Wirkung jeder einzelnen Funktionalität auf abhängige Variablen wie der Anzahl oder die Richtigkeit der Kommentare untersucht werden konnte, bleibt immer noch die Frage wie die Funktionalitäten wirken, wenn sie miteinander kombiniert werden, z.B. die Möglichkeit des Pausierens und die Anzeige von Kommentaren anderer Experten. Weitere Tests der Usability müssten hierzu durchgeführt werden.

Zuletzt ist zu erwähnen, dass die Probanden im Interview möglicherweise den Eindruck hatten, dass sie die eben genutzte Software ganz besonders im Hinblick auf Unzulänglichkeiten kritisieren sollen, weshalb eventuell eher negative Gefühle geäußert wurden. Allerdings lieferten die Verbesserungsvorschläge konstruktive Anregungen für die Weiterentwicklung der KNoTen-Station, wie beispielsweise die Ermöglichung von Audiokommentaren oder die Möglichkeit des Taggens; d.h. der Nutzer klickt auf die kritische Stelle, z.B. die Person, im Video und diese wird dann mit dem Kommentar in Verbindung gesetzt.

Literaturverzeichnis

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes,* 50, 179-211.

Asendorpf, J. (1988). Videoanalyse von Verhaltensprozessen: Neue Möglichkeiten durch Video-Mikrocomputer-Koppelung. *Psychologische Rundschau*, *39*, 1-12.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In: K. W. Spence & J. Taylor Spence: *The psychology of learning and motivation: Advances in research and theory*, *Volume 2*, 89-195. New York: Academic Press Inc.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. & Morgan M. T. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research*. 61, 213-238.

Bell, R. (2006). Die Beeinträchtigung von Arbeitsgedächtnisleistungen durch auditive Distraktoren im Altersvergleich. URL: docserv.uni-duesseldorf.de/servlets/Derivate Servlet /Derivate-3630/1630.pdf. [07.04.2013].

Berti, S., & Schröger, E. (2003). Die Bedeutung sensorischer Verarbeitung und Aufmerksamkeitssteuerung für Arbeitsgedächtnisfunktionen. *Zeitschrift für Psychologie / Journal of Psychology*, *211*, 193-201.

Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin; Psychological Bulletin*, *94*, 265-292.

Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. 6. Auflage. Heidelberg: Springer Medizin Verlag.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation: für Human-und Sozialwissenschaftler*. 4. Auflage. Berlin / Heidelberg: Springer.

Bortz, J. & Lienert, G. A. (2008). *Kurzgefasste Statistik für die klinische Forschung*. Berlin: Springer.

Bromme, R., Jucks, R., & Rambow, R. (2004). Experten-Laien-Kommunikation im Wissensmanagement. *Der Mensch im Wissensmanagement: Psychologische Konzepte zum besseren Verständnis und Umgang mit Wissen*, 176-188.

Bronner, R. (1974). Steß und Leistung. Ergebnisse experimenteller Leistungsforschung. *Zeitschrift für Organisation – Neue Betriebswirtschaft* 43, 363-368.

Brunner, R., & Zeltner, W. (1980). *Lexikon zur pädagogischen Psychologie und Schulpädagogik: Entwicklungspsychologie, Lehr-und Lernpsychologie, Unterrichtspsychologie, Erzie-*

hungspsychologie, Methoden der pädagogischen Psychologie, Methodik, Didaktik, Curriculumtheorie. München: Ernst Reinhardt.

Bryan, J. F., & Locke, E. A. (1967). Parkinson's law as a goal-setting phenomenon. *Organizational Behavior and Human Performance*, *2*, 258-275.

BSU (2012). *Jahresstatistik 2011*. Verfügbar unter URL: http://www.bsu-bund.de/SharedDocs/pdf/DE/Jahresstatistik/Jahresstatistik_2011.pdf?__blob=publicationFile. [11.02.2013].

Bühner, M. & Ziegler, M. (2009) *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.

Christmann, U. (1999). Wahrnehmung, Kognition und Metakognition. Kölner Psychologische Studien–Beiträge zur natur-, kultur-, sozialwissenschaftlichen Psychologie, 4, 6-12.

Christopherson, K. M. (2007). The positive and negative implications of anonymity in Internet social interactions:“On the Internet, nobody knows you’re a dog”. *Computers in Human Behavior*, *23*, 3038-3056.

Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: the impact of distraction. *Psychology and aging*, *6*, 533-541.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.

Durham, C. C., Locke, E. A., Poon, J. M., & McLeod, P. L. (2000). Effects of group goals and time pressure on group efficacy, information-seeking strategy, and performance. *Human Performance*, *13*, 115-138.

Emad, G., & Roth, W. M. (2008). Contradictions in the practices of training for and assessment of competency: A case study from the maritime domain. *Education+ Training*, *50*, 260-272.

Fengler, J. (1975). *Selbstkontrolle in Gruppen. Theorie, Praxis, Evaluation*. Stuttgart: Kohlhammer.

Fengler, J. (2010). Feedback als Interventionsmethode. *Gruppendynamik und Organisationsberatung*. 4, 5-20.

