

**FORSCHUNGSBERICHTE
des
PSYCHOLOGISCHEN INSTITUTS
der
ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG I. BR.**

Nr. 156

**Evaluation des
negativen Retrospektionseffektes.
Untersuchungen mit MONITOR**

**Jochen Fahrenberg, Kristina Bolkenius,
Sebastian Maier, Miriam Schmidt,
Friedrich Foerster, Paul Hüttner,
Christoph Käppler und Rainer Leonhart**

**Forschungsgruppe Psychophysiologie
Psychologisches Institut**

**Belfortstrasse 20
D-79085 Freiburg i. Br.**

Evaluation des negativen Retrospektionseffektes

Untersuchungen mit MONITOR

Inhaltsverzeichnis

	Zusammenfassung	5
1	Einleitung	9
1.1	Erinnerungsfehler und Retrospektionseffekte	9
1.2	Selbstberichte mit computer-unterstützter Methodik	11
2	Empirische Untersuchungen	13
2.1	Erste Untersuchung	13
2.2	Zweite Untersuchung	15
2.3	Dritte Untersuchung	15
2.4	Vierte Untersuchung	19
2.5	Fünfte Untersuchung	25
2.6	Sechste Untersuchung	25
2.7	Erklärungsversuche (Skalierungsprobleme) und Exkurse zu Methodenproblemen	32
2.8	Zusammenfassung	36
3	Fragestellungen der neuen Untersuchung	38
4	Methodik	39
4.1	Versuchsplan	39
4.2	Untersuchungsteilnehmer	39
4.3	Termine	39
4.4	Items	40
4.5	Skalierung	40
4.6	Skalierungstraining als experimentelle Bedingung	40
4.7	Hand-held PC und MONITOR	41
4.8	Durchführung	42
4.9	Fehlende Daten	42
4.10	Statistische Analyse	43
5	Ergebnisse	45
5.1	Variabilität u. a. deskriptive Befunde	45
5.2	Skalierungstraining und Ganztags-/Halbtags-Rückblick	45
5.3	Retrospektionseffekt	47
5.4	Relative Sicherheit	48
5.5	Prospektive Einstufung	50
5.6	Weitere deskriptive Ergebnisse	50
6	Diskussion und Ausblick	53
6.1	Akzeptanz	53
6.2	Negativer Retrospektionseffekt	53
6.3	Kritische Überlegungen zur internen und externen Validität	54
6.4	Ausblick	55
	Literaturverzeichnis	57
	ANHANG	
	Instruktionen	61
	Itemlisten und Tabellen	63

Zusammenfassung

Durch die computer-unterstützte Datenerhebung können viele Themen besser untersucht werden als bisher. Die Selbstberichte über Befinden, Beschwerden, Testergebnisse u. a. Daten können mit hand-held PC technisch und zeitlich zuverlässiger sowie mit höherer „ökologischer Validität“ unter alltagsnahen Bedingungen“ erfasst werden als mit den üblichen Fragebogen. Diese Vorzüge haben der neuen Methodik verschiedene Anwendungsgebiete erschlossen.

Dieser Forschungsbericht gibt eine Übersicht über eine Untersuchungsserie mit dem in Freiburg entwickelten Programm MONITOR für hand-held PC Psion Serie 3. Diese Untersuchungen hatten außer ihren speziellen Aufgaben auch eine gemeinsame Fragestellung, die von praktischem und theoretischem Interesse ist: In wie weit stimmen der am Abend gegebene Rückblick auf den vergangenen Tag mit den aktuellen, über den Tag verteilten Selbstberichten überein? Gibt es systematische Erinnerungstäuschungen?

Ein *negativer Retrospektionseffekt* zeigt sich, wenn der Tageslauf und das eigene Befinden im Rückblick auf den Tag oder die Woche negativer bewertet werden, als es aus den aktuellen Einstufungen während des Tages zu erwarten war. Dieser Sachverhalt wurde in sechs Untersuchungen hinsichtlich mehrerer Items, vor allem jener mit „negativem“ Inhalt, festgestellt.

Auf den verwendeten siebenstufigen Adjektivskalen haben negative Items (z. B. „bedrückt“) relativ niedrige Mittelwerte und positive Items (z. B. „körperlich wohl“) relativ hohe Mittelwerte. Die abends oder am nächsten Morgen bzw. nach einer Woche gegebenen rückblickenden Einstufungen lagen näher zur Skalenmitte. Es handelt sich um einen *robusten Befund mit mittlerer bis großer Effektstärke*. Deshalb liegt die Annahme nahe, dass es sich – zumindest teilweise – um ein psychometrisches Artefakt oder eine triviale Urteilsheuristik handeln könnte.

Die verschiedenen, einander nicht ausschliessenden Erklärungshypothesen sind:

(1) Die *Tendenz, Vergangenes negativer zu bewerten* als es aktuell erlebt wurde, ist ein bekanntes Phänomen. In Anamnesen wird u. U. von *Aggravation* oder *Klagsamkeit* gesprochen, wenn ein Patient zur negativen Überbetonung und Akzentuierung von Ereignissen oder Beschwerden neigt – vielleicht um Aufmerksamkeit zu gewinnen oder aus anderen Gründen.

(2) Ein *negativer Retrospektionseffekt* ergibt sich bei der verwendeten Skalierungsweise dann, wenn der Skalenwert des Rückblicks *größer* als der Mittelwert der aktuellen Einstufungen des Tages ist. Falls die Tages-Mittelwerte, wie es durchweg der Fall ist, am oberen oder unteren Ende der Skala liegen, wird eine *weiter in der Mitte liegende* retrospektive Einstufung zu diesem negativen Retrospektionseffekt führen. Da die Verteilungen der Itemwerte in der Regel nicht symmetrisch um die Skalenmitte 4, sondern links- oder rechtsschief sind, lässt der negative Retrospektionseffekt eine Verschiebung zur Skalenmitte hin erwarten: bei positiv formulierten Items (z. B. „aktiv“ Stufe 7 völlig zutreffend) eine Abnahme zur Skalenmitte hin, bei negativ formulierten Items (z. B. „bedrückt“ gar nicht zutreffend) eine Zunahme zur Skalenmitte hin.

Eine Tendenz zur Mitte kann jedoch aus verschiedenen Gründen zustande kommen:

(a) Eine *statistische Regression zur Mitte* ist durch Messfehler bedingt. Diese führen dazu, dass der Wert einer zweiten Messung zum Mittelwert (Dichtezentrum) der Verteilung tendiert.

(b) Beim Rückblick steht grundsätzlich der gesamte Range der Skala von 1 bis 7 zur Verfügung. Unter der Annahme, dass sich die rückblickende Einstufung wahrscheinlich mehr oder minder an einem gedachten Tagesdurchschnitt orientiert, könnte es zu einer Varianzeinengung des Rückblicks im Vergleich zum Tages-Mittelwert, d. h. einer *Restriktion des Bereichs*, kommen (bei einem gedachten Tages-Mittelwert 2, würde der Rückblick nicht unter das Skalende bei 1 gehen können).

(c) Die Selbsteinstufung des Befindens auf einer Skala ist, wenn diese Aufgabe gründlich aufgefasst wird, mit semantischen Schwierigkeiten und mit der Wahl eines Bezugssystemen verbunden. In dieser Situation der Unsicherheit könnte eine *Urteilsheuristik* eingesetzt werden, die mittlere, d. h. unentschiedene Skalenstufe „mittelmäßig“ zu bevorzugen.

(3) Zwischen der rückblickenden Einstufung am Abend und den vorausgegangenen Einstufungen kann eine *Ausgangswert-Abhängigkeit* (ähnlich dem Ausgangswertproblem bei physiologischen Messwiederholungen) bestehen (siehe unten).

Neben diesen psychometrischen und kognitionspsychologischen Erklärungshypothesen steht

(4) eine persönlichkeitspsychologische Interpretationen. Es scheint ein Zusammenhang zwischen dem Betrag des Retrospektionseffektes und bestimmten Persönlichkeitseigenschaften zu bestehen, insbesondere der Skala *Emotionalität* (FPI-R). In diesem Ergebnis stimmten zwei Untersuchungen überein: Emotional labile, selbstunsichere, sich stark beansprucht fühlende Personen tendieren dazu, Tagesereignisse und Befinden im Rückblick negativer zu bewerten als andere Personen. Dieser Zusammenhang ist über die gegenwärtige Fragestellung hinaus als ein Validierungshinweis für die FPI-R-Skalen im Hinblick auf wiederholte Selbstberichte unter Alltagsbedingungen interessant. Die Korrelation war jedoch nicht konsistent bei allen Items vorhanden und konnte in einer dritten Untersuchung nur tendenziell repliziert werden.

Personen mit hoher Ausprägung der Emotionalität haben bekanntlich die Tendenz, über viele körperliche Beschwerden zu klagen, geringe Lebenszufriedenheit zu äußern und ein negatives Krankheitsverhalten zu entwickeln. Dies wurde bevölkerungsrepräsentativ (siehe Fahrenberg, 1994b; Fahrenberg et al., 2001a) und in großen klinischen Studien (Myrtek, 1998) belegt.

Sollte es sich beim negativen Retrospektionseffekt aber nicht nur um einen psychometrischen, sondern um einen „psychologischen“ Effekt handeln, wäre dies zweifellos wichtig für jede psychologisch-diagnostische Untersuchung, die sich auf Selbstberichte in Fragebogen und Interviews stützt.

In einer Übersicht wurde eine Serie von bisher sechs Untersuchungen zu diesem Thema referiert und verschiedene Voraussetzungen und methodische Aspekte diskutiert. Dazu gehören u. a. der Nachweis einer ausreichenden Variabilität (intra- und inter-individuell), die Exploration der statistischen Verteilungsformen der Itemwerte, Unterschiede zwischen „positiv“ und „negativ“ formulierten Adjektivskalen, die Generalisierbarkeit über verschiedene Untersuchungsgruppen, die Akzeptanz der Methodik und die methodenbedingte Reaktivität. Die Untersuchungsserie regte außerdem zur Weiterentwicklung des Itempools und der Ablaufsteuerung für MONITOR an.

Die verschiedenen Erklärungshypothesen und Untersuchungsergebnisse legten eine weitere Untersuchung nahe, insbesondere im Hinblick auf die vermutete Urteilsheuristik. Ist diese Tendenz zur Mitte weniger ausgeprägt, wenn (1) der Rückblick nur für *halbe* Tage statt für den ganzen Tag verlangt wird, (2) die aktuellen Einstufungen mit relativ *großer Sicherheit* gegeben werden und (3) eine bessere Differenzierung durch ein vorgeschaltetes *Skalierungstraining* erreicht wird? Ist die Tendenz zur Mitte stärker, wenn (4) Unsicherheit provoziert wird, indem eine *prospektive* Einstufung für den folgenden Tag erbeten wird.

Die neue Untersuchung mit 64 Studierenden (52 Frauen, 12 Männer) umfasste vier Termine am ersten und sechs Termine am zweiten Erhebungstag. Jede Abfrage beinhaltete drei Fragen zum Setting, eine Aufforderung, besondere Ereignisse durch Texteingabe zu protokollieren, sowie acht Adjektivskalen zum Befinden. Hier wurden visuelle Analogskalen mit einer für die Einstufer nicht sichtbaren 21-stufigen Auflösung verwendet. Die Hälfte der Teilnehmer erhielt die Fragen zur retrospektiven Einstufung um 13 und 21 Uhr, die andere Hälfte nur abends für den gesamten Tag. Außerdem gab es abends sowohl eine Frage zur relativen Sicherheit der Einstufungen als auch eine prospektive Frage zum Befinden am nächsten Tag. Die aktuellen und die retrospektiven Einstufungen wurden mit computerunterstützter Methodik, d. h. dem Programm MONITOR für hand-held PC PSION Serie 3, unter Alltagsbedingungen gewonnen. Vor dem ambulanten Monitoring erhielt die Hälfte der Teilnehmer ein kurzes Skalierungstraining.

Der negative Retrospektionseffekt konnte bei vier von acht Items repliziert werden: bei den negativ gepolten Items („anstrengend, belastend“; „aufgeregt, nervös“; „ärgerlich, gereizt“; „bedrückt“). Eine zwei-faktorielle Varianzanalyse ergab *keine* Mittelwertunterschiede der aktuellen und der retrospektiven Einstufungen zwischen den Gruppen mit „Skalierungstraining“ bzw. mit „Retrospektion in Tagesabschnitten“. Der beobachtete negative Retrospektionseffekt war unabhängig von der subjektiven Unsicherheit der Selbsteinstufungen und war bei der naturgemäß besonders unsicheren Vorhersage für den nächsten Tag nicht stärker ausgeprägt. Diese negativen Ergebnisse sprechen insgesamt *gegen die Annahme von Urteilstendenzen bzw. einer Urteilsheuristik* als hinreichender Erklärung der Tendenz zur Mitte.

Als Ergebnis der Serie von Untersuchungen kann immerhin der *Geltungsbereich* des negativen Retrospektionseffektes eingeschränkt werden. Es handelt sich zwar um einen *robusten Effekt*, aber er tritt vorwiegend bei bestimmten Adjektivskalen auf: es sind Items mit *negativem* Inhalt und unter diesen vor allem die Items „anstrengend, belastend“, „bedrückt“ und „Erlebten Sie seit der letzten Eingabe Stress?“ Vielleicht regen gerade diese Iteminhalte an, zustimmende Antworten zu geben. Die Wochenstudie zeigte jedoch, dass die Stabilität (Reproduzierbarkeit) des negativen Retrospektionseffektes über die Tage bemerkenswert gering ist. Dies spricht gegen die Annahme einer persönlichkeitsabhängigen habituellen Urteilstendenz. In der Wochenstudie zeigte sich die Beziehung des negativen Retrospektionseffektes mit dem Konstrukt „Emotionalität“ nur am ersten Tag. Die persönlichkeitspsychologische Hypothese müsste folglich modifiziert werden.

Zusammenfassend werden die konkurrierenden Erklärungen des negativen Retrospektionseffektes diskutiert und mögliche Fragestellungen weiterführender Studien skizziert.

Der vorliegende Forschungsbericht konzentrierte sich auf die Evaluation des negativen Retrospektionseffektes (mit Reanalysen zu bestimmten Fragen) und auf die Dokumentation dieser Untersuchungsserie. Jede der sechs Untersuchungen hatte noch eine Anzahl anderer Fragestellungen. Die meisten dieser Ergebnisse wurden bereits veröffentlicht. Hier werden jedoch außer einer Zusammenfassung zahlreiche Tabellen für Vergleichszwecke und ggf. Reanalysen mitgeteilt.

Deskriptoren

Ambulantes Monitoring, Assessment, Emotionalität, Erinnerungstäuschungen, Feldstudie, Freiburger Persönlichkeitsinventar (FPI-R), (negativer) Retrospektionseffekt, Rückschaufehler, Selbstbericht, Selbsteinstufungen, Tagebuch.

Summary

A series of monitoring (ambulatory assessment) studies was reviewed. All used MONITOR, that is a computer-assisted methodology for Psion hand-held PC Series 3. This methodology is especially suited to assessing daily moods, daily hazzles, physical complaints etc., under naturalistic conditions. Although the individual studies were related to a number of different issues, they had in common the evaluation of the negative retrospection effect in self-report data.

A retrospection-effect (retrospection bias, re-call error) is present when the evaluation of the previous day, that was obtained in the evening, differs from the actual self-ratings.

Consistently a negative retrospection effect of average to high effect size was found, that is, in the evening, participants rated the day more strenuous and annoying, and their mood more depressed, tired, nervous etc., than could be expected on the basis of their average actual ratings recorded during the day.

Based on the previous investigations a new study was designed to investigate certain issues that seemed to be involved, for example, regression to the mean, restriction of range, initial values dependencies, and heuristic strategies in judgement – being confronted with uncertainty. Furthermore, different intervals of retrospection, a pre-monitoring training, and a prospective rating were used to investigate possible influences of various degrees of uncertainty in self-reporting.

Findings in 64 student subjects indicated the presence of a negative retrospection effect in four „negative” adjective scales in contrast to „positive” items which showed no such bias. A number of hypotheses pertaining to various aspects of the process of responding and judgement were investigated. The findings did not suggest any sufficient psychometric explanation or interpretation based on cognitive aspects. The negative retrospection effect appears to be a continuing challenge to method studies.

This research report is provided as a summary of previous and recent findings and, thus, includes a number of tables, which may be helpful for further research and meta-analysis.

Descriptors

Ambulatory assessment, diary study, emotionality, Freiburg Personality Inventory (FPI-R), field study, re-call error, retrospection bias, (negative) retrospection effect, self-report.



1 Einleitung

1.1 Erinnerungsfehler und Retrospektionseffekte

Wer sich an frühere Ereignisse zu erinnern versucht, lang zurückliegende oder gerade vergangene, kann u. U. *Erinnerungstäuschungen* unterliegen. Eindeutige *Erinnerungsfehler* sind dann festzustellen, wenn unabhängige und objektive Beobachtungen einen anderen als den erinnerten Sachverhalt belegen. Zweifellos wird es auch beim Rückblick auf subjektive Erlebnisse, Emotionen und Stimmungen oder bei der Schilderung von körperlichen Beschwerden häufig Erinnerungstäuschungen geben können. Trifft der Rückblick zu oder nicht? In der Regel wird die Frage nach solchen Abweichungen wegen fehlender objektiver Kriterien nicht genau zu beantworten sein. Deswegen wird bei Selbstberichten statt von Erinnerungsfehlern („recall error“) besser von *Retrospektionseffekten* („retrospection bias“) gesprochen, um eine in der Rückschau auftretende Abweichung bzw. Verzerrung zu kennzeichnen.

Die Untersuchung der Retrospektionseffekte verlangt nach einer geeigneten Methodik, damit die retrospektiven mit den aktuellen Einstufungen verglichen werden können. Darüber hinaus wird hier jeder wissenschaftliche Untersuchungsansatz mit grundsätzlichen Vorbehalten verbunden sein müssen, da schwierige quantitative Vergleiche zwischen introspektiven Aussagen angestellt und über viele Personen statistisch zusammengefasst werden.

Erinnerungstäuschungen und Retrospektionseffekte sind nicht allein ein Thema der Gedächtnisforschung und der allgemeinen Psychologie der Urteilsbildung mit Blick auf Bezugssysteme, Urteilsheuristiken und andere Teilprozesse der Informationsverarbeitung. Die Retrospektionseffekte bilden ein fundamentales Problem aller psychologischen Diagnostik, die sich auf Selbstberichte stützt.

Allgemeinbefinden, emotionale Zustände, psychische Symptome oder körperliche Beschwerden werden in der Regel mit Selbsteinstufungsskalen bzw. durch Fragebogen oder durch ein Interview erfasst. Häufig wird mittels Stimmungsskalen und Symptomlisten sowohl nach dem *gegenwärtigen* Zustand als auch nach einem *Rückblick* auf den vergangenen Tag oder auf die letzte Zeit (Woche, Monat) gefragt.

Hier ist zwischen zwei differentiell-psychologischen Perspektiven zu unterscheiden: Erstens können die von Person zu Person verschiedenen, *individuellen* Inhalte und Erinnerungstäuschungen über zurückliegende Erlebnisse und Beschwerden interessieren. Zweitens kann untersucht werden, ob darüber hinaus *systematische*, für viele Personen zutreffende Trends und Tendenzen zu erkennen sind. Um diese zweite Aufgabe geht es in der vorliegenden Arbeit. Die Untersuchungen verlangen gründliche methodische Überlegungen, um die verschiedenen Komponenten unterscheiden zu können.

Es wird hier die These vertreten, dass das Thema systematischer Retrospektionseffekte bei den alltäglichen Selbstberichten über Erlebnisse, Emotionen und Beschwerden – im Unterschied zur experimentellen Gedächtnisforschung – bisher zu wenig erforscht wurde.

Retrospektionseffekte

Über Retrospektionseffekte wurde in einer Anzahl klinisch-psychologischer Untersuchungen berichtet (siehe auch De Beurs, Lange & Van Dyck, 1991; Hedges, Jandorf & Stone, 1985; Smith & Safer, 1993; Margraf & Jacobi, 1997; Käßler, Brügger & Fahrenberg, 2001). Einige der Befunde werden hier kurz referiert.

Die Erinnerung an positive und negative Lebensereignisse scheint sowohl bei depressiven Patienten als auch bei Studenten von der aktuellen, positiven oder gedrückten Stimmung abhängig zu sein (Clark & Teasdale, 1982; Teasdale & Fogerty, 1979). Aktuelle Einstufungen, wie bestimmte Belastungen (Stress) bewältigt werden, stimmen mit retrospektiven Einstufungen nicht gut überein (Stone et al., 1998). Die Erinnerung an Menstruationsbeschwerden kann von impliziten Konzepten der Frauen

über ihren körperlichen und emotionalen Status beeinflusst sein und ist bei labilen Frauen u. U. retrospektiv stärker ausgeprägt als initial eingestuft (Ross, 1989). Intraindividuelle Vergleiche und Korrelationen zeigen, dass u. U. Kontrasteffekte, d. h. auffällige Übergänge zwischen vermehrter Tagesbeanspruchung („daily hassles“) und besserer Stimmung am folgenden Tag auftreten können (De Longis, Folkman & Lazarus, 1988). Die Genauigkeit der Erinnerung an Rückfälle, negative Emotionen und die Anzahl der Zigaretten während der Raucherentwöhnung (im Vergleich zu den Tagebuchdaten) hängt mit dem aktuellen Status als Raucher/Nichtraucher zusammen (Shiffman et al., 1997). Das aktuelle Ausmass von Schmerzen beeinflusst die Erinnerung an vorausgegangene Schmerzzustände (Smith & Safer, 1993). Zum Thema „falsche Erinnerungen“ existiert eine umfangreiche Literatur, die sich vor allem auf das langfristige autobiographische Gedächtnis oder auf Erinnerungstäuschungen im Rahmen der forensischen Psychologie bezieht (Ross, 1991; Rubin, 1986).

Beim Vergleich zwischen der von Patienten *akut* erlebten Angst und der am Abend *rückblickend* geschilderten Angst beobachteten Margraf, Taylor, Ehlers, Roth und Agras (1987; Margraf & Jacobi; 1997) eine auffällige Tendenz, im Rückblick tendenziell eine relativ größere Angst anzugeben. Der beobachtete Effekt könnte jedoch zumindest teilweise auf Unterschiede der Methodik zurückzuführen sein, denn die Einstufungen wurden mit drei Methoden gewonnen: einem Tagebuch, einem Fragebogen und einem Interview. Zur Untersuchung möglicher Retrospektionseffekte ist jedoch die ambulante Datenerhebung mit einem hand-held PC die Methode der Wahl, um die aktuellen und die retrospektiven Selbsteinstufungen methodisch vergleichbar und alltagsnah zu erfassen. Die Methodenentwicklung auf dem Gebiet des ambulanten Assessment kann gerade für diese Fragestellung gut genutzt werden (siehe unten).

Es liegt nahe, nach den Bezugssystemen der aktuellen und der retrospektiven Einstufungen zufragen. So haben z. B. Hedges et al. (1985) einen systematischen Effekt in rückblickenden Selbstbeurteilungen der Stimmung beschrieben. Die Probanden stufte zu vier Zeitpunkten am Tag, d. h. 9, 13, 16 und 19 Uhr, ihre Stimmung ein und gaben außerdem um 20 Uhr eine Einstufung des gesamten Tages. Dieser abendliche Rückblick hing eher mit den Maxima der aktuellen Einstufungen als mit deren Mittelwert zusammen.

Retrospektive Urteile können u. a. durch die zeitliche Nähe und emotionale Bedeutung, durch die Formulierung der Fragen und Antwortkategorien, durch implizite Konzepte und Erwartungen beeinflusst werden. Für Erinnerungstäuschungen gibt es viele mögliche Quellen (siehe Hank, Schwenkmezger & Schumann, 2001; Hufford et al., 2001; Ross, 1991; Schwarz, 1990; Suh, Diener & Fujita, 1996; Thomas & Diener, 1990; Wheeler & Reis, 1991). Erinnerungstäuschungen können auch in differentiell-psychologischen und klinisch-psychologischen Untersuchungen, die sich – wie mehrheitlich der Fall – auf Interviews, Fragebogen, Tagebücher oder Self-Monitoring stützen, ein großes Methodenproblem bedeuten. Die retrospektiven Einstufungen stellen zwar ebenfalls eine psychologische Wirklichkeit dar, doch kann es für viele der anschließenden diagnostischen Entscheidungen und Maßnahmen außerordentlich wichtig sein, ob nachträglich z. B. eine gravierende Neubewertung und Gewichtung – als Nivellierung (Verringerung) oder Aggravation (Verstärkung) stattgefunden hat.

Von einem Rückschaufehler („hindsight bias“) wird gesprochen, wenn eine frühere Aussage an zwischenzeitlich bekannt gewordene Ereignisse angeglichen wird (Hawkins & Hastie, 1990). An dieser und Stelle kann nicht die komplizierte Diskussion über Bezugssysteme, mögliche Verankerungen und Urteilsheuristiken aufgenommen werden (z. B. Hank et al., 2001; Mogel, 1990; Mussweiler, Förster, & Strack, 1997; Ruoss, 1997; Schwarz, 1990). Die experimentellen Untersuchungen zum Rückschaufehler (hindsight bias) bieten hier aus zwei Gründen kaum geeignete Erklärungsansätze: es gibt hinsichtlich der Befindlichkeit keine „Lösungsinformation“, die gedächtnisbasiert oder mittels hypothetischer Urteile geprüft werden könnte, und die Urteile über die aktuelle Befindlichkeit erfolgen unter naturalistischen Bedingungen. Die Konzepte und Operationalisierungen der experimentellen Rückschau-Paradigmen sind folglich kaum übertragbar. Erwähnenswert ist jedoch, dass dem subjektiven Wissen und der subjektiven Repräsentation von Abweichungen in den neueren Inferenzmodellen der Urteilsbildung offensichtlich mehr Bedeutung eingeräumt wird (Werth, 1998).

Die Rekonstruktion der Befindlichkeit im Tageslauf – und das selbstreflektierte Wissen davon – werden höchstens in einigen Aspekten exploriert werden können. Wie repräsentativ ist aus Sicht der Probanden jede der aktuellen Einstufungen und wie interpretieren die Probanden ihren Rückblick und die negativen Abweichungen?

Der Begriff „Retrospektionseffekt“ bezieht sich also auf ein größeres Problemfeld mit mehreren Fragestellungen und Anwendungsperspektiven. Die empirischen Studien können hier nicht die Prägnanz einer experimentalpsychologischen Untersuchung von Erinnerungsfehlern im engeren Sinn erreichen, sondern werden noch sehr viel mehr davon abhängen, welche Komponenten eines Selbstberichts ausgewählt und welche Operationalisierungen verwendet werden.

Negativer Retrospektionseffekt

Die eigenen Untersuchungen haben sich auf einen bestimmten Ausschnitt dieses Problemfeldes konzentriert: den *negativen* Retrospektionseffekt. *Negativer Retrospektionseffekt bedeutet, dass ein früherer Zustand, ein Erlebnis, ein Tageslauf oder eine ganze Lebensphase im Rückblick negativer geschildert oder auf andere Weise abgewertet wird.* Diese Urteilstendenz ist durch den Vergleich zwischen den aktuell und den retrospektiv erhobenen Selbsteinstufungen zu erfassen. Die Serie der Freiburger Untersuchungen, die sich in der Regel auf Tages-Rückblicke beziehen, sollte u. a. die folgenden Fragen beantworten: Wie gut ist der Effekt replizierbar und generalisierbar? Welche Effektstärke zeigt sich? Welche Erklärungsversuche lassen sich gewinnen? Darüber hinaus hatte jede dieser Untersuchungen zusätzliche Fragestellungen, die in der anschließenden Übersicht ebenfalls kurz referiert werden.

1.1 Selbstberichte mit computer-unterstützter Methodik

Seit preiswerte hand-held PC verfügbar sind, müssen schriftliche Tagebücher und Protokollblätter weitgehend als überholt angesehen werden. Schriftliche Protokollierungen sind zwar kostengünstig, jedoch unzuverlässiger und für viele Zwecke zu unflexibel. Als spezifische Nachteile sind zu nennen:

- Unsicherheit über den tatsächlichen Zeitpunkt des Protokolls, denn u. U. sehr verspätete oder erst am Abend summarisch nachgetragene Selbstberichte können nicht von pünktlichen Eingaben unterschieden werden;
- die vorausgegangenen Protokolle bleiben zugänglich, so dass es eher zu Transfereffekten und Stereotypisierungen kommen könnte;
- ein Formular ist für sich verzweigende Fragen und für kompliziertere Abfragen weniger geeignet als ein flexibles Computerprogramm;
- die Daten eines Protokollblattes müssen in zeitraubender Weise übertragen werden, wobei es Transferfehler geben kann.

Aus diesen Gründen ist heute oft der PC als Methode der Wahl für die Protokollierung von Selbstberichten unter Alltagsbedingungen anzusehen. Die meisten, wenn nicht sogar alle Anwendungen sind mit einem hand-held PC technisch zuverlässiger und methodisch ergiebiger auszuführen (siehe auch Buse & Pawlik, 1996, 2001; Fahrenberg, Hüttner & Leonhart, 2001; Fahrenberg et al., 2002; Hufford et al., 2001; Myrtek & Fahrenberg, 2001; Myrtek, Foerster & Brügger, 2001).

Die systematische Anwendung von programmierbaren hand-held PC zur Protokollierung von Selbstberichten unter naturalistischen Bedingungen wurde von Pawlik und Buse (1982) eingeleitet. In den angloamerikanischen Ländern sind entsprechende Projekte mit einem *computer-assisted diary* („*electronic diary*“), trotz einiger Vorläufer, erst in deutlichem Abstand zu verzeichnen (Hedges et al., 1990; Stone & Shiffman, 1994; Taylor, Fried & Kennardy, 1990; siehe auch Hufford et al., 2001). AMBU (Pawlik & Buse, 1999) und MONITOR (Brügger, 1998) sind Software-Entwicklungen, die als vollständiges System oder als Software-Lizenz allgemein zugänglich sind. Hier muss hervorgehoben werden, dass hand-held PC natürlich nicht allein für Selbstberichte, sondern darüber hinaus für die Protokollierung von Verhaltensbeobachtungen, für Verhaltensmessungen und für das Monitoring physiologischer Veränderungen im Alltag genutzt werden können.

Vorzüge der computer-unterstützten Methodik für Selbstberichte:

- automatische Alarm- und Erinnerungsfunktionen;
- flexible Gestaltung des Layouts der Fragen, maßgeschneiderte Antwortmöglichkeiten und verschiedene Antwortmodi (Skalen, Listen, Texteingabe, Verzweigungen usw.) in sequentiell oder hierarchischem Aufbau;
- zuverlässige Ablaufsteuerung mit vielen Optionen;
- Unzugänglichkeit der vorausgegangenen Antworten und Unmöglichkeit nachträglicher Eingaben oder Änderungen;
- zuverlässiges zeitliches Protokoll aller Eingaben (einschließlich der Antwortlatenzen);
- Vermeidung von Retrospektionseffekten;
- automatische Datensicherung;
- direkter Datentransfer für statistische Weiterverarbeitung (Ökonomie und Reduktion von Fehlerquellen).

Diese Vorzüge der Methodik legen eine breitere Anwendung nahe. Das computer-unterstützte Assessment mit hand-held PC hat – wie jede psychologische Methodik auch – einige Nachteile bzw. Besonderheiten, die unter bestimmten Bedingungen störend sind (eine Diskussion der Möglichkeiten und Grenzen dieser Methodik siehe Fahrenberg et al., 2001b, 2002).

Eine Programmbeschreibung von MONITOR für hand-held PC der PSION-Serie und weitere Informationen sind veröffentlicht worden (siehe Fahrenberg et al, 2002, Anhang; Hüttner, 2001) sowie huettnr@psychologie.uni-freiburg.de.

2 Empirische Untersuchungen

Übersicht

Die hier referierten Untersuchungen hatten verschiedene Ziele. Zu diesen gehörten u. a. Befindlichkeit im Tagesverlauf unter Alltagsbedingungen, Blutdruckverhalten und Befindlichkeit von Hypertonikern (Versuche zur Symptom-Kontext-Analyse), Kovariation von Aufmerksamkeit und subjektiver Wachheit/Müdigkeit im Tageslauf mit dem Thema Morgentyp/Abendtyp. Alle Untersuchungen waren außerdem Methodenstudien, welche die Differenzierungsleistung der Items u. a. testmethodische Eigenschaften sowie die Akzeptanz und die methodische Reaktivität bei computer-unterstützten Erhebungen dieser Art erkunden sollten.

Die Frage nach dem negativen Retrospektionseffekt war also nur eine, wenn auch besonders interessant erscheinende Fragestellung. Um diesen Effekt adäquat prüfen zu können, muss zunächst gezeigt werden, dass die verwendete Methodik überhaupt geeignet ist, die *Variabilität* der Befindlichkeit abzubilden. Deswegen werden in den nächsten Abschnitten (und durch Tabellen im Anhang) u. a. die statistischen Verteilungen der Antworten (Schiefe, Kurtosis) und die relative Anteile der Varianzkomponenten (Personen, Tage, Termine und deren Wechselwirkungen) beschrieben. Die Tabelle 2.1 vermittelt eine Übersicht über diese Studien.

Diese Untersuchungen haben viele Gemeinsamkeiten, doch wurden einige vielleicht maßgebliche Details variiert. In allen Untersuchungen dienten Adjektivskalen zur Selbsteinstufung des Befindens (siehe Anhang), in einigen Untersuchungen auch Tests zur Prüfung der Aufmerksamkeitsleistung. Außer einigen Standarditems, die unverändert blieben, wurden zusätzliche Items entworfen, um verschiedene Aspekte dieser aktuellen und retrospektiven Selbstprotokollierung zu erfassen.

In den folgenden Abschnitten werden die wichtigsten Ergebnisse der sechs Untersuchungen und die zugrundeliegenden Überlegungen relativ ausführlich dargestellt, um den Untersuchungsplan der im Kapitel 3 bis 5 beschriebenen neuen Untersuchung vorzubereiten. Der vorliegende Bericht dient außerdem der Dokumentation und ermöglicht damit weitere Re-Analysen.

2.1 Erste Untersuchung

Käppler (1994) konnte im Rahmen einer Untersuchung zum ambulanten psychophysiologischen Blutdruck-Monitoring mehrmals am Tag aktuelle Selbstberichte erheben und mit der retrospektiven Selbsteinstufung des gesamten Tages vergleichen. Hier wurde zum ersten Mal hand-held PC zur Untersuchung von Retrospektionseffekten verwendet. Diese computer-unterstützte Untersuchungsmethodik hat den Vorzug der relativen Alltagsnähe und ermöglicht eine höhere technische Zuverlässigkeit als ein Fragebogen.

In dieser Monitoringstudie wurden bei 42 Studierenden an einem Tag durchschnittlich 29 Selbstprotokolle mit jeweils vier Fragen zum Setting sowie mit zwölf siebenstufigen Skalen zum Befinden, zur Stimmung und zur Beanspruchung erhoben. Beim Vergleich der gemittelten Tageswerte mit dem morgendlichen Rückblick anhand derselben Fragen zeigten sich in sieben von zehn Items zum Befinden signifikante ($p = .05$) bis hoch signifikante ($p = .01$) Unterschiede: rückblickend wurde der Tag als belastender, körperlich und geistig anspannender, und die Stimmung als stärker aufgeregt/nervös und ärgerlich/gereizt beschrieben als aufgrund der Mittelwerte der über den Tag verteilten Einstufungen zu erwarten war. Dieser Befund wurde als Retrospektionseffekt mit einer einheitlich *negativen* Verzerrungstendenz beschrieben (siehe auch Käppler, Becker & Fahrenberg, 1993). Im Rückblick wurden außerdem die erlebten Situationen als typischer bewertet und die Bewegungsaktivität am vergangenen Tag wurde im Vergleich zum Mittelwert der aktuellen Einstufungen höherreingestuft.

Diskrepanzen zwischen aktuellen und retrospektiven Selbstberichten über Befinden und Beschwerden im Tageslauf wurden auch von anderen Untersuchern berichtet (siehe De Beurs et al., 1991; Hedges et al., 1985). Die dort verwendeten Papier-und-Bleistiftverfahren haben jedoch methodische Nachteile (siehe auch Hank, Schwenkmezger & Schuman 2001).

An diese erste Freiburger Studie zum negativen Retrospektionseffekt schlossen sich ähnliche Untersuchungen an mit dem Ziel, den Effekt zu replizieren und wichtige Bedingungen aufzuklären.

Tabelle 2.1: Übersicht über die Untersuchungen zum Retrospektionseffekt

Untersuchung	Hauptsächliche Fragen	Personenauswahl	Anzahl	Besonderheiten	Zeitraaster	Items	Hauptsächliche Ergebnisse
Käppler (1994), Käppler, Becker & Fahrenberg (1993)	Psychophysiologisches Blutdruck-Monitoring Retrospektionseffekt, zeitliche Stabilität	Studierende verschiedener Fakultäten, Teilnehmehonorar	N = 41 (M)	Zahlreiche Selbstberichte wegen spezieller Honorierung	1 Tag durchschnittl. 29 Termine	4 Settingmerkmale, 12 Adjektivskalen	Negativer Retrospektionseffekt in 7 von 10 Befindens-Items,
Frank, Herrmann & Fahrenberg (1996), Fahrenberg, Foerster & Franck (1995), Fahrenberg et al. (1995), Foerster, 1995)	Blutdruckverhalten im Tageslauf, Reaktionsskalierung, Ausgangswert-Abhängigkeit, Stabilität	Hypertoniker in einer Reha-Klinik	N = 50 (M)	Versuche zur Symptom-Kontext-Analyse	3 Tage durchschnittl. 25 Termine	7 Settingmerkmale, 7 Adjektivskalen	Negativer Retrospektionseffekt in 7 Items, Individualität der Blutdruckverläufe, emotionale Blutdruckepisoden relativ selten
Fahrenberg, Brügger, Foerster & Käppler, (1999), Käppler, Brügger & Fahrenberg (2001)	Befinden und Aufmerksamkeit im Tageslauf, Replikation des Effektes, Persönlichkeitsmerkmale FPI-R	Studierende verschiedener Fakultäten, Teilnehmehonorar	N = 61 (37 F, 24 M)	2 Aufmerksamkeits-tests MANOVA (Tage, Termine), Trends	2 Tage, 5 Termine Rückblick abends und morgens	4 Settingmerkmale, 10 Adjektivskalen, 2 offene, 1 Metafrage (störend)	Negativer Retrospektionseffekt in 5 von 6 negativen Befindens-Items, mittlere bis große Effektstärke, (bei positiven Items fehlend), Effekt korreliert mit Skala Emotionalität N (FPI-R),
Käppler & Rieder (2001)	Generalisierbarkeit bzw. interkulturelle Unterschiede des Effektes	Studierende verschiedener Fächer, Teilnehmehonorar	N = 40 (21 F, 19 M)	2 Aufmerksamkeits-tests	2 Tage, 5 Termine	4 Settingmerkmale, 10 Adjektivskalen, 1 Metafrage (störend)	Negativer Retrospektionseffekt in 7 von 8 Befindens-Items repliziert
Scheibehenne et al. (2000)	Generalisierbarkeit des Effektes, Korrelation mit FPI-R, Replikation	Berufstätige Teilnehmehonorar	N = 59 (34 F, 25 M)	Freie Texteingaben, Metafragen	1 Tag, 5 Termine	3 Settingmerkmale, 14 Adjektivskalen, 8 Metafragen	Negativer Retrospektionseffekt in 8 von 14 Items repliziert, Korrelation mit Skala N z.T. bestätigt, Metafragen liefern psychologische Hinweise
Fahrenberg, Leonhart & Foerster (2001), Leonhart (2001)	Befindlichkeit im Wochenverlauf, zeitliche Trends, Replikation des Effektes und Bedingungsanalyse	Studierende der Psychologie (überwiegend 2. FS) Testat für Praktikum	N = 33 (33 F)	1 Aufmerksamkeits-test, Wochenrückblick, MANOVA (Tage, Termine), Trends	7 Tage (versetzt), 6 Termine	3 Settingmerkmale, 14 Adjektivskalen, 8 Metafragen	Negativer Retrospektionseffekt in 4 von 6 negativen und (gegenläufig) in 1 von 4 positiven Befindens-Items, mittlere Effektstärken, im Wochenrückblick zunehmend
Aktuelle Untersuchung (siehe Kapitel 3 bis 4)	Trainingsbedingung, Ganztags- und Halbtags-Rückblick, Einstufung von Sicherheit, prospektive Einstufung	Studierende verschiedener Fächer (überwiegend Psychologie im 2. FS), Testat für Praktikum	N = 64 (52 F, 12 M)	Visuelle Analogskala (20 Stufen)	2 Tage, 4 + 6 Termine	3 Settingmerkmale, 10 Adjektivskalen, Sicherheit und prospektive Einstufung	Negativer Retrospektionseffekt in 4 von 4 negativen und 1 von 4 positiven Befindens-Items, Halbtags-/Ganztags-Bedingung, Training, subj. Sicherheit ohne Einfluss, Korrelation mit Skala N nur teilweise bestätigt

Anmerkung: In allen Untersuchungen wurden hand-held PC (CASIO PB 1000, PSION Serie 3) mit verschiedenen Entwicklungsstufen der MONITOR-Software verwendet.

2.2 Zweite Untersuchung (Hypertoniker)

Mit Patienten einer Rehabilitationsklinik wurde ein psychophysiologisches Monitoring durchgeführt. Beim ersten Untersuchungstermin ca. 10 Tage nach Beginn des Reha-Verfahrens wurde eine 2 x 24-Stunden-Registrierung vorgenommen, eine weiteres 24-Stunden-Monitoring folgte durchschnittlich 10 Tage später. Die Patienten waren Männer (22 – 60 Jahre, $M = 47.7$) und hatten eine essentielle Hypertonie verschiedener WHO-Stadien (Franck, Herrmann & Fahrenberg, 1996). Von 51 Untersuchungsteilnehmern lagen mit dem hand-held PC CASIO PB 1000 protokollierte Selbstberichte vor. Das physiologische Mess-System Physioport (Par-Port, Berlin) löste im Intervall von 30 ± 5 Minuten (nachts 60 Minuten) Blutdruckmessungen aus. Anschließend sollten die Patienten Settingmerkmale und Befinden eingeben (computer-gestützte Patientenführung siehe Anhang). Es wurden durchschnittlich 25 Eingaben am Tag registriert.

Die Untersuchung hatte mehrere Fragestellungen: die psychophysiologische Analyse des Blutdruckverhalten und der Kovariation mit dem Befinden, Versuche zur Symptom-Kontext-Analyse von Blutdruckepisoden, Vergleich der Blutdruck-Reaktivität zwischen standardisierten Settings (Tests) und ambulantem Monitoring („Labor-Feld-Vergleich“ innerhalb der Klinik), Skalierung von Blutdruckreaktionen unter Bezug auf Messungen während der Nacht, Analyse von Ausgangswert-Abhängigkeiten, Stabilität der Monitoring-Befunde bzw. Veränderungen im Verlauf des Rehabilitationsverfahrens (Frank, Herrmann & Fahrenberg, 1996; Fahrenberg, Foerster & Franck, 1995; Fahrenberg et al., 1995; Foerster, 1995).

Der Versuch zur Replikation des negativen Retrospektionseffektes war ein weiteres Thema. Die Patienten wurden im Post-Monitoring-Interview gebeten, anhand derselben Items wie in ihren aktuellen Selbstberichten retrospektive Einstufungen des vergangenen Tags zu geben. In allen sieben Befindens-Items wurde, aggregiert über die drei Tage, ein signifikanter negativer Retrospektionseffekt festgestellt.

2.3 Dritte Untersuchung (Zwei-Tage-Untersuchung)

In einer zweiten Untersuchung (Fahrenberg et al., 1999; Käppler et al., 2001) sollte versucht werden, die Befunde zum negativen Retrospektionseffekt zu replizieren. Wenn dieser Effekt bestätigt werden könnte, müsste eine generelle Antworttendenz angenommen und auf deren angemessene Berücksichtigung in der Diagnostik gedrängt werden.

Zum negativen Retrospektionseffekt ist genauer zu fragen, ob er in gleicher Weise für den abendlichen wie für den morgendlichen Rückblick auf den vergangenen Tag gilt. Ist er in gleicher Weise auch an einem anschließenden zweiten Tag festzustellen? Gibt es den Retrospektionseffekt nur bei negativ oder auch bei positiv formulierten Komponenten (Items) des Allgemeinbefindens? Bestehen Zusammenhänge zwischen dem Betrag des Retrospektionseffektes und Persönlichkeitsmerkmalen, insbesondere der Dimension Emotionalität? Theoretisch ist zu erwarten, dass emotional labile im Vergleich zu emotional stabilen Personen generell ihr Befinden negativer schildern und außerdem eine größere Differenz zwischen Rückblick und Tagesmittelwert (negativer Retrospektionseffekt) aufweisen.

