

Entwicklung und Validierung einer Experimentalumgebung zur Messung mentaler Beanspruchungszustände

Yannick FUNK, Henrike HAASE, Julian REMMERS, Barbara DEML

*Institut für Arbeitswissenschaft und Betriebsorganisation,
Karlsruher Institut für Technologie
Engler-Bunte-Ring 4, D-76131 Karlsruhe*

Kurzfassung: In dieser Arbeit wurde eine Experimentalumgebung zur Messung eines kognitiven Belastungsspektrums entwickelt und validiert. Mentale Belastung wird dabei mit einer Überwachungstätigkeit erzeugt, deren Bearbeitungsgeschwindigkeit randomisiert variiert. Zusätzlich wird die Komplexität der Tätigkeit schrittweise durch eine visuelle und eine auditive Nebenaufgabe erhöht. Die von den Versuchspersonen für die so variierten Belastungsgrade subjektiv empfundene mentale Beanspruchung wird mit Hilfe der *Rating Scale Mental Effort* (RSME) nach Zijlstra (1993) bewertet. Zwei erste Validierungsstudien (N = 17 & N = 8) zeigen, dass unsere Experimentalumgebung Belastungsgrade mit einem quasilinearen Anstieg von *wenig* bis *stark anstrengend* induzieren kann.

Schlüsselwörter: mentale Beanspruchung, RSME, mental Workload

1. Hintergrund

Im Rahmen des vom BMBF geförderten Verbundvorhabens Fahrerkabine 4.0 (Geimer, 2019) wird am KIT eine adaptive Mensch-Maschine-Schnittstelle für Landmaschinen entwickelt, die in der Lage ist, das aktuelle Beanspruchungsniveau der Fahrer*Innen mit Hilfe physiologischer Daten zu detektieren. Daraus sollen Handlungsempfehlungen abgeleitet werden: so können z.B. bei geringer Beanspruchung während einer automatisierten Erntefahrt zusätzliche Büroaufgaben, die andernfalls am Ende eines langen Tages bearbeitet werden müssen vorgeschlagen werden. Bei hoher Beanspruchung wie etwa bei Wendemanövern lassen sich unkritische Teilaufgaben oder Informationen verzögern.

Damit ein solches System zunächst im Labor entwickelt und anschließend an einem Demonstrator erprobt werden kann, entwickelt das ifab eine Experimentalumgebung, mit deren Hilfe sich die dafür benötigten mentalen Beanspruchungszustände reproduzierbar induzieren lassen. Beanspruchung wird in diesem Zusammenhang als eine individuelle Reaktion auf externe Belastung definiert (vgl. Rohmert, 1983). Sie ist demnach eine Funktion der Belastung, und hängt darüber hinaus von den individuellen Fähigkeiten und Eigenschaften der belasteten Menschen ab (Schlick et al., 2018): Eine gleiche Belastung kann zu individuell unterschiedlicher Beanspruchung führen. Man versucht deshalb sowohl aus gesundheitlichen Gründen als auch zur Optimierung der Produktivität, Arbeitsplätze an ihre individuellen Benutzer*Innen anzupassen.

Der Versuchsaufbau geht von De Waards (1996) Zusammenhang zwischen Leistung und mentaler Beanspruchung aus (vgl. Abbildung 1). Hiernach sollte es möglich sein, ein Beanspruchungsspektrum von Unterforderung über optimale Belastung bis hin zu Überforderung in Proband*Innen unterbrechungslos zu induzieren, dabei die

Leistung zu verfolgen und die individuellen physiologischen Reaktionen auf die Belastung zu messen. Ziel unserer Studien ist es, einen solchen Versuchsaufbau zu realisieren.