Frey, B., Savage, D., Schmidt, S. & Torgler, B. (2011). Implications of power on survival in tragic events: a comparison of the Titanic and Lusitania sinking [Auswirkungen von Macht auf das Uberleben in Extremsituationen: Ein Vergleich der Titanic und Lusitania Schiffskatastrophen]. *Koelner Zeitschrift fuer Soziologie und Sozialpsychologie*, *63*, 237-254.

Früh, W. (2007). *Inhaltsanalyse: Theorie und Praxis*. 6. Auflage. Konstanz: UVK.

Galassi, J. P., Galassi, M. D., & Litz, M. C. (1974). Assertive training in groups using video feedback. *Journal of Counseling Psychology*, *21*, 390-394.

Gerstenberger, H. (2002). Cost elements with a soul. In *Proceedings of International Association of Maritime Economists-International Conference* (pp. 13-15).

Gilliland, S. W., & Schmitt, N. (1993). Information redundancy and decision behavior: A process tracing investigation. *Organizational Behavior and Human Decision Processes*.54, 157-180.

Gajewski, P. (2005). *Enkodierungsprozesse beim Aufgabenwechsel.* Verfügbar unter URL: http://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-3216/1216.pdf [15.04.2013].

Greif, R., & Breckwoldt, J. (2012). Warum lebenslanges Lernen ohne effektives Feedback nicht wirkungsvoll ist. *Notfall+ Rettungsmedizin*, 15, 193-197.

Greif, S. (1991). Stress in der Arbeit – Einführung und Grundbegriffe. In: S. Greif, E. Bamberg & N. Semmer (Hrsg.). *Psychischer Streß am Arbeitsplatz* (S. 1-28). Göttingen / Toronto / Zürich: Hogrefe.

Groff, B. D., Baron, R. S., & Moore, D. L. (1983). Distraction, attentional conflict, and drivelike behavior. *Journal of Experimental Social Psychology*, *19*, 359-380.

Hagendorf, H., Krummenacher, J., Müller, H. J., & Schubert, T. (2011). Wahrnehmung und Aufmerksamkeit: Gemeinsam zum Ziel. In: H. Hagendorf, J. Krummenacher, H. J. Müller & T. Schubert (Hrsg.) *Wahrnehmung und Aufmerksamkeit (*S. 1-10). Berlin / Heidelberg: Springer.

Hahne, J., Baaske, G., Sedlaček, D., & Schubert, J. F. (2002). *Risikomanagement in Notfallsituationen an Bord von Seeschiffen*. Dortmund / Berlin / Dresden: Wirtschaftsverlag NW, Verlag für Neue Wissenschaft.

Hahne, J., Tober, H., & Brühe, B. (1997). *Rettung aus Seenot*. Bornheim / Bonn: Deutscher Kommunal-Verlag Naujoks und Behrendt.

Halisch, F., & Geppert, U. (2012). Wohlbefinden im Alter: Der Einfluss von Selbstwirksamkeit, Kontrollüberzeugungen, Bewältigungsstilen und persönlichen Zielen. Ergebnisse aus der Münchner GOLD-Studie. Verfügbar unter URL: http://psydok.sulb.uni-saarland.de/volltexte/2012/3567/. [16.04.2013].

Hasebrook, J. (1995). *Multimedia-Psychologie: Eine neue Perspektive menschlicher Kommunikation*. Heidelberg / Berlin / Oxford: Spektrum, Akademischer Verlag.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*. 77, 81-112.

Heuer, H. (2010). Wahrnehmung und Arbeitshandeln. In: U. Kleinbeck & K. H. Schmidt (Hrsg.) *Arbeitspsychologie* (Enzyklopädie der Psychologie, D/III/1) (S. 177-214). Göttingen / Bern / Toronto / Seattle: Hogrefe.

Hintz, A. J. (2011). *Erfolgreiche Mitarbeiterführung durch soziale Kompetenz*. Wiesbaden: Gabler.

Ilgen, D. R. Fischer, C. D. & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64, 349-371.

Imhof, M. (2003). *Zuhören: Psychologische Aspekte auditiver Informationsverarbeitung*. Band 4. Göttingen: Vandenhoeck &Ruprecht.

ISV (2013). *Homepage des Instituts für Schiffssicherheit und Sicherheitstechnik e.V.* Verfügbar unter URL: http://www.schiffssicherheit.de/info.php. [07.04.2013].

Jonas, K., & Tanner, C. (2006). Effekte sozialer Förderung und Hemmung. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik* (pp.167–186). Heidelberg: Springer-Verlag.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of personality and social psychology*, *65*, 681.

Kebeck, G. (1997). *Wahrnehmung. Theorien, Methoden und Forschungsergebnisse der Wahrnehmungspsychologie,* 2. Auflage*,* Weinheim / München: Juventa.

Kiefer, J. (2002). *Auswirkungen von Ablenkung durch gehörte Sprache und eigene Handlungen auf die Sprachproduktion.* Verfügbar unter: http://www.prometei.de/fileadmin/prometei.de /publikationen/Kiefer_thesis2002.pdf [15.04.2013].

Kjellberg, A., Landström, U., Tesarz, M., Söderberg, L. & Akerlund, E. (1996). The effects of nonphysical noise characteristics, ongoing task and noise sensitivity on annoyance and distraction due to noise at work. *Journal of Environmental Psychology*, *16*, 123-136.