Hier wurde MONITOR, ein weiter entwickeltes, flexibles Programm für den hand-held PC PSION, eingesetzt (Brügner, 1998). Zu 5 Terminen am Tag wurden von 61 Studierenden verschiedener Fächer (37 F, 24 M) Selbstberichte gewonnen. Die Untersuchung erstreckte sich über zwei Tage und den Morgen des dritten Tages, so dass insgesamt 11 Protokolle vorlagen. Dazu gehörten Settingmerkmale wie Ort und Tätigkeit, Befindlichkeit (10 Adjektivskalen) sowie zwei Aufmerksamkeitstests (Go/No-Go Paradigma und Arbeitsgedächtnis). Zusätzlich wurden am Abend des ersten und des zweiten Tages sowie am Morgen des zweiten und des dritten Tages ein Rückblick als Gesamteinschätzung des protokollierten Tages erhoben. Die Items für die aktuelle und die rückblickende Einstufung des Befindens unterschieden sich bis auf diesen zeitlichen Bezug nicht in der Formulierung (siehe Tabelle 2.2). Die Itemskalen waren siebenstufig angelegt: 1 = gar nicht 7 = völlig (zutreffend).

Das zweite Ziel dieser Methodenstudie war die Evaluation der computer-unterstützten Methodik der Datenerhebung mit dem Programm MONITOR zur Erfassung der Befindlichkeit und des Aufmerksamkeitsverlaufs unter Alltagsbedingungen. In wie weit diskriminieren die ausgewählten Items

und Tests zwischen Personen, Tagen und Tageszeiten? Eine hinreichende Varianz dieser Daten ist eine notwendige Voraussetzung für die Analyse des Retrospektionseffektes.

Die Compliance bei der Selbstprotokollierung sollte durch die zeitliche Steuerung mit MONITOR gewährleistet sein. Nach dem Start des Programms lag der Ablauf (ggf. bis auf die Eingabe freier Texte) fest. Allerdings kann das stufenweise lauter werdende Signal des PC unabsichtlich oder absichtlich überhört werden. Die Akzeptanz der gesamten Methodik wurde in der Regel durch ein Post-Monitoring-Interview und einen Fragebogen erkundet, wobei auch Hinweise auf eine methodenbedingte Reaktivität und mögliche Zusammenhänge mit dem Retrospektionseffekt interessierten.

Diese Untersuchung hatte noch weitere Fragestellungen. Dazu gehörten die intra-individuelle Kovariation von Selbsteinstufungen der Befindlichkeit und mit den Ergebnissen der zwei Aufmerksamkeitstests (Go/No-Go-Paradigma und Arbeitsgedächtnis). Der individuelle Tagesverlauf könnte mit der Gruppierung als „Morgentyp“ bzw. „Abendtyp“ zusammenhängen. Diese Ergebnisse sind an anderer Stelle veröffentlicht (Fahrenberg et al., 1999).

Neben der wichtigen Prüfung des Retrospektionseffektes werden auch die Untersuchungsergebnisse hinsichtlich der Variabilität der Einstufungen, der relativen Varianzanteile und Trends sowie hinsichtlich der Akzeptanz der Methodik und der Compliance referiert. Es sind wichtige Aspekte dieser Methodik.

Variabilität

Die 14 Items und die verschiedenen Testwerte der Aufmerksamkeitsprüfungen unterschieden sich deutlich in den Verteilungen. Die Item-Mittelwerte variierten zwischen $M = 1.85$ (ärgerlich, gereizt) und $M = 5.79$ (Situation im Griff), die Standardabweichungen zwischen $SD = 1.23$ (bedrückt) und $SD = 1.83$ (Programmbearbeitung störend). Die Verteilungen der Itemwerte wiesen in der Regel eine deutliche Schiefe auf: bei den „negativen“ Items wurden die Stufen 1 und 2 (geringe Ausprägung) und bei den „positiven“ Items die Stufen 6 und 7 (hohe Ausprägung) bevorzugt.

Bemerkenswerte Unterschiede gab es auch in den relativen Varianzanteilen. Die Personenvarianz war mit mehr als 20 % der Gesamtvarianz am höchsten bei „Situation im Griff“, „ärgerlich, gereizt“, „bedrückt“, „körperlich wohl“, „Stress“ und antizipierte Testleistung, und war kleiner als 10 % bei „aufgeregt, nervös“ und „geistig, angespannt“. Der Varianzanteil, welcher durch die Tageszeiten, d. h. durch generelle Verläufe bedingt ist, war im allgemeinen gering, z. B. 3 % bei „anstrengend, belastend“ und „Stress“ und noch geringer bei den übrigen Items. Dagegen war der Varianzanteil der Wechselwirkung Personen x Tageszeiten, welcher auf die individuellen Tagesläufe zurückzuführen ist, sehr viel größer: „geistig angespannt“ (74 %), antizipierte Leistung (75 %), „aufgeregt, nervös“ (65 %). Auf der anderen Seite war diese Varianzquelle nicht vorhanden bei „bedrückt“ (0 %) oder nur geringfügig bei „anstrengend, belastend“ (12 %). In diesen Fällen war allerdings die Residualkomponente, welche Fehlervarianz und Wechselwirkungen höherer Ordnung enthält, relativ groß.

Die MANOVA ergab eine Anzahl von Haupteffekten zwischen Tagen und Tageszeiten. Am zweiten Tag stuften die Teilnehmer die Situation als gewohnter ein, fühlten sich weniger aufgeregt oder geistig angespannt und bewerteten ihr körperliches Befinden und ihre Testleistungen höher. Außerdem hatten sie in beiden Tests bessere Ergebnisse. Tagesläufe und Veränderungen zeigten sich in den Items „anstrengend, belastend“, „ärgerlich, gereizt“, „geistig angespannt“, „Stress“, welche alle einen grob als \cap -förmig zu klassifizierenden Verlauf hatten. Die Koeffizienten der Retest-Reliabilität vom ersten zum zweiten Tag lagen bei den Befindlichkeits-Items erwartungsgemäß nur in mittlerer Größenordnung (zwischen .44 und .71) und erreichten bei den beiden Aufmerksamkeitstests hohe Werte (.88 und .90).

Die Ergebnisse zeigen deskriptiv eine deutliche Variabilität in den Selbsteinstufungen der Befindlichkeit. Einige der Items diskriminieren besser zwischen Personen, andere beschreiben besser die generellen und individuellen Zustandsänderungen im Tageslauf. Zwischen den Erhebungstagen und zwischen verschiedenen Tageszeiten ergaben sich Unterschiede, die teils als Übungs- und Gewöhnungseffekte, teils als zirkadiane Effekte interpretierbar sind.

Tabelle 2.2: Ambulante Datenerhebung mittels Pocket-Computer und MONITOR

Assessment-Variablen

Erhebungstage (aufeinanderfolgend):		2
Erhebungszeitpunkte:		5/Tag
durch Piepser getriggert:		um 8:00, 12:00, 15:00, 18:00, 21:00 Uhr
Retrospektive Einstufungen:	1. Abendlicher Rückblick	um 21:00 Uhr
	2. Morgendlicher Rückblick	um 8:00 Uhr

Setting-Variablen

		Kategorien
1	Wo befinden Sie sich gerade?	13
2	Welche Personen sind im Moment anwesend?	6
3	Welche Position/Lage nimmt Ihr Körper ein?	4
4	Was haben Sie gerade vor dem Piepsen getan?	22

Zustands-Variablen

		Ratings
5	Ist die momentane Situation für Sie gewohnt, typisch?	1-7
6	Ist die momentane Situation für Sie anstrengend, belastend?	1-7
7	Glauben Sie die momentane Situation "im Griff" zu haben?	1-7
8	Fühlen Sie sich momentan eher aufgeregt, nervös	1-7
9	Ist Ihre augenbl. Stimmung eher ärgerlich, gereizt	1-7
10	Ist Ihre augenblickliche Stimmung eher bedrückt	1-7
11	Fühlen Sie sich momentan geistig angespannt	1-7
12	Fühlen Sie sich momentan eher erschöpft, müde	1-7
13	Fühlen Sie sich momentan körperlich wohl	1-7

Bitte geben Sie mögliche Beschwerden in Stichworten ein:

Zusammenfassende Einstufungen

14	Gab es seit der letzten Eingabe besondere Ereignisse?	NEIN JA
	Bitte beschreiben Sie das Ereignis in Stichworten:	
	<hr/>	
15	Wie stark erlebten Sie seit der letzten Eingabe Stress?	1-7

Tests

16	Wie gut wird Ihre Leistung in den Tests schätzungsweise sein? Konzentration (GO/NOGO) Arbeitsgedächtnis (AG)	1-7
17	Wie gut war Ihre Leistung in den Tests schätzungsweise?	1-7

Methodenbewertung

18	Wie störend war insgesamt die Programm-Bearbeitung?	1-7
----	---	-----

Rückblick 1 + 3: Retrospektive Einstufungen am selben Abend

5	War der heutige Tag für Sie gewohnt, typisch?	1-7
6	War der heutige Tag für Sie anstrengend, belastend? etc.	1-7
18	Wie störend war heute insgesamt die Programm-Bearbeitung?	1-7

Rückblick 2 + 4: Retrospektive Einstufungen am nächsten Morgen

5	War der gestrige Tag für Sie gewohnt, typisch?	1-7
6	War der gestrige Tag für Sie anstrengend, belastend? etc.	1-7
18	Wie störend war gestern insgesamt die Programm-Bearbeitung?	1-7

Negativer Retrospektionseffekt

Der Retrospektionseffekt war auch in dieser dritten Untersuchung festzustellen. Der t-Test und der Wilcoxon-Test stimmten darin überein, dass in acht der neun Befindens-Items ein signifikanter Mittelwertunterschied bestand. Das Item „Situation gewohnt, typisch“ ist als Sonderfall anzusehen, weil hier ein initialer Effekt auf die Neuheit des ersten Protokollierens zurückzuführen ist. Die Null-Hypothese ist nur hinsichtlich des Items „körperlich wohl“ beizubehalten. In den acht anderen Items manifestierte sich ein negativer Retrospektionseffekt: Im abendlichen Rückblick wurden der Tag als belastender, mit mehr „Stress“ erlebt und das Allgemeinbefinden als aufgeregter, ärgerlicher, bedrückter, geistig angespannter, erschöpfter geschildert als es aus dem Mittelwert der fünf aktuellen Einstufungen zu erwarten war. Diese Items beschreiben in gleichsinnig gepolter Weise (7 = vollkommen zutreffend) negatives Befinden. Beim positiven Item „Situation im Griff“ muss die umgekehrte Bedeutungsrichtung berücksichtigt werden. Auch hier kennzeichnet der relativ niedrige Wert am Abend eine tendenziell negative Einschätzung des Tages (siehe Abbildung 2.1).

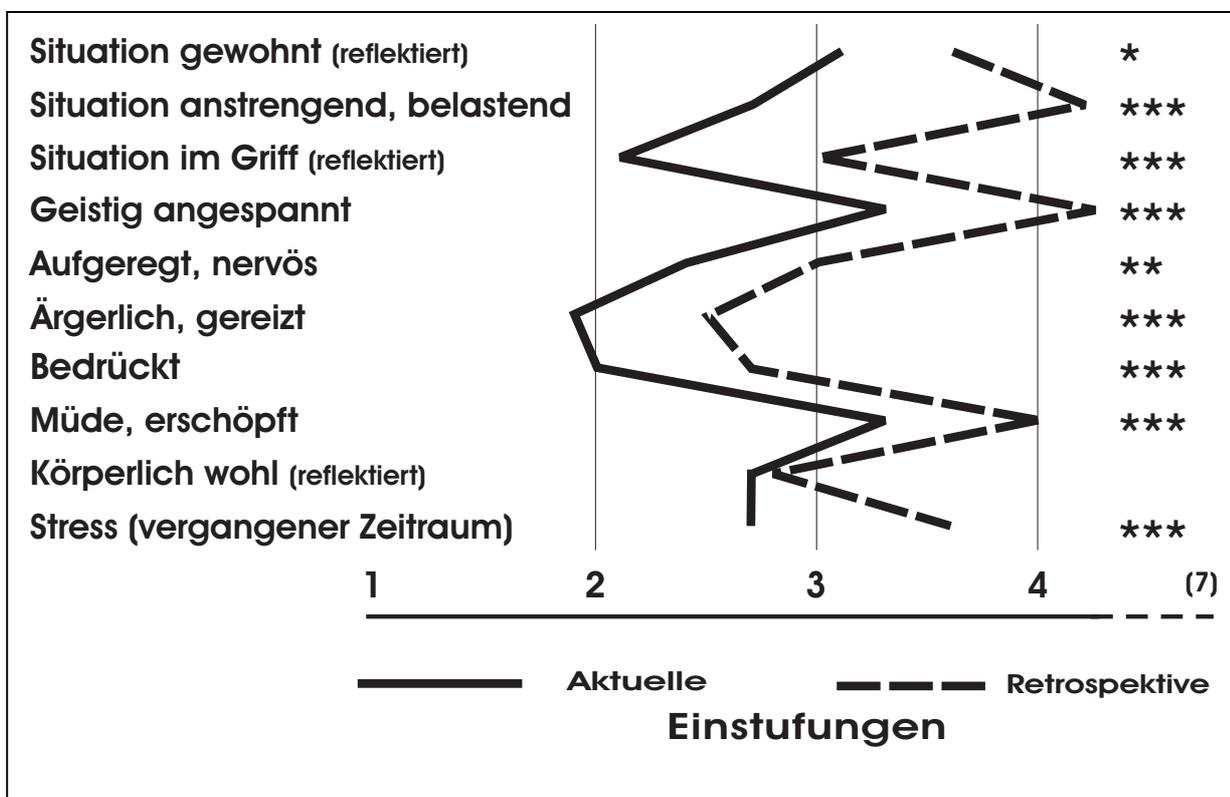


Abbildung 2.1: Negativer Retrospektionseffekt (adaptiert aus Fahrenberg, Brügger, Foerster & Käppler, 1999, S. 652). * $p < .05$, ** $p < .01$; *** $p < .001$

Der negative Retrospektionseffekt ist folglich repliziert worden. Seine Effektstärke wurde anhand der Differenz zwischen dem Mittelwert der aktuellen Einstufungen und dem Rückblick, bezogen auf die Standardabweichung der aktuellen Einstufungen, geschätzt. Die Effektstärke für die neun Items mit signifikantem Retrospektionseffekt liegt zwischen $d = 0.47$ und $d = 1.55$, d. h. es wurden *mittlere bis große Effektstärken* festgestellt. Die entsprechenden Tests für den zweiten Tag sowie für die beiden morgendlichen Rückblicke auf den vergangenen Tag ergaben eine fast deckungsgleiche, hohe Übereinstimmung.

Negativer Retrospektionseffekt und Emotionalität

Der Betrag des Retrospektionseffektes (Differenzwert zum Mittelwert des Tages) und zum Vergleich auch der Mittelwert des Tages wurden mit den Testwerten für *Extraversion* und *Emotionalität* des FPI-R (Fahrenberg, Hampel & Selg, 2001) korreliert. Für den *Index*, der durch Aggregation der beiden Items „bedrückt“ und „erschöpft, müde“ gewonnen wurde, ergab sich eine signifikante Korrelation mit der Dimension Emotionalität ($r = .35$; $p < .01$, $df = 56$).

Wenn in den Index fünf weitere Items, die negatives Befinden und negative Situationsbewertungen erfassen, einbezogen wurden, blieben der Zusammenhang zwischen Retrospektionseffekt und Emotionalität abgeschwächt bestehen ($r = .28$, $p > .05$). Außerdem korrelierte der Tagesmittelwert dieses Aggregats mit der Extraversion. Auch hinsichtlich des Index für positives Befinden korrelierte der Retrospektionseffekt positiv mit Emotionalität ($r = .35$, $p < .01$), d. h. das Ausmaß der rückblickenden Minderung in der positiven Einschätzung des Tages hängt mit diesem Persönlichkeitsmerkmal zusammen. Die Ergebnisse besagen hypothesengerecht, dass emotional labile Personen einen größeren negativen Retrospektionseffekt auf ihren Tageslauf erkennen lassen als emotional stabile Personen. Tendenziell sind diese Zusammenhänge auch am zweiten Untersuchungstag zu sehen, signifikant bzw. sehr signifikant sind sie jedoch nur am ersten Tag. Vielleicht hat es eine Rolle gespielt, dass diese Protokollierungen am ersten Tag ungewohnt waren.

Akzeptanz und Reaktivität

Aufgrund einer ausführlichen Nachbefragung konnte von einer guten Akzeptanz der computer-unterstützten Untersuchungsmethodik ausgegangen werden. Die Probanden gaben an, ihr Verhalten in Reaktion auf die Registrierung „gar nicht“ oder allenfalls „ein bisschen“ (Skalenstufen 1 und 2, zusammen 70% der Antworten) geändert zu haben. Weitere 20% der Personen berichteten moderate Effekte (Skalenstufen 3 – 5), lediglich 10% stärkere Einflüsse. Zwei Drittel der Untersuchungsteilnehmer gaben an, nicht mehr als sonst auf ihren inneren Zustand geachtet zu haben. Die große Mehrzahl (über 80%) der Studienteilnehmer erlebten in ihrer Umwelt keine oder nur geringfügig negative Reaktionen auf die Felduntersuchung. Die Nachbefragung ergab insgesamt, dass die durch das ambulante Assessment hervorgerufenen Reaktivitätseffekte von den Teilnehmern selbst als eher gering angesehen wurden. Es gibt jedoch einzelne Probanden, die ihr Verhalten änderten, höhere Selbstaufmerksamkeit und unangenehme Reaktionen erlebten.

Die auf die Untersuchungsmethodik im engeren Sinne bezogene Akzeptanz lässt sich anhand der folgenden sechs Items bewerten. Ein Drittel der Teilnehmer empfand die wiederholten Eingaben als „gar nicht“ oder „geringfügig“ störend, die überwiegende Mehrheit (60%) als „mittelmäßig“ und lediglich vier Personen (7%) als stärker beeinträchtigend (Skalenstufe 6). Fast die Hälfte der untersuchten Personen fanden die Untersuchungszeitpunkte „gar nicht“ zu häufig, weitere 15% „ein bisschen“ und nur wiederum 7% deutlich zu häufig. Die Teilnehmer unterschieden sich dabei offensichtlich im Muster ihrer Bewertungen, denn die Interkorrelationen dieser zehn Items, die zwischen $r = .01$ und $r = .47$ liegen, wiesen in der Mehrzahl keinen signifikanten Zusammenhang auf.

Die Akzeptanz der Untersuchungsmethodik, die in der vorliegenden Studie eine fünfmalige Unterbrechung des Tagesverlaufs für durchschnittlich zehn Minuten über zwei aufeinanderfolgende Tage erforderte, wird zusätzlich durch die hohe Wiederteilnahmebereitschaft der untersuchten Personen belegt: 87% antworteten auf diese Frage mit „ja“, weitere 9% mit „vielleicht“.

2.4 Vierte Untersuchung (Berufstätige)

In der vierten Untersuchung wurde geprüft, in wie weit der negative Retrospektionseffekt über die Population der Studierenden hinaus auch bei berufstätigen Personen festzustellen ist. Durch Veränderung methodischer Details sollten außerdem mehr Hinweise auf mögliche Erklärungshypothesen gewonnen werden. Da mit dieser computer-unterstützten Methodik bisher überwiegend Studierende untersucht worden sind, war es wichtig, die Akzeptanz für den PC und das Programm MONITOR bei einer anderen Personengruppe zu untersuchen. Wie reagieren Berufstätige im mittleren Lebensalter auf diese Methodik und welche Hinweise auf eine methodenbedingte Reaktivität ergeben sich dabei? Zu dieser Fragestellung wurden in den Rückblick einige Items eingefügt, um zu erkunden, ob die Tagesplanung geändert wurde, ob eine größere Selbstaufmerksamkeit berichtet wurde, wie störend die

Beantwortung der Fragen war und ob die Bereitschaft besteht, noch einmal an einer ähnlichen Untersuchung teilzunehmen.

Die neue Untersuchung soll mehr Items, vor allem mehr Items mit positiver Formulierung enthalten, um die genannten differentiellen Veränderungen erfassen zu können. Für die psychologische Interpretation des Retrospektionseffektes war die Korrelation zwischen dem Betrag des negativen Retrospektionseffektes und der Persönlichkeitsskala Emotionalität wichtig. Lässt sich diese Beziehung bestätigen? Explorativ sollen noch zwei andere Wege beschritten werden:

Meta-Fragen

Nachdem sie den Rückblick auf den vergangenen Tag gegeben haben, könnten die Personen gefragt werden, wie sie zu ihrer Gesamtbeurteilung gekommen sind. Es ist zwar fraglich, inwieweit ein direktes Wissen über diesen Prozess der Selbsteinstufungen vorhanden ist, doch können vielleicht aufschlussreiche Hinweise gewonnen werden. Deshalb soll in mehreren Fragen erkundet werden, ob sich der Rückblick an besonderen Ereignissen orientierte, an der momentanen Stimmung, am Durchschnitt der Antworten im Tagesverlauf. Wie gut stimmen Tagesrückblick und Antworten im Tagesverlauf überein? Ist der Rückblick eher positiver oder negativer? Diese Fragen beziehen sich also nicht auf die Befindlichkeit, sondern als Meta-Fragen auf den Prozess der Introspektion und Retrospektion.

Besondere Ereignisse

Besondere Tagesereignisse könnten nicht nur einen Einfluss auf die aktuellen Selbsteinstufungen, sondern auch auf den Rückblick haben. Das Ereignis hätte dann zu einer neuen Verankerung des individuellen Bezugssystems, in dem die Befindlichkeit beurteilt wird, geführt. Bei einer alltagsnahen Datenerhebung sind solche besonderen Ereignisse — zumindest bei einigen Personen — zu erwarten. Falls deren Anzahl ausreicht, soll neben einer Einzelfall-Betrachtung auch eine statistische Exploration versucht werden. MONITOR bietet zwei Möglichkeiten, solche Bedingungen zu erfassen: durch die regelmäßige Frage nach besonderen Ereignissen (ggf. mit der Bitte um eine erläuternde Texteingabe) und indirekt durch die allgemein gehaltene Frage, ob seit der letzten Eingabe „Stress“ vorhanden war.

Untersuchungsplan

Im Unterschied zu den vorausgegangenen Freiburger Studien mit der computer-unterstützten Selbstprotokollierung (MONITOR) wurden hier nicht Studierende oder Patienten, sondern berufstätige Personen untersucht. Diese breitere Auswahl war sowohl für die Frage nach Zusammenhängen von aktuellem Befinden und Persönlichkeitseigenschaften als auch für die Replikation und die vertiefte Analyse der Urteilstendenzen beim negativen Retrospektionseffekt wichtig. An der Untersuchung nahmen 59 berufstätige Personen teil (34 Frauen, 25 Männer, Alter zwischen 19 und 58, $M = 34$, $SD = 11$ Jahre).

Das allgemeine Format der vorausgegangenen Untersuchung wurde beibehalten, um den Vergleich der Ergebnisse zu erleichtern. Der erste Selbstbericht wurde morgens – mit Rücksicht auf den individuellen Tageslauf der Berufstätigen – zwischen 6 und 10 Uhr erbeten. Die morgendliche Abfrage erfolgte unmittelbar nach dem Aufstehen – niemand sollte durch den Computer geweckt werden. Weitere Befragungszeitpunkte waren um 12, 15, 18 und 21 Uhr (mit einem Intervall von ± 7 Minuten). Die Beantwortung der gesamten Fragesequenz dauerte wenige Minuten.

Die Auswahl der Items orientierte sich im wesentlichen an der vorausgegangenen Studien. Der Itempool wurde um einige Fragen erweitert (siehe Anhang). Die Liste bestand aus 19 Fragen. Die ersten fünf Fragen sollten das behaviorale Setting erfassen: Aufenthaltsort („Wo befinden Sie sich gerade“, 4 Kategorien), sozialer Kontext („Welche Personen sind im Moment anwesend“, 6 Kategorien), momentane Tätigkeit („Was haben Sie gerade vor dem Piepsen getan“, 22 Kategorien) und besondere Ereignisse („Was hat sich seit der letzten Eingabe ereignet“, freie Texteingabe). Anschließend wurden 14 Adjektivskalen vorgegeben (siehe Anhang).

Die 5 „positiven“ Befindensfragen lauteten: „Fühlen Sie sich momentan aktiv, leistungsfähig?“, „Fühlen Sie sich momentan körperlich wohl?“, „Fühlen Sie sich momentan eher ausgeglichen?“, „Fühlen Sie sich momentan selbstsicher, selbstbewusst?“, „Sind Sie momentan eher guter Stimmung?“

An diese Einstufungen schlossen sich um 21 Uhr weitere 22 Fragen an: 13 Fragen zum Rückblick und Fragen zu Antworttendenzen und zur Reaktivität. Die 5 Fragen zu möglichen Antworttendenzen

waren: „Orientierte sich Ihr Rückblick auf den heutigen Tag an besonderen Tagesereignissen?“ (ja/nein), „Orientierte sich Ihr Rückblick auf den heutigen Tag an der momentanen Stimmung/ Befinden?“ (ja/nein), „Orientierte sich Ihr Rückblick auf den heutigen Tag am Durchschnitt der Antworten im Tagesverlauf?“ (ja/nein), „Wie gut glauben Sie, stimmt Ihr Tagesrückblick mit den Antworten im Laufe des Tages überein?“ (eher gut/ eher schlecht) und „Neigen Sie dazu, die Tage im Rückblick anders einzuschätzen?“ (ja/nein). Zum Abschluss der Untersuchung folgten 4 Fragen zur Reaktivität: „Haben Sie Ihre Tagesplanung wegen der PC-Fragen geändert?“ (ja/nein), „Haben Sie heute auf Ihr Verhalten mehr als sonst geachtet?“ (ja/nein), „Wie störend war die Beantwortung der Fragen insgesamt?“ (1-7) und „Würden Sie an einer ähnlichen Untersuchung noch einmal teilnehmen?“ (ja/nein).

Variabilität

Alle 14 verwendeten Befindensitems erwiesen sich hinsichtlich Range und Streuung für die weitere Auswertung als geeignet. Sie wiesen jedoch wie zuvor Boden- und Deckeneffekte auf anhand derer sich die Items in zwei Gruppen aufteilen lassen. Die Items „Situation im Griff“, „Situation gewohnt, typisch“, „selbstsicher, selbstbewusst“, „körperlich wohl“, „gute Stimmung“, „ausgeglichen“ und „aktiv, leistungsfähig“ unterschreiten im Mittel über alle Eingaben, mit Ausnahme des Items „ausgeglichen“ (Mittelwert = 4.97), nicht den Wert 5. Die Frage „Situation im Griff“ wird im Mittel sogar mit einem Wert von über 6 beantwortet. Für die Items „Stress“, „geistig angespannt“, „Situation anstrengend, belastend“, „abgelenkt, gestört“, „aufgeregt, nervös“, „bedrückt“ und „ärgerlich, gereizt“ ergaben sich Tagesmittelwerte von weniger als 3. Für die Items „ärgerlich, gereizt“ und „bedrückt“ lagen die Werte mit 1.68 bzw. 1.81 sogar deutlich darunter. Diese Verteilungen legen die Gruppierung in positive und negative Items nahe, da alle Items mit Deckeneffekt positive Emotionen betreffen, während sich die Items mit Bodeneffekt auf negative Emotionen beziehen. Das Item „erschöpft, müde“ weist mit einem Tagesmittelwert von 3.37 weder einen Decken- noch einen Bodeneffekt auf und ist daher keiner der beiden Gruppen zuzuordnen (Tabelle im Anhang).

Die Items zeigen über die fünf Tageszeitpunkte einen unterschiedlichen Verlauf. Alle negativen Items haben ihr Maximum in der Zeit zwischen 12 und 15 Uhr mittags und fallen gegen Abend ab. Mit Ausnahme der Items „störend“ und „Stress“ lagen die abendlichen Werte unterhalb der Werte am Morgen. Der Verlauf der positiven Items war weniger einheitlich. Die Items „selbstsicher, selbstbewusst“ und „ausgeglichen“ blieben über den gesamten Tag hinweg relativ konstant. Das Item „aktiv, leistungsfähig“ zeigte einen ähnlichen Verlauf wie die negativen Items, d.h. es gab ein Maximum zwischen 12 und 15 Uhr und die Werte um 18 und 21 Uhr lagen unter denen am morgen. Die Items „Situation im Griff“ und „Stimmung“ stiegen zur Mittagszeit hin an, hatten ihr Minimum um 15 Uhr und stiegen gegen Abend hin an, wo sie ihr Maximum erreichten. Das Item „gewohnt, typisch“ zeigte einen untypischen Verlauf. Die Werte hatten am morgen ihr Maximum, fielen um 12 stark ab um dann bis um 18 noch einmal anzusteigen und um 21 wieder leicht zu fallen. Das Item „körperlich wohl“ erreichte um 15 Uhr das Maximum).

Die über 59 Personen gepoolten intra-individuellen Korrelationskoeffizienten zeigten enge Zusammenhänge ($r > 0.50$) zwischen „aktiv, leistungsfähig“ und „müde“ sowie zwischen „ausgeglichen“ und „gute Stimmung“ sowie „selbstsicher“; negative Zusammenhänge zwischen „gereizt“ und „Situation im Griff“ und „gute Stimmung“.

Zusammenhänge mit Persönlichkeitsvariablen

Die Auswertung des von allen Teilnehmern ausgefüllten Freiburger Persönlichkeitsinventars (FPI-R) zeigte im Mittelwert-Profil keine Auffälligkeiten. Sowohl in den zehn FPI-Skalen Lebenszufriedenheit, Soziale Orientierung, Leistungsorientierung, Gehemmtheit, Erregbarkeit, Aggressivität, Beanspruchung, Körperliche Beschwerden, Gesundheitsorgen und Offenheit als auch in den Persönlichkeitsdimensionen Extraversion und Emotionalität lagen die Mittelwerte innerhalb des Normbereiches (Stanine 4 bis 6). Zwischen den aggregierten Tageswerten der Befindlichkeit und dem FPI-R bestanden in mehreren Skalen signifikante Beziehungen (siehe Anhang), die auch als Validierungshinweise zum FPI-R interessant sind (Fahrenberg et al., 2001).

Die explorative Untersuchung von Zusammenhängen zwischen Befindlichkeit und Persönlichkeitseigenschaften ergab zwischen den 14 Items und den 12 Skalen des FPI-R mehr als 50 signifikante Koeffizienten ($p < .05$). Auffällig ist die Häufung dieser Koeffizienten bei den Skalen Lebenszufriedenheit, Beanspruchung, Gesundheitsorgen und Emotionalität. Dagegen scheinen andere Skalen des FPI-R wie Soziale Orientierung, Aggressivität, Offenheit und Extraversion kaum Entsprechungen in

den hier für MONITOR ausgewählten Items zu haben. Bemerkenswert ist, dass mit der Skala Extraversion nur das Item „selbstsicher, selbstbewusst“ korreliert.

Als Beispiel werden hier die Korrelationen zur Skala Emotionalität (Skala N des FPI-R) zitiert: Personen mit hohen Stanine-Werten auf dieser Skala waren bedrückter ($r = .50$), schlechterer Stimmung ($r = -.44$), weniger ausgeglichen ($r = -.46$) und erlebten mehr Stress ($r = .46$). Die Personen beschrieben sich als müder ($r = .40$), nervöser ($r = .39$), gereizter ($r = .38$), weniger aktiv ($r = -.37$), belasteter ($r = .35$) und angespannter ($r = .34$). Außerdem gaben sie an, weniger selbstsicher zu sein ($r = -.30$) und die Situation weniger im Griff zu haben ($r = -.29$). Kein Zusammenhang bestand zu dem Ausmaß der durch die Untersuchung empfundenen Störung und zur Beurteilung der Situation als gewohnt bzw. typisch.

Retrospektionseffekt

Der Vergleich zwischen den aggregierten Tagesmittelwerten und der retrospektiven Einschätzung ergab bei 8 von 14 Items rückblickend einen signifikant ($p < .01$) schlechteren Wert, als es die aggregierten Tagesmittelwerte eigentlich erwarten ließen. Die Abweichungen haben eine mittlere bis große Effektstärke. Sie betragen je nach Item zwischen $d = 1.27$ und $d = 0.45$ (Mittelwert-Differenz bezogen auf die SD der aktuellen Werte). Der *negative Retrospektionseffekt wurde bestätigt* (siehe Abbildung 2.2):

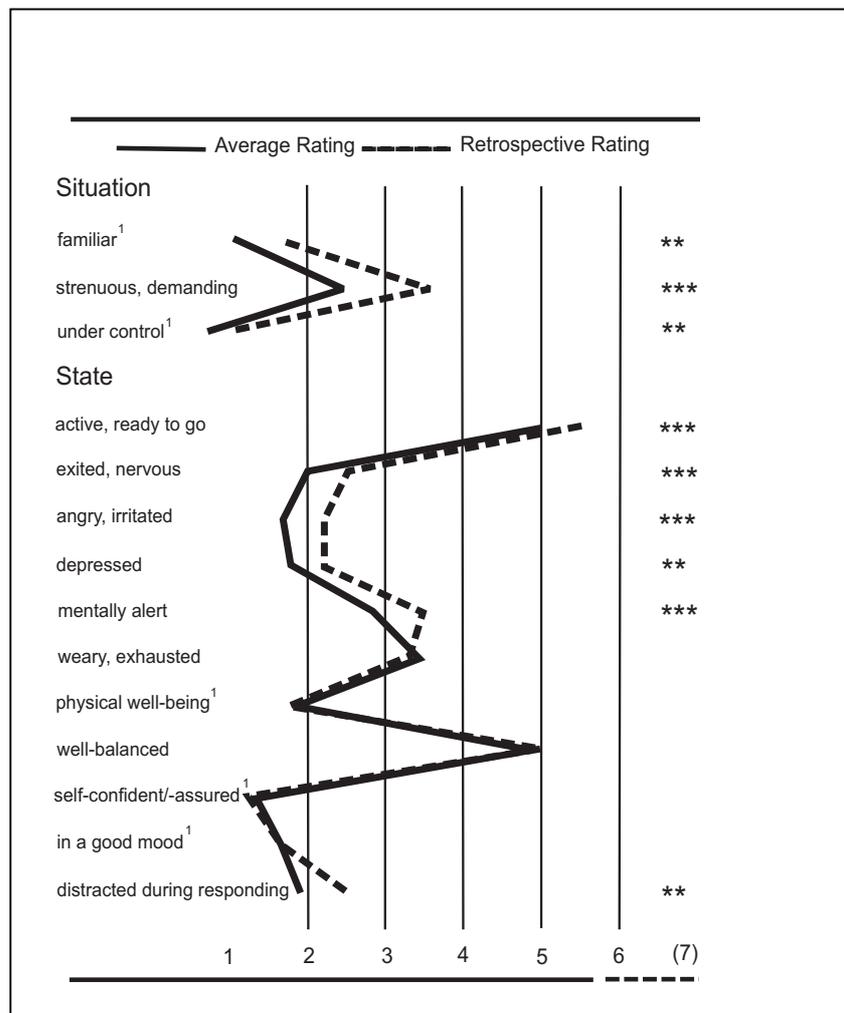


Abbildung 2.2: Negativer Retrospektionseffekt.

¹Vorzeichen der Adjektivskala reflektiert (aus Fahrenberg, Hüttner & Leonhart, 2002, S. 102). * $p < .05$, ** $p < .01$; *** $p < .001$

Die Unterschiede ergaben sich:

- Bei 5 von 6 „negativ“ formulierten Befindens-Items (mit Ausnahme von „erschöpft, müde“) als höherer Skalenwert, d. h. die Situationen wurden rückblickend als beanspruchender und das Befinden als aufgeregt-nervöser, ärgerlich-gereizter, bedrückter sowie geistig angespannter (Ausnahme „erschöpft“).
- Die Störung durch die Abfragen wurde rückblickend stärker betont als aktuell.
- Bei 2 von 7 positiv formulierten Items „Situation gewohnt, typisch“ und „Situation im Griff“) als niedrigerer Skalenwert. Diese beiden Items sind jedoch keine typischen Befindens-Items; sie haben auch die deutlichsten Deckeneffekte. Bei 4 positiv formulierten Items zeigten sich keine Unterschiede. Für das positive Item „aktiv, leistungsfähig“ ergab sich ein *signifikanter Retrospektionseffekt, jedoch in entgegengesetzter Richtung, d. h. im Rückblick wurde diese Einstufung noch deutlicher bejaht*.

Während sich also bei den Items, die sich auf eher negative Aspekte der Befindlichkeit beziehen, mehrheitlich (6 von 7 Items) ein negativer Retrospektionseffekt bestand, fehlte dieser Effekt – wenn von den beiden Situationsbewertungen abgesehen wird – bei den Items positiver Befindlichkeit oder wurde sogar zu einem positivem Effekt bei „aktiv, berufstätig“. Aus diesen differentiellen Veränderungen folgt, dass die Formulierung eines Items als negativer oder positiver Aspekt des Befindens eine Rolle spielt.

Bemerkenswert ist, dass bei drei von den vier Items, deren Mittelwerte am nächsten zur Skalenmitte 4 liegen (3.41, 4.94, 4.97, 5.19), kein signifikanter Retrospektionseffekt beobachtet wurde, allerdings waren darunter drei positive Items.

Retrospektionseffekt und Persönlichkeitseigenschaften

Die Skala „Emotionalität“ des FPI korrelierte signifikant ($p < .05$) mit dem *Betrag* des Retrospektionseffektes (Tages-Mittelwert minus Rückblick) bei den Items „anstrengend, belastend“ ($r = .28$) und „aufgeregt, nervös“ ($r = .27$). Emotional labile Personen stufen ihr Befinden am vergangenen Tag negativer ein. Signifikante Korrelationen gab es auch zwischen der Skala N und „ausgeglichen“ und „guter Stimmung“, doch waren hier die Retrospektionseffekte selbst nicht signifikant: (siehe Tabelle im Anhang). Die Korrelationskoeffizienten dieser Untersuchung unterschieden sich etwas von denen der früheren Untersuchung an Studierenden (dort wurden nur die Skalen E und N verwendet). In jener Untersuchung gab es signifikante Korrelationen zwischen der Skala N und einem aus 2 Items („bedrückt“ und „erschöpft, müde“) bzw. einem aus 7 Items gebildeten Index (jedoch nur am ersten von zwei Tagen) von $r = .35$ und $r = .28$. Jetzt lauten die Korrelationen für den ersten Index (2 Items) $r = .28$ ($p = .03$, $N = 59$) und für den zweiten Index (7 negative Items) $r = .34$ ($p = .008$, $N = 59$). Die zuvor berichteten Befunde wurden also im Prinzip bestätigt.

In der neuen Untersuchung ergaben sich signifikante Zusammenhänge mit weiteren Skalen des FPI-R. Das Ausmaß der körperlichen Beschwerden korrelierte negativ mit dem Retrospektionseffekt der Items „anstrengend, belastend“ ($r = .37$), „geistig angespannt“ ($r = .35$) und „aufgeregt, nervös“ ($r = .27$). Je größer das Ausmaß der körperlichen Beschwerden, desto stärker war bei diesen Items die negative Abweichung der retrospektiven Werte vom aggregierten Tagesmittelwert. Die FPI-Skala „Beanspruchung“ korrelierte negativ mit dem Ausmaß der Verzerrung für das Item „anstrengend, belastend“ ($r = .44$). Die negative Abweichung des Rückblicks war bei diesem Item also um so größer, je stärker sich die Personen subjektiv beansprucht fühlen. Ein vergleichbarer Zusammenhang ergab sich für das Item „aufgeregt, nervös“ ($r = .30$). Aggressive Personen schätzten das Ausmaß ihrer Erschöpfung bzw. Müdigkeit während des Tages rückblickend größer ein, während weniger aggressive eher dazu tendieren, ihre Erschöpfung rückblickend zu unterschätzen ($r = .43$). Ein vergleichbarer Zusammenhang für dieses Item besteht in Verbindung mit der FPI-Skala „Leistungsorientierung“ ($r = .34$). Die FPI-Skala „Offenheit“ stand im Zusammenhang mit der retrospektiven Abweichung für das Item „gewohnt, typisch“ ($r = .35$). Offene Personen neigten eher dazu, den Tag rückblickend als typisch einzuschätzen, während ihn weniger offene Personen Tag als eher untypisch bezeichneten. Ein ähnlicher Zusammenhang bestand zwischen Lebenszufriedenheit und dem Item „ausgeglichen“ ($r = .33$). Das Geschlecht scheint keinen Einfluss auf das Ausmaß der retrospektiven Verzerrung zu haben.

Metafragen zur Urteilsbildung

Im Anhang sind die Antwortverteilungen bei den *fünf Metafragen zur Urteilsbildung im Rückblick* enthalten. Die Frage „Wie gut, glauben Sie, stimmt Ihr Tagesrückblick mit den Antworten im Laufe des Tages überein“ beantworteten fast alle mit „eher gut“, so dass keine differentielle Auswertung stattfinden konnte. Die Möglichkeit einer systematischen Erinnerungstäuschung bzw. Verzerrung scheint diesen Untersuchungsteilnehmern nicht bewusst zu sein. Hier gibt es jedoch einen Widerspruch, denn auf die Frage „Neigen Sie dazu, den Tag im Rückblick anders einzuschätzen?“ antworteten 23 Personen (39 %) mit „ja“. Statistisch waren jedoch die Antworten zu den vier Metafragen (mit hinreichender Varianz) weitgehend unabhängig voneinander. Nur die bejahte Rolle der Tageseinflüsse korrelierte mit der Neigung, die Tage im Rückblick anders einzuschätzen (Cramer's $V = .29$, $p = .026$, $df = 1$).

Die 23 Teilnehmer mit der Neigung, den Tag im Rückblick anders einzuschätzen, unterschieden sich in ihren Einstufungen signifikant von den 36 verneinenden Teilnehmern in fünf von 14 Items: es gab größere Retrospektionseffekte (Tageswert minus Rückblick) in den Items, „ärgerlich, gereizt“, „aufgeregt, nervös“, „bedrückt“ und „geistig angespannt“ und einen geringeren Retrospektionseffekt im Item „ausgeglichen“ (siehe Anhang).

Auf die Frage: „Orientierte sich Ihr Rückblick auf den heutigen Tag am Durchschnitt der Tages“ antworteten 27 Personen (46 %) mit „ja“. Hinsichtlich des berechneten Retrospektionseffektes zeigten sich keine Unterschiede zu den verneinenden Teilnehmern.

Nur bei diesem Item zeigten sich Korrelationen mit den Skalen Beanspruchung, Körperliche Beschwerden und Emotionalität des FPI-R: Wer sich am Durchschnitt des Tages orientierte, beschrieb sich im Persönlichkeitsfragebogen als habituell weniger beansprucht ($r = .31$, $p = .018$, $df = 57$) und weniger emotional labil ($r = .36$, $p = .005$) und mit relativ wenigen Beschwerden ($r = .39$, $p = .002$).

Die zusätzlich erhobenen Informationen ergaben also Hinweise auf bestimmte Urteilstendenzen, liessen aber auch die Schwierigkeiten und die möglichen Verständnisprobleme solcher Selbstbeurteilungen erkennen. Der Zusammenhang zwischen der Neigung, die Tage im Rückblick eventuell neu zu bewerten, und dem errechneten Betrag des Retrospektionseffektes, bedeutet, dass es Personen gibt, die solche Strategien der Gewichtung bewusst verwenden. Diese Strategie erklärt jedoch noch nicht die dominierende negative Richtung der rückblickenden Bewertung. Psychologisch ist weiterhin interessant, dass sich emotional labile Menschen weniger am Durchschnitt des Tages zu orientieren scheinen, sondern u. U. an einem anderen Bezugssystem. Dies scheinen aber nicht die besonderen Ereignisse zu sein, denn die Antworten zu der betreffenden Frage waren unergiebig bzw. inkonsistent.

