

Forschungsberichte aus dem Projekt:

"Qualität des Erlebens in Arbeit und Freizeit"

Untersuchungen mit der Experience Sampling Method

Nr. 6

Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS)

Theoretische und methodische Grundlagen, Konstruktvalidität und psychometrische
Eigenschaften bei der Beschreibung intra- und interindividueller Unterschiede

Urs Schallberger

2005



Psychologisches Institut der Universität Zürich

Fachrichtung Angewandte Psychologie

Universitätstr. 84, CH-8006 Zürich

Die in dieser Reihe erscheinenden Arbeits- bzw. Forschungsberichte stammen aus dem Projekt „Qualität des Erlebens in Arbeit und Freizeit. Untersuchungen mit der Experience Sampling Method“, das in der Anlaufzeit vom Schweizerischen Nationalfonds zur Förderung wissenschaftlicher Forschung finanziert wurde (Kredit Nr. 11-50903.97). Leitung des Projekts: Prof. Dr. Urs Schallberger

Schallberger, U. (2005). *Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS). Theoretische und methodische Grundlagen, Konstruktvalidität und psychometrische Eigenschaften bei der Beschreibung intra- und interindividueller Unterschiede*. Forschungsberichte aus dem Projekt „Qualität des Erlebens in Arbeit und Freizeit“, Nr. 6. Zürich: Fachrichtung Angewandte Psychologie des Psychologischen Instituts der Universität.

Korrespondenzadresse:

Prof. Dr. U. Schallberger
Universität Zürich
Psychologisches Institut/Angewandte Psychologie
Universitätstr. 84, CH-8001 Zürich
Email: schallberger@angpsy.unizh.ch

Inhaltsübersicht:

Zusammenfassung	5
1. Einleitung: Problemrahmen, Thema und Vorblick	7
2. Theoretische und methodische Grundlagen	9
2.1 Spezifische Anforderungen an Skalen und Items für ESM-Studien	9
2.1.1 Die Quellen spezifischer Anforderungen	9
2.1.2 Konsequenzen für die Item- und Skalenkonstruktion	10
2.1.3 Konsequenzen für die Skalenüberprüfung	12
2.1.4 Bilanz	12
2.2 Die anvisierten Konstrukte: Bedeutung und Probleme	12
2.2.1 Positive und Negative Aktivierung (PA und NA)	12
2.2.2 Die Beziehung von PA und NA zur Valenz (VA)	15
2.3 Die "Positive and Negative Affect Schedule" (PANAS) und Probleme ihrer Anwendung in ESM-Studien	18
2.4 Die PANAVA-KS: Entwicklungsgeschichte und Leitgesichtspunkte	20
2.5 Ausblick auf die empirischen Untersuchungen	24
3. Studie 1: Konstruktvalidierung der PANAVA-KS durch einen Vergleich mit PANAS, AD-ACL und MDBF-GS in einer Ein-Punkt-Untersuchung	25
3.1 Zielsetzungen und Untersuchungskonzeption	25
3.2 Datengrundlage	26
3.3 Ergebnisse	27
3.3.1 Interne Konsistenz der Skalen im Vergleich	27
3.3.2 Konvergente Korrelationen und die Rolle des Itemformats	28
3.3.3 Explorative Faktorenanalyse aller acht Skalen	30
3.3.4 Konfirmatorische MTMM-Analyse der PA/EA- und NA/TA-Skalen	31
3.3.5 Konfirmatorische Analysen zur Beziehung der Valenzskalen zum PA- und NA-Faktor	35
3.4 Zusammenfassung und Bilanz aus Studie 1	38
4. Studie 2: Untersuchungen zu psychometrischen Eigenschaften der PANAVA-KS in ESM-Daten	41
4.1 Zielsetzungen und Vorblick	41
4.2 Die Datengrundlage	43
4.3 Augenscheinvalidität und Akzeptanz der PANAVA-Items	45
4.4 Verteilungskennwerte und Intraclasskorrelationen der Items und Skalen	45
4.5 Explorative Faktorenanalysen der Items auf der Zeitpunkt- und der Personenebene	47
4.6 Untersuchungen zur Reliabilität der Skalen auf der Zeitpunkt- und der Personenebene	50
4.6.1 Vorbemerkungen	50
4.6.2 Mehrebenenanalytische Konsistenzkoeffizienten	50
4.6.3 Interindividuelle Unterschiede in der Konsistenz auf der Zeitpunktebene (intraindividuelle Konsistenzen)	51
4.6.4 Retest-Reliabilitäten bzw. Stabilitätskoeffizienten auf der Personenebene	53
4.7 Interkorrelationen der Skalen auf der Zeitpunkt- und der Personenebene	55

4.8	Zusatzuntersuchungen zu den Interkorrelationen und zur differentiellen Validität der PANAVA-KS auf der Personenebene	58
4.8.1	Eine Untersuchung zum Aggregationseffekt	59
4.8.2	Zur differentiellen Validität der PANAVA-KS auf der Personenebene	62
4.9	Tages-, Wochen- und Jahresrhythmen in den PANAVA-KS	65
4.9.1	Einleitung	65
4.9.2	Circadiane Rhythmen	66
4.9.3	Wochenrhythmen	70
4.9.4	Jahresrhythmen	71
4.9.5	Schlussbemerkungen	72
4.10	Zusammenfassung und Bilanz aus Studie 2	72
5.	Rückblick und Ausblick	75
	Literaturverzeichnis	81
	Anhang A: Mehrebenenanalytische Modelle	85
A1	Aufteilung der Varianzen, Kovarianzen und Korrelationen auf die Person- und die Zeitpunktebene	85
A2	Das Modell von Raudenbush et al. (1991) zur Bestimmung ebene-spezifischer Konsistenzkoeffizienten	85
	Anhang B: Zusatztabelle	87

Zusammenfassung:

Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS). Theoretische und methodische Grundlagen, Konstruktvalidität und psychometrische Eigenschaften bei der Beschreibung intra- und interindividueller Unterschiede

Psychologische Skalen, die für den Einsatz in Experience Sampling (ESM) Studien geeignet sein sollen, müssen sowohl hinsichtlich Konstruktmerkmalen als auch hinsichtlich psychometrischer Überprüfung eine Reihe spezifischer Anforderungen erfüllen, die einleitend dargestellt werden. Die PANAVA-Kurzskalen (PANAVA-KS) wurden speziell im Hinblick auf diese Anforderungen entwickelt. Nach der Beschreibung theoretischer Probleme, die mit den drei erfassten Konstrukten Positive Aktivierung (PA), Negative Aktivierung (NA) und Valenz (VA) verbunden sind, werden zwei empirische Untersuchungen vorgestellt: In Studie 1 geht es um eine konfirmatorische Überprüfung der Konstruktvalidität der drei neuen Kurzskalen im Vergleich zu fünf konventionellen Skalen im Rahmen einer Ein-Punkt-Untersuchung ($N = 269$). Als Vergleichsinstrumente dienen die zwei Skalen des PANAS (Watson et al., 1988), die zwei Skalen der AD-ACL (Thayer, 1989) und die Skala MDBF-GS (Steyer et al., 1997). In Studie 2 werden verschiedenartige psychometrische Eigenschaften der PANAVA-KS in ESM-Daten untersucht ($N = 530$ Personen mit total $N = 23'509$ Messungen). Die hierarchische Struktur solcher Daten wird in der Form eines Zwei-Ebenen-Modells (Zeitpunktebene verschachtelt in der Personenebene) berücksichtigt und die Analysen werden – soweit sinnvoll – parallel auf beiden Ebenen durchgeführt. Thematisch geht es dabei um die Akzeptanz der PANAVA-KS, die Verteilungseigenschaften der Items- und Skalen, deren Intraclass-Korrelationen, die Faktorenstruktur der Items, die Skalenreliabilität (Konsistenz- und Retest-Koeffizienten), die Skaleninterkorrelationen, Aggregationseffekte, die differenzielle Validität der drei Skalen sowie die Frage nach zeitlichen Zyklen. Insgesamt zeigt sich, dass die bei der Entwicklung der PANAVA-KS verfolgten Ziele weit gehend erreicht werden konnten: Die Skalen stossen in ESM-Studien auf eine gute Akzeptanz; sie weisen auf beiden untersuchten Ebenen befriedigende bis sehr gute psychometrische Kennwerte auf, und die konfirmatorische Analyse spricht für eine – gemessen an den Vergleichsskalen – hohe Konstruktvalidität.

1. Einleitung: Problemrahmen, Thema und Vorblick

Methoden, die auf die häufig wiederholte Gewinnung von Daten direkt aus dem natürlichen Erlebens- und Verhaltensstrom von Personen in ihrem natürlichen Lebenskontext abzielen, erfreuen sich in der psychologischen Forschung zunehmender Beliebtheit. Formulierungen wie "capturing life as it is lived" (Bolger, Davis & Rafaeli, 2003) oder "a window into daily life" (Feldman Barrett & Barrett, 2001) versuchen, das spezifische Erkenntnisinteresse auszudrücken, das damit verbunden ist. Dasselbe gilt für die verschiedenartigen Bezeichnungen, die für solche Datensammelungsstrategien verwendet werden, zum Beispiel (in alphabetischer Reihenfolge) Ambulatory Assessment (Fahrenberg & Myrtek, 2001), Ecological Momentary Assessment (EMA; Stone, Shiffman & DeVries, 1999), Experience Sampling Method (Csikszentmihalyi, Larson and Prescott, 1977), Field Assessment (Pawlik, 1988), Self-monitoring (Perrez & Reicherts, 1996), Time Sampling Diary (Brandstätter & Elias, 2001) etc. Auch wenn mit diesen verschiedenen Begriffen teilweise Vorgehensunterschiede verbunden sind¹, etwas haben die entsprechenden Verfahren gemeinsam: Im Unterschied zu konventionellen Ein-Punkt-Untersuchungen wird von den Untersuchungspersonen (UPn) erwartet, dass sie über eine längere Untersuchungsperiode meist täglich mehrmals Daten über den je aktuellen Moment liefern. Insofern es dabei – unter Umständen neben anderen Datenarten – in aller Regel auch um Daten über das aktuelle Erleben und Befinden geht, scheint es im Kontext der nachstehend beschriebenen Problemstellung gerechtfertigt, den gut eingebürgerten Begriff "*Experience Sampling Method*" (*ESM*) in einem weiten Sinne des Wortes als Oberbegriff zu verwenden.

Ausgangspunkt der Forschungsarbeit, die in diesem Papier präsentiert wird, ist ein Problem, das rasch einmal bewusst wird, wenn man eine dertige Studie plant. Es besteht darin, dass psychologische Skalen, die für konventionelle Untersuchungen entwickelt und überprüft worden sind, in der Regel für ESM-Studien nicht geeignet sind. Entsprechend wird in der heutigen ESM-Forschung meist mit Ad-hoc-Operationalisierungen der interessierenden Konstrukte gearbeitet, und zwar selbst dann, wenn für die betreffenden Konstrukte psychologische Standardinstrumente zur Verfügung stehen – mit den bekannten ungünstigen Folgen für die Vergleichbarkeit bzw. Kumulierbarkeit der gewonnenen Einsichten aus verschiedenen Studien. Diese unbefriedigende Situation kann offensichtlich nur überwunden werden, wenn eine neue Generation von Skalen zur Verfügung steht, welche den spezifischen Anforderungen von ESM-Studien genügen *und* deren Konstruktvalidität – insbesondere deren Beziehung zu allenfalls vorhandenen Standardskalen – geklärt ist.

Gegenstand dieses Papiers ist ein Instrument, bei dessen Entwicklung versucht wurde, die genannten Forderungen einzulösen. Es entstand im Rahmen einer Serie von ESM-Studien bei berufstätigen Erwachsenen, mit der 1994 begonnen wurde (Schallberger, Langemann, Sperisen

¹ Quellen solcher Unterschiede ergeben sich vor allem aus dem intraindividuell angewandten Stichprobenplan (z.B. Zeit- vs. Ereignisstichprobe) und dem eventuellen Miteinbezug von objektiven Daten (Reaktionszeiten, Leistungsdaten und psychophysiologische Messungen; vgl. Fahrenberg & Myrtek, 1996, 2001).

& Spruit, 1995). Es dient der Erfassung von drei Konstrukten. Im Zentrum stehen die Konstrukte *Positive (PA)* und *Negative Aktivierung (NA)* im Sinne von Watson und Tellegen (1985) bzw. Watson, Wiese, Vaidya und Tellegen (1999). Das Standardinstrument zur Erfassung dieser Konstrukte ist die "Positive and Negative Affect Schedule" (PANAS; Watson, Clark & Tellegen, 1988), die das heute wohl bekannteste und am meisten angewandte Instrument zur dimensional Beschreibung der affektiven Befindlichkeit darstellt (Schumacher, Klaiberg & Brähler, 2003). So fanden Egloff, Schmuckle, Burns, Kohlmann und Hock (2003) kürzlich im Social Sciences Citation Index über 1600 Zitationen dieses Instruments. Wie aber noch zu zeigen sein wird, weist die PANAS verschiedene Eigenschaften auf, die für einen Einsatz in ESM-Studien ungünstig sind. Bei der Entwicklung einer Alternative, bei der die spezifischen Anforderungen von ESM-Studien zu berücksichtigen versucht wurden, wurde zwar am PANAS angeknüpft. Zusätzlich wurde aber noch eine dritte Dimension einbezogen, nämlich die *klassische Valenz (VA)*; Russell, 1980; auch gute-schlechte Stimmung, hedonischer Ton, pleasantness etc. genannt). Aus den entsprechenden Entwicklungsarbeiten resultierten nach mehreren Vorformen (vgl. Schallberger, 1999) die nach den Abkürzungen der drei Dimensionen benannten *PANAVA-Kurzskalen (PANAVA-KS)*. Sie umfassen total 10 Items, die sich im Rahmen einer ESM-Studie in ca. 30 bis 40 Sekunden bearbeiten lassen.²

Die *Grobstruktur des Papiers* sieht wie folgt aus: Zunächst (Abschnitt 2) werden die theoretischen und methodischen Grundlagen der PANAVA-KS dargestellt sowie – vor diesem Hintergrund – die Skalen selbst und ihre Entwicklung beschrieben. Dann folgen zwei empirische Studien: Im Zentrum der Studie 1 (Abschnitt 3) steht die Frage nach der Konstruktvalidität der PANAVA-KS. Dabei interessiert naturgemäss vor allem die Beziehung zur PANAS, aber auch jene zu anderen, theoretisch sehr verwandten Befindlichkeitsskalen. In Studie 2 (Abschnitt 4) werden die psychometrischen Eigenschaften der PANAVA-KS in ESM-Daten untersucht, wobei der spezielle Charakter solcher Daten – jeder Datenpunkt beschreibt gleichzeitig intra- und interindividuelle Unterschiede – dazu führt, dass die entsprechenden Analysen wesentlich vielschichtiger ausfallen als dies bei konventionellen Skalenanalysen der Fall ist. Das Papier endet mit einem Rück- und Ausblick (Abschnitt 5).

² Diese Skalen wurden (zum Teil allerdings in Vorformen) bereits in mehreren ESM-Untersuchungen mit verschiedenartigen Untersuchungsgruppen und Fragestellungen eingesetzt (z.B. Aellig, 2003; Engeser, 2005; Pfister, 2002; Rau & Riedel, 2004; Rheinberg, Manig & Vollmeyer, in Vorb.; Schallberger & Pfister, 2001; Triemer, 2003; Triemer & Rau, 2001a, 2001b; Venetz, Schmid & Schallberger, 2001).

2. Theoretische und methodische Grundlagen

Im Rahmen der folgenden Besprechung der Grundlagen der PANAVA-KS geht es um vier Themen: Erstens ist die in der Einleitung geäußerte Behauptung, ESM-Studien würden spezifische Anforderungen an Items und Skalen stellen, zu erläutern (Abschnitt 2.1). Zweitens werden die in den PANAVA-KS operationalisierten Konstrukte eingeführt, wobei auch auf einige der in der Literatur diskutierten offenen Fragen zu diesen Konstrukten einzugehen ist (Abschnitt 2.2). Drittens wird die PANAS, die am Ausgangspunkt der Entwicklung der PANAVA-KS stand, vorgestellt und ihre Probleme bei der Verwendung in ESM-Studien erörtert (Abschnitt 2.3). Vor diesem Hintergrund wird dann schliesslich die Konzeption und die Entwicklungsgeschichte der PANAVA-KS beschrieben (Abschnitt 2.4) und das Anliegen der beiden nachfolgenden empirischen Studien (Abschnitt 2.5) präzisiert.

2.1 Spezifische Anforderungen an Skalen und Items für ESM-Studien

2.1.1 Die Quellen spezifischer Anforderungen

Um die besonderen Anforderungen einer ESM-Studie an Skalen und Items zu erkennen, ist es nützlich, sich den Ablauf einer derartigen Studie konkret zu vergegenwärtigen. Als Beispiel sei der Ablauf einer ESM-Studie im engeren Sinne des Wortes (Csikszentmihalyi et al., 1977) geschildert. Es handelt sich dabei um ein signalkontingentes Zeitstichprobenverfahren: Die Untersuchungspersonen (UPn) werden über einen längeren Zeitraum (z.B. eine oder mehrere Wochen) mehrmals täglich (z.B. 3- bis 10-mal) an zufällig ausgewählten Zeitpunkten durch ein Signal (z.B. eines Pagers) aufgefordert, das Setting und das eigene Erleben/Verhalten im betreffenden Moment in standardisierter Weise zu registrieren. Das entscheidende Hilfsmittel ist dabei die so genannte "*Experience Sampling Form*" (ESF), d. h. der auf Papier oder mittels eines Taschencomputers (neuerdings auch per Handy, vgl. Collins, Kashdan & Gollnisch, 2003) präsentierte Fragebogen.

Vergleicht man diese Untersuchungsmethode mit konventionellen Ein-Punkt-Erhebungen (oder einzelnen zeitlich distanten Messwiederholungen) ergeben sich Unterschiede in zwei ganz verschiedenen Hinsichten, die für den Problembereich der Skalenkonstruktion von Relevanz sind: Erstens impliziert die Teilnahme an einer ESM-Studie offensichtlich eine wesentlich grössere *Belastung* der UPn. Sie haben in relativ kurzen Zeitabständen immer wieder den normalen Tagesablauf zu unterbrechen, um dieselben Fragen zu bearbeiten. Zweitens resultiert aus einer ESM-Untersuchung eine wesentlich komplexere *Datenstruktur*. Indem jede Person Angaben zu "vielen" Zeitpunkten macht, sind in jedem Datenpunkt *mindestens* zwei Varianzquellen kontaminiert, die hierarchisch ineinander "verschachtelt" (nested) sind: Die eine Varianzquelle wird durch die *intraindividuellen* Unterschiede (Unterschiede in den States eine Person) konstituiert, die zweite durch *interindividuelle* Unterschiede im Antwortverhalten (personabhängige Charak-

teristika wie Niveau, Streuung etc. der Antworten).³ Das heisst, die Daten weisen eine typische *Mehrebenenstruktur* auf: Auf der ersten (untersten) Ebene, im Folgenden *Zeitpunktebene* genannt, interessieren intraindividuelle Unterschiede, auf der zweiten Ebene, im Folgenden *Personenebene* genannt, interindividuelle Unterschiede. Eine zusätzliche Komplikation bildet dabei, dass die Daten der Zeitpunktebene den Charakter von *Zeitreihen* haben.

Diese beiden spezifischen Charakteristika einer ESM-Untersuchung haben zwei Typen von Konsequenzen: Die einen betreffen die Kriterien der Item- und Skalenkonstruktion, die andern die Problemlage bei der Skalenüberprüfung. Diese beiden Themen sollen nun je einzeln besprochen werden.

2.1.2 Konsequenzen für die Item- und Skalenkonstruktion

Konsequenzen für die Item- und Skalenkonstruktion ergeben sich vor allem aus der oben erwähnten ersten Besonderheit von ESM-Untersuchungen, nämlich aus der Tatsache, dass sie für die UPn eine beträchtliche Belastung darstellen. Diese Belastung wirkt sich in zwei Hinsichten aus: Zum einen spielt sie – als antizipierte Belastung – bereits beim Entscheid einer Person, an der Studie teilzunehmen, eine wesentliche Rolle. Zum andern stellt die erlebte Belastung im Verlauf der Untersuchung eine zentrale Determinante der Datenqualität dar, indem sie mitbestimmt, wie viele ESF ausgefüllt werden und wie sorgfältig dies geschieht (z.B. Gershuny, 2004). Beide Aspekte – erfolgreiche Rekrutierung von UPn und hohe Compliance, d.h. regelkonformes Teilnahmeverhalten – sind für den Erfolg einer ESM-Untersuchung aber zentral. Eine Grunderfordernis einer ESM-Studie ist somit, die (antizipierte bzw. erlebte) Belastung möglichst gering zu halten. Hauptdeterminanten dieser Belastung sind einerseits die Untersuchungsdauer und die Zahl der Signale pro Tag, die meist durch die Zielsetzung der Studie vorgegeben sind. Evidenterweise kommt andererseits aber auch dem Umfang und Inhalt der ESF, die immer wieder und in den verschiedenartigsten Alltagssituationen zu bearbeiten ist, eine Schlüsselrolle zu. Zwar fehlt noch eine systematische Auseinandersetzung mit den dabei relevanten Gesichtspunkten. Aufgrund vereinzelter Hinweise in der Literatur und der eigenen Erfahrungen aus ESM-Studien an bisher über 700 vorwiegend berufstätigen Erwachsenen lassen sich hauptsächlich die folgenden Kriterien ausmachen:

Die auffälligste Anforderung an eine ESF unter dem Aspekt der Belastung ist sicher, dass die erforderliche *Bearbeitungszeit* im Rahmen einer normalen Lebensführung noch bewältigbar ist (z.B. Bolger et al., 2003; Scollon, Kim-Prieto & Diener, 2003). Die kritische, noch zumutbare obere Grenze scheint (nach den Erfahrungen in unseren Studien) z.B. bei berufstätigen Erwachsenen bei etwa 3 bis höchstens 4 Minuten zu liegen. Wenn verschiedene Aspekte des je aktuellen Moments erfasst werden sollen, hat dies naturgemäss zur Folge, dass pro Themenbereich/

³ Neben diesen zwei Quellen bzw. Ebenen (Personen und Zeitpunkte) liessen sich leicht weitere einführen, z.B. durch Einbezug des Wochentags oder von Personengruppen. Die Berücksichtigung solcher weitergehender Differenzierungen würden aber die Zielsetzung dieses Papiers klar sprengen.

Konstrukt nur wenige Fragen bzw. Items möglich sind. Eine ESM-taugliche Skala darf damit nur wenige Items umfassen. An diesem Kriterium der *Ökonomie* scheitern bereits die meisten psychologischen Standardskalen; sie umfassen schlicht zu viele Items.

Dieses Kriterium der Kürze reicht aber noch nicht aus. Zusätzlich sollten die Items (bezogen auf alle möglichen Alltagssituationen) von den UPn als sinnvoll erlebt werden können – für die meisten Menschen eine notwendige Bedingung für die Motivation, die Fragen immer wieder gewissenhaft zu beantworten. Eine erste Voraussetzung dafür ist eine hohe *Augenscheinvalidität* der Fragen für die Beschreibung des Alltagslebens im Hinblick auf das deklarierte Untersuchungsziel. Weiter sollten (echte oder scheinbare) *Redundanzen* in den Items vermieden werden; sie werden von den UPn als lästig erlebt (Schimmack, 2003), führen u.U. zu Rückfragen bei der Untersuchungsleitung über den exakten Bedeutungsunterschied und/oder – besonders unter Zeitdruck – zu Auslassungen einzelner Items. Das Analoge gilt für Items, die kaum eine *Antwortstreuung* zulassen, d.h. im Alltag meist auf dieselbe Art und Weise beantwortet werden müssen (z.B. Items mit J- bzw. L-Verteilungen). Beide Typen von Items können in konventionellen Skalen durchaus eine Funktion erfüllen, z.B. zur Konsistenzhöhung (Fragen nach sehr ähnlichen Sachverhalten, aber in unterschiedlicher Formulierung) oder zur Erhöhung der Differenzierungsfähigkeit in einem bestimmten Skalenwertbereich (sehr "schwierige" oder sehr "leichte" Items). Durch die oft wiederholte Bearbeitung führen sie in einer ESM-Studie dagegen früher oder später zu einer Irritation der UPn – ganz abgesehen davon, dass sie mit ihrem beschränkten Informationsgehalt eigentlich eine Verschwendung des kostbaren Raums in einer – wie erwähnt – in jedem Fall kurz zu haltenden ESF darstellen.⁴

Die zweite Besonderheit von ESM-Studien – die komplexe *Datenstruktur* – hat ebenfalls eine Implikation für die Item- und Skalenkonstruktion. Sie betrifft die Rolle des *Antwortstils* (response style, z.B. Akquieszenz). Obwohl immer wieder kontrovers diskutiert, dominiert seit der klassischen Übersichtsarbeit von Rorer (1965) die Ansicht, dass dessen Rolle als Artefaktquelle in konventionellen Befragungen zumindest nicht überschätzt werden darf.⁵ In einer kürzlich erschienen Arbeit von Watson und Tellegen (2002) wird aber recht überzeugend demonstriert, dass dies für in ESM-Studien gewonnenen Skalenwerte auf der Personenebene nicht zu gelten scheint: Durch die Aggregation der vielen Messungen pro Person kann der in einer Einzelmessung vielleicht belanglose Antwortstil zu einer potenten Artefaktquelle werden. Dies bedeutet, dass bei Skalen für ESM-Studien vorsichtshalber darauf zu achten ist, dass der Antwortstil einer Person explizit oder implizit kontrolliert wird.

⁴ Zu diesen Regeln für die Itemkonstruktion seien zwei Dinge angemerkt: Erstens gibt es natürlich Ausnahmen: So kann z.B. das explizite (und deklarierte) Untersuchungsziel, ein seltenes Phänomen zu untersuchen, dazu führen, dass bewusst Items mit J-Verteilungen in Kauf genommen werden. Zweitens gelten sie – mutatis mutandis und mit dem gemachten Vorbehalt – auch dann, wenn pro Zeitpunkt jeweils nur eine Zufallsauswahl aller Items präsentiert wird.

⁵ Für eine Übersicht über diese Diskussion im Bereich der Affektmessung siehe z.B. Schimmack, Böckenholt und Reisenzein (2002).