Klebl, U. (2006). *Effekte von Feedback-Interventionen in Development-Centern*. München und Mering: Rainer Hampp.

Lange, E. B. (2005). *Der Fokuseffekt. Über die Ablenkung der Aufmerksamkeit durch irrelevante Reize.* Verfügbar unter: http://opus.kobv.de/tuberlin/volltexte/2005/1080 [15.04.2013].

Langer, L. F. Schulz von Thun en R. Tausch (1981). *Sich verständlich ausdrücken*. München: Ernst Reinhardt Verlag.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology*, *37*, 822-832.

Latham, G. P. & Locke, E. A. (1991) Self Regulation through Goal Setting. *Organizational Behavior and Human Decision Processes.* 50, 212-247.

London, M. (2003) *Job feedback: giving, seeking, and using feedback for performance improvement* (2nd edition). Mahwah, NJ: Erlbaum.

Martins, E., Pundt, A., Breyer, T. & Nerdinger, F. W. (2011). Betriebsräte als Förderer betrieblicher Innovationsprozesse. In: B. Schallock & H. Jacobsen (Hrsg.). *Innovationsstrategien jenseits traditionellen Managements. Wissenschaftliche und praktische Ergebnisse des Förderschwerpunktes* (S. 289-297). Stuttgart: Fraunhofer Verlag.

Matthews, G., Davies, D. R., Westerman, S. J. & Stammers, R. B. (2000). *Human performance: Cognition, stress, and individual differences*. Hove, East Sussex: Psychology Press.

Mayring, P. (2008). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 10. Auflage. Weinheim: Beltz.

Mittag, H. J. (2011). *Statistik*. Berlin / Heidelberg: Springer.

Möhring, W. & Schlütz, D. (2010). *Die Befragung in der Medien-und Kommunikationswissenschaft: eine praxisorientierte Einführung*. 2., überarbeitete Auflage. Wiebaden: VS Verlag für Sozialwissenschaften.

Musch, J. (1999). Die Gestaltung von Feedback in computergestützten Lernumgebungen: Modelle und Befunde. *Zeitschrift für Pädagogische Psychologie,* 13, 148-160.

Nerdinger, F. W. (2013). *Arbeitsmotivation und Arbeitshandeln*. Kröning: Asanger.

Nerdinger, F. W. (2012). *Grundlagen des Verhaltens in Organisationen*. 3. Auflage. Stuttgart: W. Kohlhammer Verlag.

Nestler, S., Back, M. D., & Egloff, B. (2011). Psychometrische Eigenschaften zweier Skalen zur Erfassung interindividueller Unterschiede in der Präferenz zum Alleinsein. *Diagnostica*, *57*, 57-67.

Nivison, M. E., & Endresen, I. M. (1993). An analysis of relationships among environmental noise, annoyance and sensitivity to noise, and the consequences for health and sleep. *Journal of behavioral medicine*, *16*, 257-276.

Oldenburg, M., & Jensen, H. J. (2012). Merchant seafaring: a changing and hazardous occupation. *Occupational and Environmental Medicine*, *69*, 685-688.

Oldenburg, M., Jensen, H. J., Latza, U., & Baur, X. (2009). Seafaring stressors aboard merchant and passenger ships. *International journal of public health*, *54*, 96-105.

Orvis, K. A., Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior video-game experience on performance and motivation in instructional videogames. *Computers in Human behavior*, *24*, 2415-2433.

Pierre, M. S., Hofinger, G., & Buerschaper, D. P. C. (2011). *Notfallmanagement*. Berlin / Heidelberg: Springer.

Pohl, R., & Haracic, I. (2005). Der Rückschaufehler bei Kindern und Erwachsenen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *37*, 46-55.

Posner, M. I. (1994). Attention: The mechanisms of conciousness. *Proceedings of the National Academy of Science, USA*, 91, 7398-7403.

Raab, G., Unger, A., & Unger, F. (2010). *Marktpsychologie*. Grundlagen und Anwendung. 3. Auflage. Wiesbden:Gabler.

Rafaeli, S., & Noy, A. (2002). Online auctions, messaging, communication and social facilitation: a simulation and experimental evidence. *European Journal of Information Systems*, *11*, 196-207.

Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2010). Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. 3., erweitere Auflage. Wiesbaden: Springer.

Römer, R. (2009). Robuste Spracherkennung auf der Basis recheneffizienter auditiver Modelle. Verfügbar unter URL: http://www.mmk.ei.tum.de/publ/pdf/roemer.pdf [15.04.2013].

Rosenstiel, L. v. & Nerdinger, F. W. (2011). Grundlagen der Organisationspsychologie : Basiswissen und Anwendungshinweise. 7. überarbeitete Auflage. Stuttgart: Schäffer-Poeschel Verlag.

Roter, D. L., Larson, S., Shinitzky, H., Chernoff, R., Serwint, J. R., Adamo, G., & Wissow, L. (2004). Use of an innovative video feedback technique to enhance communication skills training. *Medical Education*, *38*, 145-157.

Sampson, H. (2004). Romantic rhetoric, revisionist reality: the effectiveness of regulation in maritime education and training. *Journal of Vocational Education and Training*, *56*, 245-267.

Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta psychologica*, *53*, 61-97.

Schneider, B. (2011). *Die Simulation menschlichen Panikverhaltens*. Wiesbaden: Vieweg+Teubner.

Tomasi, D., Caparelli, E. C., Chang, L., & Ernst, T. (2005). fMRI-acoustic noise alters brain activation during working memory tasks. *Neuroimage*, *27*, 377-386.

Tonidandel, S., Quiñones, M. A. & Adams, A. A. (2002). Computer-Adaptive Testing: The Impact of Test Characteristics on Perceived Performance and Test Takers' Reactions. *Journal of Applied Psychology*, 87, 320-332.

Tversky, A. & Kahneman, D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5, 207-232.

Wahlser, W., Blocker, A., Blocker, J., Stopp, E. & Speiser, H. (1998). Ressourcenadaptierte Objektlokalisation: Sprachliche Raumbeschreibung unter Zeitdruck. *Kognitionswissenschaft*, 7, 111-117.

Williams, K., Harkins, S. G., & Latané, B. (1981). Identifiability as a deterrant to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology*, *40*, 303-311.

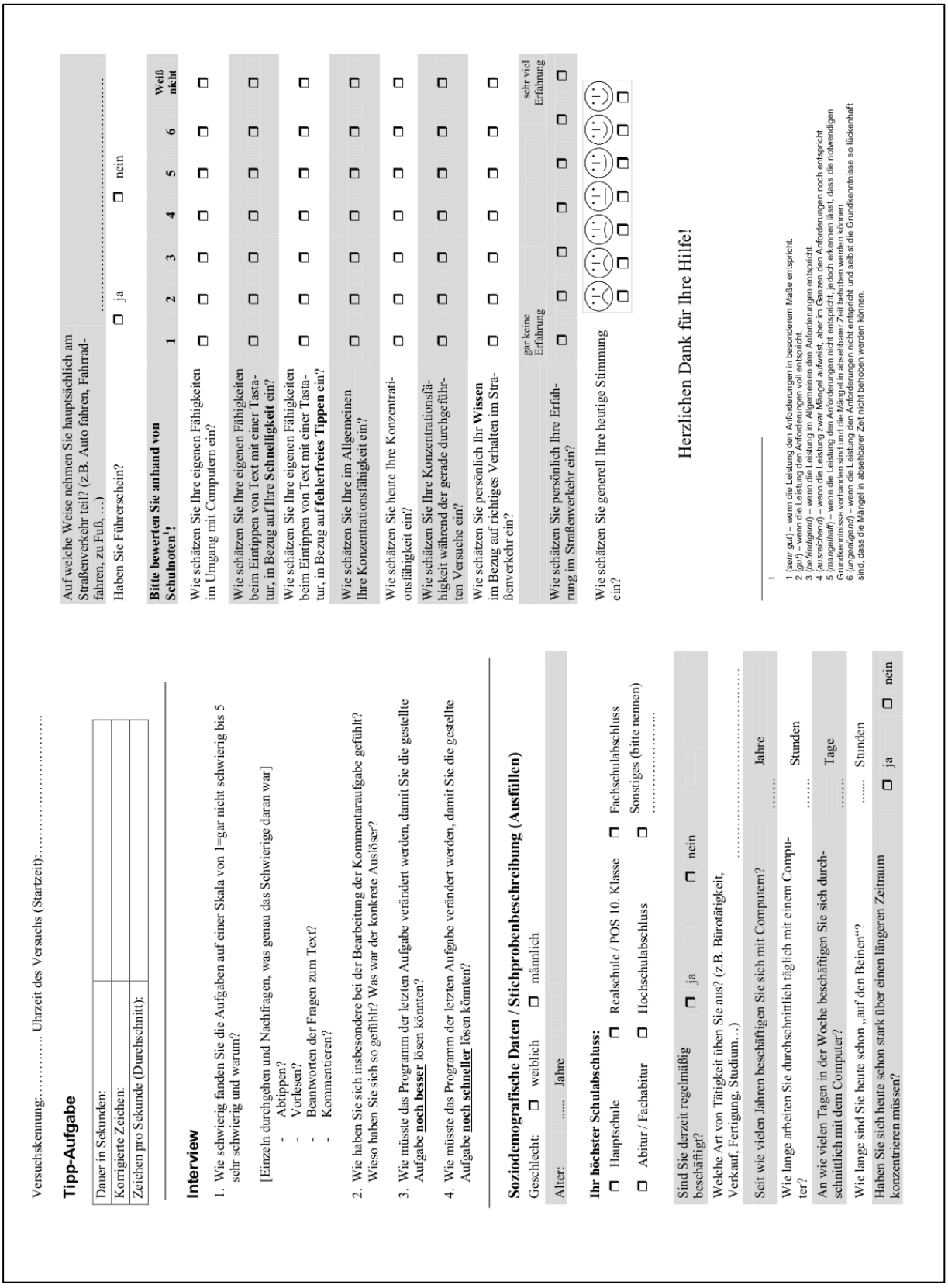
Zajonc, R. B. (1965). Social facilitation. *Science*, 149, 269-274.

Zajonc, R. B., & Sales, S. M. (1966). Social facilitation of dominant and subordinate responses. *Journal of Experimental Social Psychology*, *2*, 160-168.