Besondere Ereignisse

Von 55 der 59 Teilnehmer lagen zwischen 1 und 5 Texteingaben zu der Frage nach besonderen Ereignissen im Tageslauf oder – meist – zu körperlichen Beschwerden. Prägnante Ereignisse mit deutlicher positiver oder negativer Valenz, welche u. U. den Tagesrückblick neu gewichten lassen könnten, waren jedoch nur in wenigen Fällen dokumentiert. Eine statistische Auswertung im Hinblick auf eine mögliche Veränderung des Bezugssystems war deshalb nicht möglich.

Akzeptanz und Reaktivität

Nur 2 der 59 Teilnehmer meinten, Ihre Tagesplanung wegen der Untersuchung geändert zu haben. Mehr auf ihr Verhalten geachtet zu haben, gaben 10 Teilnehmer an. Alle würden an einer ähnlichen Studie noch einmal teilnehmen.

Zusammenfassung

Es bleibt auch nach dieser vierten Untersuchung offen, ob der negative Retrospektionseffekt eher als eine Tendenz zur Mitte oder darüber hinaus auch als eine persönlichkeitsabhängige Disposition zu interpretieren ist. Besser dokumentiert wurde der bereits früher beschriebene Befund, dass bei „positiver“ Formulierung des Iteminhalts ein negativer Retrospektionseffekt kaum zu beobachten ist. Der in der vorausgegangenen Studie bei Studierenden der Universität Freiburg beobachtete Zusammenhang zwischen Retrospektionseffekt und Emotionalität und konnte repliziert und hinsichtlich mehrerer Skalen des Freiburger Persönlichkeitsinventars erweitert werden. Damit wird die „psychologische“ und nicht ausschliesslich psychometrische Interpretation des negativen Retrospektionseffektes gestützt. Emotio-

nal labile, beanspruchte und sich körperlich beeinträchtigt fühlende Personen schildern rückblickend ihren Tageslauf negativer als es aufgrund der aktuellen Protokolle zu erwarten war. Nicht auszuschließen ist eine Erklärungshypothesen, die beide Interpretationen verbindet: Emotional Labile könnten in stärkerem Maße zu der Urteilsheuristik tendieren, bei Unsicherheit die Skalenmitte zu bevorzugen. Die Metafragen zur Erkundung von Urteilstendenzen lieferten wenige, aber überwiegend konsistente Hinweise auf Urteilsstrategien und einen Hinweis auf mögliche Persönlichkeitskorrelate der Tendenz, sich am Tagesdurchschnitt zu orientieren.

Auch bei einer Gruppe von berufstätigen Personen hat sich die computer-unterstützte Methodik – wie die Ergebnisse und die Akzeptanz durch die Untersuchungsteilnehmer bestätigen – erneut bewährt. Diese Methodik hat viele Vorzüge gegenüber den herkömmlichen Fragebogen bzw. Skalen.

2.5 Fünfte Untersuchung (Brasilianische Studierende)

In einer weiteren Untersuchung wurde die Möglichkeit einer Erhebung in einem anderen Land und die damit verbundene Chance eines – vorsichtigen – interkulturellen Vergleichs genutzt (Käppler & Rieder, 2001). Diese Untersuchung war von der Frage geleitet, ob diese negative Rückblicke eventuell für deutsche Verhältnisse typisch sind und eventuell unter anderen kulturellen Verhältnissen nicht zu finden sind oder anders geprägt sind.

Für diese Untersuchung konnten 40 (21 F, 19 M, 18 – 38 Jahre, $M = 24$) Studierende verschiedener Fakultäten der Universität Belo Horizonte in Brasilien gewonnen werden. MONITOR wurde mit den sorgfältig in die portugiesische Sprache übersetzten Items adaptiert und dann an zwei Wochentagen, jeweils mit 5 Terminen am Tag verwendet. Außerdem wurden in der zweiten Untersuchung (siehe oben) verwendeten beiden Aufmerksamkeitstests gegeben.

In 7 von 8 Befindens-Items wurden negative Retrospektionseffekte beobachtet (ausgenommen das einzige positiv formulierte Items „körperlich wohl“), so dass dieses Phänomen auch in einem anderen kulturellen Kontext festgestellt werden konnte.

Der deskriptive Vergleich der Daten mit den Daten der Freiburger zweiten Untersuchung ergab u. a. Hinweise auf Unterschiede der Befindlichkeitsprofile der deutschen und der brasilianischen Gruppe sowie ein größeres Interesse an der Untersuchung und eine grössere Akzeptanz der Methodik. Bemerkenswert ist die relativ größere Prozentsatz fehlender Daten bei den brasilianischen Studierenden. Hierfür können jedoch auch triviale Gründe verantwortlich sein. Der durchschnittlich höhere Lärmpegel in der Umgebung der brasilianischen Studierenden ist wahrscheinlich höher, und dies könnte leichter zu einem Überhören des akustischen Signals geführt haben.

2.6 Sechste Untersuchung (Wochenstudie)

Mit MONITOR wurden Selbstberichte von Studierenden der Psychologie im Laufe einer Woche erhoben. An jedem Tag wurden sechs Eingaben verlangt. Die Zeiten wurden an den typischen Tageslauf dieser Gruppe angepasst: um 9 Uhr, 11 Uhr, 13 Uhr, 15 Uhr, 18 Uhr und 21 Uhr, jeweils mit einer Zufallskomponente von ± 12 Minuten. Die Intervalle betragen hier drei Stunden, doch wurde ein zusätzlicher Mittagstermin um 13 Uhr eingeschoben. An der Untersuchung im Wintersemester 2000/2001 und im anschließenden Frühjahr nahmen 33 Frauen (19-35 Jahre, $M = 23$, $SD = 5$) teil; sie erhielten dafür einen Beleg über Praktikumsstunden. Der Untersuchungsbeginn wurde über die Woche verteilt: Der erste Tag war 9 mal ein Montag, 13 mal ein Mittwoch und 11 mal ein Freitag. Die Tabellen im Anhang vermitteln einige deskriptive Befunde dieser Wochenstudie.

Wochenverlauf

Diese Untersuchung sollte den Verlauf der Befindlichkeit (und der Reaktionszeit) über den Zeitraum einer Woche beschreiben und mehr Informationen über die Variabilität und die Varianzquellen liefern als die bisherigen Studien. Diese Informationen geben einen Einblick in die intra- und interindividuelle Diskriminationsleistung, insbesondere die Änderungssensitivität der Items (siehe Fahrenberg et al., 2001). Weiterhin wurden die Autokorrelationen der Zeitreihen analysiert. Im Hinblick auf den Retrospektionseffekt bot sich die Chance, den Effekt an aufeinanderfolgenden Tagen und im Wochenrückblick zu untersuchen.

Da die Untersuchungsteilnehmerinnen eine relativ homogene Gruppe bilden, könnte dies zu einer Restriktion der Varianzen führen. Andererseits hat diese Homogenität auch mehrere methodische Vorteile hinsichtlich der Vergleichbarkeit und der praktischen Durchführung. Alle Teilnehmerinnen waren, wie in den vorausgegangenen Untersuchungen darüber informiert, dass durch die anonyme Auswertung der Daten Datenschutz hergestellt wurde und dass ein Abbrechen der Untersuchung grundsätzlich möglich wäre.

Beim Vergleich der angegebenen Aufenthaltsorte, der sozialen Kontexte und Tätigkeiten ist zu bedenken, dass in diese Statistiken sowohl die fünf Werktage als auch das Wochenende eingingen. Die Studierenden waren in dieser Woche vorwiegend zu Hause, im Hörsaal oder unterwegs, sie waren oft allein oder mit einer vertrauten Person zusammen. Als Tätigkeiten gaben sie vorwiegend Schreibtischarbeiten, Besuch von Lehrveranstaltungen, Essen und Trinken sowie Gespräche an. Das durchschnittliche Befinden wurde überwiegend als positiv angegeben, d. h. in guter Stimmung, sich körperlich wohl fühlend. In 12.7 % der Protokolle wurden besondere Ereignisse verschiedener Art erwähnt, u. a. hinsichtlich Freunden/Familie ($n = 72$), Studium ($n = 30$), Partnerschaft ($n = 15$) und Gesundheit ($n = 11$). Außerdem wurde in 25.1 % der Protokolle über körperliche und andere Beschwerden wie Erkältung ($n = 95$), Rückenschmerzen und Verspannungen ($n = 74$), Kopfschmerzen ($n = 56$), Bauchschmerzen und Übelkeit ($n = 38$), Müdigkeit und Erschöpfung ($n = 40$) geklagt.

Wahrscheinlich ist bei Studierenden der Unterschied zwischen Werktagen und Wochenenden geringer ausgeprägt als bei Berufstätigen mit sehr festgelegter Arbeitszeit. Die Selbstberichte über die fünf Wochentage unterschieden sich von den zwei Tagen des Wochenendes in vier Items. Am Wochenende waren die Teilnehmerinnen weniger „aufgeregt, nervös“ und weniger „erschöpft, müde“, sie gaben weniger „Stress seit der letzten Eingabe“ an und fanden das Signal des PC weniger „störend, ablenkend“ ($p < .001$). Diese Befunde passen zusammen, denn sie beschreiben eine Phase relativer Erholung. In den Reaktionszeiten zeigten sich keine Unterschiede.

Ein zuverlässiges Zeitbudget (vgl. Bundesministerium für Familie, Senioren, Frauen und Jugend, 1996) ist jedoch aus diesen Statistiken noch nicht zu entnehmen. Dafür müssten statt des gewählten Zeitrasters zufällige Abfragen und kürzere Intervalle vorgesehen und mehr Untersuchungsteilnehmer berücksichtigt werden.

Variabilität

In dieser Gruppe von Studierenden sind die von Tagen und Tageszeiten (= Abfragen, Termine) abhängigen Varianzquellen nicht groß. Unterschiede ergaben sich zwischen den Wochentagen, insbesondere ein verbessertes Befinden am Wochenende. Die *Tageszeiten* bedingen nur wenige Prozent der Gesamtvarianz. Die Personenvarianz macht hier ca. 20 bis 30 % und die Wechselwirkung *Personen x Tageszeiten* ca. 5 bis 10 % aus. Eine entsprechende ANOVA zum Vergleich der Varianzquelle *Tage* statt der *Tageszeiten* lieferte ein sehr ähnliches Bild. Die relativen Anteile der Tage an der Gesamtvarianz waren eher noch geringer, doch hängt dies auch davon ab, ob die Daten nach den *Wochentagen* oder nach den *laufenden Tagen* der Untersuchungswoche sortiert wird.

In einigen Aspekten des Befindens wie „bedrückt“ und „körperlich wohl“ sowie in den Reaktionszeiten gab es relativ mehr habituelle Unterschiede als z. B. hinsichtlich „aufgeregt, nervös“, „Stress seit letzter Eingabe“ und „Signal störend, ablenkend“. In diesen Items zeigte sich dagegen mehr Wechselwirkungsvarianz, d. h. von der Person und der Tageszeit abhängige Unterschiede. Die Varianzquelle *Tageszeiten* repräsentiert die allen Teilnehmerinnen gemeinsamen Variationen des Befindens im Tageslauf.

Den größten Anteil an der Gesamtvarianz hatte das *Residuum*. Es wäre jedoch vorschnell, in diesem Rest nur Fehlervarianz zu sehen, denn hier ist auch die Wechselwirkung höherer Ordnung enthalten: *Personen x Tage x Tageszeiten*. Diese Varianzkomponente enthält die individuell von Tagen und Tageszeit abhängigen Veränderungen des Befindens und der Aufmerksamkeit — also viel von dem was neben den habituellen Unterschieden als die Individualität eines Tages- bzw. Wochenablaufs erscheint.

Wenn mit Personen (33), Abfragen (42) und Wochentagen (7) drei Faktoren berücksichtigt werden, können drei Wechselwirkungen erster Ordnung unterschieden werden: *Personen x Abfrage*, *Personen x Wochentag* und *Abfrage x Wochentag*. Unter dieser Perspektive ergeben sich nur geringe Varianzanteile für Abfragen, *Wochentage*, *Personen x Abfrage* und *Abfrage x Wochentage*. Dagegen ist die Varianzquelle *Abfrage x Wochentage* fast in der Größenordnung der Zwischen-Personen-Varianz: die

Selbstberichte im Laufe der Woche variieren stark von Person zu Person. Die Residuen sind etwas kleiner.

Die Ergebnisse zeigen die Variabilität in den Selbsteinstufungen der Befindlichkeit und Unterschiede in den relativen Varianzanteilen. Viele der hier verwendeten Items diskriminieren relativ gut zwischen Personen. Hinsichtlich der Änderungssensitivität bestehen mehr oder minder deutliche Unterschiede zwischen den Items bzw. Testwerten. In der Wochenstudie haben die meisten Items einen Varianzanteil von ca. 15 bis 22 %, der nicht auf Unterschiede zwischen Personen oder das Residuum zurückzuführen ist. Bereits eine einfache Varianzanalyse lässt hier die relativ geringen allgemeinen (zirkadianen) Unterschiede zwischen den Tageszeiten und zwischen den Tagen erkennen. Mehr Varianz tragen die individuellen Tagesläufe bei. Diese Schlussfolgerung stützt sich auf die einfache Wechselwirkung Personen x Tage/Tageszeiten. Ein ähnlicher, vielleicht größerer Beitrag der höheren Wechselwirkungen ist zu vermuten, aber methodisch kaum zu erfassen.

Solche Analysen geben einen Einblick in die Eigenschafts- und Zustandsaspekte in Selbstberichten und können für die Beurteilung der Änderungssensitivität in ihren tageszeitlichen und täglichen Aspekten nützlich sein. Diese Untersuchungsergebnisse machen also verschiedene Komponenten der within-subject Varianz und Aspekte der Reliabilität deutlich. Aufgrund dieser Daten ist eine Itemselektion für vergleichbare Untersuchungsgruppen möglich. Andere Itemlisten und Gruppen erfordern entsprechende Analysen.

Hier sind Unterschiede zwischen dieser Wochenstudie und der zweiten Untersuchung zu erkennen, denn dort war der Varianzanteil Personen x Tageszeiten erheblich größer. Dies kann an den anderen Verhältnissen einer 2-Tage- und einer 7-Tage-Untersuchung oder an dem Unterschied zwischen Berufstätigen und Studierenden liegen. Diese Unterschiede machen darauf aufmerksam, wie viel Vorbehalte beim Vergleich verschiedener Designs notwendig sind. Die Ergebnisse über die Varianzverhältnisse können eine grobe Orientierung geben; besonders nützlich sind sie beim Vergleich innerhalb einer Untersuchung. Auf diese Weise können für jedes Item bzw. für jeden Testwert die Aspekte der Änderungssensitivität und der Anteil der habituellen Differenzen getrennt werden.

Psychologisch interessante Befunde zeigten sich in der Korrelationsmatrix über die 42 Selbstberichte einer Woche. So ist die Aufmerksamkeitsleistung nur mit den Selbsteinstufungen „aktiv“ und „erschöpft“ assoziiert. Die Aussage, durch den PSION gestört oder abgelenkt zu sein, korreliert nur geringfügig mit anderen Items, am höchsten noch mit „aktiv“ (0.14). Relativ hohe positive bzw. negative Interkorrelationen ergaben sich für „aktiv“ und „erschöpft“ sowie „bedrückt“ und „gute Stimmung“, doch variieren die meisten Aspekte des Befindens in den hier verwendeten Items relativ unabhängig voneinander.

An den Zeitreihen der 33 Teilnehmerinnen wurden nach missing data-Ersetzung (Zellenschätzung im zweifaktoriellen ANOVA-Design) bei 20 Personen Autokorrelationen mit lag 1 bis lag 12 berechnet und über z' gemittelt. Bei lag 1 betragen die Koeffizienten zwischen $-.05$ und $.31$; bei lag 2 zwischen $-.12$ und $.21$; bei lag 3 zwischen $-.09$ und $.08$ und bei lag 6, d. h. nach 24 Stunden, zwischen $-.02$ und $.13$ (ausgenommen $.22$ für die mittlere Reaktionszeit). Die vergleichsweise höchsten Koeffizienten der Autokorrelation ergaben sich für die Items „körperlich wohl“, „gute Stimmung“ und „bedrückt“, d. h. den eher stationären Stimmungen, welche eine geringere Dynamik haben als die anderen Zustandsmerkmale. Bereits bei lag 2, d. h. nach wenigen Stunden, geht die mittlere Autokorrelation gegen 0. Dementsprechend sind auch die Korrelationen zwischen den Wochentagen (berechnet über Abfragen innerhalb des Tages und gepoolt über Personen) nur gering. Diese Ergebnisse sprechen dafür, dass die schwache Autokorrelation in Zeitreihen dieser Art und mit dem hier gewählten Zeitraster vernachlässigt werden kann (siehe Fahrenberg et al., 2002).

Negativer Retrospektionseffekt

Außer dem abendlichen Rückblick, der an jedem der sieben Tage erfolgte, wurden am letzten Tag zusätzlich 13 Fragen zum Rückblick auf die gesamte Woche gestellt. Schließlich erhielten die Teilnehmerinnen bei der Rückgabe des hand-held PC einen Fragebogen (siehe Anhang) mit einer Übersicht über die vergangene Woche. Sie wurden gebeten, jeden der sieben Tag auf einer Skala von 1 bis 7 einzustufen, „Wie anstrengend, belastend war dieser Tag?“ und „Waren Sie an diesem Tag in eher guter Stimmung?“ Die Teilnehmerinnen beantworteten alle das Freiburger Persönlichkeitsinventar FPI-R (Fahrenberg et al., 2001) und die Freiburger Beschwerdenliste FBL-R (Fahrenberg, 1994b):

Zur Analyse des Retrospektionseffektes konnten mehrere Vergleiche gerechnet werden (siehe auch Leonhart, 2001), unter anderem:

1. Ein Vergleich zwischen jedem Tages-Mittelwert des Items und dem zugehörigen Tagesrückblick;
2. ein Vergleich zwischen dem Mittelwert der sieben Tages-Mittelwerte und dem mittleren Rückblick;
3. ein Vergleich zwischen dem Mittelwert der sieben Tages-Mittelwerte und dem Wochenrückblick;
4. ein Vergleich zwischen den mit MONITOR erhobenen Einstufungen für die Tage bzw. die gesamte Woche und den Einstufungen in dem Fragebogen.

In dieser Wochenstudie konnte geprüft werden, in wie weit die Einstufungen und auch die Rückblicke bzw. die negativen Retrospektionseffekte konsistent und zeitlich stabil sind. Weiterhin kann der über sieben Tage aggregierte Rückblick und der Rückblick über ein Intervall bis zu einer Woche untersucht werden. Diese differenziertere Analyse der zeitlichen Bedingungen sollte die früheren Untersuchungen vertiefen. Außerdem könnte mittels der beiden Fragebogen erneut die Abhängigkeit des negativen Retrospektionseffektes von bestimmten Persönlichkeitseigenschaften geprüft werden.

Der Vergleich mittels t-Test und Wilcoxon-Test (1) zwischen den Tages-Mittelwerten der Items und den zugehörigen Tagesrückblicken, (2) über 7 Tage kombiniert und (3) zum Wochenrückblick ergab systematische Retrospektionseffekte, die in 4 von 6 negativen Items überwiegend signifikant waren („anstrengend, belastend“, „ärgerlich, gereizt“, „bedrückt“ und „Stress“) sowie in 1 von 4 positiven Items („positive Veränderung“). Bei drei weiteren Items zeigten sich einzelne Signifikanzen. Auffällig sind die beiden positiven Items „aktiv, leistungsfähig“ und „körperlich wohl“, bei denen keinerlei Signifikanz auftrat (siehe Tabelle 2.3 und zusammenfassend Tabelle 2.4).

Alle signifikanten Veränderungen gingen in Richtung der Skalenmitte, d. h. es gab Veränderungen in die psychologisch „negative“ Richtung. Die einzige Ausnahme bildete die Einstufung der „positiven Veränderungen“, denn hier stieg der geringe Itemwert im Rückblick (bezüglich der Mittelwerte der 7 Tage von 2.7 auf 3.3, bezüglich des Wochenrückblicks von 2.7 auf 3.6) an, d. h. es wurden mehr positive Veränderungen ausgesagt. Die Einstufungen der „positiven Veränderungen“ haben also eine Sonderrolle.

Die Differenzen der Mittelwerte der aktuellen Einstufungen Tag 1 - 7 zu den Rückblicken Tag 1 – 7 waren in allen 10 Items kleiner als die Differenzen zum Wochenrückblick (Tabelle 2.3). Die Effektstärken lagen hier meist im mittleren Bereich zwischen 0.5 und 1.0 (Mittelwertdifferenzen bezogen auf SD der Tages-Mittelwerte). Die Varianzen der Rückblicke (Tag und ebenso Woche) waren durchweg größer als die Varianzen der Tages-Mittelwerte.

Bei größerer zeitlicher Distanz scheinen die Verzerrungen ins Negative anzusteigen – oder es nahm die Tendenz zur Mitte zu. Für die zweite Interpretation spricht, dass die negativen Retrospektionseffekte desto höher ausfallen, je extremer der Mittelwert der aktuellen Tageseinstufungen am Skalende lagen.

Vergleichender Wochen-Rückblick (Post-Monitoring Rückblick)

Von 30 Teilnehmerinnen konnte nach Ablauf der Wochenstudie ein *vergleichender Wochen-Rückblick* mit einem Fragebogen (siehe Anhang) erhoben werden. Für diese Einstufungen wurden zwei Items ausgewählt: „Wie anstrengend, belastend war dieser Tag?“ und „Waren Sie an diesem Tag in eher guter Stimmung?“ Auch hier wurden siebenstufige Skalen vorgegeben von „gar nicht“ bis „völlig“.

Die Abbildungen 2.3 a bis 2.3 d lassen den Wochenverlauf im Rückblick für die sieben Tage des Monitoring erkennen. Für die erste Abbildung sind die Tage nach dem laufenden Untersuchungstag geordnet. Die folgenden Abbildungen zeigen die differentiellen Verläufe für die drei Kohorten mit dem Untersuchungsbeginn am Montag, am Mittwoch bzw. am Freitag. Der Wochenendeffekt ist gut zu erkennen.

Tabelle 2.3: Mittelwerte der Tageswerte und Tages-Retrospektioneffekte für die einzelnen Tage, außerdem über die 7 Tage gemittelte Tageswerte verglichen einmal mit dem mittleren Tages-Rückblick und einmal mit dem Wochenrückblick (N = 33).

Item	Aktiv Leistungsfähig	Anstrengend Belastend	Aufgeregt Nervös	Ärgerlich Gereizt	Bedrückt	Erschöpft Müde	Guter Stimmung	Körperlich wohl	Stress	Positive Veränderung
Datenbasis										
Mittelwert, Tag 1	3.79	2.47	2.46	1.91	2.13	3.84	4.60	4.56	2.87	2.81
Retro-Effekt, Tag 1	-0.30	-1.77	-0.09	-0.48	-0.41	-0.40	-0.10	0.22	-0.56	-0.59
t-Test und Wilcoxon		***		*	**	*			**	*
Mittelwert, Tag 2	3.92	2.53	2.35	2.05	2.25	3.77	4.71	4.81	2.63	2.89
Retro-Effekt, Tag 2	-0.35	-1.26	-0.14	-0.47	-0.17	-0.08	-0.14	0.15	-0.31	-0.63
t-Test und Wilcoxon		***		*						**
Mittelwert, Tag 3	4.28	2.46	2.32	1.90	2.32	3.46	4.62	4.90	2.60	2.86
Retro-Effekt, Tag 3	-0.26	-1.08	-0.24	-0.26	-0.36	0.05	-0.18	0.12	-0.68	-0.39
t-Test und Wilcoxon		***			*				***	
Mittelwert, Tag 4	3.88	2.45	2.09	1.89	2.41	3.70	4.46	4.58	2.53	2.73
Retro-Effekt, Tag 4	-0.25	-1.34	-0.54	-0.62	-0.12	-0.07	0.16	0.06	-0.53	-0.58
t-Test und Wilcoxon		***	*	**					***	*
Mittelwert, Tag 5	3.87	2.46	2.20	1.98	2.59	3.83	4.28	4.77	2.52	2.46
Retro-Effekt, Tag 5	-0.10	-1.18	-0.28	-0.08	-0.41	0.02	-0.20	0.25	-0.35	-0.45
t-Test und Wilcoxon		***			*			*	*	
Mittelwert, Tag 6	3.79	2.66	2.04	1.91	2.49	3.83	4.24	4.50	2.71	2.72
Retro-Effekt, Tag 6	-0.03	-0.98	-0.33	-0.24	-0.21	-0.33	0.03	0.25	-0.62	-0.67
t-Test und Wilcoxon		***				*			***	**
Mittelwert, Tag 7	3.64	2.48	2.08	2.00	2.36	3.89	4.12	4.50	2.42	2.75
Retro-Effekt, Tag 7	-0.14	-1.04	-0.45	-0.28	-0.44	-0.07	-0.08	-0.04	-0.72	-
t-Test und Wilcoxon		***	**		**				***	
Mittelwert, Tag 1-7	3.88	2.50	2.21	1.95	2.36	3.76	4.43	4.66	2.61	2.74
Retro-Effekt, Tag 1-7	-0.19	-1.25	-0.30	-0.35	-0.32	-0.14	-0.06	0.16	-0.55	-0.54
t-Test und Wilcoxon		***	**	***	***			*	***	***
Retro (Wochenrück.)	-0.27	-1,59	-0.36	-0.51	-0.65	-0.33	-0.17	0.47	-0.97	-0.93
t-Test und Wilcoxon		***		**	**			**	***	***

Anmerkung: Signifikanzmarkierungen, wenn t-Test und Wilcoxon's Rang-Test * p<.05, ** p<.01, *** p<.001.

Tabelle 2.4: Auszählung signifikanter Differenzen von aktuellen und retrospektiven Einstufungen an 7 Tagen mit jeweils 6 Abfragen (Anzahl der Tests pro Variable = 42).

Variable	p<.01		p<.05		p<.10	
	Anzahl	%	Anzahl	%	Anzahl	%
Aktiv, leistungsfähig	0	0.0	4	9.5	10	23.8
Anstrengend, belastend	33	78.6	41	97.6	42	100.0
aufgeregt, nervös	2	4.8	7	16.7	13	31.0
Ärgerlich, gereizt	4	9.5	10	23.8	16	38.1
Bedrückt	8	19.0	11	26.2	12	28.6
Erschöpft, müde	4	9.5	8	19.0	12	28.6
Guter Stimmung	0	0.0	2	4.8	7	16.7
Körperlich wohl	1	2.4	3	7.1	8	19.0
Stress	13	31.0	19	45.2	24	57.1

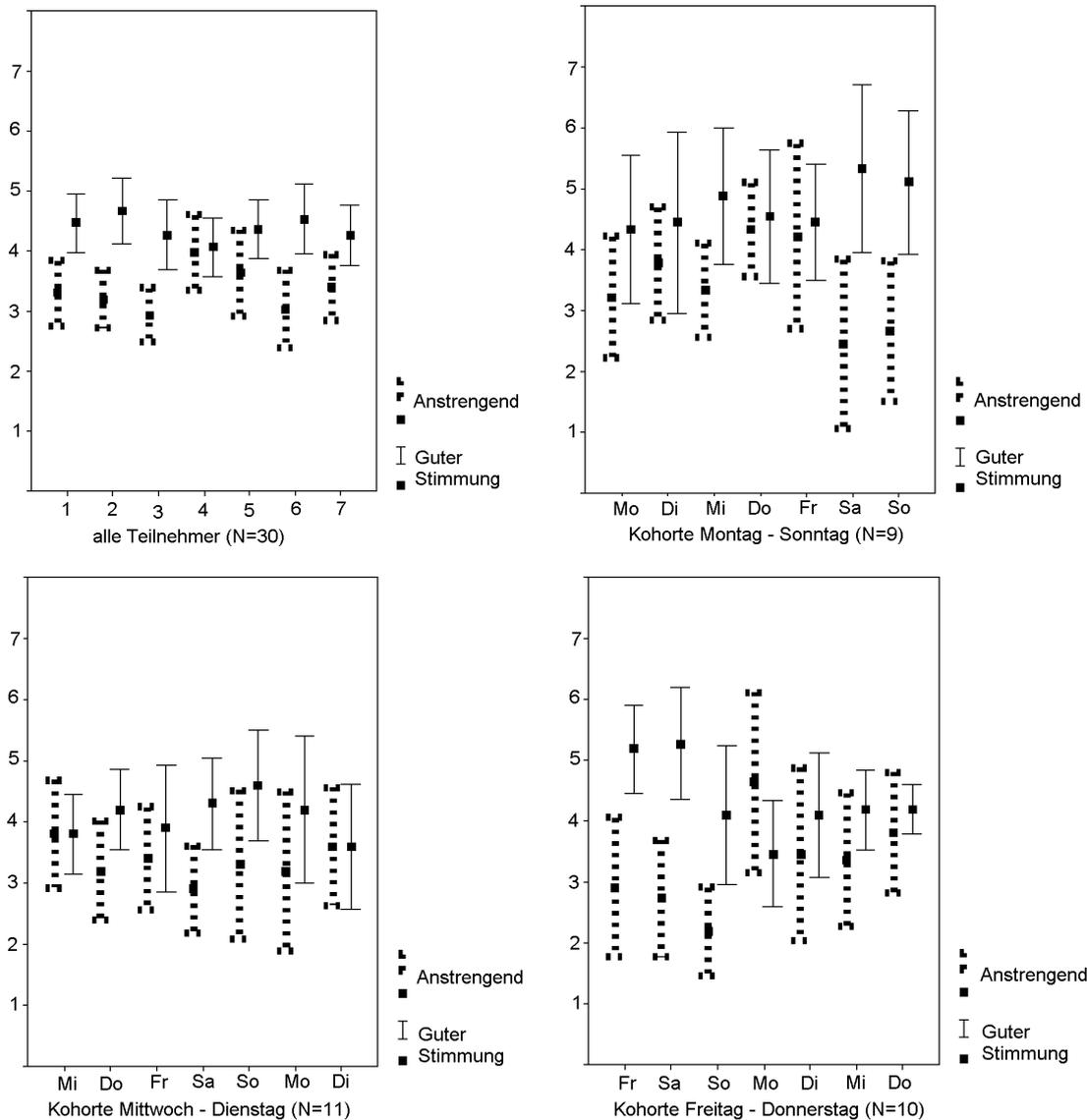


Abbildung 2.3 a bis 2.3 d: Boxplots mit 95 % Konfidenzintervall der Tageswerte beim vergleichenden Wochen-Rückblick: a. Gesamt, b. Kohorte Montag-Sonntag, c. Kohorte Mittwoch-Dienstag und d. Kohorte Freitag-Donnerstag.

Die Tabelle 2.3 zeigt die Tagesmittelwerte, die Tages-Rückblicke und den vergleichenden Wochen-Rückblick für die laufenden Untersuchungstage. Bei der Interpretation ist zu berücksichtigen, dass die differentiellen Verläufe der drei Kohorten gemittelt sind, wobei der Kohorteneffekt aufgrund der ähnlichen Anzahl der Personen (9, 11 und 10) näherungsweise ausgeglichen ist.

Tabelle 2.5: Tages-Mittelwerte, Tages-Rückblicke und vergleichender Wochen-Rückblick (N= 30)

	1. Tag	2. Tag	3. Tag	4. Tag	5. Tag	6. Tag	7. Tag
Wie anstrengend, belastend war dieser Tag?							
Tages-Mittelwert	2.56	2.58	2.39	2.42	2.42	2.61	2.43
Tages-Rückblick	4.37	3.73	3.52	3.70	3.57	3.53	3.34
Vergleichender Wochen-Rückblick	3.30	3.20	2.93	3.97	3.63	3.03	3.40
Waren Sie an diesem Tag in eher guter Stimmung?							
Tages-Mittelwert	4.54	4.74	4.63	4.56	4.30	4.40	4.22
Tages-Rückblick	4.67	4.93	4.76	4.48	4.53	4.37	4.28
Vergleichender Wochen-Rückblick	4.47	4.67	4.27	4.07	4.37	4.53	4.27

Tabelle 2.6: Korrelationskoeffizienten zwischen Tages-Mittelwert und entsprechender Einstufung im vergleichenden Wochen-Rückblick (N = 30)

	1. Tag	2. Tag	3. Tag	4. Tag	5. Tag	6. Tag	7. Tag
anstrengend, belastend	.09	.05	.48	.26	.52	.57	.40
eher guter Stimmung	.19	.54	.25	.09	.59	.53	.56

Anmerkung: zwischen den vergleichenden Einstufungen des Wochen-Rückblicks und dem 1. Tag liegen 6 Tage, zu den folgenden Tagen entsprechend kürzere Intervalle.

Je größer der zeitliche Abstand zwischen einem Tag und den vergleichenden Einstufungen des Wochen-Rückblicks, desto geringer ist die Korrelation. Für den ersten Tag, aber auch für den auffälligen 4. Tag erreichen die Korrelationskoeffizienten keine Signifikanz.

Stabilität des Retrospektionseffektes

Die Daten der Wochenstudie erlaubten es, die Reproduzierbarkeit des Effekts zwischen benachbarten Tagen als einfache Korrelationskoeffizienten sowie als Intraklass-Korrelation (ANOVA) genauer zu bestimmen (siehe Tabelle 2.6).

Zusammenfassend ist festzustellen, dass der Retrospektionseffekt auch in Selbsteinstufungen, die sich über mehrere Tage erstrecken, und in einem Wochenrückblick zu finden sind. Die Effektgrösse scheint mit der Zeit zuzunehmen.

Die Stabilität zwischen den Tagen erreichte nur selten Koeffizienten mittlerer Grösse, die verallgemeinerte Stabilität (Intraklassenkorrelation im zweifaktoriellen ANOVA-Design *Personen x Tage*) reichte von .07 bis .21. Die aus der folgenden neuen Studie (siehe unten) errechnete Kurzzeit-Stabilität (Vormittag vs. Nachmittag, N=29) lag nur für „aktiv“ und „körperlich wohl“ im mittleren Bereich (s. Tabelle 2.5 und Anhang). Dennoch sind diese Koeffizienten hervorzuheben, da sie auf Differenzen beruhen, die gewöhnlich nur eine geringe Stabilität aufweisen.

Tabelle 2.7: Stabilität des Retrospektionseffekts: Tages- und Kurzzeit-Stabilitätskoeffizienten, sowie Intraklassen-Koeffizient (Global-Stabilität) bei der Wochenstudie.

Variable	Wochenstudie							Neue Studie
	Stabilitätskoeffizienten benachbarter Tage						Intraklass-Korrelation (ANOVA)	Stabil.-Koeff. Vormittag * Nachmittag
	1*2	2*3	3*4	4*5	5*6	6*7		
Aktiv	0.26	0.34	0.58	-0.06	0.30	0.18	0.19	0.47
Anstrengend im Griff	-0.04	0.33	0.31	0.58	0.22	0.46	0.20	0.26
Aufgeregt	-0.07	0.16	-0.13	0.42	0.52	0.46	0.21	-0.02
Ärgerlich	0.22	0.18	0.11	-0.18	0.29	0.29	0.15	0.24
Bedrückt	-0.00	0.37	0.13	0.33	0.30	0.35	0.16	-0.15
Erschöpft	0.15	0.16	0.16	0.24	-0.07	0.07	0.13	
Gute Stimm	-0.13	0.13	0.20	0.35	0.19	0.08	0.09	0.19
Körp. wohl	-0.03	0.08	0.40	0.08	0.19	0.49	0.07	0.57
Stress	0.42	0.38	0.45	0.28	0.05	-0.09	0.14	

Retrospektionseffekt und Persönlichkeitsmerkmale

Auch in dieser Untersuchung gab es Belege für die psychologische Erklärungshypothese, doch handelte es sich *nicht um ein konsistentes* Muster. Am ersten Untersuchungstag korrelierten die Beträge der Retrospektion (Differenz Tagesmittelwert – Rückblick) beim Item „bedrückt“ $r = -.40$ ($p = .022$, $N = 33$) mit der Skala N des FPI-R, d. h. emotional labile Personen wiesen wie erwartet größere negative Differenzen auf. Zum Vergleich sind die Koeffizienten für den Tagesmittelwert dieses Items $r = .29$ und für den Rückblick $r = .42$ interessant. Die entsprechenden Korrelationskoeffizienten für das Item „erschöpft“ lauten: Retrospektionsbetrag (Differenz) $r = -.54$, Tagesmittelwert $r = .23$ und Rückblick $r = .50$. Demnach wäre der frühere Befund bestätigt.

Auch die Rückblicke korrelierten an den Tagen 2 bis 7 nur noch geringfügig mit der Skala N, während sich die Koeffizienten zwischen den Tagesmittelwerten und N in der Größenordnung .30 bis .40 bewegten. Es könnte sein, dass die Beziehung zwischen dem negativem Retrospektionseffekt und dem Persönlichkeitsmerkmal Emotionalität mit den initialen Bedingungen zusammenhängt, d. h. spezifisch für den ersten Tag ist.

Insgesamt ergab sich eine Anzahl signifikanter ($p < .05$) Koeffizienten zwischen der Skala N und den *Mittelwerten der Tagesrückblicke*: aktiv, leistungsfähig; anstrengend, belastend; erschöpft, müde; guter Stimmung sowie den *Gesamt-Mittelwerten der Items* (aggregiert über sieben Tage): aufgeregt, nervös; ärgerlich, gereizt; bedrückt; erschöpft, müde; guter Stimmung (negativ); körperlich wohl (negativ).

2.7 Erklärungsversuche (Skalierungsprobleme) und Exkurse zu Methodenproblemen

Der negative Retrospektionseffekt im Vergleich zu den aktuellen Einstufungen des vergangenen Tages wurde in allen referierten Untersuchungen festgestellt. Deshalb liegt die Annahme nahe, dass es sich – zumindest teilweise – um ein psychometrisches Artefakt oder eine triviale Urteilsheuristik handeln könnte. Auf der anderen Seite spricht der – weniger konsistent – beobachtete Zusammenhang u. a. mit der Persönlichkeitsskala Emotionalität für eine psychologische Interpretation. Die negative, aggravierende und klagsame Tendenz passt zu dieser Persönlichkeitseigenschaft.

Die verschiedenen Erklärungsversuche schliessen einander nicht aus.

(1) Die *Tendenz, Vergangenes negativer zu bewerten* als es aktuell erlebt wurde, ist ein bekanntes Phänomen. In Anamnesen wird u. U. von *Aggravation* oder *Klagsamkeit* gesprochen, wenn ein Patient zur negativen Überbetonung und Akzentuierung von Ereignissen oder Beschwerden neigt – vielleicht um Aufmerksamkeit zu gewinnen oder aus anderen Gründen.

Bemerkenswert ist jedoch, dass die meisten Untersuchten ihr Befinden als gut einstufen (bei positiven Adjektiven oft auf der Skalenstufe 5, bei negativen Adjektiven oft auf der Skalenstufe 2. Dies gilt nicht nur für die gesunden jungen Studierenden, sondern auch in den Gruppe der Berufstätigen und der Patienten. Dies hat zur Folge, dass die individuellen Einstufungen nicht um die Skalenmitte streuen, sondern zu den Skalenenden *schiefe* Verteilungen aufweisen.

(2) Ein *negativer Retrospektionseffekt* ergibt sich bei der verwendeten Skalierungsweise dann, wenn der Skalenwert des Rückblicks *größer* als der Mittelwert der aktuellen Einstufungen des Tages ist. Falls die Tages-Mittelwerte, wie es durchweg der Fall ist, am oberen oder unteren Ende der Skala liegen, wird eine *weiter in der Mitte liegende* retrospektive Einstufung zu diesem negativen Retrospektionseffekt führen. Da die Verteilungen der Itemwerte in der Regel nicht symmetrisch um die Skalenmitte 4, sondern links- oder rechtsschief sind, lässt der negative Retrospektionseffekt eine Verschiebung zur Skalenmitte hin erwarten: bei positiv formulierten Items (z. B. „aktiv“ Stufe 7 völlig zutreffend) eine Abnahme zur Skalenmitte hin, bei negativ formulierten Items (z. B. „bedrückt“ gar nicht zutreffend) eine Zunahme zur Skalenmitte hin.

Eine *Tendenz zur Mitte* kann jedoch aus verschiedenen Gründen zustande kommen:

(a) Eine *statistische Regression zur Mitte* ist durch Messfehler bedingt. Diese führen dazu, dass der Wert einer zweiten Messung zum Mittelwert (Dichtezentrum) der Verteilung tendiert.

(b) Beim Rückblick steht grundsätzlich der gesamte Range der Skala von 1 bis 7 zur Verfügung. Unter der Annahme, dass sich die rückblickende Einstufung wahrscheinlich mehr oder minder an einem gedachten Tagesdurchschnitt orientiert, könnte es zu einer Varianzeinengung des Rückblicks im Vergleich zum Tages-Mittelwert, d. h. einer *Restriktion des Bereichs*, kommen (bei einem gedachten Tages-Mittelwert 2, würde der Rückblick nicht mehr unter das Skalenende bei 1 gehen können. Ob diese mittelungsbedingte Bereichsrestriktion den Retrospektionseffekt beeinflusst ist jedoch fraglich.

(c) Die Selbsteinstufung des Befindens auf einer Skala ist, wenn diese Aufgabe gründlich aufgefasst wird, mit semantischen Schwierigkeiten und mit der Wahl eines Bezugssystemen verbunden. In dieser Situation der Unsicherheit könnte eine *Urteilsheuristik* eingesetzt werden, die mittlere, d. h. unentschiedene Skalenstufe „mittelmäßig“ zu bevorzugen.

(3) Zwischen der rückblickenden Einstufung am Abend und den vorausgegangenen Einstufungen kann eine *Ausgangswert-Abhängigkeit* (ähnlich dem Ausgangswertproblem bei physiologischen Messwiederholungen) bestehen (siehe unten).

Neben diesen *psychometrischen* und *kognitionspsychologischen* Erklärungshypothesen steht also eine *persönlichkeitspsychologische* Interpretationen. Es scheint ein Zusammenhang zwischen dem Betrag des Retrospektionseffektes und bestimmten Persönlichkeitseigenschaften zu bestehen, insbesondere der Skala *Emotionalität* (FPI-R). In diesem Ergebnis stimmten drei Untersuchungen überein: Emotional labile, selbstunsichere, sich stark beansprucht fühlende Personen tendieren dazu, Tagesereignisse und Befinden im Rückblick negativer zu bewerten als andere Personen. Dieser Zusammenhang ist über die gegenwärtige Fragestellung hinaus als ein Validierungshinweis für die FPI-R-Skalen im Hinblick auf wiederholte Selbstberichte unter Alltagsbedingungen interessant. Die Korrelation war jedoch nicht konsistent bei allen Items vorhanden und konnte in einer dritten Untersuchung nur tendenziell repliziert werden. Der psychologisch interessante Zusammenhang zeigte sich nicht systematisch, sondern hauptsächlich am ersten Tag der Untersuchung. Es könnte also sein, dass sich die Persönlichkeitseigenschaft Emotionalität im Sinne einer negativen Bewertung nur auf den ersten Tag der computer-unterstützten Erhebung auswirkt und mit zunehmender Gewöhnung verschwindet.

Personen mit hoher Ausprägung der Emotionalität haben bekanntlich die Tendenz, über viele körperliche Beschwerden zu klagen, geringe Lebenszufriedenheit zu äußern und ein negatives Krankheitsverhalten zu entwickeln. Dies wurde bevölkerungsrepräsentativ (siehe Fahrenberg, 1994b; Fahrenberg et al., 2001a) und in großen klinischen Studien (Myrtek, 1998) belegt.

Kritisch bleibt jedoch anzumerken, dass es sich um eine spezielle Personenauswahl handelte. Die Verteilung der Testwerte der Skala N des FPI-R in dieser Untersuchungsgruppe entspricht zwar ungefähr der Testnorm, jedoch könnte, trotz der zugesicherten Anonymität der Ergebnisse, gerade unter den Studierenden der Psychologie eine Tendenz zur sozialen Erwünschtheit bestimmter Antworten vorhanden sein.

Ein Vorbehalt ergibt sich jedoch aus der Überlegung, dass sich eine habituelle Tendenz zur negativen Bewertung der Tagesereignisse und der Befindlichkeit bereits auf die *aktuellen* Einstufungen auswirken müsste (wie es z. B. für die Selbstberichte depressiv-ängstlicher Patienten zutreffen wird). In den geschilderten Untersuchungen korrelieren die Tages-Mittelwerte vieler Items (und die Rückblicke) tatsächlich substantiell mit den Persönlichkeitsskalen. Der negative Retrospektionseffekt ist aber durch die Bildung von Differenzen von den Tagesmittelwerten und Rückblicken statistisch abhängig.