2.1.3 Konsequenzen für die Skalenüberprüfung

Auch wenn die komplexe Struktur von ESM-Daten also bereits bei der Skalenkonstruktion berücksichtigt werden sollte (Antwortstilkontrolle), liegen ihre Hauptkonsequenzen in der Art und Weise, wie die psychometrische Qualität einer Skala zu überprüfen ist. Um informativ zu sein, muss sie – der Natur der ESM-Daten entsprechend – *mehrebenenanalytisch* angelegt sein, also Aussagen über die Bewährung der Skalen (mindestens) sowohl auf der Zeitpunktebene als auch auf der Personenebene enthalten. Es wird zwar seit langem immer wieder hervorgehoben, wird aber trotzdem oft nicht berücksichtigt, dass sich die Gesetzmässigkeiten zwischen den beiden Ebenen wesentlich unterscheiden können (vgl. z.B. Schmitz, 2000). Dies gilt naturgemäss auch für messtechnische Eigenschaften von Skalen. So betrachten es Bolger et al. (2003) als ein Hauptproblem der aktuellen ESM-Forschung, dass es bei den in solchen Untersuchungen (wegen der Nichteignung konventioneller Skalen) typischerweise eingesetzten Ad-hoc-Kurzskalen meist keine Gewähr dafür gebe, "... that these measures are reliable and valid for assessing within-person variability" (a.a.O., S. 608).

2.1.4 Bilanz

Zusammenfassend ergeben sich somit folgende spezifischen Anforderungen, die bei der Entwicklung von Skalen für ESM-Studien zu berücksichtigen sind: Sie sollten – *erstens* – möglichst *kurz* (ökonomisch) sein und Items umfassen, die möglichst *augenscheinvalide* sind, möglichst *wenig redundant* erscheinen und im Verlaufe des Alltagslebens eine *deutliche Streuung der Antworten* erlauben. Sie sollten zudem so formuliert sein, dass die Skalenwerte möglichst unabhängig vom individuellen *Antwortstil* sind. *Zweitens* sollte bei der Überprüfung der psychometrischen Qualität der Skalen der *Mehrebenenstruktur* von ESM-Daten explizit Rechnung getragen werden. Und *drittens* gilt schliesslich, dass – wie in der Einleitung ausgeführt – die *Beziehungen* solcher neuer Skalen zu allenfalls vorhandenen *Standardskalen* für dieselben bzw. für verwandte Konstrukte geklärt werden sollten.

2.2 Die anvisierten Konstrukte: Bedeutung und Probleme

Wie in der Einleitung erwähnt, knüpfte die Entwicklung der PANAVA-KS am Affekt-Modell von Watson und Mitarbeitern an, zu dem heute eine äusserst umfangreiche Literatur existiert. Es ist im Rahmen dieses Papiers weder möglich noch nötig, diese Literatur vollständig aufzuarbeiten. Für die Entwicklung der PANAVA-KS sind vor allem zwei Themen relevant: Die Bedeutung der Konstrukte Positive und Negative Aktivierung (PA und NA) sowie ihr Verhältnis zur Valenz (VA).

2.2.1 Positive und Negative Aktivierung (PA und NA)

Bei der Einführung der beiden Dimensionen PA und NA knüpften Watson und Tellegen (1985) explizit am "Circumplex-Modell of Affect" von Russell (1980) an (vgl. Abbildung 1). Danach

lassen sich affektive Konzepte – schematisch und vergrößernd - kreisförmig in einer Ebene anordnen. Als (weitgehend) orthogonale Grunddimensionen dieser Struktur favorisiert Russell (1980) die beiden klassischen Dimensionen valence (auch pleasantness, hedonischer Ton etc. genannt) und activation (auch arousal, engagement etc. genannt) (vgl. auch Russell & Carroll, 1999).

Auf der Basis eigener Reanalysen einer Vielzahl einschlägiger empirischer Datensätze argumentieren Watson und Tellegen (1985) hingegen, dass ein um ca. 45 Grad gedrehtes, ebenfalls orthogonales Achsenpaar eine wesentlich stabilere und robustere Faktorenstruktur ergibt (vgl. Abbildung 1). Diese neuen Achsen nannten sie *Positive Affect (PA)* und *Negative Affect (NA)*, wobei die spezifische Bedeutung dieser Begriffe zu beachten ist (vgl. insbesondere Larsen & Diener, 1992): PA bzw. NA stehen nicht etwa für alle positiven (= ganzer rechter Halbkreis) bzw. alle negativen (= ganzer linker Halbkreis) affektiven Zustände, sondern nur für höchst spezifische Subklassen *je* sowohl positiver *und* negativer Zustände: Hoher PA umfasst (nur) positive Zustände, die mit einem hohen Mass an Aktivierung verbunden sind (z.B. begeistert). Ihnen stehen – als niedriger Pol von PA – negative Zustände gegenüber, die durch eine geringe Aktivierung gekennzeichnet sind (z.B. gelangweilt). Analog steht hoher NA (nur) für hoch aktivierte negative Zustände (z.B. nervös); der Gegenpol – niedriger NA – wird von positiven Zuständen mit geringer Aktivierung gebildet (z.B. entspannt).⁶

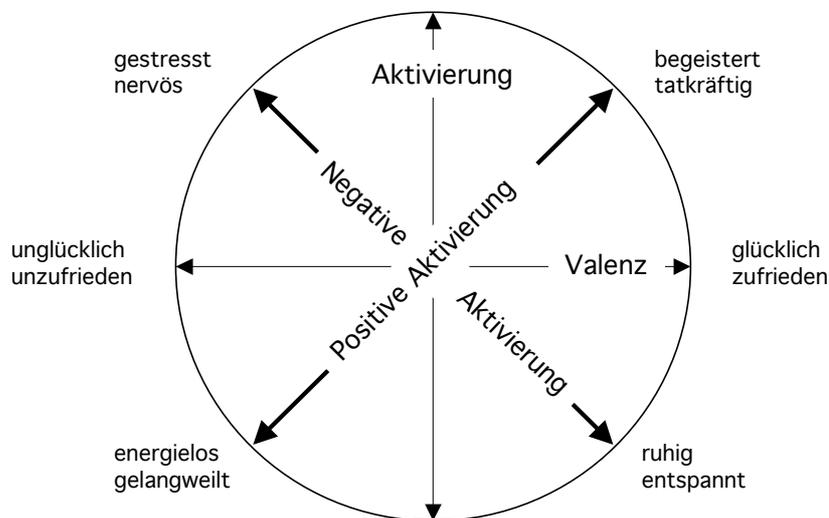


Abbildung 1. Das Circumplex-Modell affektiver Zustände (adaptiert nach Watson und Tellegen, 1985; die Markiertvariablen sind den PANAVA-KS entnommen)

Diese komplexe Bedeutung der Dimensionen PA und NA wurde und wird oft übersehen, was immer wieder zu Missverständnissen über die Implikationen des Modells für die Frage der Un-

⁶ Im deutschen Sprachraum existiert ein sehr ähnliches Modell, das von Abele-Brehm und Brehm (1986) stammt und weniger missverständliche Begriffe benutzt: PA entspricht dort "Positive Spannung vs. negative Lösung", NA "Negative Spannung vs. Positive Lösung".

abhängigkeit (independence) vs. Bipolarität positiver und negativer Affekte *im Allgemeinen* geführt hat. Das Modell impliziert, dass diese Frage so nicht beantwortbar ist: PA und NA sind zwar als orthogonal konzipiert, sind also in diesem Sinne voneinander unabhängig; gleichzeitig umfassen beide Dimensionen in sich aber je einen positiven und einen negativen Pol, sind also (zumindest deskriptiv, siehe unten) je bipolar.

Diese Missverständlichkeit der Begriffe bildete einer der Gründe für die 1999 erfolgte Umbenennung der beiden Dimensionen in *Positive* und *Negative Aktivierung* (Activation; Tellegen, Watson & Clark, 1999a, S. 298). Der andere, nach Watson et al. (1999, S. 827) wichtigere Grund ist eine theoretische Evolution im Verständnis der beiden Dimensionen. Danach sind PA und NA nicht nur faktorenanalytisch begründete, sparsame und robuste Beschreibungsdimensionen des affektiven Raumes, sondern "... reflect the operation of two broad, evolutionarily adaptive motivational systems that mediate goal-directed approach and withdrawal behaviors" (a.a.O., S. 829f.), was durch eine Reihe von theoretischen und empirischen Argumenten gestützt wird (a.a.O., S. 828-835), die nicht zuletzt auf die einflussreiche Konzeption dieser beiden Systeme von Gray (1982) Bezug nehmen. Watson et al. (1999) stehen mit dieser Interpretation nicht allein: Sehr ähnliche Argumentationen findet man neuerdings auch bei vielen anderen Autoren (z.B. Brockner & Higgins, 2001; Cacioppo, Gardner & Berntson, 1999; Carver, Sutton & Scheier, 2000; Gable, Reis & Elliot, 2003 u.a.). Für diese beiden Systeme werden verschiedene Bezeichnungen benutzt. Für eine Psychologie des Alltagslebens scheinen die Begriffe "*Annäherungs-*" und "*Vermeidungssystem*" besonders geeignet zu sein. Etwas vergröbernd ausgedrückt, ist das Annäherungssystem dann aktiv, wenn es um Ziele geht, die als solche belohnenden Charakter haben, das Vermeidungssystem hingegen dann, wenn Ziele darin bestehen, Konsequenzen, die bestrafenden Charakter haben, zu vermeiden.⁷ Aus dieser Sicht ist dann PA Ausdruck bzw. subjektiv erlebte Komponente der Aktiviertheit bzw. Desaktiviertheit des Annäherungssystems, NA analog Ausdruck bzw. die subjektive Komponente der Aktiviertheit oder Desaktiviertheit des Vermeidungssystems. Mit dieser funktionalen Deutung der affektiven Befindlichkeit wird – wie Watson et al. (1999, S. 827) hervorheben – auch offensichtlich, dass PA und NA eine enge Verwandtschaft mit den beiden Dimensionen *Energetic Arousal* (EA; energetische Aktivierung) und *Tense Arousal* (TA; Spannungsaktivierung) von Thayer (1989) aufweisen, eine Verwandtschaft, auf die Thayer schon vor der Umbenennung hingewiesen hat (z.B. Thayer, 1996). Die Untersuchung von Yik et al. (1999) spricht sogar für eine weit gehende empirische Identität der beiden Konzeptionen (siehe dazu auch unten, Studie 1).

Ein letzter Punkt, der hier angesprochen sei, betrifft die Frage nach der empirischen Beziehung von PA und NA. Während ursprünglich (vgl. Abbildung 1) von Orthogonalität ausgegangen wurde, sprechen Watson et al. (1999) auf Grund verschiedener Untersuchungen von einer "mo-

⁷ Diese Formulierung macht auch deutlich, dass dieselbe ("objektive") Handlung durch das eine oder andere System reguliert sein kann: Man kann ein Fachbuch lesen, weil das Thema spannend ist – oder weil man an der bevorstehenden Prüfung nicht durchfallen will. Es ist auch anzunehmen, dass in einer konkreten Handlung oft beide Systeme aktiviert sind (z.B. Carver et al., 2000).

derat negativen" Korrelation in der Höhe von rund $-.5 \leq r_{PA,NA} \leq -.3$ (Russell & Carroll, 1999, Table 5; Tellegen et al., 1999a; Watson et al., 1999). Die Befundlage ist aber alles andere als einheitlich (vgl. z.B. Schmuckle, Egloff & Burns, 2002). Der Hauptgrund für diese unübersichtliche Situation ist, dass Schätzungen der Korrelation von PA und NA von vielen untersuchungstechnischen Einzelheiten abhängen, die in verschiedenen Untersuchungen in unterschiedlicher Weise oder gar nicht berücksichtigt werden. In der einschlägigen Literatur diskutierte Varianzquellen der Stärke der Korrelation von PA und NA sind beispielsweise: Personenstichprobe, Operationalisierungen, Itemformate, Mitberücksichtigung möglicher Einflüsse von Antwortstilen, Analyse auf der Ebene manifester oder latenter Variablen, Zeitrahmen bei Ein-Punkt-Untersuchungen (z.B. Befinden im Moment, Befinden heute, Befinden im Allgemeinen), intra- vs. interindividuelle (state- vs. trait-) Ebene (oder Vermengung der beiden Ebenen) in ESM-Untersuchungen usw. Offensichtlich haben verschiedene dieser Faktoren primär eine methodische Bedeutung (z.B. Itemformate, Antwortstilkontrolle, Berechnung auf der Ebene manifester oder latenter Variablen), sprechen also gewissermassen Störfaktoren bei der empirischen Bestimmung "der wahren" Korrelation von PA und NA an. Bei anderen Faktoren (z.B. Zeitrahmen, intra- vs. interindividuelle Ebene) ist hingegen von Varianzquellen die Rede, die mit effektiven Unterschieden in der Korrelation von PA und NA verbunden sein könnten, was impliziert, dass es "die eine wahre" empirische Korrelation gar nicht gibt. Reich, Zautra und Kollegen bringen – ebenfalls in diesem Sinne – in ihrem "Dynamic Model of Affect" (DMA) noch weitere Varianzquellen ins Spiel, nämlich kontextuelle Gegebenheiten (z.B. Stress- vs. Nonstress-Situationen) und Persönlichkeitszüge (z.B. kognitive Simplizität bzw. Komplexität, Mood Clarity) (vgl. der Überblick in Reich, Zautra und Davis, 2003).

Zusammenfassend gesehen sprechen also viele Argumente dafür, dass die Orthogonalität der Beziehung von PA und NA in Abbildung 1 nicht "wörtlich" genommen werden darf, diese Abbildung somit als schematische Abstraktion zu betrachten ist. Zu beachten ist ferner, dass die oben beschriebene Neuinterpretation von PA und NA als Aktivierungsdimensionen die Konsequenz hat, dass die beiden Dimensionen auf der Konstruktebene eigentlich als unipolar zu verstehen sind (fehlende bis starke Aktivierung) und höchstens – weil für beide Pole sprachliche Marker existieren – als "*deskriptiv bipolar*" bezeichnet werden können (Watson et al., 1999). Diese Schlussfolgerung ist allerdings nicht unumstritten (z.B. Carver et al., 2000; Carver 2001; vgl. dazu Abschnitt 2.2.2, Auffassung E).

2.2.2 Die Beziehung von PA und NA zu Valenz (VA)

Auch in der Frage nach der Stellung von Valenz zu PA und NA darf Abbildung 1 aus heutiger Sicht nur als schematische Abstraktion verstanden werden. Zwei Gründe sind dafür verantwortlich: Erstens sprechen die empirischen Korrelationen gegen die in Abbildung 1 angenommene Symmetrie; VA scheint vielmehr wesentlich stärker (negativ) mit NA zusammenzuhängen als (positiv) mit PA. Drei prominente Untersuchungen, die sich unter anderem ausdrücklich dieser Frage widmen, liefern zum Beispiel auf der Ebene latenter Variablen für NA $-.96 \leq r_{NA,VA} \leq -.70$

und für PA $.49 \leq r_{PA,VA} \leq .65$ (Feldman Barrett & Russell, 1998, Table 9; Schimmack & Grob, 2000, Figure 2; Yik, Russell & Feldman Barrett, 1999, Table 8). Auch hier sind natürlich alle Einflussfaktoren relevant, die im vorigen Abschnitt aufgelistet wurden.⁸ Zu beachten ist ferner, dass in allen genannten Untersuchungen zur Affektstruktur nicht mit dem PANAS gearbeitet wird, sondern mit (je nach Untersuchung etwas anderen) Markiertvariablen. In der Untersuchung von Schimmack und Grob (2000) geht es zudem um zwei anders genannte Dimensionen (awake-sleepiness, tension-relaxation), die aber ausdrücklich mit jenen von Thayer (EA und TA) und implizit auch mit jenen von Watson et al. (1999) (PA und NA) parallelisiert werden.⁹

Die zweite Problematisierung der Darstellung der Position von VA in Abbildung 1 ergibt sich aus der neueren Diskussion um die Faktorenstruktur von PA, NA und VA. Hierzu existieren im Wesentlichen fünf verschiedene Auffassungen:

A) Das ursprüngliche, zweidimensionale Modell von Abbildung 1 (das z.B. von Russell und Mitarbeitern auch aktuell noch vertreten wird, z.B. Russell & Carroll, 1999) impliziert, dass die beiden Achsenpaare PA/NA und Valenz/Aktivierung im Prinzip austauschbare faktorenanalytische Konzeptualisierungen derselben Struktur darstellen. Dies impliziert, dass das eine Paar auf das andere Paar rückführbar ist bzw. – technisch gesprochen – die Dimensionen des einen Paares (z.B. VA) durch Doppelladungen im jeweils andern Paar (im Beispiel PA und NA) vollständig erklärt werden können.¹⁰ Empirische Tests dieser Hypothese liefern im allgemeinen in konfirmatorischen Faktorenanalysen aber einen auffällig schlechten Fit (z.B. Schimmack & Reisenzein, 2002; Watson et al., 1999).

B) Tellegen et al. (1999a; vgl. auch Tellegen et al., 1999b) favorisieren daher eine andere Sichtweise, nämlich eine hierarchische: Aus dieser Sicht repräsentiert VA – als Faktor dritter Ordnung – gewissermassen den Generalfaktor des affektiven Raumes. PA und NA stellen dagegen (die) zwei Faktoren zweiter Ordnung dar, während auf der untersten Ebene die einzelnen diskreten Emotionen lokalisiert werden. Die Plausibilität dieses hierarchischen Modells wird unter anderem mit exploratorischen Faktorenanalysen zu belegen versucht (a.a.O.).

C) Eine andere Lösung des Problems des schlechten Fits des in Abbildung 1 repräsentierten Modells schlagen einige Autoren vor, die an der kontinentaleuropäischen Tradition eines Modells mit drei Faktoren anknüpfen (vgl. Matthews, Jones & Chamberlain, 1990; Steyer et al., 1997; zusammenfassend: Schimmack & Grob, 2000). Dieses Modell umfasst neben der klassischen Valenzdimension zwei – wie bereits erwähnt – zwar anders benannte, aber als zu PA bzw.

⁸ Tellegen et al. (1999a 1999b) berichten z.B. von eher symmetrischen Beziehungen.

⁹ Es handelt sich um die im Mehrdimensionalen Befindlichkeitsfragebogen (MDBF; Steyer, Schwenkmezger, Notz & Eid, 1997) operationalisierten Dimensionen "wach-müde" und "ruhig-unruhig". Die genannte Parallelisierung steht übrigens in einem gewissen Gegensatz zu der vom Erstautor der genannten Studie früher (Schimmack, 1999) vertretenen Auffassung, PA und NA seien Mischungen aus je einer dieser beiden Aktivierungsdimensionen und Valenz.

¹⁰ Die Erklärbarkeit von VA durch EA und TA (die nach Watson et al., 1999, und Thayer – wie oben erwähnt – weitgehend PA und NA entsprechen) ergibt sich auch aus der Auffassung von Thayer (z.B. 1996).

EA und NA bzw. TA parallel aufgefasste Aktivierungsdimensionen. Wie nach den berichteten Befunden über die Beziehung von PA und NA zu VA nicht anders zu erwarten, ist dabei die Aktivierungsdimension, die NA entspricht, mit Valenz (z.T. extrem) hoch korreliert. Empirische Tests des Modells liefern jedoch einen sehr guten Fit (Schimmack & Grob, 2000).

D) Eine vierte Auffassung der Beziehung von PA, NA und VA hat nicht den Status eines expliziten Modells, sondern äusserst sich eher nebenbei in der Zusammenstellung von Items in Ad-hoc-Skalen. So wird z.B. das Item "Happy" – eigentlich ein prototypischer Indikator von Valenz – wohl wegen seiner Positivität gelegentlich der Dimension PA subsumiert (z.B. Williams, Suls, Alliger, Learner & Wan, 1991). Andererseits schlagen Autoren, die sich eher nach den empirischen Itemkennwerten ausrichten (die gemäss den oben berichteten Befunden meist deutlich asymmetrisch ausfallen), typische Valenzitems (umgepolt) NA zu (z.B. Triemer & Rau, 2001). Als faktorenanalytische Modellvorstellung formuliert, bedeutet dies eine Verschärfung von Modell A in dem Sinne, dass VA bereits durch je eine der Dimensionen PA und NA "erklärt" werden kann.

E) Eine fünfte Position in der hier diskutierten Frage findet sich bei Carver (2001, 2003). Ihr zentraler Gedanke ist, dass sich die Frage nach der "richtigen" Abbildung der Affektstruktur *allein* empirisch-faktorenanalytisch nicht sinnvoll beantworten lasse. Notwendig sei vielmehr der Einbezug von Überlegungen über die *funktionale Architektur* des Affektsystems. Dies liefert einen zusätzlichen Gesichtspunkt bei der theoretischen Strukturierung der Problemlage. Ausgangspunkt der entsprechenden Überlegungen ist, dass die – als je in sich echt (also nicht nur deskriptiv) – bipolar aufgefassten Dimensionen PA bzw. NA Ausdruck des subjektiv registrierten *Erfolgs* der Handlungen, die im Dienste des Annäherungs- bzw. Vermeidungssystems ausgeführt werden, darstellen (Carver et al., 2000).¹¹ Die Frage nach der Stellung der Valenz transformiert sich dann in jene nach der gemeinsamen Funktion der positiv valenzierten Pole von PA und NA (Carver, 2003). Und hier liegt der Gedanke nahe, dass VA eine Art übergeordnete, bilanzierend-evaluative Funktion zukommt, also *funktional* gewissermassen auf einer anderen Ebene liegt (vgl. auch Cacioppo & al., 1999; kritisch dazu: Russell, 2003). Im Vergleich zur hierarchischen Modellvorstellung (Modell B) verläuft hier damit die Abhängigkeitsrichtung in der gegenteiligen Richtung (von PA und NA zu VA), wobei in dieser funktionalen Sichtweise auch die Möglichkeit zeitverschobener Effekte in Betracht zu ziehen wäre.

Zusammenfassend ist also festzuhalten, dass – wie bereits die Beziehung von PA und NA untereinander (vgl. Abschnitt 2.2.1) – auch die Beziehungen von PA und NA zu VA als unklar zu betrachten sind. Abbildung 1 bleibt zwar eine für manche Zwecke nützliche grobe Illustration der gegenseitigen Beziehungen der Dimensionen PA, NA und VA. Weder die Frage nach der

¹¹ Es besteht also ein deutlicher Unterschied zur oben dargestellten Auffassung von Watson et al. (1999): Während dort PA und NA den *Grad der Aktiviertheit* der beiden Systeme widerspiegeln, widerspiegeln sie hier den *Erfolg der Handlungen*, die im Dienste der beiden Systeme ausgeführt werden.

Grösse der Winkel zwischen den Dimensionen noch die Frage nach der Dimensionalität des durch die drei Dimensionen aufgespannten Raumes ist aber definitiv geklärt.

2.3 Die Positive and Negative Affect Schedule (PANAS) und Probleme ihrer Anwendung in ESM-Studien

Wie in der Einleitung erwähnt, stellt die PANAS (Watson et al., 1988) das Standardinstrument zur Erfassung der Positiven und Negativen Aktivierung dar. Sie entstand auf Grund von Hauptkomponentenanalysen eines grossen Itempools. Bei der Auswahl der Items wurde vor allem auf faktorielle Reinheit geachtet (Nebenladungen $\leq .25$; Watson et al., 1988, S. 1064). Entsprechend enthält die PANAS keine Valenzitems, da diese Doppelladungen erzeugen, was im Sinne des ursprünglichen Modells (Abbildung 1) gedeutet wurde.¹² Total umfasst die PANAS 20 Items, je 10 für PA und NA.¹³ Bei jedem Item ist anzugeben, ob bzw. wie stark es zutrifft (1 = very slightly or not at all; 2 = a little; 3 = moderately; 4 = quite a bit; 5 = extremely). Dabei können sieben verschiedene Instruktionen verwendet werden, die unterschiedliche Zeiträume ansprechen (z.B. "Moment": "indicate to what extent you feel right now"; "Today": "indicate ... you felt today" bis "General": "indicate ... you generally feel".) Als Bearbeitungszeit geben Krohne und Egloff (2003) rund fünf Minuten an.

Trotz (oder wegen) seiner verbreiteten Anwendung (vgl. Einleitung) ist der PANAS in ganz verschiedenartigen Hinsichten kritisiert worden (siehe Fussnote 12). Im gegebenen Zusammenhang sei einzig das spezielle Konstruktionsprinzip der PANAS-Skalen erwähnt: Sie umfassen jeweils nur Items, welche die hohen Pole von PA und NA markieren ("high-end markers"). Ursprünglich (z.B. Larsen & Diener, 1992) wurde vor allem kritisiert, dass damit PA und NA eigentlich unipolar konzipiert seien und ihre je bipolare Interpretation (vgl. Abbildung 1) letztlich als ungeprüfte Annahme zu gelten habe. Dieser Einwand hat in der Zwischenzeit an Bedeutung verloren, weil die neueren Studien zur Affektstruktur jeweils auch Markiervariablen des niedrigen Pols enthalten und die Annahme der (zumindest deskriptiven) Bipolarität von PA und NA – vor allem auf der Ebene latenter Variablen – heute als gut gestützt angesehen werden kann (z.B. Feldman Barrett & Russell, 1998; Russell & Carroll, 1999; Watson & Clark, 1997; Watson et al., 1999; Yik et al., 1999).

Zwar ist die PANAS – sogar deren Langform, die PANAS-X (Watson & Clark, 1994) – schon in mehreren ESM-Untersuchungen eingesetzt worden. Im Allgemeinen handelt es sich dabei

¹² Das Fehlen solcher geradezu prototypischer Befindlichkeitsitems wurde – insbesondere im Kontrast zur Berücksichtigung von Items wie "strong" oder "active", die eher keine Emotionen darstellen – immer wieder kritisiert. Watson und Clark (1997) nehmen zu diesem – und anderen – Kritikpunkten ausführlich Stellung.

¹³ Die Items lauten (deutsche Übersetzung nach Krohne, Egloff, Kohlmann & Tausch, 1996; in der Reihenfolge der Faktorladungen der deutschen Version) für PA: inspired/angeregt, enthusiastic/begeistert, excited/freudig erregt, active/aktiv, interested/interessiert, strong/stark, attentive/aufmerksam, determined/entschlossen, alert/wach, proud/stolz und für NA: afraid/ängstlich, upset/verärgert, jittery/durcheinander, irritable/gereizt, nervous/nervös, scared/erschrocken, hostile/feindselig, distressed/bekümmert, ashamed/beschämt, guilty/schuldig.

aber um monothematische Studien zur Affekt-Psychologie, mit 1 bis 2 Messungen pro Tag und mit Studierenden als Untersuchungspersonen, die im Rahmen bestimmter Kurse damit Credits erwerben konnten (siehe z.B. Watson et al., 1999). Für polythematische Untersuchungen über das Alltagsleben bei völlig freiwillig mitarbeitenden UPn ist aber die PANAS – wie die meisten konventionellen Skalen – bereits wegen des Zeitbedarfs eindeutig nicht geeignet. Auch wenn die oben referierte Durchführungszeit (5 Minuten) gemäss unseren eigenen Erfahrungen zumindest bei wiederholter Bearbeitung im Rahmen von ESM-Studien eine oberste Grenze darstellen dürfte, umfasst sie für eine ESF, in der noch mehrere andere Aspekte des je aktuellen Moments erfragt werden sollen, zu viele Items.

