

# VOM SINN UND UNSINN DER MESSUNG PSYCHISCHER BELASTUNG UND BEANSPRUCHUNG

· *Belastung* · *Beanspruchung* · *Messmethoden* · *Bewertungskriterien*

## ZUSAMMENFASSUNG

Es wird dargelegt, dass zahlreiche der in jüngerer Zeit vorgelegten Untersuchungen zum Problemkreis der psychischen Belastung und Beanspruchung auf Methoden zurückgreifen, die sich in der Vergangenheit als untauglich erwiesen haben. Daher sind die so erhobenen Befunde ungeeignet, als Bezugssystem für zulässige Beanspruchung zu dienen.

## PRAKTISCHE RELEVANZ

Die Flut an Publikationen kann nicht darüber hinwegtäuschen, dass die Arbeitswissenschaft derzeit über keinen verlässlichen Methodensatz zur Messung der psychischen Belastung und Beanspruchung verfügt, selbst wenn enthusiastische Autoren häufig eine Bewertung mit einer Messung verwechseln.

## A PROPOS DU SENS ET DU NON-SENS DES MESURES DE LA CONTRAINTES ET DE L'ASTREINTE PSYCHIQUES

· *contrainte* · *astreinte* · *méthode de mesure* · *critères  
d'évaluation*

## RESUME

Il sera démontré que beaucoup d'expériences faites récemment concernant la problématique de la contrainte et de l'astreinte psychiques, se basent de nouveau sur des méthodes qui ont été considérées non fiables dans le passé. De ce fait, les résultats obtenus ne peuvent être utilisés comme données de référence pour l'évaluation de l'astreinte psychique maximum acceptable.

## IMPORTANCE PRATIQUE

Malgré le flot des publications, on ne peut pas nier le fait que l'ergonomie ne possède pas actuellement de méthodes de mesure fiables pour mesurer la contrainte et l'astreinte psychique ; même si pour quelques auteurs enthousiastes, il n'existe pas de différence entre l'évaluation et la mesure.



# SENSE AND NONSENSE OF THE MEASUREMENT OF MENTAL STRAIN AND WORKLOAD

· strain · stress · testing methods · criteria for evaluation

## SUMMARY

Assured knowledge in the field of mental strain and workload can be found only if the object of research is approachable by measurement. A measurement of mental or informational strain has been impossible up to now because of the lack of a frame of reference for the extremely diversity of tasks. Muscular strain can be measured in N or N/m and the workload in kJ or Watt. But we are helpless if we are asked to nominate or measure the strain factors behind the solution of a mathematical problem or a translation task. Even a clear distinction between physiological and psychological strain seems to be problematical.

But not only in the field of strain we have measurement problems. Also we find problems in the measurement of mental workload. This can be seen by the titles of the articles in this issue. The authors speak from „acquisition“, „determination“ and „stipulation“, they are using terms which should conceal the inability of a measurement in a physical-technical sense. This is not a criticism but it demonstrates the insight of the authors that we are far away from a real measurement.

If we try to measure mental workload we need a working hypothesis about the reaction of workload on the central nervous system and other physiological functions. The second step is the selection of measurement tools. Here we are between the devil and the deep blue sea. In the last hundred years nearly all physiological functions which could have a covariance with workload have been tested upon their aptitude. The results were overmodest. Neither the heart rate nor the cardiac arrhythmia has supplied information which are in a strong relation to type and duration of the preceded informational task. Also the attempts to extract indicators out of the complex functions of the EEG have been not very successful. The only information with high reliability which could be derived from the spontaneous EEG of healthy subjects was the differentiation between open or closed eyes! We believe in cheaper ways of getting this result.

The third step is the selection of tasks. Many tasks used in experiments look quite artificial compared with real tasks in industry or office environment. Which insight we get from the result of an experiment that shows that a workload indicator will react positive with increasing task complexity and test duration and the greatest ratio of variance can be ascribed to differences between test subjects? The outcome of many experiments in this area is equal to the use of common sense. Frequently researchers have found correlation between a deflection of a workload indicator and working time but to be afflicted with a lot of statistical uncertainty. Using a watch for measuring working time would lead to much more precise results!

The last step is the selection of test subjects. If we look in literature usually subjects in workload studies are students. They can be re-

cruited more easily than employees from industry. But it has to be questioned if it is acceptable to generalise test results attained with students because their motivation and emotional status seems to be quite different to these of employees in industry and office work.