Weiterhin ist kritisch zu überlegen, ob nicht symmetrisch zur Negativierung, Aggravation und Klagsamkeit bei einigen Personen auch eine Tendenz zur positiven Bewertung, hochgestimmten Überbewertung anzunehmen wäre, welche sich beim statistischen Vergleich ausmitteln könnten.

Sollte es sich beim negativen Retrospektionseffekt aber nicht nur um einen psychometrischen, sondern um einen „psychologischen“ Effekt handeln, wäre dies zweifellos wichtig für jede psychologisch-diagnostische Untersuchung, die sich auf Selbstberichte in Fragebogen und Interviews stützt.

Es gibt offensichtlich verschiedene Erklärungsansätze, wie es zu einem negativen Retrospektionseffekt kommt. Ist es überhaupt möglich, die verschiedenen hypothetischen Komponenten des Antwortprozesses zu trennen: den negativen Retrospektionseffekt als Komponente einer Persönlichkeitseigenschaft, die Urteilsheuristik bei Unsicherheit und die statistische Regression zur Mitte u. a. methodische Probleme der Veränderungsmessung?

Exkurs zu Ausgangswert-Beziehungen

Aus methodologischer Sicht hat das hier geschilderte Problem formale Ähnlichkeiten mit der sog. *Ausgangswert-Abhängigkeit*. Bei Personen, die bereits relativ hohe Blutdruckwerte haben, wird der Blutdruck bei einer Belastung meist um *kleinere* Beträge zunehmen als bei Personen mit geringerem Blutdruck. Die Diskussion über Ausgangswert-Beziehungen führte zur Entwicklung verschiedener

Meß- bzw. Fehlermodelle, die bestimmte Annahmen über die Reliabilität der Messungen machen (siehe u. a. Foerster, 1995). Auch in diesem Ausgangswertproblem sind offenbar die allgemeingültige statistische Regression zur Mitte, rechnerische Abhängigkeiten aufgrund der Korrelation einer Differenz (Veränderungswert) mit dem Ausgangswerten sowie *eine biologische Funktion* (Homöostase-Regulation zur Erhaltung eines Gleichgewichts) konfundiert (Fahrenberg, Foerster & Wilmers, 1995). Diese biometrischen Analysen beziehen sich fast ausschließlich auf inter-individuelle Vergleiche. Der von Wilder ursprünglich gemeinte Fall der intra-individuellen Veränderungen ist damit nicht berührt. Eine befriedigende Aufklärung und die Unterscheidung biologischer und biometrischer Komponenten des Problems ist nur von einer systematischen Variation der Ausgangswerte vor der Veränderungsmessung zu erwarten. Eine experimentelle Versuchsanordnung wäre jedoch sehr schwer zu realisieren.

Auch hinsichtlich der Veränderungsmessung mit psychologischen Tests gibt es eine breite Diskussion über psychometrische Fragen. Beim Retrospektionsproblem handelt es sich zwar nicht um eine Messwiederholung oder eine Veränderungsmessung im engeren Sinn, sondern um eine zusammenfassende Bewertung im Rückblick, doch werden ebenfalls Zusammenhänge zwischen Differenzen, d. h. Veränderungswerten, und vorherigen Werten untersucht.

An den Daten der Wochenstudie wurden Analysen der Ausgangswert-Beziehungen vorgenommen (Foerster, 1994, 1995). Die Tabellen im Anhang zeigen, dass solche Beziehungen bei einigen Items vorhanden sind, jedoch nicht systematisch. Die folgende Tabelle 2.6 fasst die Hauptergebnisse über die sieben Tage zusammen.

Tabelle 2.8: Zusammenfassung der statistischen Befunde über Ausgangswert-Beziehungen (Anzahlen signifikanter Befunde $p < .05$).

	Aktiv	Anstren- gend	Aufge- regt	Ärger- lich	Be- drückt	Er- schöpft	Guter Stim- mung	Körperl. wohl	Stress
Ausgangswert- Abhängigkeit	3	2	4	2	4	5	1	0	4
Signifikante Retrospektionseffekte bei verschiedenen Veränderungswerten									
Differenz	0	7	2	3	4	2	0	1	6
Wahre Differenz	0	7	2	3	4	1	0	1	6
Kovarianzanal. Verfahren ALS	0	7	2	3	4	2	0	1	6

Anmerkung: Die Anzahl signifikanter Ergebnisse für die Items wurde in den im Anhang enthaltenen Tabellen für die sieben Tage ausgezählt.

Die Items „anstrengend, belastend“ und „Stress“ ragen durch das systematische Auftreten, die drei positiven Items „aktiv, leistungsfähig“, „guter Stimmung“ und „körperlich wohl“ durch das fast völlige Fehlen von signifikanten Effekten hervor.

Der Einfluss des Typs der Veränderungswerte auf den Retrospektionseffekt ist sehr gering. Da die wahre Differenz die statistische Ausgangswertabhängigkeit, der ALS auch die echte Ausgangswertabhängigkeit eliminiert, kann geschlossen werden, dass Ausgangswertabhängigkeiten den Retrospektionseffekt nur wenig beeinflussen.

Grundsätzliche Fragen und allgemeine Skalierungsprobleme

Angesichts dieser speziellen Skalierungsprobleme dürfen die grundsätzlichen Probleme nicht übersehen werden:

- Die verwendeten Adjektive sind semantisch nicht eindeutig, z. B. kann das Adjektiv „angespannt“ unterschiedlich interpretiert werden: geistig, emotional, körperlich angespannt, mit tendenziell positiver oder mit deutlich negativer Bewertung.
- Die Stufen der Adjektivskalen entsprechen zunächst nur Ordinalmessungen, d. h. größer–kleiner–Verhältnissen in der Ausprägung von subjektiven Zuständen. Die interindividuellen Vergleiche und die statistischen Berechnungen implizieren jedoch die Annahme einer Intervallskalierung.
- Auch hinsichtlich der Anzahl der Skalenstufen gibt es konkurrierende Argumente. Eine Skala mit einer ungeraden Anzahl von Stufen hat eine mittlere Stufe, z. B. die Stufe 4 bei der siebenstufigen

Skala; bei gerader Anzahl von Stufen fehlt die Mitte, so dass eine Entscheidung zu einer der Seiten erzwungen wird. Dieses Format stösst bei manchen Personen auf Kritik, weil „mittelmäßige“ Ausprägungen nicht zugelassen werden. Außerdem könnte die zur Mitte tendierende Urteilsheuristik mit einer solchen Skala nicht grundsätzlich vermieden werden, sie würde nur auf zwei Stufen aufgeteilt werden.

- Die behauptete *Urteilsheuristik* bei Unsicherheit müsste zu einer Bevorzugung der Skalenmitte („unentschieden“) führen, wenn diese die Mitte der psychologischen Skala repräsentiert. Bei einer siebenstufigen Skala ist die Stufe 4 numerisch und auch psychologisch die Skalenmitte. Möglich wären Skalen, bei denen durch zusätzliche verbale Anker oder durch ungleiche Skalenintervalle dafür gesorgt wird, dass die Mitte des psychologischen Kontinuums der Selbstbeschreibung nicht mit der Mitte der vorgegebenen Skala zusammenfällt. Wenn solche Skalen verwendet würden, bliebe jedoch zweifelhaft, ob diese Absicht von den Untersuchungsteilnehmern tatsächlich so verstanden wird. Vielleicht wird die anschauliche Mitte der manifesten Skala weiterhin als „die Mitte“ aufgefasst. Praktisch könnten solche Skalierungsvarianten noch mehr Unsicherheiten zur Folge haben.
- Die meisten der verwendeten Items haben entweder einen positiven (eher erwünschten) oder einen negativen (eher unerwünschten) Inhalt. Bei der Skalierung von „gar nicht zutreffend“ (1) bis „völlig zutreffend“ (7) werden folglich die Mittelwerte der Verteilung eher rechts bzw. eher links der theoretischen Skalenmitte liegen. Dieser von der „psychologischen“ Polung abhängige Sachverhalt bedingt also einerseits eher Deckeneffekte, andererseits eher Bodeneffekte. (Dieser Unterschied könnte z. B. für eine graphische Darstellung des negativen Retrospektionseffektes in Abbildung 2.1 *nachträglich* durch eine Umpolung der negativen Items verändert werdend, d. h. Stufe 7 wäre dann „gar nicht zutreffend“.) Bei den *aktuellen* Einstufungen könnte jedoch ein Wechsel der Polung für positive und negative Items sehr irritierend wirken und zu neuen Missverständnissen führen.
- Die Skalierung geschieht ohne *ausdrückliche Instruktion* zur Verankerung in einem bestimmten Bezugssystem (Überlegungen zu Bezugssystemen und Erfahrungsorganisation siehe u. a. Mogel, 1990; Schwarz, 1990). So bleibt fraglich, ob z. B. der aktuelle Zustand „sehr“ angespannt (völlig zutreffend, Stufe 7) im Vergleich zum *gewöhnlichen* Befinden gemeint sein könnte oder die Aussage vielleicht eine spezielle Referenz hat (ein früher erlebten Zustand oder einen abstrakt vorstellbarer Zustand *extremer* Anspannung)? Durch Instruktionen könnte nach einer ipsativen Skalierung (intra-individuell im Vergleich zu der stärksten bisher selber erlebten Anspannung) oder eine allgemeine Skalierung (inter-individuell im Vergleich zu anderen Personen) gefragt werden, doch wären die Skalierungsprobleme und Unsicherheiten dadurch nicht prinzipiell zu lösen. Frühere Versuche mit solchen „Skalierungsübungen“, um beide Bezugssysteme zu erläutern und zu differenzieren führten häufig in eine schwierige Diskussion und schienen eher zu verwirren als die Aufforderung zu einer spontanen („naiven“) Selbsteinstufung des Befindens.
- Handelt es sich bei diesen Selbsteinstufungen vielleicht um eine unscharfe, subjektive, inkonsistente Verbindung von Bezugssystemen des aktuellen Erlebens und der habituellen Grundstimmung einer Person sowie von Bezugssystemen des *intra-individuellen* und des *inter-individuellen* Vergleichs?
- Spielen in den Antwortprozess weitere Urteilsheuristiken hinein, die sich an Tagesereignissen (u. a. an dem Maximum oder am Range der Erlebnisse) oder auch an den vermuteten Hypothesen der Untersucher und an anderen Annahmen orientieren? Die Rekonstruktion der Befindlichkeit im Tageslauf – und das selbstreflektierte Wissen darüber – werden empirisch höchstens in einigen Aspekten exploriert werden können. Welchen Einfluss haben besondere Tagesereignisse als mögliche Anker? Gibt es Hinweise auf anschließende Adjustierungen? Wie repräsentativ ist aus Sicht der Probanden jede der aktuellen Einstufungen und wie interpretieren die Probanden ihren Rückblick und die negativen Abweichungen?
- Wegen der relativ häufig verlangten Selbstberichte können nur wenige Items verwendet werden. Längere Itemlisten, wie sie in einigen standardisierten Stimmungsskalen verfügbar wären, sind hier ungeeignet. Sie würden hinsichtlich der Skores andere Verteilungen liefern als einfache Adjektivskalen, aber keine grundsätzlichen Vorteile bieten. Auch andere Itemformate, z. B. dichotome Items, Skalen mit gradzahliger Anzahl der Stufen, Itemformate mit ungleichen Intervallen oder die Verwendung spezieller verbaler Anker (zur Beeinflussung der Verteilungen) scheinen hier keinen Ausweg zu öffnen.

Es ist wichtig, auf mehrere formale Bedingungen zu achten, wenn der Retrospektionseffekt (aber auch andere Fragestellungen) verfolgt werden: die *Änderungssensitivität* (die Diskriminationsleistung zwischen Zuständen) und *Verteilungsparameter*. Die statistischen Verteilungen der Itemantworten sollten möglichst ohne große Boden- und Deckeneffekte sein. Diese Verteilungen hängen von der Formulierung der Items ab, u. U. auch von der untersuchten Population. Im allgemeinen liegen jedoch die Mittelwerte der Items solcher Selbsteinstufungen der Befindlichkeit bei positiver Formulierung (z. B. „Situation im Griff“) im oberen Bereich und bei negativer Formulierung (z. B. „bedrückt“) im unteren Bereich der siebenstufigen Skala. Die Schiefe dieser Verteilungen kann nicht als Artefakt angesehen werden, denn offensichtlich geht es den meisten Untersuchungsteilnehmern gut. Bei einzelnen Personen und bei bestimmten Patienten ergibt sich durchaus ein anderes Bild in Richtung einer „normalen“ Verteilung. Transformationen der Verteilungen wären hier jedoch problematisch. So wird eine Untersuchung von gesunden jungen Teilnehmern zu anderen Verteilungen führen als die Untersuchung von depressiven Patienten.

Verteilung der Itemantworten bei „positiven“ und „negativen“ Items

In den vorausgegangenen Untersuchungen konnte der negative Retrospektionseffekt deutlich repliziert werden. Die Interpretation dieses Effektes blieb jedoch offen. Eine befriedigende Klärung würde voraussetzen, dass zumindest drei hypothetische Tendenzen, d. h. der *emotional negative Retrospektionseffekt*, die *Tendenz zur Mitte aufgrund einer Urteilsheuristik bei Unsicherheit* und die *statistische Tendenz zur Mitte*, unterschieden werden können.

Die naheliegende Idee, diese hypothetischen Tendenzen methodisch zu trennen, führt aus den genannten Gründen in Schwierigkeiten. Im allgemeinen ist zu erwarten, dass die drei Tendenzen in dieselbe Richtung gehen. Ausnahmen wären nur bei relativ „schwierigen“ positiven Items und relativ „leichten“ negativen Items (jeweils mit Mittelwerten etwa bei Stufe 4 der siebenstufigen Skala) zu erwarten, denn in diesen Fällen müsste der negative Retrospektionseffekt zur Abweichung von der Mitte (bzw. von der Tendenz zur Mitte) führen. Die Hinzunahme weiterer positiv formulierter Items lieferte keinen klaren Befund, da auch bei diesen Items in der Regel sehr schiefe Verteilungen auftraten. Die hier verfügbaren Daten von Patienten der Rehabilitationsklinik liessen kein wesentlich schlechteres Befinden erkennen, so dass die hier aus psychometrischen Gründen erwünschten Verteilungen nicht gefunden wurden.

Die Untersuchungen sollten klären, ob der negative Retrospektionseffekt gleichermaßen für positiv formulierte Aspekte des Befindens (unter Berücksichtigung dieser Polung) besteht wie für die — in solchen Itemlisten oft dominierenden — negativen Aspekte (siehe auch Hank & Schwenkmezger, 1996). Der Effekt ist bei negativen Items zweifellos sehr viel stärker und konsistenter ausgeprägt, doch trat er gelegentlich auch bei positiven Items auf (teils gleichsinnig, teils gegensinnig). Die Formulierung (Polung) eines Items hat also einen wichtigen Einfluss, scheint aber keine ausreichende Erklärung für den negativen Bias zu geben.

Die Untersuchungen sprechen dafür, dass es sich bei den Erinnerungstäuschungen nicht nur um unsystematische Fehler handelt, doch sind Urteilsstrategien, Messfehlervarianzen, individuelle und situative Unterschiede wahrscheinlich zu vielgestaltig, um einfache Aussagen zu ermöglichen. In Anschlussarbeiten könnten u. a. die Einflüsse von Urteilstendenzen durch methodische Variationen differenzierter untersucht werden. Wenn außerdem statt einer siebenstufigen Skala eine visuelle Analogskala verwendet würde, könnten sich u. U. andere Verteilungen der Einstufungen ergeben.

Noch differenzierter angelegte Untersuchungen werden vielleicht besser aufklären können, inwieweit rückblickende Einstufungen der Befindlichkeit eine Akzentuierung durch den aktuellen Zustand oder durch Ereignisse, d. h. durch eine mögliche Veränderung des Bezugssystems, oder durch die Disposition „Emotionalität“ (im Sinne Eysencks) erhalten.

2.8 Zusammenfassung

Die computer-unterstützte Methodik hat sich in solchen Tageslaufuntersuchungen grundsätzlich bewährt. Bei Studierenden verschiedener Fächer und bei berufstätigen Personen und Patienten war die Akzeptanz der Methodik gut. Wenn es zu Datenausfällen kam, lag dies, wenn das jeweilige Programm ausgetestet war, in der Regel am Überhören des Signals. Dies kann verschiedene Ursachen haben. Da

ein genaues zeitliches Protokoll der Eingaben vorliegt, können u. U. verspätete, durch Selbstaktivierung des hand-held PC nachgeholte Eingaben noch akzeptiert werden, wenn sie innerhalb eines bestimmten Intervalls erfolgen.

In jedem Fall liefern die Untersuchungen starke Argumente für die PC-gestützte, und deswegen zeitlich genaue Aufzeichnung aktueller Selbstberichte. Die computer-unterstützte Selbstprotokollierung ist die Methode der Wahl, wenn es um mögliche Verzerrungen des Rückblicks, d. h. um den Vergleich zwischen aktuellen und retrospektiven Einstufungen geht. Die von anderen Autoren in ähnlichen Untersuchungen, z. B. über die Erinnerung von Emotionen oder Copingstrategien, verwendete Fragebogenmethodik ist gerade bei diesem Thema von zweifelhafter Validität.

Aus der hier dargestellten Serie von Untersuchungen über das Befinden unter Alltagsbedingungen mit MONITOR können mehrere Schlussfolgerungen gezogen werden:

Die Einstufungen der Befindlichkeit auf den siebenstufigen Adjektivskalen liessen in inter-individueller und in intra-individueller Hinsicht eine befriedigende Variabilität erkennen. Die meisten Verteilungen waren bei den untersuchten Gruppen gesunder Personen entweder linksschief oder rechtsschief (je nach psychologischer Polung des Items) und sprachen für eine überwiegend positive oder sehr positive Bewertung des Befindens und der Tagesereignisse. Varianzanalysen ergaben, dass die Items eine unterschiedliche Diskriminationsleistung bzw. Änderungssensitivität aufweisen. Einige Items differenzierten relativ besser zwischen Personen, andere liessen deutliche zirkadiane Verläufe erkennen, andere zeigten eher individualcharakteristische Verläufe an.

Die am Abend erhobenen Rückblicke auf den vergangenen Tag unterschieden sich systematisch vom Mittelwert der aktuellen Einstufungen: sie lagen näher bei der Skalenmitte. Psychologisch entspricht dies einer stärker negativen Bewertung der vergangenen Tages. Damit konnte der methodisch wichtige *Befund des negative Retrospektionseffektes repliziert werden*. Der Betrag dieses negativen Retrospektionseffektes in einigen Items (und der aus mehreren Items aggregierten Indizes) war mit der Persönlichkeitseigenschaft bzw. dem Konstruktbereich „Emotionalität“ korreliert. Auch wenn sich diese Beziehung nicht für alle Items oder in allen Untersuchungen repliziert werden konnte, ist das Ergebnis über mehrere Untersuchungen konsistent. Die Tendenz, sich beim Rückblick *nicht* am Tagesdurchschnitt zu orientieren (Metafrage) korreliert ebenfalls mit der Skala N.

Dieser Effekt tritt vorrangig bei Items mit *negativen* Beschreibungen des Befindens auf und nur selten bzw. mit psychologisch gegenläufiger Tendenz bei positiv formulierten Items. Der negative Retrospektionseffekt ist zwischen aufeinanderfolgenden Tagen reproduzierbar, doch erreichen die Stabilitätskoeffizienten dieser Differenzbeträge höchstens eine mittlere Größenordnung. Die Prüfung möglicher Ausgangswert-Abhängigkeiten ergab keine Befunde, die für eine wesentliche Rolle solcher Zusammenhänge sprachen.

Für diesen robusten Befund des negativen Retrospektionseffektes gibt es verschiedene Erklärungshypothesen. Die Mehrzahl der signifikanten Effekte könnte weiterhin als urteilsheuristische Tendenz bzw. statistische Regression zur Mitte gedeutet werden.

3 Fragestellungen der neuen Untersuchung

In der neuen Untersuchung sollen bisher weniger beachtete methodische Aspekte untersucht bzw. variiert werden. Die methodischen Änderungen und Neuerungen betreffen die folgenden Aspekte:

Die Hälfte der Teilnehmer erhält ein *Training*, das zu einer differenzierteren Beantwortung sowohl der aktuellen als auch der rückblickenden Befindenseinstufungen anregen soll. Es werden zwei verschiedene Alltagssituationen vorgegeben, um mit der Bedeutung der Adjektive, mit der Skalierung und mit der Differenzierung zwischen Situationen vertrauter zu machen. Die verstärkte Reflexion dieser Selbstberichte könnte die Varianz der Befindenseinstufungen erhöhen und die Auswirkung der Urteilsheuristik abschwächen.

In einer zweiten experimentellen Bedingungsvariation wird das *Intervall für den Rückblick* verändert: ganzer Tag oder halber Tag. Die Hälfte der Teilnehmer wird wie in den vorangehenden Studien um 21 Uhr nach einer Einschätzung des gesamten Tages gefragt; die andere Hälfte der Teilnehmer wird zweimal täglich zu einer rückblickenden Einschätzung aufgefordert: um 13 Uhr wird nach einer rückblickenden Einschätzung des Vormittages gefragt und um 21 Uhr nach einem Rückblick auf den Nachmittag, wobei in jedem Halbtage drei aktuelle Selbstberichte gewonnen werden. Diese Aufteilung des Rückblick soll der erwähnten Urteilsheuristik begegnen. Erwartet wird, dass aufgrund der größeren Varianz der Befindlichkeit des gesamten Tages im Tages-Rückblick *eher mittlere Werte* angegeben werden als beim Rückblick auf halbe Tage.

Die mögliche Bedeutung der Urteilheuristik wird noch auf andere Weise untersucht. Diese Tendenz müsste bei größerer subjektiver *Unsicherheit der Selbsteinstufungen* stärker ausgeprägt sein. Deswegen werden die Teilnehmer an die retrospektiven Angaben anschließend noch gebeten, die relative Sicherheit der Einstufungen zu bewerten. Es wird erwartet, dass sich eine größere Unsicherheit in der stärkeren *Tendenz zu mittleren Werten* widerspiegelt.

Eine weitere Neuerung ist die *prospektive* Einschätzung des folgenden Tages, die ebenfalls nach jeder abendlichen Retrospektion erhoben wird. Diese Einschätzung muss psychologisch schwieriger sein als ein Rückblick auf den vergangenen Tag und müsste deshalb eher zu der oben beschriebenen Urteilsheuristik führen, die Skalenmitte zu wählen. Außerdem kann der prospektive Wert vom ersten Abend mit dem Tagesmittelwert des zweiten Tages verglichen werden. Für diesen Zweck wurde, um die Itemliste nicht zu sehr anwachsen zu lassen, nur die Frage „Wie aktiv/ leistungsfähig werden Sie sich morgen fühlen?“ ausgewählt.

In der Absicht, größere Varianzen der Itemwerte zu erreichen und vielleicht auch die Tendenz zur Mitte zu reduzieren, wurde statt der bisher verwendeten Skala mit 7 Stufen eine *visuelle Analogskala* eingesetzt.

Durch diese experimentelle Bedingungen und die anderen methodischen Änderungen sollten verschiedene Aspekte der Urteilssicherheit bzw. Unsicherheit der Einstufungen variiert werden. Diese Veränderungen sollten sich (1) auf die Verteilungen bzw. die Varianz der Einstufungen und (2) entweder abschwächend oder verstärkend auf die hypothetische Urteilsheuristik im Falle von Unsicherheit auswirken. Durch diese Modifikationen sollen neue empirische Grundlagen zur Evaluation der Urteilsheuristik bei Unsicherheit mit Tendenz zur Mitte gewonnen werden. Ein Persönlichkeits-Fragebogen oder ein Test der Aufmerksamkeit wurden in dieser Untersuchung nicht verwendet.

4 Methodik

4.1 Versuchsplan

Die Variation der Bedingung *Skalierungs-Training/kein Training* und der Bedingung *Halbtags-Rückblick/Ganztags-Rückblick* führte zu einem zweifaktoriellen Plan. Im Hinblick auf den Gesamtaufwand der Untersuchung wurde vorgesehen, jeweils zwei Gruppen zusammenzufassen, falls sich bei der betreffenden Bedingungsvariation kein Effekt ($p > .05$) ergeben sollte.

Der negative Retrospektionseffekt wird wieder als Differenz zwischen dem Mittelwert der aktuellen Einstufungen des Tages und dem Rückblick berechnet. Der Effekt hat bei Items mit „negativem“ Inhalt (z. B. „bedrückt“) ein negatives Vorzeichen, falls der Tag rückblickend negativer beurteilt wird als aus den aktuellen Einstufungen zu vermuten ist, d. h. der Rückblick liegt näher zur Skalenmitte.

Für die Versuche zur Differenzierung der hypothetischen Einflüsse sind die Verteilungen der Itemwerte und ihr Abstand von der Skalenmitte interessant.

4.2 Untersuchungsteilnehmer

Untersuchungsteilnehmer waren Studierende verschiedener Fachrichtungen, ganz überwiegend Studierende der Psychologie im 2. Fachsemester. Diese erhielten ein Testat über Praktikums-Stunden innerhalb des Empiriepraktikums. Geworben wurden die Teilnehmer in erster Linie über einen Aushang im Psychologischen Institut. Zusätzlich wurden einige Studierende persönlich angesprochen. Die Datenerhebung fand im Zeitraum vom Mai 2002 bis Juni 2002 statt.

4.3 Termine

Der erste Tag der Untersuchung soll mit vier Terminen (13:00, 15:00, 18:00 und 21:00 Uhr), d. h. ohne den Vormittag, zur Eingewöhnung dienen. Der zweite Tag ist der eigentliche Untersuchungstag mit sechs Terminen (8:30 oder 9:00 oder 9:30, 11:00, 13:00, 15:00, 18:00, 21:00), wobei der erste Termin am Morgen individuell festgelegt wurde. So sollte vermieden werden, dass die Teilnehmer durch den PSION geweckt würden. Der Zeitpunkt des tatsächlichen Signals (Piepsen) bei den übrigen Terminen erfolgte in einem Intervall von ± 15 Minuten um den vorgesehenen Termin, um zu verhindern, dass sich die Studierenden auf einen zu erwartenden Zeitpunkt einstellen könnten. Die Beantwortung der Fragen dauerte etwa zwei bis drei Minuten.

Tabelle 4.1: Termine der Untersuchung

	Uhrzeit	Termin
1. Tag	13:00 Uhr	1
	15:00 Uhr	2
	18:00 Uhr	3
	21:00 Uhr	4
2. Tag	8:30/ 9:00/ 9:30 Uhr	5
	11:00 Uhr	6
	13:00 Uhr	7
	15:00 Uhr	8
	18:00 Uhr	9
	21:00 Uhr	10

Anmerkung: Das tatsächliche Signal kam innerhalb eines Bereichs von ± 15 Minuten. Der erste Termin am 2. Tag wurde individuell vereinbart und in MONITOR eingegeben.

4.4 Items

Die Items wurden aus dem Itempool der vorangegangenen Studie ausgewählt (Fahrenberg et al., 2002, Käppler et al., 2001, Leonhart, 2001, Scheibehenne et al., 2000). Die Itemzahl wurde auf acht reduziert (zuzüglich 3 Setting-Fragen, siehe Tabelle 4.2). Die Abfolge der Fragen war zu jedem der 10 Termine gleich und wurde lediglich um 21 Uhr (Ganztags-Rückblick) bzw. um 13 Uhr und um 21 Uhr (Halbtags-Rückblick) um den retrospektiven Fragenkatalog erweitert.

Durch die Reduktion auf nur acht Items zur Einstufung der Befindlichkeit (im engeren Sinn, zuzüglich Frage nach Stress in der Zwischenzeit und nach Störung) sollte der Zeitaufwand für das Protokoll möglichst gering gehalten werden, da die Frage nach einem Halbtagsrückblick und die Fragen nach Sicherheit der Einstufung und die prospektive Frage ohnehin einen zusätzlichen Aufwand mit sich brachten. Gesichtspunkte für die Itemauswahl waren die Erwartung einer möglichst guten Verständlichkeit der Adjektive für eine aktuelle und für eine retrospektive Einstufung sowie die relative Unabhängigkeit der Itemantworten, d. h. Informationen über die Höhe der intra-individuellen Kovariation solcher Items (siehe Fahrenberg et al., 2002; Scheibehenne et al., 2000; und Anhang).

4.5 Skalierung

Statt der bisher verwendeten Skala mit 7 Stufen (bei MONITOR in Gestalt von miteinander verbundenen Kästchen) wurde eine visuelle Analogskala VAS mit einer verbalen Verankerung der Skalenenden (gar nicht – völlig) eingesetzt. Diese Skala erschwert es, die genaue Mitte einzugeben, da keine Zahlenwerte sichtbar sind. Die Cursor-Position auf diesem Strahl wurde in 21 Stufen aufgelöst. Diese VAS verringert darüber hinaus etwaige Übertragungseffekte bei der Beantwortung der aufeinander folgenden Items (Ziffern könnten leichter behalten werden). Auch die Varianz der Antworten sollte von der 21-fachen (anstelle einer 7-fachen) Stufung der Skala profitieren.

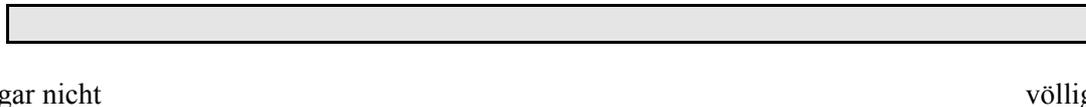


Abbildung 4.1: Visuelle Analog-Skala (intern mit Auflösung in 21 Stufen)

4.6 Skalierungstraining als experimentelle Bedingung

Das Training soll zu einer differenzierteren Beantwortung sowohl der aktuellen als auch der rückblickenden Befindenseinstufungen anhand der ausgewählten Items anregen. Die Einstufungen sind an drei *vorgestellten Situationen* zu üben. Diese Situationen sollten möglichst verschieden und zugleich alltagsnah sein. Den Teilnehmern sollte bewusst werden, wie sehr die Befindenskomponenten in verschiedenen Situationen variieren können. Um den Range der Skalierung zu verankern, bestünde die Möglichkeit, extreme Situationen (z. B. einen Autounfall oder einen Sechser im Lotto) vorzugeben – hier mangelt es jedoch wiederum an Alltagsnähe, und die Angaben wären problematisch, weil sie auf hypothetischen Überlegungen basierten. Außerdem könnte interindividuell die Varianz der Bezugssysteme der Skalierung eher noch größer werden.

So wurden drei alltägliche Situationen für das Training ausgewählt: „Weg zur Uni“, „Schlangestehen im Supermarkt“ und „Schwarzfahren in der Straßenbahn“ (siehe Anhang: Training). Erwartet wurde, dass die Teilnehmer, die ein Skalierungstraining erhielten, eine *größere Varianz* in ihren Antworten zeigen. Das Training sollte überdies zu einer differenzierteren Beantwortung ermutigen und so einer Tendenz der Versuchsteilnehmer entgegenwirken, aus Unsicherheit die Skalenmitte zu wählen.

Hand-held PC und MONITOR

Zur Datenerhebung diente der hand-held PC (Taschencomputer) PSION Serie 3. Der aufklappbare Computer wiegt 275 Gramm. Er verfügt neben einer kleinen Tastatur über einen monochromen LCD-Bildschirm (44 mm x 127 mm, 480 x 160 Pixel), auf dem maximal 25 Zeilen mit jeweils 80 Zeichen dargestellt werden können. Darüber hinaus ist er mit einer Uhr und einem kleinen akustischen Signalgeber (Alarm) ausgestattet. Der PC hat einen 16 Bit Prozessor und einen Arbeitsspeicher mit einer Kapazität zwischen 500 KB bis 2MB. Mit einem Batterie-Set ist der hand-held PC über mehrere Tage hinweg einsatzbereit.

Auf diesem PC wurde die von Brügger (1998) und Hüttner (2000) entwickelte Software MONITOR installiert (siehe auch Fahrenberg et al., 2002). Das Programm wurde den Untersuchungsbedingungen entsprechend modifiziert: Der neue Fragenkatalog wurde eingegeben und in zweifacher Version auf die PSION übertragen: einmal mit den Fragen zum Ganztags-Rückblick und einmal mit den Fragen zum Halbtags-Rückblick. MONITOR veranlasste zu vorgegebenen Zeitpunkten den Programmstart durch ein alternierendes, sich im Pegel in drei Stufen steigendes Alarmsignal. Wurde der dritte Alarm zweimal ignoriert, so schaltete der Computer nach etwa zwei Minuten automatisch ab und meldete sich erst wieder zur nächsten, zeitlich vorgegebenen Abfrage. Dennoch bestand für die Teilnehmer die Möglichkeit, die verpasste Abfrage nachzuholen, indem sie das Programm durch Tastendruck selbst starteten. Dieser verspätete Zeitpunkt wurde vom Programm ebenso wie die verpasste Eingabe festgehalten.

Tabelle 4.2: Liste der Setting-Merkmale und der Items zur Selbsteinstufung

Setting-Merkmale

Aufenthaltsort (Wo befinden Sie sich gerade?, 13 Kategorien)

Sozialer Kontext (Welche Personen sind im Moment anwesend?, 6 Kategorien)

Aktuelle Tätigkeit (Was haben Sie gerade vor dem Piepsen getan?, 22 Kategorien).

Fragen zur aktuellen Befindlichkeit

Fühlen Sie sich momentan aktiv/ leistungsfähig?

Ist die momentane Situation für Sie anstrengend/ belastend?

Glauben Sie, die momentane Situation „im Griff“ zu haben?

Fühlen Sie sich im Augenblick eher aufgeregt/ nervös?

Ist Ihre momentane Stimmung eher ärgerlich/ gereizt?

Ist Ihre momentane Stimmung eher bedrückt?

Fühlen Sie sich momentan körperlich wohl?

Sind Sie momentan guter Stimmung?

Wie stark erlebten Sie seit der letzten Eingabe Stress?

Gab es seit der letzten Eingabe besondere Ereignisse? (freie Texteingabe)

Wurden Sie beim Beantworten der Fragen abgelenkt oder gestört?

Ganztagesrückblick (21 Uhr) bzw. Halbtags-Rückblick (um 13 und um 21 Uhr)

Fühlten Sie sich heute (Vor-/Nachmittag) aktiv/ leistungsfähig?

War der heutige Tag (Vor-/Nachmittag) für Sie anstrengend/ belastend?

Glauben Sie, dass Sie den heutigen Tag (Vor-/Nachmittag) „im Griff“ hatten?

Fühlten Sie sich am heutigen Tag (Vor-/Nachmittag) eher aufgeregt/ nervös?

Waren Sie am heutigen Tag (Vor-/Nachmittag) eher ärgerlich/ gereizt?

Waren Sie am heutigen Tag (Vor-/Nachmittag) eher bedrückt?

Fühlten Sie sich am heutigen Tag (Vor-/Nachmittag) körperlich wohl?

Waren Sie heute (Vor-/Nachmittag) guter Stimmung?

Sicherheit der Einstufung

Wie sicher waren Sie sich bei der Beantwortung dieser Fragen? (21 Uhr)

(4 Kategorien: sehr unsicher, ziemlich unsicher, ziemlich sicher, sehr sicher).

Prospektive Frage

Wie aktiv/ leistungsfähig werden Sie sich morgen fühlen? (21 Uhr)

Alle zur Beantwortung der Fragen nötigen Informationen wurden auf dem Bildschirm mitgeteilt. Alle Eingaben (Ziffern und Texte) und die jeweiligen Anfangs-, und Endzeiten wurden automatisch gespeichert. Die Daten wurden später mit Hilfe eines Verbindungskabels und eines Hilfsprogramms auf einen stationären PC übertragen. Dort wurden sie im Microsoft „Notepad“ geöffnet und geordnet und dann in Excel eingelesen. Diese Daten wurden schließlich in das Statistikprogramm SPSS und SAS eingelesen und ausgewertet.

4.8 Durchführung

Die Selbstberichte wurden in dieser Studie ausschließlich an Werktagen und nicht am Wochenende erhoben. Die Übergabe des PC erfolgte am Vormittag des ersten Erhebungstages in einem Laborraum des Psychologischen Instituts Freiburg. Nur in Ausnahmefällen wurde die erste Eingabe um 13:00 Uhr ausgelassen, wenn es einem Teilnehmer nicht möglich war, einen der gegebenen Vormittagstermine wahrzunehmen. Der PSION und der Untersuchungsablauf wurden von einem der Versuchsleiter erläutert, wobei meist ein zweiter anwesend war, welcher sich jedoch im Hintergrund hielt.

Die Untersuchungsteilnehmer wurden zufällig einer der beiden Bedingungen ‚Retrospektion ganztags‘ bzw. ‚Retrospektion halbtags‘ und ‚mit Training‘ bzw. ‚ohne Training‘ zugeordnet.

In der Regel erfolgte die Rückgabe des PC nach der letzten, abendlichen Abfrage am zweiten Tag. Dabei gab es noch eine kurze Nachbefragung bezüglich Akzeptanz und möglicher Probleme bei der Durchführung.

Jedem Teilnehmer wurde eine standardisierte Form der Instruktionen vorgelesen (siehe Anhang). Anschließend wurde die Handhabung des Taschencomputers erklärt. Jeder absolvierte unter Aufsicht des Versuchsleiters einen Probedurchlauf. Inhalt dieses Probedurchlaufs war der Fragenkatalog zur Bewertung der aktuellen Befindlichkeit, nicht jedoch der retrospektive Fragenkatalog. Dieser Schritt diente lediglich zur Übung, und der Datensatz wurde anschließend gelöscht. Außerdem wurden die ungefähren Zeitpunkte genannt, zu denen die Befragungen stattfinden würden, und die Teilnehmer wurden gefragt, wann der PSION am nächsten Morgen zum ersten Mal piepsen sollte (zur Auswahl standen 8.30, 9.00 oder 9.30 Uhr). Daraufhin wurden in MONITOR die speziellen Einstellungen hinsichtlich des ersten Termins und hinsichtlich der Bedingung Ganztags-/Halbtags-Rückblick vorgenommen. Die Teilnehmer wurden gebeten, die Abfrage selbst zu starten, falls sie einen Messzeitpunkt versäumt hatten.

Training

Der Hälfte der Teilnehmer erhielt anschließend ein Skalierungstraining. Die drei auf einzelnen Din-A-4-Blätter beschriebenen Situationen wurden jeweils mit einem Antwortblatt, das die 8 Befindensitems aus der Fragenliste mit visuellen Analogskalen enthielt, vorgelegt (siehe Anhang). Die Teilnehmer wurden gebeten, sich in die jeweilige Situation hineinzusetzen und ihre Befindlichkeit anhand der Itemliste einzuschätzen.

4.9 Fehlende Daten

Im Anschluss an die Datenübertragung vom PSION auf den stationären PC erfolgte die Formatierung in EXCEL. Daraufhin wurden die Datensätze inspiziert und nach der Vollständigkeit bzw. dem Grad der Compliance der Teilnehmer gruppiert (siehe Tabelle 4.3). Die Datensätze wurden in vier Gruppen eingeteilt:

Tabelle 4.3: Gruppierung der Probanden nach Vollständigkeit der Daten.

Gruppe	Tag 1	Tag 2	Retrospektion am Tag 2
1	vollständig	vollständig	vorhanden
2	nicht vollständig	vollständig	vorhanden
3	nicht vollständig	1 Termin fehlend	vorhanden
4	nicht vollständig	mehr als 1 Termin fehlend	Vorhanden/fehlend

Fehlender Termin bedeutet, dass das Piepsen des PSION entweder völlig überhört wurde (z. B. wegen zu starken Umgebungslärms) oder die Beantwortung abgebrochen wurde. In einigen Fällen wurde die Befragung – nachdem das Piepsen überhört wurde – von den Teilnehmern selbst aktiviert. Die entsprechende Verzögerung der Programmaktivierung wurde im Datensatz protokolliert. Um die Intervalle der Termine möglichst homogen zu halten wurden nur solche Eingaben berücksichtigt, die im Zeitraum ± 30 Minuten um den geplanten Termin lagen. Beim ersten Termin am Morgen wurde ein größerer Spielraum zugelassen, da dieser Termin individuell dem Tagesrhythmus angepasst werden sollte. Das Intervall zulässiger selbstaktivierter Eingaben wurde hier auf die Zeit zwischen 8.00 Uhr und 10.00 festgelegt.

Für die statistische Analyse des Retrospektionseffektes und des Trainings wurden *nur die Daten des 2. Tages* verwendet. Hier wurde maximal ein fehlender Messzeitpunkt für die Mittelwertbildung toleriert. Die Daten der Kategorie 4 wurden aus der weiteren Analyse ausgeschlossen. Demnach waren 64 von 87 Datensätzen auswertbar. (Einen Überblick gibt die Tabelle 4.4).

Tabelle 4.4: Gruppierung der Datensätze (Teilnehmer) nach Compliance

Gruppe	Häufigkeit	Prozent	Kumuliert
1 vollständig	32	36.8	36.8
2 fehlende Daten am Tag 1	15	17.2	54.0
3 ein Termin Tag 2 fehlt, Retrospektion vorhanden	17	19.5	73.5
4 eliminiert	23	26.4	100.0
Gesamt	87	100.0	

Im Hinblick auf den zweifaktoriellen Untersuchungsplan ergab sich schliesslich die folgende Anzahl von Personen für jede der Bedingungen:

Tabelle 4.5: Verteilung auf die Untersuchungsbedingungen (N = 64)

	Ganztags-Rückblick	Halbtags-Rückblick
Training	16	17
Kein Training	19	12

4.10 Statistische Analyse

In dieser Untersuchung sind die statistischen Verteilungen der Itemwerte, d. h. Mittelwert, Standardabweichung, Schiefe und Kurtosis, von besonderem Interesse sowie die möglichen Veränderungen der Verteilungsform und ihre zu- oder abnehmende Distanz von der Skalenmitte.

Die Abweichung der rückblickenden Antworten vom Mittelwert der aktuellen Einstufungen werden mit einem t-Test für abhängige Stichproben geprüft. Außerdem wird der nicht-parametrische Wilcoxon-Test eingesetzt, um die zweifelhaften Intervallskaleneigenschaft der Selbsteinstufungen zu berücksichtigen. Hinsichtlich dieser Strategie der Absicherung gibt es jedoch ebenfalls Vorbehalte wegen der zahlreich vorkommenden „verbundenen Ränge“.

Die Effektstärke des erwarteten negativen Retrospektionseffektes wird als Mittelwertdifferenz der aktuellen und der retrospektiven Einstufungen auf die Standardabweichung der aktuellen Einstufungen des 2. Tags bezogen.

Die Gruppenunterschiede (mit Training/ ohne Training; Retrospektion ganztags/ Retrospektion halbtags) werden mit einer zweifaktoriellen Varianzanalyse mit Messwiederholung geprüft. Im Falle signifikanter Unterschiede sind auch hier nicht-parametrische Vergleiche geplant. Die Unterschiedshypothesen hinsichtlich der Gruppierung nach sicherer/unsicherer Einstufung und hinsichtlich der Abweichungen zwischen prospektiver Einstufung und Skalenmittelwert werden ebenfalls mit t-Tests für unabhängige bzw. abhängige Stichproben geprüft.

5 Ergebnisse

5.1 Variabilität und andere deskriptive Befunde

Die Antworten auf die acht Items unterschieden sich deutlich in ihren Verteilungen. Die Itemmittelwerte variierten auf der 21-stufigen Skala zwischen $M = 3.1$ (ärgerlich/gereizt) und $M = 14.7$ (Situation im Griff). Die Standardabweichungen zwischen $SD = 4.46$ (ärgerlich/gereizt) und $SD = 5.68$ (Stress seit der letzten Eingabe). Die Boxplots lassen die links- und rechtsschiefen Verteilungen der „positiv“ und der „negativ“ formulierten Items erkennen. Die Items „aktiv, leistungsfähig“ und „gute Stimmung“ weisen am ehesten eine „normale“ Verteilung auf (siehe Tabellen im Anhang).

5.2 Skalierungstraining und Ganztags-/Halbtags-Rückblick

Vor den geplanten Analysen war zu prüfen, ob sich die Bedingungsvariationen „Training“ und „Zeitraum der Retrospektion“ auf den Retrospektionseffekt ausgewirkt haben oder ob diese Gruppen eventuell zusammengefasst werden können.

