

geführt. Aufgrund der Ergebnisse dieses Experiments konnte ein Verfahren entwickelt werden, mit dem die Artefakte automatisch erkannt und eliminiert werden können. Dieses Experiment 1 wird im folgenden Kapitel dargestellt.

### **3. Experimente**

#### **3.1 Experiment 1: Versuch zur Identifizierung von Artefakten**

Zur Entwicklung eines Verfahrens für die Verbesserung der Datenbasis war eine sorgfältige Analyse dieser Datenbasis notwendig. Um Artefakte aus den Rohdaten ausschließen zu können, mussten diese zuerst identifiziert werden. Bei der Vermessung der Pupille mit der Infrarotkamera RED II wird der dunkelste Punkt des aufgezeichneten Bildes als Pupille definiert und vermessen. Wird das Auge bei einem Lidschlag geschlossen, so ist die Pupille für die Kamera nicht mehr sichtbar, da sie vom Augenlid verdeckt wird. In diesem Moment können zwei verschiedene Messfehler entstehen. Bei Personen mit dunklen Wimpern oder mit geschminkten Wimpern formen diese sich beim geschlossenen Auge zu einer Linie, die der Färbung der Pupille entspricht. Da die Messung auf der horizontalen Ausdehnung der Pupille basiert, wird der gemessene Wert sehr hoch und verzerrt den Mittelwert nach oben. Der umgekehrte Fall tritt bei Personen ein, die sehr helle Wimpern haben. Wenn diese Personen die Augen schließen, gibt es keine dunkle Stelle in der Augenaufzeichnung, die unter dem eingestellten Schwellwert liegt. Der Messwert wird sich also Null annähern und der Mittelwert wird in diesem Fall nach unten verzerrt. Da keine anderen Fälle auftreten können, erfolgt eine Verzerrung des Mittelwerts bei einer einzelnen Person immer in eine bestimmte Richtung.

Nach Bortz und Döring gehören diese Phänomene zu Artefakten physiologischer Herkunft. Um die Validität der Messung zu steigern, sollte diesen Artefakten mit geeigneten Mitteln begegnet werden (Bortz & Döring, 1995). Da der Lidschlag in einigen Untersuchungen als Indikator von Aktivierungsprozessen identifiziert wurde (Andrés, 1996; Galley, 1998), ist mit einer höheren Verzerrung der Messwerte durch Artefakte zu rechnen, wenn Versuchspersonen schwierige Aufgaben lösen, als dies bei einfachen Aufgaben der Fall ist. Um diesem Messfehler zu begegnen, mussten diese Artefakte identifiziert und von der weiteren Datenverarbeitung ausgeschlossen werden.

Das hier beschriebene Experiment 1 diente der Identifizierung solcher Artefakte. Durch das künstliche Herstellen dieser Artefakte wurde geprüft, ob es eine Möglichkeit gibt, diese Artefakte durch einen einfachen Algorithmus auszuschließen. Die Grundannahme lautete hierbei, dass der Wert eines Artefakts außerhalb des Bereichs liegt, der durch die zweifache Standardabweichung über und unter dem Mittelwert einer Messung definiert ist. Zur Überprüfung dieser Hypothese wurde den Versuchspersonen zuerst für 5 Sekunden ein weißer Monitor gezeigt, daraufhin für weitere 5 Sekunden ein schwarzer Monitor. In den folgenden 5 Sekunden wurden sie aufgefordert die Augen geschlossen zu halten und danach wieder zu öffnen. Dieser Vorgang wurde so oft wiederholt, bis von jeder Versuchsperson 4 Sets mit geschlossenen Augen, 4 Sets mit geöffneten Augen, ein Set mit einem weißen Monitor und ein Set mit einem schwarzen Monitor erzeugt wurden.

Im Verlauf des Experiments wurde nun untersucht, ob die Mittelwerte der Sets, bei denen die Augen geschlossen waren, über oder unter dem Bereich liegen, der durch zwei Standardabweichungen vom Mittelwert definiert ist. Der zugrunde liegende Mittelwert und die resultierende Standardabweichung ergaben sich aus den Mittelwerten der Pupillengröße bei der Betrachtung des Monitors mit einem weißen und des Monitors mit einem schwarzen Bild. Es

wurde hierbei vorausgesetzt, dass im vorhandenen Setting der größte Bereich der lichtbedingten Pupillenbewegung von einem Monitor mit einem weißem und mit einem schwarzen Bild hervorrufen wird.

In diesem Experiment wurden 12 Versuchspersonen untersucht, dabei wurden für jede Person 4 Sets mit geschlossenen und weitere 6 Sets mit geöffneten Augen erzeugt. Die Versuchspersonen waren zwischen 22 und 29 Jahren alt ( $M = 24,58$  Jahre). 9 Versuchspersonen waren weiblich und 3 Versuchspersonen waren männlich.

### **3.1.1 Hypothesen**

Die Hypothese des Experiments 1 lautet: Die Mittelwerte ( $\mu_Z$ ) der Messdaten, bei denen die Versuchspersonen die Augen geschlossen haben, liegen außerhalb des Wertebereichs, der durch zwei Standardabweichungen (SD) vom Mittelwert der Pupillenweite ( $\mu_{SW}$ ) definiert ist. Dieser Mittelwert ( $\mu_{SW}$ ) ergibt sich aus den mittleren Pupillengrößen bei schwarzem und weißem Monitorbild. Die Standardabweichung der lichtbedingten Pupillenreaktion wird dadurch maximal gehalten.

Die statistische Hypothese lautet:

$$H_1: \mu_Z \neq \mu_{SW} \pm 2 \text{ SD}$$

$$H_0: \mu_Z = \mu_{SW} \pm 2 \text{ SD}$$

Die verwendeten Variablen sind:

- Zustand des Auges (offen oder geschlossen)
- Monitorabstrahlung (schwarz und weiß)
- gemessene Pupillengröße (abhängige Variable)

### ***3.1.2 Versuchsdurchführung***

Für die Beantwortung der Fragestellung, ob die Mittelwerte der Messdaten außerhalb des Bereichs des Mittelwerts  $\bar{x} \pm 2 \text{ SD}$  liegen, war es notwendig, alle Versuchspersonen einzeln zu analysieren, da eine Zusammenfassung der Messdaten über Mittelwerte die Varianzen nicht mehr berücksichtigt. Um die Hypothese  $H_0$  zu verwerfen, ist hierbei von einem Signifikanzniveau  $p < .01$  auszugehen. Neben der Prüfung, ob die Daten bei geschlossenen Augen außerhalb des Bereichs von  $2\text{SD}$  liegen, wurde des Weiteren geprüft, ob die Messdaten sich bei geöffneten Augen innerhalb dieses Bereichs befinden. Sämtliche Messwerte, die bei geöffneten Augen erzeugt wurden, fielen in diesen Bereich der  $2\text{SD}$ . Die Einzelpersonenauswertungen sind im Anhang zu finden.

Im ersten Schritt wurden pro Versuchsperson der Mittelwert und die Standardabweichung aus der Pupillenmessung von 5 Sekunden mit geöffneten Augen vor dem schwarzen Monitor und 5 Sekunden vor dem weißen Monitor ermittelt. Hieraus wurden dann die Prüfgrößen Mittelwert +  $2\text{SD}$  und Mittelwert –  $2\text{SD}$  ermittelt. Die Werte, welche in den Sets mit geschlossenen Augen gemessen wurden, wurden daraufhin mit dem Independent-Sample T-Test geprüft.

### 3.1.3 Ergebnisse

Die Ergebnisse des Experiments zeigt die Tabelle 4 in der Übersicht. Bei 12 Versuchspersonen gab es lediglich einen Fall, der die Hypothese  $H_0$  stützt. Hier lag der Mittelwert der Messung bei geschlossenem Auge näher als zwei Standardabweichungen an dem Mittelwert der Pupillengröße des geöffneten Auges.

**Tabelle 4: Ergebnisse des T-Tests**

VPN	$M_{vp_i} \neq M_{sw} \pm 2s$	Signifikanz
VP001	über	$p = .000$
VP002	unter	$p = .000$
VP003	unter	$p = .000$
VP004	über	$p = .000$
VP005	über	$p = .000$
VP006	unter	$p = .000$
VP007	über	$p = .000$
VP008	über	$p = .000$
VP009	über	$p = .000$
VP010	$= M_{sw} \pm 2s$	
VP011	über	$p = .000$
VP012	über	$p = .000$

Zur statistischen Absicherung wurde ein Chi-Quadrat-Test durchgeführt, durch den die Verteilung der Messergebnisse überprüft wurde. Die Anzahl der Sets, bei denen die Artefakte außerhalb des Wertebereichs von zwei Standardabweichungen lagen, wurde mit dem Prädikat *trifft zu* versehen. Hiernach ergaben sich 44 Sets, die die Alternativhypothese stützen und lediglich 4 Sets, welche die Nullhypothese bestätigen. Das Ergebnis der statistischen Prüfung zeigt eine hoch signifikante Absicherung der gefundenen Ergebnisse gegen eine zufällige Verteilung. Die Prüfung mittels Chi<sup>2</sup>-Test ergab ein  $\chi^2(1, N = 48) = 33.33, p < .00$ .

Als Ergebnis des Experiments 1 musste die Hypothese H0 verworfen werden. Die Hypothese H1 wurde somit bestätigt. Die Annahme, dass Werte über oder unter zwei Standardabweichungen des Mittelwerts einer Messung als Artefakte anzusehen sind wird somit angenommen und wurde für die alle weiteren Messungen verwendet.

Es wurde aber auch deutlich, dass Werte innerhalb des beschriebenen Bereichs liegen können, die als Artefakte anzusehen sind. Da diese Artefakte durch ihre Nähe zum Mittelwert diesen nicht übermäßig verzerren, können diese in der Datenbasis verbleiben.

### ***3.1.4 Schlussfolgerungen aus dem Experiment 1***

Das Experiment 1 hat gezeigt, dass Werte, die zwei Standardabweichungen von einem Mittelwert einer Messung entfernt liegen (positiv oder negativ), als Lidschlag zu interpretieren sind. Wurden diese Werte beibehalten, so kam es zu einer Verzerrung des Mittelwerts der Pupillengröße durch den Lidschluss. Da diese Artefakte aber nicht gleichmäßig auftraten, ist

nicht davon auszugehen, dass sie sich entweder aufheben oder sämtliche Messwerte in gleicher Weise verzerren. Durch den Ausschluss dieser Werte kam es somit in jedem Fall zu einer Verbesserung der Datenbasis durch die Verringerung des Standardmessfehlers.

Wie in dem hier beschriebenen Experiment 1 gezeigt wurde, ist diese Methode der Identifizierung und des Ausschlusses von Lidschlägen aber nicht in jedem Fall sicher. So war im Datensatz von Versuchsperson 10 der Messwert bei geschlossenem Auge nur ca. eine Standardabweichung vom Mittelwert entfernt.

Trotz dieses Falls wurde an dem beschriebenen Vorgehen aus zwei Gründen festgehalten. Zum einen ist die Auftretenswahrscheinlichkeit eines solchen Falls relativ gering und zum anderen ist die Verzerrung der Datenbasis durch einen Wert im Bereich der Standardabweichung nicht erheblich und kann somit vernachlässigt werden.

In der Auswertung der folgenden Experimente hat sich das Verfahren zur Identifizierung der Artefakte bewährt. Durch die zeitgleiche Aufzeichnung der Blickrichtung stand ein weiteres Kontrollinstrument zur Identifizierung von Artefakten zur Verfügung. Sobald die Pupille von der Kontrolleinheit nicht mehr erkannt wurde, war auch die Blickaufzeichnung unterbrochen. In keinem Fall wurden Messdaten als Artefakte identifiziert, bei denen nicht auch die Blickrichtungsaufzeichnungen unterbrochen waren. Insofern ist es ausgeschlossen, dass gültige Messdaten versehentlich als Artefakte ausgeschlossen werden.

### ***3.1.5 Verfahren der Datenbereinigung***

Zur weiteren Verarbeitung der Daten ist es aus den genannten Gründen notwendig, diese von Artefakten zu bereinigen. Zu Beginn der Untersuchungen wurden diese Bereinigungen noch per Hand durchgeführt. Es stellte sich jedoch bald heraus, dass diese Arbeitsleistung bei der Datenflut von ca. 40.000 Messwerten pro VP nicht mehr zu leisten war. Die Lösung dieses Problems lag in der Automatisierung der Datenbereinigung. Hierzu entwickelte der Autor ein Excel-Makro, das in einem Drei-Schritt-Verfahren die Bereinigung anhand eines einfachen Algorithmus bewerkstelligen sollte.

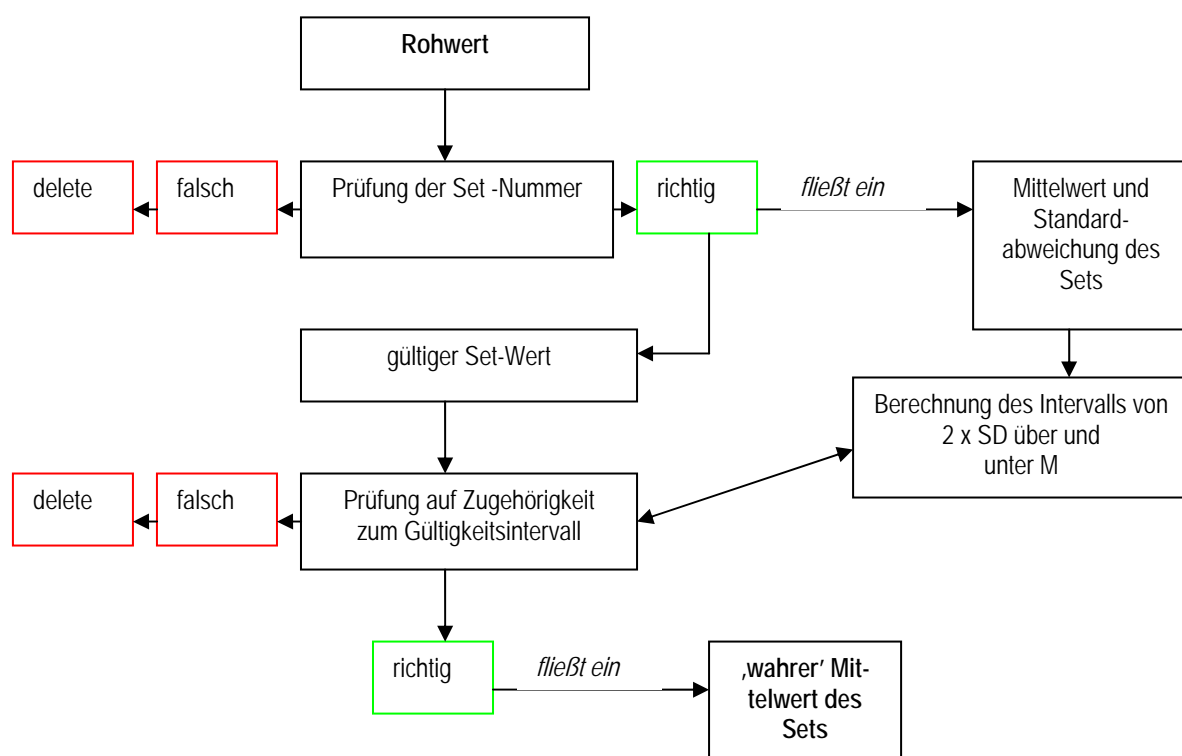
Zuerst wurde der Mittelwert der Daten für ein bestimmtes Intervall (Set) und die Standardabweichung berechnet. War der gemessene Einzelwert größer oder kleiner als die zweifache Standardabweichung des Intervalls, so wurde dieser Wert ignoriert. Die verbleibenden gültigen Werte standen somit zu einer weiteren Berechnung zur Verfügung. Die folgende Grafik verdeutlicht die Arbeitsweise der automatisierten Datenbereinigung.

Die manuelle Überprüfung von 50 Datensätzen ergab, dass 100 Prozent der automatisch gefundenen und eliminierten Werte Artefakte waren. Wie viele Artefakte nicht identifiziert wurden, wurde nicht untersucht.

Die Automatisierung dieses Prozesses begann mit dem Einlesen des Datensatzes aus der Rohdatendatei in eine Excel-Tabelle. Jeder Wert wurde nun zuerst auf seine Zugehörigkeit zu dem betreffenden Intervall (Set) geprüft. Fiel diese Prüfung positiv aus, so wurde der Wert für dieses Intervall gültig. Aus allen intervallgültigen Werten wurden dann Mittelwert und doppelte Standardabweichung berechnet. Durch eine Zwischenrechnung mit einer Zwischenvariablen, die nicht kleiner als Null werden kann, wurde verhindert, dass die doppelte Standard-



abweichung einseitig in den negativen Bereich abrutschen kann. Die intervallgültigen Werte wurden daraufhin geprüft, ob sie sich im Rahmen der doppelten Standardabweichung über oder unter dem Mittelwert befanden. Fielen die Werte aus diesem Raster, so wurden sie mit ‚delete‘ gekennzeichnet und in der weiteren Rechnung ignoriert. Aus den verbleibenden gültigen Werten wurde dann der ‚wahre‘ Mittelwert berechnet. Dieser ‚wahre‘ Mittelwert wich in extremen Fällen bis zu 20% von dem der Rohdaten ab.



**Abbildung 11: Algorithmus zur Selektion gültiger Messwerte**

### ***3.1.6 Zusammenfassung des Experiments***

Im Laufe zahlreicher Messungen fielen immer wieder Werte auf, die durch ihre extremen Abweichungen vom Mittelwert herausragten. Diese Werte konnten das Vielfache des Mittelwerts oder aber auch einfach Null annehmen. In dem zuvor beschriebenen Experiment 1 wurde der Fragestellung nachgegangen, ob diese extremen Werte, die sich nicht mit der Physiologie der Pupille in Einklang bringen lassen, als Artefakte zu betrachten sind, die durch Lidschläge hervorgerufen werden: In dem Moment, wo das Auge geschlossen war, wurden statt der Pupille die Wimpern oder überhaupt nichts vermessen.

In einem experimentellen Setting wurden diese Artefakte künstlich produziert und gemessen. Das Ergebnis der Auswertung bestätigte, dass diese Ausreißerwerte tatsächlich von geschlossenen Augen hervorgerufen werden, und zwar entweder als Nullwerte oder besonders geringe Werte bei hellen Wimpern von Versuchspersonen, oder als extrem hohe Werte, im Fall von dunklen oder geschminkten Wimpern bei Probanden.

Weiterhin konnte in dem Experiment gezeigt werden, dass eine Datenbereinigung automatisiert stattfinden kann, wenn man den Mittelwert einer Messung als Grundlage nimmt, die zweifache Standardabweichung addiert und subtrahiert und alle Werte eliminiert, die sich nicht in diesem Bereich befinden.

Es konnte aber auch gezeigt werden, dass durch diese Prozedur nicht alle Artefakte identifiziert werden konnten. In Einzelfällen traten Artefakte auf, die einen Wert in der Nähe des Mittelwerts der Messung annahmen. Es ist aber dennoch sicher, dass die Datenbasis verbes-

sert werden kann, wenn die identifizierten Artefakte eliminiert werden. Andererseits verzerren Artefakte, die mit diesem Verfahren nicht aufgespürt werden, aufgrund ihrer Nähe zum Mittelwert die Datenbasis auch nicht wesentlich.

### 3.2 Experiment 2: Herleitung der Funktion für die lichtbedingte Pupillenreaktion

In den vorhergehenden Kapiteln wurde bereits auf das Problem der Monitorstrahlung als Störvariable eingegangen. Die Störvariable Licht stellte bisher jeden Forscher, der auf dem Gebiet der Pupillometrie tätig war, vor große Probleme, entsprechend häufig findet sich dieses Argument in der Kritik zu verschiedenen Arbeiten (vgl. Galley, 2001). Die Messung der Pupillenbewegung an einem Bildschirmarbeitsplatz scheint aufgrund der Monitorstrahlung sogar eher abwegig. Wahrscheinlich liegt hier auch der Grund, warum bisher kein Forscher die Pupillenmessung am unkontrollierten Monitor erfolgreich durchführen konnte.

Das nachfolgend beschriebene Experiment hat einen zentralen Stellenwert für das gesamte Untersuchungsvorhaben. Ziel des Experiments war die Messung der Pupillenreaktion, die direkt auf die Monitorstrahlung zurückzuführen ist. Die hierbei ermittelte Pupillenreaktion soll im weiteren Verlauf dieser Arbeit dazu genutzt werden, die erhobenen Pupillendaten, die durch die Monitorabstrahlung beeinflusst wurden, zu bereinigen. Da es keine Interaktion zwischen der Helligkeit und mentalen Prozessen gibt, wie dies durch Rößger (1993, 1997) nachgewiesen wurde, kann diese Art der Bereinigung in den behandelten Fällen angewendet werden. Die Formel zur Datenbereinigung setzt sich aus den drei relevanten Werten gemessener Pupillendurchmesser ( $PD_{ges}$ ), lichtbedingter Pupillendurchmesser ( $PD_{licht}$ ) und mental bedingter Pupillendurchmesser ( $PD_{men}$ ) zusammen. Der enthaltene Messfehler bleibt hierbei unverändert. Diese Werte werden in der folgenden Formel gegeneinander berechnet:

$$\text{Formel 4: } PD_{ges} - PD_{licht} = PD_{men} + PD_{fehler}$$

Es liegen zwar vereinzelt bereits Versuche vor, das Licht nicht nur zu kontrollieren, sondern auch als Variable zu registrieren und dann zur Datenbereinigung heranzuziehen (Rößger, 1997), doch scheinen Grundannahmen des Zusammenhangs von Licht und Pupillenreaktion nur begrenzt berücksichtigt. Rößger (1997) bezieht sich hier beispielsweise auf Boff und Lincoln (1988) und postuliert einen linearen Zusammenhang von Licht und Pupillengröße.

Da in den folgenden Experimenten das Licht als Monitorstrahlung großen Einfluss auf die Pupille nehmen wird, scheint eine gründliche Betrachtung der Zusammenhänge angebracht. Das folgende Experiment dient der Herleitung einer Funktion, die später zur Bereinigung der Pupillendaten um den Faktor Licht dienen soll. Da das Außenlicht in der Laborsituation konstant gehalten werden kann, stellt die Monitorstrahlung die einzige intervenierende Variable aus diesem Bereich dar. Es ist davon auszugehen, dass die lichtbedingte Pupillenreaktion nicht linear verläuft.

### ***3.2.1 Hypothese***

Die Grundannahme, die mit dem hier beschriebenen Experiment bestätigt werden soll, ist die Existenz eines exponentiellen Zusammenhangs zwischen der Lichtstärke und der Pupillenbewegung. Hierbei ist davon auszugehen, dass die Exponentialfunktion lichtbedingte Varianzen in der Pupillenbewegung besser aufklären kann, als eine lineare Funktion.

### **3.2.2 Stichprobenbeschreibung**

An dem Experiment 2 nahmen  $N=55$  Versuchspersonen teil. Die Stichprobe unterteilte sich in 37 (67,3 Prozent) weibliche und 18 (32,7 Prozent) männliche Probanden. Das Alter der Probanden reichte von 17 Jahren bis 60 Jahren, wobei der Mittelwert zum Zeitpunkt der Untersuchung 24,1 Jahre betrug. 18 Probanden (32,7 Prozent) gaben an Brillenträger zu sein. Eine Probandin gab außerdem an, morphiumhaltige Medikamente eingenommen zu haben. Da aber dieser Datensatz durchaus im Bereich der vergleichbaren Werte lag, wurde er bei der Datenanalyse nicht ausgeschlossen.

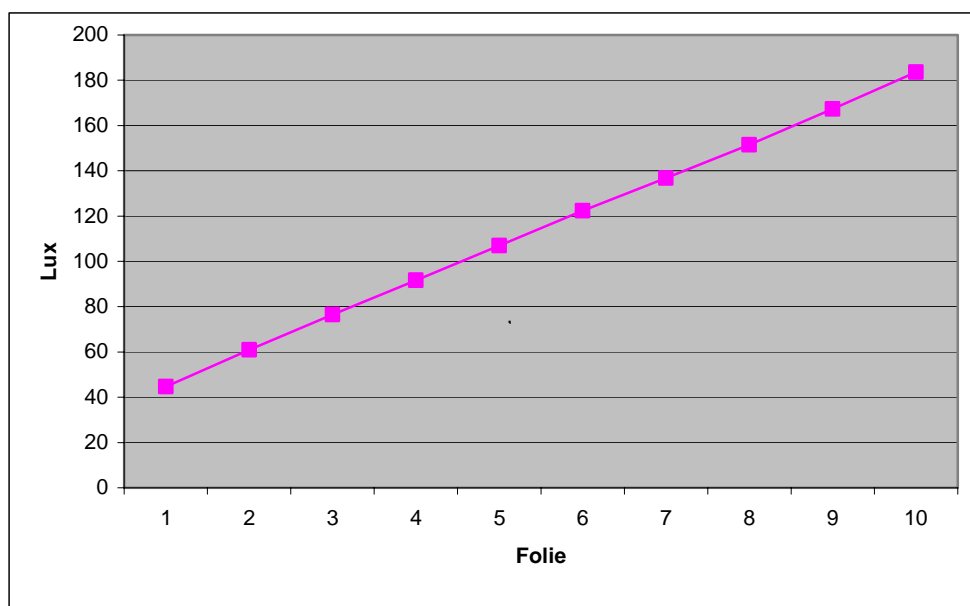
### **3.2.3 Versuchsdurchführung**

Für das Experiment zur Ermittlung der Lichtreaktion der Pupille auf Monitorstrahlung wurden 10 Bilder erstellt, die den Probanden nacheinander präsentiert wurden. Die 10 Stimulusbilder unterschieden sich in ihrem Gehalt von Grauanteilen und reichten von weißen Folien bis zu totalen Schwarzfolien. Da nach Miyao et al. (1991) Farben am Computermonitor nicht über die gemessenen Lux-Werte hinaus auf die Pupillenbewegung wirken, ist die Verwendung von Grauabstufungen in diesem Experiment als ausreichend anzusehen. Die Folien wurden jeweils mit einem Fixationskreuz in der Folienmitte versehen, um den Probanden einen visuellen Haltspunkt zu geben. Ausgehend von einer Folie mit 100 Prozent Grauanteil (schwarz) wurde der Deckungsgrad mittels Grauanteilveränderung mit dem Grafikprogramm ADOBE Photoshop solange vermindert, bis der korrekte LUX-Wert der linearen Verteilung erreicht war. So entstanden 10 Folien, deren Leuchtabstand der in Diagramm 1 dargestellten gradlinigen Funk-

tion folgte. Die Lux-Werte wurden mit dem Testo 545 Luxmeter erfasst. Der Monitor wurde vor Beginn der Prozedur kalibriert, damit die Farbtemperatur konstant gehalten werden konnte. Die Grauwerte verteilten sich wie in Tabelle 5 dargestellt.

**Tabelle 5: Grauanteile der Bildschirmvorlagen**

<i>Folien</i>	Folie	Folie	Folie	Folie	Folie	Folie	Folie	Folie	Folie	Folie
	1	2	3	4	5	6	7	8	9	10
<i>Grauanteil</i>	100%	85%	73%	64%	57%	48%	32%	20%	13%	0%
<i>Lux</i>	44,7	61	76,5	91,7	107	122,4	136,8	151,5	167,3	183,6



**Diagramm 1: Funktion der Lichtabstrahlung der Bildschirmvorlagen**

Durch die Messung der Lichtabstrahlung der verschiedenen Folien konnte ein lineares Abfallen des Faktors Licht gewährleistet werden. Um Störeffekte zu minimieren, wurde ein Vorexperiment durchgeführt, in dem sich die Probanden an das Setting gewöhnen konnten.

Jede der 10 Folien wurde für den jeweiligen Probanden 5 Sekunden angezeigt. Während dieser Zeit wurde die Pupillengröße vermessen. Da die Lichtreaktion der Pupille zwischen 250 ms und 450 ms beträgt, sind 5 Sekunden als Reaktionszeit ausreichend. Die Folienpräsentation wurde mit dem Programm *Presentation* der Firma *Neurobehavioral Systems* durchgeführt. Dieses Programm unterscheidet sich von vergleichbaren Programmen durch die besonders genaue zeitliche Steuerung der Stimuli. Diese erfolgt über ein PCL-Skript, welches die Daten zuerst in den Cache des Computers lädt und daraufhin millisekundengenau anzeigt. Bei einem Wechsel der Folie wird ein Stromimpuls von 5 Volt für 55 Millisekunden an den Parallelport gesendet. Dieser Port ist entsprechend mit dem Port des Analyserechners verbunden, der den Stromimpuls aufnimmt und registriert. In der Datenauswertung kann somit auf wenige hundertstel Millisekunden genau die Zeit des Folienwechsels festgestellt und dokumentiert werden.