Zimbardo, P. G. & Gerrig, R. J. (2004). *Psychologie.* 16., aktualisierte Auflage. München: Pearson.

Anhang

Anhang 1: Interviewleitfaden und Fragebogen



Versuchskennung:..... Uhrzeit des Versuchs (Startzeit):.....

Tipp-Aufgabe

Dauer in Sekunden:	
Korrigierte Zeichen:	
Zeichen pro Sekunde (Durchschnitt):	

Interview

1. Wie schwierig fanden Sie die Aufgaben auf einer Skala von 1=gar nicht schwierig bis 5 sehr schwierig und warum?

   [Einzeln durchgehen und Nachfragen, was genau das Schwierige daran war]
   - Abtippen?
   - Vorlesen?
   - Beantworten der Fragen zum Text?
   - Kommentieren?

2. Wie haben Sie sich insbesondere bei der Bearbeitung der Kommentaraufgabe gefühlt? Wieso haben Sie sich so gefühlt? Was war der konkrete Auslöser?

3. Wie müsste das Programm der letzten Aufgabe verändert werden, damit Sie die gestellte Aufgabe **noch besser** lösen könnten?

4. Wie müsste das Programm der letzten Aufgabe verändert werden, damit Sie die gestellte Aufgabe **noch schneller** lösen könnten?

Soziodemografische Daten / Stichprobenbeschreibung (Ausfüllen)

Geschlecht: ☐ weiblich ☐ männlich

Alter: ..... Jahre

Ihr höchster Schulabschluss:

☐ Hauptschule ☐ Realschule / POS 10. Klasse ☐ Fachschulabschluss

☐ Abitur / Fachabitur ☐ Hochschulabschluss ☐ Sonstiges (bitte nennen) .....

Sind Sie derzeit regelmäßig beschäftigt? ☐ ja ☐ nein

Welche Art von Tätigkeit üben Sie aus? (z.B. Bürotätigkeit, Verkauf, Fertigung, Studium…) .....

Seit wie vielen Jahren beschäftigen Sie sich mit Computern? ..... Jahre

Wie lange arbeiten Sie durchschnittlich täglich mit einem Computer? ..... Stunden

An wie vielen Tagen in der Woche beschäftigen Sie sich durchschnittlich mit dem Computer? ..... Tage

Wie lange sind Sie heute schon „auf den Beinen"? ..... Stunden

Haben Sie sich heute schon stark über einen längeren Zeitraum konzentrieren müssen? ☐ ja ☐ nein

Auf welche Weise nehmen Sie hauptsächlich am Straßenverkehr teil? (z.B. Auto fahren, Fahrradfahren, zu Fuß, …) .....

Haben Sie Führerschein? ☐ ja ☐ nein

Bitte bewerten Sie anhand von Schulnoten[1]:	**1**	**2**	**3**	**4**	**5**	**6**	**Weiß nicht**
Wie schätzen Sie Ihre eigenen Fähigkeiten im Umgang mit Computern ein?	☐	☐	☐	☐	☐	☐	☐
Wie schätzen Sie Ihre eigenen Fähigkeiten beim Eintippen von Text mit einer Tastatur, in Bezug auf Ihre **Schnelligkeit** ein?	☐	☐	☐	☐	☐	☐	☐
Wie schätzen Sie Ihre eigenen Fähigkeiten beim Eintippen von Text mit einer Tastatur, in Bezug auf **fehlerfreies Tippen** ein?	☐	☐	☐	☐	☐	☐	☐
Wie schätzen Sie Ihre im Allgemeinen Ihre Konzentrationsfähigkeit ein?	☐	☐	☐	☐	☐	☐	☐
Wie schätzen Sie heute Ihre Konzentrationsfähigkeit ein?	☐	☐	☐	☐	☐	☐	☐
Wie schätzen Sie Ihre Konzentrationsfähigkeit während der gerade durchgeführten Versuche ein?	☐	☐	☐	☐	☐	☐	☐
Wie schätzen Sie persönlich Ihr **Wissen** im Bezug auf richtiges Verhalten im Straßenverkehr ein?	☐	☐	☐	☐	☐	☐	☐

	gar keine Erfahrung						sehr viel Erfahrung
Wie schätzen Sie persönlich Ihre Erfahrung im Straßenverkehr ein?	☐	☐	☐	☐	☐	☐	☐

Wie schätzen Sie generell Ihre heutige Stimmung ein?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Herzlichen Dank für Ihre Hilfe!

[1]

1 (*sehr gut*) – wenn die Leistung den Anforderungen in besonderem Maße entspricht.
2 (*gut*) – wenn die Leistung den Anforderungen voll entspricht.
3 (*befriedigend*) – wenn die Leistung im Allgemeinen den Anforderungen entspricht.
4 (*ausreichend*) – wenn die Leistung zwar Mängel aufweist, aber im Ganzen den Anforderungen noch entspricht.
5 (*mangelhaft*) – wenn die Leistung den Anforderungen nicht entspricht, jedoch erkennen lässt, dass die notwendigen Grundkenntnisse vorhanden sind und die Mängel in absehbarer Zeit behoben werden können.
6 (*ungenügend*) – wenn die Leistung den Anforderungen nicht entspricht und selbst die Grundkenntnisse so lückenhaft sind, dass die Mängel in absehbarer Zeit nicht behoben werden können.