Der nahe liegende und in der Literatur mehrfach beschrittene Ausweg besteht darin, eine PANAS-Kurzform mit weniger Items zu verwenden (z.B. Williams et al., 1991). Dagegen spricht aber die Tatsache, dass die PANAS-Items zwei weitere der oben diskutierten Anforderungen für ESM-Skalen ebenfalls eindeutig nicht erfüllen:

1.) *Fehlende Antwortstreuung*: Wie in der Literatur schon oft beschrieben worden ist, führen unipolare Items der in der PANAS verwendeten Form in aller Regel zu ausgesprochenen J-Verteilungen, insbesondere wenn es sich um NA-Items handelt. Der Modus solcher Items liegt in aller Regel bei "nicht vorhanden", das heisst, die meisten UPn sehen sich (in den meisten Zeitpunkten) gezwungen, anzugeben, den entsprechenden Affekt nicht zu erleben (vgl. z.B. Schimmack, 2003, Tabelle 1). Das Phänomen schlägt sich dann naturgemäss auch in den Skalenwerten nieder: So beträgt z.B. der Mittelwert der NA-Skala bei Watson et al. (1988, Tabelle 1) bei der Moment-Instruktion – bei einem theoretischen Streubereich der Skala von 10 bis 50 – lediglich 14.8; bei der deutschen Version 14.67 (Krohne et al., 1996, Tabelle 3). Diese extreme Rechtsschiefe ist natürlich sowohl aus der Sicht der UPn als auch aus skalentheoretischer Sicht unerwünscht und hätte bei einer Kurzskala höchst unerfreuliche Auswirkungen auf die Skalenstreuung.

2.) *Fehlende Kontrolle des Antwortstils*: Aus der Eigenart des PANAS, dass nur High-endmarker abgefragt werden, folgt, dass die Akquieszenz nicht kontrollierbar ist. Unipolare Antwort- (Rating-) Skalen der beschriebenen Art werden zwar auch in den meisten anderen modernen Befindlichkeitsinstrumenten eingesetzt (z.B. Steyer et al., 1997; Thayer, 1989). In aller Regel werden aber (idealerweise ausbalanciert) Items für beide Pole einer Dimension vorgegeben und dann (teilweise umgepolt) miteinander verrechnet. Dies eliminiert naturgemäss einen allfälligen Akquieszenzeffekt. Dass dies sinnvoll ist, wird z.B. durch die Tatsache illustriert, dass Steyer et al. (1997) in mehreren explorativen Faktorenanalysen der MDBF-Items bereits in Ein-Punkt-Untersuchungen regelmässig einen bedeutenden Akquieszenzfaktor fanden. Wie weiter oben (Abschnitt 2.1.2) gesagt, wird zwar die Stärke der Rolle der Akquieszenz als Störfaktor in solchen Untersuchungen immer noch diskutiert. Wie dort ebenfalls erwähnt, haben aber Watson und Tellegen (2002) überzeugende Argumente dafür beigebracht, dass sie in aggregierten ESM-Daten, also auf der Personenebene, zu massiven Artefakten führen kann: Je mehr Zeitpunktdaten

der Personen aggregiert werden, umso mehr verändern sich beispielsweise negative Korrelationen zwischen gegenteilig valenzierten Items in positive Richtung – ein Effekt, der verschwindet, wenn ein ebenso stark aggregiertes inhaltsfreies Akquieszenzmass auspartialisiert wird (siehe dazu auch Abschnitt 4.8.1, wo die Studie von Watson & Tellegen ausführlicher referiert wird).¹⁴ Aus dieser Sicht scheint der Einsatz der (ohnehin zu langen) PANAS in ESM-Untersuchungen ohne eine zusätzliche Akquieszenzskala nicht angezeigt zu sein.

2.4 Die PANAVA-KS: Entwicklungsgeschichte und Leitgesichtspunkte

Als ich mit meiner damaligen Arbeitsgruppe in der Mitte der neunziger Jahre mit ESM-Untersuchungen über die Qualität des Erlebens in Arbeit und Freizeit begann, war zunächst ein Entscheid für ein bestimmtes theoretisches Befindlichkeitsmodell zu fällen. Die Überlegungen, die zur Wahl des Modells von Watson und Tellegen (vgl. Abschnitt 2.2) führten, sind anderswo ausführlich dargestellt (Schallberger, 1999, v. a. S. 13ff. und S. 25ff). Zusammenfassend gaben vor allem folgende Gründe den Ausschlag:

Im Hinblick auf die Zielsetzung, das Befinden im Alltag zu untersuchen, erschien der in der Befindlichkeits- bzw. Glücksforschung oft verwendete "Ein-Item-Ansatz" (Beispiele siehe Mayring, 1991), in dem im Wesentlichen die Valenz in einem bestimmten Moment abgefragt wird, zu undifferenziert (vgl. Abbildung 1). Der extreme Gegensatz, der Einsatz eines – naturgemäss umfangreichen – Fragebogens zur Erfassung diskreter Stimmungen/Emotionen (neuere Beispiele sind etwa für Stimmungen: Schimmack, 1997; für Emotionen: Schmidt-Atzert & Hüppe, 1996) kam wegen der Vielzahl von Items nicht in Frage. Als Zwischenlösung bot sich die Verwendung eines sparsamen, zwei bis drei Dimensionen umfassenden dimensional Modells an.¹⁵ Mitte der neunziger Jahre standen vor allem die in Abbildung 1 vorgestellten Modelle im Vordergrund (Valenz und Aktivierung bzw. PA und NA). Gemäss der damals vorherrschenden Meinung (Modellvorstellung A in Abschnitt 2.2.2) waren die beiden Modelle austauschbar und der Entscheid für das eine oder das andere Modell primär eine Frage der Nützlichkeit für die leitenden Forschungsfragen (vgl. z.B. Larsen & Diener, 1992). Im Hinblick auf die Beschreibung des Befindens im Alltag scheint unter diesem Gesichtspunkt das Valenz/Aktivierungsmodell dem PA/NA-Modell eher unterlegen: Aussagen über die von der Aktivierung unabhängig gedachte Valenz und über die von der Valenz unabhängig gedachte Aktiviertheit sind deutlich abstrakter und alltagsferner als Aussagen im Bezugssystem von PA und NA. Als einen Beleg dafür kann man auch die Tatsache ansehen, dass – wie schon oft festgestellt (z.B. Schimmack & Grob, 2000) – valenzfreie Marker für die klassische Aktivierungsdimension in der

¹⁴ Eine Tendenz derselben Art lässt sich der Tabelle II in Schimmack (2003) entnehmen. Dass der Effekt durch das dort verwendete Antwortstilmass nicht kompensiert werden kann, ist eventuell der Tatsache zuzuschreiben, dass dieses nicht einen analogen Aggregationslevel wie die dort verwendete PA- und NA-Skala aufweist.

¹⁵ Der Vollständigkeit halber sei erwähnt, dass ein solches sparsames Modell naturgemäss die Befindlichkeit in einem bestimmten Moment nur relativ grob wiedergeben kann. Der Entscheid für ein solches Modell ist primär ein forschungspragmatischer Entscheid – und nicht etwa ein Entscheid gegen die Relevanz des diskreten Ansatzes (vgl. dazu z.B. Larsen & Diener, 1992; Scherer & Wallbot, 1994).

Alltagssprache weit gehend fehlen. (In Abbildung 1 fehlen daher solche Items.) Dies kann nach der bekannten psycholinguistischen Hypothese, dass für alles, worüber man reden will auch Begriffe gebildet werden, als Ausdruck davon verstanden werden, dass diese Dimension für die alltägliche Befindlichkeitsbeschreibung weitgehend irrelevant ist (vgl. dazu auch Watson et al., 1999).¹⁶

Ausgangspunkt der Entwicklungsarbeiten der PANAVA-Kurzskalen bildete die damals gerade im Druck befindliche deutschsprachige Version des PANAS (Krohne et al., 1996). Zwar waren zu diesem Zeitpunkt manche der in den vorstehenden Abschnitten besprochenen Probleme dieses Instruments bzw. des zugrunde liegenden Modells noch nicht so klar bewusst wie heute. Vieles zeichnete sich aber schon ab. Daher wurde umgehend mit der Entwicklung eines neuen, speziell für ESM-Untersuchungen konzipierten Instruments begonnen, das den in Abschnitt 2.1 beschriebenen Anforderungen genügen sollte (d.h. Kürze bzw. Ökonomie, Augenscheinvalidität, Vermeidung redundanter Items, gute Antwortstreuung und Antwortstilkontrolle). Die Unterschiede zum PANAS lassen sich in drei Punkten zusammenfassen:

1.) *Einbezug der Valenzdimension:* Der Entscheid, die Valenzdimension einzubeziehen, hatte zwei Gründe. Erstens schien dies schon aus Gründen der Augenscheinvalidität angezeigt: Ein Instrument zum Alltagsbefinden, das nicht auch das generelle Wohlbefinden (z.B. glücklich vs. unglücklich) thematisiert, kann den UPn leicht als unvollständig erscheinen. Zweitens sollte dieser Einbezug ermöglichen, der damals weit gehend hypothetisch beantworteten Frage nach der Beziehung von PA und NA zur Valenz (VA) auch empirisch nachzugehen. In jedem Fall war aber vorgesehen, Valenz auch weiterhin als eigenständige Dimension zu behandeln, um die Kompatibilität der PANAVA-KS zum PANAS zu gewährleisten. Auf einen Einbezug der Dimension Aktivierung wurde angesichts des oben erwähnten Fehlens guter Marker zu Gunsten der Kürze des Instruments von vornherein verzichtet.

2.) *Verwendung bipolarer Items:* Ein zweiter Grundsatzentscheid betraf die Wahl eines bipolaren Itemformats (auch Ratingskalen oder semantisches Differential genannt). Dafür sprachen drei Gründe: *Erstens* kann damit die Kritik am PANAS aufgefangen werden, dass nur ein Pol der (zumindest in deskriptiver Hinsicht; vgl. Abschnitt 2.2.1) bipolar gedachten Dimension abgefragt wird (High-end-marker). Voraussetzung für die Berechtigung bipolarer Fragen (z.B. gelangweilt - begeistert) ist aber natürlich, dass sich die beiden Pole des Items – zwar nicht unbedingt auf der semantischen, aber auf der Konstruktebene – wirklich gegenüber stehen. Die in Abschnitt 2.3 erwähnte empirische Forschung zum Circumplex enthält heute viele überzeugende

¹⁶ Die Festlegung auf das PA/NA-Modell impliziert natürlich auch indirekt eine "Definition" des Begriffes Befindlichkeit – eine Frage, die zu weit vom Thema dieses Papiers wegführen würde. Danach steht dieser Begriff für die Ausprägung von zwei unterschiedlichen Formen von Aktiviertheit. Thayer (1996) bezeichnet denselben Sachverhalt als "mood" (Stimmung). Auf diesen Begriff wird hier aber bewusst verzichtet, weil er die klassische Abgrenzung von Emotionen impliziert: Diese Abgrenzung ist nämlich im Zusammenhang mit alltäglichen Befindensbeschreibungen, wie sie mit der ESM provoziert werden, nicht nützlich, weil in solchen Beschreibungen je aktuelle Stimmungen und aktuelle Emotionen untrennbar kontaminiert sind (Brandstätter, 1994; Schmidt-Atzert & Hüppe, 1996; vgl. auch Schallberger, 1999).

Belege dafür, dass dies – insbesondere wenn systematische und zufällige Messfehler kontrolliert werden – im Falle von PA und NA tatsächlich der Fall ist. Entsprechend vertreten heute verschiedene Autoren – zum Teil auch wegen der anderen, weiter unten angesprochenen Vorteile des bipolaren Antwortformats – die Auffassung, dass für eine routinemässige Erfassung des Befindens bipolar gefragt werden sollte (z.B. Schimmack, 2003; Yik et al., 1999). (Dabei ist allerdings der exakte Sinn der in den verschiedenen Dimensionen gemeinten Bipolarität zu beachten.) *Zweitens* ermöglicht das bipolare Antwortformat, das bei unipolaren Items so virulente Problem der J-Verteilungen (und damit der Streuungseinschränkungen; vgl. Abschnitt 2.3) zu korrigieren, indem die Beurteilungsskala explizit bis zum Vorhandensein des gegenteiligen Zustandes ausgedehnt wird. Um einigermaßen symmetrische Verteilungen zu erhalten, ist es – wie die Erfahrungen mit Vorformen der PANAVA-KS zeigten – dabei manchmal notwendig, den einen Pol etwas extremer zu formulieren als den anderen (z.B. lustlos – *hoch* motiviert). *Drittens* sind – wie in der lexikographischen Tradition der Persönlichkeitsforschung schon lange bekannt (z.B. Ostendorf, 1990) – bipolar formulierte Items weniger anfällig für Akquieszenz-Effekte. Werden solche Items zusätzlich hinsichtlich Links-Rechts ausbalanciert dargeboten, dürfte ein Optimum an impliziter Antwortstilkontrolle gegeben sein. Die in der Befindlichkeitsmessung heute üblichere andere Kontrolltechnik – das unipolare Abfragen beider Pole je für sich, verbunden mit einem nachträglichen Verrechnen der beiden Pole – ist nicht nur weniger ökonomisch (Verdoppelung der Itemzahl), sondern erzeugt auch eher den Eindruck von Redundanz.

3.) *Ökonomie bzw. Kürze*: Diese Zielsetzung ergibt sich unmittelbar daraus, dass die neu entwickelten Skalen in ESM-Untersuchungen eingesetzt werden und daher hinsichtlich Bearbeitungszeit möglichst ökonomisch sein sollen. Ausgehend von längeren Vorformen, in denen verschiedene Itemvarianten erprobt wurden, resultierte schliesslich ein Instrument mit 10 Items, zwei für VA und je vier für PA und NA. Diese Zahlen sind wie folgt begründet: In der Literatur wird oft die Ansicht vertreten, die Valenz sei bereits mit einem Item relativ reliabel zu erfassen ("Ein-Item-Ansatz; vgl. oben). Trotzdem wurden zwei Items beigezogen, was einerseits eine empirische Reliabilitätsbestimmung, andererseits – durch eine ausbalancierte Darbietung der beiden Items hinsichtlich Links-Rechts – eine implizite Antwortstilkontrolle ermöglicht.¹⁷ Bei den komplexeren Konstrukten PA und NA war von Anfang an klar, dass mehrere Items notwendig sind. Im Hinblick auf die Antwortstilkontrolle wurde ebenfalls hier eine gerade Itemzahl angestrebt. Die empirische Überprüfung von längeren Vorformen sprachen dafür, dass bereits mit vier Items eine ausreichende Reliabilität erreicht werden kann.

Kasten 1 enthält die für die PANAVA-KS definitiv ausgewählten Items, so wie sie – abgesehen von den kursiv gedruckten Einfügungen – den UPn präsentiert werden. Die beiden Valenzitems stammen aus der deutschen Übersetzung der klassischen Valenzskala von Mehrabian und

¹⁷ Beide Argumente sprechen auch gegen die von Schimmack (2003) vertretene Auffassung, Ein-Item-Skalen seien in ESM-Studien durchaus brauchbar, weil die Aggregation über viele Zeitpunkte automatisch Reliabilität garantiere (was aber ohnehin nur für Skalenwerte auf der Personenebene gilt).

Wie fühlten Sie sich **unmittelbar vor** dem Signal?

		sehr		unentschieden			sehr		
(1: VA-)	zufrieden	3	2	1	0	1	2	3	unzufrieden
(2: PA-)	energiegeladen	3	2	1	0	1	2	3	energielos
(3: NA-)	"gestresst"	3	2	1	0	1	2	3	entspannt
(4: PA)	müde	3	2	1	0	1	2	3	hellwach
(5: NA)	friedlich	3	2	1	0	1	2	3	verärgert
(6: VA)	unglücklich	3	2	1	0	1	2	3	glücklich
(7: PA)	lustlos	3	2	1	0	1	2	3	hoch motiviert
(8: NA)	ruhig	3	2	1	0	1	2	3	nervös
(9: PA-)	begeistert	3	2	1	0	1	2	3	gelangweilt
(10: NA-)	besorgt	3	2	1	0	1	2	3	sorgenfrei

Kasten 1. Die PANAVA-Items in der präsentierten Form

Hinzugefügt sind die Itemnummern, die Skalenzuordnung und die notwendigen Umpolungen (-)

Russell (1974). Sie weisen im Original und in der deutschen Version (Hamm & Vaitl, 1993) die höchsten Ladungen im Valenzfaktor auf. Die Items für PA und NA wurden in Anlehnung an jene des PANAS (Krohne et al., 1996; Watson et al., 1988) entworfen, haben aber auf Grund der Erfahrungen in mehreren sukzessiven ESM-Untersuchungen verschiedene Transformationen durchgemacht (Schallberger, 1999). Ziel war dabei, Items zu finden, die einerseits die Konstrukte PA und NA möglichst breit abdecken, andererseits von den UPn in den unterschiedlichsten Situationen ihres Alltags als sinnvoll, d.h. augenscheinlich und nicht redundant erlebt werden und auch vorteilhafte Antwortverteilungen erzeugen. Diese Ziele waren zum Teil nur durch wesentliche Umformulierungen zu erreichen. Ein Beispiel soll dies illustrieren: "Ängstlich" ist in der deutschen Version des PANAS das Item mit der höchsten Ladung im Faktor NA (Krohne et al., 1996) und war daher ein wichtiger Kandidat auch für die PANAVA-KS. Wie das Item aber auch formuliert wurde; es ergab in ESM-Daten immer extrem schiefe Antwortverteilungen. Die meisten Menschen haben offenbar im normalen Alltag nur sehr selten Momente, in denen sie sich als auch nur tendenziell ängstlich erleben. Um diesen Aspekt in den PANAVA-KS doch repräsentiert zu haben, wurde daher schlussendlich das Item "besorgt - sorgenfrei" gewählt: Fasst man Sorge als Antizipation einer möglicherweise Angst erzeugenden Situation auf, handelt es sich dabei ebenfalls um einen, allerdings schwächer formulierten Marker von NA, der aber eine deutlich ausgewogenere Antwortverteilung erzeugen sollte, was sich in Vorstudien auch bestätigte.

Die Instruktion hat ebenfalls verschiedene Wandlungen durchgemacht und sich in der angegebenen Form (vgl. Kasten 1) am besten bewährt. Im Rahmen der bei ESM-Untersuchungen üblichen (und notwendigen) persönlichen Einführung der UPn sind auch weitere Erläuterungen abzugeben, z.B. die Aufforderung, sich möglichst gut in den bewussten Zeitpunkt zurückzusetzen, dann aber möglichst spontan zu antworten und sich an möglicherweise "schiefe" erscheinenden Gegensätzen nicht zu stören (siehe zum praktischen Vorgehen und weiteren Regeln für das Ausfüllen des ESF: Roduner, Venetz, Aellig et al., 2001).

2.5 Ausblick auf die empirischen Untersuchungen

Bei der Entwicklung der PANAVA-KS war – wie beschrieben – mit verschiedenen Mitteln versucht worden, jene Anforderungen an ESM-taugliche Skalen zu erfüllen, die sich auf die Item-Konstruktion beziehen: Ziel war die Entwicklung kurzer (ökonomischer) Skalen mit Items, die augenscheinvalide und möglichst wenig redundant erscheinen, eine "gute" Antwortstreuung erzeugen und dem individuellen Antwortstil möglichst geringe Angriffsfläche bieten. Die beiden nachstehend beschriebenen Studien sind nun den zwei weiteren, in Abschnitt 2.1 genannten Anforderungen an solche Skalen gewidmet, welche die *Skalenüberprüfung* betreffen: In *Studie 1* (Abschnitt 3) geht es um die bei solchen Neuentwicklungen notwendige Abklärung der Beziehungen zu den Standardskalen der angezielten Konstrukte (Konstruktvalidität). *Studie 2* (Abschnitt 4) befasst sich unter ganz verschiedenen Gesichtspunkten mit der psychometrischen Qualität der PANAVA-KS in ESM-Daten. Dabei interessieren natürlich auch empirische Indizien dafür, ob und inwieweit die bei der Skalenentwicklung verfolgten Ziele (z.B. hinsichtlich Augenscheinvalidität, Antwortverteilungen, Akquieszenzkontrolle) tatsächlich erreicht worden sind.

3. Studie 1: Konstruktvalidierung der PANAVA-KS durch einen Vergleich mit PANAS, AD-ACL und MDBF-GS in einer Ein-Punkt-Untersuchung

3.1 Zielsetzungen und Untersuchungskonzeption

Das Ziel von Studie 1 ist, die PANAVA-Kurzskalen darauf hin zu überprüfen, ob bzw. wie gut es gelungen ist, die anvisierten Konstrukte Positive Aktivierung (PA), Negative Aktivierung (NA) und Valenz (VA) zu operationalisieren. Dies geschieht in Form eines empirischen Vergleichs mit Standardskalen, die als valide Indikatoren dieser Konstrukte gelten. Folgende Skalen werden einbezogen (vgl. Überblick in Tabelle 1, in der auch Angaben zur Anzahl der Items zu finden sind): 1.) die zwei Skalen der Positive and Negative Affect Schedule von Watson et al. (1988) (PANAS; deutsche Version von Krohne et al., 1996; abgekürzt: PANAS-PA und PANAS-NA), die den Ausgangspunkt der Entwicklung der PA- und der NA-Skala der PANAVA-KS bildeten, 2.) die zwei Skalen der Activation-Deactivation Adjective Checklist von Thayer (1989; AD-ACL; deutsche Version von Imhof, 1998) – "Energetic Arousal" (EA; Energetische Aktivierung) und "Tense Arousal" (TA; Spannungsaktivierung) – die wie in Abschnitt 2.2.1 ausgeführt – als zu PA und NA weit gehend äquivalent angesehen werden können (abgekürzt: ADAACL-EA und ADAACL-TA) 3.) Da weder PANAS noch AD-ACL eine Valenzskala enthalten, wird zusätzlich die Skala "Gute-Schlechte Stimmung" (GS) aus dem Mehrdimensionalen Befindlichkeitsfragebogen (MDBF; Steyer, Schwenkmezger, Notz & Eid, 1997) berücksichtigt (abgekürzt: MDBF-GS). Um die Untersuchungspersonen nicht zu überfordern, wurde lediglich eine einmalige Bearbeitung der genannten Skalen verlangt. Es handelt sich also um eine konventionelle Ein-Punkt-Untersuchung. Auf deren Aussagekraft für ESM-Studien wird in der abschliessenden Bilanz aus Studie 1 (Abschnitt 3.4) eingegangen.

Angestrebt wird im Folgenden eine vollständige Konstruktvalidierung im Sinne der simultanen Überprüfung der konvergenten und der diskriminanten Validität (Campbell & Fiske, 1959), die gemäss heutiger Auffassung am besten mittels konfirmatorischer Modelle vorgenommen wird (Marsh & Grayson, 1995; Wothke, 1996). Angesichts der in Abschnitt 2.2 besprochenen Vielfalt von Modellvorstellungen über die Faktorenstruktur der einbezogenen Variablen geht es dabei um *zwei Hauptfragestellungen*: Erstens interessiert, welches Modell die Struktur der Daten am besten widerspiegelt, und zweitens – als Kernfrage der Konstruktvalidierung –, wie gut sich die PANAVA-Kurzskalen in diese Struktur einfügen.

Um diese beiden Fragen zu beantworten, wird wie folgt vorgegangen: Der Ausgangspunkt von Analysen zur konvergenten und diskriminanten Validität bilden die Interkorrelationen bzw. Kovarianzen der einbezogenen Skalen. Deren Grösse wird aber insbesondere auch durch die Reliabilität der Skalen beeinflusst. Daher werden – nach der Beschreibung der Datengrundlage (Abschnitt 3.2) – zunächst die internen Konsistenzen der einbezogenen Skalen vergleichend untersucht (Abschnitt 3.3.1). Danach werden die konvergenten Korrelationen diskutiert, die als solche

bereits als Indikatoren der Konstruktvalidität betrachtet werden können (Abschnitt 3.3.2). Dabei wird auch auf die Rolle der unterschiedlichen Itemformate der verschiedenen Skalen eingegangen – neben den Skalenreliabilitäten der zweite, potentiell wichtige Einflussfaktor für die Grösse der Korrelationen. Anschliessend werden die total acht Skalen einer explorativen Faktorenanalyse unterzogen, um einen ersten Einblick in die Datenstruktur zu erhalten (Abschnitt 3.3.3). Schliesslich werden die konfirmatorischen Analysen vorgenommen. Dabei ist zu berücksichtigen, dass PANAS und AD-ACL an sich zweidimensional konzipiert sind. Daher werden in einem ersten Schritt nur jene Skalen analysiert, die diese Dimensionen betreffen (PA- bzw. EA- und NA- bzw. TA-Skalen; Abschnitt 3.3.4). Erst in einem zweiten Schritt wird dann auch die Valenzdimension (PANAVA-VA und MDBF-GS) einbezogen (Abschnitt 3.3.5).

3.2 Datengrundlage

Die *Stichprobe* umfasst (nach Ausschluss von zwei Personen, welche die Skalenbearbeitung vorzeitig abgebrochen haben) 269 Personen (37 % Männer; 63 % Frauen) mit einem Durchschnittsalter von $M = 27$ Jahren ($SD = 7.1$ Jahre). Diese Personen wurden bei verschiedenartigen Gelegenheiten (Vorlesungen, Restaurants, Warteräume, Begegnungen in der Privatsphäre etc.) von acht verschiedenen Versuchsleitern angesprochen und gebeten, ihr Befinden entweder im aktuellen oder – zu einem späteren Zeitpunkt – in einem frei gewählten Moment zu beschreiben. Die *Instruktion* erfolgte standardisiert und war dem Fragebogenbündel beigegeben.

Im Hinblick auf Fragestellungen, die über die Thematik dieses Papiers hinausreichen, wurden total fünf Instrumente mit 13 Skalen eingesetzt, was im Interesse der Datenqualität nur mittels eines partiellen *Rotationsdesigns* möglich war. Drei *Instrumente* mit acht Skalen wurden von allen UPn bearbeitet, nämlich PANAVA-KS, PANAS und AD-ACL. Die übrigen Skalen wurden jeweils nur einem Teil der Stichprobe vorgelegt. Im Folgenden wird von diesen Zusatzskalen nur die Skala GS des MDBF berücksichtigt, die von 121 Personen bearbeitet worden ist. Die Reihenfolge der Instrumente wurde systematisch variiert, um eine Konzentration der – mit Sicherheit vorhandenen – Ermüdungs- und Sättigungseffekte auf bestimmte Skalen zu vermeiden.