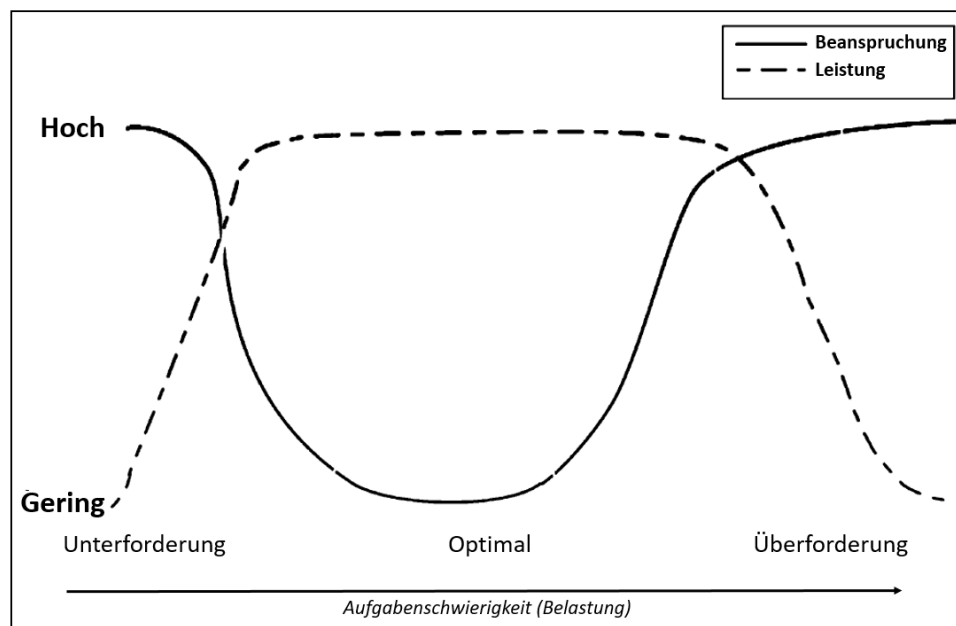


Abbildung 1: Zusammenhang von Beanspruchung und Leistung nach De Waard (1996, S. 24).

2. Methode

Unsere Experimente erzeugten mentale Belastungen durch eine bildschirmbasierte Überwachungstätigkeit in fünf randomisierten Geschwindigkeitsstufen (Studie 1), ergänzt um eine visuelle Nebenaufgabe (Studie 2). Die Bewertung der subjektiven Beanspruchung erfolgt jeweils auf Basis der Rating Scale Mental Effort (RSME; vgl. Zijlstra 1993). In einer geplanten dritten Studie wird die Belastung durch eine zusätzliche, auditive Nebentätigkeit weiter erhöht werden, um auch die in den Studien 1 und 2 noch nicht erreichten Überforderungszustände zu untersuchen.

Studie 1 sollte zunächst validieren, dass eine kürzere Ereignisfrequenz zu einer höher empfundenen Beanspruchung führt. In Studie 2 sollte untersucht werden, inwieweit das Hinzufügen einer Nebenaufgabe die Beanspruchung zusätzlich beeinflusst.

2.1 Stichprobe

An Studie 1 nahmen 17 Personen (11 Frauen, 6 Männer) im Alter von 22 bis 42 Jahren ($M = 26.5$, $SD = 4.55$) teil. Studie 2 umfasste 8 Personen (3 Frauen, 5 Männer) im Alter von 23 bis und 29 Jahren ($M = 26.71$, $SD = 2.85$). Beide Stichproben wurden aus freiwilligen Studierenden und Mitarbeitenden des KIT rekrutiert. Die Studien wurden durch die Ethikkommission des KIT genehmigt. Alle Proband*Innen wurden vorab über den Untersuchungsablauf, mögliche Risiken, ihre Rechte und die Anonymität der Daten informiert. Einverständniserklärungen liegen vor. Es wurde darauf geachtet, dass die Versuchspersonen keine Einschränkungen des Sehvermögens aufwiesen bzw. eine geeignete Sehhilfe für die Experimente benutzt wurde.