Anhang 2: Signifikanzwerte des Shapiro-Wilk-Tests auf Normalverteilung der Variablen zu Erfahrungen mit PC und Straßenverkehr

Variablen	Grundvariante	Experten-kommentare	Experten-kommentare mit Signal	Pausieren	Ton variierbar
Tippgeschwindigkeit	,144	,516	,047	,507	,068
Dauer des Vorlesensim Konzentrationstest	,499	,000	,273	,002	,315
Richtige Antworten im Konzentrationstest	,026	,005	,046	,000	,023
Jahre der PC-Erfahrung	,000	,390	,016	,990	,095
Computertage pro Woche	,000	,000	,001	,000	,000
Subjektive Einschätzung der PC-Fähigkeiten	,003	,005	,003	,038	,004
subjektives Wissen Straßenverkehr	,007	,000	,218	,022	,002
Erfahrung im Straßenverkehr	,013	,004	,004	,000	,003

Anhang 3: Signifikanztests zu Unterschieden hinsichtlich Erfahrungen mit PCs und Straßenverkehr

Gruppenvergleiche der Grundvariante mit …		U-Test			Levene-Test auf Varianzhomogenität		t-Test		
		Mann-Whitney-U	Z	Signifikanz	F-Wert	Signifikanz	T-Wert	df	Signifikanz
Tippgeschwindigkeit	Expertenkommentare				0,593	0,445	0,458	50,0	0,649
	Expertenkommentare mit Signal	243,0	-0,462	0,644					
	Pausieren				1,509	0,225	0,347	49,0	0,730
	Ton variierbar				1,264	0,267	0,006	47,0	0,995
Dauer des Vorlesens	Expertenkommentare	311,0	-0,918	0,359					
	Expertenkommentare mit Signal				1,589	0,214	0,705	47,0	0,484
	Pausieren	288,0	-1,122	0,262					
	Ton variierbar				0,065	0,800	0,694	49,0	0,491
Richtige Antworten	Expertenkommentare	282,0	-1,469	0,142					
	Expertenkommentare mit Signal	271,5	-0,572	0,567					
	Pausieren	336,0	-0,283	0,777					
	Ton variierbar	267,5	-1,128	0,259					
Jahre der PC-Erfahrung	Expertenkommentare	282,5	-1,210	0,226					
	Expertenkommentare mit Signal	253,5	-0,706	0,480					
	Pausieren	310,0	-0,507	0,612					
	Ton variierbar	184,0	-2,338	**0,019**					

Anhang 3 Fortsetzung

Gruppenvergleiche der Grundvariante mit …		U-Test			Levene-Test auf Varianzhomogenität		t-Test		
		Mann-Whitney-U	Z	Signifikanz	F-Wert	Signifikanz	T-Wert	df	Signifikanz
Computertage pro Woche	Expertenkommentare	318,5	-0,824	0,410					
	Expertenkommentare mit Signal	284,0	-0,316	0,752					
	Pausieren	328,5	-0,427	0,669					
	Ton variierbar	299,5	-0,255	0,798					
Subjektive Einschätzung PC-Fähigkeiten	Expertenkommentare	324,0	-0,731	0,465					
	Expertenkommentare mit Signal	200,5	-1,942	0,052					
	Pausieren	314,0	-0,697	0,486					
	Ton variierbar	280,0	-0,672	0,502					
subjektives Wissen Straßenverkehr	Expertenkommentare	270,5	-1,566	0,117					
	Expertenkommentare mit Signal	265,0	-0,468	0,640					
	Pausieren	303,5	-0,672	0,501					
	Ton variierbar	233,0	-1,642	0,101					
Erfahrung im Straßenverkehr	Expertenkommentare	154,0	-0,637	0,524					
	Expertenkommentare mit Signal	99,5	-0,255	0,799					
	Pausieren	157,0	-0,132	0,895					
	Ton variierbar	140,5	-0,667	0,505					