Research in the area under discussion can be only successful if the measurement methods are validated on a defined criterion. Further on it must be clear which criterion has to be selected. The criterion (e.g. reasonability or tolerability) must be defined explicit. The measurement methods must be sufficient sensitive and methods gather only single aspects of workload lead to wrong conclusions.

It makes sense if in basic research it will be searched for methods which can differentiate with high reliability between different types of mental workload. It makes also sense to define the validation criteria very thoroughly. But it makes no sense as an effect of careless literature inquiry to warm up measuring methods which have shown no contributions to problem solving. As a matter of fact it seems to be unaccountable to derive from the here criticised research approaches standards on the tolerable human workload by informational tasks.

Finally the prevalence of project research forces the researcher to publish preliminary reports, progress reports, abstracts, and final reports. This increases the individual list of publications enormous. But who believes that 20 or 30 publications of an author per year can be based on serious research even if the same content will be printed in different journals with a slightly modified title? Because as a general rule only such research will be sponsored which has positive results in terms of the financier. This constraint can lead to omit unsuitable data - called as artefacts - or to invent data. This fraudulent falsification of data particularly in biomedical science has been subject-matter during the last years. But the increasing quantity of publications has another negative effect: it stimulates desinformation. Even in a narrow special field the researcher can read frequently only titles. Unfortunately titles describe in many cases the content consciously incomplete or wrong. A citation on the basis of titles can be beneficial for the impact factor of a journal but does not stimulate information transfer within the scientific community. The evaluation of research results on the basis of the impact factor of the publishing journal might be proved as the largest slip-up within science of present time.

## PRACTICAL RELEVANCE

The rising flow of publications can not deceive thereover that in ergonomics presently no valid methods for measurement of mental strain and workload are available even if the authors in youthful enthusiasm frequently mistake evaluation for measurement.



## 1 EINLEITUNG

Es ist nicht überraschend, dass sich in einer Zeit überbordender elektronischer Datenverarbeitung die Frage nach einer Quantifizierung der psychischen oder mentalen oder informatorischen Belastung und Beanspruchung immer dringlicher stellt. Selbst in der industriellen Produktion nimmt der Anteil schwerer Muskelarbeit kontinuierlich ab und wird von Aufgaben der Überwachung von Anlagen und der Kontrolle von Gütern abgelöst. Der Gesetz- und Verordnungsgeber sowie die für die Herausgabe von Richtlinien, Regelwerken und Normen zuständigen nationalen und internationalen Institutionen verlangen, bei der Gestaltung von Arbeitsplatz, Arbeitsumgebung und Arbeitsinhalt die gesicherten arbeitswissenschaftlichen Erkenntnisse anzuwenden. Gesicherte Erkenntnisse existieren über die Auswirkungen schwerer dynamischer und statischer Muskelarbeit und über die gesundheitlichen Risiken einer Reihe von Umweltparametern, wobei allerdings stets hinterfragt werden muss, unter welchen Bedingungen und wann eine wissenschaftliche Erkenntnis als gesichert gelten kann. Für mentale Belastung und Beanspruchung durch Aufnahme, Verarbeitung und Umsetzung von Informationen gibt es aber bis heute keine gesicherten Erkenntnisse. So verwundert es nicht, dass die Herausgeber der Zeitschrift für Arbeitswissenschaft diesem Fragenkomplex ein Themenheft gewidmet haben.

„Gesicherte Erkenntnisse“ im Bereich psychischer (informatorischer) Belastung und Beanspruchung lassen sich nur gewinnen, wenn der Forschungsgegenstand einer wie auch immer gearteten Messung zugänglich ist. Einer solchen Messung hat sich die Komponente *Belastung* bisher als unzugänglich erwiesen. Dies liegt u.a. am Fehlen eines Bezugssystems für die außerordentliche Vielfalt an Aufgaben, die durch mentale Tätigkeit zu bewältigen sind. Während wir muskuläre Belastung in N bzw. N/m und die entsprechende Beanspruchung in kJ bzw. Watt oder Lärmbelastung in dB(A) und die Beanspruchung z.B. durch die TTS (temporary threshold shift) messen können, sind wir ratlos, wenn die hinter der Lösung einer mathematischen Aufgabe oder die hinter der Übersetzung eines fremdsprachlichen Textes stehenden Belastungsfaktoren benannt oder gar gemessen werden sollen. Hinzu kommt, dass auch eine klare und durchgehende Unterscheidung zwischen physiologischer und psychischer Belastung problematisch ist. Die Vermutung im Beitrag „Untersuchung der Sensitivität von NASA-TLX

und NASA-TLX-ZEIS bei einer Flugführungsaufgabe“ in diesem Heft, dass mit diesen Befragungsmethoden „verallgemeinernde Aussagen zu der bei der Durchführung von Flugführungsaufgaben entstehenden Belastungs- und Beanspruchungshöhe mit befriedigender Genauigkeit“ machen zu können, muss zumindest hinsichtlich der Belastungskomponente zurückgewiesen werden. Offenbar werden die Begriffe *Belastung* und *Beanspruchung* synonym verwendet!