Die *Datenaufbereitung* umfasste mehrere Schritte: Bei der visuellen Überprüfung der Verteilungen der Skalenwerte zeigte sich das aus der Literatur (z.B. Staudinger, 2000) gut bekannte Phänomen einer Verschiebung in Richtung des positiv valenzierten Pols, was sich bei vier der acht Skalen in hoch signifikanten Schiefekennwerten ausdrückt (PANAVA-VA; PANAS-NA; ADA-CL-TA und MDBF-GS; alle $p < .001$). Mit einer Ausnahme liegt die Schiefe aber mit Absolutbeträgen deutlich unter .8 weit unter dem Betrag von 2.0, der von Curran, West und Finch (1996) als Indikator für eine "moderate" Abweichung von der Normalverteilung betrachtet wird. Die Ausnahme (Schiefe = 1.5) findet sich – gemäss den in Abschnitt 2.3 besprochenen Befunden über die bei NA-Items üblichen J-Verteilungen nicht überraschend – bei PANAS-NA, wo ein Bodeneffekt auftritt, d.h. der Minimalwert gleichzeitig den Modus darstellt, was sich dann

auch in einer auffällig kleinen Streuung manifestiert (vgl. Tabelle 1). Um das Problem abzuschwächen, wurde PANAS-NA einer Flächentransformation unterzogen, was zu einer unauffälligen Schiefe = .3 ($p > .05$) führte. Die dabei resultierenden z-Werte wurden in eine Verteilung mit dem ursprünglichen Mittelwert und der ursprünglichen Standardabweichung zurücktransformiert.¹⁸

Bei der *Bearbeitung der fehlenden Werte* (MV) kamen zwei verschiedene Techniken zum Einsatz: Die nur vereinzelt vorhandenen MV in den verschiedenen Instrumenten wurden mittels des Expectation-Maximation-Algorithmus (Schafer & Graham, 2004) ersetzt. Zur Kompensation der durch den Rotations-Design bedingten fehlenden Werte bei MDBF-GS wurde mittels der "Full Information Maximum Likelihood" (FIML) Methode (Arbuckle, 1996; Arbuckle & Wothke, 1999, S. 331-358) eine vervollständigte Varianz-Kovarianz-Matrix geschätzt. Die Schätzungen unterscheiden sich nur wenig von den empirischen Werten, die sich in der reduzierten Stichprobe ($n = 121$) ergeben. Beispielsweise beträgt die empirische Korrelation von MDBF-GS mit PANAVA-VA in der Teilstichprobe $r = .86$, in der mittels FIML ergänzten Matrix für die Gesamtstichprobe $r = .83$ (vgl. Tabelle 1). Der Einsatz von FIML ermöglicht aber, die bei den andern Skalen vorhandene Information vollumfänglich auszunützen.

3.3 Ergebnisse

Einen Überblick über die Verteilungskennwerte, internen Konsistenzen und Interkorrelationen aller acht Skalen findet sich in Tabelle 1.

3.3.1 Interne Konsistenz der Skalen im Vergleich

Wie die dritte Spalte von Tabelle 1 zeigt, liegen die Konsistenzkoeffizienten der acht Skalen in einem Bereich zwischen .74 und .94. Sie sind also von einer befriedigenden bis sehr guten Grössenordnung. Der niedrigste Koeffizient ($\alpha = .74$) findet sich bei der Skala PANAVA-VA, der höchste Koeffizient ($\alpha = .94$) bei MDBF-GS. Beide zielen auf das Konstrukt Valenz ab; MDBF-GS umfasst jedoch acht Items, PANAVA-VA lediglich zwei, was den Reliabilitätsunterschied weit gehend erklärt: Rechnet man den Koeffizienten von PANAVA-VA mittels der Formel von Spearman-Brown auf die Itemzahl von MDBF-GS hoch, ergibt sich auch für PANAVA-VA $\alpha = .92$.

Weniger gross sind die Unterschiede bei den andern sich entsprechenden Skalen: Bei PA/EA liegen die Koeffizienten im Bereich $.83 \leq \alpha \leq .91$, bei NA/TA im Bereich $.76 \leq \alpha \leq .88$, wobei die niedrigsten Werte jeweils von den PANAVA-Skalen stammen. Auch hier dürfte die Anzahl

¹⁸ PANAS-NA ist auch die einzige Skala, die eine hoch signifikante Kurtosis aufweist. Durch die beschriebene Transformation wurde sie aber ebenfalls auf einen nicht signifikanten Wert reduziert. Im Hinblick auf die konfirmatorischen Faktorenanalysen (Abschnitt 3.3.4f.) sei noch erwähnt, dass die multivariate Kurtosis der Skalen nicht signifikant von 0 abweicht (multivariate Kurtosis = 1.8; $p > .05$; Mardia-Test).

Tabelle 1

Verteilungskennwerte, Cronbach- α und Interkorrelationen der Skalen von PANAS, AD-ACL, PANAVA und MDBF-GS

Skala (Anzahl Items)	<i>M</i>	<i>SD</i>	α	Interkorrelationen						
				PANAS		AD-ACL		PANAVA-KS		
				PA	NA	EA	TA	PA	NA	VA
PANAS:										
PANAS-PA (10)	2.9	0.73	.88	-						
PANAS-NA (10)	1.5	0.46	.83	-.26	-					
AD-ACL:										
ADACL-EA (10)	2.7	0.74	.91	.75	-.36	-				
ADACL-TA (10)	2.0	0.66	.88	-.25	.71	-.23	-			
PANAVA:										
PANAVA-PA (4)	4.3	1.26	.83	.77	-.35	.84	-.23	-		
PANAVA-NA (4)	3.3	1.21	.76	-.36	.70	-.32	.82	-.36	-	
PANAVA-VA (2)	5.3	1.15	.74	.54	-.56	.51	-.49	.57	-.59	-
MDBF:										
MDBF-GS (8)	4.0	0.76	.94	.60	-.64	.57	-.66	.60	-.71	.83

Anmerkungen: $N = 269$. Mittelwerte (*M*) und Standardabweichungen (*SD*) beruhen auf Skalenwerten, die als durchschnittliche Itemantwort ausgedrückt sind (Antwortvorgaben: AD-ACL: 1-4; PANAS und MDBF: 1-5; PANAVA: 1-7). Die fett gedruckten Zahlen sind konvergente Korrelationen. Alle in der Tabelle enthaltenen Korrelationen sind hoch signifikant ($p < .001$). *Abkürzungen:* PA = Positive Aktivierung; NA = Negative Aktivierung; EA = energetische Aktivierung (energetic arousal); TA = Spannungsaktivierung (tense arousal); GS = Gute - Schlechte Stimmung. Die Verteilungskennwerte und Korrelationen von MDBF-GS basieren auf der mittels der FIML-Methode ergänzten Kovarianzmatrix (vgl. Abschnitt 3.2).

Items (AD-ACL und PANAS je 10; PANAVA je 4) die entscheidende Erklärung liefern: Bei einer Hochrechnung der PANAVA-Konsistenzen auf die Itemzahl der Vergleichsskalen, ergeben sich geschätzte Werte, die jene der Vergleichsskalen übertreffen (PANAVA-PA: $\alpha = .92$; PANAVA-NA: $\alpha = .89$).

Im Hinblick auf die angestrebten empirischen Skalenvergleiche ist aus dieser Konsistenzanalyse vor allem festzuhalten, dass – wie zu erwarten war – die deutlich kürzeren PANAVA-Skalen etwas niedrigere Reliabilitäten aufweisen als die Vergleichsskalen. Der grösste Unterschied findet sich bei PANAVA-VA und MDBF-GS. Insbesondere bei einem Vergleich der Korrelationen dieser zwei Skalen mit Drittvariablen wird der Reliabilitätsunterschied mit in Rechnung zu stellen sein.

3.3.2 Konvergente Korrelationen und die Rolle des Itemformats

Als konvergente Korrelationen bezeichnet man Korrelationen von Skalen, die dasselbe Konstrukt erfassen sollen. Solche Koeffizienten sind in Tabelle 1 fett gedruckt. Ihre Grösse stellt naturgemäss einen ersten Hinweis auf die Konstruktvalidität der Skalen dar. Wie in Abschnitt 3.1 erwähnt, sind dabei zwei weitere Einflussfaktoren zu berücksichtigen: die im vorherigen Abschnitt diskutierten, unterschiedlichen Skalenreliabilitäten sowie Effekte des unterschiedlichen

Itemformats. Unter dem Aspekt des Itemformats lassen sich drei Gruppen von Skalen ausmachen: 1) Die PANAVA-Skalen mit bipolar formulierten Items; 2.) die AD-ACL-Skalen und die Skala MDBF-GS, die eine jeweils ganz oder annähernd¹⁹ ausbalancierte Anzahl unipolar formulierte Items für beide Pole der jeweiligen Dimension umfassen; 3.) die PANAS-Skalen, deren Items ebenfalls unipolar formuliert sind, jedoch nur den hohen Pol von PA und NA ansprechen (sog. High-end-marker).

Wie aus Tabelle 1 zeigt, fallen die konvergenten Korrelationen zwischen den PANAVA-KS und den ausbalancierten Skalen sehr hoch aus ($r_{PANAVA-PA,ADACL-EA} = .84$; $r_{PANAVA-NA,ADACL-TA} = .82$; $r_{PANAVA-VA,MDBF-GS} = .83$). Berechnet man minderungskorrigierte Koeffizienten, ergeben sich Werte von praktisch 1.0 (in der obigen Reihenfolge: $r = .97$, 1.0 und 1.0). Ob man die beiden Pole der Dimensionen je einzeln abfragt und nachträglich miteinander verrechnet (AD-ACL; MDBF-GS) oder ob man bereits die Item bipolar formuliert (PANAVA-KS), scheint keine wesentliche Rolle zu spielen; die Skalen erfassen – berücksichtigt man die nicht-perfekte Reliabilität – dasselbe Konstrukt.

Etwas niedriger sind die konvergenten Korrelationen zwischen PANAVA-KS und PANAS ($r_{PANAVA-PA,PANAS-PA} = .77$; $r_{PANAVA-NA,PANAS-NA} = .70$). Dasselbe gilt aber auch für den AD-ACL ($r_{ADACL-EA,PANAS-PA} = .75$; $r_{ADACL-TA,PANAS-NA} = .71$). Die entsprechenden minderungskorrigierten Korrelationen liegen denn auch deutlich unter 1.0 (in obiger Reihenfolge: $r = .90$, $.88$, $.84$ und $.83$). Der Verdacht liegt nahe, dass dies damit zusammenhängt, dass die PANAS-Skalen nur High-end-marker enthalten und damit der Antwortstil (Akquieszenz) nicht kontrolliert wird (vgl. Abschnitt 2.3). Dieser Verdacht lässt sich dadurch überprüfen, dass aus den AD-ACL-Skalen PANAS-analoge Subskalen gebildet werden, die ebenfalls nur jene Items enthalten, die den hohen Pol betreffen.²⁰ Ist der Verdacht über die Rolle des Itemformats richtig, müssten diese AD-ACL-Subskalen mit dem PANAS stärker, mit den PANAVA-KS jedoch schwächer korrelieren als die AD-ACL-Originalskalen. Beides trifft – zieht man minderungskorrigierte Korrelationen bei – zu.²¹ Die Korrelation der PANAS-analogen ADAACL-EA-Subskala mit PANAS-PA beträgt $r = .88$ (vs. $r = .84$ bei der EA-Originalskala), jene mit PANAVA-PA ebenfalls $r = .88$ (vs. $r = .97$ bei der EA-Originalskala). Das Analoge gilt für die TA-Subskala, die nur aus den High-end-markern besteht (PANAS-NA: $r = .87$ statt $r = .83$; PANAVA-NA: $r = .94$ statt $r = 1.0$).

Auch wenn man den Aussagewert dieser auf Schätzungen beruhenden Berechnungen nicht überstrapazieren darf, liefern sie doch ein starkes Indiz dafür, dass die etwas niedrigeren konvergenten Korrelationen zwischen PANAS einerseits und AD-ACL und PANAVA-KS andererseits mindestens zum Teil mit dem speziellen Konstruktionsprinzip des PANAS zusammenhängen.

¹⁹ Die Skala ADAACL-EA enthält 6 Items für starke und 4 Items für fehlende Energetische Aktivierung.

²⁰ Diese Subskalen stimmen nicht mit jenen von Thayer überein, da die letztgenannten in sich wiederum (tendentiell) ausbalanciert sind.

²¹ Minderungskorrigierte Korrelationen sind deswegen notwendig, weil die AD-ACL-Subskalen natürlich kürzer sind als die AD-ACL-Originalskalen.

Dass dabei zusätzlich die konvergenten Korrelationen zu PANAS-NA systematisch niedriger ausfallen als jene zu PANAS-PA dürfte durch den Bodeneffekt von PANAS-NA bedingt sein, der durch die vorgenommene Flächentransformation natürlich nicht vollständig kompensiert wird.

3.3.3 Explorative Faktorenanalyse aller acht Skalen

Um einen ersten Einblick in die Faktorenstruktur der acht untersuchten Skalen zu erhalten, wurden sie einer Hauptkomponentenanalyse unterzogen. Der resultierende Eigenwertverlauf (4.82, 1.64, 0.48, 0.39 ...) spricht sowohl nach dem Kaiser-Guttman-Kriterium als auch nach dem Scree Test eindeutig für eine Zwei-Faktorenlösung. Eine schiefwinklige Rotation (Promax; vgl. Thompson, 2004) liefert zwei mit $-.49$ stark korrelierte Komponenten (vgl. Tabelle 2).²² Die Pattern-Koeffizienten von PANAS, AD-ACL und PANAVA-KS (ohne VA) zeigen eine prägnante Einfachstruktur; lediglich ADACL-TA besitzt eine substantielle Nebenladung (.21). Dieser Befund ist nicht völlig neu: Die früher zitierte Studie von Yik et al. (1999) kommt zwar zum Schluss, dass die beiden Dimensionenpaare, die der AD-ACL und der PANAS zu Grunde

Tabelle 2

Promax-rotierte Komponenten von PANAS, AD-ACL, PANAVA-KS und MDBF-GS
(Korrelationen der Faktoren 1 und 2: $r = -.49$)

Skala	Pattern-Koeffizienten		Struktur-Koeffizienten		h^2
	Faktor 1	Faktor 2	Faktor 1	Faktor 2	
PANAS:					
PANAS-PA	.06	.93	-.39	.90	.81
PANAS-NA	.88	.04	.86	-.39	.74
AD-ACL:					
ADACL-EA	.08	.95	-.38	.91	.84
ADACL-TA	1.00	.21	.90	-.28	.84
PANAVA:					
PANAVA-PA	.06	.96	-.41	.93	.87
PANAVA-NA	.94	.05	.91	-.41	.84
PANAVA-VA	-.53	.43	-.74	.69	.69
MDBF:					
MDBF-GS	-.65	.40	-.85	.72	.84

Anmerkungen: $N = 269$. Abkürzungen: PA = Positive Aktivierung; NA = Negative Aktivierung; EA = energetische Aktivierung (energetic arousal); TA = Spannungsaktivierung (tense arousal); GS = Gute - Schlechte Stimmung.

²² Diese Korrelation sinkt von $-.49$ auf $-.35$, wenn die beiden Valenzskalen nicht in die Faktorenanalyse einbezogen werden. Dies ist Folge davon, dass bei der Rotation eine Optimierung der Einfachstruktur auch für diese Skalen angestrebt wird. Am Muster der Pattern- und Struktur-Koeffizienten ändert sich durch diese Weglassung aber nichts Grundsätzliches.

liegen "... for most practical purposes ... al most interchangeable ..." seien (a.a.O., S. 617). Jedoch erwiesen sich die AD-ACL-Dimensionen – im Bezugssystem der Abbildung 1 – gegenüber den PANAS-Dimensionen ganz leicht Richtung Aktivierungsachse verschoben, was nichts anderes bedeutet, als dass in einer gemeinsamen Faktorenanalyse kleine Nebenladungen zu erwarten sind.

Das prägnante Ladungsmuster (Pattern-Koeffizienten) macht die Interpretation der beiden Faktoren leicht: Der erste Faktor beschreibt die Negative Aktivierung (bzw. Spannungsaktivierung), der zweite Faktor die Positive Aktivierung (bzw. Energetisch Aktivierung). Die entsprechenden beiden PANAVA-Kurzskalen (PANAVA-PA und -NA) erweisen sich dabei als klare und eindeutige Repräsentanten dieser Faktoren. Die Valenzskala (PANAVA-VA) hat dagegen – ganz analog zu MDBF-GS – eine Doppelladung. Wie die Strukturkoeffizienten zeigen, sind die Valenzskalen mit *beiden* Faktoren stark korreliert, und zwar – wie auf Grund von Abschnitt 2.2.2 zu erwarten – etwas stärker mit Faktor 1 (NA/TA) als mit Faktor 2 (PA/EA).

Insgesamt ergibt sich damit aus dieser Analyse ein Bild, das der Theorie des Circumplex-Modells (vgl. Abschnitt 2.2) entspricht: Eine klare Zweifaktorenstruktur mit einer (asymmetrischen) Zwischenstellung von Valenz, die eine (statistische) Erklärbarkeit von VA durch PA und NA nahelegt (Modell A in Abschnitt 2.2.2). Ob dieses Modell auch konfirmatorischen Analysen standhält, ist Thema der beiden folgenden Abschnitte.

3.3.4 Konfirmatorische MTMM-Analyse der PA/EA- und NA/TA-Skalen

Wie in Abschnitt 3.1 begründet, werden die konfirmatorischen Analysen in einem ersten Schritt lediglich auf das zweidimensionale Modell des PANAS bzw. der AD-ACL bezogen, die beiden Skalen PANAVA-VA und MDBF-GS also vorerst ausgeklammert. Die zentralen Fragen lauten dann, ob sich dieses zweidimensionale Modell in den Daten der verbleibenden sechs Skalen (PANAS-PA und -NA; ADAACL-EA und -TA sowie PANAVA-PA und -NA) bestätigen lässt und wie gut sich die beiden PANAVA-Skalen in die gefundene Struktur einfügen.

Sieht man von den beiden Valenzskalen ab, stellt Tabelle 1 eine klassische Multitrait-Multimethod- (MTMM-) Matrix dar, wie sie seit Campbell und Fiske (1959) zur vollständigen (d.h. die diskriminante Validität und eventuelle Methodeneffekte mitberücksichtigende) Konstruktvalidierung benutzt werden. Zur Auswertung solcher Matrizen stehen verschiedene Strukturgleichungsmodelle zur Verfügung, die allerdings bei der gegebenen Datenstruktur (zwei Traits, drei Methoden) in den meisten Fällen nur mit Zusatzannahmen identifiziert sind. Dies gilt nicht für das "Correlated Traits Correlated Uniqueness"- (CTCU-) Modell, das auch andere vorteilhafte Eigenschaften aufweist (Marsh & Grayson, 1995) und hier deswegen angewandt wird. Es postuliert – bezogen auf die gegebenen Daten – dass die sechs Skalen zwei korrelierte Traits erfassen, wobei die Uniqueness (Residuals) der einzelnen Skalen innerhalb eines Instruments methodenbedingt korreliert sein dürfen. Auf Grund des oben referierten Befundes von Yik et al.

(1999), der sich in der explorativen Faktorenanalyse zumindest für TA bestätigt hat, wird zusätzlich zugelassen, dass ADACL-EA und -TA neben einer eindeutigen Hauptladung auch eine Nebenladung aufweisen. Das resultierende Modell – *Modell 0* genannt – ist in Abbildung 2 festgehalten. Die Parameterschätzungen wurden mit der Maximum Likelihood-Methode vorgenommen. Um alle Faktorladungen auf Signifikanz testen zu können, wurden die Varianzen der beiden Faktoren auf 1 fixiert.²³

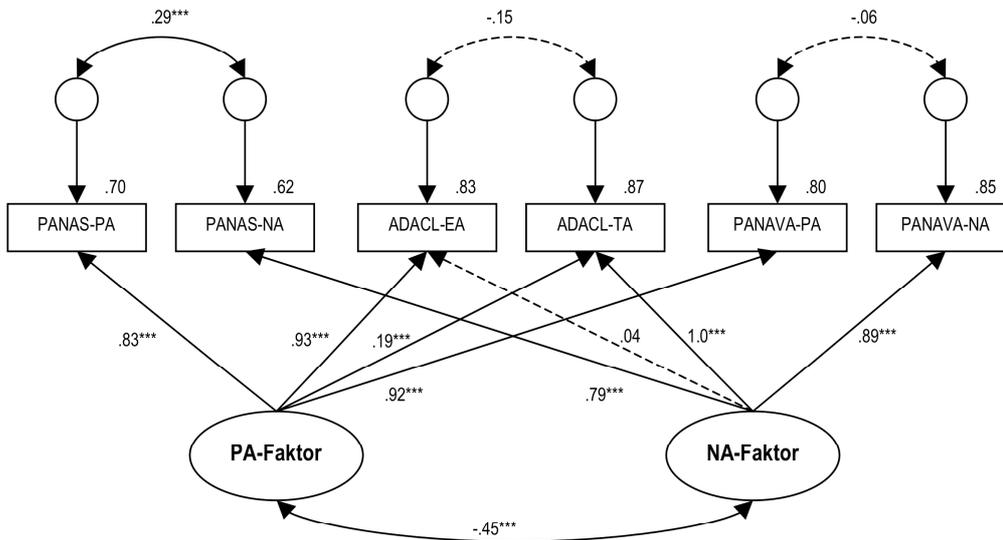


Abbildung 2. Modelle 0 und 1 mit standardisierten Parameterschätzungen aus Modell 0

Nicht signifikante Parametern entsprechende Pfeile sind gestrichelt; durch ihre Weglassung entsteht Modell 1. Die Zahlen oberhalb der Rechtecke geben die durch die beiden Faktoren erklärten Varianzanteile an.

Die standardisierten Parameterschätzungen in Abbildung 2 bestätigen die theoretischen Erwartungen und die bisherigen Einzel-Einsichten: Alle sechs Skalen haben hoch signifikante, inhaltlich den Erwartungen entsprechende Hauptladungen von rund .8 bis 1.0.²⁴ Die beim AD-ACL zugelassenen Nebenladungen sind mit .04 (EA) und .19 (TA) massiv kleiner und nur bei TA signifikant ($p < .001$). Die beiden Faktoren sind mit $r = -.45$ ($p < .001$) relativ stark korreliert. Konsistent mit den Befunden über die Rolle des speziellen Itemformats (Abschnitt 3.3.2) erklären sie bei den PANAS-Skalen mit 70 % (PANAS-PA) und 62 % (PANAS-NA) deutlich geringere Varianzanteile als bei den andern Skalen (zwischen 80 % und 87 %; vgl. Abbildung 2). Die Residuen der PANAS-Skalen sind auch die einzigen, die signifikant korrelieren ($r = .29$, $p < .001$). Nach den weiter oben berichteten Befunden liegt es nahe, darin einen Ausdruck der im PANAS nicht kontrollierten Akquieszenz zu sehen.

Das in Abbildung 2 dargestellte Modell 0 liefert auch gute bis sehr gute Fit-Indizes, wie Tabelle 3 illustriert (zu den verwendeten Kriterien siehe Arbuckle & Wothke, 1999, S. 395ff.; Byrne,

²³ Sämtliche konfirmatorische Analysen in diesem und dem folgenden Abschnitt wurden mit dem Programm Amos 5.0 (Arbuckle, 2003) durchgeführt.

²⁴ Bekanntlich beziehen sich die Signifikanztests in Strukturgleichungsmodellen auf die unstandardisierten Koeffizienten. Trotzdem wird hier – wie üblich – die einfachere Ausdrucksweise benutzt.

Tabelle 3
Fit-Indizes der getesteten Modelle

χ^2 ; df ; p	χ^2/df	CFI	RMSEA; PCLOSE	AIC
MTMM-Analyse der PA/EA- und NA/TA-Skalen:				
<i>Modell 0:</i> volles CTCU-Modell mit Nebenladungen des AD-ACL				
5.5; 3; .14	1.83	1.00	.06; .36	41.5
<i>Modell 1:</i> Modell 0, nicht-signifikante Parameter auf 0 fixiert				
8.3; 6; .22	1.38	1.00	.04; .57	38.3
<i>Modell 2:</i> Modell 1 ohne Nebenladung von ADACL-EA				
27.0; 7; <.001	3.86	.98	.10; .01	55.0
<i>Modell 3:</i> Modell 1 als Trait-only-Modell				
22.6; 7; .002	3.23	.99	.09; .05	50.6
Alternativmodelle für die PA/EA- und NA/TA-Skalen:				
<i>Modell 4:</i> Generalfaktormodell				
470.0; 8; <.001	58.75	.60	.46; <.001	496.0
<i>Modell 5:</i> drei Instrumentenfaktoren				
379.2; 6; <.001	63.20	.68	.48; <.001	409.2
Modell 1 unter Einbezug von PANAVA-VA und MDBF-GS:^a				
<i>Modell 6:</i> 2-Faktorenmodell mit Doppelladungen von VA und GS				
125.4; 30; <.001	4.18	.95	.11; <.001	137.4
<i>Modell 6a:</i> 2-Faktorenmodell mit Ladungen von VA und GS nur in PA				
385.2; 32; <.001	12.04	.80	.20; <.001	393.2
<i>Modell 6b:</i> 2-Faktorenmodell mit Ladungen von VA und GS nur in NA				
259.5; 32; <.001	8.11	.87	.16; <.001	267.5
<i>Modell 7:</i> 3-Faktorenmodell mit PA-, NA- und VA-Faktor				
31.2; 31; .46	1.01	1.00	.01; .97	41.2
<i>Modell 8:</i> hierarchisches Modell: VA-Faktor PA und NA übergeordnet				
35.9; 29; .18	1.24	1.00	.03; .86	49.9
<i>Modell 9:</i> VA-Faktor von PA- und NA-Faktor abhängig				
31.2; 29; .36	1.07	1.00	.02; .95	45.2

Anmerkungen: Erläuterungen der Modelle und Fit-Indizes im Text.

^a Alle Parameter aus Modell 1 wurden auf die dort geschätzten Werte fixiert.

2001, S. 79ff.): Der χ^2 -Test, der die Abweichung der Daten von den Modellimplikationen testet, ist nicht signifikant ($\chi^2(3, N = 269) = 5.49, p = .14$). Das Verhältnis $\chi^2/df = 1.83$ liegt unter 2.0, dem üblichen Grenzwert für ein akzeptables Modell. Der CFI beträgt praktisch 1 (als Minimalwert für gute Anpassung gilt .95). Auch Kennwerte, welche die Komplexität des Modells mit in

Rechnung stellen, fallen sehr gut aus: RMSEA beträgt .056, was (allerdings knapp) noch im von Hu und Bentler (1999) angegebenen Bereich von 0.0 bis 0.06 liegt, der für gute Modellanpassung spricht. PCLOSE, der p -Wert aus dem Test der Nullhypothese, dass RMSEA in der Population nicht grösser als .05 ist, beträgt .36, was die gute Angepasstheit des Modells an die Daten auch inferenzstatistisch absichert. Der letzte in Tabelle 3 angegebene Kennwert, AIC, dient dem Vergleich verschiedener Modelle (je kleiner umso besser) und kann für ein einzelnes Modell nicht sinnvoll interpretiert werden.