Die erfassten Datensätze waren nun darauf zu prüfen, ob in der Pupillenbewegung eine Exponentialverteilung oder eine stetig lineare Verteilung vorliegt. Zu diesem Zweck wurden die Daten mit einem Plot-Programm unter Verwendung des Non-Linear-Least-Square-Fit-Test sowie durch die SPSS-Erweiterung Curve Estimation analysiert. Dabei werden die Daten iterativ auf eine zu bestimmte Funktion geprüft. Die erhaltene Funktion wird dann wiederum auf die Ausgangsdatsätze angelegt, um die Abweichung zu prüfen, mit der die Messwerte zu dieser Funktion passen. Eine totale Übereinstimmung kann hierbei aber niemals erreicht werden, da die Pupille permanenten Spontanfluktuationen (Galley, 2001) unterliegt. Diese Bewegungen von 0,6 bis zu 1,2 Hz haben weder lichtbedingte Ursachen, noch sind die Ursachen in

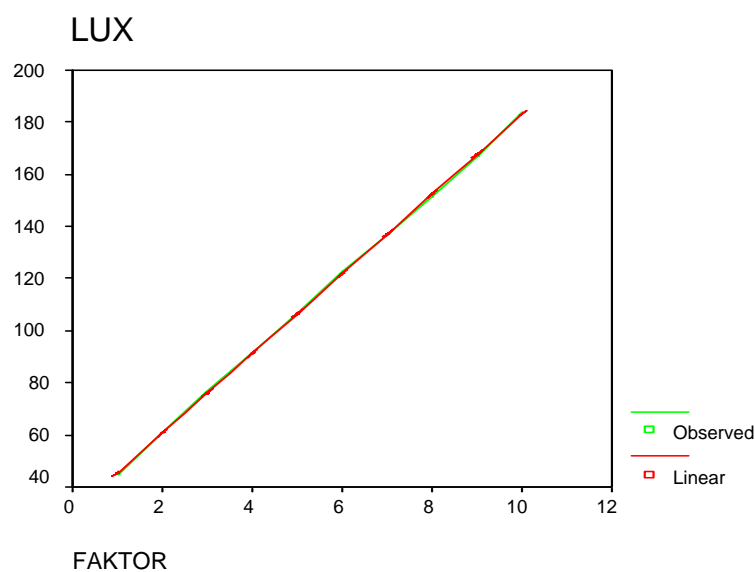


den hier untersuchten mentalen Vorgängen zu finden. Da diese Bewegungen nicht kontrolliert werden können, sind sie als Störvariable in sämtlichen Messdaten vorhanden.

Als abhängige Variable wurden neben den gemessenen Pupillenrohdaten auch die harmonisierten (Kapitel 2.2 Datenaufbereitung durch Harmonisierung) und die standardisierten Messdaten (Kapitel 2.3 Datenaufbereitung durch Standardisierung) verwendet. Die so ermittelten Funktionen finden sich im Anhang.

Gemäß gängiger Literaturmeinung (Rößger, 1997) wäre nun also eine Pupillenreaktion zu erwarten, die der Lichtfunktion geradlinig folgt. Hierbei ist die ermittelte Lichtfunktion:

$$\text{Formel 5: } y = 869,161 + (-1,6987 \times lux)$$



**Diagramm 2: Funktion der Lux-Werte über den Faktor Graustufen**

### ***3.2.4 Ergebnisse der Prüfung der Voraussetzungen***

Die Auswertung der Messdaten erfolgte in drei Schritten. Zuerst wurden die Messdaten von Artefakten befreit, wie dies in Kapitel 3.1 Experiment 1: Versuch zur Identifizierung von Artefakten) beschrieben ist.

In einem zweiten Schritt wurde überprüft, ob die Annahme richtig ist, dass bei inhaltslosen Folien keine bedeutsamen mentalen Pupilleneffekte zu erwarten sind. Wäre diese Annahme korrekt, dürften sich die Mittelwerte der Pupillendurchmesser und die Minimalwerte der Pupillendurchmesser nicht wesentlich unterscheiden.

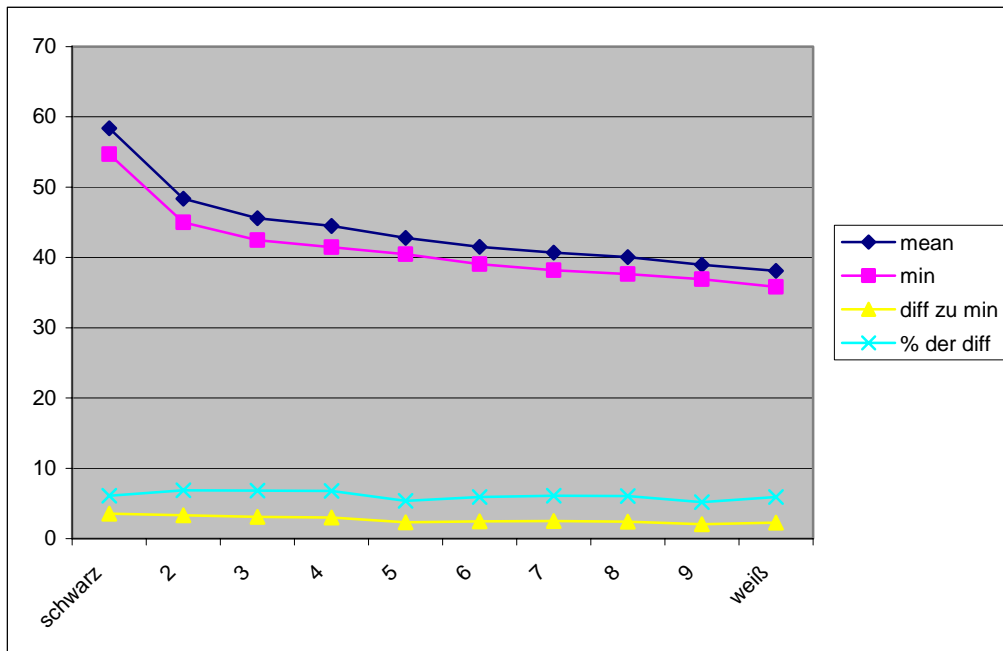
Die Daten von 24 Versuchspersonen erzeugte Werte wurden hinsichtlich Mittelwert und Minimalwert verglichen. Die Stichprobe bestand dabei aus 14 weiblichen und 10 männlichen Versuchspersonen im Alter von 17 bis 30 Jahren mit einem Mittelwert von 21,68 Jahren ( $SD=3,12$ ). 8 Versuchspersonen gaben an, Brillenträger zu sein. 3 Versuchspersonen hatten vor dem Versuch Medikamente eingenommen. Eine dieser Personen fühlte sich krank, alle anderen Personen gaben an, gesund zu sein.

Die Hypothese dieses Vorexperiments lautet, dass sich der Minimalwert der Messwerte pro Folien nicht signifikant vom Mittelwert unterscheidet. Die statistische Hypothese lautet:

$H_0: \mu = \text{Minimalwert}$

$H_1: \mu \neq \text{Minimalwert}$

Das Diagramm 3 veranschaulicht die Differenzen zwischen Mittelwert und Minimalwert und zeigt darüber hinaus die Differenz der Wertepaare und die prozentuale Differenz an.



**Diagramm 3: Verhältnis von Mittelwert zum Minimalwert**

Was bereits durch das Diagramm 3 ersichtlich wurde, wird auch durch die statistische Prüfung deutlich. Der Mittelwert unterscheidet sich im T-Test für abhängige Stichproben hoch signifikant vom Minimalwert. Dies bedeutet, dass die Hypothese  $H_0$  zugunsten der Hypothese  $H_1$  verworfen werden muss. Allerdings scheint es angebracht, die Abweichung beider Werte näher zu betrachten. Die Mittelwerte liegen im Bereich von 2,03 bis zu 3,56 Pixel über dem Minimalwert ( $M=2,7$  px). Dies entspricht einer Abweichung von durchschnittlich 6,13 Prozent bei einer Standardabweichung  $SD=0,55$  Prozent. Wie das Diagramm 3 veranschaulicht, liegt der Mittelwert mit einem relativ konstanten Abstand über dem Minimalwert.

**Tabelle 6: Vergleich der Parameter Mittelwert und Minimalwert**

<b>Folie</b>	<b>mean</b>	<b>Min</b>	<b>diff zu min</b>	<b>% der diff</b>
schwarz	58,37	54,71	3,56	6,10
2	48,37	45,00	3,33	6,88
3	45,58	42,46	3,12	6,85
4	44,49	41,46	3,03	6,81
5	42,78	40,46	2,31	5,40
6	41,5	39,04	2,46	5,93
7	40,7	38,21	2,49	6,12
8	40,04	37,63	2,42	6,04
9	38,95	36,92	2,03	5,21
weiß	38,1	35,83	2,26	5,93
<b>Mittelwert</b>	<b>43,89</b>	<b>41,17</b>	<b>2,70</b>	<b>6,13</b>
<b>SD</b>			0,49	0,55

Die Differenz zwischen Minimalwert und Mittelwert der Messung muss aus Störvariablen resultieren. Diese ließen sich im vorhandenen Setting nicht kontrollieren. Denn natürlich kann nicht mit absoluter Sicherheit ausgeschlossen werden, dass Probanden während der Messung außergewöhnliche mentale Prozesse bewältigten - obwohl dies ausdrücklich nicht erwünscht war. Eine Kontrolle war hierbei aber nicht möglich. Da das Experiment jedoch nur 50 Sekunden dauerte, ist ein mentaler Einfluss in dieser sehr kurzen Zeit eher unwahrscheinlich. Allerdings muss die Laborsituation als weiterer denkbarer Störfaktor berücksichtigt werden: Die ungewohnte und stark technisierte Umgebung und das Gefühl „vermessen“ zu werden sind durchaus geeignet, eine Fehlervarianz zu erzeugen.

Zu den nicht kontrollierbaren Einflüssen der Pupillenbewegung gehören aber auch der Hippus (plötzlich auftretende, rhythmische Schwankungen der Pupillenweite) und die Reaktion auf Atmung und Herzschlag. Eine Reaktion auf unterschiedliche Wellenlängen des Lichts aufgrund von Farbänderungen kann ausgeschlossen werden, da die Leuchtdichte des Monitors mit dem Luxmeter zuvor bestimmt wurde. Die Leuchtdichte entsprach hierbei der subjektiv wahrgenommenen Lichtstrahlung und schließt einen Farbeffekt aus (Rößger, 1997).

Der gemessene Minimalwert kann auch systematischen Fehlern unterliegen. Es ist z. B. denkbar, dass Lidschläge die Werte dann beeinflussten, wenn sie Werte produzierten, die sich im Bereich von zwei Standardabweichungen vom Mittelwert befanden und aus diesem Grund nicht herausgefiltert wurden. Der Mittelwert hat hier also den Vorteil, dass sich diese Fehlmessungen zum Teil herausmitteln und ist somit als der robustere Wert anzusehen. Obwohl die Hypothese der ersten statistischen Überprüfung nicht standhalten konnte, ist nach Prüfung der beteiligten Einflussfaktoren doch der Mittelwert als der Grundlagenwert zur Berechnung der Regression anzusehen. Letztlich ist aber darauf hinzuweisen, dass es bei der Regressionsanalyse darum geht, eine Funktion zu beschreiben, die die lichtbedingte Pupillenreaktion abbildet. Diese Funktion hat einen kleinsten und einen größten Wert der Pupillenausdehnung als Grundlage. Ob diese Werte ein höheres oder ein niedrigeres Niveau haben, ist dabei unerheblich. Von Bedeutung sind lediglich die Steigung und die Form der Funktion. Diese Steigung beider Funktionen unterscheidet sich nur geringfügig, da die Standardabweichung der Differenz zwischen Mittelwert und Minimalwert mit 0,49 Pixel als eher unerheblich einzustufen ist. Beide Kurven nehmen einen fast gleichen Verlauf und unterscheiden sich nur geringfügig in der Höhe der Werte. Deutlich wird dies in der hochsignifikanten Korrelation ( $p=.000$ ) der Wertepaare, die von  $r=.957$  bis  $r=.995$  reicht.

Im dritten Schritt erfolgt nun die Analyse der Messdaten durch verschiedene Funktionen, wie im folgenden Kapitel 3.2.5 beschrieben.

### ***3.2.5 Operationalisierung der Variablen und Methodik der Auswertung***

In dem hier beschriebenen Experiment 3 zur systematischen Herleitung einer Funktion zur Lichtbereinigung von Pupillendaten werden verschiedene Variablen verwendet. Die unabhängige Variable der Monitorabstrahlung (lux) wurde bereits in den vorhergehenden Kapiteln eingehend besprochen. Die abhängige Variable der Pupillenbewegung (rohpupil) stellt die gemessene Pupillengröße dar, welche nach den Verfahren aus dem Kapitel 3.1.5 lediglich von Artefakten bereinigt wurde.

Im Verlauf der Analyse der Daten werden aus diesem Pupillenrohwert weitere abhängige Variablen generiert. Die folgende Auflistung (Tabelle 7) gibt einen Überblick über die Bezeichnungen dieser Variablen.

**Tabelle 7: Bezeichnungen der abhängigen Variablen**

<b>Abhängige Variable</b>	<b>Beschreibung</b>
rohupuil	gemessene Rohpupille (von Artefakten bereinigt)
harpupi	harmonisierte Pupillenbewegung gem. Kapitel 2.2 Datenaufbereitung durch Harmonisierung
pup_std	standardisierte Pupillenbewegung gem. Kapitel 2.3 Datenaufbereitung durch Standardisierung
pup_r_l	linear lichtbereinigte Rohpupille
pup_r_e	exponentiell lichtbereinigte Rohpupille
pup_r_sl	exponentiell lichtbereinigte Rohpupille (square least)
pup_r_dl	doppel-linear lichtbereinigte Rohpupille
pup_h_l	linear lichtbereinigte harmonisierte Pupille
pup_h_e	exponentiell lichtbereinigte harmonisierte Pupille
pup_h_dl	doppelt-linear lichtbereinigte harmonisierte Pupille
pup_s_l	linear lichtbereinigte standardisierte Pupille
pup_s_e	exponentiell lichtbereinigte standardisierte Pupille
pup_s_dl	doppel-linear lichtbereinigte standardisierte Pupille
pup_s_q	quadratisch lichtbereinigte standardisierte Pupille

Zur Ermittlung der jeweiligen Funktionen wurden Regressionsanalysen und Kurvenanpassungen durchgeführt. Die Prüfung der Lichtbereinigung erfolgte über den Korrelationskoeffizienten nach Pearson sowie über lineare Regressionsgleichungen. Die Varianzaufteilung wurde durch eine einfaktorielle Varianzanalyse und in einigen Fällen auch durch univariate Kovarianzanalysen ermittelt. Hierbei wurde der Determinationskoeffizient  $R^2$  (in einigen Fällen auch

Eta<sup>2</sup>) als der Quotient der Quadratsumme der Varianz, welche durch die Regression aufgeklärt wird, mit der Quadratsumme der Varianz der Residuen gebildet. Die Signifikanzprüfung erfolgte durch die Prüfgröße F und das zugeordnete Signifikanzniveau von  $\alpha = .01$ .

### ***3.2.6 Vergleich von Lichtbereinigungsverfahren für Pupillendaten***

In den folgenden Abschnitten werden verschiedene Verfahren zur Bereinigung von Pupillendaten vorgestellt und verglichen. Ziel dieses Vorgehens ist die Ermittlung des jeweils besten Verfahrens für Pupillenrohdaten (Kapitel 3.2.6.1), für harmonisierte Pupillendaten (Kapitel 3.2.6.2) und für standardisierte Pupillendaten (Kapitel 3.2.6.3). Das Kapitel 3.2.6 endet nach diesem Vergleich mit einer tabellarischen Übersicht der wichtigsten Parameter. Die jeweiligen Kurvenanpassungen sind im Anhang beigefügt.

Im ersten Schritt wird die Bereinigung der Pupillenrohdaten anhand der bereits eingeführten linearen Funktion beschrieben (a). Danach folgt die Lichtbereinigung durch eine non-lineare Regressionsgleichung (b) mittels einer Exponentialfunktion über Mittelwerte und die Lichtbereinigung durch eine weitere Exponentialfunktion (c), welche mit der Methode der kleinsten Quadrate (Non-Linear-Least-Square-Fit-Test) ermittelt wurde.

Ein weiteres Verfahren, welches der Annahme nachgeht, dass sich die lichtbedingte Pupillenreaktion mit zwei unterschiedlichen, aufeinander folgenden Linearfunktionen beschreiben ließe, wird hier nicht weiter besprochen. In diesem doppel-linear-Verfahren wurde der begründeten Annahme nachgegangen, dass eine leicht steigende lineare Pupillenbewegung ab



einem Lichteinfluss von 77 Lux in eine steile lineare Funktion überwechseln würde. Diese Annahme ließ sich jedoch nicht halten.

Die Herleitung der Funktionen zur Lichtbereinigung wurde anhand der Daten des hier beschriebenen Experiments durchgeführt. Da die Variablen bestmöglich kontrolliert wurden, ist davon auszugehen, dass hauptsächlich die Monitorabstrahlung auf die Veränderung der Pupillengröße wirkte.

Im Anhang zum Kapitel 3.2.6 sind die Beziehungen der verschiedenen Pupillenvariablen zu der Monitorabstrahlungsvariablen (lux) aufgeführt. Hierbei handelt es sich im Einzelnen um die Rohpupillenwerte (rohPupil), um die harmonisierten Pupillenwerte (harmPupil) und um die standardisierten Pupillenwerte (pup\_std).

Die Rohpupillenwerte korrelieren mit den Lux-Werten mit  $r = -.410$ ,  $p < .000$ . Eine Regressionsanalyse mit varianzanalytischer Prüfung ergibt hierbei eine Pupillenvarianz von 16,8 Prozent ( $p < .000$ ), die durch die Monitorabstrahlung verursacht wurde.

Durch das Verfahren der Harmonisierung ist die Korrelation zwischen den harmonisierten Pupillenwerten und den Luxwerten bereits bei  $r = -.724$ ,  $p < .000$  angelangt. Die Prüfung auf gemeinsame Varianz ergibt hierbei bereits eine lichtbedingte Pupillenvarianz von 52,4 Prozent ( $p < .000$ ).

Da die standardisierten Pupillenwerte neben einer künstlichen Baseline auch eine gemeinsame Variationsbreite vorweisen, ist die Fehlervarianz in dieser Variablen am geringsten. Die lichtbedingte Pupillenvarianz ist mit 61,4 Prozent ( $p < .000$ ) in diesem Fall am deutlichsten ausge-

prägt. Die Korrelation der standardisierten Pupillenwerte mit der Monitorabstrahlung beträgt  $r = -.783$ ,  $p < .000$ .

Um den Erfolg der Lichtbereinigung überprüfen zu können, wurde der neue lichtbereinigte Netto-Pupillenwert zuerst mit der ursprünglichen Monitorabstrahlung Lux korreliert. Fiel diese Korrelation gering aus, so war dies ein erster Hinweis auf eine geglückte Lichtbereinigung. Um zu prüfen, ob der Zusammenhang der bereinigten Daten mit den mentalen Reizen möglichst unverändert bestehen geblieben war, musste ein weiterer Datensatz aus dem Experiment 3 zur Kontrolle herangezogen werden.

Sämtliche Funktionen, die im weiteren Verlauf dieser Arbeit errechnet wurden, wurden daher anhand dieses Datensatzes auf ihre Güte geprüft. Die Ausgangsdaten dieses Experiments 3 bestanden aus der Pupillenreaktion auf 40 Bilder, die mit verschiedenen Reizgehalten und verschiedenen Lichtabstrahlungen variiert wurden (Näheres im entsprechenden Kapitel 3.3, Experiment 3). Aus diesen 40 Stimulusbildern wurden 8 Bilder (2 Bilder pro Bedingung) ausgewählt, bei denen die Versuchspersonen vergleichbare subjektive Einschätzungen des Reizgehaltenes abgaben. So konnte die Varianz in der Variablen Reiz innerhalb der Gruppen möglichst klein gehalten werden und zwischen den Gruppen möglichst groß ausfallen. Die somit etwas eindeutiger ausfallenden Werte der Variablen Reiz erleichtern den Vergleich. Eine Analyse über alle 40 Bilder folgt dann im entsprechenden Kapitel 3.3 (Experiment 3). Die Ausgangslage der Korrelationen zeigt vorab die folgende Tabelle 8.

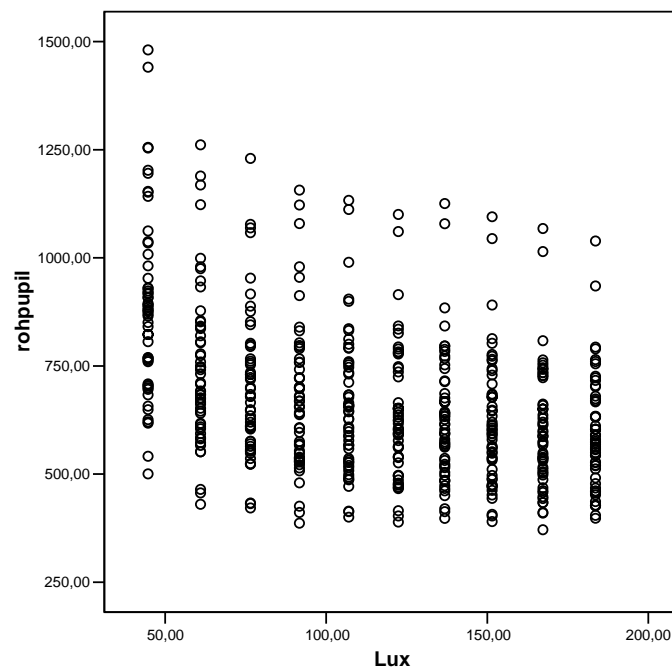
**Tabelle 8: Ausgangslage der Zusammenhänge von Licht und Reiz auf die Pupillenweite im Experiment 3 Pupillenreaktion auf bildliche Reize (Kap. 4.3).**

		REIZ	LUX
ROHPUPIL	Pearson Correlation	,200*	-,250**
	Sig. (2-tailed)	,011	,001
	N	160	160
harmonisierte Pupillengröße über alle Werte	Pearson Correlation	,435**	-,543**
	Sig. (2-tailed)	,000	,000
	N	160	160
PUP_STD	Pearson Correlation	,302**	-,382**
	Sig. (2-tailed)	,000	,000
	N	160	160

Eine geglückte Lichtbereinigung zeichnet sich dadurch aus, dass der bereinigte Pupillenwert nur noch möglichst schwach mit den Werten der Monitorabstrahlung korreliert und trotzdem weiterhin das Niveau der Korrelation mit dem mentalen Reiz hält.

### 3.2.6.1 Verfahren zur Lichtbereinigung von Pupillenrohdaten

Die Verteilung der Pupillenrohwerte (dies sind im weiteren Verlauf Pupillenrohdaten, die durch die in Kapitel 2 (Abschnitt II) beschriebenen Verfahren von Artefakten bereinigt wurden) über die verschiedenen Beleuchtungsbedingungen zeigt das Diagramm 4. Diese Daten dienen als Ausgangsbasis zur Ermittlung der verschiedenen Lichtbereinigungsfunktionen auf Rohdatenbasis. Die Korrelation der Monitorabstrahlung in Lux und der Pupillenbewegung beträgt  $r = -.410$ ,  $p < .000$  (16,8 Prozent Varianzaufklärung).



**Diagramm 4: Verteilung der Pupillenrohwerter über Beleuchtungsbedingungen**

a) Lineare Lichtbereinigung von Pupillenrohdaten

Mit dem Verfahren der non-linearen Regressionsanalyse wurde eine lineare Kurvenanpassung durchgeführt, die folgende Funktion ergab:

$$\text{Formel 6: } y = 869,161 + (-1,6987) \times lux$$

Sie ist mit  $F(538) = 108,46$ ;  $p < .000$  hoch signifikant. Im Ergebnis der Lichtbereinigung der Pupillenrohwerter durch die Linearfunktion zeigt die neue Variable (pup\_r\_1) erwartungsgemäß keine Korrelation mit den Lux-Werten der Monitorabstrahlung. Die gesamte Lichtvarianz entfällt auf die Zwischenvariable Licht, die mit  $r = 1$ ;  $p < .000$  einen absoluten Zusammenhang mit der Monitorbeleuchtung beschreibt.

Das Ergebnis der Analyse linear lichtbereinigter Pupillenwerte zeigt allerdings lediglich, dass den neuen Pupillennettowert keine gemeinsame Varianz mehr mit der Monitorabstrahlung verbindet. Dieses Ergebnis ist zwar ein Hinweis darauf, dass die Lichtbereinigung geglückt ist, als Beweis reicht sie jedoch nicht aus. Hierzu muss darüber hinaus geprüft werden, inwieweit die lichtbereinigten Pupillenwerte noch mit den mental verursachten Pupillenreaktionen korrelieren. Dafür werden die Daten des Experiments 3, welches in Kapitel 3.3 näher beschrieben wird, verwendet.

Nach der linearen Lichtbereinigung der Daten aus dem Experiment 3 (Kap. 3.3) ergeben sich Zusammenhänge aus der lichtbereinigten Pupillenweite  $pup\_r\_1$  mit der Monitorabstrahlung  $lux$  von  $r = .032$ ;  $p = .688$ . Dies bedeutet, dass die Lichtbereinigung auch hier zufrieden stellend verlaufen ist. Der geringe Zusammenhang zwischen der Monitorabstrahlung und der Pupillenweite nach der Bereinigung, welcher zudem nicht signifikant ist, zeigt dies deutlich.

Die Schwäche dieses Verfahrens wurde eingangs bereits vermutet. Da die Weitung der Pupille laut der hier aufgestellten Hypothese nicht linear verläuft, könnte dieses Verfahren die Datenbasis bei der Bereinigung insofern verschlechtern, als auch die Korrelation mit dem mentalen Reiz in Mitleidenschaft gezogen wird. Dieser mentale Reiz korrelierte in den Ausgangsdaten mit  $r = .200$  signifikant mit der Pupillenweite. Nach der linearen Lichtbereinigung korrelierte dieser Reizwert mit  $r = .208$ ;  $p < .01$ , also etwas höher als in der Ausgangssituation.

Durch die Eliminierung der lichtbedingten Varianz fällt die verbleibende Reizvarianz prozentual sogar etwas höher ins Gewicht. Zusätzlich ist diese Korrelation nun auch hoch signifikant geworden.

Das Verfahren der Lichtbereinigung mittels einer linearen Funktion kann also als zufrieden stellend bezeichnet werden. Dabei bleiben die reizbedingten Varianzen in der Pupillenbewe-

gung ausreichend erhalten. Gemäß der eingangs aufgestellten Hypothese wird im Folgenden eine Exponentialfunktion entwickelt, welche auf ihre Güte zur Lichtbereinigung der Pupillenrohwertere untersucht wird. Es ist daraufhin zu prüfen, welche der Funktionen eine bessere Lichtbereinigungsleistung hervorbringt.

b) Exponentielle Lichtbereinigung von Pupillenrohdaten

Durch das Verfahren der non-linearen Regressionsanalyse wurde nun eine exponentielle Kurvenanpassung durchgeführt, die folgende Exponentialfunktion ergab:

$$\text{Formel 7: } y = 856,857 \cdot e^{-0,0024 \cdot lux}$$

Sie ist mit  $F(538) = 108,87$ ;  $p < .000$  hoch signifikant. Nach der Lichtbereinigung der Pupillenrohwertere zeigte sich nur noch eine geringe und nicht mehr signifikante Korrelation der Pupillenwerte (pup\_r\_e) von  $r = -.034$ ,  $p = .426$  und eine sehr hohe und sehr signifikante Korrelation des durch die Funktion errechneten Lichtanteils (licht\_r\_e) mit der Monitorabstrahlung ( $r = -.999$ ,  $p = .000$ ).

Somit bleibt in diesem Fall das Ergebnis der Lichtbereinigung im Vergleich leicht hinter dem Ergebnis der linearen Lichtbereinigung zurück. Nun wird die Lichtbereinigung erneut auf das Experiment 3 aus dem Kapitel 3.3 angewendet.

