

4. Bestimmung von mentaler Beanspruchung , Maße und Gütekriterien

4.1 Arbeitsdefinition von mentaler Beanspruchung

Der Begriff der mentalen Beanspruchung beschreibt den Intensitätsaspekt der menschlichen Informationsverarbeitung (Kahnemann, 1973). Der Grad der Beanspruchung hängt von zwei Komponenten ab. Zum einen von der Komplexität der Aufgabe (Aufgabenanforderung) und zum anderen von den individuellen Leistungsvoraussetzungen (Fähigkeiten, Fertigkeiten, motivationalen Einstellungen). Mentale Beanspruchung ist also Folge einer komplexen Interaktion zwischen den Merkmalen einer Person und den Anforderungen einer Aufgabe unter dem Aspekt der Motivation (Manzey, 1998).

Das Ziel der psychophysiologischen Untersuchungen mentaler Beanspruchungszustände bei der Arbeit an einem Computerarbeitsplatz mittels Pupillometrie ist die Identifikation zentralnervöser und peripher-physiologischer Indikatoren, durch die ein Beanspruchungsgrad bei der Bewältigung kognitiver Aufgaben reliabel und valide abgebildet werden kann. Es geht hierbei um die Erforschung mentaler Ressourcen und ihrer Inanspruchnahme (Buld, 2000). Da dem Menschen für die Informationsaufnahme und -verarbeitung nur begrenzte Ressourcen zur Verfügung stehen, sinkt die Kapazität mit steigender Aufgabenschwierigkeit. In Anlehnung an Hancock und Meshaki (1988) ergibt sich daraus folgende Definition:

Mental Workload ist definiert als der Anteil der Kapazität, der durch eine Aufgabe gebunden ist.

Mental Workload ist dabei mit mentaler Beanspruchung gleichzusetzen, sofern man Beanspruchung nach Schönflug (1987) als Handlung zur Bewältigung einer Belastung definiert. Dies ist insbesondere bei Tätigkeiten der Fall, die einen hohen mentalen Anteil aufweisen. Da im gegebenen Setting eines Einzelcomputernutzers kein globales Arbeitsplatzsystem untersucht wird, ist unter diesem speziellen Fokus eine Unterscheidung von Mental Workload und mentaler Beanspruchung nicht notwendig (vgl. Buld, 2000).

4.2 Güterkriterien

Zu den beschreibenden, theoretischen und methodischen Zugängen der verschiedenen Konzepte haben sich auch unterschiedliche Messmethoden herausgebildet. Es lassen sich nach Facaoaru und Friedling (1985) vier Hauptklassen identifizieren. Hierbei handelt es sich um Verhaltensmaße, physiologische-, subjektive und analytische Maße.

Die Eignung einer bestimmten Methode für eine bestimmte Anwendungssituation wird anhand allgemeiner Beurteilungskriterien festgestellt. Als relevante Dimensionen gelten neben den üblichen Gütekriterien wie Objektivität, Validität und Reliabilität auch die Kriterien Sensitivität, Diagnostizität, Intrusion, Reaktionszeit, Benutzerakzeptanz und Erfassungsaufwand (Kramer, 1991; O'Donnel & Eggmeier, 1986; Tsang & Wilson, 1997; Wickens & Hollands, 2000).

Validität: Diese Dimension beschreibt, in welchem Ausmaß die Methode ausschließlich auf Variationen der mentalen Beanspruchung reagiert. Die Validität ist in dem Umfang verringert, in dem sich andere Faktoren (Muskeltonus, Emotionen, etc.) in den Messungen widerspiegeln. Da die mentale Beanspruchung ein hypothetisches Konstrukt ist, lässt sich die Konstruktvalidität nicht bestimmen. Zur Validierung wird somit üblicherweise auf Kriteriums- oder auf Vorhersagevaliditäten ausgewichen (Tsang & Wilson, 1997).