Trotz dieser guten Anpassung sei noch das sparsamere *Modell 1* geprüft, das sich aus Modell 0 ergibt, wenn die drei nicht signifikanten Parameter (in Abbildung 2 gestrichelte Linien) auf Null fixiert werden. Die aus diesem Modell resultierenden Schätzungen der verbleibenden Parameter entsprechen praktisch jenen in Abbildung 2 (Abweichungen $< .03$). Die Angepasstheit dieses Modells 1 (vgl. Tabelle 3) ist zwar nicht signifikant besser als jene von Modell 0 ($\chi^2_{\text{Diff}} = 2.8$, $df = 3$, $p > .05$), jedoch fallen alle Fit-Indizes ausser dem CFI, der gleich bleibt, deskriptiv noch besser aus.

Zur Absicherung dieses Modells 1 wird es in Tabelle 3 noch mit einigen anderen Modellen konfrontiert. Die Modelle 2 und 3 stellen Vereinfachungen von Modell 1 dar: In *Modell 2* ist gegenüber Modell 1 die Nebenladung von ADACL-EA auf den NA-Faktor Null gesetzt, was zu einer hoch signifikanten Verschlechterung des Fit führt ($\chi^2_{\text{Diff}} = 18.7$, $df = 1$, $p < .001$). *Modell 3* ist ein "Trait-only" Modell (Marsh & Grayson, 1995), was hier bedeutet, dass keine Fehlerkorrelationen zugelassen werden, d.h. im Modell 1 die Uniqueness-Korrelation innerhalb des PANAS Null gesetzt wird. Auch hier resultiert eine hoch signifikante Fit-Verschlechterung ($\chi^2_{\text{Diff}} = 14.3$, $df = 1$, $p < .001$).

Nicht nur Variationen, sondern Alternativhypothesen stellen die Modelle 4 und 5 dar: *Modell 4* postuliert, dass alle 6 Skalen auf einem Generalfaktor laden, *Modell 5* das Gegenteil, nämlich dass die drei Instrumente drei zwar korrelierte, jedoch instrumentenspezifische Faktoren erfassen. Wie die äusserst schlechten Fit-Indizes dieser Modelle (Tabelle 3) zeigen, werden sie den Daten keinesfalls gerecht.

Dies alles bedeutet zusammenfassend, dass *Modell 1* (Abbildung 2 unter Weglassung der gestrichelten Linien) *das* optimale Modell für die sechs PA/EA- und NA/TA-Skalen darstellt.²⁵ Die zwei Faktoren lassen sich – berücksichtigt man die messtechnische Probleme des PANAS – problemlos als "PA-Faktor" und "NA-Faktor" interpretieren, könnten aber (beinahe, vgl. die Nebenladung von ADACL-TA) als EA- und TA-Faktor bezeichnet werden. Die Ladungen von PANAVA-PA (.92) und -NA (.89) zeigen zudem, dass diese beiden PANAVA-Kurzskalen sehr gute Repräsentanten dieser beiden Faktoren darstellen.

²⁵ Dies bestätigte sich übrigens auch in einem heuristischen Specification-Search (Arbuckle, 2003, S. 13ff.), bei dem alle denkbaren Nebenladungen und (instrumentenspezifischen) Uniqueness-Korrelationen zugelassen wurden.

3.3.5 Konfirmatorische Analysen zur Beziehung der Valenzskalen zum PA- und NA-Faktor

Es bleibt aber noch die zentrale Frage, wie sich die bisher ausgeklammerten Valenzskalen (PANAVA-VA und MDBF-GS) in die Struktur von Abbildung 2 einfügen. Dazu existieren in der aktuellen Literatur implizit oder explizit ganz verschiedene Modellvorstellungen, die in Abschnitt 2.2.2 vorgestellt wurden. Die wichtigsten Modelle sind in Abbildung 3 – in vereinfachter Form – graphisch illustriert. Ihre Fit-Indizes sind im untersten Bereich von Tabelle 3 zu finden. Bei der Überprüfung dieser Modelle wurden sämtliche Parameter, für die aus dem Modell 1 Schätzungen zur Verfügung stehen und die im jeweiligen Modell ebenfalls vorhanden sind, fixiert, um die Ausgangslage bei den verschiedenen Modellen vergleichbar zu halten.²⁶ Bei den Drei-Faktormodellen (Modelle 7 bis 9) wurde zusätzlich die Varianz der Uniqueness von MDBF-GS auf Null fixiert: Diese Varianz erwies sich in allen diesen Modellen als sehr klein (in Modell 8 sogar als negativ) und nicht signifikant. Das heisst: MDBF-GS scheint den dritten Faktor praktisch fehlerfrei zu messen, was mit der Tatsache, dass diese Skala den grössten Konsistenzkoeffizienten liefert (.94; vgl. Tabelle 1), zumindest kompatibel ist.

Modell 6 (vgl. Abbildung 3) geht davon aus, dass die Annahme eines dritten Faktors unnötig ist. Es handelt sich also um ein Zwei-Faktorenmodell (PA-Faktor, NA-Faktor), wobei PANAVA-VA und MDBF-GS je eine Doppelladung in diesen beiden Faktoren aufweisen. Dies entspricht strukturell der Theorie des Circumplex (Modell A in Abschnitt 2.2.2), die sich ja in der explorativen Faktorenanalyse (Abschnitt 3.3.3) zu bewähren schien. Die Parameter des Modells 6, die den Ladungen der beiden Skalen in den beiden Faktoren entsprechen (vgl. Abbildung 3), stimmen denn auch praktisch mit jenen überein, die in der explorativen Faktorenanalyse gefunden wurden (Tabelle 2; Differenzen $\leq .03$). Die Fit-Indizes (ausser vielleicht CFI) dieses Modells sind aber – in Übereinstimmung mit den früher erwähnten konfirmatorischen Befunden aus anderen Untersuchungen – klar ungenügend (vgl. Tabelle 3: $\chi^2(30, N = 269) = 125.4, p < .001$; $\chi^2/df = 4.18$; CFI = .95; RMSEA = .11; PCLOSE < .001; AIC = 137.4.).

Der noch weit schlechtere Fit der *Modelle 6a und 6b* ist vor diesem Hintergrund nicht überraschend (vgl. Tabelle 3; die Modelle sind in Abbildung 3 nicht enthalten): Es geht hier um die in Abschnitt 2.2.2 als Modell D beschriebene Praxis, Valenzitems *entweder* als Indikatoren von PA *oder* als Indikatoren für den negativen Pol von NA zu verwenden. Formal handelt es sich dabei um Vereinfachungen des Modells 6, indem *entweder* nur Ladungen im PA-Faktor (Modell 6a) *oder* im NA-Faktor (Modell 6b) zugelassen werden. Wie die Fit-Indizes in Tabelle 3 zeigen, sind diese Modelle völlig unakzeptabel.

²⁶ Alle Berechnungen wurden auch mit freien Parametern durchgeführt. Dadurch verschlechterten sich aus begrifflichen Gründen (Verlust an Freiheitsgraden) sämtliche Fit-Indizes. Die Verschlechterung ändert aber nichts Wesentliches am Bild, das aus Tabelle 3 ersichtlich ist. Die Indizes lauten beispielsweise beim Modell 7, dem besten Modell aller acht Skalen: $\chi^2(16, N = 269) = 30.0, p = .2$; $\chi^2/df = 1.9$; CFI = .99; RMSEA = .06; PCLOSE = .32; AIC = 70.0; beim hier besonders wichtigen Zwei-Faktorenmodell 6: $\chi^2(15, N = 269) = 119.4, p < .001$; $\chi^2/df = 8.0$; CFI = .94; RMSEA = .16; PCLOSE < .001; AIC = 161.4.

Modellvorstellungen, die lediglich zwei Faktoren postulieren, sind somit nicht in der Lage, die Kovarianzstruktur der Daten angemessen zu rekonstruieren; die explorative Faktorenanalyse führt hier offenbar in die Irre. Es ist notwendig, einen dritten Faktor, einen VA-Faktor, zuzulassen, was im Folgenden in drei Varianten geschieht (vgl. Modelle 7 bis 9 in Abbildung 3):

Modell 7 entspricht dem Modell C in Abschnitt 2.2.2, das von der Annahme von drei einander nebengeordneten, korrelierten Faktoren besteht. Dieses Modell liefert (deskriptiv) die besten Fit-Indizes aller geprüften Modelle: χ^2 (31, N = 269) = 31.2, $p = .46$; $\chi^2/df = 1.01$; CFI = 1.00; RMSEA = .01; PCLOSE = .97; AIC = 41.2.

Modell 8 entspricht Modell B in Abschnitt 2.2.2 und nimmt an, der VA-Faktor sei den beiden anderen Faktoren (PA- und NA-Faktor) hierarchisch übergeordnet. Im Unterschied zu üblichen hierarchischen Modellen wird hier aber davon ausgegangen, dass auch für den übergeordneten Faktor empirische Indikatoren existieren. Die Fit-Indizes dieses Modells liegen ebenfalls noch in einem sehr akzeptablen Bereich, sind aber etwas schlechter als bei Modell 7: χ^2 (29, N = 269) = 35.9, $p = .18$; $\chi^2/df = 1.24$; CFI = 1.00; RMSEA = .03; PCLOSE = .86; AIC = 49.9.²⁷

Modell 9 schliesslich ist die einfachste denkbare Umsetzung der in Abschnitt 2.2.2 als Modell E skizzierten Vorstellung, der Valenz-Faktor habe eine summativ-evaluative Funktion. Es handelt sich nicht mehr um ein reines Faktorenmodell; vielmehr wird hier angenommen, die latente Variable "VA-Faktor" sei (unter anderem) von den latenten Variablen "PA-Faktor" und "NA-Faktor" (kausal) abhängig. Die Fit-Indizes dieses Modells sind ebenfalls sehr befriedigend und liegen zwischen jenen der Modelle 7 und 8: χ^2 (29, N = 269) = 31.2, $p = .36$; $\chi^2/df = 1.07$; CFI = 1.00; RMSEA = .02; PCLOSE = .95; AIC = 45.2. Erwähnt sei noch, dass der PA-Faktor und der NA-Faktor zusammen 74 % der Varianz des VA-Faktors aufklären, was zwar nicht unerheblich ist, aber doch auch auf die Existenz anderer Einflussfaktoren hinweist.

Alle drei Modelle, die in dieser oder jener Form drei Faktoren annehmen, zeigen somit sehr akzeptable Fit-Werte. Allein auf Grund dieser Werte (vgl. v. a. die AIC-Werte) ist zwar eine Rangierung möglich (Modell 7 vor Modell 9 vor Modell 8), aber kaum ein definitiver Entscheid zu rechtfertigen. Der Hauptbefund aus den Analysen dieses Abschnitts besteht vielmehr darin, dass die Drei-Faktorenmodelle (Modelle 7 bis 9) den Zwei-Faktorenmodellen (Modelle 6, 6a und 6b) eindeutig überlegen sind. Dies heisst, dass – bei der gegebenen Datengrundlage – die Annahme eines eigenständigen VA-Faktors sinnvoll ist.

Abbildung 3 liefert noch zwei wichtige Zusatzinformationen: Die erste betrifft die Beziehung der drei Faktoren untereinander: In Übereinstimmung mit den in Abschnitt 2.2.2 berichteten Befunden aus anderen Untersuchungen ist von starken und asymmetrischen Beziehungen zwischen

²⁷ Wären die beiden Modelle ineinander verschachtelt (sie sind es nicht), ergäbe der Signifikanztest des Fit-Unterschieds kein signifikantes Ergebnis ($\chi^2_{\text{Diff}} = 4.7$, $df = 2$, $p > .05$).

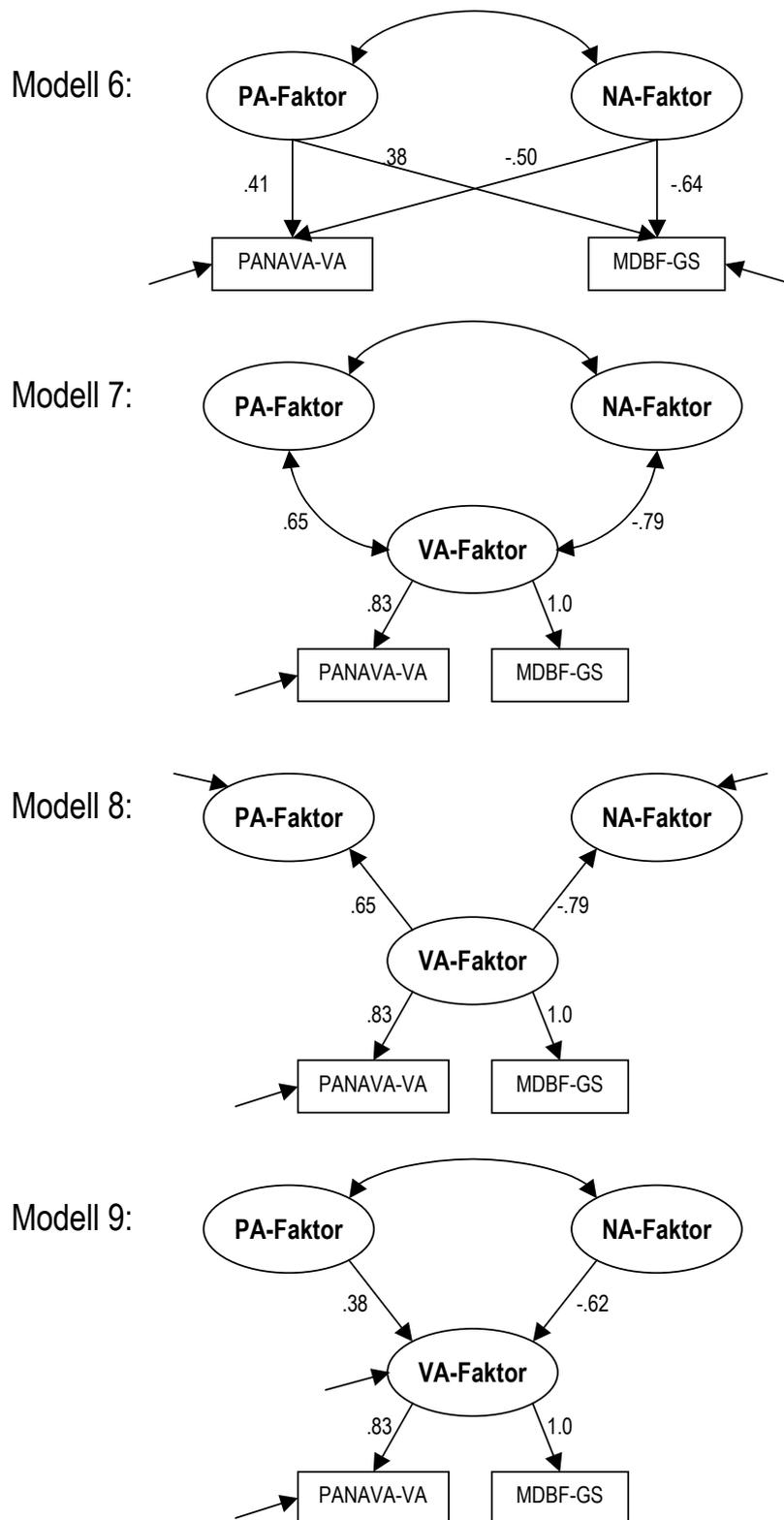


Abbildung 3. Modelle 6 bis 9 mit Parameterschätzungen (standardisiert; für alle angegebenen Parameter gilt $p < .001$)

Vereinfachte Darstellung: ohne die Messmodelle für PA und NA (Modell 1); die Uniqueness-Terme sind mit schrägen Pfeilen angedeutet.

dem PA- und dem NA-Faktor einerseits und dem VA-Faktor andererseits auszugehen: Die Korrelationen in Modell 7 betragen $r_{NA,VA} = -.79$; $r_{PA,VA} = .65$; die standardisierten Ladungen im hierarchischen Modell (Modell 8) sind von exakt derselben Grösse. Auch die Pfadkoeffizienten von Modell 9 liefern ein analoges Bild: Danach "beeinflusst" der NA-Faktor (-.62) den VA-Faktor deutlich stärker als der PA-Faktor (.38). Die Valenz ist somit stärker von der *Abwesenheit* von Negativer Aktivierung abhängig als von der *Anwesenheit* positiver Aktivierung.

Die zweite, hier wichtige Zusatzinformation betrifft die Ladungen von PANAVA-VA: Es überrascht nicht, dass die (viermal so lange) Skala MDBF-GS ein deutlich besserer Indikator des VA-Faktors ist als die Skala PANAVA-VA, die mit ihren lediglich zwei Items aber auch noch eine durchaus akzeptable Ladung von .83 aufweist (Modelle 7 bis 9).²⁸

3.4 Zusammenfassung und Bilanz aus Studie 1

Grundanliegen der Studie 1 war, abzuklären, ob und wie gut die PANAVA-Kurzskalen tatsächlich jene Konstrukte erfassen, die sie erfassen sollen. Zu diesem Zweck wurde ein Vergleich mit den PANAS- und AD-ACL-Skalen sowie der Skala MDBF-GS durchgeführt, die als relevante Standardskalen betrachtet werden können. Zwei Hauptfragestellungen standen dabei im Vordergrund: Die eine richtete sich auf die zugrunde liegende Konstruktstruktur, die andere darauf, wie gut die PANAVA-Kurzskalen die Konstrukte dieses Strukturmodells erfassen. Vorbereitende Analysen befassten sich mit den internen Konsistenzen der Skalen im Vergleich, mit der Rolle des Itemformats für die Skaleninterkorrelationen und mit einer explorativen Faktorenanalyse aller Skalen.

Die Hauptergebnisse lassen sich wie folgt zusammenfassen: 1.) Die interne Konsistenz der PANAVA-Kurzskalen ist – als Folge ihrer Kürze – etwas niedriger als jene der Vergleichsskalen, liegt aber (absolut gesehen) in einem für Forschungszwecke durchaus noch genügenden Bereich. 2.) Das im Bereich der Befindlichkeitsforschung eher unübliche bipolare Itemformat der PANAVA-KS wirkt sich nicht negativ aus, wenn zum Vergleich Skalen benutzt werden, die mit unipolaren Items jeweils beide Skalenpole abfragen (AD-ACL und MDBF-GS). Hingegen scheint der PANAS, der nur High-end-marker enthält, wegen dieser Tatsache im Rahmen sämtlicher einbezogener Instrumente eine Sonderstellung einzunehmen, die aller Wahrscheinlichkeit nach durch die fehlende Kontrolle der Akquieszenz zu erklären ist. 3.) Soweit sich dies mit einem Skalenvergleich überhaupt feststellen lässt, scheinen die PANAVA-KS die anvisierten Konstrukte auf sehr valide Weise zu erfassen. Folgende Teilbefunde sind hierbei von besonderer Bedeutung: a) Die MTMM-Analyse der PANAS- und der AD-ACL-Skalen sowie der Skalen PANAVA-PA und -NA zeigt, dass sich das Modell mit den zwei Konstrukten Positive und

²⁸ Die Ladung von 1.0 bei MDBF-GS (Modelle 7 bis 9) ist natürlich Folge davon, dass die Varianz der Uniqueness dieser Skala aus den früher angegebenen Gründen auf Null gesetzt wurde. Lässt man (in den Modellen 7 und 9, wo dies eine zulässige Lösung erzeugt) auch diesen Parameter unfixiert, treten aber nur sehr geringe Veränderungen auf (Ladung von MDBF-GS: .99; von PANAVA-KS: .84).

Negative Aktivierung klar bewährt, und dass PANAVA-PA und -NA mit Ladungen um .9 sehr valide Indikatoren dieser Konstrukte darstellen. b) Werden zusätzlich die Skalen PANAVA-VA und MDBF-GS in die konfirmatorischen Analysen einbezogen, zeigt sich zudem, dass die Annahme von drei Faktoren notwendig ist, um die Daten befriedigend zu erklären, die Valenzskalen also einen eigenständigen Varianzanteil des Befindens beschreiben. Dieser Befund steht im Gegensatz zum Ergebnis der explorativen Faktorenanalyse, dessen Zwei-Faktorenmodell jedoch die Bedingungen für ein gut angepasstes Modell nicht erfüllt. Offen bleiben musste allerdings die Frage, wie die Beziehung des VA-Faktors zum PA- und NA-Faktor genau beschaffen ist: Die drei geprüften Drei-Faktorenmodelle unterscheiden sich im Fit nur unwesentlich. Klar ist hingegen, dass PANAVA-VA – trotz der Kürze dieser Skala – diesen dritten Faktor mit einer Ladung von .83 ebenfalls valide erfasst.

Insgesamt fallen somit die Ergebnisse von Studie 1 für die PANAVA-KS sehr positiv aus: Offenbar muss man von einem Drei-Faktorenmodell des Befindens ausgehen – und es scheint gelungen, diese drei Faktoren in den PANAVA-KS konstruktvalide zu operationalisieren.

Bei all diesen Befunden ist aber die Natur dieser Untersuchung zu berücksichtigen: Es handelt sich um eine Ein-Punkt-Studie, weil ein Vorlegen so vieler Skalen in einer ESM-Studie nicht sinnvoll möglich wäre.²⁹ Der Preis dafür ist, dass in den hier bearbeiteten Daten inter- und intraindividuelle Varianz untrennbar konfundiert sind: Pro Person wurde ja nur ein – eventuell sehr spezifische Charakteristika aufweisender – Moment beschrieben. Ob die Ergebnisse auf ESM-Daten übertragbar sind, muss damit letztlich offen bleiben. Es ist gut denkbar, dass – wenn die beiden Ebenen getrennt untersucht würden – auf der Person- und/oder der Zeitpunktebene andere Ergebnisse resultieren als in dieser Studie. Entsprechende Hypothesen haben jedoch unterschiedliche Plausibilität: So ist es z.B. völlig plausibel, zu erwarten, dass die hier gefundenen Konsistenzkoeffizienten, die auf einem "Mix" der beiden Ebenen basieren, nicht direkt auf die beiden Ebenen übertragbar sind. Es ist vielmehr zu erwarten, dass die Skalen bei der Erfassung von Personmerkmalen (Personenebene) grössere Reliabilitäten aufweisen als bei der Beschreibung von States (Zeitpunktebene) (vgl. dazu auch Studie 2). Unterschiede in der einen oder anderen Richtung sind ebenfalls bei den Interkorrelationen der Skalen denkbar, wobei – wie bei den Reliabilitätskoeffizienten – auf der Zeitpunktebene auch mit interindividuellen Differenzen zu rechnen ist. Hingegen scheint es kaum gute Gründe dafür zu geben, anzunehmen, dass die *strukturellen Beziehungen* zwischen den verschiedenen Skalen, wie sie hier konfirmatorisch untersucht wurden, in ESM-Daten völlig anders aussehen. Dies bedeutet, dass die zentralen Befunde zur Verwandtschaft der PANAVA-Kurzskalen mit den einbezogenen etablierten Standardskalen wohl auch auf ESM-Daten übertragbar sein dürften. Wie weit dies auch bei anderen Ergebnissen der Studie 1 der Fall ist, wird sich in der nachfolgend beschriebenen Studie 2 erweisen, in der die Bewährung der PANAVA-KS in ESM-Daten untersucht wird.

²⁹ Es sei hier daran erinnert, dass die in Abschnitt 2 referierten Vergleichsstudien in der Regel nicht die Originalskalen benützen, sondern lediglich eine Auswahl mehr oder weniger plausibler Markieritems.

4. Studie 2: Untersuchungen zu psychometrischen Eigenschaften der PANAVA-KS in ESM-Daten

4.1 Zielsetzungen und Vorblick

Ziel der Studie 2 ist es, die Bewährung der PANAVA-Kurzskalen in ESM-Daten zu überprüfen. Im Unterschied zur psychometrischen Überprüfung konventioneller Skalen existieren für diesen Fall (noch) keine Standards. Von der Problemlage her gesehen sind praktisch beliebig viele Analysen möglich. Dies hat vor allem drei Gründe: Der erste besteht in der Mehrebenenstruktur von ESM-Daten (vgl. Abschnitt 2.1), wobei man diese unterschiedlich differenziert berücksichtigen kann, beispielsweise in Form eines Zwei-Ebenenmodells: Zeitpunkte verschachtelt (nested) in Personen – oder eines Drei-Ebenenmodells: Zeitpunkte verschachtelt in Wochentage verschachtelt in Personen. Es ist denkbar, ja zu erwarten, dass sich die psychometrischen Kennwerte von Ebene zu Ebene unterscheiden. So können zum Beispiel Item- und Skaleninterkorrelationen auf der Personenebene anders ausfallen als auf der Zeitpunktebene (z.B. Schmitz, 2000) oder in der Menge der Zeitpunkte von entspannten Wochenenden anders als in jenen von stressreichen Arbeitstagen (Reich et al., 2003).

Der zweite Grund besteht darin, dass die Kennwerte pro Ebene keine Konstanten darstellen müssen. Vielmehr ist damit zu rechnen, dass sie sich zwischen den Einheiten der nächst höheren Ebene unterscheiden. Beispielsweise dürften im Zwei-Ebenenmodell die psychometrischen Kennwerte auf der Zeitpunktebene (z.B. Itemmittelwerte und -standardabweichungen, Skalenreliabilitäten und -interkorrelationen pro Person) von Person zu Person verschiedene Werte annehmen. Zu jedem Kennwert existieren somit Verteilungen, die interessante Fragestellungen aufwerfen, die in konventionellen Analysen keine Parallele haben. Am offensichtlichsten ist dies bei den individuellen Mittelwerten von Items und Skalen, die in ESM-Untersuchungen seit jeher als diagnostisch bedeutsame Personkennwerte Berücksichtigung fanden. Neuerdings wird aber oft betont, dass ebenfalls die individuellen Standardabweichungen als Persönlichkeitsmerkmale mit eigenständigem Informationsgehalt zu betrachten sind (z.B. Penner et al., 1994; McConvill & Cooper, 1999; Eid & Diener, 1999; Ghisletta et al., 2002). Es gibt keinen Anlass, a priori anzunehmen, dass dies nicht auch für andere individuelle Kennwerte (z.B. individuelle Korrelationen oder Reliabilitäten) gilt. Dies bedeutet aber, dass eigentlich die psychometrischen Eigenschaften all dieser Kennwerte untersucht werden könnten.

Der dritte Grund für die mögliche Inflation psychometrischer Analysen in ESM-Daten wurzelt darin, dass solche Daten nicht nur hierarchisch verschachtelt sind, sondern auf der Zeitpunktebene zusätzlich den Charakter von Zeitreihen haben. Das heisst, dass auf dieser Ebene Phänomene wie Autokorrelationen, Trends und/oder Zyklen zu beobachten sein dürften (vgl. z.B. Fitzmaurice, Laird & Ware, 2004) – Phänomene, die aus verschiedenen Perspektiven von Interesse sind: als generelle Phänomene (z.B.: Gibt es einen Tageszyklus im Befinden?), als differenzielle Phänomene (z.B. Frage nach interindividuellen Unterschieden in solchen Zyklen oder

in Trends) sowie als methodische "Störfaktoren" bei anderen Analysen. So zeigen beispielsweise West und Hepworth (1991) sehr eindrücklich, wie Zeiteffekte Korrelationen zwischen zwei Variablen (z.B. wenn beide Variablen den gleichen zeitlichen Trend aufweisen) erhöhen oder (z.B. wenn die eine der beiden Variablen einen Tageszyklus aufweist, die andere jedoch nicht) senken können.