Das Ergebnis dieser Bereinigung ist gleich dem Ergebnis der linearen Pupillenbereinigung ( $r = .032$ ,  $p = .688$ ). Wie schon bei der linearen Lichtbereinigung ist das Ergebnis in Bezug auf den Zusammenhang von Pupillenbewegung und mentalem Stimulus ( $r = .208$ ,  $p = .008$ ) ausreichend. Um zu überprüfen, ob diese Ergebnisse noch verbessert werden können, wird im fol-

genden Abschnitt eine weitere Exponentialfunktion getestet, welche sich durch die Verwendung eines iterativen Ansatzes in der Herleitung etwas unterscheidet. Weiterführende Informationen zu der Herleitung sind im Anhang angegeben.

c) Exponentielle Lichtbereinigung von Pupillenrohdaten (kleinste Quadrate)

Im Gegensatz zu der Exponentialfunktion aus dem vorangegangenen Abschnitt b) wird die folgende Funktion nicht über die Mittelwerte berechnet, sondern über die kleinsten Quadrate gewichtet. Zugrunde gelegt wurde eine exponentielle Abnahme des Pupillendurchmessers ( $y$ ) in Abhängigkeit von der Leuchtstärke ( $lux$ ):

$$\text{Formel 8: } y = a + b \cdot e^{-k \cdot lux}$$

Dabei ist  $a$  der Pupillendurchmesser bei extremer Helligkeit ( $lux = \text{unendlich}$ ),  $a+b$  der Pupillendurchmesser bei absoluter Dunkelheit ( $lux = 0$ ).  $k$  beschreibt, wie schnell sich der Pupillendurchmesser bei zunehmender Leuchtstärke dem Minimalwert annähert.

Die Parameter  $a, b$  und  $k$  wurden mit einem Nonlinear-Least-Square-Fit-Test mit den Daten der 55 Versuchspersonen und dem Gesamtmittel berechnet.

Zum einen zeigt das Ergebnis eine extrem schlechte Identifizierung des Lichtanteils ( $r=-.134$ ,  $p=.002$ ) von lediglich 1,8 Prozent lichtbedingtem Varianzanteil in der Variablen Licht ( $li\_r\_sl$ ) und zum anderen eine weiterhin sehr hohe Korrelation der bereinigten Pupillenwerte mit der Monitorabstrahlung von  $r=-.664$ ,  $p=.000$ . Die lichtbedingte Varianz in der bereinigten

Pupillenweite liegt hier mit 44,1 Prozent deutlich über dem Ausgangswert von 16,8 Prozent in den Pupillenrohwerten.

Die Überprüfung des Ergebnisses anhand der Daten des Experiments 3 aus Kapitel 3.3 zeigt fast eine Verdopplung der Korrelation der Ausgangswerte (für den mentalen Reiz  $r=.406$ ,  $p=.000$  und für die Lichtabstrahlung  $r=-.453$ ,  $p=.000$ ). Insgesamt ist das Verhältnis von mentalem Reiz zur Monitorabstrahlung von ursprünglich 1:0,8 auf 1:1,07 gestiegen. Trotzdem bringt dieses Verfahren der Lichtbereinigung keine substantielle Verbesserung der Datenqualität - hinsichtlich des mentalen Reizes - im Vergleich zum Ausgangsmaterial.

Das Verfahren der Lichtbereinigung von Pupillendaten mittels einer Exponentialfunktion, die auf der Grundlage der kleinsten Quadrate ermittelt wurde, ist den zuvor genannten linearen und exponentiellen Verfahren in ihrer Leistungsfähigkeit eindeutig unterlegen.

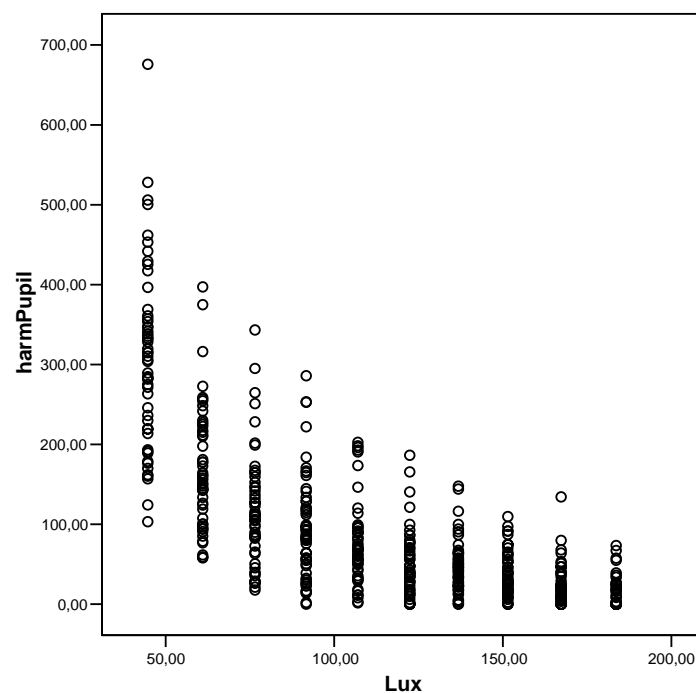
Beim Vergleich der Verfahren zur Lichtbereinigung von Pupillenrohdaten schneidet die lineare Lichtbereinigung (a) mit knappem Vorsprung vor der exponentiellen Lichtbereinigung (b) ab. Trotz der relativ guten Löschung der Lichtanteile in den Pupillenrohwerten fällt die reizbedingte Varianz in den bereinigten Pupillendaten mit 4,3 Prozent etwas höher aus als im Ausgangsniveau mit 4 Prozent. Nach dem Wegfall der Lichtanteile wäre eine prozentual höhere Wirkung der Reizvariable wünschenswert. Offensichtlich liegt die Schwierigkeit der Lichtbereinigung von Pupillenrohdaten in der unterschiedlichen Variationsbreite der Messdaten, bedingt durch den fehlenden Nullpunkt. Da dieser bei der Messung - mit den zur Verfügung stehenden Geräten - jedoch nicht erfasst werden kann, ist eine künstliche Erzeugung dieser Baseline, wie dies im Kapitel 2.2 beschrieben wurde, notwendig.



### 3.2.6.2 Verfahren zur Lichtbereinigung von harmonisierten Pupillendaten

Aufgrund der nicht ganz überzeugenden Leistungen der unter 3.2.6.1 vorgestellten Verfahren für Pupillenrohdaten werden in diesem Abschnitt Verfahren untersucht, die aus einer harmonisierten Datenbasis errechnet wurden. Nachdem sämtliche Werte mit dem in Kapitel 2.2 Datenaufbereitung durch Harmonisierung) beschriebenen Verfahren mit einer Baseline von 0 versehen wurden, wurde eine erneute lineare Kurvenanpassung durchgeführt.

Die Verteilung der harmonisierten Pupillenwerte in Abhängigkeit von den verschiedenen Beleuchtungsbedingungen zeigt das Diagramm 5. Diese Daten dienen als Ausgangsbasis zur Ermittlung der verschiedenen Lichtbereinigungsfunktionen auf Basis harmonisierter Pupillenwerte. Die Korrelation der Monitorabstrahlung in Lux und der Pupillenbewegung betrug hierbei  $r = -.724$ ,  $p < .000$ , wobei die lichtbedingte Varianz in den Pupillenwerten 52,4 Prozent ausmachte.



**Diagramm 5: Verteilung der harmonisierten Pupillenwerte über Beleuchtungsbedingungen**

Untersucht wird im Folgenden die Lichtbereinigung durch eine lineare Funktion (a) und die Bereinigung durch eine Exponentialfunktion (b).

a) Lineare Lichtbereinigung von harmonisierten Pupillendaten

Nach der Durchführung einer linearen Kurvenanpassung für die harmonisierten Pupillenwerte über die Monitorabstrahlung wurde folgende Funktion ermittelt:

$$\text{Formel 9: } y = 286,795 + (-1.6987) \cdot lux$$

Sie ist mit  $F(538) = 592,43$ ;  $p < .000$  hoch signifikant. Zur Überprüfung der Güte dieser Funktion zur Lichtbereinigung der harmonisierten Pupillendaten wurde für jeden Lux-Wert der Versuchsdaten ein entsprechender lichtbedingter Pupillenwert ( $li\_h\_l$ ) ermittelt. Dieser Wert wurde daraufhin vom harmonisierten Pupillenausgangswert abgezogen, um den Nettowert der Pupillenbewegung ( $pup\_h\_l$ ) zu erhalten.

Es ist keine Korrelation der neuen, lichtbereinigten Pupillenbewegung mit der Lichtvariablen Lux mehr zu erkennen ( $r=0$ ,  $p=1$ ). Die Variable  $li\_h\_l$ , welche den identifizierten Lichtanteil darstellt, hat hingegen 100 Prozent der Lichtvarianz aufgenommen. Die Prüfung der Qualität dieser Lichtbereinigungsfunktion wurde nun wieder an den Daten des Experiments 3 aus Kapitel 3.3 vorgenommen.

Das Ergebnis der Lichtbereinigung zeigt eine durchaus zufrieden stellende Wirkung der Lichtbereinigung. Die lichtbedingte Varianz in der Pupillenweite ist von 18,9 Prozent auf 0,64 Prozent zurückgegangen. Dieser Wert ist nicht signifikant ( $r=.08$ ,  $p=.314$ ). Die reizbe-

dingte Varianz in der Pupillenbewegung ist hingegen von 43,5 Prozent auf nur auf 26,9 Prozent abgesunken ( $r=.519$ ,  $p=.000$ ). Hier ist also von einer Verschlechterung der Qualität in den Daten auszugehen. Im folgenden Abschnitt wird nun geprüft, ob eine exponentielle Lichtbereinigung ein besseres Ergebnis erzielen kann.

*b) Exponentielle Lichtbereinigung von harmonisierten Pupillendaten*

Nach der Durchführung einer exponentiellen Kurvenanpassung ergibt sich folgende Funktion:

$$\text{Formel 10: } y = 918,842 \times e^{-0,0271 \times \text{lux}}$$

Sie ist mit  $F(538) = 621,84$ ,  $p < .000$  hoch signifikant. Zur Überprüfung der Güte dieser Funktion zur Lichtbereinigung der harmonisierten Pupillendaten wurde für jeden Lux-Wert der Versuchsdaten ein entsprechender lichtbedingter Pupillenwert ( $li\_h\_e$ ) ermittelt. Dieser Wert wurde daraufhin vom harmonisierten Pupillenausgangswert abgezogen, um den Nettowert der Pupillenbewegung ( $pup\_h\_e$ ) zu erhalten.

Der Lichtanteil der Pupille ( $li\_h\_e$ ) korreliert hoch signifikant mit den Lux-Werten ( $r = -.894$ ,  $p = .000$ ), wohingegen die lichtbereinigten Pupillenwerte ( $pup\_h\_e$ ) nicht mehr signifikant korrelieren ( $r = -.006$ ,  $p = .898$ ). Es gibt also keinen statistisch bedeutsamen Zusammenhang mehr zwischen den lichtbereinigten Pupillenwerten und dem Licht.

Das Ergebnis der Lichtbereinigung bleibt in diesem Fall etwas hinter dem Ergebnis der linearen Lichtbereinigung (a) zurück. Nun wird anhand des Datensatzes des Experiments 3 (aus Kapitel 3.3) untersucht, ob die Lichtbereinigung der Pupillendaten mit dieser Exponentialfunktion ein besseres Ergebnis bezüglich des mentalen Reizes erzielen kann. Das Ergebnis der

Korrelation beträgt  $r=.528$ ,  $p=.000$  für mentalen Reiz und lichtbereinigten Pupillenwert, sowie  $r=-.171$ ,  $p=.03$  für die Lichtabstrahlung und den lichtbereinigten Pupillenwert.

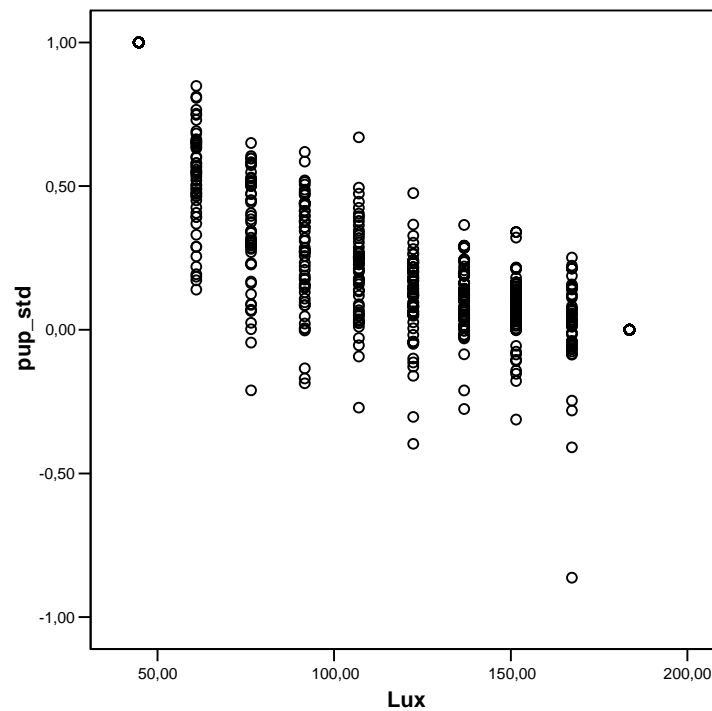
Die Varianzaufklärung der Pupillenbewegung durch den mentalen Reiz ist nach dieser Form der Lichtbereinigung mit 27,9 Prozent etwas besser ausgefallen als bei der linearen Lichtbereinigung. Die Lichtbereinigung ist aber dafür deutlich schlechter ausgefallen, da die lichtbereinigte Pupillenvariable immer noch mit 2,9 Prozent signifikant von der Lichtvariablen beeinflusst wird. Insofern stellt die hier ermittelte Exponentialfunktion keine wesentliche Verbesserung gegenüber der Linearfunktion dar.

In der Gesamtbetrachtung sind die Lichtbereinigungsverfahren für harmonisierte Pupillendaten den Verfahren der Lichtbereinigung für Pupillenrohdaten deutlich überlegen. Problematisch ist jedoch die Tatsache, dass die entsprechenden Werte zur Harmonisierung der Daten erst nach dem Abschluss der Messung vorliegen. Für eine Anwendung im Bereich adaptiver Systeme können diese Daten somit nicht verwendet werden, sie können lediglich zur Analyse vorhandener Daten genutzt werden. Da es aber ein ausdrückliches Ziel dieser Arbeit ist, Verfahren für die Echtzeitanalyse von Daten bereitzustellen, werden im folgenden Abschnitt Lichtbereinigungsverfahren für standardisierte Pupillenwerte diskutiert.

### 3.2.6.3 Verfahren zur Lichtbereinigung von standardisierten Pupillendaten

Die Grundlage dieser Lichtbereinigung ist die Standardisierung der Pupillenbewegungen gemäß Kapitel 2.3. Durch die Standardisierung der Pupillenwerte wird neben der Schaffung einer einheitlichen Baseline, die nicht unbedingt mit dem kleinsten Wert zusammenfallen muss, auch die Variabilität eingeschränkt. Die standardisierten Werte werden dadurch - über

die Versuchsperson hinaus - gut vergleichbar. Die Streuung der standardisierten Werte zeigt das Diagramm 6.



**Diagramm 6: Verteilung der standardisierten Pupillenwerte über Lux-Werte**

In den nächsten Abschnitten werden Lichtbereinigungsfunktionen hergeleitet und auf ihre Qualität untersucht. Zuerst wird eine lineare Lichtbereinigungsfunktion (a) aus den standardisierten Pupillenwerten errechnet, die dann mit einer Exponentialfunktion (b) verglichen wird. Im letzten Abschnitt (c) wird dann eine quadratische Funktion hergeleitet, die auf einer Gewichtung der standardisierten Werte nach ihren Varianzen beruht. Das Verfahren der Gewichtung von Abweichungen kann nur bei standardisierten Werten sinnvoll angewendet werden.

a) Lineare Lichtbereinigung von standardisierten Pupillendaten

Nach der Durchführung einer linearen Kurvenanpassung für die standardisierten Pupillenwerte über die Monitorabstrahlung (lux) wurde folgende Funktion ermittelt:

$$\text{Formel 11: } y = 1,1264 + (-0.0058) \cdot lux$$

Sie ist mit  $F(538) = 902,61$ ;  $p < .000$  hoch signifikant. Die Lichtbereinigung erfolgte nun mit dieser ermittelten Linearfunktion. Die hierdurch erzeugte neue Variable  $pup\_s\_1$  stellt dabei die Pupillenbewegung dar, die von den linearen Lichtanteilen bereinigt wurde. Die Korrelation der identifizierten lichtbedingten Pupillenwerte ( $li\_s\_1$ ) mit den Werten der Monitorabstrahlung ist mit  $r = -.790$ ,  $p < .000$  sehr stark negativ. Trotzdem ist die Korrelation der lichtbereinigten Pupillenwerte mit der Monitorabstrahlung (lux) mit  $r = -.465$ ;  $p < .000$  noch immer stark negativ. Die lineare Lichtbereinigung scheint in diesem Fall also keine geeigneten Ergebnisse zu liefern.

Die Überprüfung der linearen Lichtbereinigungsfunktion anhand des Datensatzes des Experiments 3 (Kapitel 3.3) bestätigte das negative Abschneiden der Lichtbereinigungsfunktion. Die Korrelation der Pupillenweite mit dem mentalen Reiz ist mit  $r = .335$ ;  $p < .000$  zwar noch recht nah am Ausgangswert, die Korrelation mit der Monitorabstrahlung von  $r = -.209$ ,  $p < .05$  ist aber noch zu hoch. Dies entspricht immerhin einer signifikanten Varianzproportion von 4,4 Prozent.

Insgesamt ist die Qualität der linear lichtbereinigten Pupillenwerte als nicht ausreichend anzusehen. Im nächsten Abschnitt wird analog zu den vorherigen Kapiteln untersucht, ob eine Exponentialfunktion zur Lichtbereinigung der Pupillenwerte bessere Ergebnisse erzielen kann.

b) Exponentielle Lichtbereinigung von standardisierten Pupillendaten

Nach der Durchführung einer exponentiellen Kurvenanpassung über die standardisierten Pupillenwerte ergab sich folgende Funktion:

$$\text{Formel 12: } y = 1,5395 \times e^{-0,0126 \cdot \text{lux}}$$

Sie ist mit  $F(538) = 1016,26$ ;  $p < .000$  hoch signifikant. Die Lichtbereinigung erfolgte nun mit dieser ermittelten Exponentialfunktion. Die hierdurch erzeugte neue Variable `pup_s_e` stellt dabei die Pupillenbewegung dar, die von den exponentiellen Lichtanteilen bereinigt wurde. Die Korrelation der identifizierten Lichtwerte (`li_s_e`) mit den Werten der Monitorabstrahlung ist mit  $r = -.770$ ,  $p < .000$  sehr stark negativ. Trotzdem ist die Korrelation der lichtbereinigten Pupillenwerte mit der Monitorabstrahlung (lux) mit  $r = -.476$ ,  $p < .000$  noch immer stark negativ. Die exponentielle Lichtbereinigung ist in diesem Fall noch weniger geeignet, als dies bei der linearen Lichtbereinigung in der vorhergehenden Analyse der Fall war.

Eine Überprüfung der Lichtbereinigung anhand der Daten des Experiments 3 bestätigt diesen Trend. Die Pupillenweite korreliert mit  $r = -.273$ ;  $p < .000$  deutlich höher mit den Lux-Werten als die Pupillenwerte nach der linearen Lichtbereinigung. Die Korrelation mit den mentalen Reizen nimmt dabei einen fast gleichen Wert von  $r = .332$ ;  $p < .000$  an.

Es zeigt sich in diesem Versuchsabschnitt, dass die exponentielle Lichtbereinigung keine Verbesserung gegenüber der linearen Lichtbereinigung bringt. Ganz im Gegenteil ist auf der Grundlage der standardisierten Pupillenwerte ein besseres Abschneiden der linearen Lichtbereinigung festzustellen.

Da die beiden Lichtbereinigungen für standardisierte Pupillenwerte kein zufriedenstellendes Ergebnis erbrachten, wird im folgenden Abschnitt eine Verbesserung der Datenbasis der standardisierten Pupillenbewegungen durch die Gewichtung von Abweichungen vorgenommen, um hierdurch eine exaktere Lichtbereinigung entwickeln zu können.

### c) Quadratisch- gewichtete Lichtbereinigung von standardisierten Pupillendaten

Die bisherigen Versuche der Erfassung lichtbedingter Varianzen in den Pupillenbewegungen führten zu keinem befriedigenden Ergebnis. Insofern scheinen die nicht lichtbedingten Varianzen in den Daten zu Fehlern in der Kurvenanpassung zu führen, die das Identifizieren der Lichtvarianz nur ungenügend zulassen. Aus diesem Grund wird der bisher beschrittene Weg nun verlassen, der darauf beruhte, die Kurvenanpassungen über aggregierte Mittelwerte vorzunehmen. Um den Messfehler möglichst auszuschalten, wird in diesem Abschnitt eine non-lineare Regressionsanalyse beschrieben, deren Messwerte über eine Gewichtsschätzung ermittelt werden.

Wie bereits in den Streudiagrammen deutlich wurde, vergrößert sich die Varianz der abhängigen Variablen (Pupillenbewegung) mit der Steigung der unabhängigen Variablen (Lux). In



diesem Falle ist es angebracht, die einzelnen Datenpunkte so zu gewichten, dass die Messungen mit den größeren Varianzen einen geringeren Einfluss auf den Mittelwert haben (Bühl & Zöfel, 2002). Eine lineare Regressionsanalyse ergibt eine gemeinsame Varianz der Variablen Lux und der standardisierten Pupillenweite von 61,3 Prozent ( $R^2=.613$ ).

Bei den Verfahren der Varianzen-Gewichtung der unabhängigen Variablen wird der reziproke Wert der Varianz in jeder Gruppe der abhängigen Variablen als Gewichtungsfaktor genutzt. Die optimale Potenz nach der Likelihood-Funktion beträgt -2. Sie wird zur Gewichtung der Einzelwerte eingesetzt. Der Korrelationskoeffizient und der Determinationskoeffizient ( $R^2=.971$ ) haben sich durch diese Gewichtung deutlich verbessert.

Insbesondere die deutliche Abnahme des Standardfehlers von 0,20192 auf 0,05466 ist für die Verbesserung der Qualität der neuen gewichteten und standardisierten Variablen von besonderer Bedeutung.

Auf diese neue Variable wird nun wieder eine non-lineare Regressionsanalyse angewendet und eine Kurvenanpassung durchgeführt. Ergebnis dieser Kurvenanpassung ist die Ermittlung der folgenden quadratischen Funktion, welche mit  $F(537)= 1,2E+18$ ;  $p<.000$  hoch signifikant ist.

$$\text{Formel 13: } y = -0,630 + (-2, e - 16) \times lux + 3,2e - 05 \times lux^2$$

Diese quadratische Funktion über gewichtete standardisierte Pupillenwerte wird nun auf ihre Güte geprüft, indem sie zur Lichtbereinigung der standardisierten Pupillenwerte eingesetzt wird.

Das Ergebnis dieser Lichtbereinigung zeigt eine nur noch minimale Korrelation der lichtbereinigten Pupillenwerte mit der Monitorabstrahlung ( $r = -.035$ ,  $p = .414$ ), welche nicht signifikant ist. Die Lichtbereinigung ist also sehr zufriedenstellend verlaufen, da der größte Teil der lichtbedingten Varianz in der Variablen *li\_s\_q* aufgegangen ist ( $r = .985$ ,  $p = .000$ ). Um nun zu prüfen, ob dieses Vorgehen der Lichtbereinigung der Pupillendaten die Varianzen unangetastet lässt, welche nicht vom Licht erzeugt wurden, wird nun wieder der Datensatz des Experiments 3 aus Kapitel 3.3 geprüft.

Wie bereits im ersten Teil der Überprüfung deutlich wurde, ist auch hier eine gute Lichtbereinigung geglückt. Die Korrelation von  $r = -.036$ ;  $p > .05$  ist nicht nur minimal, sondern auch nicht signifikant. Der Determinationskoeffizient von  $R^2 = .001$  zeigt eine nicht signifikante Wirkung der unabhängigen Variablen *lux* von 1 Prozent auf die Varianz der Pupillenweite.

Die Korrelation zwischen dem mentalen Reiz und der Pupillenweite ist hingegen noch deutlicher als in der Ausgangsdatenlage. Durch das Eliminieren der Lichtvarianz konnte die Korrelation mit  $r = .320$ ;  $p < .000$  über den Ausgangswert von  $r = .302$ ;  $p < .000$  gelangen. Eine Regressionsanalyse zeigt einen Einfluss des mentalen Reizes auf die Varianz der Pupillenbewegung von 10,2 Prozent.

Das Verfahren der quadratischen Pupillen-Lichtbereinigung ist somit sämtlichen vorhergehend beschriebenen Verfahren deutlich überlegen. Darüber hinaus genießt dieses Verfahren noch eine sehr hohe praktische Relevanz, da die notwendigen Parameter einer lichtbedingt kleinsten und einer lichtbedingt größten Pupillenweite durch eine Kalibrierung mittels schwarzem und weißem Monitor vor einer Messung schnell und einfach erhoben werden

können. Somit stehen, anders als bei der Harmonisierung, sämtliche notwendigen Werte für eine Echtzeitanalyse zur Verfügung.

### ***3.2.7 Ergebnisse und Diskussion***

In den vorhergehenden Abschnitten konnte gezeigt werden, dass es für die drei verschiedenen Arten von Pupillenbewegungsdaten verschiedene Möglichkeiten der Lichtbereinigung gibt, welche in der Qualität der Lichtbereinigung unterschiedlich ausfallen. Die eingangs aufgestellte Hypothese, die von einer Überlegenheit der exponentiellen Lichtbereinigung ausging, kann nicht ohne weiteres gehalten werden. Vielmehr müssen die drei verschiedenen Arten der Pupillendaten gesondert betrachtet werden, da es keine Lichtbereinigung gibt, die auf alle Fälle gleichermaßen passt.

Für den Bereich der Pupillenrohwerte konnte gezeigt werden, dass eine lineare Lichtbereinigung die besten Ergebnisse liefert. Allerdings sind diese Ergebnisse kaum besser, als die der Exponentialfunktion, da diese der Linearfunktion sehr ähnlich ist. Für die Anwendung macht es hier keinen Unterschied, welche der beiden Funktionen zur Lichtbereinigung verwendet wird. Außer Frage steht jedoch die mangelhafte Lichtbereinigungsleistung der Exponentialfunktion, welche auf der Grundlage kleinster Quadrate ermittelt wurde. Sie konnte zwar die besten Ergebnisse bei dem Verhältnis mentaler Reize zur Pupillenweite erzielen, hatte dabei jedoch noch eine stark negative Beziehung zu den Werten der Monitorabstrahlung.

Für den Bereich der harmonisierten Daten ergibt sich ein deutlich anderes Bild. Das Bereinigungsverfahren auf der Grundlage der linearen Lichtbereinigung schneidet hierbei besser ab.

Es liefert eine deutlich besser Lichtbereinigung, wobei jedoch eine leichte Verschlechterung bei den Werten der Varianz des mentalen Reizes zu beklagen ist. Die exponentielle Lichtbereinigung leistet in diesem Punkt etwas mehr, sie schneidet jedoch bei der Lichtbereinigung inakzeptabel ab.

Das Verfahren der Datenharmonisierung, für das der Minimalwert der Messung benötigt wird, ist für eine Echtzeitanalyse nicht einsetzbar. Es kann lediglich für Analysen nach der Beendigung einer Messung eingesetzt werden. In diesem Fall kann aber davon ausgegangen werden, dass die verlässlichsten Daten generiert werden. Ist die Echtzeitbedingung also nicht unbedingt erforderlich, sollten in jedem Fall harmonisierte Daten zur Analyse genutzt werden.

Die interessanteste Datenbasis dieses Kapitels ist die standardisierte Pupillenweite. Durch eine Skalenangleichung bietet sie sehr gute Vergleichsmöglichkeiten zwischen den einzelnen Datensätzen. In der Lichtbereinigung über die standardisierte Datenbasis zeigt die lineare Funktion ebenso wie die Exponentialfunktion erhebliche Schwächen. Nach der Lichtbereinigung zeigten beide Variablen noch mäßige Korrelationen mit den Monitorabstrahlungswerten, die im linearen Fall signifikant und im exponentiellen Fall sogar sehr signifikant waren. Hier zeigte sich die quadratische Funktion weit überlegen. Sie war imstande fast die gesamte Lichtvarianz in der Pupillenvariablen aufzudecken, ohne dabei die mental reizbedingte Varianz zu verkleinern.