Reliabilität: Diese Dimension betrifft die Zuverlässigkeit einer Methode, also ihre Stabilität und Konsistenz. Nur für die wenigsten Beanspruchungsmaße existieren explizite Reliabilitätswerte (Kramer, 1991; Xie & Salvendy, 2000). Die Reliabilität einer Methode wird eher daraus abgeleitet, dass ein konsistenter Zusammenhang von Maß und Beanspruchung aus verschiedenen vorherigen Studien vorliegt.

Sensitivität: Je feiner sich Schwankungen in der mentalen Beanspruchung in dem gewählten Maß abbilden, desto sensitiver ist die Methode. Nach Beatty (1982) sollte ein geeignetes Maß in Bezug auf die Sensitivität folgende drei Bedingungen erfüllen: Es soll interindividuelle Fähigkeitsunterschiede widerspiegeln, es sollte für Schwierigkeitsunterschiede zwischen strukturell identischen Aufgaben sensitiv sein und Schwierigkeitsunterschiede zwischen strukturell unterschiedlichen Aufgaben abbilden.

Diagnostizität: Diese Dimension beschreibt die Fähigkeit einer Methode neben der kognitiven Beanspruchung auch denjenigen Aspekt in der Mensch-Maschine-Interaktion zu identifizieren, welcher für die festgestellte Beanspruchung verantwortlich ist. Der Grad, in dem ein Maß entsprechende Identifikationen ermöglicht, wird als Diagnostizität bezeichnet. Ein hoch diagnostisches Maß hat den Nachteil, dass es nicht

sensitiv ist. Die Diagnostizität kommt dadurch zustande, dass die Messung selektiv nur diejenigen Beanspruchungen abbildet, die durch bestimmte Aspekte der Aufgabe hervorgerufen werden. Es wird keine Information über die Beanspruchungshöhe anderer Aspekte der Aufgaben gegeben (Insensivität). Es hängt also von der Situation und dem Ziel der Untersuchung ab, welcher Dimension man die höhere Wichtigkeit zuschreibt.

Intrusion: Diese Dimension beschreibt die Tendenz, mit der die eingesetzte Methode die Merkmalsausprägung behindert oder stört. Bei Tests mit Zweitaufgaben lässt sich ein störender Einfluss der Zweitaufgabe auf die Erstaufgabe beispielweise nicht ausschließen.

Reaktionszeit: Anahnd dieser Dimension wird die Geschwindigkeit festgehalten, mit der sich eine Veränderung in der Beanspruchung auch in dem abgeleiteten Maß niederschlägt.

Benutzerakzeptanz: Diese Dimension beschreibt den Grad, mit dem eine Person die Erfassungsmethode als sinnvoll und nicht störend empfindet. Bei geringer Akzeptanz besteht die Gefahr, dass aufgrund geringer Motivation Messungen durch Verzerrungen unbrauchbar werden.

Erfassungsaufwand: Das Verhältnis von Kosten, Schwierigkeit der Datenauswertung und entsprechender Nutzen wird durch diese Dimension beschrieben.

Die erfassbaren reaktionszentrierten und personenabhängigen Merkmale mentaler Beanspruchung werden nach Facaoaru & Friedling (1985) in physiologische Parameter, Verhaltens- und Leistungsdaten und Erlebens- und Befindensdaten (Tab. 1) unterglie-

dert. Die entsprechenden Messmethoden werden analog hierzu von O'Donnell und Eggmeier (1986) in die drei großen Klassen subjektive Messungen, Leistungsmessungen und physiologische Messungen unterteilt.