Anhang 4: Deskriptive Statistiken der abhängigen Variablen je Versuchsgruppe

		Grundvariante	Experten-kommentare	Experten-kommentare mit Signal	Pausieren	Ton variier-bar
Anzahl der Kommentare	Mittelwert	8,80	10,22	11,74	12,50	9,42
	Median	8,00	9,00	11,00	10,50	8,00
	Standardabweichung	5,71	5,37	6,83	8,82	5,96
	Minimum	2	3	3	2	1
	Maximum	29	25	33	39	22
Anzahl einbezogener Videos	Mittelwert	3,36	3,41	3,78	3,42	3,42
	Median	4,00	3,00	4,00	4,00	4,00
	Standardabweichung	0,91	0,57	1,09	0,81	0,83
	Minimum	1	2	2	2	1
	Maximum	4	4	8	4	4
Ausführlichkeit (Anzahl Zeichen / Kommentar)	Mittelwert	31,25	30,13	28,65	43,10	30,37
	Median	30,50	27,25	29,43	44,16	28,92
	Standardabweichung	8,21	10,77	7,95	11,59	9,93
	Minimum	15,50	14,40	13,00	16,50	14,27
	Maximum	46,83	59,86	46,71	61,64	54,63
Tippgeschwindigkeit (Zeichen / Sekunde)	Mittelwert	2,80	2,64	3,13	2,12	2,57
	Median	2,72	2,53	3,15	2,06	2,48
	Standardabweichung	1,01	0,90	1,00	0,88	0,90
	Minimum	0,64	0,73	1,81	0,90	0,35
	Maximum	4,81	4,30	5,56	5,79	4,46
Versuchsdauer (Sek.)	Mittelwert	487,26	485,68	483,21	811,31	484,89
	Median	482,46	481,91	481,68	764,52	481,32
	Standardabweichung	8,56	12,76	3,98	224,70	7,77
	Minimum	480,43	480,51	480,46	501,16	480,46
	Maximum	510,03	547,24	497,00	1372,00	513,43
Anzahl richtig erkannter Verstöße	Mittelwert	5,16	6,04	5,91	7,08	5,42
	Median	5,00	6,00	5,00	5,50	4,50
	Standardabweichung	3,09	3,09	2,89	6,49	3,19
	Minimum	1	1	1	1	0
	Maximum	14	13	12	31	13
Verständlichkeit	Mittelwert	6,28	4,33	4,88	5,91	5,43
	Median	4,45	3,91	3,88	4,88	4,85
	Standardabweichung	5,38	2,96	4,29	4,61	3,60
	Minimum	1,17	0,29	0,30	1,18	0,50
	Maximum	24,00	12,00	19,33	21,50	15,75

Anhang 5: Tests auf Normalverteilung für die abhängigen Variablen

		Shapiro-Wilk		
		Statistik	df	Signifikanz
Anzahl der Kommentare	Grundvariante	,834	25	**,001**
	Expertenkommentare	,926	27	,055
	Expertenkommentare mit Signal	,906	26	**,021**
	Pausieren	,883	24	**,010**
	Ton variierbar	,849	23	**,003**
Anzahl einbezogener Videos	Grundvariante	,706	25	**,000**
	Expertenkommentare	,726	27	**,000**
	Expertenkommentare mit Signal	,687	26	**,000**
	Pausieren	,721	24	**,000**
	Ton variierbar	,631	23	**,000**
Ausführlichkeit (Anzahl Zeichen / Kommentar)	Grundvariante	,974	25	,749
	Expertenkommentare	,916	27	**,032**
	Expertenkommentare mit Signal	,965	26	,490
	Pausieren	,916	24	**,048**
	Ton variierbar	,976	23	,835
Tippgeschwindigkeit (Zeichen / Sekunde)	Grundvariante	,984	25	,955
	Expertenkommentare	,976	27	,751
	Expertenkommentare mit Signal	,714	26	**,000**
	Pausieren	,981	24	,907
	Ton variierbar	,937	23	,152
Versuchsdauer (Sek.)	Grundvariante	,769	25	**,000**
	Expertenkommentare	,383	27	**,000**
	Expertenkommentare mit Signal	,925	26	,060
	Pausieren	,613	24	**,000**
	Ton variierbar	,660	23	**,000**
Anzahl richtig erkannter Verstöße	Grundvariante	,899	25	**,017**
	Expertenkommentare	,944	27	,156
	Expertenkommentare mit Signal	,733	26	**,000**
	Pausieren	,880	24	**,008**
	Ton variierbar	,953	23	,334
Verständlichkeit	Grundvariante	,798	25	**,000**
	Expertenkommentare	,944	27	,151
	Expertenkommentare mit Signal	,730	26	**,000**
	Pausieren	,919	24	,055
	Ton variierbar	,763	23	**,000**

Anhang 6: Ergebnisse der Signifikanztests der abhängigen Variablen.

AV	Gruppenvergleich der Grundvariante mit ...	U-Test			Levene-Test			T-Test						N der Grund-variante
		N	Signifikanz	Z	Mann-Whitney-U	Signifikanz	F-Wert	Mittlere Differenz	M2	M1	Signifikanz	df	T	
Anzahl der Kommen-tare	Expertenkommentare	28	0,142	-1,467	279,5									26
	Expertenkommentare mit Signal	23	**0,038**	**-2,071**	**196,0**									
	Pausieren	27	0,151	-1,436	270,5									
	Ton variierbar	25	0,664	-0,435	302,0									
Anzahl der Videos	Expertenkommentare	28	0,977	-0,029	362,5									26
	Expertenkommentare mit Signal	23	0,235	-1,187	246,5									
	Pausieren	27	0,781	-0,278	337,0									
	Ton variierbar	25	0,604	-0,519	300,5									
Ausführlich-keit	Expertenkommentare	26	0,622	-0,493	335,5									26
	Expertenkommentare mit Signal	23				0,628	0,237	2,32	28,65	30,97	0,32	47	1,005	
	Pausieren	27	**0,000**	**-3,781**	**138,5**									
	Ton variierbar	25	0,651	-0,452	301,0									
Tippgeschw in-digkeit	Expertenkommentare	28				0,618	0,252	0,12	2,68	2,8	0,646	51	0,462	25
	Expertenkommentare mit Signal	23	0,348	-0,939	242,0									
	Pausieren	27				**0,151**	**2,123**	**0,73**	**2,08**	**2,8**	**0,008**	**50**	**2,757**	
	Ton variierbar	25				0,443	0,597	0,21	2,59	2,8	0,431	48	0,794	
Gesamt-dauer	Expertenkommentare	28	0,441	-0,77	319,5									26
	Expertenkommentare mit Signal	23	0,288	-1,062	246,0									
	Pausieren	27	**0,000**	**-6,209**	**2,0**									
	Ton variierbar	25	0,187	-1,319	255,0									
Richtigkeit	Expertenkommentare	28	0,165	-1,39	284,5									26
	Expertenkommentare mit Signal	23	0,218	-1,233	238,0									
	Pausieren	27	0,366	-0,905	300,5									
	Ton variierbar	25	0,654	-0,448	301,5									
Verständ-lichkeit	Expertenkommentare	27	0,197	-1,29	278,5									26
	Expertenkommentare mit Signal	23	0,237	-1,182	240,0									
	Pausieren	26	0,993	-0,009	337,5									
	Ton variierbar	24	0,808	-0,243	299,5									