Betrachtet man die drei abhängigen Variablen nun im Vergleich, so ist ein eindeutiger Trend von den Rohpupillenwerten über die harmonisierten Pupillenwerte bis hin zu standardisierten Pupillenwerten zu erkennen, was die Steigerung der Qualität angeht. Bei der Auswahl eines geeigneten Verfahrens scheiden die Bereinigungsverfahren für die Rohpupille aus, da sie im Vergleich nur mäßige Ergebnisse erbrachten. Die harmonisierten Daten sind für eine Echtzeit-

analyse nicht geeignet, da der Minimalwert einer Messung in das Harmonisierungsverfahren einfließt, welcher aber erst nach Abschluss einer Messung zur Verfügung steht. Diese Daten können also nicht für eine Prozesssteuerung eingesetzt werden.

Angesichts dieser Erwägungen bleibt nur die quadratische Lichtbereinigung auf der Grundlage gewichteter Pupillenwerte als geeignetes Verfahren übrig. Da das Verfahren der Wertegewichtung nur für standardisierte Daten anwendbar ist, fehlen hier die direkten Vergleichsmöglichkeiten. Fest steht jedoch, dass die Ergebnisse der Lichtbereinigung bei weitem besser ausfallen, als bei jedem anderen hier getesteten Verfahren.

Die von Rößger (1997) sowie Boff und Lincoln (1988) vertretene These, dass sich ein linearer Trend in Pupillenbewegungen über geringe Leuchtdichteunterschiede abzeichnen würde ist somit widerlegt. Diese Annahme lässt sich lediglich für Pupillenrohwerte aufrechterhalten, die aber aus Gründen geringer Vergleichbarkeit möglichst nicht zu verwenden sind. Fortgeschrittene Verfahren der Datenvorverarbeitung schließen eine lineare Lichtbereinigung aus.

Da sich vor dem Beginn einer Messung durch das Kalibrieren der Pupillenvariabilität zwischen dem dunkelsten und dem hellsten Lichtwert alle notwendigen Informationen zum Einsatz der quadratischen Lichtbereinigungsfunktion schnell und einfach abnehmen lassen, ist dieses Verfahren als Lichtbereinigungsverfahren für weitere Pupillenanalysen anzuwenden.

Tabelle 9: Vergleichsübersicht der Lichtbereinigungsfunktionen

(\*=signifikant, \*\*= sehr signifikant)

## Übersicht Regressionsanalysen

Standardisierung	Funktion	Rsq	Ausgangswert gr	Experiment 2	Ausgangswert	Experiment 3
3.2.6.1 Rohpupille	a) linear AV: pup_rl	.168	lux: r= -.410** R2= .168**	lux: r= 0 R2= .000** harmPupil: r= .448** std: r= 218**	lux: r= -.250** R2= .062** reiz: r= .200* R2= .040* verh. 1:0,8	reiz: r= .208** R2= .043** lux: r= .032 R2= .001 verh. 1:650
	b) exponential AV: pup_re	.168		lux: r= -.034 R2= .001** harmPupil: r= .468** std: r= .339**		reiz: r= .208** R2= .043** lux: r= .032 R2= .001 verh. 1:650
	c) square least AV: pup_rsl			lux: r= -.664** Eta2= .441** harmPupil: r= .973** std: r= .883**		reiz: r= .435** R2= .189** lux: r= .406** R2= .165** verh. 1:1,07
	doppellinear AV: pup_rdl	.046 .151		lux: r= -.004 R2= .000 harmPupil: r= .371** std: r= .127**		reiz: r= .206** R2= .043** lux: r= .094 verh. 1:219,14
3.2.6.2 harmonisierte Pupille	a) linear AV: pup_h_l	.524	lux: r= -.724** R2= .524**	lux: r= 0 R2= .0 harmPupil: r= .640** std: r= .450**	lux: r= -.543** R2= .294** reiz: r= .435** R2= .189** verh. 1:0,8	reiz: r= .519** R2= .269** lux: r= .080 R2= .006 verh. 1:648,75
	b) exponential AV: pup_h_e	.536		lux: r= -.006 R2= .000 harmPupil: r= .592** std: r= 258**		reiz: r= .528** R2= .279** lux: r= .171* R2= .029* verh. 1:3,08
	doppellinear AV: pup_h_dl	.523 .434		lux: r= -.012 R2= 0 harmPupil: r= .578** std: r= 271**		reiz: r= .505** R2= .255** lux: r= .229** R2= .053** verh. 1:2,2
3.2.6.3 standardisierte Pupille	a) linear AV: pup_s_l	.654	Lux: r= -.783** R2= .614**	lux: r= -.465** R2= .216** harmPupil: r= .539** std: r= .703**	lux: r= -.382** R2= .146** reiz: r= .302** R2= .085** verh. 1:0,8	reiz: r= .335** R2= .112** lux: r= .209* R2= .044* verh. 1:6
	b) exponential AV: pup_s_e (c2)	.677		lux: r= -.476** R2= .227** harmPupil: r= .515** std: r= 669**		reiz: r= .332** R2= .110** lux: r= .273** R2= .075** verh. 1:1,22
	c) quadratisch AV: pup_s_q	1.0		lux: r= -.035 R2= .001 harmPupil: r= .106* std: .235**		reiz: r= .320** R2= .102** lux: r= .036 R2= .001 verh. 1:888,89
	doppellinear AV: pup_s_dl	.347 .728		lux: r= -.373** R2= .139** harmPupil: r= .344** std: .474**		reiz: r= .319** R2= .102** lux: r= .328** R2= .107** verh. 1:0,97

### **3.3 Experiment 3: Pupillenreaktion auf bildliche Reize**

Das Experiment zur Pupillenreaktion auf bildliche Reize wurde mit dem Ziel durchgeführt die mental bedingte Reaktion der Pupille, zu untersuchen und sie mit der Reaktion auf lichtbedingte Veränderungen zu vergleichen. Es sollte hierbei gezeigt werden, dass der mentale Reiz den Lichtreiz überlagert, und dass demnach auch in der dunkelsten Umgebung im Setting (also mit der größten lichtbedingten Pupillenweitung) eine mentale Reaktion der Pupille erfolgt, welche sie über das lichtbedingte Niveau vergrößert.

Als Stimulus wurde visuelles Material angefertigt, welches unterschiedlich starke Reize auf die Probanden ausübte. Die Reizstärke wurde nach dem Experiment über eine individuelle Einschätzung der Probanden validiert. Das Stimulusmaterial wurde außerdem über die Faktoren *Hell* und *Dunkel* variiert.

#### ***3.3.1 Hypothesen***

Die dem Experiment zugrunde liegende Hypothese lautet:

Die durch mentale Vorgänge innervierte Pupillenweitung überlagert den Effekt der lichtinnervierten Pupillenweitung.

Hierbei bleibt die Art des mentalen Reizes und die dadurch angesprochene psychische Dimension unberücksichtigt. Zur weiteren Analyse werden die Variablen in den folgenden Experimenten separiert und kontrolliert.

Die Unterhypothesen lauten:

1. Der Mittelwert der Pupillendurchmesser bei reizhaltigen hellen Vorlagen ( $\mu_{rv\_h}$ ) ist kleiner als der Mittelwert bei reizhaltigen dunklen Vorlagen ( $\mu_{rv\_d}$ ).
2. Der Mittelwert der Pupillendurchmesser bei reizarmen hellen Vorlagen ( $\mu_{ra\_h}$ ) ist kleiner als der Mittelwert bei reizarmen dunklen Vorlagen ( $\mu_{ra\_d}$ ).
3. Der Mittelwert der Pupillendurchmesser bei hellem und wenig reizhaltigem visuellen Material ( $\mu_{ra\_h}$ ) ist geringer als bei reizhaltigem hellem Material ( $\mu_{rv\_h}$ ).
4. Der Mittelwert der Pupillendurchmesser bei dunklem und wenig reizhaltigen visuellen Material ( $\mu_{ra\_d}$ ) ist geringer als bei reizhaltigem dunklem Material ( $\mu_{rv\_d}$ ).
5. Der Mittelwert der Pupillendurchmesser bei reizarmen hellen Vorlagen ( $\mu_{ra\_h}$ ) ist kleiner als der Mittelwert der Pupillendurchmesser bei reizhaltigen dunklen Vorlagen ( $\mu_{rv\_d}$ ).

### ***3.3.2 Operationalisierung und Methoden***

Bei dem durchgeführten Experiment handelt es sich um ein Messwiederholungsdesign auf zwei über zwei Faktoren. Folgende Variablen wurden dabei erhoben:

- Pupillenweite (abhängige Variable)



- Valenz der Reizvorlage (unabhängige Variable)
- Monitorabstrahlung in Lux (unabhängige Variable)
- Einschätzung des Reizgehaltes für positive Vorlagen (Ordinalskala 0 bis 4)
- Einschätzung des Reizgehaltes für negative Vorlagen (Ordinalskala 0 bis 4)
- Mediennutzung in Stunden

Für die statistische Auswertung wurden mehrfaktorielle Messwiederholungsverfahren des allgemeinen linearen Modells verwendet. Darüber hinaus kamen auch Varianzanalysen und Mittelwertvergleiche zum Einsatz.

### **3.3.3 Stichprobenbeschreibung**

An dem Experiment *Pupillenreaktion auf bildliche Reize* nahmen insgesamt  $N=32$  Versuchspersonen teil. Diese Stichprobe untergliederte sich in 11 männliche und 21 weibliche Versuchspersonen. Das Alter der Versuchspersonen reichte von 17 bis 38 Jahren mit einem Mittelwert von  $M=25,8$  Jahren mit einer Standardabweichung von  $SD=4,4$  Jahren. Als weitere Variable wurde der durchschnittliche wöchentliche Fernsehkonsum mit  $M=6,9$  Stunden in der Woche und die durchschnittliche Computernutzung mit  $M=12,8$  Stunden in der Woche erfasst.

### **3.3.4 Versuchsdurchführung**

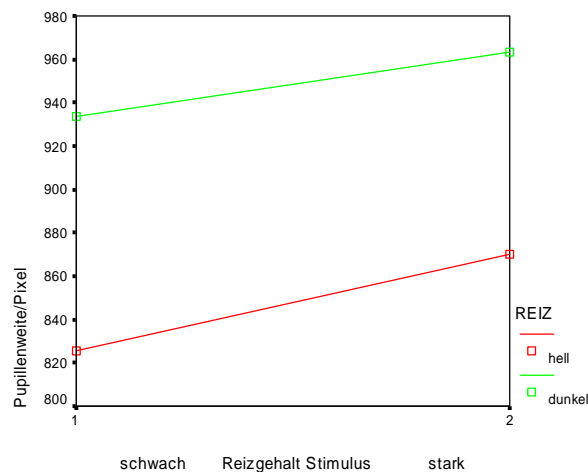
Jeder Versuchsperson wurden am Monitor 43 Folien verschiedenen Inhalts vorgeführt. Das Experiment startete mit einer schwarzen Folie, gefolgt von einer 50-prozentigen Graufolie und schließlich einer Weißfolie. Diese ersten drei Folien dienten der späteren Kalibrierung der individuellen lichtbedingten Pupillengröße. Anschließend wurden 40 Folien mit verschiedenen Fotos in zufälliger Reihenfolge präsentiert. Die Folien variierten über die vier Faktoren hell, dunkel, reizhaltig und reizarm mit je 10 Bildern pro Faktor. Die Fotografien nahmen dabei ca. 30% des Bildschirmplatzes ein. Die 20 Fotos mit hellen Motiven wurden auf einem weißen Hintergrund präsentiert, um den Lichteffect zu steigern. Die 20 Fotos mit dunklen Motiven wurden auf einem schwarzen Hintergrund präsentiert um die Monitorstrahlung zu minimieren. Die Leuchtstärke variierte dabei von  $M_h=174$  Lux bei hellen Vorlagen und bis  $M_d=103$  Lux bei dunklen Vorlagen.

Um den individuellen Reizgehalt der Bilder zu ermitteln, wurde den Probanden nach dem Experiment eine Übersicht der Bilder gezeigt und sie wurden dazu aufgefordert, die Bilder in ihrer Wirkung positiv oder negativ zu bewerten. Weiterhin sollten die Versuchspersonen auf einer Skala von 0 bis 4 den individuellen Reizgehalt bewerten. Zur Analyse der Pupillenweitung über die Lichtwerte wurden die Messdaten der 8 Bilder aggregiert und ausgewertet, für die einheitliche Einschätzungen vorlagen. Für alle anderen Auswertungen wurden sämtliche Bilder verwendet.

### 3.3.5 Ergebnisse

In dem Experiment *Pupillenreaktion auf bildliche Reize* konnten mittels einer Varianzanalyse in einem zweifaktoriellen Experimentaldesign (N= 32) über zwei Faktoren mit Messwiederholung in zwei systematisch variierten Gruppen Effekte von verschiedenen visuellen Reizmaterialien auf den Pupillendurchmesser nachgewiesen werden, die nicht auf den Einfluss des Raum- oder Monitorlichtes zurückgeführt werden konnten. Aufgrund dieses Versuches sowie der gängigen Expertenmeinung (Beatty, 1982; Galley, 2001; Hess, 1977; Kahnemann, 1973; Lang et al. 2004; Manzey, 1998; Rößger, 1997) konnte eine lichtunabhängige Pupillenbewegung repliziert und für das bestehende Setting bestätigt werden.

Wie die folgende Grafik (Diagramm 7) verdeutlicht, konnte eine Vergrößerung der Pupille gemessen werden, sobald reizhaltiges visuelles Material präsentiert wurde. Sowohl bei dunklen als auch bei hellen Vorlagen reagierte die Pupille bei reizhaltigem Material stärker. Bei reizarmen hellen Vorlagen betrug der Mittelwert der Pupillenweite  $M=825$  Pixel,  $SD=158,57$ , während bei reizhaltigem hellem Material  $M=870$ ,  $SD=172,38$  Pixel gemessen wurden. Bei reizarmen dunklen Stimuli konnten  $M=934$ ,  $SD=180,68$  Pixel und bei reizhaltigen dunklen Stimuli konnten  $M=963$ ,  $SD=196,30$  Pixel gemessen werden. Die Leuchtstärke des Monitors variierte dabei von  $M_h=174$  Lux bis zu  $M_d=103$  Lux. Die Effekte *Licht* und *mentaler Reiz* konnten - für Licht mit  $F(1,30)=205,539$ ,  $p<0,00$  und für den mentalen Reiz mit  $F(1,30)=24,755$ ,  $p<0,00$  - statistisch hoch signifikant abgesichert werden. Somit beträgt die auf die Lichtschwankung eines Monitors begrenzte Differenz der Mittelwerte zwischen den Effekten im hellen Umfeld 45 Pixel und im dunklen Umfeld 39 Pixel. Hier wird also deutlich, dass sich die Effekte leicht überlagern. Somit muss in dunkler Umgebung mit leicht verminderten Effektstärken bei mentalen Reizreaktionen gerechnet werden, während diese im hellen Umfeld höhere Werte annehmen.



**Diagramm 7: Effekte der Strahlungsstärke des Monitors sowie des Reizgehaltes der Stimuli**

Das Experiment verdeutlicht, dass ein nicht unerheblicher Pupilleneffekt durch die mentale Belastung der Versuchspersonen erzeugt wird. Dieser Effekt überlagert den Lichtreflex der Pupille auf der Schwankungsbreite der Strahlung eines handelsüblichen Röhrenmonitors.

In einer einfaktoriellen univariaten Kovarianzanalyse zeigte sich über alle Bilder ein hoch signifikanter Effekt des Reizes auf die Pupillenwerte von  $\eta^2 = .037$ ;  $p < .000$ . Der Effekt des Lichtes auf die Pupillenwerte wurde mit  $\eta^2 = .285$ ;  $p < .000$  ebenfalls hoch signifikant. Diese Ergebnisse sollen hier nicht weiter vertieft werden, da sie bereits im Kapitel 3.2.6 ausführlich besprochen wurden.

Für die Fragestellung ist hingegen folgender Befund von Bedeutung: Im Bereich der hellen Bildvorlagen, die bei den Versuchspersonen zu kleineren Pupillengrößen führten, liegt der Effekt des mentalen Reizes auf die Pupille mit  $\eta^2 = .056$ ;  $p < .000$  deutlich über dem Durchschnittsniveau, wohingegen der Reizeffekt bei dunklen Vorlagen mit  $\eta^2 = .024$ ;  $p < .000$  deutlich unter dem Durchschnittswert liegt. Es ist also zu vermuten, dass der mentale Reiz auch

bei sehr dunklen Vorlagen eine Pupillenweitung hervorruft, diese aber geringer ausfällt, als bei dunklen Vorlagen, also lichtbedingt bereits geweiteten Pupillen. Dieser Interaktionseffekt nimmt die Stärke von  $\eta^2 = .012$ ;  $p < .000$  an.

Ein weiterer Effekt auf die Pupillengröße konnte über die Variable *Mediennutzung* aufgedeckt werden. Der mit  $\eta^2 = .027$ ;  $p < .000$  hoch signifikant ausfallende Effekt zeigt, dass Personen mit größerer Mediennutzungsdauer auch geringer auf die reizhaltigen Stimuli reagieren. Bei der Beachtung der Mediennutzung als Kovariate stieg somit der Effekt des mentalen Reizes auf die Pupille auf  $\eta^2 = .040$ ;  $p < .000$  an.

Vor der statistischen Hypothesenprüfung werden die Daten gemäß Kapitel 2.2 mit einer Baseline versehen. Betrachtet man nun die Ergebnisse in Hinblick auf die eingangs formulierten Hypothesen, so kann Folgendes festgehalten werden. Die Prüfung homogener Untergruppen mit dem Tukey-B Test, dessen Ergebnisse im Anhang dargestellt sind, bestätigt die eingangs formulierten Hypothesen. Im Einzelnen bedeutet dies:

**Hypothese 1:** Der Mittelwert der Pupillendurchmesser bei reizhaltigen hellen Vorlagen ( $\mu_{RV\_h}$ ) ist mit  $M=109,02px$  ( $SD=62,0$ ) kleiner als der Mittelwert bei reizhaltigen dunklen Vorlagen ( $\mu_{RV\_d}$ ) mit  $M=197,59px$  ( $SD=96,35$ ). Somit kann die  $H_0$  verworfen werden und die  $H_1$  wird angenommen.

**Hypothese 2:** Der Mittelwert der Pupillendurchmesser bei reizarmen hellen Vorlagen ( $\mu_{ra\_h}$ ) ist mit  $M=72,99px$  ( $SD=64,93$ ) kleiner als der Mittelwert bei reizarmen dunklen Vorlagen ( $\mu_{ra\_d}$ ) mit  $M=172,93px$  ( $SD=78,91$ ),  $p < .01$ . Somit kann die  $H_0$  verworfen werden und die  $H_1$  wird angenommen.

**Hypothese 3:** Der Mittelwert der Pupillendurchmesser bei hellem und wenig reizhaltigem visuellen Material ( $\mu_{ra\_h}$ ) ist mit  $M=72,99$ px ( $SD=64,93$ ) geringer als bei reizhaltigem hellem Material ( $\mu_{rv\_h}$ ) mit  $M=109,02$ px ( $SD=62,0$ ),  $p<.01$ . Somit kann die  $H_0$  verworfen werden und die  $H_1$  wird angenommen.

**Hypothese 4:** Der Mittelwert der Pupillendurchmesser bei dunklem und wenig reizhaltigem visuellen Material ( $\mu_{ra\_d}$ ) ist mit  $M=172,93$  px ( $SD=78,91$ ) geringer als bei reizhaltigem dunklen Material ( $\mu_{rv\_d}$ ) mit  $M=197,59$  px ( $SD=96,35$ ),  $p<.01$ . Somit kann die  $H_0$  verworfen werden und die  $H_1$  wird angenommen.

**Hypothese 5:** Der Mittelwert der Pupillendurchmesser bei reizarmen hellen Vorlagen ( $\mu_{ra\_h}$ ) ist mit  $M=72,99$  px ( $SD=64,93$ ) kleiner als der Mittelwert der Pupillendurchmesser bei reizhaltigen dunklen Vorlagen ( $\mu_{rv\_d}$ ) mit  $M=197,59$  px ( $SD=96,35$ ),  $p<.01$ . Somit kann die  $H_0$  verworfen werden und die  $H_1$  wird angenommen.

### ***3.3.6 Zusammenfassung***

In der Zusammenfassung der Analysen aus dem Experiment 3 kann an dieser Stelle festgehalten werden, dass die eingangs formulierte Hypothese bestätigt wird. Unabhängig von der Monitorabstrahlung konnten Effekte der mentalen Stimuli unter allen untersuchten Bedingungen festgestellt werden. Insbesondere konnte im verwendeten Setting kein Deckeneffekt für besonders dunkle Beleuchtungsbedingungen festgestellt werden. Obwohl die Pupille bereits

durch das fehlende Licht besonders groß war, konnte der psychologische Reiz eine weitere Vergrößerung provozieren. Einschränkend muss aber auch erwähnt werden, dass die Effekte der psychologischen Stimuli unter den lichtarmen Bedingungen kleiner ausfielen, als die Effekte unter den hellen Bedingungen. Insofern könnte die Datenbasis durch die Einführung eines lichtbedingten Korrekturfaktors verbessert werden, d. h. in dunkler Umgebung gemessene Effekte müssten höher gewichtet werden, als die Effekte, welche in hellen Umgebungen erfasst werden. Um diesen Gewichtungsfaktor näher spezifizieren zu können, müssten allerdings umfangreiche systematische Untersuchungen durchgeführt werden, die im Rahmen dieser Arbeit nicht erfolgen können. Da das pupillometrische Messverfahren allerdings auch ohne diese Effektkorrektur gut verwertbare Daten liefert, wird auf weitere Untersuchungen in dieser Richtung vorerst verzichtet.

Als interessantes Ergebnis ist auch der Zusammenhang zwischen Medienkonsum und Pupillenreaktion von  $r = -.175$ ,  $p < .01$  anzusehen. Hier konnte beobachtet werden, dass Personen weniger heftige Pupillenreaktionen zeigten, wenn sie einen hohen Medienkonsum haben. Es scheint hier einen Gewöhnungseffekt zu geben, der pupillometrisch erfassbar ist. Durch diese Gewöhnung an mediale Darstellungen fallen die Aufmerksamkeit und die Verarbeitungstiefe (Involvement) der Rezipienten offensichtlich geringer aus, als bei Personen, die nur einen geringen Medienkonsum angeben. Erklärbar ist dieser Effekt mit dem *Elaboration-Likelihood-Model* von Petty und Cacioppo (1983). Hiernach ist das *Involvement* für den Grad der Informationsverarbeitung entscheidend. Bei hohem Medienkonsum ist die Wahrscheinlichkeit höher, vergleichbare Inhalte bereits mehrfach rezipiert zu haben. In diesem Fall würde eine geringere Aufmerksamkeit ausschlaggebend sein. Nach Schenk et al. (1990) ist die Aufmerksamkeit die Vorstufe der Wahrnehmung. Sie entscheidet über die Reizauswahl und die Verarbeitungstiefe und geht mit einer körperlichen Aktivierung einher (Kroeber-Riel & Weinberg, 1999).

### **3.4. Experiment 4: Kognitiver Fähigkeitstest (KFT)**

Das Untersuchungsziel bei dem Experiment 4 war die Erfassung und Registrierung der Pupillenreaktion, die auf kognitive Beanspruchung zurückzuführen ist. Diese mentale Beanspruchung sollte durch eine kognitive Belastung der Probanden kontrolliert und abgestuft erzeugt werden. Als Stimulusmaterial diente der Kognitive Fähigkeitstest KFT 4-12 + R (Heller & Perleth, 2000). Hierbei sollte der Fragestellung nachgegangen werden, ob sich die Pupillengrößen beim Bearbeiten einfacher und schwieriger Aufgaben am Monitor jeweils signifikant unterscheiden. Aus den Ergebnissen bisheriger Untersuchungen lässt sich nach gängiger Literaturmeinung (s. Abschnitt I dieser Arbeit) ableiten, dass die Pupillenweite bei der Bearbeitung schwieriger Aufgaben größer sein sollte, als bei der Bearbeitung einfacher Aufgaben. Die Lichtverhältnisse wurden während des Experiments auf zwei Stufen variiert. Die Beleuchtungseffekte wurden für bestimmte statistische Verfahren, welche in Kapitel 3.4.5 Statistische Auswertung) näher beschrieben sind, durch die Verwendung des in Kapitel 2. hergeleiteten Verfahrens bereinigt.

#### ***3.4.1 Der Intelligenztest KFT***

Der KFT steht in der Tradition des Lorge-Thorndike Intelligence Tests (Lorge, Thorndike & Hagen, 1964) bzw. des Cognitive Abilities Tests von Thorndike und Hagen (Thorndike & Hagen, 1971) und ist ein differenzieller Intelligenztest zur Ermittlung der kognitiven Ausstattung von Schülern. Der Test besteht aus einem verbalen, einem nonverbalen und einem qualitativen Teil. Jeder Teil besteht wiederum aus 3 Subtests. Für die Bearbeitung der Tests liegen



zeitliche Beschränkungen vor, so dass der KFT als kombinierter Power-Speed-Test anzusehen ist. Die Autoren Heller und Perleth (2000) legen den Schwerpunkt jedoch eindeutig auf den Bereich *Power*. In der hier verwendeten Form dient der Test der differenziellen Bestimmung kognitiver Fähigkeitsdimensionen. Folgende Fähigkeitsbereiche werden mit dem Test erfasst:

- Sprachverständnis
- Sprachgebundenes Denken
- Arithmetisches Denken
- Rechenfähigkeit
- Anschauungsgebundenes Denken
- Konstruktive Fähigkeiten

Da der KFT in seinem Umfang für die im Experiment zu beobachtenden Effekte zu lang erscheint, wurde auf seine Kurzform zurückgegriffen. Es handelt sich hierbei um eine zeitsparende Variante, für die jedoch keine gesonderte Normierung vorliegt. Es wird aber, gerade in Bezug auf eine Computernutzung, auf die Rasch-Skalierbarkeit hingewiesen. Die meisten Items der Subtests werden als Rasch-homogen angesehen.

Für das hier beschriebene Experiment wurden die Subskalen *Nonverbaler Teil* (N1) sowie *Quantitativer Teil* (Q1 und Q2) verwendet. Von der Verwendung des verbalen Teils wurde abgesehen, da das Sprechen zu Bewegungen des Kopfes führen kann. Diese Bewegungen können Messungen mit den vorhandenen Geräten deutlich verschlechtern oder auch unmöglich machen.

Da die verwendeten Subskalen nur in Teilen eingesetzt wurden, sind die Angaben zur Reliabilität, zur inneren Konsistenz, zur Normierung und zur Validität nicht mehr anwendbar. Aus diesem Grund wurde die Aufgabenschwierigkeit zusätzlich durch Kontrollvariablen (Befra-

gung der Versuchspersonen) validiert. Als Kontrollvariablen wurden zusätzlich die Dauer der Aufgabenbearbeitung und die gemachten Fehler erhoben. Die ausgewählten Aufgaben aus dem KFT wurden für das Experiment als Computervorlage erstellt.

### ***3.4.2 Hypothesen***

Im Rahmen dieses Experiments wurden 5 Hypothesen geprüft. Mittels der Haupthypothese wurde geprüft, ob sich die Pupillengröße bei der Bearbeitung von Aufgabenreihen des KFT, welche eine höhere mentale Beanspruchung hervorrufen, von der Pupillengröße bei der Bearbeitung von Aufgabenreihen unterscheiden, welche eine geringe Beanspruchung hervorrufen.

Um die mentale Beanspruchung in den Aufgabenreihen trennen zu können wurde des Weiteren die Hypothese geprüft, dass die Bearbeitung von gering belastenden Aufgaben weniger Zeit in Anspruch nimmt, als die Bearbeitung belastender Aufgaben. Korrespondierend hierzu wurde die Hypothese geprüft, dass die Lösungswahrscheinlichkeit gering belastender Aufgaben höher ist, als die Lösungswahrscheinlichkeit belastender Aufgaben. Hieraus kann die aufgabenevozierte die mentale Beanspruchung abgeleitet werden.