Dass es auch mit einer Messung der *Beanspruchung* durch derartige Aufgaben nicht zum Besten bestellt ist, weisen schon die Überschriften der in diesem Heft abgedruckten Beiträge aus. Hier wird von „Erfassung“ „Ermittlung“ und „Bestimmung“ gesprochen. Es werden also Termini gewählt, die möglicherweise das Unvermögen einer Messung im physikalisch-technischen Sinne kaschieren sollen. Eine solche Aussage diskriminiert die Autoren nicht, sondern belegt nur deren Einsicht, dass wir auch von einer echten Beanspruchungsmessung noch weit entfernt sind.

## 2 ZUR KRITIK DES METHODISCHEN ANSATZES

Will man Beanspruchung messen, dann muss man zunächst wenigstens eine *Arbeitshypothese* darüber haben, wie der menschliche Organismus auf eine von außen einwirkende Belastung reagiert. Hat man eine solche, dann wird ein *Messinstrument* benötigt, das diese Reaktion zu erfassen gestattet. Ferner bedarf es einer *Aufgabe*, von der man glaubt, dass sie Prozesse der Aufnahme, Verarbeitung und Umsetzung von Informationen beinhaltet und auch *Menschen*, die diese Aufgabe ausführen. Schließlich müssen die *Messergebnisse* ausgewertet und auf ihre Aussagefähigkeit getestet werden.

### ♦ Arbeitshypothese

Mit der Formulierung einer Arbeitshypothese sind wir noch auf der relativ sicheren Seite, wenn wir unterstellen, dass informatorische Belastung Rückwirkungen auf zentralnervöse und andere physiologische Funktionen hat und sie auch emotionale und kognitive Faktoren beeinflusst.

### ♦ Messinstrument

Mit der Auswahl des Messinstrumentariums beginnt bereits das Dilemma. In den

vergangenen hundert Jahren wurden nahezu alle physiologischen Funktionen, von denen man eine Kovariation mit der Beanspruchung erwarten konnte, auf ihre Eignung getestet. Die Ergebnisse waren eher ernüchternd. Natürlich ist es trivial, dass die menschliche Physiologie auf Stress reagiert. Die zahllosen, über Jahrzehnte immer wiederholten Versuche, aus der Herzschlagfrequenz oder der Herzschlagfrequenz-Arrhythmie (oder -Variabilität) einen Kennwert zu extrahieren, der mit Art und Dauer der vorausgegangenen informatorischen Tätigkeit kovariert, haben insgesamt gesehen nur einen bescheidenen Erfolg gehabt, wenn man den Ausdruck „bescheiden“ im wahrsten Sinne des Wortes interpretiert. Dieser Sachverhalt wird aber - gründliches Literaturstudium vorausgesetzt - offenbar verdrängt und hält die Forscher nicht von dem Bemühen ab, das Rad immer wieder neu zu erfinden. Aber kann man gründliches Literaturstudium immer voraussetzen? Gelegentlich entsteht der Eindruck, dass bei manchen Forschern ein schlichtes Nichtwissen vorliegt. Dieses Nichtwissen besteht darin, „dass das zu berücksichtigende Wissen bereits vorliegt, und zwar sogar innerhalb der eigenen Disziplinengruppe, aber von den Individuen, die sich mit der zu lösenden Frage befassen, noch nicht gekannt wird“ (Zimmerli 1993).

Vergleichsweise bescheidenen Erfolg hatten auch die zahlreichen Bemühungen, aus den komplexen Funktionen des EEG Indikatoren über die Vorbeanspruchung des Zentralnervensystems zu extrahieren. Hier haben uns Forscher einzureden versucht, dass es bei einer geschickten Auswahl von EEG-Parametern möglich ist, eine Voraussage dahingehend zu machen, ob ein Mensch in einer definierten Situation auf ein in wenigen Sekunden erscheinenden Reiz reagieren wird oder nicht. Leider haben die Autoren in ihrer Präsentation keinen Hinweis auf die außerordentlich geringe Stabilität des ausgewählten Parametersatzes und damit auf die geringe Reliabilität dieses Messkonzeptes gegeben. Was wirklich aus dem Spontan-EEG gesunder Probanden zuverlässig abgeleitet werden konnte, war die Feststellung, ob diese die Augen offen oder geschlossen hatten (Hecker et al. 1980). Ob diese Feststellung die außerordentlich aufwendigen Experimente zu rechtfertigen vermag, möge jeder für sich entscheiden. Nun ist hier nicht der Ort, alle physiologischen Methoden aufzuführen; aber aus Gründen der Fairness soll nicht unerwähnt bleiben, dass auch der Autor dieses Beitrages (im Verein mit so renommierten Physiologen wie Ernst Simonson,