Die zweifaktorielle Varianzanalyse mit Messwiederholung ergab keinerlei signifikanten Haupt- oder Interaktions-Effekt. Die assoziierten F-Werte hatten sämtlich einen p -Wert $> .10$, so dass die Nullhypothesen beibehalten wurden. Auf eine Wiedergabe der Tabellen kann hier verzichtet werden. Das Skalierungstraining und die Verkürzung des Zeitraums für den Rückblick auf den Vormittag bzw. Nachmittag hat keinen Einfluss auf die Mittelwerte der aktuellen Einstufungen am Tag und den Retrospektionseffekt. Auch aus den Histogrammen des Ganztags-Rückblicks und des Halbtags-Rückblicks (aus Vormittag und Nachmittag gemittelt) sind keine Hinweise auf einen deutlichen Einfluss des Zeitraums der Retrospektion auf die Verteilung der retrospektiven Werte zu entnehmen.

Diese Ergebnisse rechtfertigen die Zusammenfassung der Untersuchungsgruppen. Die folgenden Analysen beziehen sich auf diesen *aggregierten Datensatz*

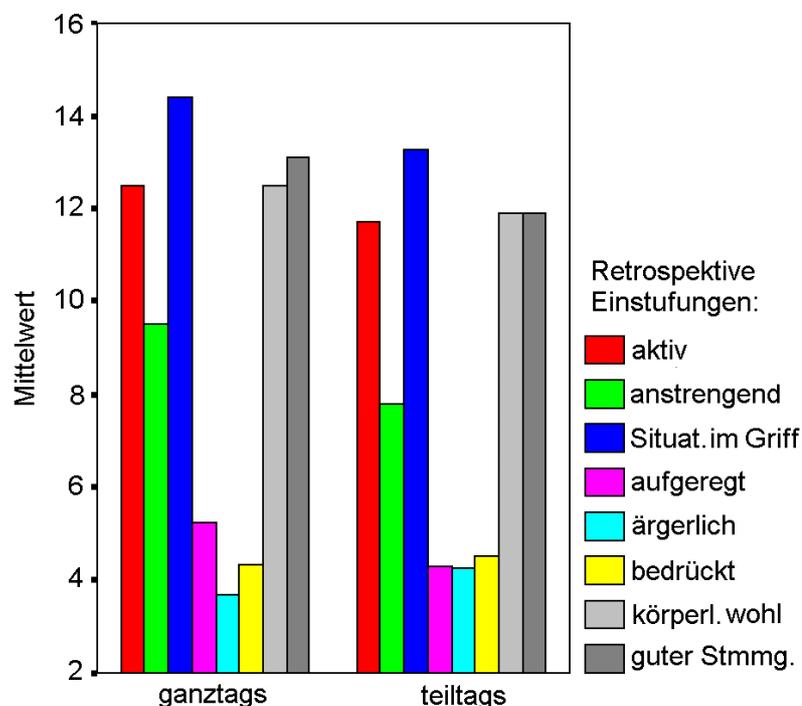


Abbildung 5.1: Retrospektive Einstufungen ganztags ($N = 35$) und halbtags ($N = 29$). Anmerkung: Beim Rückblick wurde das hier unpassende Item „Stress während der letzten Zeit“ nicht verwendet.

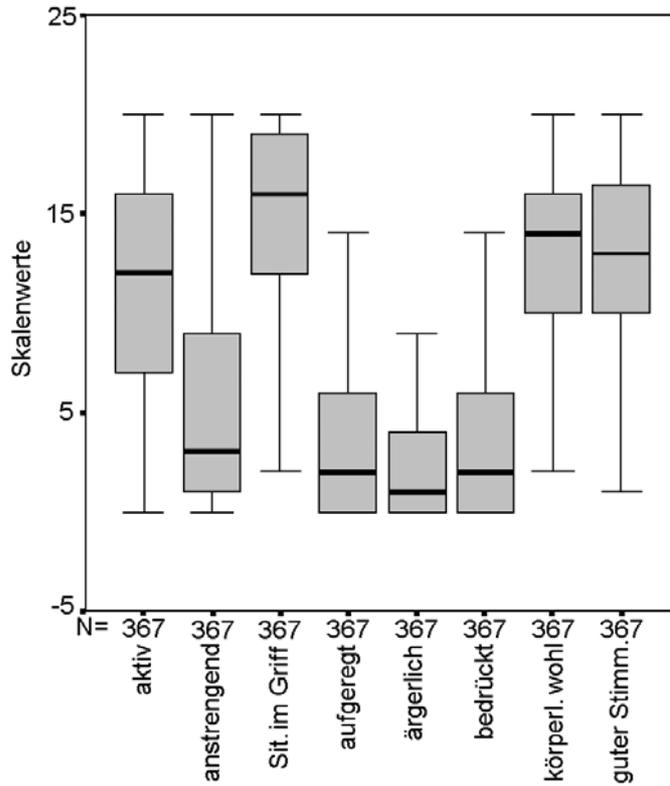


Abbildung 5.2: Boxplots der Tagesmittelwerte (2. Tag, N = 64)

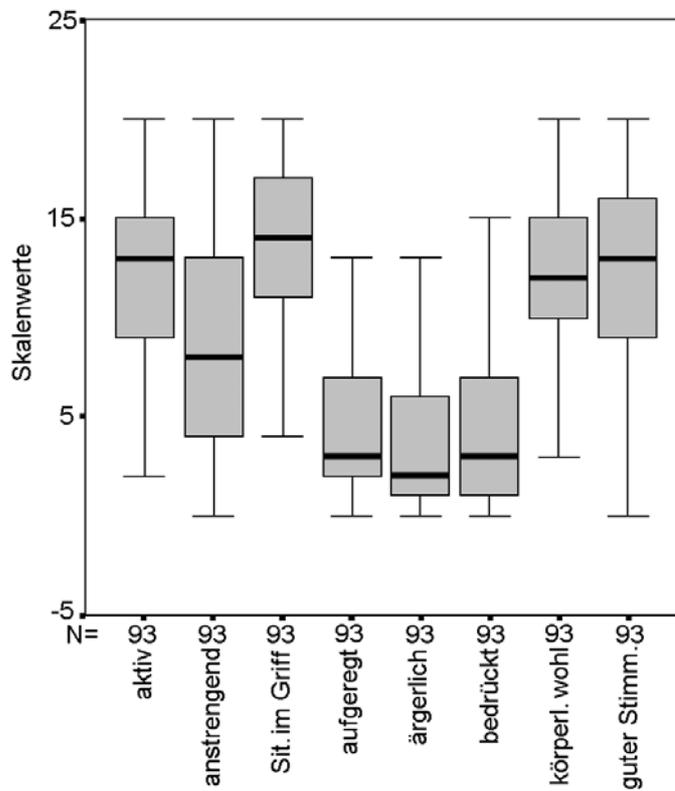


Abbildung 5.3: Boxplots der retrospektiven Einstufungen (Abend des 2. Tages, N=64). Anmerkung: Die 93 Retrospektionen ergeben sich aus n = 35 Ganztags- und n = 29 x 2 Halbtags-Rückblicken

Die Boxplots der retrospektiven Einstufungen entsprechen weitgehend den Boxplots der Tages-Mittelwerte (Beim Rückblick wurde das hier unpassende Item „Stress während der letzten Zeit“ nicht verwendet. Die meisten Verteilungen lassen – über alle Itemwerte betrachtet – eine befriedigende Differenzierung innerhalb des Skalenbereichs von 1 bis 21 erkennen. Bei mehreren Items weisen die Verteilungen jedoch eine deutliche Schiefe auf (Schiefe und Kurtosis siehe Tabelle im Anhang).

Die Standardabweichungen der retrospektiven Einstufungen sind – im Unterschied zu früheren Untersuchungen – nicht systematisch größer als die SD der aktuellen Einstufungen. Sämtliche Mittelwerte der retrospektiven (und der aktuellen) Einstufungen sind hochsignifikant von der Skalenmitte 10.5 entfernt (Ein-Stichproben t-Tests).

Die retrospektiven Einstufungen für den Vormittag und den Nachmittag unterscheiden sich bei einigen Personen deutlich voneinander, wobei nur in Einzelfällen aus den zusätzlichen Texteingaben zu sehen war, dass besondere Ereignisse vorlagen. Im allgemeinen korrelieren die Einstufungen der Halbtage bei den N = 29 Personen dieser Untersuchungsbedingung relativ hoch in der Größenordnung von $r = .70$ (Ausnahmen „aufgeregt, nervös“ .31, „aktiv“ .59 und „gute Stimmung“ .66). Am höchsten ist die Stabilität bei „körperlich wohl“ .76, einem relativ überdauernden Befindensaspekt.

5.3 Retrospektionseffekt

Der Mittelwert der sechs aktuellen Einstufungen am 2. Tag sowie der abendliche Rückblick und die Differenz beider Werte sind in Tabelle 5.1 dargestellt.

Der t-Test und der Wilcoxon-Test stimmen darin überein, dass in vier von vier negativen Befindens-Items ein signifikanter oder sehr signifikanter Mittelwertunterschied besteht: „anstrengend, belastend“, „aufgeregt, nervös“, „ärgerlich, gereizt“ und „bedrückt“. Bei zwei weiteren Items ist ein Effekt nur durch einen der beiden Tests oder durch einseitige Prüfung zu sichern: „Situation im Griff“ und „körperlich wohl“. Die Null-Hypothese ist hinsichtlich der Items „aktiv, leistungsfähig“ und „guter Stimmung“ beizubehalten (siehe Tabelle 5.1).

Im Rückblick wurde der Tag als anstrengender erlebt, das Allgemeinbefinden wurde als aufgeregter, gereizter, bedrückter und das körperliche Wohlbefinden als schlechter eingestuft, als es aufgrund der Mittelwerte der 6 aktuellen Einstufungen zu erwarten war. Der negative Retrospektionseffekt konnte folglich für die Mehrzahl der 8 Items repliziert werden. Die Effektstärken dieser Unterschiede liegen zwischen niedrigen (0.21 bei „bedrückt“) und hohen Werten (1.06 bei „anstrengend, belastend“). Hervorzuheben ist, dass bei drei von vier positiv formulierten Items *keine* auffälligen negativen Retrospektionseffekte auftraten.

Tabelle 5.1: Mittelwerte der aktuellen Einstufungen am Tag 2, retrospektive Einstufung am Abend und Retrospektionseffekt (Differenz) mit zweiseitigem t-Test und Wilcoxon-Test

Item	Einstufung	Mittelwert	SD	t-Test t	p	Wilcoxon p
aktiv/ leistungsfähig	Mittelwert	11.48	3.04			
	Retrospektion	12.16	4.10			
	Differenz	-0.68	3.11	-1.738	.087	.074
anstrengend/ belastend	Mittelwert	4.96	3.55			
	Retrospektion	8.73	5.16		***	***
	Differenz	-3.77	4.28	-7.035	.000	.000
Situation im Griff	Mittelwert	14.58	2.96			
	Retrospektion	13.89	3.83			*
	Differenz	+0.68	2.84	1.926	.059	.036
aufgeregt/ nervös	Mittelwert	3.79	3.24			
	Retrospektion	4.83	3.93		**	**
	Differenz	-1.04	2.94	-2.825	.006	.002
ärgerlich/ gereizt	Mittelwert	2.84	2.68			
	Retrospektion	3.95	4.33		**	**
	Differenz	-1.10	2.69	-3.278	.002	.006
bedrückt	Mittelwert	3.71	3.31			
	Retrospektion	4.41	4.20		*	*
	Differenz	-0.71	2.33	-2.431	.018	.014
körperlich wohl	Mittelwert	12.95	3.22			
	Retrospektion	12.24	4.22		*	
	Differenz	+0.70	2.60	2.162	.034	.062
guter Stimmung	Mittelwert	12.90	3.20			
	Retrospektion	12.57	3.96			
	Differenz	+0.33	2.59	1.022	.311	.643

Anmerkung: Die Mittelwerte beziehen sich auf die 6 aktuellen Einstufungen (von 0 = gar nicht bis 20 = völlig) des zweiten Untersuchungstages (* $p \leq .050$, ** $p \leq .010$, *** $p \leq .000$, $N = 64$).

5.4 Relative Sicherheit

Die Mehrzahl der Teilnehmer fühlte sich bei der rückblickenden Einschätzung des Tages relativ sicher (siehe Tabelle 5.2). Da die Kategorie „sehr unsicher“ nicht vorkam, verblieben drei Gruppen.

Tabelle 5.2: relative Sicherheit der Einstufungen (Angaben bei der Befragung um 21 Uhr des zweiten Erhebungstages)

	Häufigkeit	Prozent	Kumuliert
sehr unsicher	0.0	0.0	0.0
ziemlich unsicher	7	10.9	10.9
ziemlich sicher	43	67.2	78.1
sehr sicher	14	21.9	100.0
Gesamt	64	100.0	

Der Vergleich der Gruppen „ziemlich unsicher“, „ziemlich sicher“ und „sehr sicher“ lieferte deutliche Unterschiede in den Mittelwerten der retrospektiven Einstufungen. Die Abbildung 5.4 zeigt, dass in

den Gruppen, die sich bei der rückblickenden Beantwortung sicherer waren, extremere Werte angegeben wurden, als bei der unsichereren Gruppe.

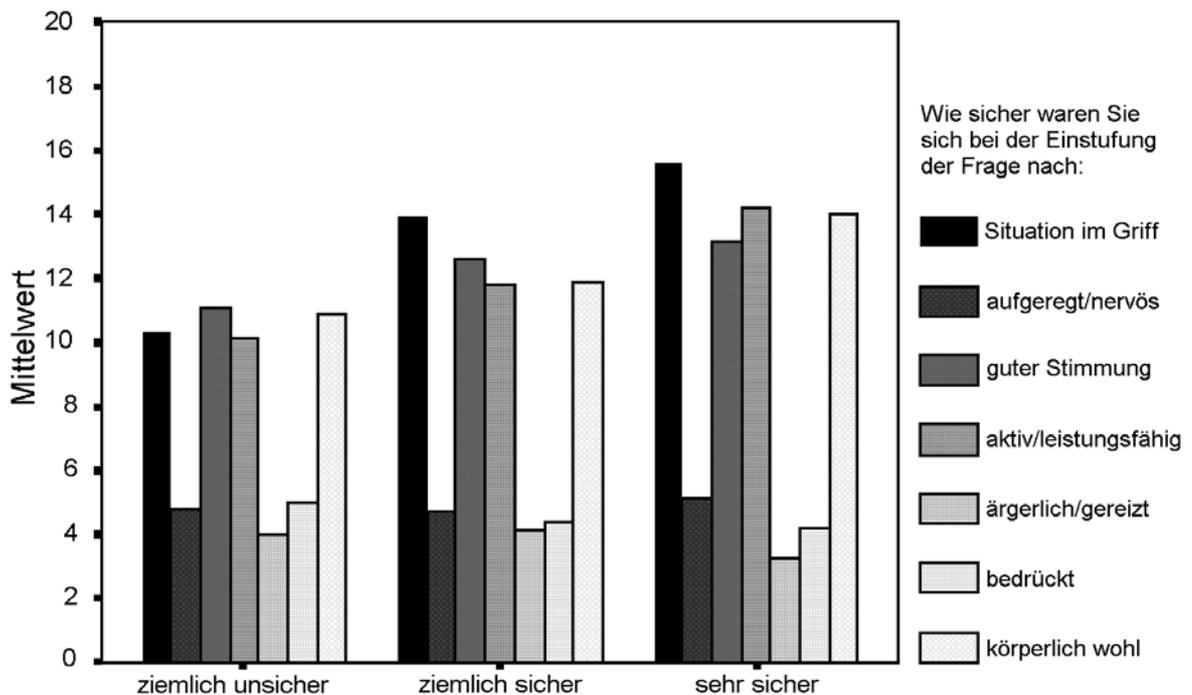


Abbildung 5.4: Vergleich der retrospektiven Einstufungen gruppiert nach der relativen Sicherheit der Einstufung bei der letzten Befragung am Tag 2.

Es wurde erwartet, dass Personen mit einer höheren Sicherheit ihrer Einstufungen einen geringeren Retrospektionseffekt und eine geringere Tendenz zur Mitte erkennen lassen würden. Ein Mittelwertvergleich mit dem t-Test und dem nicht-parametrischem Kruskal-Wallis-Test wurde wegen der zu geringen Besetzungen nur zwischen zwei Gruppen vorgenommen: (1) Einstufung „ziemlich unsicher“ zusammengefasst mit „ziemlich sicher“ (N = 50) und (2) Einstufung „sehr sicher“ (N = 14).

Die t-Test ergaben Unterschiede im Retrospektionsbetrag bei 2 der 8 Items: „aktiv, leistungsfähig“ (sehr sicher M = -2.16, sicher M = -0.26; p = .043, df = 52) und „aufgeregt, nervös“ (sehr sicher M = -2.45; sicher M = 0.64; p = .041, df = 62). Der F-Test zeigte ungleiche Varianzen im Item „anstrengend, belastend“ (sehr sicher SD = 5.73; sicher SD = 3.67; F = 8.78, p = .004). Sehr sichere Personen hatten also entgegen der Erwartung einen etwas größeren negativen Retrospektionseffekt.

Bei den sich *sehr sicher* fühlenden Teilnehmern müsste der Mittelwert eigentlich weiter von der Skalenmitte entfernt liegen als bei den weniger sicheren Einstufern, die aufgrund der hypothetischen Urteilsheuristik die unentschiedene Position bevorzugen. Es ergaben sich für diese „sehr sicher“ einstufernden 14 Teilnehmer. Tendenziell wurde diese Annahme bestätigt, aber nur bei den Items „Situation im Griff“ und „guter Stimmung“ war der Unterschied knapp signifikant. Bei dieser Gruppe waren die Retrospektionsbeträge in den Items „aktiv, leistungsfähig“ und „aufgeregt, nervös“ relativ größer als bei den nur sicheren Einstufern. Die Erwartungen wurden also nur tendenziell bzw. in inkonsistenter Weise gestützt: Die sehr sicheren Einstufer lagen geringfügig weiter von der Skalenmitte entfernt und zeigten in zwei Items einen größeren negativen Retrospektionseffekt.

Weiterhin wurde erwartet, dass Personen mit Rückblick-Intervallen von halben Tagen sich bei ihren Einstufungen relativ sicherer fühlen würden als Personen mit Ganztagsrückblick. Der t-Test ergab, dass die Null-Hypothese beizubehalten ist.

Die statistischen Analysen zu verschiedenen Aspekten der *Sicherheit der Einstufungen* ergaben nur wenige Befunde, die für einen Einfluss dieser Variable sprechen. Diese Schlussfolgerung ist jedoch mit dem Vorbehalt verbunden, dass sich große Mehrzahl der Teilnehmer bei ihren Einstufungen sicher fühlten.

5.5 Prospektive Einstufung

Angesichts der Unsicherheit einer Vorhersage des Befindens am kommenden Tag wurden hier eine deutliche Tendenz zur Mitte und eine größere Varianz der prospektiven Einstufungen erwartet.

Der Mittelwert der prospektiven Einschätzung am Abend des 1. Tages („Wie aktiv/leistungsfähig werden Sie sich morgen fühlen?“) war $M = 12.19$ ($SD = 4.31$, $N = 59$). Der Mittelwert der aktuellen Einstufungen am 2. Tag lag tendenziell unter diesem Wert: $M = 11.50$ ($SD = 5.02$). Die beiden Werte unterschieden sich jedoch nicht signifikant ($t = 1.612$, $p = .112$; $df = 58$). Die Varianz der prospektiven Einstufungen war tendenziell sogar geringer und nicht größer. Auffällige Verteilungsanomalien bestanden nicht.

Diese prospektive Einstufung korrelierte mit dem Mittelwert der aktuellen Einstufungen in diesem Item am folgenden Tag ($r = .428$; $p < .001$), doch beträgt die gemeinsame Varianz nur ca. 20 %. Die prospektive Einstufung am Abend des 1. Tages (für den 2. Tag) korreliert zu $r = .33$ ($p = .01$, $N = 59$) mit der prospektiven Einstufung am Abend des 2. Tages für den dritten Tag ($M = 12.44$, $SD = 4.47$, $N = 64$).

Die durch die notwendige Vorhersage des Befindens am folgenden Tag geschaffene Unsicherheit führte *nicht* zu einer deutlichen Verschiebung in Richtung Skalenmitte oder zu einer Varianzerweiterung, wenn die tatsächlichen Einstufungen als Vergleich gewählt werden.

5.6 Weitere deskriptive Ergebnisse

Geschlechtsunterschiede

Explorativ wurde geprüft, ob in den Tagesmittelwerten, Rückblicken oder Retrospektionsbeträgen Mittelwertunterschiede zwischen den 51 Frauen und den 13 Männern bestanden. Die t-Tests ergaben keine konsistenten Unterschiede.

Sicherheit der Einstufung

Bei den 13 Teilnehmern mit den sehr sicheren Einstufungen der aktuellen Tageswerte konnte eventuell erwartet werden, dass sie auch ihre prospektive Einstufungen relativ sicherer geben würden als die 50 nur „sicheren“ Einstufer.

Tabelle 5.3: Subjektive Sicherheit der prospektiven Einstufungen

	M	SD
Sehr sichere Einstufer		
Prospektive Einstufung (4) am Abend des 1. Tages	13.54,	5.25;
Einstufung (10) am Abend des 2. Tages	14.86,	4.33;
Sichere Einstufer		
Prospektive Einstufung (4) am Abend des 1. Tages	11.80,	3.99;
Einstufung (10) am Abend des 2. Tages	12.76,	4.31.

Das signifikante Ergebnis des t-Test auf Mittelwertunterschiede am Abend des 1. Tages (bei Annahme gleicher Varianzen, $p = .021$) spricht für die Annahme, dass die subjektive Sicherheit sich auch auf die prospektiven Einstufungen erstreckt.

Diskrepante Tageshälften

Es könnte sein, dass es eine Auswirkung auf die rückblickenden Einstufungen hat, wenn eine der Tageshälften deutlich anstrengender oder aufregender erlebt wurde als die andere. Die Teilnehmer wurden danach gruppiert, ob diese Differenz zwischen dem Mittelwert der drei aktuellen Einstufungen am

Vormittag im Vergleich zu dem Mittelwert der drei Einstufungen am Nachmittag gleich oder größer als 3 war (eine post hoc Dichotomisierung). Dies war beim Item „anstrengend, belastend“ bei 13 und bei dem andren für diese Exploration ausgewählten Item „aufgeregt, nervös“ bei 14 Personen der Fall.

Bei der *Gruppierung nach „anstrengend, belastend“* zeigten die t-Tests (unter Berücksichtigung gleicher/ungleicher Varianzen) – abgesehen von der verwendeten Variablen selber – keine signifikanten Unterschiede in den Tagesmittelwerten, in den retrospektiven Einstufungen oder den Differenzen. Eine Ausnahme bildete nur das Item „körperlich, wohl“. Hier wurden die Gruppenunterschiede hinsichtlich der retrospektiver Einstufung ($p = .017$) und der Differenz aktueller und retrospektiver Einstufungen ($p = .029$) signifikant.

Bei der *Gruppierung nach „aufgeregt, nervös“* gab es aufgrund dieser Dichotomisierung in den Tagesmittelwerten der Items „anstrengend, belastend“, „ärgerlich, gereizt“, „bedrückt“, „Situation im Griff“ und „gute Stimmung“ signifikante oder sehr signifikante Unterschiede, sowie in der retrospektiven Einstufung hinsichtlich „ärgerlich, gereizt“. In keinem Fall unterschieden sich die Retrospektionsbeiträge signifikant.

Diese Befunde lassen sich aus der Korrelation der verwendeten Items erklären und sprechen für die innere Konsistenz der Einstufungen. Nach dieser explorativen Analyse scheint die Ausprägung des Retrospektionseffektes nicht nachhaltig von der Dynamik des Befindens im Tageslaufs beeinflusst zu sein.

Item-Interkorrelationen

Einige der Items korrelierten im Querschnitt (inter-individuell, R-Technik) untereinander relativ hoch. So korreliert zum Beispiel das Item „anstrengend/belastend“ negativ mit den Items „Situation im Griff“ ($r = -.50$; $p \leq .000$) und „guter Stimmung“ ($r = -.60$; $p \leq .000$) und positiv mit den Items „aufgeregt/nervös“ ($r = .51$; $p \leq .000$), „ärgerlich/gereizt“ ($r = .54$; $p \leq .000$), „Stimmung bedrückt“ ($r = .51$; $p \leq .000$) und „Stress seit der letzten Eingabe“ ($r = .50$; $p \leq .000$).

Für die Beurteilung der gemeinsamen Varianz (und möglichen Redundanz) der Items wichtiger ist die Korrelation im Längsschnitt (intra-individuell, P-Technik). Hier wird jedoch auf eine Tabelle verzichtet, da die Korrelationsmatrix der Wochenstudie mit 7 x 6 Terminen aussagefähiger ist (siehe Anhang).

6 Diskussion der neuen Studie und Ausblick

6.1 Akzeptanz

Auch in dieser Studie hat sich der Einsatz der PSION-Taschencomputer mit dem Programm MONITOR zur Erfassung der Befindlichkeit im Alltag bewährt. Die Kommentare der Teilnehmer ergab, dass sie die Beantwortung der Fragen als unaufwendig und wenig störend empfunden hatten. Auch die Handhabung des Computers bereitete ihnen keine Schwierigkeiten. Auf die Frage: „Wurden Sie beim Beantworten der Fragen abgelenkt oder gestört?“ entsprach die mittlere Antwort der Skalenstufe von 1.8 bei 21 Stufen der visuellen Analogskala (Stufe 1 = gar nicht).

Die relativ große Anzahl fehlender Termine kam meist dadurch zustande, dass der Alarm des PSION aufgrund großen Umgebungslärms (z.B. in der Mensa) überhört wurde. Trotz der Aufforderung, in solchen Fällen den Computer selbst zu starten, geschah dies entweder nicht oder sehr viel später. Wenn die Untersuchungsteilnehmer den Alarmton des Taschencomputers nicht gehört hatten oder anderweitig in Anspruch genommen waren, sollten sie die Abfrage selbst starten, sobald sie ihr Versäumnis bemerkt hatten oder sich dem Taschencomputer widmen konnten. Dies geschah jedoch oft viel später. Um die Intervalle der Befragungszeitpunkte möglichst homogen zu halten, wurden nur Eingaben berücksichtigt, welche im Zeitraum ± 30 Minuten um die festgelegte Uhrzeit lagen. Ein liberaleres Kriterium hätte eine größere Anzahl von Datensätzen erreichen lassen.

6.2 Negativer Retrospektionseffekt

Der negative Retrospektionseffekt wurde auch in dieser Untersuchung repliziert. In fünf von acht Items gab es signifikante oder sehr signifikante Effekte mit mittlerer bis hoher Effektstärke. Bei den „positiven“ Items „Situation im Griff“ und „guter Stimmung“ zeigte sich ebenfalls eine schwache Tendenz in negativer Richtung, sie wurde jedoch nicht signifikant. Bei negativ formulierten Adjektivskalen ist dieser Effekt robust, bei positiv formulierten Adjektivskalen inkonsistent, fehlt entweder oder weist mal in die eine, mal in die andere Richtung.

In den vorausgegangenen Kapiteln, insbesondere im Abschnitt 2.7, wurden die verschiedenen Erklärungsversuche des negativen Retrospektionseffektes diskutiert: *die psychometrischen Erklärungen der Tendenz zur Mitte, die persönlichkeitspsychologischen Interpretationen und die kognitionspsychologischen Erklärungsversuche.*

In der vorliegenden Untersuchung ging es vor allem um mögliche *Urteilstendenzen bzw. Urteilsheuristiken* sowie um die *subjektive Sicherheit* bei der Selbsteinstufung der Tagesereignisse und der Befindlichkeit.

Wer den gesamten Tag rückblickend bewerten soll, bildet wahrscheinlich gedanklich einen Durchschnitt zwischen einem beispielsweise eher positiven und den eher negativen Tagesabschnitten und zwischen den intensiveren und den schwächeren Eindrücken. Dass dieser „subjektive Mittelwert“ *psychologisch* gebildet wird und daher anders ist als das arithmetische Mittel der erhobenen Daten, liegt nahe. In diesem Falle müsste der Rückblick in kleineren Abschnitten, d. h. getrennt für Vor- und Nachmittag, die Möglichkeit bieten, eine differenziertere und sicherere Beantwortung zu leisten. Die Differenzierung von *Halbtags-Rückblicken im Vergleich zu Ganztags-Rückblicken* hatte jedoch keinen Einfluss auf die Befunde zum Retrospektionseffekt.

Ein *Training der Skalierung* hinsichtlich der drei vorgestellten Situationen könnte zu differenzierteren Selbstbeurteilungen führen und auf diese Weise den negativen Retrospektionseffekt verringern. Durch die „Gedankenreisen“ im Training sollte den Teilnehmern nahegebracht werden, wie sich das Befinden in verschiedenen Situationen ändern kann – und sie dadurch zur differenzierteren Beantwortung ermutigen, d. h. sich wirklich für eine Seite (‘gar nicht’ bzw. ‘völlig’) zu entscheiden. Dies könnte sich auf die Verteilungen, d. h. auf die Position des Mittelwertes (auch dessen Abweichung von der Skalenmitte) und auf die Varianz, die Schiefe und die Kurtosis der Verteilungen auswirken.

Wenn diese Versuchsbedingung ebenfalls keinen Einfluss auf den negativen Retrospektionseffekt hatte, kann das bedeuten, dass dieses Training noch unzureichend war. Oder die Teilnehmer waren von vorn herein bemüht, möglichst genaue Angaben zu machen – sowohl bei den aktuellen als auch

bei den rückblickenden Einstufungen. Auch dies würde dagegen sprechen, dass der Retrospektionseffekt nur durch undifferenzierte Angaben zustande kommt.

Hinsichtlich der selbst eingestuften *relativen Sicherheit der Antworten* ließ sich zwischen den Gruppen kein signifikanter Unterschied in der Einschätzung der Befindlichkeit feststellen, mit Ausnahme des Items „Situation im Griff“. Eine relativ größere subjektive Unsicherheit bei den Einstufungen führte nicht generell zu einer stärkeren Tendenz zur Mitte oder zu einer Varianzerweiterung. Sehr sichere Personen hatten entgegen der Erwartung einen etwas größeren negativen Retrospektionseffekt.

Die *prospektive Einstufung* des kommenden Tages hinsichtlich „aktiv, leistungsfähig“ zeigte, dass selbst bei objektiv erhöhter Unsicherheit nicht die Skalenmitte bevorzugt wird. Vielmehr war am nächsten Tag der Mittelwert der aktuellen Einstufungen sogar der Skalenmitte näher als die prospektive Einstufung. Es ist natürlich möglich, dass die Ergebnisse bei anderen Items divergieren.

Die Gestaltung der Skala als siebenstufige Likert-Skala oder als visuelle Analog-Skala scheint keine wesentliche Rolle zu spielen.

Zusammenfassend ist festzustellen, dass die verschiedenen Hypothesen zur Rolle von Urteilstendenzen bzw. zur Auswirkung einer Urteilsheuristik empirisch nicht bestätigt wurden. Zwar lässt sich die Möglichkeit einer Urteilsheuristik als Ursache für den Retrospektionseffekt weiterhin nicht völlig ausschließen, doch reicht dieser Erklärungsversuch offensichtlich nicht aus.

In *psychometrischer Hinsicht* wurden keine Verteilungsauffälligkeiten, die im Zusammenhang mit dem negativen Retrospektionseffekt stehen könnten, beobachtet. Gegen eine „elementare“ psychometrischen Erklärungsversuch, im negativen Retrospektionseffekt ein Artefakt zu sehen, spricht auch der Sachverhalt, dass der Effekt nicht streng systematisch auftritt, sondern auch bei negativ formulierten Adjektivskalen (mit unauffälliger Verteilung) fehlen kann. Die statistischen Analysen lieferten keine Befunde, die für eine systematische Ausgangswert-Abhängigkeit sprechen. *Auch die psychometrischen Überlegungen, so berechtigt sie u. a. hinsichtlich der Skalierungsprobleme sind, reichen als Erklärungsversuch nicht aus.*

Als Ergebnis der Untersuchungen kann immerhin der *Geltungsbereich* des negativen Retrospektionseffektes eingeschränkt werden. Es handelt sich zwar um einen *robusten Effekt*, aber er tritt vorwiegend bei bestimmten Adjektivskalen auf: es sind Items mit *negativem* Inhalt und unter diesen vor allem die Items „anstrengend, belastend“, „bedrückt“ und „Erlebten Sie seit der letzten Eingabe Stress?“ Vielleicht regen gerade diese Iteminhalte an, zustimmende Antworten zu geben. Die Wochenstudie zeigte jedoch, dass die Stabilität (Reproduzierbarkeit) des negativen Retrospektionseffektes über die Tage bemerkenswert gering ist. Dies spricht gegen die Annahme einer persönlichkeitsabhängigen habituellen Urteilstendenz. In der Wochenstudie zeigte sich die Beziehung des negativen Retrospektionseffektes mit dem Konstrukt „Emotionalität“ nur am ersten Tag. Die persönlichkeitspsychologische Hypothese müsste folglich modifiziert werden.

6.3 Kritische Überlegungen zur internen und externen Validität

Die neue Untersuchung zeichnet sich durch einige methodische Verbesserungen aus. Dazu gehörten die Verwendung einer visuellen Analogskala, die Unterscheidung von Rückblicksintervallen (ganztags/halbtags) und die Berücksichtigung der subjektiven Sicherheit der Einstufungen. Diese Modifikationen sollten vor allem der gründlicheren Evaluation der kognitionspsychologischen Hypothesen hinsichtlich Urteilsheuristiken, Urteilstendenzen und Urteilssicherheit dienen. Diese Bedingungsvariationen waren jedoch kaum wirksam. So bleibt offen, ob diese Bedingungen tatsächlich nur einen geringen Einfluss hatten oder ob es nicht gelang, diese Bedingungsvariation in hinreichendem Maße zu realisieren.

Auch das Training hat nicht in der beabsichtigten Weise gewirkt. Die drei Situationen wurden von den Versuchspersonen als alltagsnah und realistisch empfunden, was für die Auswahl der Situationen spricht. Möglicherweise war jedoch das hier entworfene Training zu kurz, um zu einer differenzierteren Skalierung des Befindens anzuregen. Verbesserungen wären vielleicht durch Hinzufügen weiterer Situationen zu erreichen oder durch eine Vergleichsmöglichkeit mit den Einstufungen der anderen Teilnehmer, um die Varianz des Befindens deutlicher zu machen. Künftige Untersuchungen könnten diese Aspekte des Trainings weiter ausbauen bzw. andere Formen eines Trainings entwickeln.

Wie bei allen Untersuchungen der Befindlichkeit könnte die interne Validität der Datenerhebung durch eine mit der Anzahl der Abfragen zunehmende Sensibilisierung, durch erhöhte Selbstaufmerk-

samkeit, aber auch durch abnehmende Motivation beeinträchtigt sein. Manche Teilnehmer scheinen die Befragung so schnell wie möglich hinter sich bringen zu wollen, wenn der PC sie zu ihnen ungelegenen Zeitpunkten zur Eingabe aufforderte. Differenzierte Angaben wären dann schwieriger. Insbesondere bei den Messzeitpunkten um 13 Uhr und 21 Uhr, an die sich, je nach Untersuchungsbedingung, eine retrospektive Einschätzung des Befindens anschloss, ist zu fragen, ob die Teilnehmer zu diesen Zeiten besonders eilig, müde oder abgelenkt waren. So könnten sie um 13 Uhr gerade aus einer Lehrveranstaltung kommen und auf dem Weg zum Mittagessen sein. Dies könnte sich sowohl auf die aktuelle als auch auf die retrospektive Befindenseinschätzung ausgewirkt haben.

Zu jedem Termin wurde nach besonderen Ereignissen in Form einer freien Texteingabe gefragt. Diese sollten vor allem bei Auffälligkeiten ausgewertet und mit den korrespondierenden Befindenseinschätzungen verglichen werden. Extreme Werte in der Befindenseinschätzung sollten so mit eventuellen – positiven oder negativen – besonderen Ereignissen in Beziehung gesetzt werden. Jedoch wurden nur selten Angaben zu besonderen emotionalen Ereignissen gemacht. Deshalb konnte dieser Aspekt noch nicht gründlicher analysiert werden. Auch in den vorausgegangenen Studien gab es relativ wenig Zusatzinformationen dieser Art. Durch Aggregation über mehrere dieser Untersuchungen kann künftig vielleicht mehr zu dieser Frage ausgesagt werden.

Hinweise auf besondere Versuchleitereffekte sind nicht deutlich geworden. Die Teilnehmer wurden von einem Versuchsleiter instruiert, wobei (zumindest in den ersten zwei Wochen) ein weiterer Versuchsleiter anwesend war, sich allerdings im Hintergrund hielt. So nahm jeder der drei Versuchsleiter mehrmals bei der Versuchsdurchführung durch einen der beiden anderen Versuchsleiter teil.

Die Generalisierbarkeit der Untersuchungsergebnisse im Hinblick auf die Datenerhebung ist wegen der naturalistischen Bedingungen („ökologische Validität“) hoch zu bewerten – zumindest im Vergleich zu Fragebogendaten. Dagegen ist die Generalisierbarkeit auf andere Personen eher gering, da hier ganz überwiegend Studierende der Psychologie im zweiten Fachsemester teilnahmen. Die vorausgegangenen Untersuchungen an berufstätigen Personen sowie an Studierenden aus anderen Fächern hatten jedoch zur Kernfrage des negativen Retrospektionseffektes gleichlautende Ergebnisse.

Auch die Itemauswahl bleibt weiterhin zu diskutieren. Dies gilt hinsichtlich der semantischen Aspekte und hinsichtlich der Anzahl der Items. In dieser Studie konnte der Retrospektionseffekt für alle negativ gepolten Items repliziert werden, jedoch nur für ein positiv gepoltes Item. Es gelang nicht, durch bestimmte Items die Mittelwerte der Verteilungen wesentlich zu beeinflussen, um eine befriedigendere Verteilung um die Skalenmitte zu erreichen.

6.4 Ausblick

Die berichteten Untersuchungsergebnisse über den negativen Retrospektionseffekt stimmen im Prinzip mit den früheren Befunden überein. Für eine Reihe von Erklärungsversuchen gab es insgesamt noch keine hinreichenden empirischen Bestätigungen.

Zur persönlichkeitspsychologischen Interpretation des negativen Retrospektionseffektes als Ausdruck einer habituellen Disposition zur emotionalen Labilität und Klagsamkeit konnte die neue Untersuchung nichts beitragen. Die vorausgegangenen Untersuchungen zeigten, dass dieser Befund nicht konsistent auftritt, vielleicht auch von der Personenauswahl abhängt. Hohe Testwerte der Skala FPI-R sind in mehreren Items auch mit höheren Tagesmittelwerten und höheren rückblickenden Einstufungen (und nicht nur mit deren Differenz als Betrag der Retrospektion) assoziiert. Aus der Wochenstudie stammt das Ergebnis, dass die Korrelation zwischen dem Betrag der Retrospektion und der Skala N nur am ersten Untersuchungstag deutlich zu beobachten war, also u. U. einen der Gewöhnung unterliegenden Bewertungsprozess repräsentiert.

Ein wiederholter – wenn auch nicht völlig konsistenter – Befund war der Unterschied zwischen den positiv und den negativ formulierten Adjektivskalen. Enthält die negative Formulierung bereits eine suggestive Komponente, im Rückblick zu generalisieren? Oder ist die Lage der Verteilung relativ zur Skalenmitte wesentlich? Mit den bisher verwendeten und z. T. gezielt modifizierten Items konnte dieser Aspekt nicht befriedigend untersucht werden. Vielleicht kann durch neu entworfene Items die Absicht, sowohl für negative wie für positive Zustände Verteilungen um die Skalenmitte zu erreichen, verwirklicht werden. Dies ist eine wichtige Aufgabe für künftige Untersuchungen.

Angesichts der beim gegenwärtigen Stand relativ unergiebigem Erklärungsversuche kognitionspsychologischer und psychometrischer Art, könnte sich nun der Blick wieder mehr den differentiell-

psychologischen Perspektiven zuwenden. Hierbei wären intra-individuelle Explorationen des Bezugssystems und der u. U. aufgrund von Tagesereignissen modifizierten rückblickenden Einstufungen möglich. Anhand der inzwischen angewachsenen Datenbasis können eher statistische Vergleiche zwischen Datensätzen, die explizit auf besondere, positive oder negative, emotionale Ereignisse verweisen, durchgeführt werden. Die geringe Ergiebigkeit der in der vierten Untersuchung verwendeten „Metafragen“ zu diesen Aspekten, d. h. dem möglichen Wissen über solche Urteilstendenzen, lassen jedoch die Schwierigkeiten solcher Analysen erkennen.

Der negative Retrospektioneffekt erweist sich, trotz der bisherigen Bemühungen weiterhin als eine besondere Herausforderung für kritische Methodenstudien.

Literaturverzeichnis

- Brügner, G. (1998). MONITOR: Ein flexibles Programm zur Datenerhebung mittels Pocket-PC. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 19, 145-147.
- Buse, L. & Pawlik, K. (1996). Ambulatory behavioral assessment and in-field performance testing. In: J. Fahrenberg & M. Myrtek (Eds.). *Ambulatory assessment: computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 29-50). Seattle, WA: Hogrefe & Huber.
- Buse, L. & Pawlik, K. (2001) Computer-assisted ambulatory performance tests in everyday situations: Construction, evaluation, and psychometric properties of a test battery measuring mental activation. In: J. Fahrenberg & M. Myrtek (Eds.). *Progress in ambulatory assessment* (pp. 3-23). Seattle, WA: Hogrefe & Huber Publishers.
- Clark, D. M. & Teasdale, J. D. (1982). Diurnal variation in clinical depression and accessibility of memories of positive and negative experiences. *Journal of Abnormal Psychology*, 91, 87-95.
- De Beurs, E., Lange, A. & Van Dyck, R. (1991). Self-monitoring of panic attacks and retrospective estimates of panic: Discordant findings. *Behaviour Research and Therapy*, 30, 411-413.
- DeLongis, A., Folkman, S. & Lazarus, R. S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality and Social Psychology*, 54, 486-495.
- Fahrenberg, J. (1994a). Ambulantes Assessment. *Diagnostica*, 40, 195-216.
- Fahrenberg, J. (1994b). *Die Freiburger Beschwerdenliste (FBL)*. Göttingen: Hogrefe
- Fahrenberg, J. (1997). Ambulantes psychophysiologisches Assessment in der Klinischen Psychologie. In: B. Rockstroh, T. Elbert & H. Watzl (Hrsg.) *Impulse für die Klinische Psychologie* (S. 143-165). Festschrift für Rudolf Cohen. Göttingen: Hogrefe.
- Fahrenberg, J., Brügner, G., Foerster, F. & Käppler, C. (1999). Ambulatory assessment of diurnal changes with a hand-held computer: Mood, attention, and morningness – eveningness. *Personality and Individual Differences*, 26, 641-656.
- Fahrenberg, J., Foerster, F. & Franck, M. (1995a). *Response scaling: Night-time baselines, resting baselines, and initial-value dependencies* (Forschungsbericht Nr. 121). Freiburg i. Br.: Albert-Ludwigs-Universität, Psychologisches Institut.
- Fahrenberg, J., Foerster, F. & Wilmers, F. (1995b). Is elevated blood pressure level associated with higher cardiovascular responsiveness in laboratory tasks and with response specificity? *Psychophysiology*, 32, 81-91.
- Fahrenberg, J., Hampel, R. & Selg, H. (2001a). *Freiburger Persönlichkeitsinventar FPI-R*. (7. Auflage). Göttingen: Hogrefe.
- Fahrenberg, J., Hüttner, P. & Leonhart, R. (2001b). Psychological assessment in everyday life by hand-held PC: Applications of MONITOR. In: J. Fahrenberg & M. Myrtek (Eds.). *Progress in ambulatory assessment* (pp. 93-112). Seattle, WA: Hogrefe & Huber Publishers.
- Fahrenberg, J., Leonhart, R. & Foerster, F. (2002). *Alltagsnahe Psychologie. Datenerhebung im Feld mit Hand-held PC und physiologischem Meßsystem*. Bern: Huber.
- Fahrenberg, J. & Myrtek, M. (Eds.). (2001). *Progress in ambulatory assessment*. Seattle, WA: Hogrefe & Huber Publishers.
- Foerster, F. (1994). Über die Probleme von Ausgangswertabhängigkeiten und Reaktions-Skalierungen. *Forschungsberichte des Psychologischen Instituts der Universität Freiburg i.Br., Nr. 104*.
- Foerster, F. (1995). On the problems of initial-value-dependencies and measurement of change. *Journal of Psychophysiology*, 9/4, 324-341.
- Franck, M., Herrmann, J. M. & Fahrenberg, J. (1996). Psychophysiologisches Blutdruck-Monitoring. *Nieren- und Hochdruckkrankheiten*, 25, 195-201.
- Hank, P. & Schwenkmezger, P. (1996). Computer-assisted versus paper-and-pencil based self-monitoring: An analysis of experimental and psychometric equivalence. In J. Fahrenberg & M. Myrtek (Eds.), *Ambulatory assessment. Computer-assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 85-99). Seattle, WA: Hogrefe & Huber Publishers.
- Hank, P., Schwenkmezger, P. & Schumann, J. (2001). Daily mood reports in hindsight: Results of a computer-assisted time sampling study. In: J. Fahrenberg & M. Myrtek (Eds.). *Progress in ambulatory assessment* (pp. 143-156). Seattle, WA: Hogrefe & Huber Publishers.
- Hawkins, S. A. & Hastie, R. (1990). Hindsight: Biased judgements of past events after outcomes are known. *Psychological Bulletin*, 107, 311-327.
- Hedges, S. M., Jandorf, L. & Stone, A. A. (1985). Meaning of daily mood assessments. *Journal of Personality and Social Psychology*, 48, 428-434.
- Hufford, M. R., Shiffman, S., Paty, J. & Stone, A. A. (2001). Ecological momentary assessment: Real-world, real-time measurement of patient experience. In: J. Fahrenberg & M. Myrtek (Eds.). *Progress in ambulatory assessment* (pp. 69-92). Seattle WA: Hogrefe & Huber Publishers.