4.3 Subjektive Messungen

Bei dieser Form der Datenerfassung wird der Operator nach der Art und Höhe seiner mentalen Beanspruchung befragt. Die Befragung kann sich dabei auf die Anforderungsstruktur oder die erlebten Folgen der Beanspruchung beziehen. Die Erfassung erfolgt üblicherweise in ein- oder mehrdimensionalen Urteilsskalen. Diese Skalen gelten als sensitiv in Bereichen außerhalb von Über- und Unterforderungen (Buld, 2000). Die Diagnostizität ist hingegen eingeschränkt. Mehrdimensionale Skalen (SWAT oder NASA TLX) können zwar beispielsweise körperliche von mentalen Bereichen der Beanspruchung trennen, eine Trennung von visueller Verarbeitung und zentraler Auslastung der Informationsverarbeitung leisten sie jedoch nicht.

Meistens werden die Befragungen nach der Aufgabenbearbeitung durchgeführt, was die Intrusion herabsetzt. Eine Befragung während der Aufgabenbearbeitung ist hingegen problematisch. Der Erfassungsaufwand und die Akzeptanz stellen in der Regel keine Probleme dar (Tsang & Wilson, 1997). Gravierende Mängel dieser Instrumente sind in den Bereichen Reaktionszeit und Validität festzustellen. Gopher und Donchin (1986) geben zu bedenken, dass Personen lediglich Dinge berichten können, die Ihnen zum Zeitpunkt der Befragung bewusst sind. Nach Tsang & Wilson (1997) kann es jedoch passieren, dass sich die Person entweder bei der Befragung nicht mehr an die Beanspru-

chung erinnern kann, oder dass sie ihr während der Aufgabenbearbeitung nicht bewusst geworden ist. Nach Wickens und Hollands (2000) ist ein Hauptfaktor bei der Beanspruchungseinschätzung die Aufgabenanzahl, die gleichzeitig bearbeitet werden soll, d.h. zwei leichte Aufgaben, die problemlos gleichzeitig bearbeitet werden können, werden als schwieriger eingeschätzt als eine schwere Aufgabe, deren Bearbeitung misslingt.

Eingesetzte Verfahren zur subjektiven Messung von mentaler Beanspruchung beruhen häufig auf dem Wokload-Konzept und greifen eher auf der aufgabenanalytischen Ebene. Eingesetzte Instrumente sind z.B.:

- NASA Task Load Index (Hart & Staveland, 1988)
- SEA-Skala (Eilers, Nachreiner & Hänecke, 1986)
- SWAT – Subjective Workload Assessment Technique (Reid & Shingledecker, 1981; Reid & Nygren, 1988)
- SWORD – Subjective Workload Dominance Technique (Vidulich, 1989)

Wie im vorhergehenden Teil bereits deutlich wurde, stellt die eingeschränkte Validität ein Problem beim Einsatz subjektiver Maße dar. Hauptargument gegen eine Verwendung unter Online-Bedingungen ist aber sicherlich die lange Reaktionszeit. Eine Befragung im Nachhinein ist natürlich per se als Offline-Instrument angelegt. Würde man nun diese Befragung zeitgleich zur Tätigkeit durchführen, so würde die Intrusion deutlich erhöht. Vergleichbar wäre diese Störquelle mit den methodischen Problemen des *Zweitufgabenmaß* (Kap. 4.4), wobei aufgrund mangelnder Einbettungsmöglichkeiten die Intrusion noch höher ausfallen würde. Auch die technische Umsetzung zur automatischen Generierung einer Beanspruchungsvariablen ist nur schwer vorstellbar.

Aus diesen Gründen ist eine Online-Erfassung subjektiver Maße im Bereich der Mensch-Maschine-Interaktion und insbesondere der Mensch-Computer-Interaktion kaum umsetzbar und methodisch äußerst unzureichend.

4.4 Verhaltens- und Leistungsmessung

Zur Bestimmung des *Workload* wird in dieser Klasse das Verhalten oder die Leistung bei der Aufgabenbearbeitung gemessen. Hierbei unterscheidet man Messungen ohne Modifikation der Aufgabensituation als Hauptaufgabenmaß und Messungen in Kombination mit zusätzlichen Aufgaben als Zweit-, Doppel- oder Nebenaufgabenmaß.