Kurt Wacholder, Etienne Grandjean u.a.) vor mehr als 5 Jahrzehnten einmal glaubte, mit dem Flimmertest den diesbezüglichen Stein des Weisen gefunden zu haben (Schmidtke H. 1949). Aber auch dieser Glaube hat sich später leider als ein Wunschtraum erwiesen.

Wenn wir schon nicht genau wissen, was *psychische Beanspruchung* eigentlich ist und dieses Phänomen darüber hinaus auch mit naturwissenschaftlichen Methoden nicht zu quantifizieren vermögen, dann liegt es doch nahe, einfach die Menschen zu befragen, wie sie die Beanspruchung in einer gegebenen Arbeitssituation mit informatorischen Inhalten subjektiv beurteilen. Die Ergebnisse von Studien mit Polaritätenprofilen oder Anstrengungsskalen - wie sie Bartenwerfer als einer der ersten verwendete - sind eher positiver zu bewerten als die mit physiologischen Methoden gewonnenen. Dies gilt auch dann, wenn wir keine detaillierten Kenntnisse darüber haben, was z.B. ein Skalenwert von 50 oder 100 wirklich bedeutet. Die unzähligen Bemühungen, mit Hilfe von Arbeitsversuchen - wie etwa dem Dükerschen Rechentest oder den Doppelarbeitsversuchen von Bornemann - dem Phänomen der psychischen Beanspruchung messtechnisch auf die Spur zu kommen, waren wissenschaftlich gesehen recht ergiebig. Sie lieferten Einsichten über den Einfluss von Aufgabenschwierigkeit und Tätigkeitsdauer auf die Leistung in diesen Testarbeiten, hatten aber leider kaum praktische Relevanz. Insbesondere lieferten sie keine Hinweise auf die Beanspruchung einer Datentypistin oder eines Radarlotsen, was wir aber gern wissen möchten.

#### ♦ Aufgabe

Bei allen Experimenten, die auf eine Analyse der psychischen Beanspruchung abzielen, werden Aufgaben gesucht, bei denen der Experimentator unterstellt, dass sie in dieser Richtung beanspruchend wirken. Viele ausgewählte Versuchstätigkeiten machen einen ausgesprochen artifiziellen Eindruck, wenn man sie mit realen Berufsarbeiten vergleicht. Welche Erkenntnis gewinnen wir, wenn ein Experiment zeigt, dass mit wachsender Komplexität einer Dekodierungsaufgabe (etwa in der Art Bild 3 des Beitrages Schütte in diesem Heft, „Bestimmung der bedingungsbezogenen Messgenauigkeit der Anstrengungsskala“) und zunehmender Versuchsdauer ein Beanspruchungsindikator positiv reagiert? Überrascht es, wenn Schütte in seiner Untersuchung findet, dass der größte Varianzanteil auf Unterschiede zwischen den Probanden in dieser Art von Experimen-

ten zurückzuführen ist? Man kann doch unterstellen, dass ein solches Versuchsergebnis - unabhängig von der Qualität des Indikators - nicht viel mehr Einsichten erbringt als der Gebrauch des gesunden Menschenverstandes! Ist es nicht so, dass als Ergebnis vieler Beanspruchungsexperimente in der Vergangenheit nicht mehr herausgekommen ist, als eine mit allen statistischen Unsicherheiten behaftete Abschätzung der Dauer der Versuchstätigkeit? Mit einer Uhr wäre ein solcher Befund genauer und billiger zu bekommen. Anders sieht es hingegen bei Simulatorexperimenten aus. Hat der Simulator eine hohe Realitätsnähe - wie z.B. die Ausbildungssimulatoren der Lufthansa -, dann ist die Distanz zwischen Berufs- und Versuchstätigkeit klein, insbesondere dann, wenn die Leistung in der Simulationsaufgabe für die Karriere der Versuchspersonen bedeutsam ist. Kovariiert hier der Beanspruchungsindikator mit der Aufgabenschwierigkeit - d.h. der Belastung - oder der Beanspruchungsdauer, dann ist ein Erkenntnisgewinn nicht auszuschließen.