Zur Kontrolle wurden darüber hinaus die Hypothesen geprüft, dass sich die Lösung heller Aufgabenvorlagen weder in der Zeit noch in der Lösungswahrscheinlichkeit von der Lösung dunkler Aufgabenvorlagen unterscheidet.

### 3.4.3 Versuchsdurchführung

Jede Versuchsperson wurde unter den gleichen kontrollierten Bedingungen untersucht. Das Experiment beinhaltete die Testbatterien *Nonverbaler Teil* (N1), *Quantitative Tests 1* (Q1) und *Quantitative Tests 2* (Q2). Den Einstieg zu jeder Messung bildete die Instruktion der Subskala, die durch ein Aufgabenbeispiel ergänzt wurde. Danach hatten die Versuchspersonen Gelegenheit, weitere Fragen an den Versuchsleiter zu stellen. Nach der Klärung offener Fragen begann die Messung mit der Präsentation der Aufgaben.

Die präsentierten Aufgaben waren nach Klassenstufen sortiert (Messwiederholungsdesign). In die Auswertung wurden die Klassenstufen 6 (leichte Aufgaben) und 12-13 (schwierige Aufgaben) übernommen. Die Zwischenstufe (Klassenstufe 9) konnte aus messtechnischen Gründen leider nicht berücksichtigt werden.

**Tabelle 10: Variablenbezeichnung**

<b>Variable</b>	<b>Name</b>	<b>Niveau</b>	<b>Messbezeichnung</b>
N1-6	Nonverbale Tests 1	Klasse 6	set0
N1-12	Nonverbale Tests 1	Klasse 12/13	set2
Q1-6	Quantitative Tests 1	Klasse 6	set3
Q1-12	Quantitative Tests 1	Klasse 12/13	set5
Q2-6	Quantitative Tests 2	Klasse 6	set6
Q2-12	Quantitative Tests 2	Klasse 12/13	set8

Pro Subskala waren 9 Aufgaben zu bearbeiten, welche die Versuchspersonen in zwei Gruppen mit systematisch über die beiden Helligkeitsunterschiede variierten Vorlagen bearbeiteten.

Um Ermüdungseffekte auszuschließen, wurde die Aufgabenreihenfolge zusätzlich variiert.

Für die Auswertung wurden die Daten später aggregiert.

Vor der Auswertung wurden die Daten - wie in Kapitel 3.1 beschrieben - bereinigt. Eine Lichtbereinigung wurde nur teilweise durchgeführt, da die Lichtverhältnisse in dem Experiment konstant über zwei Stufen gehalten wurden.

#### ***3.4.4 Stichprobenbeschreibung***

Die Stichprobe des Experiments setzte sich aus N=32 Versuchspersonen zusammen, die auf dem Campus Lankwitz der FU Berlin durch Ansprache oder über Aushänge im Fachbereich akquiriert wurden. Die Teilnahme an dem Experiment war also freiwillig. Die Zufallsstich-

probe gliederte sich in 23 weibliche und 9 männliche Probanden. Das Durchschnittsalter lag bei 25,88 Jahren (SD 4,96). Bei zwei Probanden musste die Messung abgebrochen werden, da erhebliche Probleme mit der Messtechnik auftraten.

### **3.4.5 Statistische Auswertung**

Die zu untersuchende Fragestellung, ob die Pupillen bei größerer kognitiver Belastung im Mittel auch einen signifikant größeren Durchmesser aufweisen, kann als Hypothese folgendermaßen formuliert werden (die statistischen Hypothesen finden sich im Anhang):

Der Mittelwert der Pupillengröße ist bei der Lösung schwieriger Aufgaben der Subskala N1 signifikant größer als bei der Lösung leichter Aufgaben dieser Skala. N1-6 bezeichnet die Subskala *Nonverbale Tests* im Aufgabenniveau der Klassenstufe 6 und N1-12/13 bezeichnet das Aufgabenniveau der Klassenstufe 12/13. Das Kriterium zum Verwerfen der  $H_0$  ist hierbei mit einem  $\alpha$ -Fehlerniveau von  $p < .05$  festgelegt. Die statistische Absicherung der Ergebnisse auf dem 5 Prozent- Niveau ist auch bei den weiteren Hypothesen aus dem KFT-Test vorgesehen.

Analog zu den Hypothesen der Aufgaben N1 werden auch die Subskalen der quantitativen Tests mit Mengenvergleichen und Zahlenreihen (Q1 und Q2) geprüft.

Zur Überprüfung der Hypothesen wurden folgende Variablen erfasst:

- Pupillengröße,
- objektivierte Schwierigkeit der Aufgaben nach Klassenstufen (aus KFT),

- subjektive Einschätzung der Aufgabenschwierigkeit (Kontrollvariable),
- Dauer der Aufgabenbearbeitung (Kontrollvariable),
- Fehler bei der Lösung der Aufgaben (Kontrollvariable).

Die Operationalisierung der Variablen *Objektivierte Schwierigkeit der Aufgaben nach Klassenstufen* als Hauptaufgabenmaß wurde der Skala des KFT entnommen. Die Variable *Subjektive Einschätzung der Aufgabenschwierigkeit* wurde auf einer dreistufigen Skala durch Befragung der Versuchspersonen im direkten Anschluss an die Aufgabenbearbeitung erhoben. Die Dauer der Aufgabenbearbeitung wurde bei der Aufzeichnung der Pupillenbewegung automatisch registriert.

Die Auswertung der Daten erfolgt durch Mittelwertsvergleiche (einfaktorielle Varianzanalyse der Einzelpersonen und T-Tests der aggregierten Daten). Die Ergebnisse der Varianzanalysen der Einzelpersonen sind aus Gründen der Übersichtlichkeit dem Anhang beigelegt. Da bei der Aggregation von Prozessdaten die Gefahr der Varianzaufhebung besteht (Schmitz, 1989; Suckfüll, 2004), wurden für jede Versuchsperson Einzelauswertungen durchgeführt.

Als Voraussetzung für die Verwendung des T-Tests bei abhängigen Stichproben gelten nach Bortz (1993) die Annahme der Normalverteilung sowie die Annahme der Varianzhomogenität in den Stichproben. Die Normalverteilungsvoraussetzung wurde mit dem Kolmogorov-Smirnov-Test geprüft. Die Prüfung der Varianzhomogenität erfolgte über den Levene-Test.

Bei den Einzelvergleichen in den Varianzanalysen waren einige sehr kleine Werte auffällig, die in den Klassenstufen 12/13 auftraten. Es ist leider nicht interpretierbar, ob diese aufgrund von Überforderungen der Probanden zustande kamen und in der Folge versucht wurde die Ergebnisse zu erraten oder ob die Motivation zur Aufgabenbearbeitung nicht ausreichte und

somit keine tiefe Verarbeitung stattfand. Beide Erklärungen könnten die abflachenden Pupillengrößen erklären - ohne der Literaturmeinung zu widersprechen, die einen eindeutigen Zusammenhang zwischen der Pupillengröße und der mentalen Arbeitslast postuliert.

Um fehlerhafte Ergebnisse aufgrund dieser nicht weiter eingrenzbarer Faktoren ausschließen zu können, wurden diejenigen Messergebnisse bei der Auswertung ausgeschlossen, die bei Aufgabenbearbeitungen mit über 60 Prozent Ergebnisfehlern abgenommen wurden. In diese Kategorie fielen insbesondere jene Fälle, in denen Aufgaben gar nicht erst bearbeitet wurden. Sämtliche Datensätze mit kleineren Fehlerprozenten wurden in die Auswertung übernommen, da hierbei sichergestellt war, dass die angenommene kognitive Belastungssituation bei den Versuchspersonen auch zu einer mentalen Beanspruchung i. S. d. Belastungs-Beanspruchungs-Konzepts nach Rohmert (1984) geführt hatte.

In folgender Auflistung (Tabelle 11) sind die ausgeschlossenen Datensätze aufgeführt. Sie betreffen für die Subskala *Nonverbale Tests* die Versuchspersonen 007 und 111, und für die *Quantitativen Tests 2* die Versuchspersonen 007, 008, 106, 109, 112 und 113.

**Tabelle 11: Datenausschluss aufgrund hoher Fehleranteile**

Messreihe	N1	Q1	Q2
Ausgeschlossene Daten (VP/set)	2	0	6

Die Variablen *Dauer der Aufgabenlösung*, *Subjektive Aufgabenschwierigkeit* und *Fehler bei der Aufgabenlösung* wurden als Kontrollvariablen erhoben. Da die einzelnen Subskalen des

KFT nicht genau übernommen werden konnten, ist eine Gegenprüfung über diese Kontrollvariablen zur Validierung der Schwierigkeitsstufen angebracht.

Sollten die Subskalen des KFT tatsächlich gestufte Schwierigkeitsgrade abbilden, so müssten folgende Annahmen zutreffen:

- Die Bearbeitungszeit ( $t$ ) der schwierigen Aufgaben (gem. den Klassenstufen des KFT) übersteigt die Bearbeitungszeit einfacher Aufgaben bei gleicher Aufgabenanzahl.
- Die Lösungswahrscheinlichkeit ist bei schwierigen Aufgaben kleiner als bei leichten Aufgaben, da die Menge der Fehler bei schwierigen Aufgaben größer ist als bei einfachen Aufgaben.

Ferner wurde in dem Experiment geprüft, ob sich Unterschiede bei der Lösungswahrscheinlichkeit und der Dauer der Aufgabenbearbeitung zwischen der Bearbeitung der dunklen und der hellen Vorlagen ergeben. Durch dieses Vorgehen soll ausgeschlossen werden, dass die Pupillenreaktionen, die auf die mentale Beanspruchung zurückgehen, von der Helligkeit der Vorlage beeinflusst werden. Wäre es z. B. schwieriger, eine dunkle Vorlage zu bearbeiten, so würde dies (neben der Lichtreaktion) zu einer Pupillenweitung führen, die nicht lichtbedingt ist, und das Ergebnis dieses Experiments würde verzerrt werden. Hieraus leiten sich folgende Kontrollhypothesen ab:

- Die Bearbeitungszeit der dunklen Aufgabenvorlagen unterscheidet sich nicht signifikant von der Bearbeitungszeit der hellen Aufgabenvorlagen.
- Die Lösungswahrscheinlichkeit der dunklen Aufgabenvorlagen unterscheidet sich nicht signifikant von der Lösungswahrscheinlichkeit der hellen Aufgabenvorlagen.



Die Ergebnisse der statistischen Überprüfung dieser Hypothesen werden im folgenden Kapitel beschrieben und dargestellt.

### 3.4.6 Ergebnisse

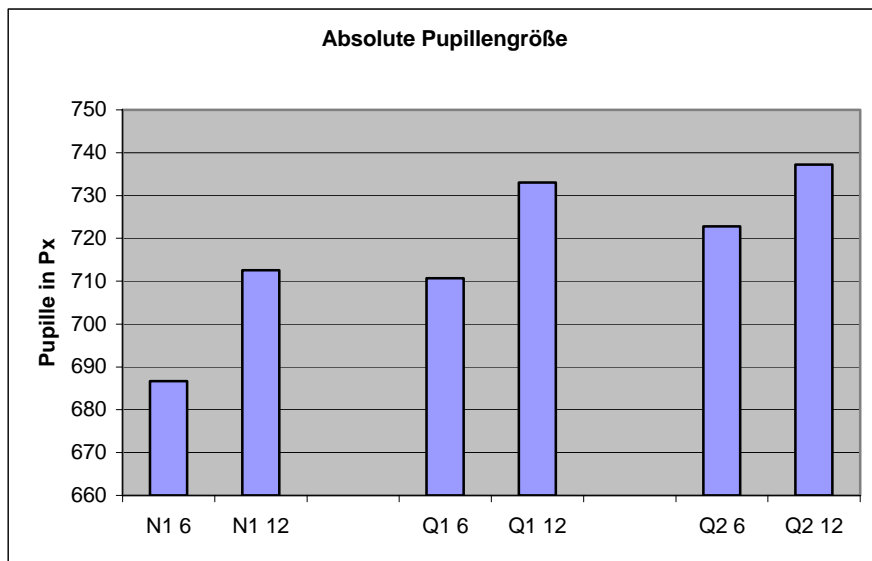
In der folgenden Ergebnisdarstellung sind die Variablen nach den Set-Nummern der Messungen benannt, set0 steht etwa für die Subskala *Nonverbale Tests* (N1) auf dem Klassenstufenniveau der 6. Klasse und set2 auf dem der 12./13. Klasse. Die Variablen set3 und set5 stehen für die *Quantitativen Tests 1* (Q1) in den Klassenstufen 6 und 12/13 und die Variablen set6 und set8 für die *Quantitativen Tests 2* (Q2) in den Klassenstufen 6 und 12/13.

**Tabelle 12: Variablenbezeichnung**

Set0	Set2	Set3	Set5	Set6	Set8
N1-6	N1-12/13	Q1-6	Q1-12/13	Q2-6	Q2 12/13

#### 3.4.6.1 Hypothesenprüfung 1.1 bis 1.3

Die ausformulierten statistischen Hypothesen finden sich im Anhang zu diesem und den folgenden Kapiteln. Die Überprüfung der Hypothese, dass die mittlere Pupillenausdehnung bei der Lösung schwierigerer Aufgaben die mittlere Pupillengröße bei der Lösung einfacherer Aufgaben signifikant übersteigt erbrachte Ergebnisse, wie sie im Diagramm 8 dargestellt sind.



**Diagramm 8: Absolute Pupillengrößen**

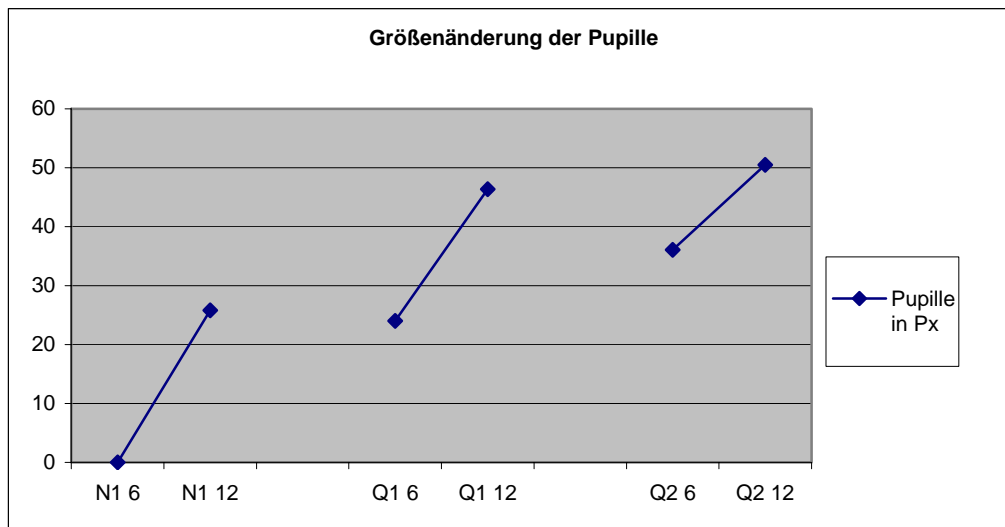
Wie Diagramm 8 zeigt, existieren Unterschiede gemäß der aufgestellten Hypothese, d. h. die Pupillendurchmesser von Probanden bei der Bearbeitung von Aufgabenvorlagen der Klassenstufe 12/13 übersteigen in allen Subskalen die Pupillendurchmesser in der Niveaustufe 6.

**Tabelle 13: Pupillengrößen der Aufgabenpaare**

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	N1_6	712,9100	28	200,04268	37,80451
	N1_12	739,2868	28	212,39891	40,13962
Paaren 2	Q1_6	723,8417	30	177,94286	32,48777
	Q1_12	758,3673	30	191,80668	35,01895
Paaren 3	Q2_6	719,4013	24	230,60697	47,07245
	Q2_12	739,3192	24	248,69100	50,76384

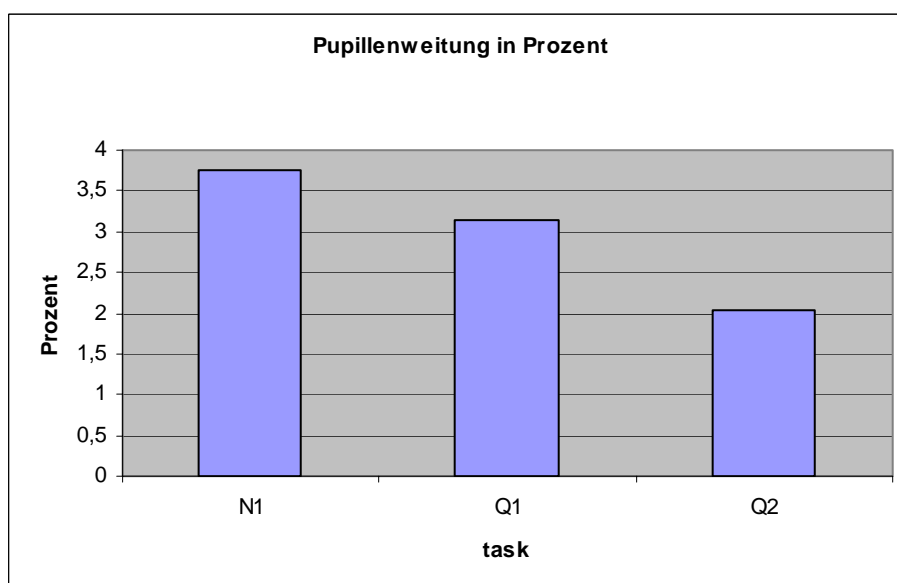
Die Paare 1 in Tabelle 13 stehen hierbei für die Hypothese 1, die Paare 2 für die Hypothese 2 und die Paare 3 für die Hypothese 3. Die Mittelwerte der Pupillenausdehnung während der Aufgabenbearbeitung auf dem Niveau der Klassenstufe 6 liegt regelmäßig unter der mittleren Ausdehnung der Pupille während der Aufgabenbearbeitung auf dem Niveau der Klassenstufe

12/13. Die absolute Differenz der harmonisierten und bereinigten Pupillengröße in px/16 zeigt das Diagramm 9.



**Diagramm 9: Differenz der Pupillengröße zwischen den Klassenstufen 6 und 12/13**

Diese absolute Differenz der Pupillengröße kann auch als prozentuale Veränderung angegeben werden. Die prozentuale Veränderung der Pupillengröße vom Anspruchsniveau der Klassenstufe 6 zum Anspruchsniveau der Klassenstufe 12/13 zeigt das Diagramm 10.



**Diagramm 10: Prozentuale Abweichung der Pupillengröße zwischen dem Aufgabenniveau 1 (Klassenstufe 6) und dem Aufgabenniveau 2 (Klassenstufe 12/13)**

Die deskriptive Auswertung des Experiments KFT ergibt einen eindeutigen Trend der Pupillenerweiterung bei der Bearbeitung schwierigerer Aufgaben. Die Pupillenweitung beträgt bei den Nonverbalen Tests 1 im Mittel 3,76 Prozent, bei den Quantitativen Tests 1 im Mittel 3,14 Prozent und bei den Quantitativen Tests 2 im Mittel 2,03 Prozent.

Zur statistischen Absicherung dieses deskriptiven Ergebnisses der Mittelwertunterschiede wurde ein T-Test für abhängige Stichproben berechnet. Als Voraussetzung muss hier die Normalverteilung angenommen werden und die Stichprobenpaare sollten miteinander korrelieren (Tabellen siehe Anhang).

Da die Hypothesen 1.1 -1.3 (analog repräsentiert durch die Paare 1-3) eingangs einseitig formuliert wurden, muss in der Berechnung der Signifikanz der Mittelwertsunterschiede der Wert  $p$  der  $\alpha$ -Fehlerwahrscheinlichkeit halbiert werden (Diehl & Arbinger, 2001).

Es zeigt sich also im Ergebnis des T-Tests ein sehr signifikanter Unterschied des Mittelwerts N1-6 (set0) von dem Mittelwert N1-12/13 (set2). Der Mittelwert von set2 übersteigt dabei den Mittelwert von set0 um 37,4 px/16. Dieses Ergebnis  $T(27)=-3,728$  ist mit  $p(\text{einseitig})<.01$  sehr signifikant.

Zu 1.1) Die  $H_0$  ( $\mu_{N1-6} = \mu_{N1-12/13}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_1$  ( $\mu_{N1-6} < \mu_{N1-12/13}$ ) wird angenommen.

Der Mittelwert Pupillengröße bei der Bearbeitung der Aufgaben Q1-12/13 (set5) übersteigt den Mittelwert der Pupille bei der Bearbeitung der Aufgaben Q1-6 (set3) um 66,22 px/16. Dieses Ergebnis ist mit  $T(29)=- 2,856$  ist mit  $p(\text{einseitig})= .004$  sehr signifikant.

Zu 1.2) Die  $H_0$  ( $\mu_{Q1-6} = \mu_{Q1-12/13}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_1$  ( $\mu_{Q1-6} < \mu_{Q1-12/13}$ ) wird angenommen.

Der Mittelwert Pupillengröße bei der Bearbeitung der Aufgaben Q2-12/13 (set8) übersteigt den Mittelwert der Pupille im bei der Bearbeitung der Aufgaben Q2-6 (set6) um 34,54 px/16. Dieses Ergebnis ist mit  $T(23)=-2,825$  ist mit  $p(\text{einseitig})= .05$  signifikant.

Zu 1.3) Die  $H_0$  ( $\mu_{Q2-6} = \mu_{Q2-12/13}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_1$  ( $\mu_{Q2-6} < \mu_{Q2-12/13}$ ) wird angenommen.

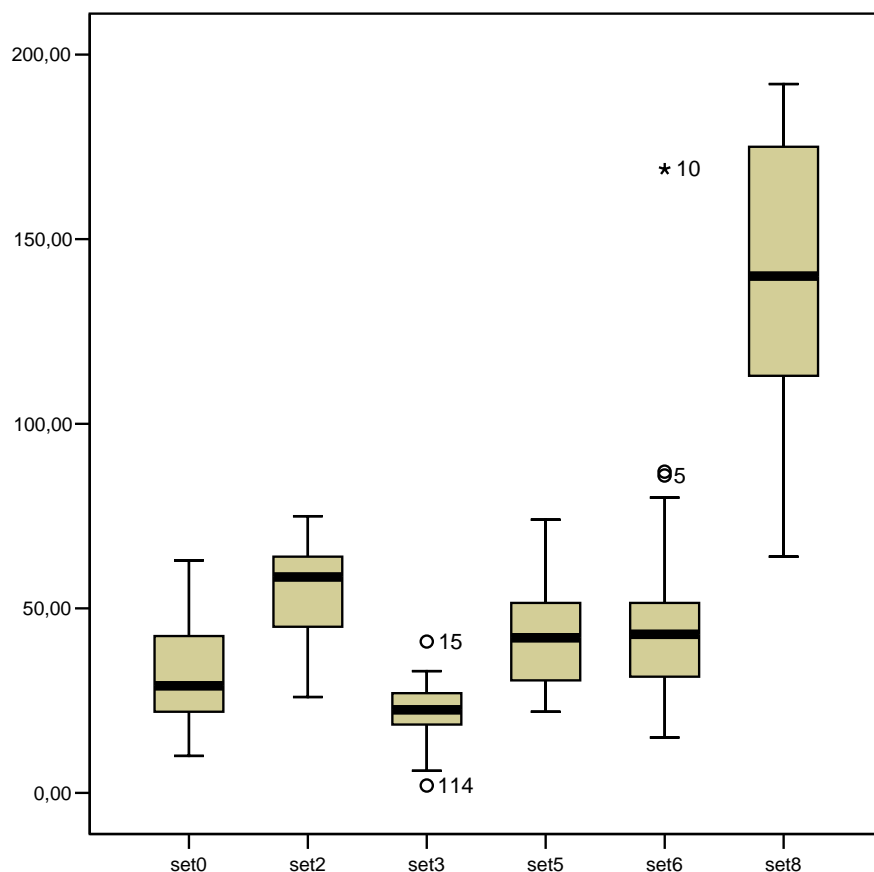
#### 3.4.6.2 Hypothesenprüfung 2.1 bis 2.3

Um die Validität der im Kapitel 3.4.6.1 beschriebenen Ergebnisse abzusichern, ist eine genaue Betrachtung der Variablen *objektivierte Schwierigkeit* notwendig. Die Skala (Klassenstufe) basiert auf dem Intelligenztest KFT. Die Subskalen konnten jedoch nicht absolut identisch auf das Experiment übertragen werden, womit die Normierung der Skala nicht mehr gegeben ist. Dies ist in dem hier beschriebenen Experiment nicht von großer Bedeutung, da hier lediglich sichergestellt sein musste, dass sich die Schwierigkeitsstufen 6 und 12/13 deutlich voneinander unterscheiden. Hierfür wurden die drei Kontrollvariablen *Dauer der Aufgabebearbeitung*, *subjektive Einschätzung der Aufgabenschwierigkeit* und *Fehler bei der Lösung der Aufgaben* erhoben. In diesem Kapitel wird die Dauer der Aufgabenlösung genauer betrachtet. Nach Iqbal, Zheng und Bailey (2004) benötigen Computernutzer für schwierigere

Aufgaben auch mehr Zeit als für leichte Aufgaben. Insofern ist die Zeit, die zur Lösung einer Aufgabe aufgewendet werden muss, ein Indikator für die Aufgabenschwierigkeit. Diese Ansicht ist jedoch nicht unbedingt mit einem Ressourcenmodell der mentalen Beanspruchung in Einklang zu bringen. Gerade die kurze Dauer einer Aufgabenbearbeitung kann danach zu einer sehr viel höheren Arbeitslast führen als die langsame Bearbeitung einer Aufgabe. Da in den hier beschriebenen Experimenten jedoch keine Zeitvorgabe festgelegt wurde, kann davon ausgegangen werden, dass die Probanden alle Aufgaben mit vergleichbarer Intensität zu lösen versuchten. Wahrscheinlich hat es im vorhandenen Setting kaum Effekte der Geschwindigkeit auf die Arbeitslast gegeben. Die vorhandene Varianz Bearbeitungsdauer von Aufgaben geht somit eher auf die Varianz der Schwierigkeit der Aufgaben zurück. Dies wird auch durch die hoch signifikante Korrelation der Dauer mit der subjektiven Schwierigkeit der Aufgabenlösung von  $r = .829$  deutlich (vgl. Kapitel 3.4.6).

Die Dauer der Aufgabenbearbeitung erscheint somit zwar aus theoretischer Sicht kein idealer Indikator für die Aufgabenschwierigkeit zu sein, in dem vorhandenen Setting kann sie aber dennoch brauchbare Daten liefern.

Im folgenden Box-Plot Diagramm werden die Zeitunterschiede (in Sekunden) explorativ dargestellt, die die Versuchspersonen zur Aufgabenlösung benötigten.



**Diagramm 11: Lösungsdauer der Aufgaben in Sekunden**

Die explorative Datenauswertung zeigt erwartungsgemäße Unterschiede in der Bearbeitungszeit der Aufgaben zwischen den Klassenstufen 6 und 12/13. Zur Bearbeitung der Nonverbalen Tests 1 auf der Klassenstufe 6 (N1-6) wurden hierbei durchschnittlich 32 Sekunden benötigt, während die Bearbeitung der Aufgaben auf dem Niveau der Klasse 12/13 (N1-12/13) durchschnittlich 54 Sekunden in Anspruch nahm. Bei den Quantitativen Tests 1 und 2 ist ein ähnliches Ergebnis zu beobachten. Hierbei benötigten die Probanden zur Lösung der Aufgaben Q1-6 im Mittel 22 Sekunden und bei den Aufgaben Q1-12/13 im Mittel 41 Sekunden. Bei der Lösung der Aufgaben Q2-6 benötigten sie im Mittel 48 Sekunden, während sie bei den Aufgaben Q2-12/13 im Mittel 140 Sekunden benötigten. Um dieses Ergebnis statistisch gegen den Zufall abzusichern, wird dieses Ergebnis anhand eines T-Tests für abhängige Stichproben überprüft.

Die Prüfung der Korrelation der einzelnen Paare als Voraussetzung für den T-Test zeigte signifikante Ergebnisse ( $p < .05$ ) für die Paare N1-6 (set0) und N1-12/13 (set2) mit  $r = .57$  sowie für die Paare Q2-6 (set6) und Q2-12/13 (set8) mit  $r = .38$ . Die Korrelation der Paare Q1-6 (set3) und Q1-12/13 (set5) mit  $r = .24$  kann statistisch nicht signifikant abgesichert werden. Die Absicherung der Korrelation ist nach Diehl und Arbinger (2001, S. 161) die Voraussetzung zur Durchführung des T-Tests für abhängige Stichproben, sollte bei kleinen Stichproben jedoch lediglich deutlich über null sein. Da die  $H_0$  aufgrund des Ergebnisses aber verworfen wird, ist die Gefahr des  $\beta$ -Fehlers praktisch nicht relevant.