- Hüttner, P. (2001). *MONITOR Manual* [Programmbeschreibung Version 9]. Freiburg i. Br.: Psychologisches Institut der Albert-Ludwigs-Universität, Forschungsgruppe Psychophysiologie.
- Käppler, C. (1994). *Psychophysiologische Bedingungsanalyse von Blutdruckveränderungen im alltäglichen Lebenskontext*. Frankfurt a. M.: Peter Lang.
- Käppler, C., Becker, H.-U. & Fahrenberg, J. (1993). Ambulantes 24-Stunden-Monitoring als psychophysiologische Assessmentstrategie: Reproduzierbarkeit, Reaktivität, Retrospektionseffekt und Bewegungskonfundierung. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *14*, 235-251.
- Käppler, C., Brügger, G. & Fahrenberg, J. (2001). Pocketcomputer-unterstütztes Assessment mit MONITOR: Befindlichkeit und Aufmerksamkeitsverlauf im Alltag, Methodenakzeptanz und die Replikation des Retrospektionseffektes. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *22*, 249-266.
- Käppler, C. & Rieder, S. (2001). Does the retrospection effect hold as a stable phenomenon? First results from a transcultural self-monitoring study of mood and cognitive state in Brazil and Germany. In: J. Fahrenberg & M. Myrtek (Eds.). *Progress in ambulatory assessment* (pp. 113-122). Seattle, WA: Hogrefe & Huber Publishers.
- Kendall, M.G. & Stuart, A. (1967). *The advanced theory of statistics* (Vol. 2) 2nd ed. London: Griffin.
- Lacey, J.I. (1956). The evaluation of autonomic responses: toward a general solution. *Annals of the New York Academy of Sciences*, *67*, 123-164.
- Leonhart, R. (2001). *Die Verwendung des Handheld-Computers in Psychologie und Medizin. Unveröffentlichte Diplomarbeit*. Albert-Ludwigs-Universität Freiburg i. Br.: Psychologisches Institut, Forschungsgruppe Psychophysiologie.
- Margraf, J. & Jacobi, F. (1997). Marburger Angst- und Aktivitätstagebuch. In G. Wilz & E. Brähler (Hrsg.). *Tagebücher in Therapie und Forschung. Ein anwendungsorientierter Leitfaden* (S. 139-153). Göttingen: Hogrefe.
- Margraf, J., Taylor, C. B., Ehlers, A., Roth, W. T. & Agras, W. S. (1987). Panic attacks in the natural environment. *Journal of Nervous and Mental Disease*, *175*, 558-565.
- Mogel, H. (1990). Bezugssystem und Erfahrungsorganisation. Göttingen: Hogrefe.
- Mussweiler, T., Förster, J. & Strack, F. (1997). Der Ankereffekt in Abhängigkeit von der Anwendbarkeit ankerkonsistenter Information: Ein Modell selektiver Zugänglichkeit. *Zeitschrift für Experimentelle Psychologie*, *44*, 589-615.
- Myrtek, M. (1998). *Gesunde Kranke – kranke Gesunde*. Bern: Huber.
- Myrtek, M., Foerster, F. & Brügger, G. (2001). *Freiburger Monitoring System (FMS). Ein Daten-Aufnahme und Auswertungs-System für Untersuchungen im Alltag — Emotionale Beanspruchung, Körperlage, Bewegung, EKG — Subjektives Befinden — Verhalten*. Frankfurt a. M.: Peter Lang.
- Pawlik, K. & Buse, L. (1982). Rechnergestützte Verhaltensregistrierung im Feld: Beschreibung und erste psychometrische Überprüfung einer neuen Erhebungsmethode. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *3*, 101-118.
- Ross, B. M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, *96*, 341-357.
- Ross, B.M. (1991). *Remembering the personal past*. New York: Oxford University Press.
- Rubin, D.C. (Ed.) (1986). *Autobiographical memory*. New York: Cambridge University Press.
- Ruoss, M. (1997). Schmerzpatienten zeigen einen höheren Hindsight Bias. *Zeitschrift für Experimentelle Psychologie*, *44*, 561-588.
- Scheibehenne, B., Saller, T., Riemann, D. & Fahrenberg, J. (2000). *Befinden im Tagesablauf: Zwei Untersuchungen mit MONITOR. Forschungsbericht Nr. 147*. Albert-Ludwigs-Universität Freiburg i. Br.: Psychologisches Institut, Forschungsgruppe Psychophysiologie.
- Schwarz, N. (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. S. Clark (Eds.). *Research methods in personality and social psychology* (pp. 98-119). Newbury Park, CA : Sage.
- Shiffman, S., Hufford, M., Hickcox, M., Paty, J. A., Gnys, M. & Kassel, J. D. (1997). Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of Consulting and Clinical Psychology*, *65*, 292-300.
- Smith, W. B. & Safer, M. A. (1993). Effects of present pain level on recall of chronic pain and medication use. *Pain*, *55*, 355-361.
- Stone, A. A., Kennedy-Moore, E. & Neale, J. M. (1995). Association between daily coping and end-of-day mood. *Health Psychology*, *14*, 341-349.
- Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M. & Paty, J. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology*, *74*, 1670-1680.
- Stone, A. A., Smyth, J. M., Pickering, T. & Schwartz, J. (1996). Daily mood variability: Form of diurnal patterns and determinants of diurnal patterns. *Journal of Applied Social Psychology*, *26*, 1286-1305.
- Stone, A. A. & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, *16*, 199-202.

- Suh, E., Diener, E. & Fujita, F. (1996). Events and subjective well-being: Only recent events matter. *Journal of Personality and Social Psychology*, 70, 1091-1102.
- Suls, J. & Martin, R. E. (1993). Daily recording and ambulatory monitoring methodologies in behavioral medicine. *Annals of Behavioral Medicine*, 15, 3-7.
- Taylor, C. B., Fried, L. & Kenardy, J. (1990). The use of a real-time computer diary for data acquisition and processing. *Behaviour Research and Therapy*, 28, 93-97.
- Teasdale, J. D. & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant und unpleasant events from episodic memory. *Journal of Abnormal Psychology*, 88, 248-257.
- Thomas, D. L. & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology*, 59, 291-297.
- Werth, L. (1998). *Ein inferentieller Erklärungsansatz des Rückschaufehlers. Der Rückschaufehler: Ein Effekt sowohl zu hoher als auch zu geringer Unsicherheit*. Hamburg: Verlag Dr. Kovac.
- Wheeler, L. & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality*, 59, 339-354.

Anhang Instruktionen

Liebe Teilnehmerin, lieber Teilnehmer,

Vielen Dank für Dein Interesse an unserer Studie.

Wir wollen untersuchen, welchen Schwankungen das Befinden im Tagesverlauf unterliegt.

Wir werden Dir einen Taschencomputer anvertrauen, der Dir heute und morgen zu bestimmten Zeitpunkten Fragen nach Deinem Befinden stellen wird.

Diesen Computer solltest Du in diesem Zeitraum ständig bei Dir tragen und gut auf ihn aufpassen.

Zunächst möchten wir Dir die einzelnen Funktionen erklären und einen kleinen Probelauf durchführen.

Sollten sich dabei Fragen ergeben, stelle sie bitte ohne Zögern !!!

Abschließend müssen wir einen Termin vereinbaren, zu dem Du uns den Computer wieder zurückbringst.

Du kannst den Versuch jederzeit ohne Angabe von Gründen und ohne Nachteile für Dich abbrechen. Deine Daten werden streng vertraulich behandelt und anonym bearbeitet.

?



Alles klar ?

Training (Situationen)

Situation 1:

Stelle Dir bitte folgende Situation vor:

Du befindest Dich in diesem Augenblick bei Edeka in der Warteschlange an der Kasse. Du hast Deinen Arbeitstag hinter Dich gebracht und freust Dich nun auf das Abendessen, dessen Zutaten vor Dir in Deinem Einkaufswagen liegen.

Wenn Du soweit bist, dass Du Dir die Situation vorstellen kannst, dann gehe jetzt bitte noch einmal die vorliegende Fragenliste durch und versuche, Deine momentane Befindlichkeit möglichst genau zu beschreiben. Markiere bitte bei jeder Frage die entsprechende Stelle auf der Skala.

Situation 2:

Stelle Dir bitte folgende Situation vor:

Du bist auf dem Weg zur Uni. Du hast gerade gefrühstückt und gehst gleich in Deine erste Vorlesung für den heutigen Tag.

Wenn Du soweit bist, dass Du Dir die Situation vorstellen kannst, dann gehe jetzt bitte noch einmal die vorliegende Fragenliste durch und versuche, Deine momentane Befindlichkeit möglichst genau zu beschreiben. Markiere bitte bei jeder Frage die entsprechende Stelle auf der Skala.

Situation 3:

Stelle Dir bitte folgende Situation vor:

Du sitzt in der Straßenbahn und möchtest bloß zwei Stationen fahren. Du hast kein Kleingeld für den Automaten und hast keine Möglichkeit, in der Bahn eine Fahrkarte zu kaufen. Da Du in Eile bist, riskierst Du es, schwarz zu fahren.

Wenn Du soweit bist, dass Du Dir die Situation vorstellen kannst, dann gehe jetzt bitte noch einmal die vorliegende Fragenliste durch und versuche, Deine momentane Befindlichkeit möglichst genau zu beschreiben. Markiere bitte bei jeder Frage die entsprechende Stelle auf der Skala.

Anhang Itemlisten und Tabellen

Dieser Anhang enthält die Itemlisten und ausgewählte Ergebnisse von Untersuchungen mit MONITOR zu verschiedenen Fragestellungen innerhalb der Untersuchung der Ereignisse, Befindlichkeit und Aufmerksamkeit unter Alltagsbedingungen. Die Befunde zum negativen Retrospektionseffekt wurden bereits im Kapitel 2 und 4 berichtet. Die folgende Dokumentation gibt darüber hinaus statistische Ergebnisse, die zur Beurteilung der Variabilität der Itemwerte und zum Vergleich mit ähnlichen Studien dienen können. Eine Übersicht über diese Untersuchungen gibt die Tabelle 1.1 im Kapitel 2.

Sind Sie gerade erschöpft?	13
gar nicht 1 2 3 4 5 6 7 völlig	
Fühlen Sie sich gerade wohl?	14
gar nicht 1 2 3 4 5 6 7 völlig	
Ist die momentane Situation ungewohnt für Sie?	15
gar nicht 1 2 3 4 5 6 7 völlig	
D a n k e !	

Tabelle 2: Retrospektionseffekt bei Patienten mit Hypertonie(N = 50)

Item	Tages-Mittelwert	Retrospektion	p
Körperlich angespannt	1.97	2.76	< .01
Guter Laune	5.12	4.79	< .01
Aufgeregt	1.76	2.15	< .01
Bedrückt	1.89	1.17	< .01
Ärger	1.38	1.75	< .01
Erschöpft	1.82	2.28	< .01
Wohl fühlend	5.13	4.76	< .01
Gewohnte Situation	4.88	4.08	< .01

Anmerkung: Aggregierte Daten über drei Untersuchungstage

Dritte Untersuchung (Fahrenberg, Brügger, Foerster & Käppler, 1999; Käppler, Brügger & Fahrenberg, 2001) „Zwei-Tage-Studie“.

Ambulante Datenerhebung mittels Pocket-Computer und MONITOR

Assessment-Variablen

Erhebungstage (aufeinanderfolgend):		2
Erhebungszeitpunkte:		5/Tag
durch Piepser getriggert:		um 8:00, 12:00, 15:00, 18:00, 21:00 Uhr
Retrospektive Einstufungen:	1. Abendlicher Rückblick	um 21:00 Uhr
	2. Morgendlicher Rückblick	um 8:00 Uhr

Setting-Variablen

		Kategorien
1	Wo befinden Sie sich gerade?	13
2	Welche Personen sind im Moment anwesend?	6
3	Welche Position/Lage nimmt Ihr Körper ein?	4
4	Was haben Sie gerade vor dem Piepsen getan?	22

Zustands-Variablen

		Ratings
5	Ist die momentane Situation für Sie gewohnt, typisch?	1-7
6	Ist die momentane Situation für Sie anstrengend, belastend?	1-7
7	Glauben Sie die momentane Situation "im Griff" zu haben?	1-7
8	Fühlen Sie sich momentan eher aufgeregt, nervös	1-7
9	Ist Ihre augenbl. Stimmung eher ärgerlich, gereizt	1-7
10	Ist Ihre augenblickliche Stimmung eher bedrückt	1-7
11	Fühlen Sie sich momentan geistig angespannt	1-7
12	Fühlen Sie sich momentan eher erschöpft, müde	1-7
13	Fühlen Sie sich momentan körperlich wohl	1-7

Bitte geben Sie mögliche Beschwerden in Stichworten ein:

Zusammenfassende Einstufungen

14	Gab es seit der letzten Eingabe besondere Ereignisse?	
----	---	--

Bitte beschreiben Sie das Ereignis in Stichworten:

15	Wie stark erlebten Sie seit der letzten Eingabe Stress?	1-7
----	---	-----

Tests

16	Wie gut wird Ihre Leistung in den Tests schätzungsweise sein? Konzentration (GO/NOGO) Arbeitsgedächtnis (AG)	1-7
17	Wie gut war Ihre Leistung in den Tests schätzungsweise?	1-7

Methodenbewertung

18	Wie störend war insgesamt die Programm-Bearbeitung?	1-7
----	---	-----

Rückblick 1 + 3: Retrospektive Einstufungen am selben Abend

5	War der heutige Tag für Sie gewohnt, typisch?	1-7
6	War der heutige Tag für Sie anstrengend, belastend? etc.	1-7
18	Wie störend war heute insgesamt die Programm-Bearbeitung?	1-7

Rückblick 2 + 4: Retrospektive Einstufungen am nächsten Morgen

5	War der gestrige Tag für Sie gewohnt, typisch?	1-7
6	War der gestrige Tag für Sie anstrengend, belastend? etc.	1-7
18	Wie störend war gestern insgesamt die Programm-Bearbeitung?	1-7

Tabelle 1: Mittelwerte der aktuellen Einstufungen am Tag 1, retrospektive Einstufung am Abend und Retrospektionseffekt (Differenz).

Item	Einstufung	Mittelwert	SD	t – T e s t		Wilcoxon
				t	p	p
Situation typisch	Mittelwert	4.85	0.94			
	Retrospektion	4.43	1.76			
	Differenz	+0.42	1.49	2.13	.037	.074
Situation belastend	Mittelwert	2.84	0.95			
	Retrospektion	4.31	1.65			
	Differenz	-1.47	1.43	7.82	.000	.000
Situation "im Griff"	Mittelwert	5.82	0.73			
	Retrospektion	4.95	1.40			
	Differenz	+0.87	1.13	5.87	.000	.000
aufgeregt, nervös	Mittelwert	2.50	1.01			
	Retrospektion	3.07	1.82			
	Differenz	-0.57	1.46	3.00	.004	.019
ärgerlich, gereizt	Mittelwert	1.88	.84			
	Retrospektion	2.45	1.48			
	Differenz	-0.56	1.04	4.14	.000	.000
bedrückt	Mittelwert	1.98	0.78			
	Retrospektion	2.57	1.61			
	Differenz	-0.59	1.24	3.64	.001	.002
geistig angespannt	Mittelwert	3.28	1.06			
	Retrospektion	4.26	1.72			
	Differenz	-0.98	1.29	5.77	.001	.000
erschöpft, müde	Mittelwert	3.28	0.99			
	Retrospektion	3.93	1.61			
	Differenz	-0.65	1.24	4.03	.000	.000
körperlich wohl	Mittelwert	5.25	1.12			
	Retrospektion	5.28	1.36			
	Differenz	-0.02	0.85	0.22	.830	.632
Stress (vergängerer Zeitraum)	Mittelwert	2.72	1.15			
	Retrospektion	3.67	1.76			
	Differenz	-0.95	1.10	6.57	.000	.000

Anmerkung: Die Mittelwerte beziehen sich auf die fünf aktuellen Einstufungen (von 1 = überhaupt nicht bis 7 = vollständig) des ersten Untersuchungstags (N = 58 wegen einiger fehlender Daten).

Tabelle 2: Korrelationen zwischen den Mittelwerten der aktuellen Einstufungen und den Retrospektioneffekten mit den Persönlichkeitsdimensionen Extraversion und Emotionalität (FPI-R)

Variable (Aggregat)	Persönlichkeitsdimensionen		
	Extraversion r	Emotionalität r	
Negatives Befinden (2 Items)			
Mittelwert	Tag 1	.06	.27*
	Tag 2	.22	.19
	Tag 1 und 2	.16	.26*
Retrospektioneffekt	Tag 1	.05	.35**
	Tag 2	.01	.20
	Tag 1 und 2	.05	.38**
Negatives Befinden (7 Items)			
Mittelwert	Tag 1	.18	.24
	Tag 2	.31*	.17
	Tag 1 und 2	.28*	.22
Retrospektioneffekt	Tag 1	.11	.28*
	Tag 2	.16	.19
	Tag 1 und 2	.16	.29*
Positives Befinden (2 Items)			
Mittelwert	Tag 1	-.18	-.38**
	Tag 2	-.41***	-.32*
	Tag 1 und 2	-.33**	-.38**
Retrospektioneffekt	Tag 1	.08	.35**
	Tag 2	.18	.20
	Tag 1 und 2	.14	.33**

Anmerkung: Der Retrospektioneffekt wurde als Differenz zwischen dem Mittelwert der aktuellen Einstufungen und dem abendlichen Rückblick berechnet. Das valide N lag hier zwischen 56 und 58. * $p < .05$ ** $p < .01$ *** $p < .001$

Tabelle 3: Methodenbedingte Reaktivität und Belastung aus Sicht der Untersuchungsteilnehmer (N=61)

Items der Nachbefragung	1 gar nicht	2	3	4	5	6	7 völ lig	M	SD
Kam es durch die Untersuchungsmethodik zu Änderungen Ihres Verhaltens?	25	18	4	6	3	4	1	2.34	1.64
Haben Sie vermehrt auf Ihren psychischen Zustand geachtet?	22	15	5	5	9	4	1	2.67	1.77
Reagiert Ihre Umwelt negativ auf Ihre Ausrüstung?	29	19	5	4	1	1	-	1.85	1.13
Waren Ihnen die Reaktionen Ihrer Umwelt unangenehm?	37	11	4	3	4	1	1	1.82	1.32
Wie störend war im allgemeinen die Programmbearbeitung?	2	18	22	9	6	4	-	3.18	1.25
Wurde die Bearbeitung zu häufig verlangt?	28	9	8	4	7	3	1	2.43	1.73
Wie lästig waren Ihnen die Selbsteinstufungen?	17	23	9	7	3	2	-	2.38	1.32
Waren es zu viele Einstufungen pro Eingabe?	28	14	4	2	8	2	2	2.37	1.78
Wie lästig war Ihnen der Konzentrationstest?	28	5	8	6	-	2	1	2.08	1.41
Wie lästig war Ihnen der Gedächtnistest?	9	13	7	14	11	4	3	3.48	1.72

Table 4: Covariation between subjective state and performance (N = 61, 11 protocols in 48 hrs).

		1	2	3	4	5	6	7	8	9	10	11	12
1	Strenuous, demanding	–											
2	Situation under control	-45	–										
3	Excited, nervous	43	-38	–									
4	Angry, irritated	39	-38	44	–								
5	Depressed	27	-32	29	42	–							
6	Mentally alert	49	-32	41	27	28	–						
7	Weary, exhausted					21		–					
8	Physically well-being		20	-22	-21	-29	-25	-32	–				
9	Stress	32	-26	37	36	25	32		-21	–			
10	Anticipated performance	-30	30	-32	-29	-22		-29	20	-31	–		
11	Rated performance										29	–	
12	Testing tiresome	26	-24	21							-36	-29	–
Go/No-Go													
	Reaction time M							20					
	Reaction time SD										-20	-24	
Working memory													
	Number of correct responses											40	-21
	Performance index											46	-22
	Number of false responses											-27	
	False responses M											-20	

Note: Within-subject correlation coefficients $r \geq .20$ (pooled across 61 subjects).

Tabelle 5: Variabilität von Befinden und Aufmerksamkeit: Effekte von Tagen (2), Tageszeiten (5), Tagen und Tageszeiten, Retest-Reliabilitäten und Trends (11 Eingaben)

Variable	Tage		Tageszeiten		Tage x Tageszeiten	Retest-Reliab.	Trends
	F	Verlauf	F	Verlauf	F	r	r
Dauer der Eingabe	41.44 ***	Abnahme	50.89 ***	U-förmig	8.45 ***	.63 ***	-.26
Situation gewohnt	20.09 ***	Zunahme	11.29 ***	initialer Anstieg	23.49 ***	.34 **	.45
Situation anstrengend	4.49 *	Abnahme	5.30 ***	∩-förmig	0.46	.55 ***	-.38
Situation "im Griff"	0.13		3.05 *	U-förmig	1.10	.44 ***	-.21
Aufgeregt	11.51 ***	Abnahme	0.91		2.64 *	.52 ***	-.70 *
Ärgerlich	0.54		3.56 **	∩-förmig	2.84 *	.58 ***	.15
Bedrückt	0.71		0.91		0.58	.57 ***	-.15
Geistig angespannt	14.26 ***	Abnahme	5.79 ***	∩-förmig	2.98 *	.63 ***	-.55
Erschöpft	2.27		3.00 *	eher ansteigend	1.69	.47 ***	.11
Körperlich wohl	10.43 **	Zunahme	0.60		1.21	.67 ***	.71 *
Stress seit letzter Eingabe	4.24 *	Abnahme	12.56 ***	∩-förmig	1.46	.63 ***	-.04
Leistung antizipiert	1.43		0.73		2.61 *	.71 ***	-.01
Leistung eingeschätzt	19.30 ***	Zunahme	5.94 ***	ansteigend	3.08 *	.67 ***	.81 **
Abfragen störend	5.13 *	Zunahme	2.07		3.50 **	.56 ***	.74 **
Go/No-Go							
Mittlere Reaktionszeit	8.47 **	Abnahme	15.85 ***	U-förmig	3.38 *	.88 ***	-.38
SD der Reaktionszeit	0.36		7.48 ***	U-förmig	6.21 ***	.55 ***	-.25
Arbeitsgedächtnis							
Mittlere Reaktionszeit	5.60 *	Abnahme	5.20 ***	U-förmig	1.36	.90 ***	-.55
SD der Reaktionszeit	10.50 **	Abnahme	3.16 *	absteigend	1.36	.67 ***	-.76 **
Anzahl richtiger Reaktionen	9.93 **	Zunahme	1.75		3.22 *	.69 ***	.56

Anmerkungen: Die Freiheitsgrade (ohne missing data) betragen für Tage 1/56, für Tageszeiten 4/53 und für die Interaktion 4/53 im Flatness-Test der MANOVA. Die Werte der Tageszeiten sind gemittelt über die beiden Tage. Der Verlauf wurde durch Inspektion der Mittelwertverläufe festgestellt. Die Retest-Reliabilitäten wurden als Korrelation der Mittelwerte des ersten und des zweiten Tages (df=59) und die Trends wurden als Korrelation zwischen dem Mittelwert einer Variable und der laufenden Nummer der Eingabe (df=9) bestimmt, * p ≤ .050, ** p ≤ .010, *** p ≤ .001.

Vierte Untersuchung (Scheibehenne, Saller & Fahrenberg, 2000) Berufstätige

Tabelle 1: Liste der Fragen/Items

Befragung zu den fünf Tageszeitpunkten:	
Wo befinden sie sich gerade?	zu Hause am Arbeitsplatz Unterwegs andere Orte
Welche Personen sind im Moment anwesend?	Allein eine vertraute Person eine nicht vertraute Person mehrere vertraute Personen mehrere nicht vertraute Personen eine größere Gruppe
Was haben Sie gerade vor dem Piepsen getan?	
(Arbeit)	körperliche Arbeit Schreibtischarbeit/ PC/ Telephon Besprechung Pause/ Essen/ WC Unterwegs Hausarbeit Sonstige Arbeiten
(Freizeit)	An, Auskleiden/ Körperpflege/ WC Essen, Trinken Gespräch/ Telefonieren Fahren (Auto, Bus, Fahrrad, Bahn) Einkaufen/ Erledigung Haushalt/ Kochen Familie Lesen Fernsehen/ Radio/ Musik hören Entspannen/ Dösen/ Schlafen Spaziergehen Veranstaltung Sport/ Hobby Geselligkeit Sonstige Freizeit
Ist die momentane Situation für Sie gewohnt, typisch?	(1-7)
Fühlen Sie sich momentan aktiv, leistungsfähig?	(1-7)
Ist die momentane Situation für Sie anstrengend, belastend?	(1-7)
Glauben Sie, die momentane Situation "im Griff" zu haben?	(1-7)
Fühlen Sie sich momentan eher aufgeregt, nervös?	(1-7)
Ist Ihre augenblickliche Stimmung eher ärgerlich, gereizt?	(1-7)
Ist Ihre augenblickliche Stimmung eher bedrückt?	(1-7)
Fühlen Sie sich momentan geistig angespannt?	(1-7)
Fühlen Sie sich momentan eher erschöpft, müde?	(1-7)
Fühlen Sie sich momentan körperlich wohl?	(1-7)
Fühlen Sie sich momentan eher ausgeglichen?	(1-7)
Fühlen Sie sich momentan selbstsicher, selbstbewußt?	(1-7)
Sind Sie momentan eher guter Stimmung?	(1-7)
Was hat sich seit der letzten Eingabe ereignet?	(1-7)
Wie stark erlebten Sie seit der letzten Angabe Streß?	(1-7)
Wurden Sie beim Beantworten der Fragen abgelenkt oder gestört?	(1-7)

Rückblick:

War der heutige Tag für Sie gewohnt, typisch?	(1-7)
Fühlen Sie sich heute aktiv, leistungsfähig?	(1-7)
War der heutige Tag für Sie anstrengend, belastend?	(1-7)
Glauben Sie, daß Sie den heutigen Tag "im Griff" hatten?	(1-7)
Fühlten Sie sich heute eher aufgeregt, nervös?	(1-7)
War Ihre heutige Stimmung eher ärgerlich, gereizt?	(1-7)
War Ihre heutige Stimmung eher bedrückt?	(1-7)
Fühlten Sie sich heute geistig angespannt?	(1-7)
Fühlten Sie sich heute eher erschöpft, müde?	(1-7)
Fühlten Sie sich heute körperlich wohl?	(1-7)
Fühlten Sie sich heute eher ausgeglichen?	(1-7)
Fühlten Sie sich heute selbstsicher, selbstbewußt?	(1-7)
Waren Sie heute eher in guter Stimmung?	(1-7)
Orientierte sich Ihr Rückblick auf den heutigen Tag an besonderen Tagesereignissen	(ja/nein)
Orientierte sich Ihr Rückblick auf den heutigen Tag an der momentanen Stimmung/ Befinden	(ja/nein)
Orientierte sich Ihr Rückblick auf den heutigen Tag am Durchschnitt der Antworten im Tagesverlauf	(ja/nein)
Wie gut glauben Sie, stimmt Ihr Tagesrückblick mit den Antworten im Laufe des Tages überein?	(ja/nein)
Neigen Sie dazu, die Tage im Rückblick anders einzuschätzen?	(eher gut/ eher schlecht)
Haben Sie Ihre Tagesplanung wegen der PC-Fragen geändert?	(ja/nein)
Haben Sie heute auf Ihr Verhalten mehr als sonst geachtet?	(ja/nein)
Wie störend war die Beantwortung der Fragen insgesamt?	(1-7)
Würden Sie an einer ähnlichen Untersuchung noch einmal teilnehmen?	(ja/nein)

Tabelle 2: Tageswerte der Befindlichkeit (N=59)

Item	Mean ¹	Mode ²	SD ³	Min ⁴	Max ⁵	Skew ⁶	Kurt ⁷
Abgelenkt, gestört	1.98	1	1.46	1	7	1.42	1.18
Aktiv, leistungsfähig	5.00	6	1.45	1	7	-0.44	-0.53
Anstrengend, belastend	2.43	1	1.61	1	6	0.93	-0.35
Ärgerlich, gereizt	1.68	1	1.22	1	6	1.97	3.05
Aufgeregt, nervös	2.00	1	1.41	1	7	1.41	1.00
Ausgeglichen	4.97	6	1.43	1	7	-0.61	-0.30
Bedrückt	1.81	1	1.29	1	7	1.82	2.86
Erschöpft, müde	3.37	2	1.77	1	7	0.26	-1.16
Geistig angespannt	2.74	1	1.67	1	7	0.71	-0.62
Gewohnt, typisch	5.85	7	1.63	1	7	-1.40	0.96
Gute Stimmung	5.29	6	1.42	1	7	-1.05	0.68
Körperlich wohl	5.22	6	1.43	1	7	-0.71	-0.23
Selbstsicher, selbstbewußt	5.66	6	1.22	1	7	-1.05	0.99
Situation "im Griff"	6.33	7	1.01	2	7	-2.01	4.54
Stress	2.86	2	1.60	1	7	0.61	-0.63

Anmerkung: Das Item "abgelenkt, gestört" bezieht sich auf die Eingaben am PC

¹ Mittelwert ² Modalwert ³ Standardabweichung ⁴ Minimum ⁵ Maximum ⁶ Schiefe ⁷ Exzeß

Tabelle3: Ergebnisse des Freiburger Persönlichkeitsinventars FPI-R (Rohwerte)

FPI-Skala	Mean ¹	Mode ²	SD ³	Min ⁴	Max ⁵	Skew ⁶	Kurt ⁷
Lebenszufriedenheit	7.32	9	3.30	0	12	-.66	-0.49
Soziale Orientierung	7.53	7	2.33	1	12	-.19	0.24
Leistungsorientierung	7.42	9	2.74	1	12	-.19	-0.75
Gehemmtheit	4.56	6	2.74	0	10	.19	-0.63
Erregbarkeit	5.83	7	3.23	0	12	.14	0.61
Aggressivität	4.19	3	2.58	0	10	.25	-0.86
Beanspruchung	6.07	6	3.27	0	11	.14	-1.09
Körperliche Beschwerden	2.49	2	1.82	0	7	.45	-0.47
Gesundheitssorgen	3.93	2	2.99	0	10	.55	-0.88
Offenheit	7.22	10	2.39	1	12	-.19	-0.62
Extraversion (E)	7.86	8	3.06	1	14	-.32	-0.08
Emotionalität (N)	5.39	5	3.16	0	12	.46	-0.75

¹ Mittelwert ² Modalwert ³ Standardabweichung ⁴ Minimum ⁵ Maximum ⁶ Schiefe ⁷ Exzeß

Tabelle 4: Ergebnisse des Freiburger Persönlichkeitsinventars FPI-R (Stanine-Werte)

FPI-Skala	Mean ¹	Mode ²	SD ³	Min ⁴	Max ⁵	Skew ⁶	Kurt ⁷
Lebenszufriedenheit	5.37	6	2.03	1	9	-.22	-.58
Soziale Orientierung	5.80	6	1.68	1	9	-.17	.23
Leistungsorientierung	5.29	4	1.84	1	9	.26	-.06
Gehemmtheit	4.54	5	1.78	1	8	-.16	-.30
Erregbarkeit	5.02	4	2.09	1	9	.06	-.46
Aggressivität	4.88	5	1.77	1	8	-.30	-.34
Beanspruchung	5.00	5	1.79	1	8	-.30	-.59
Körperliche Beschwerden	4.31	4	1.59	1	8	-.15	-.10
Gesundheitssorgen	4.02	5	2.08	1	8	.17	-.99
Offenheit	5.53	6	1.61	2	9	.00	-.36
Extraversion (E)	5.19	5	1.75	1	9	-.16	.09
Emotionalität (N)	4.54	4	4.70	1	9	.39	-.10

¹ Mittelwert ² Modalwert ³ Standardabweichung ⁴ Minimum ⁵ Maximum ⁶ Schiefe ⁷ Exzeß

Tabelle 5: Korrelation der aggregierten Tageswerte mit den Stanine-Werten des FPI (N=59)

Tag/FPI	LEB	SOZ	LEI	GEH	ERR	AGGRr	BEAN	KÖRP	GES	OFF	EXTR	EMOT
aktiv	.38 (.003)	-.05 (.699)	.32 (.012)	-.20 (.125)	-.16 (.214)	.16 (.221)	-.18 (.185)	-.26 (.050)	.02 (.912)	-.08 (.539)	.24 (.070)	-.37 (.005)
ange- spannt	-.15 (.255)	.04 (.749)	-.20 (.887)	.19 (.154)	.32 (.013)	-.13 (.321)	.24 (.064)	.14 (.277)	.23 (.080)	.01 (.917)	-.01 (.915)	.34 (.008)
ausge- glichen	.51 (.000)	.08 (.554)	.17 (.211)	-.30 (.021)	-.18 (.173)	.09 (.518)	-.39 (.002)	-.30 (.023)	-.29 (.024)	.20 (.131)	.13 (.318)	-.46 (.000)
bedrückt	-.49 (.000)	.06 (.655)	-.22 (.093)	.08 (.534)	.27 (.036)	-.03 (.808)	.29 (.028)	.14 (.305)	.18 (.175)	.02 (.907)	.04 (.763)	.50 (.000)
belastend	-.24 (.065)	.02 (.912)	-.04 (.775)	.29 (.025)	.23 (.079)	-.12 (.361)	.20 (.121)	.19 (.152)	.34 (.009)	.03 (.850)	-.06 (.670)	.35 (.008)
gereizt	-.27 (.043)	-.08 (.548)	-.04 (.738)	.13 (.345)	.20 (.129)	.01 (.930)	.16 (.215)	.28 (.032)	.32 (.014)	.00 (.960)	-.09 (.495)	.38 (.003)
Sit. ge- wohnt	-.14 (.287)	.05 (.706)	-.26 (.044)	.21 (.104)	.26 (.043)	.16 (.241)	-.04 (.766)	-.05 (.709)	-.40 (.002)	.27 (.037)	-.12 (.358)	.15 (.256)
Sit. im Griff	.20 (.132)	-.03 (.808)	.04 (.775)	-.11 (.389)	.01 (.940)	.26 (.044)	-.18 (.181)	-.21 (.120)	-.43 (.001)	.16 (.225)	.12 (.373)	-.29 (.027)
körperlich wohl	.19 (.146)	-.10 (.453)	.15 (.267)	-.10 (.455)	-.04 (.763)	.07 (.610)	-.33 (.010)	-.28 (.034)	-.14 (.304)	-.08 (.558)	.10 (.470)	-.22 (.099)
müde	-.33 (.011)	.23 (.075)	-.26 (.046)	.21 (.114)	.23 (.085)	-.14 (.306)	.36 (.005)	.18 (.174)	0.05 (.688)	.00 (.993)	-.17 (.201)	.40 (.002)
nervös	-.28 (.031)	-.10 (.464)	-.02 (.894)	.18 (.170)	.21 (.107)	-.05 (.691)	.37 (.004)	.21 (.114)	.56 (.000)	-.22 (.100)	-.06 (.681)	.39 (.002)
selbst- sicher	.43 (.001)	.02 (.901)	.36 (.005)	-.39 (.002)	-.29 (.026)	.17 (.198)	-.12 (.372)	-.30 (.022)	-.14 (.275)	-.10 (.461)	.30 (.022)	-.30 (.023)
gute Stimmung	.51 (.000)	.09 (.488)	.21 (.120)	-.24 (.067)	-.26 (.048)	.07 (.578)	-.31 (.016)	-.31 (.019)	-.22 (.093)	.09 (.519)	.14 (.276)	-.44 (.000)
störend	.19 (.151)	.12 (.362)	-.06 (.646)	.03 (.794)	-.05 (.691)	-.12 (.364)	.08 (.572)	.16 (.242)	.17 (.195)	.06 (.671)	.00 (.999)	-.05 (.737)
im Stress	-.31 (.018)	-.06 (.671)	.03 (.821)	.06 (.673)	.27 (.036)	-.04 (.782)	.52 (.000)	.39 (.002)	.40 (.002)	-.15 (.255)	.04 (.773)	.46 (.000)

Tabelle 6: Retrospektionseffekt: Vergleich zwischen aggregierten Tageswerten und Rückblick (N=59)

Item	Tag ¹	Retro ²	SD T ³	SD R ⁴	Diff. ⁵	t ⁶	p ⁷	Z ⁸	p ⁹	Corr. ¹⁰
Abgelenkt, gestört	1.99	2.48	0.83	1.28	-0.49	-3.14	.003	2.99	.003	.43
Aktiv, leistungsfähig	4.97	5.48	0.83	0.97	-0.50	-4.40	.000	3.88	.000	.54
Anstrengend, belast.	2.45	3.64	0.94	1.58	-1.19	-6.79	.000	5.20	.000	.53
Ärgerlich, gereizt	1.69	2.24	0.71	1.33	-0.55	-3.97	.000	3.71	.000	.60
Aufgeregt, nervös	2.02	2.59	0.91	1.44	-0.58	-4.03	.000	4.03	.000	.64
Ausgeglichen	4.94	5.03	1.05	1.50	-0.10	-0.68	.497	1.14	.256	.70
Bedrückt	1.82	2.22	0.89	1.46	-0.40	-3.07	.003	2.77	.006	.75
Erschöpft, müde	3.41	3.31	1.11	1.50	0.10	0.74	.464	0.80	.423	.70
Geistig angespannt	2.75	3.48	1.12	1.60	-0.73	-4.47	.000	3.86	.000	.63
Gewohnt, typisch	5.85	5.31	0.91	1.39	0.55	3.47	.001	3.19	.001	.51
Gute Stimmung	5.26	5.29	0.95	1.31	-0.03	-0.30	.764	0.73	.463	.86
Körperlich wohl	5.19	5.24	1.04	1.37	-0.05	-0.40	.692	0.69	.490	.78
Selbstsicher, selbstb	5.64	5.70	0.89	1.06	-0.05	-0.54	.592	0.47	.640	.73
Situation "im Griff	6.33	5.98	0.59	1.04	0.35	2.96	.004	2.65	.008	.50
Streß	-	-	-	-	-	-	-	-	-	-

1) Mittelwert der aggregierten Tageswerte 2) Mittelwert der Retrospektion 3) Standardabweichung der aggregierten Tageswerte 4) Standardabweichung der Retrospektion 5) Differenz aggregierte Tageswert – Retrospektionswert 6) t-Wert für abhängige Stichproben mit 7) p-Wert 8) Wilcoxon-Test Z-Wert mit 9)p-Wert 10) Korrelation zwischen Tageswert und Retrospektion

Tabelle 7: Korrelation des Retrospektionseffektes mit den Stanine-Werten des FPI (N=59)

Ret/FPI	LEB	SOZ	LEI	GEH	ERR	AGG	BEAN	KÖRPP	GES	OFF	EXTR	EMO
aktiv	-.01 (.935)	.05 (.716)	.08 (.526)	-.12 (.350)	.00 (.994)	.11 (.425)	.15 (.250)	-.05 (.701)	.06 (.671)	-.14 (.283)	.21 (.118)	.10 (.468)
ange- spannt	.10 (.442)	.08 (.538)	-.12 (.376)	-.19 (.161)	-.07 (.621)	-.04 (.793)	-.21 (.106)	-.35 (.006)	-.12 (.354)	.10 (.472)	.01 (.924)	-.20 (.139)
ausge- glichen	-.33 (.011)	-.05 (.707)	-.11 (.390)	-.13 (.323)	.15 (.261)	.06 (.682)	.19 (.161)	.08 (.538)	.03 (.813)	-.01 (.970)	.05 (.708)	.26 (.046)
bedrückt	.21 (.119)	.08 (.561)	-.02 (.882)	.04 (.792)	.06 (.635)	.06 (.632)	-.10 (.433)	-.14 (.286)	-.19 (.149)	.14 (.287)	-.01 (.942)	-.21 (.111)
belas- tend	.22 (.096)	-.07 (.590)	.03 (.839)	.04 (.788)	.01 (.967)	-.02 (.861)	-.44 (.000)	-.37 (.004)	-.06 (.642)	.25 (.059)	-.06 (.647)	-.28 (.034)
gereizt	.18 (.170)	.24 (.070)	-.24 (.066)	.06 (.640)	.17 (.206)	-.19 (.149)	-.20 (.134)	-.12 (.385)	-.21 (.117)	.18 (.175)	-.17 (.208)	-.12 (.354)
Sit. ge- wohnt	-.08 (.573)	-.03 (.839)	-.06 (.668)	.02 (.908)	-.04 (.777)	-.08 (.563)	.26 (.051)	-.01 (.963)	.16 (.230)	-.35 (.007)	.01 (.946)	.17 (.200)
Sit. im Griff	-.03 (.800)	-.08 (.571)	.07 (.581)	-.09 (.495)	.04 (.748)	.08 (.569)	.03 (.798)	.03 (.815)	-.12 (.366)	-.02 (.880)	.01 (.945)	.01 (.950)
körper- lich wohl	-.17 (.192)	-.06 (.659)	-.15 (.429)	.03 (.805)	-.08 (.527)	-.12 (.377)	.24 (.073)	.06 (.654)	.21 (.104)	-.16 (.217)	-.09 (.478)	.24 (.067)
müde	.10 (.622)	.14 (.292)	-.34 (.009)	-.15 (.269)	-.00 (.991)	-.43 (.001)	-.24 (.063)	-.28 (.030)	-.09 (.502)	-.11 (.426)	-.03 (.853)	-.24 (.072)
nervös	.08 (.548)	.08 (.561)	-.05 (.718)	-.04 (.793)	.10 (.465)	.04 (.744)	-.30 (.023)	-.27 (.042)	-.10 (.471)	.17 (.210)	-.03 (.813)	-.27 (.038)
selbst- sicher	-.06 (.670)	.02 (.908)	-.01 (.939)	-.21 (.115)	-.07 (.577)	-.11 (.398)	.09 (.477)	-.04 (.745)	-.07 (.622)	-.13 (.340)	.24 (.071)	.08 (.570)
gute Stimm.	-.24 (.064)	.06 (.648)	.12 (.375)	-.02 (.880)	-.05 (.684)	.09 (.510)	.11 (.391)	.13 (.314)	.16 (.235)	-.13 (.322)	.06 (.663)	.30 (.023)
störend	.19 (.151)	.05 (.694)	.17 (.193)	-.02 (.862)	-.03 (.842)	.11 (.405)	.04 (.739)	.13 (.322)	.04 (.772)	.06 (.680)	.23 (.084)	-.09 (.500)
im Stress	-	-	-	-	-	-	-	-	-	-	-	-

Anmerkungen: Da der Retrospektionseffekt als Differenz Tages-Mittelwert minus Rückblick berechnet wird, bedeuten höhere *negative Werte* (bei "negativen" Items) einen stärkeren Effekt. Das Item „Stress“ wurde beim Rückblick nicht gefragt.