Grundsätzlich ist jedoch auch bei der Auswahl der Versuchstätigkeiten die Versuchsdauer zu bedenken. Aus den Ergebnissen eines 30-Minuten-Versuchs auch nur indirekt auf die Beanspruchung innerhalb einer kompletten Arbeitsschicht zu schließen, ist schlicht abenteuerlich. Ein solches Vorgehen ist eigentlich nur dann zulässig, wenn es um die Analyse der Brauchbarkeit einer Methode geht.

#### ♦ Menschen

Verfolgt man die Literatur in dem hier zur Diskussion stehenden Felde, dann werden als Versuchspersonen für Experimente zur Beanspruchungsanalyse überwiegend Studenten oder Mitarbeiter des Experimentators auch dann herangezogen, wenn es sich nicht um Methodenstudien handelt. Natürlich sind diese billiger und leichter zu rekrutieren als etwa die Mitarbeiter in einem Wirtschaftsunternehmen. Aber ist es zulässig, aus den mit einem derartigen Versuchspersonenkollektiv gewonnenen Ergebnissen verallgemeinernde Rückschlüsse auf Berufstätige zu ziehen? Die im Arbeitsleben die psychische Beanspruchung modifizierenden emotionalen und motivationalen Faktoren bleiben unberücksichtigt, wenn man z.B. mit Studenten arbeitet. Möglicherweise sind auch Studenten- oder Mitarbeitergruppen homogener als die mehr oder weniger unausgelesenen Stichproben, die man in einem Betrieb vorfindet. Menschen sind sehr wohl in der Lage, zwischen verschiedenen

Beanspruchungssituationen zu differenzieren. So wird der Pilot eines modernen Kampfflugzeugs in einer Selbstbeurteilung seine psychische Beanspruchung während eines Konturenfluges mit höchster Geschwindigkeit auf jedweder Beanspruchungsskala sicher höher einschätzen, als die Beanspruchung in einer Schulungsstunde. Aber welche Einsichten gewinnen wir, wenn - wie in der Vergangenheit geschehen - Hausfrauen in einer Fremdbeurteilung die Beanspruchung z.B. eines Piloten beurteilen? Auch bei Methodenstudien ist Vorsicht angezeigt. Kann man tatsächlich unterstellen, dass eine mit Institutsmitarbeitern durchgeführte Methodenstudie auf der Basis von Flugführungsaufgaben mit positivem Ergebnis (wie bei Pfendler und Schütte in diesem Heft) auch einen gleich guten Befund zeitigen würde, wenn als Versuchspersonen Mitarbeiter der Flugsicherung zur Verfügung stünden?

#### ♦ Messergebnisse

Es ist trivial, dass Messergebnisse daraufhin überprüft werden müssen, ob sie zuverlässig (reliabel) sind, d.h. bei einer Versuchswiederholung durch den gleichen oder einen anderen Forscher gleiche oder vergleichbare Ergebnisse erbringen. Da absolut gleiche Ergebnisse in der Beanspruchungsforschung eher unwahrscheinlich sind, bleibt wohl nur der Weg, die Vergleichbarkeit der Befunde zu testen. Gleichermäßen trivial ist es, die Gewähr dafür zu fordern, dass die Messergebnisse tatsächlich das ausdrücken, was die Meßmethoden zu messen vorgeben (Validität). Für beide Prüfaufgaben bietet die Statistik einen bunten Strauß von Methoden an.

Aber auch bei der Analyse der publizierten statistischen Auswertungen von Messergebnissen sind wir keineswegs vor Überraschungen sicher. So hat sich in der medizinischen und psychologischen Literatur die Konvention eingeschlichen, einen Befund dann als „signifikant“ zu bezeichnen, wenn die Wahrscheinlichkeit eines Zufallsbefundes  $\leq 5$  Prozent ist ( $p \leq 0,05$ , Fehler erster Art). Beck-Bornholdt und Dubben fragen, ob man sich einen Feuermelder kaufen würde, der mit einer Wahrscheinlichkeit von 5 % Feuer meldet, ohne das es brennt und mit 20 % Wahrscheinlichkeit keinen Alarm gibt, obwohl ein Feuer ausgebrochen ist (Fehler zweiter Art). Auch eine große Anzahl von Tests schützt nicht vor Zufallsbefunden. Bei einer Testbatterie mit z.B. 7 Tests zur Beanspruchungsmessung wird man mit einer Wahrscheinlichkeit von 30 %

mindestens einen signifikanten Zufallsbefund mit  $p \leq 0,05$  bekommen! Fleißige Forscher werden so immer irgendwelche Signifikanzen finden, auch wenn diese absolut bedeutungslos oder gar unsinnig sind!