Im Rahmen dieses Papiers ist es weder möglich noch sinnvoll, all diese Aspekte gleichzeitig zu berücksichtigen. Vielmehr erfolgt eine Konzentration auf einige Basisinformationen, die durch die traditionellen psychometrischen Konzepte nahe gelegt werden. In diesem Zusammenhang interessieren auch Indizien dafür, inwieweit die bei der Skalenentwicklung angestrebten Ziele (z.B. Augenscheinvalidität, Antwortverteilungen, Akquieszenzkontrolle) erreicht worden sind. Die *Mehrebenenstruktur* der Daten wird dabei – wo notwendig – in der für psychometrische Analysen plausibelsten Form berücksichtigt, nämlich in Form des oben genannten Zweiebenenmodells (Zeitpunktebene verschachtelt in die Personenebene). Hingegen kann hier der *Zeitreihencharakter* von ESM-Daten nicht explizit berücksichtigt werden. Dies entspricht – mit wenigen Ausnahmen – der gängigen Praxis (vgl. z.B. die Übersicht über moderne Auswertungsstrategien von ESM-Daten bei Tennen, Affleck & Armeli, 2005). Hauptgrund für diese Praxis dürfte sein, dass es (noch) keine Verfahren gibt, mit denen sämtliche, bei der Analyse von ESM-Daten potentiell relevanten Zeiteffekte sozusagen routinemäßig kontrolliert werden können. Eine solche Kontrolle würde vielmehr eigenständige Abklärungen verlangen, die rasch einmal sehr umfangreich würden.³⁰ Um den Rahmen dieses Papiers nicht zu sprengen, wird im Folgenden daher lediglich auf einen theoretisch besonders interessanten Typ von Zeiteffekten, auf die Frage nach der Existenz von Zyklen, eingegangen. Diese (weitgehende) Nichtberücksichtigung von Zeiteffekten bedeutet, dass in den nachstehend berichteten Ergebnissen die allenfalls vorhandenen Zeiteffekte nicht als separate Effekte thematisiert, sondern mit allen anderen intra-individuell vorhandenen Varianz- (und Varianzeinschränkungs-) Quellen zusammengefasst werden.

Im Detail werden vor diesem Hintergrund im Rahmen der Studie 2 – nach der Darstellung der Datengrundlage (Abschnitt 4.2) – folgende Themen behandelt: die Augenscheinvalidität und Akzeptanz der PANAVA-Items aus der Sicht der Untersuchungspersonen (Abschnitt 4.3); die Verteilungskennwerte der Items und der Skalen (Abschnitt 4.4) und die Faktorenstruktur der Items (Abschnitt 4.5). Dann folgen Untersuchungen zu den Skalenreliabilitäten (Abschnitt 4.6) und Skaleninterkorrelationen (Abschnitt 4.7). Bei all diesen Themen werden auch Fragen aufgegriffen, die sich erst auf Grund der Befunde ergeben. Ein ganzer Themenkomplex dieser Art betrifft die Rolle der Aggregation für die Beziehung von PA, NA und VA auf der Personenebene und die differentielle Validität der Skalen auf dieser Ebene, der in Abschnitt 4.8 behandelt wird. In Abschnitt 4.9 wird schliesslich noch untersucht, ob sich Hinweise auf das Bestehen von Ta-

³⁰ Die Schwierigkeit liegt vor allem darin, dass die relevante Problemstruktur bei ESM-Daten weder mit jener bei typischen Zeitreihenanalysen noch mit jener bei modernen Growth-curve-Modellen übereinstimmt. In beiden Fällen wären Lösungen verfügbar. Es ist geplant, die angesprochenen Abklärungen später nachzuholen.

ges-, Wochen- und Jahresrhythmen ergeben. Eine zusammenfassende Bilanz beschliesst die Darstellung von Studie 2 (Abschnitt 4.10).

4.2 Die Datengrundlage

Die Datengrundlage von Studie 2 wurde in einer Serie von ESM-Studien mit sehr unterschiedlichen Zielsetzungen gewonnen, wobei die Untersuchungspersonen auf ganz unterschiedliche Weise rekrutiert worden sind (z.B. Aushänge in Betrieben, Schneeballsystem, Zeitungsinserte usw.). Die Teilnahme war in jedem Fall freiwillig und wurde zum Teil durch einen symbolischen Geldbetrag entschädigt. Alle diese Studien umfassten vier Phasen: 1.) Eine mündliche Instruktion, 2.) einen Probetag mit drei Signalzeitpunkten mit anschliessender persönlicher Besprechung allfälliger Schwierigkeiten bzw. Unklarheiten beim Ausfüllen des Experience Sampling Fragebogens (ESF), 3.) die eigentliche ESM-Untersuchung und 4.) (im Anschluss daran) das Beantworten eines konventionellen Fragebogens. Die eigentliche ESM-Untersuchung (Phase 3) dauerte in allen Fällen mindestens eine Woche. Im Folgenden wird – wenn nicht anders vermerkt – auch von längeren Untersuchungsperioden nur jeweils die erste Untersuchungswoche berücksichtigt. Die von den UPn angegebene voraussichtliche Wachzeit wurde pro Tag in sieben gleich lange Intervalle aufgeteilt. Für jedes Intervall wurde mittels eines Zufallszahlengenerators ein Signalzeitpunkt bestimmt, der einen Mindestabstand von 1/2 Stunde vom vorhergehenden Signal aufwies. Als Signalgeber dienten je nach Wunsch der Untersuchungsperson eine programmierte Uhr oder ein Pager. Die Experience Sampling Fragebogen (ESF) wurden in drei A6-Heften (zwei Hefte für je zwei Tage, eines für drei Tage), die immer mit sich zu führen waren, zusammengefasst. Pro Signal war eine Doppelseite mit rund 40 Fragen – darunter die 10 PANAVA-KS-Items – auszufüllen (Ausfüllzeit ca. 3 bis 4 Minuten).

Total haben 561 Personen mit der ESM-Untersuchung begonnen; 19 Personen (3.4 %) haben diese vorzeitig abgebrochen³¹; 2 Personen (0.4 %) teilten nach der Untersuchungswoche mit, das gesamte Material (inkl. Signalgeber) sei verloren gegangen (z.B. es sei in einem Bus liegen gelassen worden). 10 weitere Personen (2.8 %) wurden nach Durchsicht der ESF-Hefte ausgeschlossen: 2 Rentner(innen) hatten offensichtlich die Instruktion nicht richtig verstanden und bei bipolaren Rating-Skalen praktisch durchwegs zwei Kreuze gesetzt; 2 Personen hatten den Grossteil der ESF nur sehr selektiv bearbeitet und 6 Personen weniger als 30 ESF ausgefüllt. Der Grenzwert von 30 wurde auf Grund einer empirischen Verteilungsanalyse festgesetzt: Sie zeigte, dass Werte unter 30 den Charakter eigentlicher Ausreisser haben: Die meisten Untersuchungspersonen haben sehr viel mehr ESF ausgefüllt (siehe unten).

Von den ESF der verbleibenden 530 Personen wurden total 2 % ausgeschlossen. Zwei Ausschlusskriterien wurden benutzt: 1.) mehr als fünf fehlende Werte bei den PANAVA-Items (Ausschlussgrund für 1.1 % der ESF) und 2.) mehr als 120 Minuten Zeitabstand zwischen Sig-

³¹ Bei 5 dieser 19 Personen handelt es sich um Rentner(innen), von denen 3 aus gesundheitlichen Gründen und 2 wegen Hörproblemen (das Signal der Uhr wurde nicht gehört) die Teilnahme abbrachen.

nal und Zeitpunkt des Ausfüllens (Ausschlussgrund für 0.9 % der ESF). Auch hier wurden die Grenzen so gesetzt, dass es sich bei den ausgeschlossenen Fällen um klare Ausreisser handelt. Sie haben zudem Gründe, die – studiert man die betreffenden ESF genauer – in vielen Fällen nachvollziehbar sind. Beispielsweise ist es sehr verständlich, dass eine Person, die durch das Signal geweckt wurde, die Befindlichkeitsitems nicht ausfüllt. Analog ist ein Zeitabstand von mehr als 120 Minuten bei einer (sehr langen) Sitzung oder während eines Interviews (oft) nicht zu umgehen. Dasselbe gilt übrigens für viele ESF, die mit einer noch unter zwei Stunden liegenden grossen zeitlichen Verzögerung geantwortet haben. Der reale Alltag berufstätiger Personen verhindert oft ein sofortiges Ausfüllen des ESF.

Da von sämtlichen Personen auch nach dieser Datenbereinigung die oben genannte Mindestzahl von 30 ESF vorlagen, besteht die *definitive Stichprobe* aus 530 Personen (49.8 % Frauen). Das Durchschnittsalter beträgt 40.5 Jahre ($SD = 13.7$ Jahre; Streubereich: 18 bis 83 Jahre). Die meisten UPn (73 %) sind berufstätig; der Rest verteilt sich auf Lehrlinge und Studierende (13 %), Arbeitslose (4 %), Hausfrauen (2 %) und Rentner(innen) (8 %). Die berufliche Stellung der Berufstätigen verteilt sich ungefähr zu gleichen Teilen auf das obere Kader, freie Berufe und Selbständige (35 %), mittleres Kader (36 %) und qualifizierte und unqualifizierte Angestellte und Arbeiter (29%).

Die 530 Personen haben für die 49 (7 mal 7) Signalzeitpunkte im Durchschnitt 44.4 ESF ($SD = 3.95$) ausgefüllt, also knapp 91 %. Dies ergibt total 23'509 ESF. Der Zeitabstand zum Signal ist extrem schief verteilt: Der Modus (11'275 oder 48 %) liegt bei 0 Minuten (sofortiges Ausfüllen), der Median bei 1 Minute. Weil bei einer Befindlichkeitsuntersuchung potentiell störend, interessiert vor allem auch die Häufigkeit langer Zeitabstände: Während 87 % der ESF innerhalb einer Viertelstunde ausgefüllt wurden, beträgt der Prozentsatz für 16 bis 30 Minuten 6.9 %, für 31 bis 60 Minuten 4.1 % und für 61 bis 120 Minuten 1.6 %.³²

Ein letzter Punkt betrifft die Berechnung der Skalenwerte der PANAVA-Kurzskalen: Da fehlende Werte bei den PANAVA-Items (nach der oben beschriebenen Datenbereinigung) sehr selten sind – lediglich 0.9 % der einbezogenen ESF enthalten noch fehlende Werte (1 fehlender Wert: 0.8 %; 2 fehlende Werte: 0.1 %) – wurden sie nicht explizit ersetzt. Vielmehr wurden die Skalenwerte als Mittelwert der vorhandenen Items berechnet und dann auf die Gesamtzahl der Items pro Skala (PA, NA: 4; VA: 2) hoch gerechnet.

³² Man kann sich natürlich fragen, ob ein Zeitabstand von bis gegen 2 Stunden zwischen zu registrierender Situation und Protokollierung für eine Untersuchung über das Befinden im Alltag nicht zu lange ist. Delespaul (1995) empfiehlt auf Grund empirischer Indizien ein tolerierbares Zeitfenster von lediglich ca. 15 Minuten. Drei Gründe führten zu unserer grosszügigen Regelung: Erstens kann – wie in den obigen Beispielen sichtbar wird – ein derart enges Zeitfenster zum Ausschluss ganzer Situationsklassen führen. Zweitens wurden die Untersuchungspersonen ausdrücklich instruiert, sich einen Zeitpunkt, der nicht sofort protokolliert werden kann, genau zu merken ("einzufrieren"). Und drittens machen in unseren Daten – wie gezeigt – echt lange Zeitabstände nur einen relativ geringen Anteil aus. Hier liegt aber sicher noch ein Ansatzpunkt für vertiefende Untersuchungen.

4.3 Augenscheinvalidität und Akzeptanz der PANAVA-Items

Wie in Abschnitt 2.1 ausgeführt, sind in ESM-Untersuchungen im Interesse der resultierenden Datenqualität besondere Anforderungen an die Items und Skalen zu stellen. Insbesondere sollten die UPn die immer wieder zu bearbeitenden Items als sinnvoll erleben können. Entsprechend wurde bei der Entwicklung der PANAVA-Items versucht, die Befindlichkeit aus alltagspsychologischer Sicht möglichst vollständig und augenscheinvalide abzudecken sowie Items zu vermeiden, die (aus verschiedenen Gründen; vgl. Abschnitt 2.1) sozusagen zum Überspringen einladen.

Um zu überprüfen, ob bzw. wie gut diese Ziele erreicht worden sind, dienen hier zwei Kriterien: einerseits eine direkte Frage an die UPn, andererseits als objektiver Indikator die Anzahl fehlender Werte bei den einzelnen Items.

Die direkte Frage an die UPn, die im Rahmen der konventionellen Befragung nach Abschluss der ESM-Untersuchung präsentiert worden ist, lautete, ob es möglich gewesen sei, "... das Befinden im Alltag ... [im ESF, der Verf.] ... wirklichkeitsgetreu wiederzugeben". Die überwiegende Mehrzahl, nämlich 74 % der Befragten beantworteten diese Frage positiv ("ja": 17.1 %; "eher ja": 56.8 %), 16 % wählten die Antwortvorgabe "unentschieden" und nur 10 % antworteten negativ ("eher nein": 9.5 %, "nein": 0.6 %). Auch der objektive Indikator für die Akzeptanz der Items, die Anzahl fehlender Werte, liefert ein positives Bild: Pro Item gibt es – berechnet über alle ESF ($N = 23'509$) – lediglich zwischen 0.07 % und 0.17% fehlende Angaben. Wenn man sich die Belastung der UPn durch die ESM-Untersuchung vergegenwärtigt, handelt es sich dabei um eine vernachlässigbare Grössenordnung.

Insgesamt sprechen diese Ergebnisse dafür, dass es gelungen ist, mit den 10 Items ein breites Spektrum von Befindlichkeitsaspekten abzufragen, gegen deren Beantwortung in allen möglichen Alltagssituationen nichts zu sprechen scheint.

4.4 Verteilungskennwerte und Intraclass-Korrelationen der Items und Skalen

Ein anderes Ziel, das bei der Entwicklung der PANAVA-Items verfolgt wurde, war, den in der Befindlichkeitsmessung typischen, oft extrem asymmetrischen Verteilungen vorzubeugen (vgl. Abschnitt 2.3). Die in Tabelle 4 wiedergegebenen Item-Mittelwerte und -Standardabweichungen zeigen, dass dies weit gehend gelungen ist: Die *Mittelwerte* M liegen zwar alle – konsistent mit den Befunden vieler Befindlichkeitsstudien (vgl. z.B. Staudinger, 2000) – im jeweils moderat positiven Bereich, d.h. in Richtung des positiv formulierten Pols über bzw. unter der theoretischen Skalenmitte von 4. Sie liegen aber durchwegs noch so weit vom Minimum bzw. Maximum (1 bzw. 7) entfernt, dass der Streubereich nicht entscheidend eingeschränkt erscheint. In einem scharfen Kontrast zu den Verhältnissen bei unipolaren Items tritt übrigens auch in keinem Item eine J-Verteilung auf: Der Modus (in Tabelle 4 nicht angegeben) liegt in allen Fällen mindestens ein Punkt von den Extremwerten entfernt.

Bei den *Standardabweichungen* in Tabelle 4 sind drei verschiedene Berechnungsarten unterschieden:³³ Die Angabe in der Kolonne "total" bezieht sich auf die Gesamtheit der Zeitpunkte ($N = 23'509$) ohne Berücksichtigung der hierarchischen Datenstruktur. Die entsprechenden Varianzen entstammen somit zwei verschiedenen Quellen, den *interindividuellen* Unterschieden in der (durchschnittlichen) Befindlichkeit (Varianz auf der Ebene der Personen) und den *intraindividuellen* Unterschieden von Zeitpunkt zu Zeitpunkt innerhalb der Personen (Varianz auf der Zeitpunktebene). Teilt man die Totalvarianz in diese beiden Varianzkomponenten auf, ergeben sich die in den Kolonnen "Personen" und "Zeitpunkte" angegebenen Standardabweichungen. (Das für diese Aufteilung verwendete mehrebenenanalytische Modell ist in Anhang A1 beschrieben.) Sie lehren, dass die Itemstreuungen auf der Zeitpunktebene durchgängig deutlich grösser sind als auf der Personenebene. Dasselbe zeigen die *Intraclass-Korrelationen* ρ_1 (Kolonne "beobachtet"), die direkt den Anteil der "personbedingten Varianz" (= Varianz auf der Personenebene) an der Totalvarianz ausdrücken. Diese Koeffizienten liegen bei den Items im Bereich zwischen .16 und .34, wobei auffällt, dass die NA- (und VA-) Items im Allgemeinen etwas grössere Werte liefern als die PA-Items. Dieser Unterschied schlägt dann auch bei den Skalenwerten durch: Während bei PA 20 % der Varianz personbedingt sind, sind es bei VA 26 % und bei NA 28 %. Mit anderen Worten: PA ist etwas stärker situationsabhängig als NA und VA.

Bei den besprochenen Intraclass-Korrelationen ist zu berücksichtigen, dass die Varianz auf der Zeitpunktebene auch die Messfehler umfasst, welche durch die nicht-perfekte Reliabilität der Skalen entstehen. Für die Koeffizienten der Skalen ist hier eine Art Minderungskorrektur möglich, indem der Varianzanteil, der durch die Skalenunreliabilität erzeugt wird, von der Gesamtvarianz subtrahiert wird. (Das verwendete mehrebenenanalytische Modell ist in Anhang A2 dargestellt.) Dies ergibt Intraclass-Korrelationen, die sich auf die latente Variable beziehen, die durch die Items einer Skala erfasst wird. Sie sind in der letzten Spalte von Tabelle 4 angegeben und fallen naturgemäss grösser aus. Der Unterschied ist aber nicht wirklich massiv: Auch die Varianz der (messfehlerfrei gemessenen) Konstrukte ist – je nach Skala – "nur" zu 25 % bis 34 % auf Personunterschiede zurückzuführen; 66 % (NA und VA) bzw. 75 % (PA) sind durch intraindividuelle (situationsbedingte) Unterschiede zu erklären.

Wie zu erwarten (und oben implizit berücksichtigt), unterscheiden sich naturgemäss die Mittelwerte und Standardabweichungen der Items und Skalen von Person zu Person. Eine Übersicht über die Verteilungskennwerte dieser Kennwerte in der Menge der Personen findet sich in Tabelle B1 des Anhangs B.

³³ Bei den Mittelwerten ist diese Aufteilung nicht nötig, weil sich (aus logischen Gründen) auf der Zeitpunktebene und der Personenebene identische Mittelwerte ergeben, die vom Mittelwert in der Gesamtheit der ESF nur unwesentlich abweichen. Diese Abweichung ergibt sich daraus, dass die Anzahl der ESF nicht bei allen Personen dieselbe ist, bei der Bestimmung des Mittelwerts in der Gesamtheit der ESF die Personen also mit unterschiedlichem Gewicht einflüssen. In Tabelle 4 ist jeweils der Mittelwert der Mittelwerte der Personen angegeben

Tabelle 4

Verteilungskennwerte und Intraclass-Korrelationen ρ_i der Items und Skalen

Skala bzw. Item (Nummer) ^a :	M	SD			ρ_i	
		total	Personen	Zeitpunkte	beobachtet	latent
<i>Positive Aktivierung (PA)</i>	18.79	4.79	2.15	4.28	.20	.25
energieelos-energiegeladen (2*)	4.84	1.46	0.61	1.33	.17	
müde-hellwach (4)	4.34	1.84	0.79	1.66	.19	
lustlos-hoch motiviert (7)	4.86	1.37	0.57	1.24	.17	
gelangweilt-begeistert (9*)	4.75	1.25	0.50	1.14	.16	
<i>Negative Aktivierung (NA)</i>	11.04	5.01	2.63	4.27	.28	.34
entspannt-gestresst (3*)	3.08	1.70	0.73	1.54	.18	
friedlich-verärgert (5)	2.34	1.40	0.65	1.24	.22	
ruhig-nervös (8)	2.58	1.49	0.73	1.29	.24	
sorgenfrei-besorgt (10*)	3.04	1.60	0.93	1.30	.34	
<i>Valenz (VA)</i>	10.91	2.34	1.19	2.02	.26	.34
unzufrieden-zufrieden (1*)	5.63	1.35	0.56	1.22	.18	
unglücklich-glücklich (6)	5.28	1.23	0.69	1.03	.31	

Anmerkungen: $N = 23'509$ Zeitpunkte von 530 Personen mit $M = 44.4$ Zeitpunkten ($SD = 4.0$). Die Antwortvorgaben der Items sind mit 1 bis 7 codiert; 4 entspricht dem neutralen Mittel. Die Skalenwerte sind Summenscores; der mögliche Streubereich beträgt bei PA und NA 4 bis 28 (Mitte: 16), bei VA 2 bis 14 (Mitte: 8). SD total, SD Personen, SD Zeitpunkte = Standardabweichung total, auf der Person- bzw. auf der Zeitpunktebene. ρ_i = Intraclass-Korrelation. ρ_i latent = ρ_i bezogen auf latente Variablen.

^a Die Itemnummern entsprechen der Darbietungsreihenfolge (* = umgepolt dargeboten).

Zusammenfassend können als Hauptergebnisse festgehalten werden, dass die Verteilungseigenschaften der Items und Skalen der PANAVA-KS befriedigend ausfallen – ferner bei allen Items und (als Folge) bei den Skalen der Hauptanteil der Varianz auf intraindividuelle Unterschiede, d.h. auf Unterschiede auf der Zeitpunktebene *innerhalb* der Personen, zurückzuführen ist – also durch jene Unterschiede, die man in einer ESM-Studie durch die wiederholten Messungen meist speziell ins Visier nimmt.

4.5 Explorative Faktorenanalysen der Items auf der Zeitpunkt- und der Personenebene

Die leitende Frage dieses Abschnitts ist, ob die Faktorenstruktur der PANAVA-Items die Bildung der drei Skalen PA, NA und VA stützt.³⁴ Diese Frage soll hier mit Hilfe von explorativen Faktorenanalysen angegangen werden, und zwar sowohl auf der Zeitpunkt- als auch auf der Personenebene: Auf der Personenebene geht es um die Faktorenstruktur der Items bei der Beschreibung interindividueller Unterschiede, auf der Zeitpunktebene um die Faktorenstruktur der Items bei der Beschreibung intraindividuelle Unterschiede. Eine Komplikation bildet dabei die Tatsache, dass sich die Faktorenstruktur auf der Zeitpunktebene zwischen den verschiedenen Individuen unterscheiden kann. Um trotzdem einen allgemeinen Anhaltspunkt zu bekommen, wird

³⁴ Die Frage nach der Faktorenstruktur von Variablen in hierarchisch strukturierten Zeitreihen-Daten (mit ungleichen Zeitabständen) wirft eine Fülle methodischer Fragen auf (vgl. z.B. Nesselroade & Molenaar, 1999). Die folgenden Analysen haben heuristischen Charakter.

hier die Idee des in der Arbeitsgruppe von Watson (z.B. Watson & Clark, 1997; Watson et al., 1999) in solchen Fällen angewandten Vorgehens übernommen. Diese Idee besteht darin, die Matrix der *durchschnittlichen* intraindividuellen Interkorrelationen einer Faktorenanalyse zu unterziehen (was in Cattels Taxonomie einer P-Faktorenanalyse entspricht).³⁵ Das Resultat widerspiegelt dann die Faktorenstruktur der Items auf der Zeitpunktebene, wie sie beim durchschnittlichen Individuum zu finden ist.

Als Ausgangspunkt der Faktorenanalysen werden im Folgenden Item-Interkorrelationsmatrizen benutzt, die auf einer mehrebenenanalytischen Zerlegung der Varianzen und Kovarianzen auf die Person- und die Zeitpunktebene beruhen (vgl. z.B. Hox, 2002). (Das verwendete mehrebenenanalytische Modell ist in Anhang A1 beschrieben.) Etwas vereinfachend kann man sagen, dass die Matrix auf der Personenebene den Korrelationen der individuellen Itemmittelwerte in der Menge der Personen entspricht, die Matrix auf der Zeitpunktebene den durchschnittlichen intraindividuellen Iteminterkorrelationen. Diese beiden ebenenspezifischen Korrelationsmatrizen (die im Anhang B in Tabelle B2 zu finden sind) werden analog zu Studie 1 einer Hauptkomponentenanalyse mit anschliessender Promax-Rotation unterzogen.

Werden alle 10 Items faktorenanalysiert, resultieren auf beiden Ebenen Befunde, die strukturell exakt jenen der explorativen Faktorenanalyse aller acht Instrumente in Studie 1 (Abschnitt 3.3.3) entsprechen: Erstens implizieren die Eigenwertverläufe eine Zweifaktoren-Lösung (Zeitpunktebene: 4.08, 2.26, 0.68, 0.6 ...; Personenebene: 6.20, 1.36, 0.81, 0.48 ...), wobei die zwei Faktoren eindeutig als PA und NA zu interpretieren sind. Zweitens laden die VA-Items klar in beiden Faktoren, wobei die Ladungen im NA-Faktor etwas grösser sind als im PA-Faktor. Diese (hier nicht im Detail wiedergegebene) Analyse spricht also – wie die analoge Analyse beim Instrumentenvergleich – für die Notwendigkeit lediglich von zwei Skalen. Da aber der Einbezug der Valenzitems in die Messung der Konzepte Positive und/oder Negative Aktivierung klar im Widerspruch zur Konzeption von PA und NA im Sinne des PANAS stünde (vgl. Abschnitt 2.3), wurde bereits bei der Entwicklung der PANAVA-KS von einem solchen Einbezug Abstand genommen (vgl. Abschnitt 2.4) – ein Entscheid, der durch die Ergebnisse der konfirmatorischen Analysen in Abschnitt 3.3.5 zusätzlich gestützt wird.

Vor diesem Hintergrund werden im Folgenden die Valenzitems aus der explorativen Faktorenanalyse ausgeschlossen. (Die Frage nach der Redundanz der Valenzskala wird später durch eine Analyse der Skaleninterkorrelationen wieder aufgegriffen; vgl. Abschnitte 4.7 und 4.8.)

Auch bei der Beschränkung auf die acht PA- und NA-Items sprechen die Eigenwertverläufe auf beiden Ebenen für eine Zweifaktorenlösung (Zeitpunktebene: 2.98, 2.21, 0.64, 0.58 ...; Personenebene: 4.77, 1.34, 0.67, 0.44 ...). Vergleicht man die Befunde der beiden Ebenen (Tabelle 5), so

³⁵ Diese Idee wird hier allerdings durch ein anderes Vorgehen realisiert: In den genannten Arbeiten werden die durchschnittlichen Interkorrelationen geschätzt, indem die Korrelationen in der Gesamtheit der intraindividuell z-standardisierten Zeitpunktdaten berechnet werden. Hier wird dagegen mehrebenenanalytisch vorgegangen (siehe unten).