Nach Gopher und Donchin (1986) liegt dem Hauptaufgabenmaß die Annahme zugrunde, dass die Qualität der Aufgabenbewältigung ein direktes Maß für den aufgabeninduzierten *Workload* darstellt. Üblicherweise werden als Maße die Idealspur, Präzision, Frequenz, Reaktionszeiten und Anzahl von Fehlern herangezogen. Die Augenscheinvalidität der Methode ist sehr hoch, da der Zusammenhang von Leistung und Beanspruchung naheliegend erscheint. Allerdings ist es nicht immer möglich, den Einfluss physischer Beanspruchung auf das Verhalten von dem Einfluss mentaler Beanspruchung zu trennen. Verringerte Leistung kann auch in den Aufgabeneigenschaften begründet sein (Wickens & Hollands, 2000) oder auf Motivationsprobleme zurückgeführt werden (Gopher & Donchin, 1986). Das Hauptaufgabenmaß ist darüber hinaus nicht intrusiv, weswegen O'Donnel und Eggmeier (1986) eine hohe Benutzerakzeptanz vermuten. Der Erfassungsaufwand ist zwar höher als bei den subjektiven Maßen, jedoch deutlich geringer als der Erfassungsaufwand einiger physiologischer Messungen. Die Anwen-

dungsanforderungen sind mehr oder weniger komplex, dies hängt vom gegebenen Setting ab und kann von einfachen Messgeräten bis hin zu komplexen Laboreinrichtungen reichen.

Diesen Vorteilen steht jedoch der gravierende Nachteil der geringen Sensitivität gegenüber. Mentale Beanspruchung ist nicht durch einen funktionalen Zusammenhang zwischen den Eigenschaften des Mensch-Aufgaben-Systems und dem Systemverhalten beschreibbar. Eine solche Messung bleibt somit unvollständig. Hart und Wickens (1990) fanden Korrelationen von subjektiven Maßen und Leistungsmaßen der Hauptaufgabe lediglich im mittleren Bereich der Beanspruchung. Auch die Diagnostizität ist eher gering, da es sich hier um ein eher globales Maß handelt. Die Reaktionszeit dieser Methode ist sehr lang, da Beanspruchung nur mit der Rate gemessen werden kann, wie das Verhalten der Person auftritt. Es ist allerdings ein brauchbares Maß, wenn man befürchten muss, dass die Person ein subjektives Maß wegen bestimmter Interessen verzerren würde.

Da alle Ressourcen, die nicht von der Hauptaufgabe gebunden werden, theoretisch noch verfügbar sind, können diese für andere Aktivitäten eingesetzt werden. Das Nebenaufgabenmaß berechnet sich nach diesem Zusammenhang als Ausmaß der Ressourcen, die in der Nebenaufgabe gebunden werden. Aus der Leistung bei der Nebenaufgabe ergibt sich der Ressourcenverbrauch, der von der Hauptaufgabe erzeugt wird und somit lassen sich Aussagen über den *Workload* machen, der von der Hauptaufgabe verursacht wird. Sensitivität, Reaktionszeit und Diagnostizität sind bei dem Zweitaufgabenmaß deutlich besser als bei dem Hauptaufgabenmaß. Nach Tsang und Wilson (1997) besteht aber auch eine Gefahr darin, dass die falsche Auswahl der Zweitaufgabe eine andere Dimension der mentalen Beanspruchung betrifft und somit über eine hohe Bean-

spruchung durch die Erstaufgabe hinwegtäuscht. Es ist weiterhin nicht auszuschließen, dass Interaktionseffekte von Hauptaufgabe und Nebenaufgabe die Messung verzerren. Die Intrusion dieses Maßes ist besonders negativ hervorzuheben. Eine Ausnahme bilden die sogenannten eingebetteten Zweitaufgaben, welche von der Person nicht als „künstliche“ Zweitaufgaben aufgefasst werden (Mueller, Großmann-Hutter, Jameson, Rummer & Wittig, 2001).