### 3 GRUNDLAGENFORSCHUNG UND ANGEWANDTE FORSCHUNG

Grundlagenforschung betreiben heißt, etwas im Verborgenen bereits Vorhandenes aufzufinden. Grundlagenforschung in der Arbeitswissenschaft heißt aber auch, Prinzipien und Konzepte dieser Wissenschaft zu überprüfen und offen und unvoreingenommen zu diskutieren. Zu den Prinzipien der Arbeitswissenschaft gehört es, die Interaktionen zwischen Mensch und Arbeit in einer definierten physikalischen und sozialen Umwelt zu analysieren und den Analysebefund vor dem Hintergrund des geltenden Rechtes und geltender Normen mit dem Ziel zu diskutieren, eine Synthese zwischen den psychophysischen Bedürfnissen des einzelnen Menschen und den wirtschaftlichen Ansprüchen der Arbeitswelt zu finden. Untersuchungen über die psychische Beanspruchung durch informatorische Belastung können durchaus in den Bereich der *Grundlagenforschung* fallen. Grundlagenforschung ist wertfrei; sie dient ausschließlich dem Erkenntnisgewinn. Wenn z.B. gefunden wurde, dass Lärm oberhalb eines kritischen Grenzwertes bei längerer Einwirkungsdauer neben anderen Effekten zu einer dauerhaften Verschiebung der Hörschwelle (PTS) führt, so ist das eine Grunderkenntnis. Wie weit hieraus Anwendungskonsequenzen (in Gesetzen, Verordnungen, Richtlinien oder Normen) gezogen werden, steht auf einem anderen Blatt.

Der Wiener Physiker Anton Zeilinger hat diese Wertfreiheit der Grundlagenforschung unlängst an einem schönen Beispiel dargestellt: Michael Faraday erhielt eines Tages den Besuch des britischen Schatzkanzlers in seinem Labor und zeigte diesem seine Experimente über Induktion und Elektrolyse. Auf dessen Frage, wozu diese Experimente eigentlich gut seien, antwortete Faraday, er wisse es selbst nicht so genau. „Aber Euer Ehren, eines Tages werden Sie darauf Steuern erheben!“ Es ist eher unwahrscheinlich, dass unser Finanzminister eines Tages aus den Ergebnissen arbeitswissenschaftlicher Grundlagenforschung Steuereinnahmen gewinnt. Dennoch sollte akzeptiert werden, dass auch in unserer Disziplin und hier wiederum im Felde der Belastungs- und Beanspruchungsanalyse geforscht wird ohne

die Zielvorgabe, den optimalen Wirkungsgrad bei der Aufnahme, Verarbeitung und Umsetzung von Informationen zu bestimmen.

Nun ist die Arbeitswissenschaft eine andere Disziplin als die Astrophysik. Von der durch Steuergelder subventionierten Arbeitswissenschaft erwartet die Öffentlichkeit Beiträge zu einer humanen Arbeitswelt. In einer humanen Arbeitswelt sollten Bedingungen herrschen, die eine Überforderung des Einzelnen genau so ausschließen, wie eine Gesundheitsgefährdung. Die Analyse von möglichen Überforderungen oder gesundheitsgefährdenden Bedingungen ist Gegenstand der *angewandten Forschung*. Untersuchungen über die psychische Beanspruchung durch informatorische Belastung sind dann angewandte Forschung, wenn für eine konkrete Arbeitssituation zu prüfen ist, ob die Beanspruchung innerhalb der Grenzen der Ausführbarkeit, der Erträglichkeit oder der Zumutbarkeit liegt. Aber auch hier gelten die gleichen methodischen Anforderungen, wie sie in der Grundlagenforschung üblich sind. Ob eine informatorische Tätigkeit ausführbar ist, kann mittels eines Fehler- oder Auslassungskriteriums relativ sicher bestimmt werden. Anders sieht es hingegen aus, wenn entschieden werden muss, ob diese Tätigkeit auch erträglich oder gar zumutbar ist. Für die Entscheidung über Erträglichkeit oder Zumutbarkeit sind Methoden erforderlich, die an diesen Kriterien validiert wurden.