Tabelle 8: Korrelation des Retrospektionseffektes mit dem Alter und Geschlecht (N=59)

	aktiv	An- gesp	aus- gegl	be- drück t	belas- tend	ge- reizt	ge- wohn t	im Griff	körp wohl	mü- de	ner- vös	selb- stsi	gute Stim	stör- end
Alter	-.02 (.863)	.14 (.289)	-.01 (.914)	-.13 (.313)	.08 (.558)	.08 (.528)	.06 (.661)	.10 (.443)	.12 (.381)	.17 (.205)	.09 (.511)	.00 (.977)	.16 (.234)	.00 (.983)
Ge- schl	-.38 (.003)	.01 (.934)	-.06 (.657)	.02 (.893)	.21 (.105)	.13 (.333)	-.40 (.002)	-.07 (.585)	-.08 (.553)	.21 (.103)	.11 (.391)	-.34 (.008)	.09 (.504)	.19 (.148)

Tabelle 9: Antwortverteilungen bei den Metafragen am Ende der Untersuchung

"Orientierte sich Ihr Rückblick auf den heutigen Tag am Durchschnitt der Antworten im Tagesverlauf?"

Antwort	Häufigkeit	Anteil in %
Ja	27	45.8
nein	32	54.2

"Orientierte sich Ihr Rückblick auf den heutigen Tag an der/dem momentanen Stimmung/Befinden?"

Antwort	Häufigkeit	Anteil in %
ja	18	30.5
nein	41	69.5

"Orientierte sich Ihr Rückblick auf den heutigen Tag an besonderen Tagesereignissen?"

Antwort	Häufigkeit	Anteil in %
ja	33	55.9
nein	26	44.1

"Wie gut, glauben Sie, stimmt Ihr Tagesrückblick mit den Antworten im Laufe des Tages überein?"

Antwort	Häufigkeit	Anteil in %
eher gut	56	94.9
eher schlecht	3	5.1

"Haben Sie Ihre Tagesplanung wegen der PC-Fragen geändert?"

Antwort	Häufigkeit	Anteil in %
ja	2	3.4
nein	57	96.6

"Neigen Sie dazu, die Tage im Rückblick anders einzuschätzen?"

Antwort	Häufigkeit	Anteil in %
ja	23	39
nein	36	61

Haben Sie heute auf Ihr Verhalten mehr als sonst geachtet?"

Antwort	Häufigkeit	Anteil in %
ja	10	16.9
nein	49	83.1

"Würden Sie an einer ähnlichen Studie noch einmal teilnehmen?"

Antwort	Häufigkeit	Anteil in %
ja	59	100
Nein	0	0

Tabelle 10: Zusammenhang des Retrospektionseffektes mit der Neigung, den Tag rückblickend anders einzuschätzen

Item	anders ¹ (n=23)	gleich ² (n=36)	F ³	p ⁴	Levene- Test ⁵
Abgelenkt, gestört	-0.65	-0.38	0.77	.384	n.s.
Aktiv, leistungsfähig	-0.38	-0.58	0.77	.384	<.05
Anstrengend, belastend	-1.32	-1.10	0.37	.547	n.s.
Ärgerlich, gereizt	-0.90	-0.33	4.19	.045	<.05
Aufgeregt, nervös	-1.08	-0.25	9.04	.004	n.s.
Ausgeglichen	0.25	-0.32	4.26	.044	<.05
Bedrückt	-0.72	-0.19	4.17	.046	n.s.
Erschöpft, müde	-0.19	0.29	2.95	.091	n.s.
Geistig angespannt	-1.29	-0.37	8.64	.005	n.s.
Gewohnt, typisch	0.87	0.34	2.73	.104	n.s.
Gute Stimmung	-0.10	-0.03	0.44	.509	n.s.
Körperlich wohl	-0.37	-0.50	0.00	.956	n.s.
Selbstsicher, selbstbewußt	-0.05	-0.05	0.00	.994	n.s.
Situation "im Griff	0.50	0.26	0.97	.329	n.s.
Streß	-	-	-	-	-

¹ Differenz aggregierte Tageswerte – Retrospektionswerte für diejenigen Probanden, die auf die Frage: "Neigen Sie dazu, den Tag im Rückblick anders einzuschätzen" mit "Ja" geantwortet haben ² Differenz aggregierte Tageswerte – Retrospektionswerte für diejenigen Probanden, die auf die Frage: "Neigen Sie dazu, den Tag im Rückblick anders einzuschätzen" mit "Nein" geantwortet haben ³ F-Wert ⁴ p-Wert ⁵ p-Wert des Levene Tests auf Homogenität der Varianzen (zweiseitig)

Tabelle 11: Zusammenhang des Retrospektionseffektes mit der berichteten Orientierung am durchschnittlichen Tageswert

Item	Orientierung am Durchschnitt ¹ (n=27)	keine Orientie- rung am Durch- schnitt ² (n=32)	F ³	p ⁴	Levene- Test ⁵
Abgelenkt, gestört	-0.28	-0.66	1.56	.216	n.s.
Aktiv, leistungsfähig	-0.39	-0.59	0.77	.385	n.s.
Anstrengend, belastend	-0.95	-1.40	1.62	.208	n.s.
Ärgerlich, gereizt	-0.36	-0.71	1.55	.218	n.s.
Aufgeregt, nervös	-0.34	-0.78	2.46	.122	n.s.
Ausgeglichen	-0.15	-0.05	0.15	.701	n.s.
Bedrückt	-0.52	0.30	0.69	.410	n.s.
Erschöpft, müde	0.29	-0.05	1.49	.228	n.s.
Geistig angespannt	-0.60	-0.83	0.50	.482	n.s.
Gewohnt, typisch	0.34	0.72	1.42	.238	n.s.
Gute Stimmung	-0.12	0.05	0.87	.355	n.s.
Körperlich wohl	-0.12	0.02	0.36	.553	<.05
Selbstsicher, selbstbewußt	0.14	-0.21	3.49	.067	n.s.
Situation "im Griff	0.31	0.38	0.09	.765	n.s.
Streß	-	-	-	-	-

¹ Differenz aggregierte Tageswerte - Retrospektionswerte für diejenigen Probanden, die auf die Frage: "Orientierte sich Ihr Rückblick auf den heutigen Tag am Durchschnitt der Tagesantworten" mit "Ja" geantwortet haben ² Differenz aggregierte Tageswerte - Retrospektionswerte für diejenigen Probanden, die auf die Frage: "Orientierte sich Ihr Rückblick auf den heutigen Tag am Durchschnitt der Tagesantworten" mit "Nein" geantwortet haben ³ F-Wert ⁴ p-Wert ⁵ p-Wert des Levene Tests auf Homogenität der Varianzen (zweiseitig).

Sechste Untersuchung Wochenstudie (Fahrenberg, Leonhart & Foerster, 2002; Leonhart, 2001) Wochenstudie (Alltag von Studierenden der Psychologie)

Ablauf der Wochenstudie mit Fragen und möglichen Antworten

Bitte benutzen Sie zur Beantwortung der Fragen die obere Tastenreihe (Zahlen). Sie können die Eingabe korrigieren. Schließen Sie jede Eingabe mit 'Enter' ab.

Falls Sie die Eingabe beendet haben, erscheint gleich die nächste Frage. Wollen Sie die vorhergehende Frage doch anders beantworten, dann drücken Sie die Taste mit Doppelpfeil, und diese Frage erscheint wieder.

Wo befinden Sie sich gerade?

eigene Wohnung

fremde Wohnung

unterwegs

Hörsaal (UNI)

Seminarraum, Labor (UNI)

Bibliothek (UNI)

Mensa/Cafeteria (UNI)

am Arbeitsplatz (JOB)

Einkaufsgeschäft

Kneipe/Restaurant

Öffentl. Veranstaltungsraum

im Freien/ im Grünen

Sonstiges

Welche Personen sind im Moment anwesend?

allein

eine vertraute Person

eine n i c h t vertraute Person

mehrere vertraute Personen

mehrere n i c h t vertraute Personen

eine größere Gruppe

Was haben Sie gerade getan? (vor dem Piepsen)

09 2 A l l g e m e i n

03 Schlafen

04 Entspannen, Dösen

05 Körperpflege, An-, Auskleiden

06 Essen und Trinken

07 Gehen

08 Fahrradfahren

09 Fahren (Auto, Bus, Bahn)

10 Einkaufen

11 Haushalt

05 1 A r b e i t

12 Schreibtischarbeit

13 Lehrveranstaltung

14 Organisation

15 Nebenjob

16 körperliche Arbeit

07 2 F r e i z e i t

17 Lesen

18 Fernsehen/Radio

19 Gespräch

20 Geselligkeit

21 Veranstaltung

22 Sport

23 Hobby

01 1 S o n s t i g e s

24 Sonstiges

Adjektivskalen siebenstufig

Fühlen Sie sich momentan aktiv, leistungsfähig?

Ist die momentane Situation für Sie anstrengend, belastend?

Fühlen Sie sich momentan eher aufgeregt, nervös?

Ist Ihre momentane Stimmung eher ärgerlich, gereizt?

Ist Ihre momentane Stimmung eher bedrückt?

Fühlen Sie sich momentan erschöpft, müde?

Sind Sie momentan eher guter Stimmung?

Fühlen Sie sich momentankörperlich wohl?

bei Eingabe kleiner 7

Bitte mögliche Beschwerden in Stichworten eingeben!

Bitte als letzte Eingabe nur Enter drücken.

Gab es seit der letzten Eingabe besondere Ereignisse?

Ja - Nein

bei Ja-Antwort Bitte Ereignisse in Stichworten eingeben!

Bitte als letzte Eingabe nur Enter drücken.

Wie stark erlebten Sie seit der letzten Eingabe Stress?

Gab es seit der letzten Eingabe positive Veränderungen?

Wie *g u t* wird Ihre Leistung in dem Test schätzungsweise sein?

GO/NOGO-TEST

Wie *g u t* war Ihre Leistung in dem Test schätzungsweise?

Wurden Sie bei diesen Fragen und dem Testabgelenkt oder gestört?

nach 20 Uhr, jeden Tag

Geben Sie bitte rückblickend eine Gesamteinschätzung des heutigen Tages, indem Sie sich den Tageslauf noch einmal vergegenwärtigen.

War der heutige Tag für Sie gewohnt, typisch?

Fühlten Sie sich heute aktiv, leistungsfähig?

War der heutige Tag für Sie anstrengend, belastend?

Waren Sie am heutigen Tag eher aufgeregt, nervös?

Waren Sie am heutigen Tag eher ärgerlich, gereizt?

Waren Sie am heutigen Tag eher bedrückt?

Waren Sie am heutigen Tag erschöpft, müde?

Haben Sie sich am heutigen Tag körperlich wohl gefühlt?

Waren Sie heute eher in guter Stimmung?

Wie störend war insgesamt die Programm-Bearbeitung am heutigen Tag?

Wie stark haben Sie am heutigen Tag Stress erlebt?

Gab es am heutigen Tag positive Veränderungen?

am Ende der Woche nach 20 Uhr

Es folgt nun noch ein Rückblick über die ganze letzte Woche. Bitte beantworten Sie die Fragen im Bezug auf diese Zeit.

Fühlten Sie sich in der Woche aktiv, leistungsfähig?

War die Woche für Sie anstrengend, belastend?

Fühlten Sie sich in der Woche eher aufgeregt, nervös?

War Ihre Stimmung in der Woche eher ärgerlich, gereizt?

War Ihre Stimmung in der Woche eher bedrückt?

Fühlten Sie sich in der Woche erschöpft, müde?

Waren Sie in der Woche in eher guter Stimmung?

Fühlten Sie sich in der Woche körperlich wohl?

Wie stark erlebten Sie in der Woche Stress?

Gab es in der Woche positive Veränderungen?

Haben Sie Ihre Tagesplanung wegen der PC-Fragen verändert?

Haben Sie in dieser Woche auf Ihr Verhalten und Ihr Befinden mehr geachtet?

Wie störend waren Fragen und Tests insgesamt?

Würden Sie an einer ähnlichen Untersuchung noch einmal teilnehmen?

Ja-Nein

bei Nein Warum nicht?

Bitte als letzte Eingabe nur Enter drücken.

Vielen Dank für Ihre Mitarbeit !

Gerät jetzt bitte schließen und

wieder in die Tasche stecken.

Fragebogen

Rückblick

Versuchen Sie bitte, sich an die einzelnen Tage der Woche zu erinnern und geben Sie bitte Ihre Einstufung mit einem Kreuz (x) an der betreffenden Stelle für jeden Tag. (Am besten ist es, die Wochentage über die Spalten zu schreiben, um sich leichter erinnern zu können.)

Wie anstrengend, belastend war dieser Tag?

	1. Tag	2. Tag	3. Tag	4. Tag	5. Tag	6. Tag	7. Tag
1 gar nicht							
2							
3							
4							
5							
6							
7 völlig							

Waren Sie an diesem Tag in eher guter Stimmung?

	1. Tag	2. Tag	3. Tag	4. Tag	5. Tag	6. Tag	7. Tag
1 gar nicht							
2							
3							
4							
5							
6							
7 völlig							

Tabelle 1: Eine Woche im Selbstbericht von Studierenden der Psychologie (33 Teilnehmerinnen x 42 Messzeitpunkte = 1386 Daten).

Befindlichkeit (Skalen 1 bis 7)

Item	M	Ganze Woche			Tag 1-5	Tag 6-7
		SD	Schiefe	Kurtosis	M	M
Aktiv, leistungsfähig	3.88	1.44	-.17	-.62	3.85	3.96
Anstrengend, belastend	2.50	1.53	.90	-.01	2.52	2.44
Aufgeregt, nervös	2.22	1.41	1.16	.62	2.28	2.05 **
Ärgerlich, gereizt	1.94	1.35	1.52	1.65	1.94	1.93
Bedrückt	2.36	1.48	.91	-.14	2.37	2.34
Erschöpft, müde	3.75	1.63	.06	-.87	3.83	3.56 **
Guter Stimmung	4.44	1.36	-.30	-.25	4.41	4.50
Körperlich wohl	4.67	1.45	-.41	-.20	4.61	4.80
Stress seit letzter Eingabe	2.61	1.55	.78	-.29	2.71	2.35 **
Testleistung antizipiert	3.68	1.19	.15	.04	3.65	3.77
Testleistung tatsächlich	3.75	1.44	.15	-.48	3.77	3.68
Positive Veränderungen	2.73	1.60	.61	-.53	2.73	2.76
Signal störend, ablenk.	3.32	1.89	.33	-1.11	3.46	2.96 **
<i>Reaktionszeit (sec):</i>						
Mittelwert richtige Reaktionen	0.49	0.06	1.02	2.17	0.49	0.49
Stand.Abw. richtige Reaktionen	0.05	0.02	1.52	5.63	0.06	0.06

Anmerkung. N = 1351 Eingaben (Tag 1-5: 970, Tag 6-7: 381). ** p < .01 (t-Test Tag 1-5 vs. Tag 6-7)

Aufenthaltsort

Kategorie	Häufigkeit	Prozent	Tage 1-5	Tage 6-7
eigene Wohnung	622	46.2	37.7	67.7
fremde Wohnung	112	8.3	5.4	15.8
unterwegs	206	15.3	19.2	1.6
Hörsaal	107	8.0	11.1	0.0
Seminarraum, Labor	88	6.5	9.1	0.0
Bibliothek	26	1.9	2.7	0.0
Mensa/ Cafeteria	24	1.8	2.4	0.3
am Arbeitsplatz	43	3.2	3.8	1.6
Einkaufsgeschäft	26	1.9	2.3	1.1
Kneipe/ Restaurant	42	3.1	2.9	3.7
öffentl. Veranstaltungsraum	10	0.7	0.5	1.3
im Freien/ im Grünen	17	1.3	1.1	1.6
Sonstiges	23	1.7	1.8	1.6
Fehlend	40			
Gesamt	1346	100.0	100.0	100.0

Sozialer Kontext

Kategorie	Häufigkeit	Prozent	Tage 1-5	Tage 6-7
allein	528	39.1	37.9	142.0
eine vertraute Person	377	27.9	24.7	35.9
eine nicht vertraute Person	17	1.3	1.3	1.1
mehrere vertraute Personen	174	12.9	11.8	15.8
mehrere nicht vertraute Personen	101	7.5	9.6	2.1
eine größere Gruppe	154	11.4	14.6	3.2
Fehlend	35			
Gesamt	1351	100.0	100.0	100.0

Tätigkeit

Kategorie	Häufigkeit	Prozent	Tage 1-5	Tage 6-7
Schlafen	53	3.9	2.7	7.1
Entspannen, dösen	58	4.3	3.7	5.8
Körperpflege, An- Auskleiden	43	3.2	2.9	3.9
Essen und Trinken	139	10.3	8.9	13.9
Gehen	95	7.1	9.1	1.8
Fahrradfahren	16	1.2	1.6	0.3
Fahren (Auto, Bus, Bahn)	86	6.4	7.5	3.7
Einkaufen	33	2.5	2.8	1.6
Haushalt	33	2.5	1.9	3.9
Schreibtischarbeit	205	15.2	12.8	21.3
Lehrveranstaltung	144	10.7	14.9	0.0
Organisation	35	2.6	3.3	0.8
Nebenjob	34	2.5	3.0	1.3
körperliche Arbeit	11	0.8	0.7	1.1
Lesen	32	2.4	1.8	3.9
Fernsehen/ Radio	79	5.9	5.2	7.6
Gespräch	116	8.6	7.8	10.8
Geselligkeit	29	2.2	1.2	4.5
Veranstaltungen	9	0.7	0.5	1.1
Sport	6	0.5	0.6	0.0
Hobby	15	1.1	1.2	0.8
Sonstiges	77	5.7	6.0	5.0
Fehlend	38			
Gesamt	1348	100.0	100.0	100.0

Tabelle 2: Varianzkomponenten der Wochenstudie. Zweifaktorielle ANOVA Personen (33), Tageszeiten (6) und Wechselwirkung Personen x Tageszeiten, Residuum

Item	Personen	Tageszeiten	WW	Residuum
Aktiv, leistungsfähig	20.6	2.6	5.3	71.5
Anstrengend, belastend	20.0	1.7	2.5	75.8
Aufgeregt, nervös	18.6	0.0	13.2	68.2
Ärgerlich, gereizt	26.6	0.0	0.6	72.8
Bedrückt	32.0	0.0	3.2	64.8
Erschöpft, müde	23.5	3.1	6.6	66.8
Guter Stimmung	22.9	0.5	4.0	72.6
Körperlich wohl	42.1	0.0	5.6	52.3
Stress seit letzter Eingabe	16.1	2.3	11.6	70.0
Positive Veränderungen	17.9	2.8	0.5	78.8
Signal störend, ablenkend	12.6	0.0	20.0	67.4
<i>Reaktionszeit (sec):</i>				
Mittelwert richtige Reaktionen	32.1	3.0	7.0	57.9
Stand.Abw. richtige Reaktionen	18.7	3.3	3.0	75.0

Anmerkung: Varianzkomponenten in Prozent. (33 Teilnehmerinnen x 7 Tage x 6 Messzeitpunkte = 1386 Daten). Tage sind nicht die Wochentage, sondern die laufenden Tage der Untersuchung.

Tabelle 3: Varianzkomponenten der Wochenstudie dreifaktorielle Analyse

Variable	Personen	Abfragen	Wochentag	Pers x Abfrage	Pers x Wochentag	Abfr x Wochentag	Residuum
Aktiv	18	2	0	5	10	1	64
Anstrengend	17	2	0	1	11	1	68
Aufgeregt	17	0	0	3	14	3	63
Ärgerlich	23	0	0	2	9	1	65
Bedrückt	28	0	0	3	15	1	53
Erschöpft	20	3	0	3	15	1	58
Gute Stimmung	19	1	0	2	19	0	59
Körperlich wohl	39	0	0	2	17	1	41
Stress seit	14	2	1	3	12	1	67
Pos. Veränd.	15	2	0	0	5	0	78
Störend	13	0	2	0	5	3	77
<i>Reaktionszeit (s):</i>							
M richtige Reakt.	29	3	0	1	14	2	51
SD richtige Reakt.	16	3	0	1	6	1	73

Anmerkung: Varianzanalyse mit drei Random-Faktoren (SAS-Makro vaps3rrr, Foerster).

Tabelle 4: Korrelationen innerhalb Personen

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Aktiv	–	-08	06	-09	-21	-.69	36	40	-05	19	14	-25	-11
2 Anstrengend	-08	–	42	37	35	14	-40	-20	38	-13	-00	-09	01
3 Aufgeregt	06	42	–	38	31	-00	-24	-09	37	-07	08	-10	01
4 Ärgerlich	-09	37	38	–	40	12	-49	-15	36	-15	05	-02	01
5 Bedrückt	-21	35	31	40	–	24	-53	-27	30	-15	-04	-02	02
6 Erschöpft	-69	14	-00	12	24	–	-33	-36	12	-15	-13	18	08
7 Gute Stimmung	36	-40	-24	-49	-53	-33	–	36	-27	25	10	-02	-02
8 Körperl. wohl	40	-20	-09	-15	-27	-36	36	–	-13	22	04	-09	-03
9 Stress	-05	38	37	36	30	12	-27	-13	–	-01	14	-09	-05
10 Pos. Veränd	19	-13	-07	-15	-15	-15	25	22	-01	–	03	01	-02
11 Störend	14	-00	08	05	-04	-13	10	04	14	03	–	-01	10
12 Reaktionszeit M	-25	-09	-10	-02	-02	18	-02	-09	-09	01	-01	–	50
13 Reaktionszeit SD	-11	01	01	01	02	08	-02	-03	-05	-02	10	50	–

Anmerkung: Korrelation über 42 Selbstberichte (Abfragen) gepoolt über 33 Personen (R_{IP} -Matrix der Kovarianzzerlegung).

Tabelle 5: Skalen des FPI-R und der FBL-R (N = 33)

Variable		N	M	SD
ALTER	Lebensalter (Jahre)	33	23.94	5.20
LEB	FPIR1 Lebenszufriedenheit	33	6.70	2.94
SOZ	FPIR2 Soziale Orientierung	33	7.85	1.77
LEI	FPIR3 Leistungsorientierung	33	6.73	2.13
GEH	FPIR4 Gehemmtheit	33	5.85	3.11
ERR	FPIR5 Erregbarkeit	33	7.00	3.02
AGGR	FPIR6 Aggressivität	33	3.45	2.37
BEAN	FPIR7 Beanspruchung	33	6.61	3.15
KOERP	FPIR8 Körperliche Beschwerden	33	3.67	2.43
GES	FPIR9 Gesundheitsorgen	33	3.24	2.53
OFF	FPIR10 Offenheit	33	7.76	2.00
EXTR	FPIRE Extraversion	33	7.27	3.13
EMOT	FPIRN Emotionalität	33	7.61	3.39
ALL	FBL Allgemeinbefinden	33	21.79	3.40
MUED	FBL Müdigkeit	33	19.06	4.77
HKR	FBL Herz-Kreislauf	33	12.39	4.09
MDA	FBL Magen-Darm	33	17.12	3.86
KHA	FBL Kopf-Hals-Reizsyndrom	33	16.70	5.24
ANSP	FBL Anspannung	33	15.33	5.24
EMO	FBL Emotionale Reaktivität	33	21.61	4.44
SCHM	FBL Schmerz	33	18.06	5.54
SEN	FBL Sensorik	33	19.39	5.96
FBL	FBL Beschwerdensumme	33	161.45	28.61

Tabelle 6: Prüfung der Ausgangswert-Abhängigkeit zwischen retrospektiven und aktuellen Eingaben (über je 6 Abfragen eines Tages gemittelt), sowie Prüfung des Retrospektions-Effekts durch drei verschiedene Veränderungs-Scores (Differenz, wahre Differenz, ALS).

VARIABLE	Aktiv	An- strg.	Auf- geregt	Ärger- lich	Be- drückt	Er- schöpft	Guter Stimmg.	Körp. wohl	Stress
Tag 1									
Mittel aktuelle Eingaben	3.79	2.47	2.46	1.91	2.13	3.84	4.60	4.56	2.87
Std. aktuelle Eingaben	0.78	0.93	1.01	0.66	0.85	1.14	0.73	1.00	0.90
Std. Fehler aktuelle Eing.	0.54	0.59	0.55	0.47	0.46	0.58	0.48	0.45	0.58
Stabilität aktuelle Eing.	0.52	0.60	0.71	0.50	0.70	0.75	0.56	0.80	0.59
Mittel Tagesretrospekt.	4.09	4.24	2.55	2.39	2.55	4.24	4.70	4.33	3.42
Std. Tagesretrospektion	1.57	1.52	1.39	1.46	1.37	1.68	0.95	1.16	1.37
Std. Fehler Tagesretrosp.	0.36	1.09	0.65	0.29	0.17	0.71	0.54	0.73	0.81
Anzahl Personen	33	33	33	33	33	33	33	33	33
Korrelation aktuell*retro	0.70	0.54	0.74	0.69	0.83	0.78	0.62	0.69	0.62
Kovarianz aktuell*retro	0.86	0.76	1.05	0.67	0.97	1.50	0.43	0.80	0.77
λ	0.44	3.45	1.43	0.38	0.14	1.51	1.26	2.65	1.98
Steigung 1. Hauptachse	2.71	1.48	1.44	3.06	1.91	1.54	1.44	1.02	1.58
F-Wert für AWA	32.45	1.43	5.24	46.86	31.63	9.05	3.00	0.01	3.87
p-Wert für AWA	0.00	0.24	0.03	0.00	0.00	0.01	0.09	0.94	0.06
Regressionskoeffizient	1.40	0.88	1.02	1.53	1.34	1.15	0.81	0.81	0.94
Std Differenz	1.16	1.29	0.93	1.11	0.82	1.06	0.76	0.86	1.08
t-Wert für Differenz	-1.46	-7.77	-0.53	-2.47	-2.86	-2.14	-0.74	1.46	-2.93
p-Wert für Differenz	0.15	0.00	0.60	0.02	0.01	0.04	0.46	0.15	0.01
Std wahre Differenz	1.32	1.31	0.99	1.25	0.94	1.14	0.77	0.84	1.12
t-Wert für wahre Diff.	-1.29	-7.63	-0.50	-2.19	-2.50	-1.98	-0.73	1.50	-2.82
p-Wert für wahre Diff.	0.21	0.00	0.62	0.04	0.02	0.06	0.47	0.14	0.01
Std ALS-Wert	1.12	1.28	0.93	1.05	0.76	1.04	0.75	0.84	1.08
t-Wert für ALS	-1.51	-7.79	-0.53	-2.60	-3.06	-2.17	-0.76	1.50	-2.93
p-Wert für ALS	0.14	0.00	0.60	0.01	0.00	0.04	0.46	0.14	0.01
Tag 2									
Mittel aktuelle Eingaben	3.92	2.53	2.35	2.05	2.25	3.77	4.71	4.81	2.63
Std. aktuelle Eingaben	0.67	0.91	1.06	0.87	0.96	1.06	0.77	0.91	0.87
Std. Fehler aktuelle Eing.	0.53	0.58	0.48	0.50	0.46	0.54	0.43	0.44	0.53
Stabilität aktuelle Eing.	0.37	0.59	0.79	0.67	0.77	0.74	0.68	0.77	0.63
Mittel Tagesretrospekt.	4.27	3.79	2.48	2.52	2.42	3.85	4.85	4.67	2.94
Std. Tagesretrospektion	1.23	1.49	1.30	1.25	1.25	1.54	1.03	1.24	1.27
Std. Fehler Tagesretrosp.	0.77	1.37	1.05	1.10	0.66	1.11	0.51	0.79	0.45
Anzahl Personen	33	33	33	33	33	33	33	33	33
Korrelation aktuell*retro	0.48	0.30	0.53	0.39	0.75	0.60	0.72	0.68	0.74
Kovarianz aktuell*retro	0.39	0.41	0.73	0.43	0.89	0.98	0.57	0.77	0.83
λ	2.07	5.63	4.78	4.74	2.03	4.23	1.38	3.31	0.71
Steigung 1. Hauptachse	2.35	0.84	0.82	0.85	1.27	1.17	1.43	1.19	1.72
F-Wert für AWA	7.16	0.02	0.15	0.03	1.95	0.23	4.52	0.53	13.34
p-Wert für AWA	0.01	0.90	0.70	0.86	0.17	0.63	0.04	0.47	0.00
Regressionskoeffizient	0.87	0.50	0.65	0.57	0.98	0.87	0.97	0.92	1.09
Std Differenz	1.09	1.50	1.16	1.21	0.83	1.25	0.72	0.92	0.85
t-Wert für Differenz	-1.85	-4.75	-0.66	-2.19	-1.19	-0.37	-1.11	0.90	-2.06
p-Wert für Differenz	0.07	0.00	0.52	0.04	0.24	0.71	0.28	0.37	0.05
Std wahre Differenz	1.13	1.43	1.12	1.15	0.86	1.25	0.75	0.92	0.94
t-Wert für wahre Diff.	-1.77	-4.98	-0.69	-2.30	-1.16	-0.37	-1.06	0.90	-1.87
p-Wert für wahre Diff.	0.09	0.00	0.50	0.03	0.26	0.71	0.30	0.38	0.07
Std ALS-Wert	1.08	1.42	1.10	1.15	0.83	1.24	0.72	0.91	0.85
t-Wert für ALS	-1.85	-4.99	-0.69	-2.31	-1.19	-0.37	-1.11	0.91	-2.06
p-Wert für ALS	0.07	0.00	0.49	0.03	0.24	0.71	0.27	0.37	0.05

Tabelle 6 (Fortsetzung)

VARIABLE	Aktiv	An- strg.	Auf- geregt	Ärger- lich	Be- drückt	Er- schöpft	Guter Stimmg.	Körp. wohl	Stress
Tag 3									
Mittel aktuelle Eingaben	4.28	2.46	2.32	1.90	2.32	3.46	4.62	4.90	2.60
Std. aktuelle Eingaben	0.74	0.94	0.94	0.95	0.96	0.90	1.03	0.89	0.99
Std. Fehler aktuelle Eing.	0.54	0.56	0.51	0.47	0.50	0.57	0.47	0.42	0.57
Stabilität aktuelle Eing.	0.47	0.64	0.70	0.75	0.72	0.60	0.79	0.78	0.68
Mittel Tagesretrospekt.	4.53	3.56	2.59	2.19	2.72	3.38	4.75	4.81	3.31
Std. Tagesretrospektion	1.08	1.64	1.32	1.35	1.40	1.26	1.24	1.18	1.53
Std. Fehler Tagesretrosp.	0.73	1.39	0.98	0.73	0.54	0.32	0.63	0.70	0.65
Anzahl Personen	32	32	32	32	32	32	32	32	32
Korrelation aktuell*retro	0.50	0.43	0.56	0.73	0.78	0.75	0.77	0.71	0.74
Kovarianz aktuell*retro	0.40	0.67	0.69	0.94	1.05	0.85	0.98	0.75	1.13
λ	1.86	6.07	3.67	2.44	1.14	0.32	1.79	2.76	1.32
Steigung 1. Hauptachse	1.56	1.17	1.12	1.39	1.58	1.76	1.17	1.20	1.71
F-Wert für AWA	2.10	0.07	0.10	3.17	10.59	14.77	0.90	0.76	10.76
p-Wert für AWA	0.16	0.80	0.76	0.09	0.00	0.00	0.35	0.39	0.00
Regressionskoeffizient	0.74	0.75	0.78	1.05	1.14	1.06	0.92	0.94	1.15
Std Differenz	0.95	1.50	1.11	0.93	0.88	0.84	0.80	0.83	1.03
t-Wert für Differenz	-1.49	-4.10	-1.37	-1.71	-2.53	0.56	-0.93	0.57	-3.84
p-Wert für Differenz	0.15	0.00	0.18	0.10	0.02	0.58	0.36	0.57	0.00
Std wahre Differenz	0.95	1.49	1.10	0.96	0.95	0.93	0.81	0.84	1.13
t-Wert für wahre Diff.	-1.49	-4.14	-1.39	-1.64	-2.33	0.50	-0.92	0.57	-3.52
p-Wert für wahre Diff.	0.15	0.00	0.17	0.11	0.03	0.62	0.37	0.57	0.00
Std ALS-Wert	0.93	1.48	1.09	0.92	0.87	0.84	0.80	0.83	1.02
t-Wert für ALS	-1.52	-4.15	-1.39	-1.71	-2.56	0.56	-0.93	0.58	-3.88
p-Wert für ALS	0.14	0.00	0.17	0.10	0.02	0.58	0.36	0.57	0.00
Tag 4									
Mittel aktuelle Eingaben	3.88	2.45	2.09	1.89	2.41	3.70	4.46	4.58	2.53
Std. aktuelle Eingaben	0.87	0.89	0.95	0.93	1.14	1.12	0.89	1.07	1.10
Std. Fehler aktuelle Eing.	0.52	0.52	0.47	0.44	0.44	0.55	0.45	0.42	0.52
Stabilität aktuelle Eing.	0.65	0.66	0.75	0.78	0.85	0.76	0.74	0.84	0.78
Mittel Tagesretrospekt.	4.13	3.80	2.60	2.50	2.50	3.77	4.37	4.60	3.03
Std. Tagesretrospektion	1.22	1.47	1.73	1.57	1.46	1.22	1.22	1.28	1.27
Std. Fehler Tagesretrosp.	1.02	1.14	0.55	1.15	0.68	0.47	0.90	0.68	0.57
Anzahl Personen	30	30	30	30	30	30	30	30	30
Korrelation aktuell*retro	0.44	0.51	0.82	0.60	0.81	0.80	0.58	0.78	0.79
Kovarianz aktuell*retro	0.47	0.67	1.36	0.88	1.35	1.10	0.63	1.06	1.11
λ	3.88	4.81	1.33	6.98	2.42	0.72	3.95	2.60	1.23
Steigung 1. Hauptachse	0.96	1.29	2.00	1.31	1.22	1.16	1.08	1.10	1.17
F-Wert für AWA	0.00	0.35	26.60	0.49	1.85	1.24	0.04	0.27	1.11
p-Wert für AWA	0.95	0.56	0.00	0.49	0.18	0.27	0.84	0.61	0.30
Regressionskoeffizient	0.62	0.85	1.50	1.02	1.04	0.88	0.80	0.93	0.91
Std Differenz	1.15	1.27	1.09	1.25	0.85	0.74	1.01	0.81	0.79
t-Wert für Differenz	-1.18	-5.72	-2.52	-2.63	-0.56	-0.50	0.52	-0.11	-3.47
p-Wert für Differenz	0.25	0.00	0.02	0.01	0.58	0.62	0.60	0.91	0.00
Std wahre Differenz	1.10	1.28	1.22	1.27	0.87	0.74	0.99	0.81	0.79
t-Wert für wahre Diff.	-1.23	-5.70	-2.27	-2.59	-0.55	-0.50	0.53	-0.11	-3.44
p-Wert für wahre Diff.	0.23	0.00	0.03	0.01	0.59	0.62	0.60	0.92	0.00
Std ALS-Wert	1.10	1.26	0.98	1.25	0.85	0.73	0.99	0.80	0.78
t-Wert für ALS	-1.23	-5.75	-2.81	-2.63	-0.57	-0.51	0.53	-0.11	-3.50
p-Wert für ALS	0.23	0.00	0.01	0.01	0.58	0.62	0.60	0.91	0.00

¹ Sonderfall bei Schätzung der Fehlervarianz des Verlaufswerts bewirkt $b = s_y^2 / s_{xy}$

Tabelle 6 (Fortsetzung)

VARIABLE	Aktiv	An- strg.	Auf- geregt	Ärger- lich	Be- drückt	Er- schöpft	Guter Stimmg.	Körp. wohl	Stress
Tag 5									
Mittel aktuelle Eingaben	3.87	2.46	2.20	1.98	2.59	3.83	4.28	4.77	2.52
Std. aktuelle Eingaben	1.03	1.10	0.71	0.88	1.26	1.17	1.03	1.15	0.82
Std. Fehler aktuelle Eing.	0.46	0.50	0.49	0.48	0.50	0.50	0.45	0.35	0.54
Stabilität aktuelle Eing.	0.80	0.79	0.52	0.69	0.85	0.82	0.81	0.91	0.58
Mittel Tagesretrospekt.	3.97	3.64	2.48	2.06	3.00	3.82	4.48	4.52	2.88
Std. Tagesretrospektion	1.36	1.43	1.15	1.12	1.60	1.49	1.06	1.39	1.32
Std. Fehler Tagesretrosp.	0.57	1.01	0.66	0.80	0.65	0.57	0.60	0.52	-0.68
Anzahl Personen	33	33	33	33	33	33	33	33	33
Korrelation aktuell*retro	0.81	0.63	0.59	0.58	0.84	0.84	0.74	0.88	0.85
Kovarianz aktuell*retro	1.13	0.99	0.48	0.57	1.70	1.45	0.81	1.41	0.93
λ	1.55	4.02	1.80	2.75	1.72	1.32	1.76	2.27	1.61
Steigung 1. Hauptachse	1.34	1.03	1.84	1.06	1.26	1.30	0.96	1.18	2.37
F-Wert für AWA	4.96	0.01	6.68	0.04	3.80	5.08	0.06	2.64	50.45
p-Wert für AWA	0.03	0.91	0.01	0.85	0.06	0.03	0.81	0.11	0.00
Regressionskoeffizient	1.07	0.82	0.96	0.74	1.07	1.07	0.77	1.08	1.37
Std Differenz	0.80	1.13	0.93	0.94	0.87	0.82	0.75	0.66	0.75
t-Wert für Differenz	-0.68	-5.90	-1.72	-0.46	-2.67	0.10	-1.51	2.14	-2.68
p-Wert für Differenz	0.50	0.00	0.10	0.65	0.01	0.92	0.14	0.04	0.01
Std wahre Differenz	0.84	1.11	0.98	0.91	0.91	0.87	0.71	0.68	0.94
t-Wert für wahre Diff.	-0.65	-6.00	-1.63	-0.48	-2.56	0.10	-1.59	2.08	-2.12
p-Wert für wahre Diff.	0.52	0.00	0.11	0.64	0.02	0.92	0.12	0.05	0.04
Std ALS-Wert	0.79	1.11	0.93	0.91	0.87	0.82	0.71	0.66	0.68
t-Wert für ALS	-0.68	-6.00	-1.72	-0.48	-2.68	0.10	-1.59	2.16	-2.93
p-Wert für ALS	0.50	0.00	0.09	0.64	0.01	0.92	0.12	0.04	0.01
Tag 6									
Mittel aktuelle Eingaben	3.79	2.66	2.04	1.91	2.49	3.83	4.24	4.50	2.71
Std. aktuelle Eingaben	1.07	1.10	1.01	1.00	1.17	1.15	1.09	1.46	1.05
Std. Fehler aktuelle Eing.	0.45	0.49	0.38	0.36	0.43	0.51	0.37	0.36	0.55
Stabilität aktuelle Eing.	0.82	0.80	0.86	0.87	0.86	0.80	0.88	0.94	0.72
Mittel Tagesretrospekt.	3.82	3.64	2.36	2.15	2.70	4.15	4.21	4.24	3.33
Std. Tagesretrospektion	1.36	1.65	1.69	1.52	1.53	1.42	1.43	1.68	1.61
Std. Fehler Tagesretrosp.	0.84	0.84	0.83	0.80	0.65	0.34	0.47	0.64	-0.30
Anzahl Personen	33	33	33	33	33	33	33	33	33
Korrelation aktuell*retro	0.71	0.77	0.81	0.79	0.84	0.87	0.89	0.89	0.86
Kovarianz aktuell*retro	1.04	1.41	1.38	1.22	1.50	1.42	1.39	2.20	1.46
λ	3.48	2.88	4.80	4.92	2.24	0.43	1.60	3.14	0.29
Steigung 1. Hauptachse	1.10	1.45	1.58	1.38	1.28	1.34	1.31	1.09	1.84
F-Wert für AWA	0.16	5.21	8.21	3.36	4.00	8.31	8.28	0.71	29.05
p-Wert für AWA	0.69	0.03	0.01	0.08	0.05	0.01	0.01	0.41	0.00
Regressionskoeffizient	0.90	1.16	1.36	1.20	1.10	1.07	1.16	1.03	1.33
Std Differenz	0.96	1.07	1.06	0.95	0.83	0.70	0.68	0.75	0.88
t-Wert für Differenz	-0.15	-5.17	-1.75	-1.46	-1.41	-2.63	0.27	1.91	-3.97
p-Wert für Differenz	0.88	0.00	0.09	0.15	0.17	0.01	0.79	0.06	0.00
Std wahre Differenz	0.96	1.12	1.12	0.98	0.87	0.76	0.72	0.76	1.03
t-Wert für wahre Diff.	-0.15	-4.91	-1.66	-1.41	-1.35	-2.42	0.25	1.89	-3.39
p-Wert für wahre Diff.	0.88	0.00	0.11	0.17	0.19	0.02	0.80	0.07	0.00
Std ALS-Wert	0.95	1.05	1.00	0.93	0.83	0.70	0.66	0.75	0.81
t-Wert für ALS	-0.15	-5.25	-1.86	-1.49	-1.43	-2.65	0.28	1.92	-4.31
p-Wert für ALS	0.88	0.00	0.07	0.14	0.16	0.01	0.78	0.06	0.00

Tabelle 6 (Fortsetzung)

VARIABLE	Aktiv	An- strg.	Auf- geregt	Ärger- lich	Be- drückt	Er- schöpft	Guter Stimmg.	Körp. wohl	Stress
Tag 7									
Mittel aktuelle Eingaben	3.64	2.48	2.08	2.00	2.36	3.89	4.12	4.50	2.42
Std. aktuelle Eingaben	1.06	0.91	0.77	0.99	1.16	1.20	1.01	1.50	1.09
Std. Fehler aktuelle Eing.	0.47	0.50	0.42	0.47	0.39	0.51	0.39	0.31	0.52
Stabilität aktuelle Eing.	0.80	0.70	0.70	0.78	0.89	0.82	0.85	0.96	0.77
Mittel Tagesretrospekt.	3.75	3.53	2.53	2.28	2.81	3.97	4.16	4.50	3.16
Std. Tagesretrospektion	1.34	1.61	1.19	1.42	1.60	1.40	1.27	1.52	1.48
Std. Fehler Tagesretrosp.	0.64	0.91	0.72	0.44	0.41	0.32	0.65	0.74	0.82
Anzahl Personen	32	32	32	32	32	32	32	32	32
Korrelation aktuell*retro	0.79	0.69	0.66	0.84	0.91	0.88	0.79	0.85	0.73
Kovarianz aktuell*retro	1.12	1.02	0.61	1.18	1.68	1.49	1.02	1.95	1.18
λ	1.82	3.31	2.91	0.88	1.11	0.40	2.75	5.55	2.51
Steigung 1. Hauptachse	1.25	1.73	1.48	1.54	1.42	1.26	1.17	0.91	1.29
F-Wert für AWA	2.24	6.58	2.82	14.20	18.03	5.53	0.92	0.31	1.87
p-Wert für AWA	0.14	0.02	0.10	0.00	0.00	0.03	0.34	0.58	0.18
Regressionskoeffizient	1.00	1.22	1.03	1.20	1.26	1.03	0.99	0.87	1.00
Std Differenz	0.83	1.18	0.89	0.80	0.72	0.66	0.78	0.82	1.01
t-Wert für Differenz	-0.75	-5.00	-2.84	-1.96	-3.47	-0.69	-0.24	0.03	-4.06
p-Wert für Differenz	0.46	0.00	0.01	0.06	0.00	0.49	0.81	0.98	0.00
Std wahre Differenz	0.86	1.25	0.93	0.88	0.78	0.70	0.79	0.80	1.04
t-Wert für wahre Diff.	-0.72	-4.70	-2.73	-1.78	-3.20	-0.65	-0.24	0.03	-3.94
p-Wert für wahre Diff.	0.47	0.00	0.01	0.09	0.00	0.52	0.81	0.98	0.00
Std ALS-Wert	0.83	1.16	0.89	0.78	0.66	0.66	0.78	0.79	1.01
t-Wert für ALS	-0.75	-5.07	-2.84	-2.02	-3.81	-0.69	-0.24	0.03	-4.06
p-Wert für ALS	0.46	0.00	0.01	0.05	0.00	0.49	0.81	0.98	0.00

¹ Sonderfall bei Schätzung der Fehlervarianz des Verlaufswerts bewirkt $b=s_y^2/s_{xy}$

Anmerkung:

Die Tabellen enthalten im oberen Teil die Grundstatistiken der (gemittelten) aktuellen und der retrospektiven Einstufungen. Die Fehlervarianz der aktuellen Einstufungen wird aus der Innerhalb-Personen-Varianz berechnet (einfache Varianzanalyse), die Fehlervarianzen der gemittelten aktuellen Einstufungen vermindern sich um den Faktor 6 (Anzahl der Summanden, d.h. Abfragen pro Tag). Die Fehlervarianzen der retrospektiven Einstufungen werden aus dem Modell geschätzt (Struktur-Regressions-Modell nach Kendall und Stuart, 1967, Case 1; s.a. Foerster, 1994, 1995). Im Mittelteil sind die Bestimmungsstücke zur Prüfung des Ausgangswertgesetzes (AWA) angegeben: Korrelation bzw. Kovarianz zwischen aktuellen und retrospektiven Eingaben, Fehlervarianz-Verhältnis λ , Prüfgröße b (Steigung der ersten Hauptachse) mit Teststatistik F und p -Wert für die Hypothese $b=1$. Im unteren Teil wird der Retrospektions-Effekt getestet, wobei drei unterschiedliche Veränderungswerte benutzt werden: (1) einfache Differenz von aktuellen minus retrospektiven Einstufungen; (2) die durch das Modell geschätzte wahre Differenz, die den echten Anteil der AWA, nicht jedoch deren statistischen Anteil enthält und sich nur durch ihre Varianz (Skalierung) von der einfachen Differenz unterscheidet; (3) der Autonomic Liability Score (ALS) von Lacey (1956), also das Residuum des Verlaufswerts nach Auspartialisieren des Ausgangswert.