Als Maß zur Online-Verarbeitung eignet sich das Nebenaufgabenmaß sehr schlecht. Die hohe Intrusion macht den Einsatz unter Online-Bedingungen in den allermeisten Fällen unmöglich, da es nicht akzeptabel ist, Leistungseinbußen bei der Hauptaufgabe durch die Bearbeitung der Zweitaufgabe zu riskieren.

4.5 Analytische Maße

Diese relativ neue Methode der Beanspruchungsmessung beruht auf einer Modellierung der Belastungssituation in Verbindung mit Annahmen über Eigenschaften des menschlichen Informationsverarbeitungsapparates. Aus diesem komplexen Modell werden Schätzungen der mentalen Beanspruchungshöhe bei der Nutzung von Mensch-Maschine-Systemen möglich. Diese analytischen Methoden gehen auf Zeitverlaufsanalysen zurück. Zur Kritik über Zeitverlaufsanalysen geben Wickens und Hollands (2000, Kap. 11) einen Überblick.

Neuere Verfahren begegnen dieser Kritik mit einer erweiterten Zeitverlaufsanalyse, wie z.B. der Task Analysis Workload – TAWL (Hamilton & Bierbaum, 1990), der W/INDEX (North, and Riley, 1989) und der TLAP - Time Line Analysis and Prediction (Parks & Boucek, 1989). Nach Sarno und Wickens (1995) sind die Ergebnisse dieser Verfahren den klassischen analytischen Verfahren deutlich überlegen.

Da die Eigenschaften des Mensch-Maschine-Systems besonders expliziert werden müssen und die Charakteristiken der menschlichen Informationsverarbeitung so genau wie möglich zu berücksichtigen sind, ist der Erfassungsaufwand für diese Maße sehr hoch (Tsang & Wilson, 1997). Die äußerst geringe Intrusion dürfte sich aber auch in einer sehr hohen Benutzerakzeptanz auswirken. Die Reaktionszeit ist durch die frei wählbaren Phasen der Einteilung fast beliebig zu verkürzen, bleibt aber trotzdem weit hinter denen physiologischer Messungen zurück. Sensitivität, Diagnostizität und Validität lassen sich derzeit noch nicht bestimmen.

Gegen eine Onlineerfassung dieser analytischen Maße spricht im Prinzip nichts. Zu bedenken ist jedoch, dass die Beurteilung der Beanspruchung ohne Rückgriff auf die stattfindende Interaktion zwischen Mensch und Maschine abgeleitet wird. Vorhersageungenauigkeiten können durch große interindividuelle Unterschiede von Personen entstehen, aber auch durch unerwartete Interaktionssituationen. Die Probleme bei der Messgenauigkeit sprechen somit eher gegen eine Verwendung als Onlinemaß.

4.6 Physiologische Maße

Die Abhängigkeit vegetativer Begleitphänomene der Informationsverarbeitung vom retikulären Aktivierungspegel ergibt sich aus der Kopplung der *formatio retikularis* an die vegetativen Zentren des Zwischenhirns (Hypothalamus) (Luczak, 1987). Diese Begleitphänomene sind z.B. in einer Veränderung der Herzfrequenz und Blutdruckreaktion und des Sympathikustonus zu sehen. Durch diese Kopplung wird im Organismus die Voraussetzung zur gesteigerten Arbeitsbereitschaft geschaffen (Ergotropie). Verbindungen der *formatio retikularis* zum Limbischen System sind auch für eine Veränderung der Affektlage und Emotionen mitverantwortlich.