Tabelle 5
Promax-rotierte Hauptkomponenten der PA- und NA-Items

Item	Zeitpunktebene $r_{F1,F2} = -.15$					Personebene $r_{F1,F2} = -.55$				
	Patternkoeff.		Strukturkoeff.		h^2	Patternkoeff.		Strukturkoeff.		h^2
	F1	F2	F1	F2		F1	F2	F1	F2	
Positive Aktivierung:										
energielos-energiegeladen	.16	.85	.03	.83	.71	.99	.14	.92	-.40	.85
müde-hellwach	.14	.80	.02	.78	.62	.71	-.11	.77	-.50	.60
lustlos-hoch motiviert	-.12	.83	-.24	.85	.73	.93	-.03	.94	-.53	.89
gelangweilt-begeistert	-.25	.67	-.35	.71	.56	.80	-.08	.84	-.52	.71
Negative Aktivierung:										
entspannt-gestresst	.84	.09	.82	-.04	.68	.04	.90	-.45	.88	.77
friedlich-verärgert	.78	-.10	.80	-.22	.65	-.16	.78	-.59	.86	.76
ruhig-nervös	.83	.14	.81	.02	.67	.05	.94	-.46	.91	.83
sorgenfrei-besorgt	.70	-.12	.72	-.23	.53	.0	.82	-.45	.82	.67

Anmerkungen: $N = 23'509$ Zeitpunkte von 530 Personen mit $M = 44.4$ Zeitpunkten ($SD = 4.0$). F1 = Faktor 1; F2 = Faktor 2; $r_{F1,F2}$ = Korrelation der beiden Faktoren.

fallen sowohl Gemeinsamkeiten als auch Unterschiede auf: Gemeinsam ist, dass die Ladungen a der Items (Pattern-Koeffizienten) auf beiden Ebenen im Sinne der theoretischen Erwartungen ausfallen: Alle Items zeigen jeweils eine eindeutige Hauptladung ($.67 \leq a \leq .99$) in jenem Faktor, in dem auch die andern Items der jeweiligen Skala laden. Die Nebenladungen sind dagegen klein; sie erfüllen sogar das Kriterium, das bei der Itemauswahl des PANAS angewandt wurde ($|Nebenladungen| \leq .25$; vgl. Abschnitt 2.3). Die beiden Faktoren können entsprechend auf beiden Ebenen (wenn auch in unterschiedlicher Reihenfolge) eindeutig als PA und NA interpretiert werden. Ein deutlicher Unterschied zwischen den beiden Ebenen besteht aber in der Interkorrelation der beiden Faktoren: Während sie auf der Zeitpunktebene nur *schwach* negativ ($r_{F1,F2} = -.15$) korrelieren, korrelieren sie auf der Personebene *stark* negativ ($r_{F1,F2} = -.55$). Dies erklärt dann auch, warum die Struktur-Koeffizienten (Item-Faktor-Korrelationen) im jeweils anderen Faktor auf der Personebene deutlich grösser ausfallen als auf der Zeitpunktebene.

Bilanzierend kann man festhalten, dass die PA- und NA-Items sowohl auf der Person- wie (für das Durchschnittsindividuum) auf der Zeitpunktebene tatsächlich die intendierte Faktorenstruktur aufweisen, die Bildung der entsprechenden Skalen aus dieser Sicht also gerechtfertigt ist. Auf Grund der Ergebnisse von Studie 1 nicht unerwartet, nehmen dabei die Valenzitems eine Art Zwischenstellung ein und wurden daher aus der Faktorenanalyse ausgeklammert. Auf die Frage nach der Beziehung von PA und NA einerseits und VA andererseits ist daher später nochmals zurückzukommen. Dasselbe gilt für den auffälligen Unterschied der Faktorenkorrelation zwischen der Person- und der Zeitpunktebene.

4.6 Untersuchungen zur Reliabilität der Skalen auf der Zeitpunkt- und auf der Personenebene

4.6.1 Vorbemerkungen

Bei der Frage nach der Reliabilität – dem wichtigsten formalen Gütekriterium einer Skala – ist die Tatsache, dass bei ESM-Studien zwischen der Person- und der Zeitpunktebene unterschieden werden muss, besonders offensichtlich: Es ist ja anzunehmen, dass die Reliabilität auf der Personenebene, die sich erst in aggregierten Daten manifestiert, wesentlich grösser ist als auf der Zeitpunktebene, auf der es um Einzelmessungen geht. Dabei ist auch daran zu erinnern, dass die Reliabilitäten auf der Zeitpunktebene von Person zu Person variieren können. Etwas Analoges gilt bei einer streng mehrebenenanalytischen Betrachtungsweise auch für die Reliabilität auf der Personenebene: Insofern pro Person unterschiedlich viele Zeitpunkte zur Verfügung stehen und aggregiert werden können, kann die Reliabilität auch eines Personkennwerts (z.B. eines Mittelwerts) von Person zu Person variieren (Raudenbush & Bryk, 2002). Im Folgenden wird auf der Personenebene aber immer nur je ein Koeffizient angegeben, der sich auf eine Person mit einer durchschnittlichen Anzahl von Zeitpunktangaben bezieht. Wird auch auf der Zeitpunktebene nur eine Angabe gemacht, entspricht diese hingegen der durchschnittlichen intraindividuellen Reliabilität.

Die Reliabilität der PANAVA-KS wird nachfolgend in drei Hinsichten untersucht: erstens simultan auf beiden Ebenen (Abschnitt 4.6.2), zweitens im Hinblick auf interindividuelle Unterschiede auf der Zeitpunktebene (Abschnitt 4.6.3) und drittens auf der Personenebene bei Berücksichtigung unterschiedlicher Zeitabstände bzw. Lebenskontexte (Abschnitt 4.6.4).

4.6.2 Mehrebenenanalytische Konsistenzkoeffizienten

Raudenbush, Rowan und Kang (1991) haben ein umfassendes Konzept zur simultanen Analyse von Skalenreliabilitäten in Zweiebenen-Daten vorgelegt, die auf einer Zerlegung der Gesamtvarianz eines Skalenwerts in drei Komponenten beruht. Übertragen auf die vorliegende Problemstellung handelt es sich um die personbedingte Varianz, die zeitpunktbedingte (situationsbedingte) Varianz und die itembedingte Varianz. Durch die unterschiedliche Inbeziehungsetzung dieser Anteile können für jede Ebene die Anteile der (jeweils relevanten) "wahren" Varianz an der (jeweils relevanten) Gesamtvarianz bestimmt werden – was exakt der Definition von Reliabilität entspricht. (Eine genaue Beschreibung des verwendeten mehrebenenanalytischen Modells findet sich in Anhang A2.)

Die derart geschätzten Reliabilitätskoeffizienten sind in den zwei ersten Spalten von Tabelle 6 enthalten. Wie erwartet sind sie auf der Personenebene mit Werten über .9 sehr hoch, was die intuitive Erwartung bestätigt, dass eine ESM-Untersuchung auch ein ausgezeichnetes (allerdings

Tabelle 6

Konsistenzkoeffizienten der PANAVA-KS auf der Person- und der Zeitpunktebene

Skala	Mehrebenenanalytische Schätzung		Konsistenzanalysen pro Person (Zeitpunktebene)			
	Personebene	Zeitpunkt-ebene	<i>M</i>	Median	<i>Q</i> ₁	<i>Q</i> ₃
Positive Aktivierung (PA)	.92	.78	.77	.78	.72	.85
Negative Aktivierung (NA)	.94	.76	.76	.79	.70	.86
Valenz (VA)	.94	.71	.70	.74	.62	.82

Anmerkungen: $N = 23'509$ Zeitpunkte von 530 Personen mit $M = 44.4$ Zeitpunkten ($SD = 4.0$). M = Mittelwert; Q_1 = unteres Quartil; Q_3 = oberes Quartil

aufwändiges) Mittel für die Individualdiagnostik darstellen könnte. Niedrigere Koeffizienten ergeben sich auf der Zeitpunktebene: Mit Werten von .71 (VA, 2 Items) und .78 bzw. .76 (PA bzw. NA, je 4 Items) fallen sie aber noch befriedigend aus.³⁶

4.6.3 Interindividuelle Unterschiede in der Konsistenz auf der Zeitpunktebene (intraindividuelle Konsistenzen)

Die Zeitpunktebene kann man bei einer ESM-Untersuchung auch als aus einer Menge von Einzelfallstudien zusammengesetzt betrachten. Mit einer metaanalytischen Verarbeitung der Ergebnisse dieser Fallstudien erreicht man dann ebenfalls (wie bei der Mehrebenenanalyse) eine Elimination der personbedingten Varianz.³⁷ Das heisst im gegebenen Zusammenhang, dass man für jede Person ($N = 530$) eine konventionelle Konsistenzanalyse durchführen und ein Cronbach- α bestimmen kann. Der spezifische Informationsgehalt dieser Analyse besteht dann darin, dass sie auf einfache Weise zu untersuchen erlaubt, in welchem Ausmass sich die intraindividuellen Konsistenzen interindividuell unterscheiden und womit diese Unterschiede allenfalls zusammenhängen.

Tabelle 6 (rechte Spalten) enthält einige Verteilungskennwerte dieser individuellen Cronbach- α . (Die Verteilungskennwerte der dazu gehörigen Trennschärfen sind in der Tabelle B1 des Anhangs B aufgeführt.) Zunächst sei auf die annähernd perfekte Übereinstimmung der Mittelwerte der intraindividuellen Konsistenzen mit den entsprechenden mehrebenenanalytischen Schätzungen hingewiesen – ein angesichts der völlig unterschiedlichen Vorgehensweisen keineswegs trivialer Sachverhalt.³⁸ Das besondere Interesse dieser Analyse liegt jedoch bei den Verteilun-

³⁶ Die etwas niedrigere Reliabilität von VA liesse sich durch die Hinzunahme von Items leicht korrigieren. So lieferte im Rahmen von Studie 1 die probeweise Hinzunahme der zwei hinsichtlich der Pole ausbalanciert dargebotenen Items "schlecht-gut" und "wohl-unwohl" eine interne Konsistenz für VA, die jene der PA- und NA-Skala deutlich übertrifft.

³⁷ Es bleiben aber natürlich Unterschiede. Vor allem werden in der Metaanalyse in der Regel die konkreten Ergebnisse der einzelnen Fallstudien als "feste" Effekte (fixed effects) betrachtet, während die entsprechenden Daten bei der Mehrebenenanalyse nur als Zufallsstichprobe aus einer Population solcher Daten behandelt werden, d.h. als zufallsvariable (Random-) Effekte.

³⁸ Gemäss der Empfehlung von Hunter und Schmidt (1990) wurden die Konsistenzkoeffizienten ohne Anwendung der Fisherschen z -Transformation gemittelt, da diese zu einer Überschätzung des Mittelwerts führt, was sich auch im vorliegenden Fall tendenziell bestätigt (PA und NA: $\alpha = .79$; VA: $\alpha = .73$). Diese Transforma-

gen. Da diese – wegen des natürlichen Deckeneffekts bei Konsistenzkoeffizienten – asymmetrisch sind, sind in Tabelle 6 neben den Mittelwerten auch die Mediane und – als Streuungsmass – die oberen und unteren Quartile aufgeführt. Dass die Mediane leicht höher ausfallen als die Mittelwerte ist Ausdruck der linksschiefen Verteilungen. Die Quartile lehren vor allem zwei Dinge: Erstens zeigt sich, dass deutliche interindividuelle Unterschiede in der intraindividuellen Konsistenz bestehen. Zweitens macht aber das untere Quartil (das definitionsgemäss von 75 % der Personen übertroffen wird) deutlich, dass die Cronbach- α bei der überwiegenden Zahl der Personen in einem zumindest für Forschungszwecke noch akzeptablen Bereich liegen.

Eine für ESM-Studien wichtige Folgefrage, der hier in Form eines *Exkurses* mit einigen Hinweisen nachgegangen sei, ist, wie diese interindividuellen Unterschiede in der intraindividuellen Konsistenz zu erklären sind. Eine naheliegende Hypothese ist, dass es sich dabei um eine Konsequenz interindividueller Unterschiede in der (generellen) Compliance handelt: Gewisse Personen beantworten die ESF nur flüchtig und folglich inkonsistent, während andere ihre Aufgabe gewissenhaft wahrnehmen und daher auch eine grössere Konsistenz aufweisen. Wäre diese Hypothese richtig, müssten sich starke Korrelationen zwischen den intraindividuellen Konsistenzen verschiedener Skalen ergeben. Dies ist aber nicht der Fall: Die Rangkorrelation der intraindividuellen Konsistenzen von PA und NA beträgt beispielsweise lediglich $r_{\alpha_{PA}, \alpha_{NA}} = .1$ ($p < .05$), was – trotz der Signifikanz – dagegen spricht, dass die Unterschiede in der intraindividuellen Konsistenz zu einem wesentlichen Anteil auf Unterschiede in der generellen Compliance zurückzuführen sind. Vielmehr scheinen die intraindividuellen Konsistenzen etwas mit den Skaleninhalten zu tun zu haben: Wie in Abschnitt 4.7 gezeigt werden wird, gilt $r_{PA,NA} < r_{PA,VA} < r_{NA,VA}$. Genau derselben Abstufung folgen auch die Korrelationen der intraindividuellen Konsistenzen: $r_{\alpha_{PA}, \alpha_{NA}} = .1 < r_{\alpha_{PA}, \alpha_{VA}} = .27 < r_{\alpha_{NA}, \alpha_{VA}} = .45$. Das heisst: je stärker zwei Skalen korrelieren (psychologisch Verwandtes erfassen), umso stärker stimmt die Rangfolge der intraindividuellen Konsistenzen in den zwei Skalen überein. Dies muss wohl so interpretiert werden, dass die Konsistenz des Antwortverhaltens einer Person vom psychologischen Gehalt der zu bearbeitenden Skalen abhängig ist.

Natürlich suchten wir nach Korrelaten dieser (skalenspezifisch variierenden) individuellen Konsistenzen. Während sich zu klassischen Persönlichkeitsmerkmalen (z.B. Fünf-Faktoren-Modell) keine substantiellen Zusammenhänge ergaben, zeigten sich starke positive Zusammenhänge mit der intraindividuellen Streuung in der jeweils betrachteten Skala (Rangkorrelationen $> .5$, alle $p < .001$), wobei diese Streuungen ebenfalls – wenn auch etwas schwächer als die Konsistenzen – skalenspezifisch variieren. Diese Beziehung von Konsistenz und Streuung ist methodisch sehr gut nachvollziehbar (Streuungseinschränkungen wirken sich im Allgemeinen negativ auf Reliabilitätskoeffizienten aus); sie hat aber auch einen inhaltlichen Aspekt: Wie in Abschnitt 4.1 erwähnt, wird heute oft davon ausgegangen, dass nicht nur intraindividuelle Mittelwerte, sondern

tion sei zwar gut geeignet, die Verteilung von Korrelationen zu normalisieren; im Gegensatz zu den Angaben in vielen Lehrbüchern sei sie aber für Mittelwertbildungen weder gedacht noch wirklich geeignet (siehe dazu auch Field, 2001).

auch intraindividuelle Streuungen diagnostisch bedeutsam sind und eine eigenständige Klasse von Persönlichkeitsmerkmalen bilden. Nach dem eben skizzierten Befund scheinen die intraindividuellen Konsistenzen etwas Ähnliches zu erfassen wie diese neuartigen Persönlichkeitsmerkmale.

4.6.4 Retest-Reliabilitäten bzw. Stabilitätskoeffizienten auf der Personenebene

Eine andere Methode der Reliabilitätsschätzung als die bisher verwendeten, die Retest-Reliabilität, beruht auf der Idee der Testwiederholung. Die Korrelation von zwei Messungen desselben Sachverhalts widerspiegelt ebenfalls den Anteil wahrer Varianz an der Gesamtvarianz einer Messung. Auf der Zeitpunktebene ist dieser Ansatz nicht sinnvoll realisierbar; es liegt in der Natur von States, dass sie nicht zweimal erfasst werden können. Möglich wäre höchstens der Ansatz der Halbierungsreliabilität (Aufteilung der Skalen in je zwei Hälften). Dies würde aber gegenüber den im vorigen Abschnitt referierten Kennwerten der internen Konsistenz (Cronbach- α) keinen wesentlichen Erkenntnisgewinn liefern. Daher wird hier darauf verzichtet.

Anders ist es auf der Personenebene, wo die interessierenden (Person-) Kennwerte auf der Aggregation der individuellen Antworten aus einer ganzen ESM-Untersuchung beruhen. Eine solche ESM-Untersuchung ist selbstverständlich wiederholbar, die Idee der Retest-Reliabilität entsprechend anwendbar. Wie bei allen Retest-Untersuchungen muss allerdings berücksichtigt werden, dass die Reliabilität (im exakten Sinne des Begriffs) nur eine notwendige, aber nicht hinreichende Bedingung für grosse Retest-Koeffizienten ist; deren Grösse ist auch von der Stabilität der erfassten Personmerkmale abhängig. Bei ESM-Untersuchungen, die sich auf das Befinden in einzelnen Alltagsmomenten beziehen, ist dabei nicht nur an den Faktor Zeit, die sozusagen rein "organismische" Stabilität, zu denken, sondern auch an den Einfluss kontextueller Faktoren in einem weiten Sinne.³⁹ Beispielsweise gibt es Hinweise darauf, dass es einen Jahreszyklus im Befinden gibt (vgl. Abschnitt 4.9.4). Es ist gut denkbar, dass dabei interindividuelle Unterschiede existieren (z.B. Personen, die im Frühling, andere, die im Herbst "aufblühen"). Würden die erste und die zweite ESM-Untersuchung in zwei verschiedene saisonalen Kontexten (Jahreszeiten) durchgeführt, würden diese interindividuellen Unterschiede naturgemäss die Retest-Reliabilität negativ beeinflussen.

Bei den folgenden Analysen wird schliesslich noch ein anderer Gedanke benutzt: Bei einer konventionellen Skala ist eine Retest-Untersuchung nur durch eine echte Testwiederholung zu bestimmen. Bei ESM-Untersuchungen, die eine Menge von Zeitpunkten betreffen, ist es hingegen möglich, diese Zeitpunkte in zwei Hälften aufzusplintern und diese zwei Hälften als zwei verschiedene ESM-Untersuchungen zu betrachten.⁴⁰ Um mit einer "echten" Wiederholung der

³⁹ Dieses Problem entspricht jenem, das in der Persönlichkeitspsychologie meist unter dem Stichwort der trans-situativen Konsistenz behandelt wird. Der Begriff "Konsistenz" wird im gegebenen Zusammenhang aber vermieden, weil er in einem logisch engeren Sinne schon im Begriff "Konsistenzkoeffizient" verwendet wurde.

⁴⁰ Formal entspricht dies dem Vorgehen bei der konventionellen Testhalbierungsmethode. Da dort aber Items aufgeteilt werden, resultiert eine Schätzung der internen Konsistenz. Hier hingegen wird nach Zeitpunkten aufgeteilt, somit die Retest-Reliabilität abgeschätzt.

ESM-Untersuchung vergleichbare Koeffizienten zu erhalten, müssen die Korrelationen der Personkennwerte aus den beiden Untersuchungshälften noch mit Hilfe der bekannten Formel von Spearman-Brown aufgewertet werden.

Tabelle 7 enthält nun derartige Retest-Koeffizienten, und zwar nicht nur für die individuellen Mittelwerte, sondern auch für die individuellen Standardabweichungen. Die verschiedenen Untersuchungstypen sind dabei (von links nach rechts) nach der zeitlichen Distanz der beiden ESM-Untersuchungen angeordnet, die gleichzeitig auch immer grössere potentielle Unterschiede des Alltagskontextes implizieren, in dem die beiden Untersuchungen stattfanden: Diese zeitlichen/kontextuellen Unterschiede sind bei der Halbierung nach geraden/ungeraden Zeitpunkten minimal (die beiden Untersuchungen überlappen sich im Verlauf ein- und desselben Tages); bei jener nach geraden und ungeraden Tagen etwas grösser; noch grösser bei der Halbierung nach erster und zweiter Wochenhälfte (wovon die eine z.B. das Wochenende mit umfasst) und am grössten bei der einzigen vollständigen Retest-Untersuchung, bei der die eine ESM-Studie im Rahmen einer "normalen" Arbeitswoche, die andere – bei einem durchschnittlichen zeitlichen Abstand von rund 18 Tagen (siehe Anmerkung zur Tabelle 7) – im Rahmen einer Ferienwoche (meist verbunden mit einem Ortswechsel) durchgeführt wurde.

Tabelle 7

Retest-Koeffizienten für die individuellen Mittelwerte und Standardabweichungen der PANAVAKS auf der Personenebene

Skala	gerade/ungerade Zeitpunkte ^a		gerade/ungerade Tage ^a		1./2. Wochenhälfte ^a		Arbeitswoche/ Ferienwoche ^b	
	Rel(M)	Rel(SD)	Rel(M)	Rel(SD)	Rel(M)	Rel(SD)	Rel(M)	Rel(SD)
Positive Aktivierung (PA)	.95	.87	.89	.79	.82	.75	.70	.75
Negative Aktivierung (NA)	.97	.84	.92	.77	.87	.71	.73	.66
Valenz (VA)	.96	.80	.91	.69	.85	.68	.83	.57

Anmerkungen: $N = 23'509$ Zeitpunkte von 530 Personen mit $M = 44.4$ Zeitpunkten ($SD = 4.0$). Rel(M) bzw. Rel(SD) = Retest-Koeffizienten der individuellen Mittelwerte bzw. Standardabweichungen

^a Retest-Reliabilitäten auf der Grundlage der angegebenen Aufteilung der Zeitpunkte und mittels der Formel von Spearman-Brown auf eine ganze Untersuchungswoche hochgerechnet. $N = 530$.

^b Verglichen werden zwei einwöchige ESM-Untersuchungen, von denen die eine in einer Arbeitswoche, die andere in einer Ferienwoche stattfand. Zeitlicher Abstand der beiden Untersuchungswochen: $M = 17.7$ Tage; $SD = 16.4$ Tage. $N = 23$.

Auch wenn die Befunde aus der zuletzt genannten Untersuchung wegen des kleinen N mit Vorsicht zu interpretieren sind (es handelt sich um eine Teilstichprobe von $N = 23$ Personen), sind die Reliabilitätskoeffizienten klar im erwarteten Sinne parallel zur wachsenden zeitlichen/kontextuellen Distanz abgestuft.⁴¹ Bei den individuellen Skalenmittelwerten fallen die Retest-Koeffizienten in allen Fällen hoch aus ($.70 \leq r_{tt} \leq .97$) – selbst dann, wenn die beiden ESM-Untersuchungen in völlig verschiedenen Kontexten (Arbeitswoche vs. Ferienwoche) stattfinden. Das Analoge – wenn auch etwas abgeschwächt ($.57 \leq r_{tt} \leq .87$) – lässt sich für die individuellen

⁴¹ Dass die Koeffizienten im ersten Fall (Rel(M) auf der Basis gerade/ungerade Zeitpunkte) sogar die Konsistenzkoeffizienten auf der Personenebene (vgl. Tabelle 6), denen sie ungefähr entsprechen sollten, übertreffen, ist logisch an sich unmöglich, angesichts der kleinen Differenzen aber (Betrag $\leq .03$) nicht weiter von Bedeutung.

Standardabweichungen sagen. Diese Befunde sprechen dafür, dass sowohl mit den individuellen Standardabweichungen als auch – vor allem – mit den individuellen Mittelwerten tatsächlich zwar leicht kontextuell eingefärbte, aber doch relativ stabile Persönlichkeitsmerkmale erfasst werden.

4.7 Interkorrelationen der PANAVA-KS auf der Zeitpunkt- und der Personenebene

Die Frage nach den Beziehungen zwischen PA, NA und VA wurde schon im Rahmen der explorativen Faktorenanalysen in Abschnitt 4.5 angesprochen. Dort zeigte sich, dass die Faktoren, die den PA- und NA-Items entsprechen, auf beiden Ebenen negativ korreliert sind, auf der Zeitpunktebene allerdings nur schwach ($r_{F1,F2} = -.15$), auf der Personenebene hingegen stark ($r_{F1,F2} = -.55$). Die Beziehungen von PA und NA zu VA wurden nicht genau untersucht. Es wurde lediglich festgehalten, dass die VA-Items – konsistent mit den Befunden in Studie 1 – eine asymmetrische Zwischenposition einnehmen.

Tabelle 8
Interkorrelationen der PANAVA-KS

Skala	Mehrebenenanalytische Schätzungen				Interkorrelationsanalysen pro Person (Zeitpunktebene)	
	Personenebene		Zeitpunktebene		NA	VA
	NA	VA	NA	VA	Median (Q_1 ; Q_3)	Median (Q_1 ; Q_3)
Positive Aktivierung (PA)	-.57	.68	-.14	.37	-.15 (-.33; .04)	.38 (.21; .54)
Negative Aktivierung (NA)		-.80		-.71		-.71 (-.80; -.59)

Anmerkungen: $N = 23'509$ Zeitpunkte von 530 Personen mit $M = 44.4$ Zeitpunkten ($SD = 4.0$). $M =$ Mittelwert; $Q_1 =$ unteres Quartil; $Q_3 =$ oberes Quartil

Tabelle 8 enthält nun eine Interkorrelationsanalyse aller drei PANAVA-Skalen. Die mehrebenenanalytischen Schätzungen (linke Hälfte der Tabelle) wurden auf dieselbe Weise bestimmt wie die Iteminterkorrelationen (Abschnitt 4.5), die Verteilungskennwerte der intraindividuellen Korrelationen nach derselben Logik berechnet, die bereits bei den intraindividuellen Konsistenzen angewandt wurde (Abschnitt 4.6.3). Die Einsichten aus Tabelle 8 lassen sich in fünf Punkten zusammenfassen:

- 1.) Die Zeitpunktebene liefert systematisch deutlich niedrigere Interkorrelationen als die Personenebene. Der Grössenunterschied, der sich bei den Faktoreninterkorrelationen (Abschnitt 4.5) ergab, gilt also auch für die Skalen, und zwar inklusive VA. Darauf wird unten nochmals zurückgekommen.
- 2.) Die mehrebenenanalytisch bestimmten Korrelationen von PA und NA (Personenebene: $r_{PA,NA} = -.57$; Zeitpunktebene: $r_{PA,NA} = -.15$) stimmen bis auf wenige Hundertstel mit den entsprechenden Faktoreninterkorrelationen überein (Personenebene: $r_{F1,F2} = -.55$; Zeitpunktebene: $r_{F1,F2} = -.15$; vgl. Abschnitt 4.5). Dies führt zur Vermutung, dass die Skalen PA und NA weitgehend dasselbe

erfassen wie die entsprechenden Faktoren. Diese Vermutung kann allerdings beim mehrbenenanalytischen Vorgehen von Abschnitt 4.5 nicht überprüft werden, weil dieses Vorgehen keine Bestimmung von individuellen Faktorwerten erlaubt. Hingegen ist dies beim dort beschriebenen typischen Vorgehen in der Arbeitsgruppe von Watson auf der Zeitpunktebene (intraindividuelle z-Standardisierung) sowie bei einer Faktorenanalyse der Interkorrelationen der effektiven individuellen Skalenmittelwerte (Personebene) leicht möglich. Dieses Vorgehen führt zu praktisch denselben faktorenanalytischen Ergebnissen wie in Tabelle 5 und liefert Korrelationen zwischen den Skalen und den Faktorwerten von .99, was die genannte Vermutung bestätigt.