Tabelle 7: Vergleiche des Retrospektionseffekts zwischen den Tagen bei der Wochenstudie.

Variable	N	Mittelwerte				Standardabweichungen				Korrelation		Intraklass-K.	
		akt.	retro	t	p	akt.	retro	t	p	r _{ar}	p	r _{IC}	p
Tag 1 mit 2													
Aktiv	33	-0.30	-0.35	0.2	.823	1.16	1.09	0.4	.695	0.26	.144	0.27	.057
Anstrengend	33	-1.77	-1.26	1.4	.159	1.29	1.50	0.8	.407	-0.04	.841	-0.05	.616
Aufgeregt	33	-0.09	-0.14	0.2	.862	0.93	1.16	1.2	.220	-0.07	.693	-0.05	.621
Ärgerlich	33	-0.48	-0.47	0.1	.956	1.11	1.21	0.5	.608	0.22	.208	0.24	.085
Bedrückt	33	-0.41	-0.17	1.2	.256	0.82	0.83	0.1	.912	-0.00	.983	-0.01	.521
Erschöpft	33	-0.40	-0.08	1.2	.241	1.06	1.25	0.9	.359	0.15	.400	0.14	.208
Gute Stimm	33	-0.10	-0.14	0.2	.835	0.76	0.72	0.3	.756	-0.13	.457	-0.12	.750
Körp.wohl	33	0.22	0.15	0.3	.739	0.86	0.92	0.4	.720	-0.03	.866	-0.02	.537
Stress	33	-0.56	-0.31	1.3	.194	1.08	0.85	1.4	.157	0.42	.015	0.39	.010
Tag 1 mit 3													
Aktiv	32	-0.30	-0.26	0.2	.865	1.18	0.96	1.2	.256	0.03	.869	0.04	.400
Anstrengend	32	-1.75	-1.08	2.0	.051	1.30	1.50	0.8	.436	0.14	.452	0.09	.309
Aufgeregt	32	-0.14	-0.24	0.4	.700	0.90	1.11	1.2	.253	-0.16	.377	-0.14	.791
Ärgerlich	32	-0.51	-0.26	1.0	.340	1.11	0.93	1.0	.324	-0.06	.756	-0.06	.621
Bedrückt	32	-0.37	-0.36	0.1	.956	0.80	0.87	0.5	.629	0.19	.286	0.21	.119
Erschöpft	32	-0.35	0.05	1.6	.110	1.03	0.83	1.2	.246	-0.06	.741	-0.09	.684
Gute Stimm	32	-0.10	-0.18	0.4	.672	0.77	0.77	0.0	.992	-0.04	.836	-0.03	.555
Körp.wohl	32	0.23	0.12	0.5	.599	0.87	0.82	0.3	.732	0.08	.654	0.09	.301
Stress	32	-0.57	-0.68	0.4	.687	1.10	1.03	0.4	.727	-0.02	.934	-0.00	.503
Tag 1 mit 4													
Aktiv	30	-0.41	-0.25	0.5	.595	1.16	1.17	0.1	.946	0.08	.681	0.09	.313
Anstrengend	30	-1.77	-1.34	1.4	.177	1.31	1.27	0.2	.853	0.13	.485	0.12	.264
Aufgeregt	30	-0.12	-0.54	1.7	.108	0.91	1.12	1.1	.284	0.13	.483	0.10	.292
Ärgerlich	30	-0.49	-0.62	0.5	.635	1.11	1.22	0.5	.631	0.23	.219	0.24	.092
Bedrückt	30	-0.41	-0.12	1.5	.137	0.82	0.79	0.3	.803	0.20	.293	0.18	.169
Erschöpft	30	-0.30	-0.07	0.9	.394	1.04	0.80	1.5	.152	-0.13	.501	-0.12	.740
Gute Stimm	30	-0.09	0.16	1.2	.240	0.77	1.00	1.4	.169	0.24	.193	0.23	.105
Körp.wohl	30	0.23	0.06	0.8	.406	0.88	0.80	0.5	.621	0.16	.406	0.16	.191
Stress	30	-0.65	-0.53	0.5	.633	1.02	0.77	1.5	.141	-0.07	.696	-0.06	.623
Tag 1 mit 5													
Aktiv	33	-0.30	-0.10	0.8	.427	1.16	0.80	2.2	.038	-0.04	.822	-0.03	.573
Anstrengend	33	-1.77	-1.18	1.9	.063	1.29	1.13	0.7	.475	-0.02	.932	-0.06	.624
Aufgeregt	33	-0.09	-0.28	1.0	.335	0.93	0.93	0.0	.968	0.27	.125	0.27	.058
Ärgerlich	33	-0.48	-0.08	1.7	.095	1.11	0.94	0.9	.354	0.15	.407	0.12	.252
Bedrückt	33	-0.41	-0.41	0.0	.991	0.82	0.87	0.4	.713	0.22	.223	0.23	.091
Erschöpft	33	-0.40	0.02	1.6	.114	1.06	0.82	1.4	.161	-0.17	.344	-0.19	.860
Gute Stimm	33	-0.10	-0.20	0.6	.584	0.76	0.75	0.1	.935	0.08	.647	0.09	.298
Körp.wohl	33	0.22	0.25	0.2	.875	0.86	0.66	1.5	.144	0.13	.456	0.14	.205
Stress	33	-0.56	-0.35	0.9	.399	1.08	0.75	2.1	.044	-0.06	.723	-0.06	.624
Tag 1 mit 6													
Aktiv	33	-0.30	-0.03	1.2	.233	1.16	0.96	1.1	.265	0.28	.108	0.27	.058
Anstrengend	33	-1.77	-0.98	2.9	.006	1.29	1.07	1.1	.294	0.16	.388	0.05	.392
Aufgeregt	33	-0.09	-0.33	1.1	.284	0.93	1.06	0.7	.470	0.22	.217	0.22	.108
Ärgerlich	33	-0.48	-0.24	1.2	.255	1.11	0.95	0.9	.358	0.36	.039	0.35	.020
Bedrückt	33	-0.41	-0.21	1.1	.261	0.82	0.83	0.1	.900	0.24	.171	0.24	.085
Erschöpft	33	-0.40	-0.33	0.3	.737	1.06	0.70	2.4	.025	0.06	.759	0.06	.357
Gute Stimm	33	-0.10	0.03	0.7	.502	0.76	0.68	0.6	.533	-0.17	.348	-0.16	.818
Körp.wohl	33	0.22	0.25	0.2	.878	0.86	0.75	0.8	.451	-0.01	.952	0.00	.489
Stress	33	-0.56	-0.62	0.3	.775	1.08	0.88	1.2	.253	0.27	.125	0.28	.053

Tabelle 7 (Fortsetzung)

Variable	N	Mittelwerte				Standardabweichungen				Korrelation		Intraclass-K.	
		akt.	retro	t	p	akt.	retro	t	p	r _{ar}	p	r _{IC}	p
Tag 1 mit 7													
Aktiv	32	-0.24	-0.14	0.4	.657	1.13	0.83	1.7	.090	0.23	.213	0.23	.099
Anstrengend	32	-1.73	-1.04	2.3	.028	1.29	1.18	0.5	.629	0.08	.678	0.01	.474
Aufgeregt	32	-0.12	-0.45	1.7	.093	0.93	0.90	0.2	.869	0.33	.067	0.30	.044
Ärgerlich	32	-0.51	-0.28	1.0	.328	1.11	0.82	1.7	.099	0.10	.570	0.10	.289
Bedrückt	32	-0.44	-0.44	0.0	.991	0.82	0.75	0.5	.623	0.09	.623	0.11	.277
Erschöpft	32	-0.39	-0.07	1.7	.091	1.07	0.69	2.7	.010	0.37	.036	0.31	.039
Gute Stimm	32	-0.07	-0.08	0.0	.965	0.76	0.75	0.0	.990	-0.01	.941	0.00	.494
Körp.wohl	32	0.24	-0.04	1.2	.241	0.87	0.83	0.3	.784	-0.16	.388	-0.16	.821
Stress	32	-0.61	-0.72	0.6	.560	1.06	1.03	0.2	.848	0.45	.009	0.46	.003
Tag 2 mit 3													
Aktiv	32	-0.37	-0.26	0.5	.600	1.10	0.96	0.8	.427	0.34	.053	0.35	.022
Anstrengend	32	-1.24	-1.08	0.5	.613	1.52	1.50	0.1	.959	0.33	.066	0.34	.026
Aufgeregt	32	-0.13	-0.24	0.4	.682	1.18	1.11	0.3	.729	0.16	.387	0.17	.169
Ärgerlich	32	-0.51	-0.26	1.0	.314	1.21	0.93	1.5	.144	0.18	.325	0.17	.167
Bedrückt	32	-0.20	-0.36	0.9	.351	0.83	0.87	0.3	.794	0.37	.038	0.37	.017
Erschöpft	32	-0.10	0.05	0.6	.543	1.26	0.83	2.4	.025	0.16	.383	0.16	.190
Gute Stimm	32	-0.12	-0.18	0.4	.722	0.72	0.77	0.4	.706	0.13	.486	0.14	.215
Körp.wohl	32	0.18	0.12	0.3	.768	0.91	0.82	0.6	.575	0.08	.654	0.10	.295
Stress	32	-0.33	-0.68	1.8	.074	0.86	1.03	1.0	.306	0.38	.030	0.34	.024
Tag 2 mit 4													
Aktiv	30	-0.43	-0.25	1.0	.326	1.07	1.17	0.7	.520	0.63	.000	0.63	.000
Anstrengend	30	-1.25	-1.34	0.3	.797	1.56	1.27	1.1	.273	0.12	.540	0.13	.241
Aufgeregt	30	-0.17	-0.54	1.4	.186	1.20	1.12	0.4	.707	0.21	.256	0.20	.138
Ärgerlich	30	-0.50	-0.62	0.4	.713	1.26	1.22	0.2	.859	0.08	.676	0.09	.305
Bedrückt	30	-0.18	-0.12	0.3	.787	0.85	0.79	0.4	.685	0.02	.902	0.04	.416
Erschöpft	30	0.03	-0.07	0.5	.614	1.16	0.80	2.2	.034	0.40	.030	0.38	.016
Gute Stimm	30	-0.13	0.16	1.3	.207	0.73	1.00	1.7	.109	0.04	.842	0.02	.447
Körp.wohl	30	0.14	0.06	0.4	.701	0.94	0.80	0.8	.411	0.16	.386	0.18	.169
Stress	30	-0.34	-0.53	1.0	.327	0.88	0.77	0.7	.492	0.27	.144	0.27	.068
Tag 2 mit 5													
Aktiv	33	-0.35	-0.10	1.0	.316	1.09	0.80	1.8	.086	-0.15	.419	-0.14	.786
Anstrengend	33	-1.26	-1.18	0.3	.789	1.50	1.13	1.6	.109	0.29	.099	0.29	.044
Aufgeregt	33	-0.14	-0.28	0.9	.374	1.16	0.93	1.7	.106	0.63	.000	0.62	.000
Ärgerlich	33	-0.47	-0.08	1.8	.087	1.21	0.94	1.5	.135	0.33	.057	0.29	.044
Bedrückt	33	-0.17	-0.41	1.1	.264	0.83	0.87	0.3	.803	0.04	.809	0.04	.413
Erschöpft	33	-0.08	0.02	0.5	.641	1.25	0.82	2.6	.013	0.42	.014	0.40	.009
Gute Stimm	33	-0.14	-0.20	0.4	.712	0.72	0.75	0.2	.814	0.27	.132	0.28	.053
Körp.wohl	33	0.15	0.25	0.7	.520	0.92	0.66	2.0	.053	0.38	.029	0.37	.016
Stress	33	-0.31	-0.35	0.3	.771	0.85	0.75	0.8	.420	0.43	.012	0.44	.004
Tag 2 mit 6													
Aktiv	33	-0.35	-0.03	1.4	.160	1.09	0.96	0.7	.478	0.20	.264	0.18	.149
Anstrengend	33	-1.26	-0.98	1.0	.324	1.50	1.07	2.0	.055	0.27	.132	0.25	.073
Aufgeregt	33	-0.14	-0.33	0.8	.408	1.16	1.06	0.6	.583	0.32	.073	0.32	.032
Ärgerlich	33	-0.47	-0.24	1.0	.338	1.21	0.95	1.4	.159	0.28	.113	0.27	.057
Bedrückt	33	-0.17	-0.21	0.2	.860	0.83	0.83	0.0	.990	0.18	.306	0.20	.128
Erschöpft	33	-0.08	-0.33	1.0	.306	1.25	0.70	3.4	.001	0.16	.378	0.13	.223
Gute Stimm	33	-0.14	0.03	0.9	.381	0.72	0.68	0.3	.751	-0.24	.169	-0.24	.917
Körp.wohl	33	0.15	0.25	0.5	.613	0.92	0.75	1.1	.267	-0.01	.968	0.00	.490
Stress	33	-0.31	-0.62	1.9	.061	0.85	0.88	0.2	.829	0.46	.006	0.43	.005

Tabelle 7 (Fortsetzung)

Variable	N	Mittelwerte				Standardabweichungen				Korrelation		Intraclass-K.	
		akt.	retro	t	p	akt.	retro	t	p	r _{ar}	p	r _{IC}	p
Tag 2 mit 7													
Aktiv	32	-0.32	-0.14	0.8	.410	1.09	0.83	1.5	.138	0.22	.225	0.22	.110
Anstrengend	32	-1.21	-1.04	0.6	.554	1.50	1.18	1.4	.182	0.30	.091	0.30	.041
Aufgeregt	32	-0.07	-0.45	1.6	.122	1.12	0.90	1.2	.225	0.15	.403	0.13	.241
Ärgerlich	32	-0.44	-0.28	0.6	.544	1.22	0.82	2.2	.033	0.07	.699	0.08	.336
Bedrückt	32	-0.11	-0.44	1.6	.111	0.75	0.75	0.0	.993	-0.13	.468	-0.16	.813
Erschöpft	32	-0.13	-0.07	0.3	.785	1.23	0.69	3.4	.001	0.16	.375	0.15	.196
Gute Stimm	32	-0.19	-0.08	0.6	.552	0.67	0.75	0.7	.507	-0.07	.709	-0.06	.626
Körp.wohl	32	0.06	-0.04	0.5	.636	0.77	0.83	0.4	.720	0.06	.733	0.07	.338
Stress	32	-0.28	-0.72	1.8	.081	0.85	1.03	1.1	.296	-0.07	.705	-0.10	.718
Tag 3 mit 4													
Aktiv	29	-0.21	-0.22	0.0	.971	0.99	1.18	1.1	.270	0.58	.000	0.59	.000
Anstrengend	29	-1.18	-1.41	0.7	.459	1.53	1.22	1.3	.221	0.31	.102	0.31	.047
Aufgeregt	29	-0.23	-0.56	1.0	.320	1.14	1.13	0.0	.964	-0.13	.505	-0.13	.754
Ärgerlich	29	-0.26	-0.64	1.4	.184	0.95	1.23	1.4	.177	0.11	.565	0.09	.312
Bedrückt	29	-0.39	-0.13	1.2	.235	0.90	0.80	0.6	.529	0.13	.515	0.12	.268
Erschöpft	29	0.03	-0.11	0.7	.512	0.87	0.78	0.6	.584	0.16	.409	0.17	.185
Gute Stimm	29	-0.14	0.17	1.4	.166	0.79	1.01	1.3	.196	0.20	.288	0.18	.167
Körp.wohl	29	0.20	0.11	0.6	.581	0.77	0.78	0.1	.958	0.40	.031	0.41	.012
Stress	29	-0.70	-0.55	0.8	.437	1.05	0.77	1.8	.086	0.45	.014	0.44	.007
Tag 3 mit 5													
Aktiv	32	-0.26	-0.12	0.6	.550	0.96	0.79	1.1	.300	0.02	.911	0.03	.433
Anstrengend	32	-1.08	-1.21	0.5	.652	1.50	1.14	1.6	.116	0.31	.087	0.31	.040
Aufgeregt	32	-0.24	-0.30	0.2	.830	1.11	0.93	1.0	.348	0.01	.971	0.02	.452
Ärgerlich	32	-0.26	-0.08	1.2	.243	0.93	0.95	0.2	.846	0.61	.000	0.61	.000
Bedrückt	32	-0.36	-0.38	0.1	.953	0.87	0.86	0.0	.962	-0.12	.527	-0.10	.711
Erschöpft	32	0.05	0.02	0.2	.857	0.83	0.83	0.0	.999	0.24	.176	0.26	.070
Gute Stimm	32	-0.18	-0.20	0.1	.938	0.77	0.76	0.1	.939	0.32	.078	0.33	.029
Körp.wohl	32	0.12	0.24	0.6	.567	0.82	0.67	1.1	.259	-0.16	.394	-0.14	.786
Stress	32	-0.68	-0.34	2.0	.052	1.03	0.76	2.0	.058	0.51	.002	0.45	.004
Tag 3 mit 6													
Aktiv	32	-0.26	-0.05	0.9	.349	0.96	0.97	0.0	.969	0.17	.352	0.17	.168
Anstrengend	32	-1.08	-0.94	0.4	.661	1.50	1.07	1.9	.062	0.16	.369	0.17	.173
Aufgeregt	32	-0.24	-0.33	0.3	.740	1.11	1.08	0.2	.862	0.09	.614	0.11	.276
Ärgerlich	32	-0.26	-0.23	0.1	.889	0.93	0.96	0.2	.856	0.18	.322	0.20	.135
Bedrückt	32	-0.36	-0.17	1.2	.249	0.87	0.82	0.3	.737	0.43	.014	0.42	.007
Erschöpft	32	0.05	-0.29	2.0	.054	0.83	0.69	1.1	.286	0.20	.280	0.15	.203
Gute Stimm	32	-0.18	-0.00	1.2	.258	0.77	0.66	0.9	.379	0.27	.136	0.26	.070
Körp.wohl	32	0.12	0.24	0.6	.572	0.82	0.76	0.4	.670	-0.11	.532	-0.10	.718
Stress	32	-0.68	-0.63	0.2	.849	1.03	0.89	0.8	.443	0.17	.344	0.19	.148
Tag 3 mit 7													
Aktiv	31	-0.34	-0.16	0.8	.442	0.86	0.84	0.2	.873	-0.03	.861	-0.03	.557
Anstrengend	31	-1.05	-0.95	0.3	.784	1.52	1.09	1.9	.068	-0.22	.243	-0.19	.852
Aufgeregt	31	-0.20	-0.49	1.2	.239	1.10	0.89	1.1	.263	0.14	.455	0.13	.239
Ärgerlich	31	-0.21	-0.30	0.5	.634	0.90	0.83	0.4	.669	0.29	.116	0.30	.047
Bedrückt	31	-0.33	-0.42	0.5	.610	0.87	0.75	0.9	.392	0.41	.021	0.42	.008
Erschöpft	31	0.09	-0.04	0.8	.411	0.82	0.68	1.1	.281	0.36	.049	0.35	.022
Gute Stimm	31	-0.22	-0.09	0.6	.538	0.76	0.76	0.0	.994	-0.00	.994	0.01	.481
Körp.wohl	31	0.11	-0.03	0.6	.572	0.83	0.84	0.0	.962	-0.39	.030	-0.38	.984
Stress	31	-0.65	-0.68	0.1	.933	1.03	1.01	0.1	.883	-0.45	.010	-0.44	.994

Tabelle 7 (Fortsetzung)

Variable	N	Mittelwerte				Standardabweichungen				Korrelation		Intraclass-K.	
		akt.	retro	t	p	akt.	retro	t	p	r _{ar}	p	r _{IC}	p
Tag 4 mit 5													
Aktiv	30	-0.25	-0.10	0.6	.569	1.17	0.78	2.3	.031	-0.06	.753	-0.04	.593
Anstrengend	30	-1.34	-1.24	0.5	.636	1.27	1.14	0.7	.484	0.58	.000	0.59	.000
Aufgeregt	30	-0.54	-0.31	1.1	.274	1.12	0.95	0.9	.362	0.42	.021	0.41	.010
Ärgerlich	30	-0.62	-0.05	1.8	.083	1.22	0.97	1.2	.233	-0.18	.328	-0.22	.880
Bedrückt	30	-0.12	-0.42	1.7	.100	0.79	0.84	0.4	.703	0.33	.079	0.30	.051
Erschöpft	30	-0.07	0.02	0.5	.619	0.80	0.86	0.4	.666	0.24	.207	0.25	.087
Gute Stimm	30	0.16	-0.14	1.6	.127	1.00	0.76	1.6	.122	0.35	.060	0.31	.042
Körp.wohl	30	0.06	0.24	0.9	.353	0.80	0.66	1.0	.309	0.08	.664	0.08	.328
Stress	30	-0.53	-0.33	1.2	.251	0.77	0.77	0.0	.980	0.28	.137	0.27	.068
Tag 4 mit 6													
Aktiv	30	-0.25	-0.02	0.9	.370	1.17	0.97	1.0	.314	0.22	.240	0.22	.115
Anstrengend	30	-1.34	-0.91	1.6	.118	1.27	1.07	0.9	.366	0.27	.147	0.24	.093
Aufgeregt	30	-0.54	-0.36	0.8	.441	1.12	1.11	0.0	.960	0.36	.051	0.36	.021
Ärgerlich	30	-0.62	-0.24	1.2	.242	1.22	0.98	1.2	.238	-0.22	.249	-0.22	.883
Bedrückt	30	-0.12	-0.25	0.6	.527	0.79	0.86	0.5	.618	0.12	.523	0.13	.240
Erschöpft	30	-0.07	-0.34	1.3	.213	0.80	0.70	0.7	.483	-0.17	.373	-0.18	.832
Gute Stimm	30	0.16	0.09	0.3	.764	1.00	0.67	2.2	.039	0.02	.902	0.04	.420
Körp.wohl	30	0.06	0.21	0.8	.450	0.80	0.75	0.4	.718	0.07	.695	0.08	.330
Stress	30	-0.53	-0.60	0.4	.675	0.77	0.92	1.0	.311	0.39	.034	0.39	.013
Tag 4 mit 7													
Aktiv	29	-0.29	-0.13	0.6	.582	1.18	0.86	1.6	.114	-0.00	.980	0.01	.484
Anstrengend	29	-1.24	-1.07	0.6	.581	1.18	1.23	0.2	.837	0.06	.758	0.07	.352
Aufgeregt	29	-0.50	-0.46	0.2	.847	1.12	0.93	1.0	.323	0.23	.236	0.24	.099
Ärgerlich	29	-0.60	-0.21	1.4	.174	1.23	0.83	2.1	.041	0.00	.993	-0.02	.532
Bedrückt	29	-0.10	-0.44	1.8	.080	0.79	0.77	0.1	.931	0.17	.375	0.13	.241
Erschöpft	29	-0.09	-0.07	0.1	.929	0.80	0.72	0.6	.561	0.10	.618	0.11	.273
Gute Stimm	29	0.15	-0.12	1.1	.270	1.02	0.77	1.4	.166	0.00	.990	-0.00	.506
Körp.wohl	29	0.06	-0.01	0.4	.726	0.82	0.86	0.3	.767	0.14	.459	0.16	.200
Stress	29	-0.48	-0.75	0.9	.361	0.75	1.06	2.1	.047	-0.42	.022	-0.40	.985
Tag 5 mit 6													
Aktiv	33	-0.10	-0.03	0.4	.705	0.80	0.96	1.1	.288	0.30	.094	0.30	.039
Anstrengend	33	-1.18	-0.98	0.8	.404	1.13	1.07	0.3	.739	0.22	.215	0.22	.098
Aufgeregt	33	-0.28	-0.33	0.3	.789	0.93	1.06	0.9	.384	0.52	.001	0.53	.001
Ärgerlich	33	-0.08	-0.24	0.8	.405	0.94	0.95	0.1	.950	0.29	.105	0.29	.046
Bedrückt	33	-0.41	-0.21	1.1	.262	0.87	0.83	0.2	.804	0.30	.088	0.30	.043
Erschöpft	33	0.02	-0.33	1.7	.094	0.82	0.70	0.9	.378	-0.07	.687	-0.10	.718
Gute Stimm	33	-0.20	0.03	1.4	.159	0.75	0.68	0.5	.586	0.19	.286	0.17	.160
Körp.wohl	33	0.25	0.25	0.0	.984	0.66	0.75	0.7	.476	0.19	.284	0.21	.120
Stress	33	-0.35	-0.62	1.3	.192	0.75	0.88	0.9	.358	0.05	.761	0.04	.406
Tag 5 mit 7													
Aktiv	32	-0.14	-0.14	0.0	.962	0.78	0.83	0.5	.634	0.58	.000	0.59	.000
Anstrengend	32	-1.13	-1.04	0.4	.713	1.12	1.18	0.3	.743	0.25	.169	0.26	.069
Aufgeregt	32	-0.29	-0.45	0.8	.418	0.94	0.90	0.2	.805	0.26	.156	0.26	.069
Ärgerlich	32	-0.08	-0.28	1.0	.334	0.95	0.82	0.8	.407	0.17	.360	0.17	.176
Bedrückt	32	-0.39	-0.44	0.2	.830	0.88	0.75	0.9	.381	-0.05	.776	-0.04	.580
Erschöpft	32	0.02	-0.07	0.4	.681	0.83	0.69	1.1	.283	-0.18	.311	-0.17	.827
Gute Stimm	32	-0.25	-0.08	1.1	.264	0.70	0.75	0.5	.647	0.32	.072	0.32	.035
Körp.wohl	32	0.21	-0.04	1.5	.150	0.62	0.83	1.6	.112	0.22	.228	0.19	.139
Stress	32	-0.29	-0.72	1.8	.082	0.65	1.03	2.7	.012	-0.26	.149	-0.27	.936

Tabelle 7 (Fortsetzung)

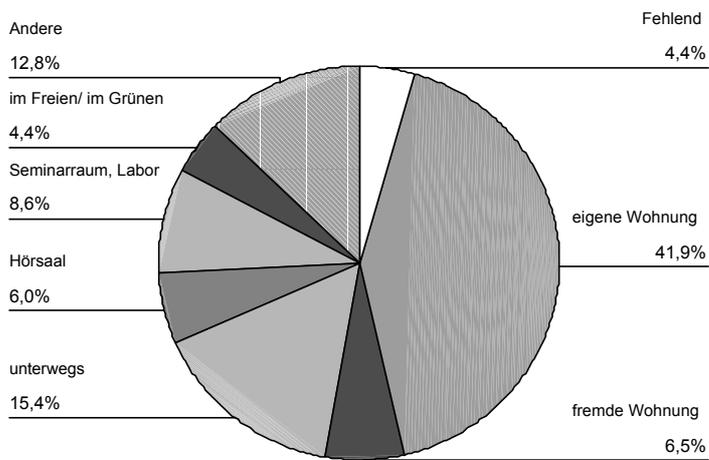
Variable	N	Mittelwerte				Standardabweichungen				Korrelation		Intraclass-K.	
		akt.	retro	t	p	akt.	retro	t	p	r _{ar}	p	r _{IC}	p
Tag 6 mit 7													
Aktiv	32	0.01	-0.14	0.7	.466	0.95	0.83	0.7	.468	0.18	.331	0.18	.152
Anstrengend	32	-1.01	-1.04	0.2	.871	1.07	1.18	0.6	.531	0.46	.007	0.47	.002
Aufgeregt	32	-0.37	-0.45	0.4	.661	1.05	0.90	0.9	.351	0.46	.008	0.46	.003
Ärgerlich	32	-0.25	-0.28	0.2	.879	0.96	0.82	0.9	.363	0.29	.102	0.30	.041
Bedrückt	32	-0.21	-0.44	1.4	.182	0.85	0.75	0.7	.470	0.35	.051	0.33	.028
Erschöpft	32	-0.28	-0.07	1.3	.206	0.66	0.69	0.2	.808	0.07	.708	0.06	.373
Gute Stimm	32	0.04	-0.08	0.7	.486	0.69	0.75	0.5	.621	0.08	.659	0.09	.311
Körp.wohl	32	0.27	-0.04	2.1	.040	0.75	0.83	0.6	.575	0.49	.004	0.44	.005
Stress	32	-0.64	-0.72	0.3	.742	0.89	1.03	0.8	.439	-0.09	.609	-0.08	.669

Tabelle 8: Varianzkomponenten der Differenzen von aktuellen und retrospektiven Einstufungen, zweifaktorielle Varianzanalyse (33 Personen, 7 Tage).

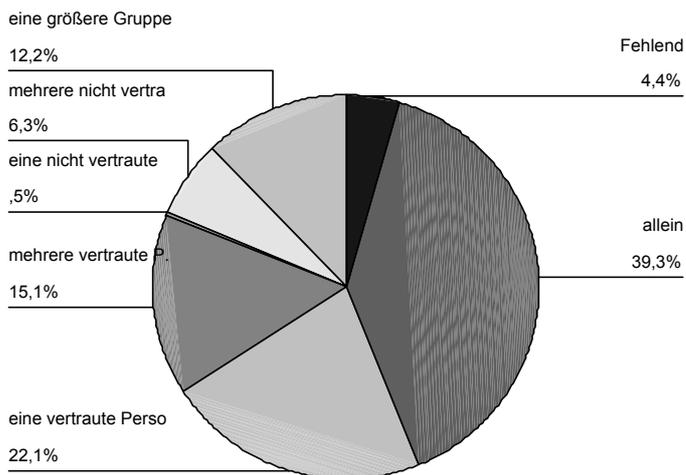
Variable	Personen		Tage		Fehler- Varianz- komponente (%)
	Varianz- komponente (%)	p	Varianz- komponente (%)	p	
Aktiv, leistungsfähig	18.6	.00004	0.0	.75792	81.4
Anstrengend, belstend	18.4	.00003	1.8	.11287	79.8
Aufgeregt, nervös	20.6	.00001	0.0	.44624	79.4
Ärgerlich, gereizt	15.0	.00044	0.5	.31259	84.5
Bedrückt	15.9	.00025	0.0	.44425	84.1
Erschöpft, müde	11.7	.00366	0.9	.23873	87.4
Guter Stimmung	9.2	.01541	0.0	.47698	90.8
Körperlich wohl	7.5	.03607	0.0	.71749	92.5
Stress	13.4	.00135	0.2	.37464	86.4

Aktuelle Untersuchung Setting-Merkmale (N = 33)

Wo befinden Sie sich gerade (Tag 2)



Welche Personen sind im Moment anwesend? (Tag 2)



Was haben Sie gerade vor dem Piepsen getan? (Tag 2)

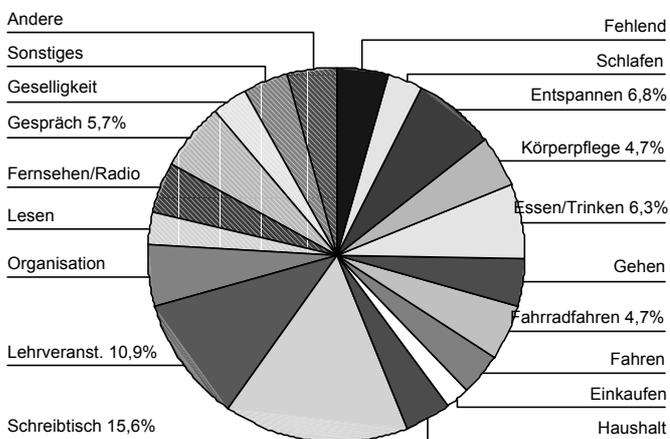


Tabelle 1: Basisstatistiken (N = 64, 6 Termine des 2. Tags)

		aktiv/ leistungsfähig	anstrengend/ belastend	Situation im Griff	aufgeregt/ nervös	ärgerlich/ gereizt	Stimmung bedrückt	körperlich wohl	guter Stimmung	Wie stark erlebten Sie seit der letzten Eingabe Stress?
Aktuelle Einstufungen										
N	Gültig	367	367	367	367	367	367	367	367	367
	Fehlend	17	17	17	17	17	17	17	17	17
Mittelwert		11,50	4,99	14,55	3,81	2,85	3,75	12,98	12,93	5,95
Standardabweichung		5,02	5,37	4,84	4,76	4,12	4,66	4,63	4,65	5,50
Varianz		25,15	28,79	23,46	22,66	16,96	21,73	21,46	21,63	30,30
Schiefe		-,310	,976	-,926	1,425	1,788	1,302	-,656	-,460	,836
Kurtosis		-,850	-,262	,208	1,072	2,783	,700	-,086	-,453	-,286
Retrospektive Einstufungen										
N	Gültig	93	93	93	93	93	93	93	93	93
	Fehlend	291	291	291	291	291	291	291	291	291
Mittelwert		12,02	8,43	13,70	4,67	4,04	4,44	12,14	12,37	
Standardabweichung		4,33	5,55	4,11	4,21	4,52	4,32	4,36	4,44	
Varianz		18,78	30,75	16,91	17,72	20,45	18,68	18,97	19,73	
Schiefe		-,506	,203	-,434	1,043	1,398	,795	-,559	-,544	
Kurtosis		-,383	-,960	-,530	,292	1,416	-,491	,292	-,219	

Anmerkung: 32 Teilnehmer gaben den Rückblick am Abend für den ganzen Tag, 32 gaben den Rückblick mittags und abends, so dass maximal 96 Rückblicke vorliegen.

Tabelle 2: Prüfung der Ausgangswert-Abhängigkeit zwischen retrospektiven und aktuellen Eingaben (über je 3 bzw. 6 Abfragen eines Tages gemittelt), sowie Prüfung des Retrospektions-Effekts durch drei verschiedene Veränderungs-Scores (Differenz, wahre Differenz, ALS).

VARIABLE	Aktiv	An- streng.	Sit. im Griff	Auf- geregt	Ärger- lich	Bedrückt	Körp. wohl	Guter Stimmg.
Vormittags-Retrospektion (N=29)								
Mittel aktuelle Eingaben	11.09	4.29	13.59	3.59	3.30	3.70	12.09	11.79
Std. aktuelle Eingaben	3.39	3.26	3.88	3.42	3.27	3.35	3.79	3.89
Std. Fehler aktuelle Eing.	2.73	2.59	2.44	1.93	2.32	1.95	2.09	2.27
Stabilität aktuelle Eing.	0.35	0.37	0.60	0.68	0.50	0.66	0.69	0.66
Mittel Halbtagesretrospekt.	12.14	6.41	13.45	3.83	4.76	4.52	12.17	12.31
Std. Halbtagesretrospektion	4.72	5.21	3.87	3.55	4.86	4.35	4.15	4.46
Std. Fehler Halbtagesretrosp.	-3.84 ¹	-3.31 ¹	2.14	1.56	-2.46 ¹	0.53	-1.01 ¹	2.77
Anzahl Personen	29	29	29	29	29	29	29	29
Korrelation aktuell*retro	0.76	0.72	0.65	0.74	0.79	0.81	0.86	0.64
Kovarianz aktuell*retro	12.21	12.21	9.70	8.98	12.58	11.73	13.48	11.05
λ	1.98	1.63	0.77	0.65	1.12	0.07	0.23	1.49
Steigung 1. Hauptachse	3.03	3.12	1.07	1.13	2.36	1.59	1.35	1.11
F-Wert für AWA	28.02	21.52	0.09	0.46	22.86	2.63	3.70	0.18
p-Wert für AWA	0.00	0.00	0.77	0.50	0.00	0.12	0.07	0.67
Regressionskoeffizient	1.06	1.15	0.65	0.77	1.17	1.05	0.94	0.73
Std Differenz	3.05	3.66	3.25	2.51	3.03	2.58	2.15	3.60
t-Wert für Differenz	-1.81	-3.07	0.23	-0.50	-2.55	-1.68	-0.20	-0.76
p-Wert für Differenz	0.08	0.00	0.82	0.62	0.02	0.10	0.84	0.45
Std wahre Differenz	3.89	4.43	2.95	2.40	3.71	2.88	2.33	3.46
t-Wert für wahre Diff.	-1.42	-2.54	0.26	-0.52	-2.08	-1.50	-0.18	-0.79
p-Wert für wahre Diff.	0.17	0.02	0.80	0.61	0.05	0.14	0.86	0.44
Std ALS-Wert	3.05	3.63	2.95	2.38	2.97	2.57	2.13	3.45
t-Wert für ALS	-1.82	-3.10	0.26	-0.52	-2.60	-1.68	-0.20	-0.79
p-Wert für ALS	0.08	0.00	0.80	0.60	0.01	0.10	0.84	0.43
Nachmittags-Retrospektion (N=29)								
Mittel aktuelle Eingaben	11.28	4.55	14.92	2.85	2.21	3.36	12.71	13.20
Std. aktuelle Eingaben	3.66	3.74	2.88	2.57	2.49	3.68	3.96	3.77
Std. Fehler aktuelle Eing.	2.11	2.43	2.27	1.99	1.79	2.04	1.85	2.10
Stabilität aktuelle Eing.	0.67	0.58	0.38	0.40	0.49	0.69	0.78	0.69
Mittel Halbtagesretrospekt.	11.31	9.14	13.10	4.79	3.76	4.48	11.66	11.52
Std. Halbtagesretrospektion	4.04	6.02	4.68	4.60	4.40	4.37	4.69	5.42
Std. Fehler Halbtagesretrosp.	3.16	2.82	-2.75 ¹	-0.68 ¹	-1.40 ¹	1.68	3.17	2.82
Anzahl Personen	29	29	29	29	29	29	29	29
Korrelation aktuell*retro	0.51	0.67	0.72	0.64	0.73	0.77	0.65	0.71
Kovarianz aktuell*retro	7.52	15.12	9.66	7.56	8.03	12.35	12.09	14.46
λ	2.23	1.35	1.47	0.12	0.62	0.68	2.92	1.81
Steigung 1. Hauptachse	0.84	1.87	3.05	2.86	2.66	1.32	0.99	1.48
F-Wert für AWA	0.22	7.37	20.57	5.46	16.09	2.68	0.00	3.89
p-Wert für AWA	0.65	0.01	0.00	0.03	0.00	0.11	0.97	0.06
Regressionskoeffizient	0.56	1.08	1.16	1.14	1.29	0.91	0.77	1.02
Std Differenz	3.83	4.47	3.30	3.56	3.09	2.82	3.67	3.82
t-Wert für Differenz	-0.04	-5.43	2.91	-2.89	-2.66	-2.11	1.52	2.33
p-Wert für Differenz	0.97	0.00	0.01	0.01	0.01	0.04	0.14	0.03
Std wahre Differenz	3.50	4.84	3.97	4.02	3.61	2.91	3.56	4.02
t-Wert für wahre Diff.	-0.04	-5.02	2.42	-2.55	-2.27	-2.03	1.57	2.22
p-Wert für wahre Diff.	0.97	0.00	0.02	0.02	0.03	0.05	0.13	0.03
Std ALS-Wert	3.47	4.46	3.27	3.54	3.00	2.80	3.56	3.82
t-Wert für ALS	-0.04	-5.45	2.94	-2.90	-2.73	-2.12	1.57	2.33
p-Wert für ALS	0.97	0.00	0.01	0.01	0.01	0.04	0.13	0.03

¹ Sonderfall bei Schätzung der Fehlervarianz des Verlaufswerts bewirkt $b=s_y^2/s_{xy}$

Tabelle 2 (Fortsetzung)

VARIABLE	Aktiv	An- streng.	Sit. im Griff	Auf- geregt	Ärger- lich	Bedrückt	Körp. wohl	Guter Stimmg.
Tages-Retrospektion (N=64)								
Mittel aktuelle Eingaben	11.48	4.96	14.58	3.79	2.84	3.71	12.95	12.90
Std. aktuelle Eingaben	3.04	3.54	2.96	3.24	2.68	3.30	3.22	3.20
Std. Fehler aktuelle Eing.	1.79	1.80	1.72	1.56	1.40	1.48	1.51	1.51
Stabilität aktuelle Eing.	0.65	0.74	0.66	0.77	0.73	0.80	0.78	0.78
Mittel Tagesretrospekt. ¹	12.16	8.73	13.89	4.83	3.95	4.41	12.24	12.57
Std. Tagesretrospektion ¹	4.10	5.16	3.83	3.93	4.33	4.20	4.22	3.95
Std. Fehler Tagesretrospe. ¹	2.39	3.87	2.13	2.47	1.42	1.52	1.92	2.01
Anzahl Personen	64.00	64.00	64.00	64.00	64.00	64.00	64.00	64.00
Korrelation aktuell*retro	0.66	0.57	0.68	0.68	0.81	0.83	0.79	0.76
Kovarianz aktuell*retro	8.18	10.43	7.67	8.69	9.38	11.57	10.70	9.59
λ	1.78	4.61	1.53	2.50	1.04	1.06	1.62	1.76
Steigung 1. Hauptachse	1.36	1.12	1.32	1.08	1.79	1.32	1.32	1.21
F-Wert für AWA	4.23	0.23	3.85	0.24	32.81	10.77	7.60	2.91
p-Wert für AWA	0.04	0.63	0.05	0.63	0.00	0.00	0.01	0.09
Regressionskoeffizient	0.89	0.83	0.87	0.83	1.30	1.06	1.03	0.94
Std Differenz	3.11	4.28	2.84	2.94	2.69	2.33	2.60	2.58
t-Wert für Differenz	-1.72	-6.98	1.91	-2.80	-3.25	-2.41	2.14	1.01
p-Wert für Differenz	0.09	0.00	0.06	0.01	0.00	0.02	0.04	0.31
Std wahre Differenz	3.18	4.25	2.89	2.89	2.99	2.47	2.72	2.63
t-Wert für wahre Diff.	-1.69	-7.03	1.88	-2.85	-2.92	-2.27	2.05	1.00
p-Wert für wahre Diff.	0.10	0.00	0.06	0.01	0.00	0.03	0.04	0.32
Std ALS-Wert	3.09	4.24	2.82	2.88	2.56	2.32	2.60	2.58
t-Wert für ALS	-1.74	-7.05	1.93	-2.86	-3.41	-2.42	2.15	1.02
p-Wert für ALS	0.09	0.00	0.06	0.01	0.00	0.02	0.04	0.31

¹ Bei Probanden mit Halbtags-Retrospektion Mittelwert von Vormittag und Nachmittag.

Anmerkung:

Die Tabellen enthalten im oberen Teil die Grundstatistiken der (gemittelten) aktuellen und der retrospektiven Einstufungen. Die Fehlervarianz der aktuellen Einstufungen wird aus der Innerhalb-Personen-Varianz berechnet (einfache Varianzanalyse), die Fehlervarianzen der gemittelten aktuellen Einstufungen vermindern sich um den Faktor 3 bzw. 6 (Anzahl der Summanden, d.h. Abfragen pro Halbtage/Tag). Die Fehlervarianzen der retrospektiven Einstufungen werden aus dem Modell geschätzt (Struktur-Regressions-Modell nach Kendall und Stuart, 1967, Case 1; s.a. Foerster, 1994, 1995). Im Mittelteil sind die Bestimmungsstücke zur Prüfung des Ausgangswertgesetzes (AWA) angegeben: Korrelation bzw. Kovarianz zwischen aktuellen und retrospektiven Eingaben, Fehlervarianz-Verhältnis λ , Prüfgröße b (Steigung der ersten Hauptachse) mit Teststatistik F und p-Wert für die Hypothese $b=1$. Im unteren Teil wird der Retrospektions-Effekt getestet, wobei drei unterschiedliche Veränderungswerte benutzt werden: (1) einfache Differenz von aktuellen minus retrospektiven Einstufungen; (2) die durch das Modell geschätzte wahre Differenz, die den echten Anteil der AWA, nicht jedoch deren statistischen Anteil enthält und sich nur durch ihre Varianz (Skalierung) von der einfachen Differenz unterscheidet; (3) der Autonomic Liability Score (ALS) von Lacey (1956), also das Residuum des Verlaufswerts nach Auspartialisieren des Ausgangswert.