Anhang 7: Test auf Normalverteilung der Variable Aufgabenschwierigkeit

	Shapiro-Wilk		
	Statistik	df	Signifikanz
Grundvariante	,781	25	,000
Expertenkommentare	,829	27	,000
Expertenkommentare mit Signal	,848	26	,001
Pausieren	,859	22	,005
Ton variierbar	,808	19	,002

Anhang 8: U-Tests zur erlebten Aufgabenschwierigkeit

Gruppenvergleich Grundvariante mit …	Mann-Whitney-U	Z	Signifikanz
Expertenkommentare	310,0	-,535	,593
Expertenkommentare mit Signal	237,5	,000	1,000
Pausieren	313,5	-,229	,819
Ton variierbar	267,0	-,180	,857

Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie

Zimmermann, J., Konrad, S. & Nerdinger, F.W. (2009). Bedarfs- und Anforderungsanalyse zur Entwicklung einer internetbasierten Kommunikationsplattform zur Unterstützung des Forschungstransfers. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 1.* Universität Rostock.

Pundt, A., Martins, E., Vetterlein, A. & Nerdinger, F. W. (2009). Betriebsräte und Mitarbeiter in betrieblichen Innovationsprozessen. Stand der Forschung und Entwicklung eines psychologischen Forschungsmodells. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 2.* Rostock: Universität Rostock, Lehrstuhl für Wirtschafts- und Organisationspsychologie.

Stracke, S. & Nerdinger, F. W. (2009). "Alles unter einen Hut bringen?" Rollen und Rollenkonflikte von Betriebsräten bei betrieblicher Innovation. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 3.* Rostock: Universität Rostock, Lehrstuhl für Wirtschafts- und Organisationspsychologie.

Beile, J., Glass, E., Röhrig, R. & Stracke, S. (2010). Betriebliche Sanierungs- und Innovationsvereinbarungen in der Metall- und Elektroindustrie: Nachhaltige Bündnisse für Innovations- und Wettbewerbsfähigkeit? *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 4.* Rostock: Universität Rostock, Lehrstuhl für Wirtschafts- und Organisationspsychologie.

Breyer, T., Curth, C., Martins, E., Pundt, A. & Nerdinger, F. W. (2010). Innovatives Verhalten - Ein Geben und Nehmen? Innovation als Austauschprozess zwischen Mitarbeitern und Unternehmen. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 5.* Rostock: Universität Rostock, Lehrstuhl für Wirtschafts- und Organisationspsychologie.

Martins, E. & Breyer, T. (2010). Der Betriebsrat als normative Referenzgruppe für innovatives Verhalten. Empirische Untersuchungen der Bedingungen und der Wirkung auf das innovative Verhalten der Mitarbeiter. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 6.* Rostock: Universität Rostock, Lehrstuhl für Wirtschafts- und Organisationspsychologie.

Sprenger, W. (2011). Trade Unions and innovation - innovative unions? Experiences from selected EU member states. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 7.* Rostock: Universität Rostock, Lehrstuhl für Wirtschafts- und Organisationspsychologie.

Müller, C.; Curth, S. & Nerdinger, F. W. (2012). Demografischer Wandel, alternde Belegschaften und betriebliche Innovation. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 8.* Rostock: Universität Rostock, Lehrstuhl für Wirtschafts- und Organisationspsychologie.

Büttner, B. C.; Maaß, S. & Nerdinger, F. W. (2012). Wissenschaftliche Weiterbildung und Öffnung für nicht-traditionelle Zielgruppen als Herausforderungen für Hochschulen – Eine empirische Untersuchung zu den Sichtweisen von Hochschullehrern und Verwaltungsmitarbeitern an der Universität Rostock. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 9.* Rostock: Universität Rostock, Lehrstuhl für Wirtschafts- und Organisationspsychologie.

Stracke, S. & Haves, J. (2013). Personalarbeit mit alternden Belegschaften. Eine Analyse betrieblicher Demografieprojekte. *Rostocker Beiträge zur Wirtschafts- und Organisationspsychologie, Nr. 10*. Universität Rostock, Lehrstuhl für Wirtschafts- und Organisationspsychologie.