Hier stehen wir vor einem weiteren Dilemma. Ich möchte versuchen, auch dieses an Hand eines Beispiels zu beschreiben, welches der englische Astrophysiker Eddington 1939 in seinem Buche „The Philosophy of Physical Science“ im Rahmen einer Diskussion über die Beziehungen zwischen Forschungsergebnissen und Lebenswirklichkeit dargelegt hat:

Ein Fischbiologe studiert die Unterwasserwelt; er wirft sein Netz aus und analysiert den Fang nach wissenschaftlichen Methoden: messen, wiegen und beschreiben. Nach zahlreichen Fängen kommt er zu folgendem Befund:

- ♦ alle gefangenen Fische sind länger als 5 cm,
- ♦ alle Fische haben Kiemen.

Er folgert daraus, dass auch alle künftigen Fänge zum gleichen Ergebnis führen. Ein kritischer Kollege, der in seiner Forschung

von einem anderen Kriterium ausgeht, stimmt zwar dem zweiten Befund mit den Kiemen zu, unterstellt jedoch, dass im Wasser zahllose Fische mit einer Länge unter 5 cm leben, die der Fischbiologe nicht fangen konnte, da dessen Netz eine Maschenweite von 5 cm hatte. Auf dieses Problem angesprochen erwiderte der Fischbiologe: „Alle Objekte, die ich nicht mit meinem Netz fangen kann, liegen prinzipiell außerhalb der Ichtologie und gehören nicht zu den Objekten, die in der Ichtologie definiert sind. Soweit es meine Forschung betrifft, können Objekte, die ich nicht mit meinem Netz fangen kann, nicht als Fische bezeichnet werden.“

Aus diesem Beispiel kann für unser Problem der Messung der psychischen Beanspruchung folgendes abgeleitet werden:

- ♦ Experimente, in denen zwei oder mehrere artifizielle Laboraufgaben hinsichtlich ihrer Beanspruchungshöhe verglichen werden, sind solange praktisch irrelevant, wie die Analysemethoden nicht an einem bestimmten Kriterium validiert worden sind.
- ♦ Es muss Klarheit darüber herrschen, welches Kriterium heranzuziehen ist.
- ♦ Das Kriterium (z.B. Erträglichkeit oder Zumutbarkeit) muss eindeutig definiert sein.
- ♦ Die Analysemethoden müssen hinreichend sensitive sein, d.h. eine möglichst geringe „Maschenweite“ haben und
- ♦ Analysemethoden, die nur Einzelaspekte der Beanspruchung zu erfassen vermögen, führen zu falschen Schlüssen.

### 4 FOLGERUNGEN BEZÜGLICH SINN UND UNSINN

Es macht Sinn, wenn im Rahmen der Grundlagenforschung nach Methoden Ausschau gehalten wird, die zwischen unterschiedlichen Formen der informatorischen Beanspruchung mit hoher Reliabilität differenzieren. Es macht auch Sinn, wenn versucht wird, die Validierungskriterien genau zu umschreiben und zu definieren. Hingegen ist es unsinnig, infolge nachlässiger Literaturrecherchen immer wieder Meßmethoden aufzuwärmen, deren Beiträge zur Problemlösung sich als gegen Null konvergierend erwiesen haben. Es ist gleichermaßen unsinnig, aus im Labor generierten Testaufgaben



auf mentale Beanspruchung im Beruf zu schließen, solange keine genauen Kenntnisse über den Grad der „Verwandtschaft“ zwischen Testaufgabe und konkreter Berufstätigkeit vorliegen. Und - last not least - erscheint es unsinnig, wenn aus kurzdauernden Versuchen mit Studenten auf die Beanspruchung innerhalb einer kompletten Arbeitsschicht oder einer langen Arbeitsperiode geschlossen wird.

Unverantwortlich dürfte es sein, aus den hier kritisierten Forschungsansätzen allgemeingültige Standards über die zulässige menschliche Beanspruchung bei informatorischer Arbeit abzuleiten. Unter dem Appell an die Bringschuld der Wissenschaft geraten leider immer wieder Wissenschaftler in die Versuchung, aus noch ungesichertem Wissen zu Regelungen beizutragen, die normativen Charakter haben. Zum Wissenschaftler gehört auch der Mut, Nichtwissen zu bekennen. Schließlich sollte es in einer angewandten Wissenschaftsdisziplin wie der Arbeitswissenschaft nicht unzulässig sein, gelegentlich auch den bereits erwähnten „gesunden Menschenverstand“ zu benutzen. Dies würde nicht nur Geld sparen, sondern zu einem fürsorglicheren Umgang mit der endlichen Lebensspanne der Forscher führen. Ob es allerdings hilfreich ist, einerseits Forderungen nach Qualitätssicherung von Mess- und Bewertungsverfahren aufzustellen, andererseits jedoch vor einer „Formulierung unverhältnismäßig hoher wissenschaftlicher Gütemaßstäbe“ zu warnen, wie bei Ullsperger und Windel in ihrem Beitrag „Psychische Belastung und Beanspruchung aus Sicht der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin“ in diesem Heft geschehen, mag dahingestellt bleiben. Das gilt auch für semantische Diskussionen, wie sie unlängst Oesterreich angestoßen hat.