Nach Haider (1969) ist die Voraussetzung und Begeleiterscheinung jeder verhaltensbezogenen Änderung eine Gesamterregung des Organismus. Hierbei lassen sich vier belastungsbezogene und hierarchisch aufgebaute Aktivierungsstadien voneinander trennen, nämlich das Schlaf-Wach-Kontinuum, der allgemeine Aktivierungsstatus (tonische Aktivierung, langsame Fluktuation), die phasische Aktivierung (schnelle Aktivierungsfluktuation) und eine vierte Ebene der Belastungsselektivität, die nach Art und Intensität einer Reizsituation allen Ebenen überlagert ist. Hieraus ergibt sich die besondere Ausgangswertproblematik, die im Ausgangswertgesetz beschrieben ist. Je stärker Organe aktiviert sind, desto stärker sprechen sie auf hemmende Reize an und desto geringer sprechen sie auf aktivierende Reize an. Veränderungswerte unterliegen somit einem systematischen Fehler und eine Baseline muss immer miterhoben werden (Buld, 2000). Somit ergibt sich ein besonderes Problem bei der Operationalisierung psychophysischer Beanspruchung. Nach Luczak (1987) entzieht sich der Beitrag einer bestimmten Messgröße zur Erkennung und Bewertung der psychophysischen Beanspruchung einer auf

allgemein anerkannten und detaillierten Theorien basierten Parameterisierung. Analysen psychophysiologischer Beanspruchung sind in der praktischen Anwendung auf die operationale Definition des unmittelbar nicht quantifizierbaren Indikandum angewiesen.

Nach O'Donnel et al. (1986) stellen physiologische Reaktionen in Zusammenhang mit der Aufgabenausführung einen hoch sensitiven Workload-Parameter dar. Von diagnostischem Wert sind jedoch nur einige Verfahren, wie z.B. ereigniskorrelierte EEG-Potenziale. Je nach technischem Ableitungsverfahren ist der Aufwand der Ableitung von physiologischen Parametern recht hoch und oftmals der Akzeptanz abträglich, sofern die Verfahren intrusiv sind. Hier ist die Pupillometrie als einziges Verfahren von geringer Intrusion herauszustellen. Im Gegensatz zu diskret erhobenen subjektiven Maßen und bedingt kontinuierlich erhobenen Leistungsmaßen können physiologische Maße sehr gut kontinuierlich dargestellt werden (Wickens & Hollands, 2000).

4.6.1 Physiologische Messverfahren

Wie in den vorangegangenen Kapiteln ausgeführt wurde, stehen das Aktivierungspotenzial des Organismus und die mentale Beanspruchung in einem direkten Zusammenhang. Einige der gebräuchlichsten physiologischen Maße, die mit dem Aktivierungszustand korrelieren, werden nach Rösler (2001) den folgenden Bereichen zugeordnet:

- Hirnaktivität,
- Herz-Kreislaufaktivität,
- Respiratorische Psychophysiologie,

- Dermale Aktivität,
- Elektromyographische Aktivität,
- Okulomotorische Aktivität.

Zu diesen Hauptbereichen wurden im Laufe der Zeit zahlreiche Messmethoden entwickelt, die gängigsten sollen im Folgenden kurz erläutert und auf ihre Tauglichkeit zum Einsatz in einer bioanalogen Computerschnittstelle diskutiert werden.

4.6.2 Gängige Verfahren, Vor- und Nachteile

Die gängigsten Verfahren zur Ableitung von Messwerten in der Biopsychologie sind:

- Elektroenzephalogramm (EEG) für die Hirnaktivität;
- Elektrokardiogramm (EKG) für die Herz-Kreislaufaktivität;
- Atemfrequenz- und Atemvolumenmessung für die respiratorische Psychophysiologie;
- Elektrodermale Aktivität (EDA) zur Messung der Haut-Aktivität;
- Elektromyographie (EMG) zur Messung der Muskel-Aktivität;
- Pupillometrie, Lidschlag-, Fixations- und Sakkadenmessung zur Analyse der okulomotorischen Aktivität.