Es zeigt sich also im Ergebnis des T-Tests ein sehr signifikanter Unterschied des Mittelwerts N1-6 (set0) von dem Mittelwert N1-12/13 (set2) in Bezug auf die Dauer der Aufgabenbearbeitung. Der Mittelwert von set2 übersteigt dabei den Mittelwert von set0 um 22,52 Sekunden. Dieses Ergebnis ist mit  $T(22) = 8,43$   $p(\text{einseitig}) = .000$  sehr signifikant.

Zu 2.1) Die  $H_0$  ( $\mu_{t-N1-6} = \mu_{t-N1-12/13}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_1$  ( $\mu_{t-N1-6} < \mu_{t-N1-12/13}$ ) wird angenommen.

Der Mittelwert der Bearbeitungszeit bei der Bearbeitung der Aufgaben Q1-12/13 (set5) übersteigt den Mittelwert der Bearbeitungszeit bei der Bearbeitung der Aufgaben Q1-6 (set3) um 19,5 Sekunden. Dieses Ergebnis ist mit  $T(29) = 7,356$   $p(\text{einseitig}) = .000$  sehr signifikant.

Zu 2.2) Die  $H_0$  ( $\mu_{t-Q1-6} = \mu_{t-Q1-12/13}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_1$  ( $\mu_{t-Q1-6} < \mu_{t-Q1-12/13}$ ) wird angenommen.



Der Mittelwert der Bearbeitungsdauer der Aufgaben Q2-12/13 (set8) übersteigt den Mittelwert der Bearbeitungsdauer bei der Bearbeitung der Aufgaben Q2-6 (set6) um 91,8 Sekunden. Dieses Ergebnis ist mit  $T(29)=12,565$   $p(\text{einseitig})= .000$  sehr signifikant.

Zu 2.3) Die  $H_0$  ( $\mu_{Q2-6} = \mu_{Q2-12/13}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_1$  ( $\mu_{Q2-6} < \mu_{Q2-12/13}$ ) wird angenommen.

Es kann also im Ergebnis festgehalten werden, dass die Lösung der Aufgaben auf dem Klassenstufenniveau 6 (nach KFT) signifikant weniger Zeit beansprucht als die Lösung der Aufgaben auf dem Niveau der Klassenstufe 12/13.

#### 3.4.6.3 Hypothesenprüfung 3.1 bis 3.3

Als zweite Kontrollvariable dient die Lösungswahrscheinlichkeit bei verschiedenen Belastungsniveaus. Nach den theoretischen Vorüberlegungen und den vorherigen Ergebnissen ist ein Unterschied in der Fehlerhäufigkeit zwischen den Belastungsstufen zu vermuten. Man kann davon ausgehen, dass auf dem Klassenstufenniveau 6 signifikant weniger Fehler gemacht werden, als dies bei der Klassenstufe 12/13 der Fall ist. Die Auswertung der erhobenen Daten zeigt folgendes Ergebnis:

**Tabelle 14: Paarvergleiche der Fehlermittelwerte**

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren	false0	,7667	30	,81720	,14920
1	false2	1,9667	30	,76489	,13965
Paaren	false3	,1000	30	,30513	,05571
2	false5	1,0667	30	,73968	,13505
Paaren	false6	,2333	30	,62606	,11430
3	false8	1,5333	30	,86037	,15708

Wie der Mittelwertsvergleich in Tabelle 14 bereits anzeigt, existieren erhebliche Unterschiede in der Anzahl der Fehler über die verschiedenen Anforderungsniveaus. Zur statistischen Absicherung wird auch hier wieder ein T-Test für abhängige Stichproben berechnet. Zur Kontrolle des  $\beta$ -Fehlers wird die Korrelation der Datenpaare berechnet, die Ergebnisse finden sich im Anhang.

Die Datenpaare korrelieren zwar nicht signifikant miteinander, sie nehmen jedoch den geforderten (Diehl & Arbinger, 2001) positiven Wert über Null an. Der nachfolgende T-Test zeigt darüber hinaus, dass die Nullhypothese verworfen wird und die Kontrolle des  $\beta$ -Fehlers keine praktische Relevanz mehr hat.

Die Prüfung der Normalverteilungsvoraussetzung (s. Anhang) zeigt, dass für die Paare false3-false5 und false6-false8 keine Normalverteilung vorliegt. Aus diesem Grund wird im Folgenden der nichtparametrische Wilcoxon-Test berechnet.

Es zeigt sich im Ergebnis des T-Tests und des Wilcoxon-Tests ein sehr signifikanter Unterschied des Mittelwerts N1-6 (set0) von dem Mittelwert N1-12/13 (set2) in Bezug auf die Feh-

leranzahl. Der Mittelwert von set2 übersteigt dabei den Mittelwert von set0 um 1,2 Fehler.

Dieses Ergebnis ist mit  $T(29)=6,378$ ,  $p(\text{einseitig})=.000$  sehr signifikant.

Zu 3.1) Die  $H_0$  ( $\mu \text{ Fehler-N1-6} = \mu \text{ Fehler-N1-12/13}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_1$  ( $\mu \text{ Fehler-N1-6} < \mu \text{ Fehler-N1-12/13}$ ) wird angenommen.

Die Differenz der Mittelwerte der Fehler im Teil Q1 zeigen, dass auf der Niveaustufe Klasse 12/13 0,97 Fehler mehr gemacht wurden, als auf der Niveaustufe 6 mit nur 0,1 Fehlern. Das Ergebnis  $Z(29)=-4,553$ ,  $p(\text{einseitig})=.000$  ist sehr signifikant.

Zu 3.2) Die  $H_0$  ( $\mu \text{ Fehler-Q1-6} = \mu \text{ Fehler-Q1-12/13}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_1$  ( $\mu \text{ Fehler-Q1-6} < \mu \text{ Fehler-Q1-12/13}$ ) wird angenommen.

In der Subskala Q2 zeigt sich eine Mittelwertsdifferenz von 1,3 Fehlern, die auf der Niveaustufe 12/13 mehr gemacht wurden. Das Ergebnis  $Z(29)=-4,782$  ist mit  $p(\text{einseitig})=.000$  sehr signifikant.

Zu 3.3) Die  $H_0$  ( $\mu \text{ Fehler-Q2-6} = \mu \text{ Fehler-Q2-12/13}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_1$  ( $\mu \text{ Fehler-Q2-6} < \mu \text{ Fehler-Q2-12/13}$ ) wird angenommen.

Im Ergebnis kann hier festgehalten werden, dass im Bereich der Klassenstufen (nach KFT) 12/13 signifikant mehr Fehler gemacht wurden, als dies in der Klassenstufe 6 der Fall war.

### 3.4.6.4 Hypothesenprüfung 4.1 bis 4.3

Damit ein eventuell auftretender verzerrender Effekt durch verschiedene Belastungsniveaus heller und dunkler Vorlagen ausgeschlossen werden kann, werden nun die Hypothesen geprüft, die den Unterschied in den hellen und dunklen Vorlagen in Bezug auf die Lösungswahrscheinlichkeit und die Dauer der Aufgabenlösung beschreiben.

Die Bearbeitungszeit der hellen und dunklen Vorlagen in den beiden Versuchsgruppen verteilt sich wie folgt:

**Tabelle 15: Mittelwerte der Aufgabenbearbeitung in Sekunden**

	Group	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
set0	1	15	27,9333	12,00278	3,09910
	2	13	36,5385	13,36423	3,70657
set2	1	15	52,8000	12,42808	3,20892
	2	13	54,4615	16,06557	4,45579
set3	1	15	21,4000	8,06934	2,08350
	2	15	22,2000	8,84146	2,28286
set5	1	15	40,0667	12,10942	3,12664
	2	15	42,5333	13,34095	3,44462
set6	1	15	53,5333	38,32542	9,89558
	2	15	42,2000	16,15638	4,17156
set8	1	15	143,2667	36,01680	9,29950
	2	15	136,0667	45,67973	11,79446

Die Helligkeitsverteilung variiert dabei folgendermaßen (Schwankungen der Lux-Werte liegen hierbei unter 3 Prozent):

**Tabelle 16: Variation hell/dunkel der Experimentalgruppen**

<b>Set Nr.</b>	<b>Group 1</b>	<b>Group 2</b>
<b>set0</b>	70 Lux	170 Lux
<b>set2</b>	70 Lux	170 Lux
<b>set3</b>	170 Lux	70 Lux
<b>set5</b>	170 Lux	70 Lux
<b>set6</b>	70 Lux	170 Lux
<b>set8</b>	70 Lux	170 Lux

Bereits die deskriptive Statistik zeigt lediglich geringe Unterschiede zwischen den Bearbeitungszeiten heller und dunkler Aufgaben. Die Überprüfung der Unterschiede in der Lösungszeit der beiden Gruppen mit hellen und dunklen Vorlagen wird mittels T-Test statistisch abgesichert. Der Kolmogorov-Smirnov-Test zur Prüfung der Normalverteilungsvoraussetzung ergab hierbei die Erfüllung der Normalverteilungsvoraussetzung.

Als Ergebnis kann hier festgehalten werden, dass keine signifikanten Unterschiede zwischen beiden Experimentalgruppen bestehen, was die Zeit der Aufgabenlösung für die einzelnen Subskalen betrifft. Die Helligkeit der Vorlage der Aufgaben hat somit keinen signifikanten Effekt auf die Bearbeitungsdauer. Zur statistischen Absicherung wird in jedem Fall Varianzhomogenität angenommen, da der Levene-Test in keinem Fall die Annahme der Varianzhomogenität widerlegen konnte.

In der Aufgabenreihe N1-6 zeigen sich Unterschiede in der Dauer der Aufgabenbearbeitung von 8,6 Sekunden. Die statistische Absicherung mit  $T(26)=-1,795$ ,  $p= .084$  weist hierbei auf

ein eher zufälliges Ergebnis hin. Im Aufgabenniveau 12/13 ergibt sich ein Unterschied von lediglich 1,6 Sekunden, welcher statistisch geprüft mit einer Irrtumswahrscheinlichkeit von  $p=.760$  nicht signifikant wurde. Die längere Bearbeitungszeit entfällt hierbei jeweils auf die helle Aufgabenvorlage.

Zu 4.1) Die  $H_1$  ( $\mu_{t-N1-6dunkel} \neq \mu_{t-N1-6hell}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_0$  ( $\mu_{t-N1-6dunkel} = \mu_{t-N1-6hell}$ ) wird beibehalten und die  $H_1$  ( $\mu_{t-N1-12/13dunkel} \neq \mu_{t-N1-12/13hell}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_0$  ( $\mu_{t-N1-12/13dunkel} = \mu_{t-N1-12/13hell}$ ) wird beibehalten.

In der Aufgabenreihe Q1-6 zeigen sich Unterschiede von 3,1 Sekunden in der Dauer der Aufgabenbearbeitung, die mit  $T(26)-0,308$ ,  $p=.798$  nicht signifikant wurden. In der Niveaustufe 12/13 des Tests ergab sich ein Unterschied von 4,7 Sekunden. Dieser Unterschied wurde mit  $p=.600$  ebenfalls nicht signifikant. Die längere Bearbeitungszeit entfiel hierbei auf die jeweils dunklere Aufgabenvorlage.

Zu 4.2) Die  $H_1$  ( $\mu_{t-Q1-6dunkel} \neq \mu_{t-Q1-6hell}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_0$  ( $\mu_{t-Q1-6dunkel} = \mu_{t-Q1-6hell}$ ) wird beibehalten und Die  $H_1$  ( $\mu_{t-Q1-12/13dunkel} \neq \mu_{t-Q1-12/13hell}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_0$  ( $\mu_{t-Q1-12/13dunkel} = \mu_{t-Q1-12/13hell}$ ) wird beibehalten.

Die Aufgabenreihe Q2 wies im Vergleich Unterschiede in der Bearbeitungszeit von 11,3 Sekunden im Niveau 6 und 7,2 im Niveau 12/13 auf (wobei die längeren Zeiten jeweils auf die dunklen Vorlagen entfielen). Die statistische Absicherung ergab hierbei einen  $\alpha$ -Fehlerwert von  $p=.300$  und  $p=.635$ .

Zu 4.3) Die  $H1$  ( $\mu_{t-Q2-6dunkel} \neq \mu_{t-Q2-6hell}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H0$  ( $\mu_{t-Q2-6dunkel} = \mu_{t-Q2-6hell}$ ) wird beibehalten und

Die  $H1$  ( $\mu_{t-Q2-12/13dunkel} = \mu_{t-Q2-12/13hell}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H0$  ( $\mu_{t-Q2-12/13dunkel} \neq \mu_{t-Q2-12/13hell}$ ) wird beibehalten.

Das hier gefundene Ergebnis bleibt auch erhalten, wenn die vorhandenen Daten so umgruppiert werden, dass die Faktoren dunkle Aufgabenvorlagen (weiße Schrift auf schwarzem Grund) und die hellen Aufgabenvorlagen (schwarze Schrift auf weißem Grund) zusammengefasst werden.

Die Bearbeitungszeit der Negativ-Vorlagen liegt in den zusammengefassten Datensätzen zwar über der Bearbeitungszeit der Positiv-Vorlagen (summiert  $t_d=342,27$  sec zu  $t_h=330,73$  sec), wird aber statistisch nicht signifikant. Der gefundene Effekt ist mit  $\eta^2= .16$  recht gering. Obwohl es einen zeitlichen Unterschied bei der Lösungszeit der verschiedenen Aufgabenvorlagen gibt, kann dieser in der weiteren Untersuchung vernachlässigt werden, da er gering ist und statistisch nicht signifikant wird. Weitere Untersuchungen zu diesem Thema sind im

Rahmen dieser Arbeit nicht relevant und unterbleiben aus diesem Grund. Trotzdem stellt dieser Effekt ein interessantes Ergebnis dar.

#### 3.4.6.5 Hypothesenprüfung 5.1 bis 5.3

Die Auswirkung der Helligkeit auf die Lösungswahrscheinlichkeit der Aufgaben wird überprüft, indem wiederum ein Vergleich beider Gruppen (hell/dunkel) mittels T-Test berechnet wird diesmal hierbei wird diesmal die Anzahl der Fehler verglichen wird. Folgende Befunde sind hier zu nennen, wobei die Variable *false* die Anzahl der Fehler wiedergibt:

**Tabelle 17: Fehler bei der Aufgabenlösung**

	Group	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
false0	1	15	,6000	,73679	,19024
	2	15	,9333	,88372	,22817
false2	1	15	2,0000	,65465	,16903
	2	15	1,9333	,88372	,22817
false3	1	15	,1333	,35187	,09085
	2	15	,0667	,25820	,06667
false5	1	15	1,0667	,79881	,20625
	2	15	1,0667	,70373	,18170
false6	1	15	,1333	,35187	,09085
	2	15	,3333	,81650	,21082
false8	1	15	1,6000	,73679	,19024
	2	15	1,4667	,99043	,25573

Die deskriptive Untersuchung der Daten ergibt auch hier nur geringe Abweichungen der Fehlermengen, welche die Werte 0 bis 9 annehmen können (s. 17).



Der Kolmogorov-Smirnov-Test zur Prüfung der Normalverteilungsvoraussetzung ergab hierbei die Nichterfüllung der Normalverteilungsvoraussetzung für die Variablen *false3* und *false6* (siehe Anhang). Aus diesem Grund kam für diese Variablen der Wilcoxon-Test zur Anwendung. Als Ergebnis kann festgehalten werden, dass die Lösungswahrscheinlichkeit offensichtlich nicht von der Helligkeit der Aufgabenvorlagen abhängt. Der Vergleich der Fehlerwerte ergibt in keinem Fall signifikante Unterschiede der Mittelwerte der Lösungsdauer. Varianzhomogenität wird hierbei angenommen, da der Levene-Test in keinem Fall die Varianzhomogenität widerlegt.

Die Hypothesenprüfung ergibt somit folgende Ergebnisse: In der Aufgabenreihe N1-6 zeigen sich Unterschiede in der Fehlerhäufigkeit von durchschnittlich 0,33 Fehlern, die bei hellen Aufgabenvorlagen mehr gemacht wurden. Die statistische Absicherung mit  $p = .271$  weist hierbei auf ein eher zufälliges Ergebnis hin. Im Aufgabenniveau 12/13 ergab sich ein Unterschied von 0,67 Fehlern, die diesmal im dunklen Milieu zusätzlich gemacht wurden. Statistisch geprüft wurde dieses Ergebnis mit einer Irrtumswahrscheinlichkeit von  $p = .816$  nicht signifikant.

Die  $H_1$  ( $\mu_{\text{Fehler-N1-6dunkel}} \neq \mu_{\text{Fehler-N1-6hell}}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_0$  ( $\mu_{\text{Fehler-N1-6dunkel}} = \mu_{\text{Fehler-N1-6hell}}$ ) wird beibehalten

und

die  $H_1$  ( $\mu_{\text{Fehler-N1-12/13dunkel}} \neq \mu_{\text{Fehler-N1-12/13hell}}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_0$  ( $\mu_{\text{Fehler-N1-12/13dunkel}} = \mu_{\text{Fehler-N1-12/13hell}}$ ) wird beibehalten.

In der Aufgabenreihe Q1-6 zeigten sich Unterschiede in der Fehlerhäufigkeit von (0,67 Fehlern im hellen Bereich), welche mit  $p=.775$  nicht signifikant wurde. In der Niveaustufe 12/13 des Tests ergab sich kein Unterschied ( $p=1$ ). Die längere Bearbeitungszeit entfiel hierbei auf die jeweils dunklere Aufgabenvorlage.

Die  $H_1$  ( $\mu_{\text{Fehler-Q1-6dunkel}} \neq \mu_{\text{Fehler-Q1-6hell}}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_0$  ( $\mu_{\text{Fehler-Q1-6dunkel}} = \mu_{\text{Fehler-Q1-6hell}}$ ) wird beibehalten

und

die  $H_1$  ( $\mu_{\text{Fehler-Q1-12/13dunkel}} \neq \mu_{\text{Fehler-Q1-12/13hell}}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_0$  ( $\mu_{\text{Fehler-Q1-12/13dunkel}} = \mu_{\text{Fehler-Q1-12/13hell}}$ ) wird beibehalten.

Die Aufgabenreihe Q2 wies in der Fehlerhäufigkeit im Vergleich Unterschiede von 0,2 Fehlern im Niveau 6 auf (die im dunklen Milieu mehr gemacht wurden), und von 0,14 Fehlern im Niveau 12/13 (die wiederum im dunklen Milieu mehr gemacht wurden). Die statistische Absicherung ergab hierbei einen  $\alpha$ -Fehlerwert von  $p=.774$  und  $p=.679$ .

Die  $H_1$  ( $\mu_{\text{Fehler-Q2-6dunkel}} \neq \mu_{\text{Fehler-Q2-6hell}}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die  $H_0$  ( $\mu_{\text{Fehler-Q2-6dunkel}} = \mu_{\text{Fehler-Q2-6hell}}$ ) wird beibehalten

und

Die H1 ( $\mu_{\text{Fehler-Q2-12/13dunkel}} \neq \mu_{\text{Fehler-Q2-12/13hell}}$ ) muss aufgrund dieses Ergebnisses verworfen werden und die H0 ( $\mu_{\text{Fehler-Q2-12/13dunkel}} = \mu_{\text{Fehler-Q2-12/13hell}}$ ) wird beibehalten.

Die Lösungswahrscheinlichkeit für die Aufgaben ergibt sich aus den Fehlern und den nicht gelösten Aufgaben und verteilt sich wie folgt:

Set0 74, 44 Prozent

Set2 34, 44 Prozent

Set3 96, 66 Prozent

Set5 64, 44 Prozent

Set6 92, 22 Prozent

Set8 48, 98 Prozent

Die Überprüfung der Korrelationen aus den vorhergehenden Ergebnissen zeigt dabei ein weniger deutliches Bild. Korreliert man die subjektive Einschätzung der Schwierigkeit mit der benötigten Lösungszeit pro Aufgabe im Rahmen einer Produkt-Moment-Korrelation nach Pearson, so ergibt sich eine Korrelation von  $r=.40$  ( $p<.05$ ) für die Lösungsdauer und die Schwierigkeitseinschätzung bei der Aufgabe Q1-12/13. Tendenziell signifikant wird die Korrelation bei der Aufgabe N1-6 mit  $r= .282$  ( $p=.073$ ), N1-12/13 mit  $r= .270$  ( $p=.083$ ) und Q2-12/13 mit  $r= .286$  ( $p= .062$ ). Die Korrelationen der Aufgaben Q1-6 und Q2-6 konnten statis-

tisch nicht als signifikant abgesichert werden. Bei beiden Variablen existiert jedoch auch fast keine Fehlervarianz, da die Lösungswahrscheinlichkeit in beiden Fällen deutlich über 90 Prozent liegt. Die entsprechenden Variablen sind hier also fast konstant, ein Bodeneffekt wirkt bei der Verteilung durch eine mangelhafte Leistungsdifferenzierung auf dem Niveau der 6. Klasse. Da diese Korrelationen aber lediglich der Validierung der Schwierigkeitsabstufung dienen, ist ihre Signifikanz Korrelationen unerheblich. Aussagen, die die Stichprobe betreffen, reichen hierbei völlig aus, denn die Ergebnisse des Experiments, die auf die Population übertragen werden sollen, beziehen sich lediglich auf die Pupillenbewegung.

Im Ergebnis kann festgehalten werden, dass die durchschnittliche Fehleranzahl bei Negativvorlagen mit 0,9111 geringer ist als die Fehleranzahl bei Positiv-Vorlagen. Die Differenz bleibt jedoch sehr klein und wird statistisch keinesfalls signifikant. Der Effekt der Vorlagenhelligkeit ist mit  $\eta^2 = .035$  sehr gering und mit  $p = .645$  nicht signifikant.

#### 3.4.6.6 Subjektive Einschätzung

Die subjektive Einschätzung der Aufgabenschwierigkeit erfolgte nach der Aufgabenlösung auf einer Skala von 1=einfach bis zu 3=schwierig. Da das Skalenniveau nicht mit den Skalenniveaus der Zeit und der absoluten Fehler vergleichbar ist, aus deren Betrachtung aber bereits ausreichende Erkenntnisse gewonnen wurden, wird die Variable *subjektive Einschätzung* aus forschungsökonomischen Gründen nur noch deskriptiv betrachtet.

**Tabelle 18: Subjektive Einschätzung der Aufgabenschwierigkeit**

	N	Minimum	Maximum	Mean	Std. Deviation
subj0	31	3,00	7,00	4,5161	1,06053
subj2	31	3,00	9,00	6,8387	1,48541
subj3	31	3,00	4,00	3,1935	,40161
subj5	31	3,00	9,00	6,0323	1,62242
subj6	31	3,00	9,00	4,0000	1,52753
subj8	31	5,00	9,00	8,4516	,85005
Valid N (listwise)	31				

Hier zeigt sich erwartungsgemäß ein ähnliches Bild wie bei den vorherigen Auswertungen. Das Aufgabenniveau der Klassenstufe 12/13 wird subjektiv als schwieriger erlebt, als das Niveau der Klassenstufe 6.

Auch hier werden die Daten umkodiert, so dass alle dunklen und alle hellen Vorlagen aggregiert werden. Die subjektive Einschätzung der Schwierigkeit der Aufgabenvorlagen verteilt sich danach wie in Tabelle 19 dargestellt. Der Effekt der Vorlagenhelligkeit ist mit  $\text{Eta}^2 = .37$  gering und wird mit  $p = .614$  nicht signifikant.

**Tabelle 19: Vergleich von subjektiven Schwierigkeitsbewertungen bei hellen und dunklen Vorlagen**

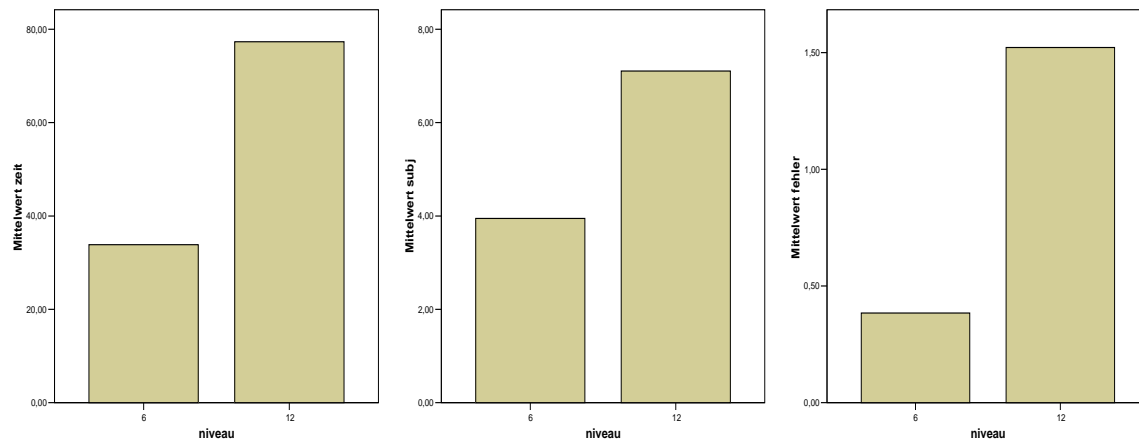
subj			
	Mittelwert	N	Standardabweichung
dunkel	5,4239	92	2,21500
hell	5,5851	94	2,14230
Insgesamt	5,5054	186	2,17417

### ***3.4.7 Zusammenfassung und Diskussion***

Um festzustellen, wie die Bearbeitung belastender kognitiver Aufgaben unter variierenden Lichtbedingungen am Computer auf die Pupille wirkt, wurde das Experiment KFT (Kognitiver Fähigkeitstest) durchgeführt. Hierzu wurden drei verschiedenen kognitive Aufgabenreihen (Subskalen für die Klassenstufe 6 und 12/13) aus dem KFT auf den Computer übertragen. Darüber hinaus sollte untersucht werden, ob die Helligkeit der Aufgabenvorlagen einen Einfluss auf das Lösungsverhalten der Probanden hat.

Wie gezeigt werden konnte, unterscheiden sich die Pupillengrößen bei der Aufgabenbearbeitung der Schwierigkeitsstufe Klasse 6 signifikant von denen der Klassenstufe 12/13. Die durchschnittliche Pupillenweitung unter der schwierigen Bedingung reichte von 2,03 Prozent bis zu 3,76 Prozent, je nach Aufgabentyp. Bei den Aufgabenreihen *Nonverbale Tests 1* (Figurenklassifikation und Figurenanalogien) betrug die Pupillenweitung 3,76 Prozent, bei den Testreihen der *Quantitativen Tests 1* (Mengenvergleiche) betrug sie 3,14 Prozent während sie bei den Testreihen *Quantitativen Tests 2* (Zahlenreihen) im Durchschnitt 2,03 Prozent betrug.

Die unterschiedlichen Schwierigkeitsniveaus wurden zur Kontrolle über die Bearbeitungszeit, die Lösungswahrscheinlichkeit und die subjektive Schwierigkeitseinschätzung validiert. Hierbei ergaben sich signifikante Unterschiede zwischen den Anforderungsniveaus.



**Diagramm 12: Übersicht über die (1) Aufgabendauer, die (2) subjektive Selbsteinschätzung und die (3) Fehler bei der Aufgabenlösung in den unterschiedlichen Niveaustufen**

Während die Bearbeitungszeit der Aufgaben vom Niveau der 6. Klasse im Mittel 33,8 Sekunden dauerte, stieg sie bei der Lösung von Aufgaben der Klassenstufe 12/13 auf 77,3 Sekunden im Mittel. Die subjektive Einschätzung der Schwierigkeit bewegte sich dabei von 3,9 (auf der Klassenstufe 6) zu 7,1 (im Niveau 12/13) auf eine Skala von 0 bis 9. Dabei machten die Versuchspersonen im Hauptaufgabenmaß auf dem Niveau der 6. Klasse im Schnitt 0,38 Fehler und im Niveau 12/13 bereits 1,52 Fehler.