2.2 Versuchsaufbau und -ablauf

Zur Induktion verschiedener mentaler Belastungszustände auf die Teilnehmenden wurde zunächst eine reine Überwachungstätigkeit in Studie 1 untersucht. Dabei wurde auf einem Computermonitor (1920x1080 Pixel, 22 Zoll) ein Video von einer Mähdrescherfahrt aus der Egoperspektive gezeigt. Während des Videos wurden in randomisierter Reihenfolge die Buchstaben „W“, „A“ oder „D“ in fünf, ebenfalls randomisierten Frequenzstufen (alle 15, 10, 5, 3 oder 2 Sekunden, im Folgenden als Stufe 1-5 bezeichnet) auf dem Bildschirm eingeblendet. Die Aufgabe der Proband*Innen bestand darin, die angezeigten Buchstaben auf einer Computertastatur mit der linken Hand zu drücken. Die Teilnehmenden wurden angewiesen, eine korrekte Antwort einer möglichst schnellen Antwort zu bevorzugen. Jede Frequenzstufe wurde zwei Minuten lang angezeigt. Nach jeder Stufe wurden 15 Sekunden lang keine weiteren Buchstaben angezeigt. In diesem Zeitfenster sollten die Versuchspersonen ihre subjektiv empfundene mentale Beanspruchung auf einer Skala von 0 bis 150 (RMSE) bewerten. Insgesamt wurde dieses Experiment zweimal pro Proband*In durchgeführt. Der erste Durchgang wurde dabei als Übungsphase gewertet.

Laut Wickens (2002) Ressourcenmodell interferieren zwei Aufgaben der gleichen Ausprägung (in diesem Fall visuell) in einem höheren Grad als Aufgaben unterschiedlicher Ausprägungen. Um eine höhere mentale Beanspruchung, als in Studie 1 bereits nachgewiesen, zu induzieren wurde daher in Studie 2 die oben beschriebene Überwachungstätigkeit um eine visuelle Nebenaufgabe ergänzt. Dabei handelte es sich um eine Mentale Rotation (MR, vgl Eggemeier & Wilson, 1991). Es wurden vier verschiedene Figurenpaare randomisiert, in einem Abstand von drei Sekunden angezeigt. Die Figuren konnten zueinander gedreht (randomisierter Winkel) oder zueinander gedreht und zusätzlich gespiegelt sein.

Die Proband*Innen sollten analog zu Studie 1 die angezeigten Buchstaben mit der linken Hand auf der Tastatur drücken und parallel dazu mit der rechten Hand anzeigen, ob es sich bei der MR um eine Drehung (linke Pfeiltaste) oder um eine Drehung mit Spiegelung (rechte Pfeiltaste) handelt. Wie in Studie 1 wurden die Proband*Innen nach jeder zweiminütigen Buchstaben-Frequenzstufe gebeten, die subjektiv empfundene mentale Beanspruchung von 0 bis 150 (RMSE) zu bewerten.

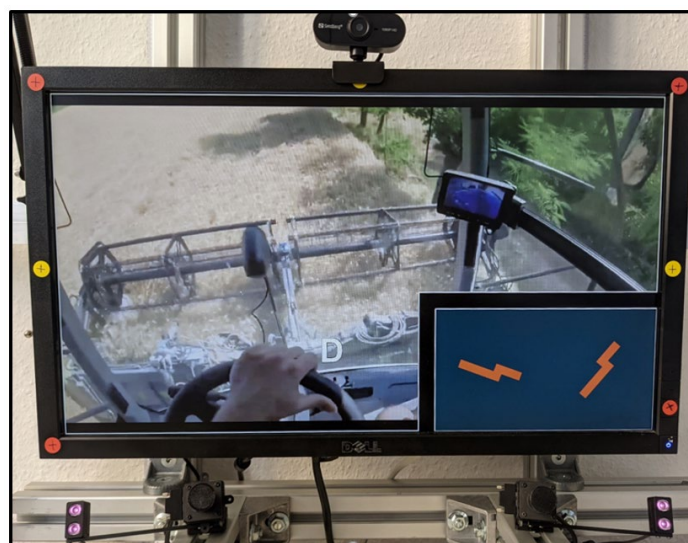


Abbildung 2: Überwachungstätigkeit und Mentale Rotation

2.3 Statistische Auswertung

Für die RMSE-Bewertungen der einzelnen Stufen wurden zunächst jeweils der zweite Durchgang innerhalb der Studien untersucht. Dazu wurden die Daten vor der Auswertung über den Shapiro-Wilk-Test auf Normalverteilung geprüft. Die Daten in Studie 1 wurden mithilfe einer einfaktoriellem ANOVA mit Messwiederholung untersucht. Aufgrund einer Verletzung der Normalverteilung durch die geringe Stichprobengröße in Studie 2 wurde hier ein Friedman-Test mit Dunn-Bonferroni-Post-Hoc-Test berechnet.