Im Hinblick auf das Ziel der vorliegenden Arbeit - der Entwicklung einer biopsychologischen Mensch-Computer-Schnittstelle - müssen sämtliche Verfahren zunächst auf die technischen Realisierungsmöglichkeiten der Online-Signableitung geprüft werden. Hierbei fallen diejenigen Methoden weg, deren Ableitungstechnik invasiven Charakter

haben. Die Verkabelung von Probanden ist in einer Laborsituation vertretbar, denkt man jedoch über die Laborphase hinaus, so ist einer non-invasiven Methode der Vorzug zu geben. Die Methode sollte den Probanden möglichst wenig belasten oder ablenken (geringe Intrusion).

Unter dieser Prämisse bleibt aus den aufgeführten Bereichen der biopsychologischen Verfahren lediglich die Messung der okulomotorischen Aktivität übrig. Die Messung des Lidschlags hat jedoch bisher für den Bereich der mentalen Aktivität zu keinen eindeutigen Ergebnissen führen können (Andrés, 1996; Birbaumer & Pauli, 2000; Galley, 1998; Lang et al. 1990).

Nach Galley (2001) dient die Sakkade, als schnelle Augenbewegung, dem schnellen Auffinden interessierender Objekte im Gesichtsfeld. Die Geschwindigkeit der Sakkade ist entsprechend dem Systemzustand immer maximal und beträgt bis zu 520°/sec. Anhand der Sakkadengeschwindigkeit kann Leistung und Aktivierung operationalisiert werden, jedoch besteht hierbei eine weitgehende Nichtübereinstimmung bezüglich Normwerten und der Bewertung von gemessenen Werten. Die Messung der Sakkadenlatenz und Fixationsdauer - als Indikator für den Schwierigkeitsgrad einer Aufgabe - scheint auch eine interessante Methode darzustellen. Galley bemerkt hierzu jedoch, dass die Erfassung der Sakkadenlatenz erforderlich macht, dass ein Zeitsignal für den Beginn der Reizexposition und die Sakkade in ausreichender Abtastrate erfasst werden, wobei ein Videotakt von 20ms zu langsam erscheint (2001). Dieses Gebiet ist jedoch in Zukunft bei der Verwendung von Hochauflösungskameras für weitere Forschungsvorhaben nicht uninteressant.

Als letztes ableitbares Biosignal verbleibt also die Pupillenmessung, deren Daten bequem und körperlos mit einer Infrarot-Kamera erfasst werden können. Im Gegensatz zu subjektiven Maßen und Zweitaufgabenmaßen beeinträchtigt die Pupillenmessung die Bearbeitung der Aufgabe nicht und sie ist auch verfügbar, wenn der Operator kein beobachtbares Verhalten zeigt. Es lassen sich darüber hinaus mentale Beanspruchungszustände differenzieren, die nicht zu einer Leistungsminderung in der Aufgabe führen. Diese Vorzüge gelten zwar für eine Reihe von physiologischen Parametern, die Pupille hebt sich aber dadurch ab, dass sie schneller reagiert, geringere Latenzzeiten aufweist, körperlos ableitbar ist und die größte Sensitivität besitzt (O'Donnell et al., 1986).

Mit der extrem hohen Sensitivität ist jedoch auch ein Nachteil verbunden, der die Validität herabsetzt. Die Pupille reagiert nicht nur auf die mentale Beanspruchung sondern auch auf andere Reize, wie z.B. den Lichtreflex. Der Wert der Pupillenanalyse wird also durch die Kontrollierbarkeit von Störreizen bedingt. Nach Luczak (1987) ist die Pupillometrie hinsichtlich einer Indikation psychophysischer Beanspruchung beachtlich valide und reliabel. Luczak bezeichnet die physiologischen Beanspruchungsindikatoren durch ihre Kopplung an die Retikularformation allgemein als konstruktvalide und augenscheinvalide. Zudem wird die hohe Durchführungs- und Auswertungsobjektivität hervorgehoben.