Als Schlussfolgerung dieser Verteilungen kann festgehalten werden, dass die Zusammenstellung der Aufgaben (nach KFT-Subskalen) zwei Anforderungsbereiche deutlich trennt. Demgemäß ist die kognitive Beanspruchung der Probanden in der Niveaustufe 6 geringer als in der Niveaustufe 12/13. Die Pupillenweitung, die in der Niveaustufe 12/13 auftrat, kann also direkt auf die Veränderung der Anforderung zurückgeführt werden.

Zur Absicherung dieser Ergebnisse wurde zusätzlich geprüft, ob sich Unterschiede im Bearbeitungsverhalten der Probanden bei hellen (schwarze Schrift auf hellem Grund) und dunklen

(weiße Schrift auf schwarzem Grund) Aufgabenvorlagen ergeben. Die Unterschiede blieben hierbei jedoch äußerst gering und konnten statistisch nicht abgesichert werden.

Um den Zusammenhang der in diesem Kapitel besprochenen Pupillendaten aufzuzeigen, wurden sie nach einer Lichtbereinigung über die Aufgabentypen korreliert. Hierbei wurde die Korrelation nach Spearman gewählt, da das Skalenniveau der subjektiven Einschätzung lediglich ordinal ist. Es ergeben sich folgende Werte für N=6 Aufgabenvorlagen:

Es besteht ein Zusammenhang von der Pupillenweite mit der Fehleranzahl von  $r=.943$ , welcher mit  $p=.002$  signifikant ist. Der Zusammenhang der Pupillengröße mit der subjektiven Schwierigkeitseinschätzung von  $r=.829$  wird mit  $p=.021$  ebenfalls signifikant. Diese beiden Ergebnisse betätigen die *neural efficiency hypothesis* (Verney, Granholm & Marshall, 2004). Hiernach sind Personen mit hohem Intelligenzquotienten in der Lage kognitive Prozesse effizienter durchzuführen. Demzufolge werden relevante Stimuli besser diskriminiert und das Tempo und die Genauigkeit beim Problemlösen sind höher als bei Personen mit geringerem IQ. Die Korrelation von Pupillengröße und IQ ist nach Verney et al. (2004) negativ.

Die Korrelation der Klassenstufe gem. KFT mit der Pupillenbewegung beträgt  $r=.878$  und ist mit  $p=.011$  signifikant. Lediglich der Zusammenhang der Pupillenweite mit der Dauer der Aufgabenbearbeitung von  $r=.600$  wird mit  $p=.104$  nur knapp nicht tendenziell signifikant.

Bezieht man nun das Bestimmtheitsmaß  $r^2$  in die Berechnung ein, so ergibt sich bei dem Zusammenhang von der Pupillenbewegung mit der subjektiven Schwierigkeitseinschätzung ein Wert von  $r^2=.687$ . Dieser Wert beschreibt die gemeinsame Varianz der Variablen von 69 Prozent. Nimmt man nun an, dass die Pupille als Prädiktor der subjektiven Schwierigkeitsein-



schätzung anzusehen ist, so klärt dieser 69 Prozent der Varianz in der subjektiven Einschätzung auf.

Der Zusammenhang zwischen der Pupillenbewegung und der Fehlerhäufigkeit ist mit einem Bestimmtheitsmaß von  $r^2 = .889$  ausgewiesen. Die gemeinsame Varianz beider Variablen beträgt hier sogar 89 Prozent. Ist nun die Pupillenbewegung als Prädiktor anzusehen, so klärt dieser 89 Prozent der Varianz der Fehlerhäufigkeit auf.

In Bezug auf die Klassenstufe nach dem KFT beträgt das Bestimmtheitsmaß mit  $r^2 = .770$ .

Die Pupillenbewegung und die Niveaustufe gem. KFT weisen also 77 Prozent gemeinsame Varianz auf.

Insgesamt bleibt festzuhalten, dass die Pupillengröße ein sehr guter Indikator für die kognitive Beanspruchung ist. Im Bereich kognitiver Belastung spiegelt sie das Beanspruchungsniveau eines Computernutzers bei der Bewältigung kognitiver Aufgaben wider. Die Beziehung zwischen Klassenstufe gem. KFT als objektiviertem Schwierigkeitsmaß hat eine gemeinsame Varianz mit der Pupillenbewegung von 77 Prozent. Die Selbsteinschätzung der Schwierigkeit und die Pupillenbewegung haben eine gemeinsame Varianz von 68 Prozent.

Dieses Ergebnis ist von größter Bedeutung für die weiteren Schritte bei der Entwicklung einer biopsychologischen Mensch-Maschine-Schnittstelle. Die Messung, Bereinigung und Auswertung von Pupillometriedaten scheint somit gut geeignet, diese Schnittstelle zu realisieren, sofern die Belastungen des Users als kognitiv klassifiziert werden können.

Ob dieses Ergebnis auch unter anderen Bedingungen Bestand haben wird, wurde im nächsten Kapitel geprüft, in dem das Pupillenverhalten bei der Textrezeption und –reproduktion über verschiedene Schwierigkeitsstufen untersucht wurde.

### **3.5 Experiment 5: Textrezeption und -reproduktion**

Bei der Betrachtung digitaler Medien ist besonders auffällig, dass der Text als Symbolsystem zur Codierung von Informationen noch immer eine zentrale Rolle in der Mensch-Computer-Interaktion einnimmt. Aus diesem Grund wurden dem Thema Text und seiner Wirkung auf den Rezipienten in dieser Arbeit ein komplettes Experiment gewidmet, in dem sich seine zentrale Rolle beim Wissenserwerb widerspiegelt. Hierbei soll geklärt werden, ob sich die Ergebnisse zur Sprachverarbeitung und Pupillenstellung mit Wörtern und Sätzen von z.B. Beatty (1982), Ben-Nun (1986), Hyönä et al. (1995) und Just und Carpenter (1993) auch auf längere Texte übertragen lassen, oder ob die Pupille die mentale Beanspruchung über diese Zeit nicht widerspiegelt, wie dies von Schultheis (2004) hervorgehoben wurde.

Es soll außerdem überprüft werden, inwieweit die subjektiv erfahrene mentale Beanspruchung (subjektives Maß) beim Lesen und Reproduzieren von Texten anhand der Pupillenbewegung bestimmt werden kann. Außerdem soll das Verhältnis von subjektiv erfahrener mentaler Beanspruchung zur objektivierten (Hauptaufgabenmaß) mentalen Beanspruchung untersucht werden.

Die wissenschaftliche Hypothese, die in diesem Experiment geprüft werden soll, lautet: Mentale Beanspruchung durch Textrezeption und –reproduktion führt zu einer allgemeinen Res-

sourcenaktivierung, die sich in einer Pupillenweitung äußert. Es wird somit erwartet, dass die Pupillenbewegung mit dem Ausmaß der subjektiv erlebten mentalen Beanspruchung korreliert.

Das Experiment *Text-Test* wurde darauf ausgerichtet Erkenntnisse über das Pupillenverhalten von Personen zu erhalten, die an einem Monitor Texte lesen und verarbeiten. Um die Verarbeitung sicherzustellen, wurden die Probanden aufgefordert, die Texte nach der Rezeption inhaltlich zu reproduzieren. Die mentale Leistung bestand also darin, Informationen in das Kurzzeitgedächtnis zu überführen und von dort aus wieder abzurufen.

Aus den bisherigen theoretischen Überlegungen und Ergebnissen ergibt sich die Annahme, dass die Pupille beim Lesen einfacher Texte relativ kleiner bleibt, als beim Lesen komplizierter Texte. Darüber hinaus kann man davon ausgehen, dass die Reproduktion eine höhere Aktivierung bedingt als die Rezeption, d. h. dass die Pupille während der Reproduktion größer ist als bei der Rezeption. Auch bei der Reproduktion müsste somit die Pupillengröße bei komplexen Texten die Pupillengröße bei einfachen Texten übersteigen.

Zur Überprüfung dieser Grundannahmen wurde das Design des Text-Tests entwickelt. Da es sich bei diesem Experiment um ein umfangreiches und kompliziertes Design handelt, wurde ein Vorexperiment vorgeschaltet, um einen optimalen Versuchsaufbau für das Hauptexperiment zu gewährleisten.

Im Folgenden wird das Versuchsdesign erläutert und der Vorversuch kurz beschrieben. Danach folgt die Beschreibung und Auswertung des Hauptexperiments.

### 3.5.1 Text-Test Vorexperiment

#### 3.5.1.1 Versuchsdesign und –durchführung des Vorexperiments

Als Versuchsdesign wurde ein Messwiederholungsdesign über 2 x 3 Faktoren gewählt. Die Faktoren waren *Textschwierigkeit* und *Rezeption vs. Reproduktion*. Der Versuch Text-Test bestand aus drei verschiedenen, gleichlangen Texten, deren Textschwierigkeit hinsichtlich Formulierungen und Inhalt in drei Stufen variierte. Die Texte wurden den Versuchspersonen am Monitor präsentiert und die Pupillenbewegung wurde über die Versuchszeit aufgezeichnet.

Die Differenzierung der Texte in unterschiedliche Schwierigkeitsstufen wurde im Vorfeld anhand einer Schwierigkeitseinschätzung vorgenommen. Sie wurden 10 Personen (40 Prozent männlich, 60 Prozent weiblich, Durchschnittsalter  $M=28,2$  Jahre,  $SD=2,71$ ) vorgelegt, die die Texte auf einer 10-stufigen Skala einzuschätzen hatten. Das Ergebnis dieser Einschätzung stimmte mit den Vorannahmen überein.

**Tabelle 20: Rating der Texte im Vorversuch mit 0=sehr leicht, 5=mittel, 10= sehr schwer**

<b>Text</b>	<b>Mittelwert</b>	<b>SD</b>
Text 1 (leicht)	1,2	0,4
Text 2 (mittel)	4,6	0,8
Text 3 (schwer)	9,4	1,2

Der Versuch begann mit der Präsentation der standardisierten Instruktion in Textform, die von den Probanden am Monitor gelesen wurde. In der Instruktion wurden die Probanden dazu aufgefordert, die folgenden Texte zu lesen und danach mündlich wiederzugeben. Es gab so- dann Gelegenheit, Unklarheiten mit dem Versuchsleiter zu besprechen. Der Versuch begann mit der 1,25 Minuten dauernden Präsentation des leichtesten Textes. Danach wurde der Text durch eine leere Folie ersetzt, die lediglich ein Fixationskreuz in der Mitte zeigte. Die Pro- banden mussten auch während der Reproduktion auf den Monitor sehen, damit eine Pupil- lenmessung durchgeführt werden konnte. Die Zeit zur Reproduktion betrug 1,5 Minuten. Der zweite Text wurde 1,5 Minuten präsentiert und der dritte Text 2 Minuten. Die Reproduktions- zeit lag auch hier bei jeweils 1,5 Minuten. Die Präsentationsreihenfolge war leicht - mittel - schwer. Sämtliche Vorlagen waren so gestaltet, dass sie identische Lichtabstrahlungen von 96 Lux produzierten. Somit wurde eine lichtbedingte Pupillenbewegung ausgeschlossen. Am Versuchsende wurden die Probanden aufgefordert, die Texte nach ihrem Schwierigkeitsgrad auf einer 10stufigen Skala einzuschätzen.

### 3.5.1.2 Stichprobe des Vorexperiments

Die Stichprobe der Voruntersuchung bestand aus  $N=5$  Personen. Diese wurden aus einer Vor- lesung rekrutiert. Die Versuchspersonen waren alle weiblich, das Durchschnittsalter lag bei 27,4 Jahren ( $SD=11,02$ ). Alle Versuchspersonen gaben an, gesund zu sein und weder Alkohol noch Medikamente zu sich genommen zu haben.

### 3.5.1.3 Ergebnisse

Vor der statistischen Auswertung wurden die Datensätze von Artefakten befreit. In die Auswertung gingen die kompletten Datensätze der Textrezeption ein. Die Datensätze der Reproduktion gingen nur gekürzt (eine Minute) in die Auswertung ein, da einige Versuchspersonen vor der festgelegten Zeit die Reproduktion abgeschlossen hatten.

Die Versuchspersonen bewerteten die Schwierigkeiten der Texte wie in der Tabelle 21 dargestellt.

**Tabelle 21: Bewertung der Textschwierigkeit durch Selbsteinschätzung**

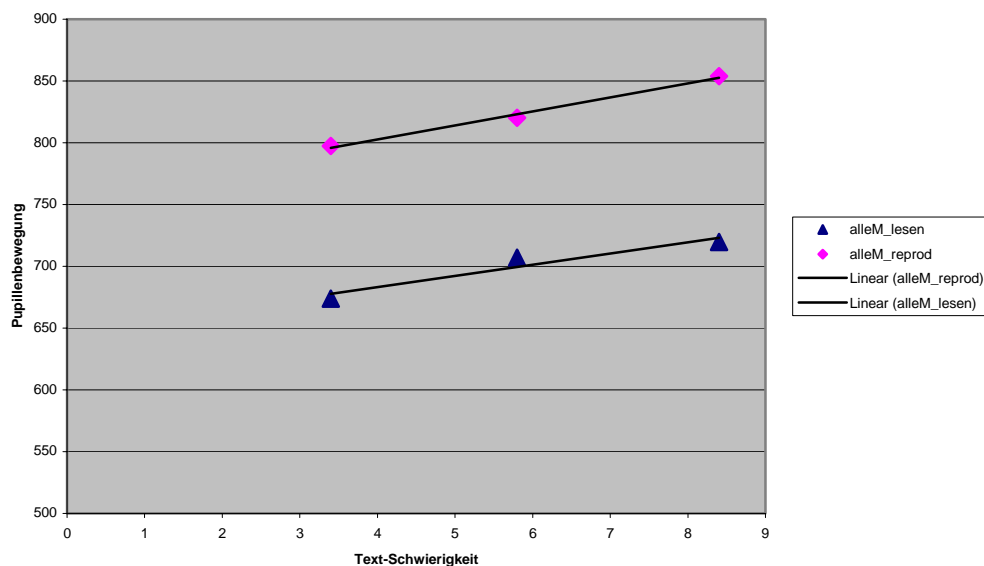
<b>Text</b>	<b>Mittelwert</b>	<b>SD</b>
Text 1 (leicht)	3,4 (1,2)	1,2
Text 2 (mittel)	5,8 (4,6)	1,33
Text 3 (schwer)	8,4 (9,4)	1,63

Die Differenzen der subjektiven Einschätzung der Textschwierigkeit zu der Vorbewertung der Texte (in der Klammer) kann wahrscheinlich durch die ungewohnte Laborsituation erklärt werden. Hierfür spricht auch die Tatsache, dass sich die Werte im Verlauf der Untersuchung angleichen. Für jeden einzelnen Datensatz wurden die Mittelwerte berechnet und jeweils über die Versuchspersonen gemittelt. Diese Mittelwerte wurden mit der gemittelten Textschwierigkeit (siehe Tabelle 21) korreliert. Die Ergebnisse finden sich in Tabelle 22 und im Diagramm 13.

**Tabelle 22: Korrelation der Textschwierigkeit (Rating der Probanden) mit den (über die Versuchspersonen gemittelten) Mittelwerten der Pupillenbewegung**

Textschwierigkeit	Mittelwerte der Pupillenbewegung Lesen	Mittelwerte der Pupillenbewegung Reproduzieren
3,4	673, 95	797, 33
5,8	707, 01	820, 12
8,4	719, 61	853, 95
Korrelation	$r = .962$	$r = .996$
Varianzaufklärung	$r^2 = .925$	$r^2 = .992$

Die Pupillenbewegung beim Lesen und auch beim Reproduzieren korreliert sehr hoch mit der (gemittelten) subjektiv erfahrenen Textschwierigkeit. Die Pupillenbewegung ist beim Reproduzieren höher als beim Lesen.

**Diagramm 13: Korrelation der Textschwierigkeit (Rating der Probanden) mit den (über die Versuchspersonen gemittelten) Mittelwerten der Pupillenbewegung**

#### 3.5.1.4 Zusammenfassung und Diskussion

Die Ergebnisse zeigen, dass die Pupillenweite mit der Schwierigkeitsstufe der verarbeiteten Texte steigt. Außerdem wurde deutlich, dass der Bereich der Textreproduktion im Gesamtniveau der Pupillengröße über dem Niveau der Textrezeption liegt. Die Aktivierung und die damit verbundene Weitung der Pupillen ist bei der Textreproduktion höher.

Wie dieser Vorversuch zeigt, gibt es ferner einen starken Zusammenhang zwischen der Pupillenbewegung und der subjektiven Einschätzung der Textschwierigkeit. Die Varianzaufklärung der Pupillenbewegung durch die subjektive Einschätzung der Schwierigkeit liegt beim Lesen bei 92,5 Prozent und bei der Textreproduktion bei 99,2 Prozent.

Die Ergebnisse dieser Voruntersuchung scheinen somit analog zu den Ergebnissen des KFT zu sein. Komplexere Anforderungen an den Computernutzer führen je nach Leistungsfähigkeit des Nutzers zu einer Weitung der Pupille, auch wenn sich die Aufgabenbearbeitung über einen längeren Zeitraum erstreckt. Die Pupillenweitung kann im hier beschriebenen Setting sehr gut als Indikator mentaler Beanspruchung angesehen werden. Um diese Ergebnisse zu bestätigen und statistisch gegen den Zufall abzusichern, wird dieses Experiment mit einer größeren Stichprobe wiederholt.

Aus den Erfahrungen des Vorversuchs ergab sich die Notwendigkeit, einige Änderungen im Versuchsdesign vorzunehmen. Um Effekte auszuschalten, die aus der Reihenfolge der Textdarbietung (Reiheneffekte) resultieren, wurden die Texte im Hauptversuch in der Reihenfolge variiert. Außerdem wurden die Zeiten für die Textreproduktion und die Textrezeption nicht mehr fest vorgegeben. Die Versuchspersonen konnten die Texte beliebig lange lesen und sig-



nalisierten dem Versuchsleiter danach, fertig zu sein. Anschließend wurden die Texte reproduziert, auch hier konnten die Probanden das Ende der Reproduktion selbst bestimmen und dem Versuchsleiter anzeigen.

### ***3.5.2 Text-Test Hauptexperiment***

Das Hauptexperiment sollte analog zu dem zuvor durchgeführten Vorversuch realisiert werden. Unter Berücksichtigung der Überlegungen, die im vorangegangenen Kapitel zu dem Verhältnis der Aufgabendauer und der mentalen Beanspruchung angestellt wurden, und unter der Annahme eines Ressourcenmodells der mentalen Aktivierung wurde im Hauptexperiment auf festgelegte Zeiten verzichtet. Dies hatte auch den Vorteil, dass, im Gegensatz zum Vorversuch, die kompletten Datensätze der Textreproduktion verwendet werden konnten.

#### *3.5.2.1 Hypothesen*

Die Haupthypothese der Voruntersuchung wurde nun statistisch überprüft.

##### 1. Hypothese

Mentale Beanspruchung durch Textrezeption und –reproduktion am Computer führt zu einer allgemeinen Ressourcenaktivierung, die sich in einer Pupillenweitung äußert. Es wird somit erwartet, dass die Pupillenbewegung mit dem Ausmaß der subjektiv erlebten mentalen Beanspruchung korreliert.

## 2. Hypothese

Mentale Beanspruchung durch Textrezeption und –reproduktion am Computer führt zu einer allgemeinen Ressourcenaktivierung, die sich in einer Pupillenweitung äußert. Es wird somit erwartet, dass die Pupillenbewegung mit dem Ausmaß der Beanspruchung (als Steigerung des Abstraktionsgrades der Texte) korreliert.

Zur statistischen Überprüfung dieser Hypothesen wurden folgende Variablen erhoben:

- Pupillenweite,
- subjektive Schwierigkeitseinschätzung gem. SEA-Skala und
- objektivierte Textschwierigkeit.

### 3.5.2.2 Stichprobenbeschreibung

Bei der verwendeten Stichprobe handelt es sich um eine Zufallsstichprobe, deren Teilnehmer auf dem Gelände der Freien Universität in Berlin Lankwitz rekrutiert wurden. Neben studierenden und lehrenden Universitätsangehörigen gingen auch Teilnehmer außeruniversitärer Ausbildungseinrichtungen (des Campus Berlin Lankwitz) in die Stichprobe ein. Darüber hinaus wurden Fußgänger der Umgebung als Probanden rekrutiert. Die Stichprobe bestand aus  $N=48$  Versuchspersonen, deren Daten aufgenommen wurden. Bei 3 weiteren Versuchspersonen wurden aufgrund von technischen Problemen keine Messungen durchgeführt. Die Stichprobe gliederte sich in 34 weibliche und 14 männliche Probanden, das Alter der Probanden reichte von 17 bis 45 Jahre mit einem Mittelwert von  $M=23,18$  Jahre ( $SD=5,019$ ). 17 Personen gaben an, manchmal eine Brille zu tragen, während der Versuchsdurchführung trug je-

doch kein Proband eine Brille. 3 Personen hatten vor dem Versuch Medikamente zu sich genommen. Da es sich jedoch lediglich um leichte Kopfschmerzmittel handelte, die keinen verzerrenden Einfluss auf die Pupille ausüben, verblieben diese Personen in der Stichprobe. Eine Versuchsperson klagte über leichte gesundheitliche Einschränkungen. Keine Versuchsperson hatte vor dem Versuch Alkohol konsumiert.

Von den  $N=48$  Probanden, bei denen die Pupillenmessung durchgeführt wurde, konnten die Datensätze von 7 Probanden (VP005, VP008, VP013, VP016, VP018, VP030 und VP034) nicht in die weitere Datenanalyse aufgenommen werden, da die Messung nicht ordnungsgemäß durchgeführt werden konnte. Ursachen waren beispielsweise extrem stark geschminkte Augen, welche eine Messung der Pupille zeitweise unmöglich machten, technische Probleme mit der Messanlage und der Software oder Fehlmessungen durch zu starke Kopfbewegungen der Probanden. Diese Umstände führten dazu, dass der Kontakt von der Pupille zur Messeinrichtung verloren ging. Ein Proband (VP040) wurde aufgrund erheblicher Verständnisschwierigkeiten der deutschen Sprache von der Messung ausgeschlossen, da er die nötige Informationsverarbeitungsleistung nicht erbringen konnte.

### 3.5.2.3 Versuchsdurchführung

Das Versuchsdesign wurde aus dem Vorexperiment als Messwiederholungsdesign drei über zwei Faktoren übernommen. Die Faktoren waren *Textschwierigkeit* und *Rezeption vs. Reproduktion*. Als Stimuli wurden die Texte aus dem Vorversuch übernommen. Der Abstraktionsgrad der Texte stieg über drei Stufen an. Der erste Text stammte aus einem Kinderbuch in Form einer konkreten Erzählung. Der zweite Text beschrieb einen wissenschaftlichen Sachverhalt in Form eines Zeitungsberichts und der dritte Text beschrieb einen hoch abstrakten

philosophischen Sachverhalt. Sowohl in der Versuchsvorbereitung als auch im Vorversuch hatte sich gezeigt, dass die Verarbeitungsleistung der Texte mit zunehmenden Abstraktionsgrad eine erhöhte mentale Beanspruchung hervorrief.

Die Texte wurden den Versuchspersonen am Monitor präsentiert und die Pupillenbewegung wurde über die Versuchszeit aufgezeichnet. Zusätzlich wurden die Probanden mit einer Videokamera in Bild und Ton aufgezeichnet, um später die individuelle Versuchssituation nachvollziehen zu können. Dabei wurde die Reihenfolge der Textdarbietung nach dem Zufallsprinzip variiert, um Reiheneffekte auszuschließen.

Die subjektive Selbsteinschätzung der Textschwierigkeit wurde in Abweichung zum Vorexperiment anhand der SEA-Skala (Eilers, Nachreiner & Hänecke, 1986) vorgenommen. Die Skala wurde bezüglich der Skalenbeschriftung verändert: In Anlehnung an das Vorexperiment wurde die ursprüngliche Skala auf die Range von 0-10 interpoliert, die Skalengüte wurde hierdurch jedoch nicht beeinträchtigt. Der Vorteil dieser Skala ist das höhere Datenniveau, welches in der Auswertung die Berechnung eines Korrelationskoeffizienten nach Pearson zulässt.

#### 3.5.2.4 Die SEA-Skala

Die Erfassung von Daten zum Beanspruchungszustand eines Menschen bei der Aufgabenbearbeitung ist im Rahmen der Ergonomieforschung (vgl. Hacker & Richter, 1984; Ulich 1994) von großer Bedeutung. In diesem Bereich wurden zahlreiche Methoden entwickelt, um verlässliche Informationen über den mentalen Zustand eines Probanden gewinnen zu können

(vgl. Kap. 4). Während der NASA Task Load Index (Hart & Staveland, 1988) die Beanspruchung mit 6 Subskalen vor dem Hintergrund eines mehrdimensionalen Beanspruchungskonzeptes erfasst, basiert die SEA-Skala („Skala zur Erfassung subjektiv erlebter Anstrengung“) auf der eindimensionalen Einschätzung der mentalen Beanspruchung. Dabei wird nach Sanders (1983) die Anstrengung (effort) als wesentliches Kriterium des Beanspruchungserlebens beurteilt. In einer Methoden vergleichenden Untersuchung (Seifert, 2002) ergaben sich für die SEA-Skala und den NASA-TLX Gesamtscore Korrelationen von  $r = .78$  und für die NASA-TLX Subskala Anstrengung und die SEA-Skala Korrelationen von  $r = .81$ .

In dem hier beschriebenen Experiment wurde auf die SEA-Skala zurückgegriffen, da diese aufgrund des geringeren Umfangs deutlich schneller zu bearbeiten ist. Hinzu kommt, dass der NASA-TLX einige Subskalen verwendet, die für das vorliegende Experiment weniger bedeutsam sind.

Die Werte der Originalskala (0-200) wurden auf die Werte 1 -10 interpoliert, ohne hierbei die Skalenabstände zu verändern. Dieses Vorgehen erlaubt einen ungefähren Vergleich der erhobenen Daten mit den Daten der Voruntersuchung.

#### 3.5.2.5 Datenaufbereitung

Vor der Datenauswertung wurden die Daten (wie in Kap. 3.1.3 beschrieben) von Artefakten bereinigt. Eine Lichtbereinigung wurde nicht durchgeführt, da die Lichtverhältnisse in dem Experiment konstant waren. Die Monitorabstrahlung lag bei genau 96 Lux. Bei der Auswertung der Daten zeigte sich nach der Datenbereinigung ein zufrieden stellendes Bild für den

Bereich der Textrezeption. Für den Bereich der Textreproduktion zeigte sich ein weniger positives Bild: Die Schwankungen in den Messwerten lagen hier deutlich höher, da sich die Probanden während des Sprechens mehr bewegten, als während des Lesens. Die Artefakte nahmen hierbei teilweise extreme Werte an. Dieser Umstand führte zu einer deutlich größeren Varianz der Messwerte und somit zu einem schlechteren Ergebnis nach der Datenbereinigung (nach dem im Kapitel 3.1.3 beschriebenen Verfahren). Um dieses Problem zu minimieren, musste eine weitere Bereinigung der Daten aus den Messreihen der Textreproduktion erfolgen.

Das Vermessen der Pupille während des Sprechens ist jedoch für die vorliegende Arbeit nicht weiter relevant, da die hier behandelten Systeme über keinerlei Spracheingabe verfügen. Die Auswertung dieser Daten diente lediglich der Überprüfung, ob tatsächlich eine Informationsverarbeitung stattgefunden hatte. Aus diesen Gründen erscheint die Entwicklung eines eigenen Verfahrens zur Bereinigung der beim Sprechen gewonnenen Pupillendaten überflüssig.

Die abgenommenen Werte wurden per Hand von den identifizierten Artefakten befreit.

Grundlage der Datenbereinigung war hierbei die Annahme, dass sich die Größenveränderung der Pupille im Bereich von 2-8 Millimetern bewegen kann. Setzt man nun den kleinsten gemessenen Wert beim Lesen mit 2 mm gleich, so ergibt die Multiplikation dieses Werts (Intervallskala) mit 4 den erreichbaren Maximalwert. Da der kleinste Wert aber sicherlich größer als 2 mm sein dürfte, da im Labor keine absolute Helligkeit herrschte, dürfte der 4fache Wert bereits über dem erreichbaren Maximalwert liegen. Es erscheint also legitim, die darüber liegenden Werte in den Messreihen der Textreproduktion zu löschen und die Varianz somit zu verkleinern. Nach dieser ersten groben Datenbereinigung wurden die Daten mit dem in Kapitel 3.1.3 beschriebenen Verfahren einer zweiten Bereinigung unterzogen, deren Ergebnisse eine deutlich realistischere Streuung aufwiesen.