Im nächsten Schritt wurden die Ergebnisse der Studien mithilfe einer einfaktoriellem ANOVA miteinander verglichen. Post-hoc wurden die Ergebnisse der Varianzanalyse durch einen Tukey-Test untersucht.

Die Berechnungen wurden in SPSS und Matlab durchgeführt.

3. Ergebnisse

Die deskriptiven Daten lassen bereits darauf schließen, dass mit steigender Frequenz auch eine höhere mentale Beanspruchung erfahren wird. Dies lässt sich sowohl innerhalb von Studie 1 (mit Ausnahme der Stufen 3 und 4), innerhalb von Studie 2, als auch über beide Studien hinweg beobachten (Vgl. Tabelle 1). Darüber hinaus wird deutlich, dass die zusätzliche visuelle Nebenaufgabe die mentale Beanspruchung in allen Stufen erhöht.

Tabelle 1: Deskriptive Daten

Studie 1	N	M	SD	Studie 2	N	M	SD
Stufe 1	17	25.76	14.22	Stufe 1	8	56.00	16.19
Stufe 2	17	28.59	13.71	Stufe 2	8	66.00	22.48
Stufe 3	17	33.24	17.19	Stufe 3	8	67.14	14.96
Stufe 4	17	31.82	16.22	Stufe 4	8	75.00	19.37
Stufe 5	17	45.29	22.49	Stufe 5	8	79.29	14.27

Eine ANOVA mit Messwiederholung und Greenhouse-Geisser-Korrektur zeigte, dass sich die RSME-Bewertung der fünf Stufen in Studie 1 statistisch signifikant unterschieden [$F(1.96,31.35) = 6.092$, $p = .006$, partielles $\eta^2 = .276$]. Die Kontraste zeigen einen statistisch signifikanten Unterschied zwischen der Stufe fünf ($M = 45.29$, $SD = 22.49$) und den Stufen:

- eins ($M = 25.76$, $SD = 14.22$, $F(1,16) = 10.251$, $p = .006$, partielles $\eta^2 = .390$),
- zwei ($M = 28.59$, $SD = 23.72$, $F(1,16) = 7.295$, $p = .016$, partielles $\eta^2 = .313$),
- drei ($M = 33.24$, $SD = 17.20$, $F(1,16) = 7.822$, $p = .013$, partielles $\eta^2 = .328$),
- und vier ($M = 31.82$, $SD = 22.50$, $F(1,16) = 6.530$, $p = .021$, partielles $\eta^2 = .290$).

Zwischen den anderen Stufen bestanden keine signifikanten Unterschiede.

Ein Friedmann-Test für Studie 2 zeigte, dass sich die RSME-Bewertung der fünf Stufen statistisch signifikant unterschieden, $\text{Chi-Quadrat}(4) = 10.96$, $p = .027$, $n = 8$. Anschließend durchgeführte Post-hoc-Tests (Dunn-Bonferroni-Tests) zeigen, dass sich die Stufen eins und fünf ($z = -3.083$, $p = .020$, Effektstärke nach Cohen: $r = .86$) signifikant unterscheiden.

Die Untersuchung beider Gruppen (Studie 1 – reine Überwachungstätigkeit und Studie 2 – Überwachung mit Mentaler Rotation) über eine einfaktoriellem ANOVA zeigte, dass sich die RSME-Bewertungen statistisch signifikant voneinander unterschieden,

$F(9,115) = 12.934$, $p < .001$, partielles $\eta^2 = .276$. Statistisch signifikante Unterschiede des Post-hoc Tukey-Tests sind in Abbildung 3 dargestellt. Stufen 1-5 (grün) beziehen sich auf Studie 1 und Stufen 6-10 (blau) auf Studie 2.