Und schließlich: Durch das Überhandnehmen von Projektforschung werden die

Forscher zur Publikation von vorläufigen Berichten, Zwischenberichten, Abstracts und Abschlußberichten gezwungen. Dadurch wächst zwar das Publikationsverzeichnis des einzelnen Forschers enorm an; aber wer glaubt eigentlich noch, dass 20 oder gar 30 Veröffentlichungen eines Autors pro Jahr auf ernsthafter Forschung basieren können, selbst wenn häufig der gleiche Inhalt mit leicht modifiziertem Titel in verschiedenen Zeitschriften gedruckt wird? Da in der Regel nur gefördert wird, was positive Ergebnisse im Sinne des Geldgebers bringt, kann dieser Zwang auch zum Weglassen unliebsamer - als Artefakte umgedeuteter - Messwerte oder gar zur Erfindung von Daten führen, d.h. zur Fälschung. Schon 1994 hat Di Trocchio auf den Anstieg von Fälschungen in der Wissenschaft hingewiesen, auf eine Entwicklung, die sich in den letzten Jahren insbesondere im Felde der biologisch-medizinischen Disziplinen erschreckend fortgesetzt hat. Aber die wachsende Publikationsflut hat noch einen anderen negativen Effekt, sie fördert die *Desinformation*. Selbst auf einem engen Spezialgebiet kann man kaum noch mehr als Überschriften lesen. Überschriften beschreiben aber nicht selten den Inhalt bewusst unvollständig oder gar falsch. Zitierungen auf Basis von Überschriften mögen zwar dem *impact factor* einer Zeitschrift und damit zugleich dem Prestige eines Autors dienlich sein, fördern aber nicht den Informationstransfer innerhalb der scientific community. Die Bewertung von Forschungsergebnissen anhand des *impact factors* der veröffentlichenden Zeitschrift dürfte sich als die größte Fehlleistung im Wissenschaftsbetrieb der Gegenwart erweisen. Nicht nur sind auch seriöse Zeitschriften keineswegs gegen ein „Seilschaftsprinzip“ innerhalb des Gutachterkreises gefeit (wovon jeder Herausgeber einer Zeitschrift ein Lied singen kann); auch die Zuerkennung z.B. der dreißigfachen Bedeutsamkeit einer Publikation in einer

Zeitschrift mit dem *impact factor* 15 gegenüber einer mit dem *impact factor* 0,5 - wie in manchen Fakultäten üblich - dürfte sich als schlichter Schwachsinn erweisen.

## LITERATUR

**Beck-Bornholdt, H.-P. und Dubben, H.H.:** Der Hund der Eier legt. Rowohlt Taschenbuch Verlag, Reinbek, 2001

**Di Trocchio, F.:** Der große Schwindel. Betrug und Fälschung in der Wissenschaft. Campus, Frankfurt a.M., 1994

**Hecker, R., Schmidtke, H., Wegener, H.:** Reliabilität und Validität spektraler EEG-Parameter als Indikatoren der psychischen Beanspruchung. *Psychologia Universalis*, Band 41, Verlag Anton Hain, Meisenheim am Glan, 1980

**Oesterreich, R.:** Das Belastungs-Beanspruchungskonzept im Vergleich mit arbeitspsychologischen Konzepten. *ZfArbwiss.* 55, 162, 2001

**Schmidtke H.:** Flimmertest und psychische Ermüdung. Diss. TH Braunschweig, 1949

**Zimmerli, W.Ch.:** Der Mensch - Ein Deus ex Machina? In: K. Weis (Hrsg.), Bilder vom Menschen in Wissenschaft, Technik und Religion. Faktum 2, TU München, 1993

## ANSCHRIFT DES VERFASSERS

**em. Univ. Prof. Dr. rer. nat. Heinz Schmidtke**  
Lehrstuhl für Ergonomie der Technischen Universität München  
Boltzmannstraße 15  
85747 Garching  
E-Mail: schmidtke@lfe.mw.tu-muenchen.de