Ein weiteres Problem stellte die Schwierigkeit der Texte dar, die für alle Probanden gleich war. Da die Leistungsfähigkeit der Probanden jedoch schwankt, ist nicht davon auszugehen, dass die Verarbeitung der Texte eine vergleichbare Ressourcenaktivierung zur Folge haben würde. Diese Annahme wurde durch die Streuung der Werte der SEA-Skala gestützt. So traten einige Fälle ein, bei denen die Texte der dritten Abstraktionsstufe nicht wiedergegeben werden konnten, da die Probanden sie nicht verstanden hatten. Die betroffenen Probanden (VP012, VP023, VP042 und VP045) äußerten sich auch dahingehend, dass sie zu einer Textreproduktion nicht in der Lage seien. Die jeweils abgeleiteten Daten zu dieser Niveaustufe zeigten somit zu niedrige Werte, da eine Informationsverarbeitung aufgrund der Überforderungssituation nicht im ausreichenden Maße stattgefunden hatte. Die Werte der Messungen auf dieser Niveaustufe wurden aus diesem Grund nicht weiter berücksichtigt.

Ein erster Vergleich der erhobenen Daten zeigte einen nicht vorhergesehenen Effekt. Die Messwerte der Pupille während der Textrezeption des jeweils zuerst rezipierten Textes waren höher als erwartet. Dieser Effekt schien auch unabhängig von der Textschwierigkeit zu sein. Im Gegensatz zum Vorversuch handelte es sich bei den Probanden offenbar nicht um Personen, die mit Laborsituationen vertraut waren. Die Pupillenweitung in diesen ersten Minuten war offensichtlich eine Reaktion auf die ungewohnte Umgebung und Situation. Diese Annahme wurde dadurch erhärtet, dass die Probanden des Vorversuchs, die bereits an vorhergehenden Versuchen teilgenommen hatten, diese Reaktion nicht zeigten. Aus diesem Grund wurde ursprünglich auch auf eine „Aufwärmphase“ vor dem eigentlichen Versuch verzichtet, um die Probanden nicht länger als notwendig zu beanspruchen. Um diese Fehleinschätzung im Nachhinein zu kompensieren, wurden die Messwerte der ersten Messreihen in der Analyse nicht berücksichtigt.

Die Anzahl der Versuchspersonen verteilte sich folgendermaßen auf die Darbietungsreihenfolge:

**Tabelle 23: Sampling der Versuchsvorlagen**

<i>Sampling der Vorlagentexte</i>	<i>Anzahl der Versuchspersonen</i>
1 – 2 – 3	7
2 - 1 – 3	7
3 -1 – 2	6
1 - 3 - 2	7
2 – 3 – 1	7
3 – 2 – 1	6

Von den insgesamt 288 aufgezeichneten Messreihen wurden durch das Entfernen der Messreihen an der Startposition insgesamt 40 Messreihen nicht berücksichtigt. Hierbei entfielen auf die erste und die zweite Niveaustufe jeweils 14 Messreihen und auf die dritte Niveaustufe 12 Messreihen, die für die weitere Datenanalyse nicht mehr berücksichtigt wurden.

### 3.5.2.6 Statistische Auswertung

Die folgende statistische Auswertung gliedert sich zwei Phasen: In der ersten Phase werden die statistischen Operationen durchgeführt, die aufgrund der Datenbasis zulässig sind. Durch die Einschränkungen der Messtechnik sind die Ergebnisse für den Zweck dieser Arbeit jedoch nicht in ausreichendem Maß interpretierbar, so dass in einer zweiten Phase das Verfahren der



Harmonisierung zur Datentransformation zum Einsatz kommt. Nach der Harmonisierung werden dann weiterführende statistische Verfahren eingesetzt, um eine bessere Näherung an den Forschungsgegenstand zu erzielen.

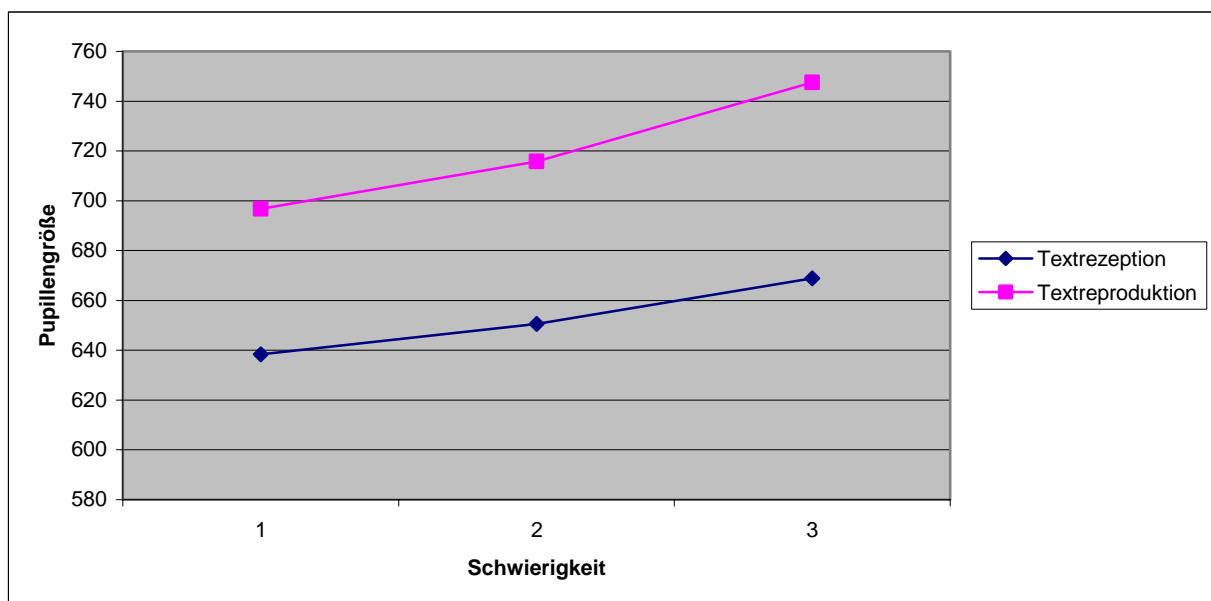
Für die erste Phase der Auswertung wurden die Daten in der Weise bereinigt, wie dies in Kapitel 3.1.3 beschrieben ist. Nun konnten die Daten über Mittelwerte aggregiert werden. Für jeden einzelnen Datensatz wurden die Mittelwerte berechnet und jeweils über die Versuchspersonen gemittelt. Dieses Verfahren eignet sich im vorliegenden Fall, da aufgrund der Arbeitsweise der Messsysteme keine Baseline für die Pupillendaten erstellt werden konnte.

Die Werte der SEA-Skala wurden ebenfalls über Mittelwerte aggregiert. Diese Mittelwerte wurden mit der gemittelten Textschwierigkeit über den Pearson-Korrelationskoeffizienten korreliert. Die Ergebnisse sind in der Tabelle 24 dargestellt.

**Tabelle 24: Korrelation der Textschwierigkeit (Rating der Probanden) mit den (über die Versuchspersonen gemittelten) Mittelwerten der Pupillenbewegung**

Textschwierigkeit	Mittelwerte der Pupillenbewegung Lesen	Mittelwerte der Pupillenbewegung Reproduzieren
2,67	638,31	696,74
5,08	650,63	715,75
8,38	668,84	747,47
Korrelation	$r = 1.0$	$r = .999$
Varianzaufklärung	$r^2 = 1.0$	$r^2 = 99.80$

Eine erste Auswertung des vorhandenen Datenmaterials zeigt eindeutig eine noch engere Beziehung der abhängigen und der unabhängigen Variablen, als dies nach dem Vorversuch zu erwarten gewesen wäre. So korrelieren die subjektiven Einschätzungen der mentalen Anforderung auf der SEA-Skala für die Textverarbeitung mit den gemessenen Pupillengrößen beim Lesen mit  $r= 1.0$  und bei der Textwiedergabe mit  $r= .999$ . Berechnet man nun den Bestimmtheitsfaktor, so ergeben sich für die subjektive Belastungseinschätzung beim Textlesen und der Pupillenweite eine gemeinsame Varianz von 100 Prozent und für die Textwiedergabe eine gemeinsame Varianz von 99,8 Prozent.



**Diagramm 14: Pupillenweite über die Versuchsbedingungen Textreproduktion und -rezeption**

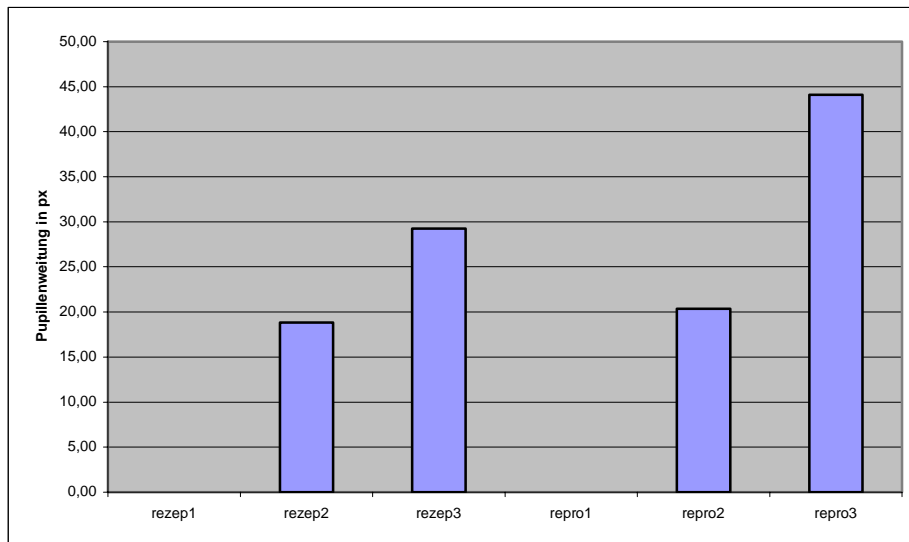
Auch die Prüfung auf einen Zusammenhang zwischen objektivierter Schwierigkeitseinschätzung aus dem Vorversuch und der Pupillenweitung (Rezeption:  $r= 1$ ,  $p=.004$  und Reproduktion:  $r= .999$ ,  $p=.014$ ) zeigt eine Bestätigung des Ergebnisses. Die Variable *Objekt* gibt dabei die objektivierte Schwierigkeitseinschätzung aus der Vorbereitung des Vorversuchs an.

Die nicht gemittelten Einzelwerte der subjektiven Einschätzung der mentalen Beanspruchung auf der SEA-Skala korrelieren dabei mit den drei Niveaustufen auf der Textschwierigkeit mit  $r = .806$ ,  $p = .000$  hoch signifikant.

Es ist demnach möglich, durch die Messung der Pupillenweite die subjektiv erlebte mentale Arbeitslast vorherzusagen. Dies gilt zu 100 Prozent für die Textrezeption und zu 99,8 Prozent für die Selbsteinschätzung bei der Textreproduktion. Pupillometrische Messverfahren scheinen also für die Analyse von Informationsverarbeitungsprozessen für den Bereich Lesen und Wiedergeben von Bildschirmtext sehr gut geeignet zu sein.

Für eine genauere Auswertung des Datenmaterials ist die Transformation der Rohdaten durch eine Harmonisierung mit Festlegung einer Baseline mit dem Wert 0 notwendig. Das Verfahren wurde bereits im Kapitel 2.2 erläutert.

Das folgende Diagramm 15 zeigt die deskriptive Übersicht der neuen Mittelwerte des harmonisierten Datenmaterials. Deutlich erkennbar ist der Nullpunkt, welcher die Pupillengröße während der Textbearbeitung auf der Niveaustufe 1 als Baseline repräsentiert. Die prozentuale Weitung der Pupille bei der Textrezeption beträgt auf der 2. Stufe 42,69 Prozent und auf der 3. Stufe bereits 66,29 Prozent. Bei der Textreproduktion weitet sich die Pupille auf der 2. Stufe um 46,16 Prozent und erreicht auf der 3. Stufe ihre maximale Weitung von 100 Prozent.



**Diagramm 15: Verteilung der Mittelwerte der Pupillenweite nach der Datenharmonisierung**

Schon bei der deskriptiven Statistik zeigt sich ein eindeutiger Trend über sämtliche Schwierigkeitsstufen in der Form, dass die Pupillenweite mit der Zunahme der Schwierigkeit größer wird. Die stärkeren Abweichungen der Variablen Textreproduktion (repro) auf der Stufe 3 gehen dabei auf die schlechtere Datenbasis zurück. Wie bereits bei der Datenbereinigung erwähnt, spiegelt sich hier die höhere Fehlerquote wider, die durch die stärkere Bewegung der Versuchspersonen beim Sprechen entsteht.

Der Trend, der bereits deskriptiv erkennbar wurde, setzt sich auch beim Einsatz varianzanalytischer Verfahren für Mittelwertsunterschiede fort. Bei der durchgeführten einfaktoriellen Varianzanalyse zeigen sich sowohl bei der Textrezeption (rezept) mit  $F(2)=44,993$ ,  $p=.000$  wie auch bei der Textreproduktion (repro) mit  $F(2)=47,452$ ,  $p=.000$  hochsignifikante Unterschiede der Mittelwerte. Dabei beträgt der Effekt der Niveaustufe  $\eta^2=.548$  bei der Textrezeption und  $\eta^2=.618$  bei der Textreproduktion.

Zur weiteren Analyse der Daten dienen die Ergebnisse der durchgeführten einfaktoriellen Varianzanalyse. Die Mehrfachvergleiche der Mittelwertsunterschiede der Post-Hoc-Tests

(Tukey-HSD-Test) zeigen hochsignifikante Unterschiede zwischen sämtlichen Mittelwerten der Variablen auf den drei Schwierigkeitsstufen.

Die Prüfung auf Homogenität der Untergruppen mittels Tukey-HSD-Test bestätigt sowohl für die Variable *Textrezeption* als auch für die Variable *Textreproduktion* den bereits in der deskriptiven Statistik sichtbaren Trend. Mit einer Irrtumswahrscheinlichkeit von unter 5 Prozent zeigen beide Variablen jeweils drei homogene Untergruppen. Diese sind jeweils in der erwarteten Reihenfolge abgebildet (s. Anhang).

Die Ergebnisse der Varianzanalyse und der Post-Hoc-Tests zeigen ein eindeutiges Bild. Das Verhältnis von Pupillenweite zu den drei dargebotenen Niveaustufen der Textschwierigkeit stützt die eingangs aufgezeigten Hypothesen. Neben den drei Niveaustufen der Textschwierigkeit ist es mit den harmonisierten Daten nun aber auch möglich, den Korrelationskoeffizienten der subjektiven Selbsteinschätzung der mentalen Beanspruchung auf der SEA-Skala mit der Pupillenweite zu bilden. Im Gegensatz zu dem Verfahren, wie es in der ersten Phase dieser Analyse durchgeführt wurde, gehen nun nicht nur die gemittelten Werte in die Berechnung ein, sondern die gesamten Messwerte der Messreihen. Die Varianzen in der SEA-Skala und in der Pupillenbewegung sind also berücksichtigt. Sowohl der Wert der Korrelation von  $r = .603$ ,  $p = .000$  bei der Textrezeption wie auch der Korrelationskoeffizient  $r = .694$ ,  $p = .000$  bei der Textreproduktion liegen nun zwar deutlich unter den Korrelationen der gemittelten Daten, sie sind aber das genauere Ergebnis. Interessant ist auch die mit  $p = .000$  höhere Signifikanz der Korrelation. Die gemeinsame Varianz der Pupillenweite liegt bei der Textrezeption bei 36,36 Prozent und bei der Textreproduktion bei 48,16 Prozent.

Etwas höher liegt die Korrelation der Pupillenweite mit den drei Schwierigkeitsstufen der Texte. Hier ergibt sich für die Textrezeption ein Korrelationskoeffizient von  $r = .738$ ,  $p = .000$  und für die Textreproduktion von  $r = .784$ ,  $p = .000$ .

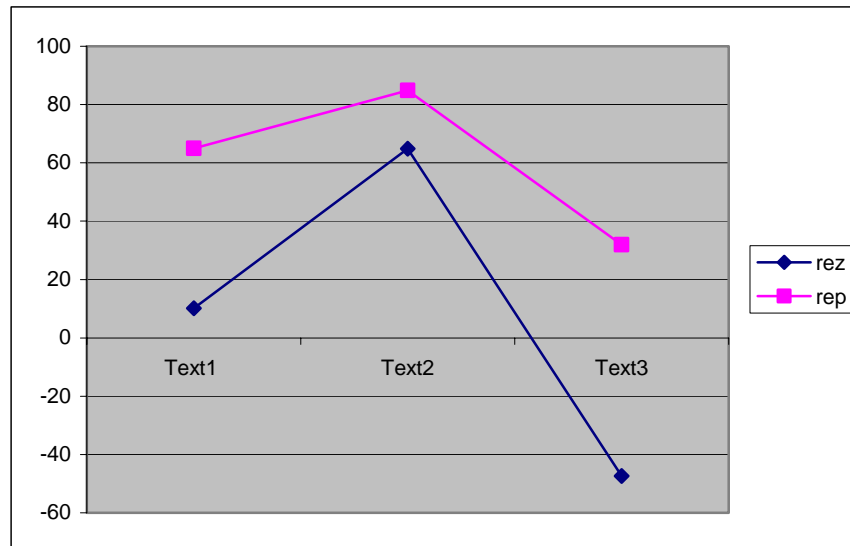
Die Verwendung harmonisierter Daten führt in der Analyse zwar insgesamt zu weniger eindeutigen Ergebnissen als die Verwendung der gemittelten Rohdaten, die Ergebnisse sind aber dafür wesentlich aussagekräftiger, da die Einzelvarianzen berücksichtigt werden.

#### 3.5.2.7 Analyse von Überforderungsreaktionen

In der Auswertung des Experiments 5 wurde bei 4 Datensätzen eine Unregelmäßigkeit festgestellt, die durch kognitive Überlastung der Probanden erklärbar wurde. Dies konnte durch Kommentare der Probanden untermauert werden, die von einer VHS-Subjekt-Kamera festgehalten worden waren. Alle 4 Probanden gaben an, den Text der Niveaustufe 3 nicht verstanden zu haben. Wie bereits erwähnt, wurden die betreffenden Datensätze aus diesem Grund von der weiteren Auswertung ausgeschlossen.

Um der Frage nachzugehen, welche spezifischen Pupillenreaktionen auf eine kognitive Überforderung hindeuten, sollen diese Datensätze an dieser Stelle allerdings etwas näher betrachtet werden. Das Diagramm 16 zeigt die Mittelwertverteilung dieser 4 Datensätze. Die Mittelwerte wurden für die Zusammenstellung standardisiert, indem von dem jeweiligen Pupillenwert der Messung ein Ausgangswert abgezogen wurde. Dieser Ausgangswert wurde vor der eigentlichen Messung abgenommen, indem jeder Versuchsperson ein leerer Bildschirm präsentiert wurde, nachdem die Instruktion beendet war. Während dieser Einblendung wurden

die Pupillendaten aufgezeichnet, wobei der leere Bildschirm die gleiche Lichtabstrahlung wie die Textfolien aufwies. Im Diagramm 16 ist somit das Ausgangsniveau auf der Ordinate gleich Null gesetzt. Abgebildet sind also nur die relativen Veränderungen vom Ausgangswert.



**Diagramm 16: Überlastungsreaktionen beim Textniveau 3 in der Übersicht**

Für den Bereich der Textrezeption sinkt der Pupillenwert bei einer Überforderungssituation deutlich unter das Ausgangsniveau ab. Im Fall der Textreproduktion sinkt der Pupillenwert zwar auch deutlich ab, bleibt jedoch eindeutig über dem Ausgangsniveau. Eine mögliche Erklärung hierfür wäre, dass in diesem Fall eine Textwiedergabe versucht wurde und durch das Sprechen bereits eine mentale Belastung stattgefunden hat, die den Pupillenwert über dem Ausgangswert geweitet hält.

Diese Ergebnisse der deskriptiven Analyse von Überforderungssituationen können jedoch aufgrund der geringen Fallzahl lediglich zur Hypothesengenerierung genutzt werden. Für eine statistische Auswertung müsste hier ein weiteres Experiment durchgeführt werden, in dem Überforderungssituationen systematisch provoziert werden.

### 3.5.2.8 Ergebnisse

Für die eingangs formulierte erste Hypothese über den Zusammenhang zwischen subjektiv erlebter mentaler Beanspruchung und der gemessenen Pupillenweite ( $H_0: \rho = 0$  und  $H_1: \rho > 0$ ) wurden mehrere Verfahren zur statistischen Überprüfung angewandt. Im Ergebnis unterscheiden sich die Analysen nicht elementar voneinander, lediglich die Effektstärke schwankt. Im Einzelnen konnten folgende Zusammenhänge aufgedeckt werden:

- Korrelation der Pupillenweite und der gemittelten SEA-Skala der Textrezeption unter Verwendung der Ursprungsdatenbasis mit  $r = 1.0$ ,  $p(\text{einseitig}) = .007$ .
- Korrelation der Pupillenweite und der gemittelten SEA-Skala der Textreproduktion unter Verwendung der Ursprungsdatenbasis mit  $r = .999$ ,  $p(\text{einseitig}) = .017$ .
- Korrelation der Pupillenweite und der SEA-Skala der Textrezeption unter Verwendung der harmonisierten Daten mit  $r = .603$ ,  $p(\text{einseitig}) = .000$ .
- Korrelation der Pupillenweite und der gemittelten SEA-Skala der Textreproduktion unter Verwendung der harmonisierten Daten mit  $r = .694$ ,  $p(\text{einseitig}) = .017$ .

Die Nullhypothese wird somit verworfen und die Alternativhypothese  $H_1$  wird angenommen.

Die Irrtumswahrscheinlichkeit liegt in allen Fällen unter 5 Prozent.



Für die eingangs formulierte zweite Hypothese über den Zusammenhang von objektivierten Textschwierigkeiten und der gemessenen Pupillenweite ( $H_0: \rho = 0$  und  $H_1: \rho > 0$ ) wurden ebenfalls Analysen mit der Ursprungsdatenbasis und mit der harmonisierten Datenbasis berechnet. Das Ergebnis dieser Berechnungen kann der folgenden Zusammenstellung entnommen werden:

- Korrelation der Pupillenweite und der objektivierten Textschwierigkeit bei der Textrezeption unter Verwendung der Ursprungsdatenbasis mit  $r = 1.0$ ,  $p(\text{einseitig}) = .004$ .
- Korrelation der Pupillenweite und der objektivierten Textschwierigkeit bei der Textproduktion unter Verwendung der Ursprungsdatenbasis mit  $r = .999$ ,  $p(\text{einseitig}) = .014$ .
- Korrelation der Pupillenweite und der objektivierten Textschwierigkeit bei der Textrezeption unter Verwendung der harmonisierten Datenbasis mit  $r = .738$ ,  $p(\text{einseitig}) = .000$ .
- Korrelation der Pupillenweite und der objektivierten Textschwierigkeit bei der Textproduktion unter Verwendung der harmonisierten Datenbasis mit  $r = .784$ ,  $p(\text{einseitig}) = .000$ .

Die Nullhypothese wird somit verworfen und die Alternativhypothese  $H_1$  wird angenommen. Die Irrtumswahrscheinlichkeit liegt dabei jeweils unter einem Prozent.

Die Hypothesen 3 bis 6 beziehen sich darauf, wie stark die Pupillenweite von der Textschwierigkeit beeinflusst wird. Mittels einfaktorieller Varianzanalysen wurden die Mittelwertsunter-

schiede der Pupillenweite geprüft, wofür die harmonisierte Datenbasis herangezogen wurde. Das Ergebnis der Varianzanalyse ergab für die Textrezeption über die Niveaustufen mit  $F(2,73)=44,203$ ,  $p=.000$  hoch signifikante Mittelwertsunterschiede. Die Mittelwertsunterschiede bei der Textreproduktion über die Niveaustufen wurden mit  $F(2,73)=58,981$ ,  $p=.000$  ebenfalls hoch signifikant. Eine weitere Analyse der Mittelwertsdifferenzen durch den Tukey-B-Test ergab in beiden Fällen eine klare Abstufung der Pupillenwerte. Der Pupillendurchmesser wurde auf der Niveaustufe 1 signifikant kleiner gemessen als auf der Niveaustufe 2, welche wiederum kleiner als die Niveaustufe 3 gemessen wurde. Diese Ergebnisse wurden für alle Fälle mit  $p<.05$  signifikant.

In Bezug auf die entsprechenden Hypothesen bedeutet dies:

- Hypothese 3: Der Mittelwert  $M_{rez2}= 14,38$  übersteigt den Mittelwert  $M_{rez1}= 0,0$  mit  $p=.000$  hoch signifikant. Die  $H_0$  wird abgelehnt und die  $H_1$  wird angenommen.
- Hypothese 4: Der Mittelwert  $M_{rez3}= 32,87$  übersteigt den Mittelwert  $M_{rez}= 14,38$  mit  $p=.000$  hoch signifikant. Die  $H_0$  wird abgelehnt und die  $H_1$  wird angenommen.
- Hypothese 5: Der Mittelwert  $M_{rep2}= 20,51$  übersteigt den Mittelwert  $M_{rep0}= 0,0$  mit  $p=.000$  hoch signifikant. Die  $H_0$  wird abgelehnt und die  $H_1$  wird angenommen.
- Hypothese 6: Der Mittelwert  $M_{rep3}= 47,65$  übersteigt den Mittelwert  $M_{rep2}= 20,51$  mit  $p=.000$  hoch signifikant. Die  $H_0$  wird abgelehnt und die  $H_1$  wird angenommen.

Der Effekt der unabhängigen Variable *Textschwierigkeit* auf die abhängige Variable *Pupillenweite* hat eine Stärke von  $\eta^2= .548$  für den Bereich der Textrezeption. Für den Bereich der Textreproduktion beträgt die Effektstärke  $\eta^2= .618$ .

### 3.5.2.9 Zusammenfassung und Diskussion

Durch die experimentelle Untersuchung der Wirkung von Textrezeption und Textreproduktion auf die Pupillenweite konnte ein eindeutiger Zusammenhang aufgezeigt werden. Da der Text bis heute eines der wichtigsten Symbolsysteme zur medialen Informationsübermittlung ist, sind die Ergebnisse dieses Experiments von großer Bedeutung.

Es konnte aufgezeigt werden, dass sowohl eine objektivierete Schwierigkeitsbewertung (Vor-Rating) von Textvorlagen als auch die subjektiv erlebte mentale Beanspruchung bei der Textverarbeitung mit der Pupillenweite hoch korrelieren. Diese Ergebnisse waren jedoch erst nach umfangreichen Maßnahmen zur Datenbereinigung der Pupillenwerte zu erzielen. Die Aufbereitung der Pupillendaten scheint der Schlüssel zur erfolgreichen Analyse von Pupillendaten zu sein.

Für die Analyse von Überforderungen konnte die begründete Hypothese aufgestellt werden, dass bei einer kognitiven Überlastung eine spontane Pupillenverengung zu beobachten ist. Es ist davon auszugehen, dass die Handlung zur Bewältigung der Belastung plötzlich wegfällt und somit bei gleicher Belastung die mentale Beanspruchung stark abfällt (vgl. Kap. 3.3, Abschnitt I) Diese speziellen Bewegungsmuster bedürfen in Zukunft jedoch weiterer systematischer Untersuchungen.

Mit den hier vorgestellten Methoden der Datenaufbereitung und –analyse konnte bewiesen werden, dass sich die Pupillenanalyse zur Bestimmung der mentalen Beanspruchung eignet. Der enge Zusammenhang der untersuchten Parameter zeigt, dass die Pupillometrie ein geeig-

netes Verfahren zur Analyse der mentalen Beanspruchung bei der Textverarbeitung am Monitor darstellt. Insofern lassen sich die Ergebnisse zur Sprachverarbeitung und Pupillenstellung mit Wörtern und Sätzen von z.B. Beatty (1982), Ben-Nun (1986), Hyönä et al. (1995) und Just und Carpenter (1993) auch auf längere Texte übertragen. Die Ergebnisse von Schultheis (2004) sind somit wahrscheinlich auf methodische Probleme zurückzuführen. Insbesondere die fehlende Artefaktbereinigung und die mangelnde Standardisierung der Messwerte könnten zu diesem Ergebnis beigetragen haben. Außerdem sind die Abstufungen in der Textschwierigkeit in keiner Weise validiert.