1	6	$p = 0.001$	2	6	$p = 0.031$	
	7	$p < 0.001$		7	$p < 0.001$	
	8	$p < 0.001$		8	$p < 0.001$	
	9	$p < 0.001$		9	$p < 0.001$	
	10	$p < 0.001$		10	$p < 0.001$	
			3	7	$p = 0.030$	
				8	$p = 0.033$	
				9	$p < 0.001$	
			4	7	$p = 0.002$	
				8	$p = 0.002$	
				9	$p < 0.001$	
				10	$p < 0.001$	
				5	9	$p = 0.032$
					10	$p = 0.002$

Abbildung 3: Post-Hoc-Vergleich der Stufen 1-5 in Studie 1 und Studie 2.

4. Diskussion und Ausblick

Aus den statistisch signifikanten Unterschieden zwischen den Stufen

- eins und fünf innerhalb beider Studien
- eins der Studie 1 und fünf der Studie 2
- fünf der Studie 1 und fünf der Studie 2

lässt sich ein quasilinearer Anstieg von wenig anstrengend bis ziemlich anstrengend (RSME-Rating, vgl. Zijlstra (1993)) ableiten (vgl. Abbildung 4). Somit können wir mit der aktuellen Version der Experimentalumgebung ein Beanspruchungsspektrum von geringer bis mittlerer/hocher Beanspruchung induzieren.

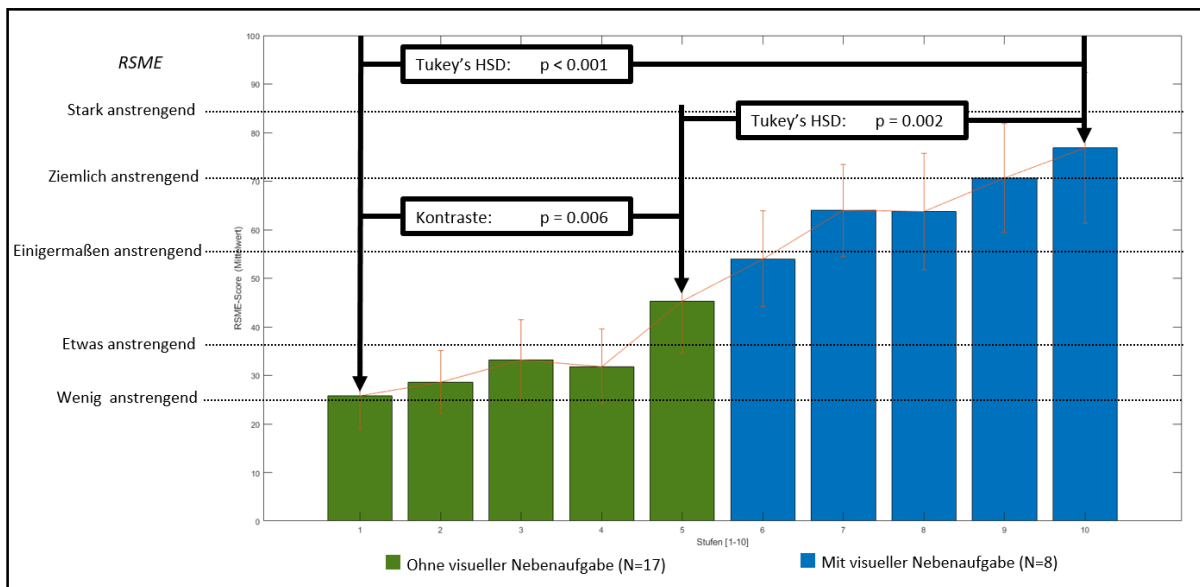


Abbildung 4: Quasilinearer Anstieg induzierter mentaler Beanspruchung in den Studien 1 und 2, jeweils im zweiten Durchgang.

Die höchste, in beiden Studien induzierte Beanspruchung (Studie 2, Stufe fünf, $M = 76.90$) entspricht der RSME-Bewertung ziemlich anstrengend. Die maximale mentale Beanspruchung auf der RSME-Skala ist jedoch außerordentlich anstrengend mit einer Bewertung von 112. Damit ein System zur Erfassung mentaler Beanspruchung in allen Bereichen evaluiert werden kann, muss sichergestellt sein, dass alle Versuchspersonen sowohl mindestens einen Unterforderungszustand (wenig anstrengend) als auch

mindestens einen Überforderungszustand (außerordentlich anstrengend) erfahren. Ziel ist es, das volle Spektrum psychophysiologischer Reaktionen auf mentale Belastungen bei den Teilnehmenden hervorzurufen.

In einer weiteren Studie 3 wird der Versuch deshalb um eine auditive Nebenaufgabe ergänzt werden. Diese Kombination aus Überwachungstätigkeit, visueller und auditiver Nebenaufgabe sollte auch hohe und sehr hohe Beanspruchungszustände generieren können. Die auditive Nebenaufgabe wird wie die visuelle eine konstante Belastung auf die Teilnehmenden ausüben. Verschiedene Schwierigkeitsgrade werden weiterhin über die Variation der Ereignisfrequenzen (Stufe eins bis fünf) induziert. Mit den Daten der Studie 3 werden auch die Ergebnisse aus Studie 2 noch einmal evaluiert werden, um den Einfluss der geringen Stichprobengröße der Studie weiter einzugrenzen.

Zusammenfassend ist festzustellen, das sich mit Hilfe der vorgestellten Experimentalumgebung Systeme zur kabellosen Echtzeiterfassung mentaler Beanspruchung in mindestens drei Ausprägungen (Unterforderung, optimale Beanspruchung, Überforderung) entwickeln und validieren lassen.

5. Literatur

- De Waard, D. (1996). The measurement of drivers' mental workload. Traffic Research Center, Groningen, Niederlande.
- Eggemeier, T. F. & Wilson, G. F. (1991). Performance-based and subjective assessment of workload in multi-task environments. In D. L. Damos (Hrsg.), Multiple-task performance (S. 217–275). Taylor & Francis, London and Washington.
- Geimer, M. (2019). Fahrerkabine 4.0. Abgerufen November 27, 2020, von <https://www.agrarsysteme-der-zukunft.de/konsortien/fahrerkabine-40>
- Rohmert, W. (1983). Belastungs-Beanspruchungs-Konzept. In Praktische Arbeitsphysiologie, (S. 9-13). Georg Thieme Verlag, Stuttgart, Deutschland
- Schlick, C., Bruder, R., & Luczak, H. (2018). Arbeitswissenschaft. Springer-Verlag, Heidelberg, Deutschland.
- Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical issues in ergonomics science, 3(2), 159-177.
- Zijlstra, F. R. H. (1993). Efficiency in work behavior: a design approach for modern tools (Dissertation). Delft University Press, Delft, Niederlande.

Danksagung: Dieser Beitrag entstand im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) gefördert Verbundvorhabens „Fahrerkabine 4.0“. Das Forschungsvorhaben wird vom Projektträger Jülich (PTJ) betreut.



Gesellschaft für
Arbeitswissenschaft e.V.

Arbeit HUMAINE gestalten

67. Kongress der
Gesellschaft für Arbeitswissenschaft

Lehrstuhl Wirtschaftspsychologie (WiPs)
Ruhr-Universität Bochum

Institut für Arbeitswissenschaft (IAW)
Ruhr-Universität Bochum

3. - 5. März 2021

GfA-Press

Bericht zum 67. Arbeitswissenschaftlichen Kongress vom 3. - 5. März 2021

**Lehrstuhl Wirtschaftspsychologie, Ruhr-Universität Bochum
Institut für Arbeitswissenschaft, Ruhr-Universität Bochum**

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Dortmund: GfA-Press, 2021
ISBN 978-3-936804-29-4

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Alle Rechte vorbehalten.

© **GfA-Press, Dortmund**

Schriftleitung: Matthias Jäger

im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Ohne ausdrückliche Genehmigung der Gesellschaft für Arbeitswissenschaft e.V. ist es nicht gestattet:

- den Kongressband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen,
- den Kongressband oder Teile daraus in Print- und/oder Nonprint-Medien (Webseiten, Blog, Social Media) zu verbreiten.

Die Verantwortung für die Inhalte der Beiträge tragen alleine die jeweiligen Verfasser; die GfA haftet nicht für die weitere Verwendung der darin enthaltenen Angaben.

Screen design und Umsetzung

© 2021 fröse multimedia, Frank Fröse

office@internetkundenservice.de · www.internetkundenservice.de