3.) Die Beziehungen von PA und NA zu VA fallen – wie gemäss Faktorenanalyse und Studie 1 zu erwarten – asymmetrisch aus: Auf beiden Ebenen korreliert VA mit NA wesentlich stärker (negativ) als mit PA (positiv) (Personebene: $r_{PA,VA} = .68$; $r_{NA,VA} = -.80$; Zeitpunktebene: $r_{PA,VA} = .37$; $r_{NA,VA} = -.71$). Besonders stark ist diese Asymmetrie auf der Zeitpunktebene: PA und VA teilen hier rund 14 % der Varianz, NA und VA hingegen 50 %, was einem Verhältnis von 1 : 3.7 entspricht. Auf der Personebene ergibt sich ein entsprechendes Verhältnis von nur ca. 1 : 1.4 (46 % : 64 %). Auch auf diesen Unterschied zwischen den beiden Ebenen ist unten zurückzukommen.

4.) Tabelle 8 (rechte Hälfte) enthält weiter die Befunde über die interindividuellen Unterschiede in den intraindividuellen Korrelationen. Die Mediane stimmen naturgemäss praktisch mit den mehrbenenanalytischen Schätzungen überein. Die Streuungsmasse (Quartile) zeigen aber, dass beträchtliche interindividuelle Unterschiede bestehen. Beispielsweise zeigen mehr als 25 % der Personen – im Unterschied zur negativen Korrelation bei den übrigen Personen – ein positives $r_{PA,NA}$ ($Q_3 = .04$). Möglicherweise verbergen sich hinter solchen Korrelationsunterschieden ebenfalls – wie hinter intraindividuellen Standardabweichungen oder Konsistenzen – neuartige Personcharakteristika.

5.) Und schliesslich drängt sich noch ein Vergleich der Interkorrelationen der PANAVA-KS mit den Befunden der konfirmatorischen Analysen in Studie 1 (Abschnitte 3.3.4 und 3.3.5) auf. Betrachtet man das dortige Modell 7– es wies die besten Fit-Kennwerte auf – ergaben sich folgende Interkorrelationen: $r_{PA-Faktor,NA-Faktor} = -.45$, $r_{PA-Faktor,VA-Faktor} = .65$ und $r_{NA-Faktor,VA-Faktor} = -.79$. Die strukturellen Verhältnisse zwischen den drei Koeffizienten ($r_{PA,NA} < r_{PA,VA} < r_{NA,VA}$) stimmen also in allen drei Untersuchungsvarianten überein (Faktorenstruktur in der Ein-Punkt-Untersuchung, Skaleninterkorrelationen auf der Person- und der Zeitpunktebene in ESM-Daten). Geht man von der (bei Studie 1 erwähnten) Tatsache aus, dass in einer Ein-Punkt-Untersuchung personbedingte und zeitpunkt- (situations-) bedingte Varianz bzw. Kovarianz vermischt sind, müsste man erwarten, dass die Interkorrelationen in dieser Untersuchung *zwischen* jenen auf der Zeitpunkt- und der Personebene liegen. Dies ist tatsächlich der Fall, wobei sie aber deutlich näher bei jenen auf der Personebene ($.01 \leq |Differenz| \leq .12$) als bei jenen auf der Zeitpunktebene ($.08 \leq |Differenz| \leq .31$) liegen. Bei dieser Asymmetrie ist in Rechnung zu stellen, dass es sich bei den Ergebnissen aus der Ein-Punkt-Untersuchung – entsprechend der Natur konfirmatori-

scher Analysen – um messfehlerbereinigte (minderungskorrigierte) Korrelationen handelt, während in Tabelle 8 empirische Koeffizienten angegeben sind.

Die Ergebnisse der Interkorrelationsanalyse bestätigen also das Bild, das sich aus den bisher berichteten Befunden ergab. Allerdings drängt sich nun definitiv die Frage auf, wie denn der Unterschied in der Grösse der Interkorrelationen auf der Person- und der Zeitpunkt-Ebene zu erklären ist. Dabei sind mindestens vier Möglichkeiten in Betracht zu ziehen, die einander aber nicht ausschliessen:

- 1.) Eine erste Möglichkeit ist, dass der Unterschied Ausdruck der Tatsache ist, dass – wie schon mehrmals erwähnt – auf der Person- und der Zeitpunkt-Ebene unterschiedliche Gesetzmässigkeiten gelten können.
- 2.) Da die Skalen auf der Zeitpunktebene eine deutlich geringere Reliabilität aufweisen als auf der Personenebene (Tabelle 6), besteht auch die Möglichkeit, dass der Grössenunterschied der Korrelationen zwischen den beiden Ebenen lediglich Folge dieses Reliabilitätsunterschieds darstellt. Schätzt man die doppelt minderungskorrigierten Interkorrelationen, bleibt aber der Ebenenunterschied bei zwei der drei Korrelationen klar erhalten, nämlich bei $r_{PA,NA}$ (Zeitpunktebene: -.18; Personenebene: -.61) und bei $r_{PA,VA}$ (Zeitpunktebene: .50; Personenebene: .73). Einzig bei $r_{NA,VA}$ ist dies nicht der Fall; hier kehrt sich das Verhältnis sogar um (Zeitpunktebene: -.97; Personenebene: -.85; vgl. dazu auch weiter unten). Trotz dieses Gegenbeispiels kann man aber aus diesen Berechnungen schliessen, dass der *systematische* Grössenunterschied der Interkorrelationen zwischen den beiden Ebenen nicht durch die Reliabilitätsunterschiede erklärt werden kann.
- 3.) Im einleitenden Abschnitt zu Studie 2 (Abschnitt 4.1) wurde erwähnt, dass der Zeitreihencharakter von ESM-Daten Einfluss auf die Grösse der Interkorrelationen nehmen kann. Dieser Zeitreihencharakter besteht aber natürlich nur auf der Zeitpunktebene. Es ist somit als dritte Möglichkeit denkbar, dass die in Tabelle 8 festgehaltenen Befunde über die Interkorrelationen der PANAVA-KS auf dieser Ebene mindestens zum Teil anders ausfallen würden, wenn die möglichen Zeiteffekte explizit kontrolliert würden.
- 4.) Als letzte Möglichkeit bleibt, dass (analog) auf der Personenebene Faktoren wirksam sind, welche die Interkorrelationen auf *dieser* Ebene beeinflussen. Ein nahe liegender Kandidat eines solchen Faktors – der Effekt der Aggregation – wird im folgenden Abschnitt 4.8 thematisiert.

Ein Nebenergebnis dieser Überlegungen über die möglichen Gründe des systematischen Grössenunterschieds der Korrelationen auf der Zeitpunkt- und der Personenebene verdient noch einen *Exkurs*. Bei der zweiten Erklärungsmöglichkeit (Reliabilitätsunterschied) ergab sich, dass die doppelte Minderungskorrektur auf der Zeitpunktebene für die Durchschnittsperson zu $r_{NA,VA} = .97$ führt. Dieses Ergebnis legt nahe, dass die Skalen NA und VA für alle Personen mit durchschnittlichem oder überdurchschnittlichem $r_{NA,VA}$ als redundant betrachtet werden können, die Skala VA also zu NA geschlagen oder weggelassen werden könnte (was einer der Varianten des

Modells D in Abschnitt 2.2.2 entspräche). Dieser Schluss wäre aber aus ganz verschiedenen Gründen wohl voreilig: Grundsätzlich ist zu bedenken, dass die Minderungskorrektur lediglich eine Schätzung darstellt, die besonders bei der gegebenen Problemlage (Existenz interindividueller Unterschiede in den Korrelations- und Konsistenzkoeffizienten) nicht unproblematisch ist. Aber auch wenn man ihr Ergebnis "wörtlich" nimmt, lehrt dasselbe Ergebnis ja, dass zumindest für einen Teil der Personen NA und VA nicht redundant sind. Und schliesslich sei noch ein empirisches Argument angefügt: Wären NA und VA tatsächlich redundant, müsste man erwarten, dass die Beziehung von VA zu Drittvariablen Null wird, wenn der Einfluss von NA kontrolliert wird. Dies ist aber zumindest bei der Drittvariablen PA nicht der Fall: Berechnet man mit Hilfe der Interkorrelationen in Tabelle 8 für die Zeitpunktebene (Durchschnittsperson) die Partialkorrelation von PA und VA unter Kontrolle von NA ergibt sich $r_{(PA,VA).NA} = .39$ ($p < .01$) (vs. $r_{PA,VA} = .37$ in Tabelle 8). Diese substantielle Partialkorrelation kann dabei nicht etwa der Unreliabilität der auspartialisierten Variablen (NA) zugeschrieben werden: Korrigiert man $r_{(PA,VA).NA}$ mit Hilfe der Bereiter-Korrektur (Stelzl, 1982) für die Unreliabilität von NA, vergrössert sich die Partialkorrelation sogar leicht (korrigiertes $r_{(PA,VA).NA} = .42$). Dies bedeutet, dass – im Gegensatz zum Befund aus der Minderungskorrektur – VA gegenüber NA auch auf der Zeitpunktebene und für die Durchschnittsperson einen eigenständigen Informationsgehalt aufzuweisen scheint.

Bilanzierend kann man festhalten, dass die Interkorrelationen der PANAVA-KS in ESM-Daten *strukturell* den Erwartungen entsprechen, die sich aus der Literatur (Abschnitt 2.2) und aus Studie 1 (Abschnitt 3.3.3) ergeben: PA und NA sind untereinander negativ korreliert und haben die erwartete asymmetrische Beziehung zu VA, wobei – wie bei allen, in der Literatur berichteten Modellen, die drei Faktoren umfassen – die Beziehung von NA und VA ausserordentlich stark ist (vgl. Abschnitt 2.2.2). Dabei zeigt sich allerdings ein erheblicher Grössenunterschied der Korrelationen auf den beiden Ebenen, für den es verschiedene Erklärungsmöglichkeiten gibt. Die zwei eben genannten Punkte (starke Korrelation von NA und VA, Grössenunterschied zwischen den zwei Ebenen) sollen im nächsten Abschnitt durch Zusatzuntersuchungen auf der Personenebene noch etwas vertieft werden.

4.8 Zusatzuntersuchungen zu den Interkorrelationen und zur differentiellen Validität der PANAVA-KS auf der Personenebene

In diesem Abschnitt werden noch zwei Zusatzuntersuchungen zu den Skaleninterkorrelationen auf der Personenebene präsentiert. Die erste knüpft am Unterschied dieser Interkorrelationen auf der Zeitpunkt- und auf der Personenebene an. Die Frage lautet, welche Rolle dem Aggregationsprozess bei der Entstehung der auf der Personenebene deutlich stärkeren Interkorrelationen zukommt (Abschnitt 4.8.1). Die zweite Untersuchung geht den Implikationen dieser starken Interkorrelationen nach. Die Frage ist hier, ob die drei PANAVA-KS auf der Personenebene überhaupt noch eine differentielle Validität haben – eine Frage, die sich vor allem im Falle von NA und VA ($r_{NA,VA} = -.80$) aufdrängt, aber auch schon bei PA und VA ($r_{PA,VA} = -.68$) virulent ist (Abschnitt 4.8.2).

4.8.1 Eine Untersuchung zum Aggregationseffekt

Die folgende Analyse ist durch die bereits früher (Abschnitte 2.1 und 2.3) erwähnte Untersuchung von Watson und Tellegen (2002) motiviert, die hier daher zunächst genauer geschildert sei. In dieser Untersuchung wird gezeigt, dass in ESM-Studien mit steigender Aggregation unipolarer Befindlichkeitsitems der individuelle Antwortstil (Akquieszenz) die Interkorrelationen immer stärker beeinflusst, diese also zunehmend Artefaktcharakter bekommen. Watson und Tellegen (2002) nennen dieses Phänomen "Aggregationsparadox": "When combining multiple responses collected from the same participant over time, aggregating beyond a certain point actually may be counterproductive; more is not necessarily better" (a.a.O., S. 596). Paradox ist dies, weil aus der Sicht der traditionellen Auffassung das Aggregieren von Einzelmessungen (nur) vorteilhafte Eigenschaften hat: Es führt zu einer zunehmenden Elimination von Messfehlern, d.h. zu einer Erhöhung der Reliabilität des aggregierten Masses. Aus dieser Sicht ist also "mehr prinzipiell besser". Dass dieser klassische Aggregationseffekt bei wiederholten Befindlichkeitsmessungen nicht gilt, erklären Watson und Tellegen dadurch, dass der individuelle Antwortstil (ein so genannter systematischer, nicht zufälliger Messfehler) eine grössere Stabilität aufweist als die States, auf welche die Befindlichkeitsmessung eigentlich abzielt. Dieser Sachverhalt hat zur Folge, dass mit zunehmender Aggregation der Befindlichkeitsmasse diese immer reliabler (auch) den individuellen Antwortstil erfassen, was sich dann natürlich auch in den Interkorrelationen der aggregierten Befindlichkeitsmasse widerspiegelt. Tatsächlich zeigt sich (a.a.O., Table 2), dass sich die Interkorrelationen mit zunehmender Aggregation (1., 1. + 2., 1. + 2. + 3. Zeitpunkt usw.) systematisch *in positiver Richtung* verändern: Positive Korrelationen, d.h. solche zwischen "same valence scales", werden immer *stärker* positiv *und* negative Korrelationen, d.h. solche zwischen "opposite valence scales", immer *schwächer* negativ.⁴² Das entscheidende Argument für die postulierte Rolle des Antwortstils in diesem Prozess liefert das Auspartialisieren eines (analog sequentiell aggregierten) Akquieszenzmasses: Dadurch verschwindet der generelle Trend in positiver Richtung: Die positiven Korrelationen nehmen zwar immer noch systematisch zu, jedoch weniger stark als ohne Kontrolle der Akquieszenz. Bei den negativen Korrelationen hingegen kehrt sich nach Auspartialisieren der Akquieszenz der Trend um: Negative Korrelationen werden (analog zu den Verhältnissen bei den positiven Korrelationen) mit zunehmender Aggregation immer stärker negativ. *Dieses* Veränderungsmuster wird aber bei Watson und Tellegen (2002) nicht mehr diskutiert, vermutlich weil es genau den Erwartungen entspricht, die man üblicherweise mit dem Aggregationsprozess verbindet: Je mehr Informationen aggregiert werden umso grösser ist die Reliabilität der Messung und um so geringer ist die Minderung (Attenuation) eines Korrelationskoeffizienten durch die Unreliabilität der Messungen.

⁴² Dieses Argument von Watson und Tellegen (2002) mag vielleicht auf den ersten Blick irritieren, ist aber natürlich richtig: Die Abschwächung einer negativen Korrelation (z.B. von -.4 auf -.2) bedeutet unter den geschilderten Umständen tatsächlich eine Veränderung in *positiver* Richtung (in Richtung auf +1.0) und kann dadurch entstehen, dass immer stärker *auch* eine gemeinsame Drittvariable erfasst wird.

Überträgt man nun diese Überlegungen und Befunde auf die PANAVA-KS, ergeben sich zwei Hypothesen:

Die *erste Hypothese* knüpft unmittelbar an der Argumentation von Watson und Tellegen (2002) über die Rolle des Antwortstils an und besagt, dass der bei unipolaren Items mit zunehmender Aggregation feststellbare Trend der Interkorrelationen in positiver Richtung bei den PANAVA-KS keine Parallele hat, da hier – durch die ausbalancierte Darbietung bipolarer Items – die Akquieszenz keine Rolle spielen kann. Somit sollte also bei zunehmender Aggregation bereits in den einfachen Korrelationen (und nicht erst nach Auspartialisieren eines Akquieszenzmasses) das Befundmuster auftreten, das Watson und Tellegen (2002) nicht mehr weiter diskutieren: Eine Zunahme der Interkorrelationen in ihrer jeweiligen Richtung.

Die *zweite Hypothese* betrifft dieses Befundmuster, das letztlich impliziert, dass die drei PANAVA-KS – trifft die erste Hypothese zu – mit zunehmender Aggregation immer mehr gemeinsame Varianz aufweisen. Überträgt man die Erklärung von Watson und Tellegen (2002) für die mit zunehmender Aggregation steigende Rolle des (stabilen) Antwortstils auf diesen Fall, stellt sich die Frage, ob es bei den PANAVA-KS einen anderen, im Verhältnis zu den States, stabilen Drittfaktor gibt, der für diesen Trend verantwortlich ist. Vergewenwärtigt man sich das bei Watson und Tellegen (2002) nach Auspartialisieren der Akquieszenz entstandene und bei den PANAVA-KS gemäss Hypothese 1 bereits in den einfachen Korrelationen erwartete Befundmuster, liegt die Vermutung nahe, dass dieser Drittfaktor in einer Eigenschaft der Person liegen könnte, die als "*Grundgestimmtheit*" zu bezeichnen wäre: Diese Disposition würde zwanglos erklären, warum mit zunehmender Aggregation VA und PA immer stärker positiv und PA und NA sowie NA und VA immer stärker negativ korrelieren. Alle drei Skalen würden immer reliabler (auch) diese generelle Grundgestimmtheit der Personen erfassen. Zur Überprüfung dieser Hypothese bietet sich eine Analogie zu dem von Watson und Tellegen (2002) benützten Vorgehen an: Die Partialkorrelationen zwischen je zwei PANAVA-KS unter Kontrolle der dritten Skala sollten den Trend zu immer stärkeren Korrelationen nicht mehr zeigen.

Tabelle 9 zeigt die Ergebnisse der Analysen zu den zwei Hypothesen. Zu beachten ist, dass es naturgemäss sehr viele Möglichkeiten (49 Fakultät) gibt, eine sequentielle Aggregation vorzunehmen. Entsprechende Experimente zeigten jedoch, dass sich damit zwar gewisse Einzelheiten in den Verläufen ändern, die allgemeinen Trends jedoch dieselben bleiben. Und allein auf diese Trends kommt es hier an.⁴³

⁴³ Zu Tabelle 9 sei noch angemerkt, dass die einfachen Korrelationen bei einem Zeitpunkt (erste Zeile) nicht etwa den "Korrelationen auf der Zeitpunktebene" aus Tabelle 8 entsprechen. Bei den "einfachen Korrelationen bei einem Zeitpunkt" (erste Zeile von Tabelle 9) handelt es sich im Prinzip um eine Ein-Punkt-Untersuchung, bei der person- und zeitpunktbedingte Varianz vermischt sind und deren Ergebnis auch von Zufälligkeiten des gewählten Zeitpunkts abhängig ist. Bei den "Korrelationen auf der Zeitpunktebene" geht es hingegen – wie früher hervorgehoben – gewissermassen um die durchschnittliche intraindividuelle Korrelation, also um etwas wesentlich Anderes.

Tabelle 9

Illustration des Effekts der Aggregation auf die Interkorrelationen der PANAVA-KS

Anzahl aggregierter Zeitpunkte	einfache Korrelationen			multiple Korrelationen R			Partialkorrelationen		
	$r_{PA,NA}$	$r_{PA,VA}$	$r_{NA,VA}$	PA	NA	VA	$r_{(PA,NA),VA}$	$r_{(PA,VA),NA}$	$r_{(NA,VA),PA}$
1	-.34	.53	-.70	.53	.70	.77	.05	.44	-.65
3	-.39	.59	-.75	.60	.75	.82	.10	.49	-.70
5	-.45	.64	-.77	.64	.77	.84	.09	.51	-.70
7 (1 Tag)	-.48	.66	-.78	.66	.78	.85	.07	.51	-.70
14 (2 Tage)	-.50	.64	-.80	.64	.80	.85	.03	.46	-.72
21 (3 Tage)	-.53	.66	-.81	.66	.81	.85	.03	.47	-.73
28 (4 Tage)	-.54	.67	-.82	.67	.82	.86	.01	.47	-.73
35 (5 Tage)	-.54	.66	-.81	.66	.81	.85	-.02	.45	-.72
42 (6 Tage)	-.54	.66	-.81	.66	.81	.85	-.03	.44	-.71
49 (7 Tage)	-.55	.66	-.81	.66	.81	.85	-.03	.44	-.71

Anmerkungen: Die Auswahl der Zeitpunkte erfolgte nach einem Zufallsprinzip: Zunächst wurden sequentiell in zufälliger Reihenfolge die Zeitpunkte des 3. Untersuchungstages aggregiert, dann – ebenfalls sequentiell – die Zeitpunkte des 5., 7., 1., 2., 6. und 4. Tages hinzugenommen. Für weitere Erläuterungen siehe Text.

Für die *Prüfung der ersten Hypothese* sind die einfachen Korrelationen in Tabelle 9 relevant. Wie erwartet zeigt sich bei der PANAVA-KS der Trend zu immer stärkeren Zusammenhängen bereits in diesen Rohkorrelationen (und nicht erst nach Auspartialisieren eines Antwortstilmasse). Die Zunahme ist relativ stark und systematisch, sodass die Hypothese als gut bestätigt angesehen werden kann. Interessant ist noch das Nebenergebnis, dass die Korrelationen bereits nach ca. 3 Tagen (21 Signalen) praktisch jenen entsprechen, die sich nach 7 Tagen ergeben. Der Aggregationseffekt scheint also nicht mit derselben Stärke ins Unendliche weiterzugehen, sondern in Richtung bestimmter Werte zu konvergieren, die bereits auf Grund von Daten über ca. drei Tage (21 Signale) recht gut erfasst werden.

Die steigenden Interkorrelationen implizieren natürlich, dass die drei Skalen immer mehr Varianz teilen. Eine logische Folge ist, dass der Skalenwert einer Skala durch jene der beiden andern Skalen immer besser vorhersagbar ist. Dies wird durch die multiplen Korrelationen in Tabelle 9 illustriert, die – parallel zu den einfachen Korrelationen – weit gehend systematisch ansteigen.

Die zur *Prüfung der zweiten Hypothese* berechneten Partialkorrelationen (Tabelle 9) legen nahe, dass auch diese Hypothese aufrecht erhalten werden kann: Seltsamerweise treten zwar bei allen drei Reihen von Partialkorrelationen – allerdings an unterschiedlichen Stellen des Aggregationsprozesses – schwache kurvilineare Trends auf. Ob sie durch zufällige Besonderheiten des angewendeten sequentiellen Aggregationsverfahrens bedingt sind oder etwas bedeuten, müsste noch überprüft werden. Sie widersprechen aber der Hypothese nicht, dass in den Partialkorrelation der systematische Trend zu immer stärkeren Interkorrelationen nicht mehr derart massiv auftritt wie bei den einfachen Korrelationen. Vielmehr bleiben die Partialkorrelationen über den ganzen Aggregationsprozess *ungefähr* in derselben Grössenordnung. Bemerkenswert ist dabei, dass diese Grössenordnungen recht gut mit jenen übereinstimmen, die sich (für die Durchschnittsperson) auf der Zeitpunktebene ergeben. Zieht man jene Partialkorrelation bei, die auf

der höchsten Aggregationsstufe (also bei den höchsten Reliabilitäten der Skalen) zu beobachten sind, ist die Partialkorrelation von PA und NA mit $r_{(PA,NA).VA} = -.03$ ebenfalls tendentiell negativ (Zeitpunktebene: $r_{PA,NA} = -.14$); die Partialkorrelation von PA und VA ebenfalls stark positiv ($r_{(PA,VA).NA} = .44$; Zeitpunktebene: $r_{PA,VA} = .37$) und die Partialkorrelation von NA und VA ebenfalls sehr stark negativ ($r_{(NA,VA).PA} = r_{NA,VA} = -.71$).⁴⁴ Dies muss wohl als starker Hinweis darauf interpretiert werden, dass die hier "Grundstimmung" genannte Personeigenschaft wesentlich zum grossen Unterschied zwischen den Interkorrelationen auf der Zeitpunkt- und Personenebene beiträgt.

Zusammenfassend kann also festgehalten werden, dass die beiden eingangs formulierten Hypothesen zutreffen: Die PANAVA-KS zeigen auch ohne Akquieszenzkontrolle bei zunehmender Aggregation immer stärkere Interkorrelationen, das heisst auch immer mehr gemeinsame Varianz (Hypothese 1). Diese Zunahme der gemeinsamen Varianz scheint mindestens teilweise dadurch bedingt zu sein, dass die drei Skalen immer reliabler eine allgemeine Personeigenschaft erfassen, die oben "Grundgestimmtheit" genannt wurde (Hypothese 2). Diese Grundgestimmtheit äussert sich auf der Ebene des Verhaltens darin, dass die Personen die Tendenz haben, die Items, welche die drei Skalen ausmachen, eher auf der Seite des positiv valenzierten Pols *oder aber* eher auf der Seite des negativ valenzierten Pols zu beantworten, was dazu führt, dass die Skaleninterkorrelationen mit zunehmender Aggregation immer stärker werden. Handelt es sich dabei wirklich um eine "echte" Persönlichkeitsdisposition⁴⁵, bedeutet dies, dass der Unterschied der Interkorrelationen auf den beiden Ebenen nicht nur ein methodisches Artefakt darstellt, sondern als theoretisch bedeutsam anzusehen ist (Erklärung 1 vs. Erklärungen 2 bis 4 im vorherigen Abschnitt).

4.8.2 Zur differenziellen Validität der PANAVA-KS auf der Personenebene

Der durch den Aggregationsprozess implizierte Anstieg der Interkorrelationen der PANAVA-KS wirft die in forschungspraktischer Hinsicht wichtige Frage auf, ob denn die drei PANAVA-Skalen auf der Personenebene überhaupt noch eine differenzielle Aussagekraft haben. Wenn diese Skalen mit zunehmender Aggregation immer reliabler (auch) ein gemeinsames Konstrukt – die Grundgestimmtheit einer Person – erfassen, kann man dann noch erwarten, dass sich die Bezie-

⁴⁴ Zu beachten ist aber, dass die Korrelationen auf der Zeitpunktebene keine Partialkorrelationen sind. Berechnet man diese Partialkorrelationen, ergeben sich keine grossen Unterschiede: $r_{(PA,NA).VA} = .19$; $r_{(PA,VA).NA} = .39$; $r_{(NA,VA).PA} = -.72$. Auffällig ist lediglich die Vorzeichenumkehr bei $r_{(PA,NA).VA}$. Sie hat aber eine Parallele in den ebenfalls positiven Partialkorrelationen $r_{(PA,NA).VA}$ der ersten Aggregationsstufen in Tabelle 9 und besagt, dass bei Kontrolle der Valenz und bezogen auf einen oder einige wenige Zeitpunkte die beiden Aktivierungsdimensionen tendenziell positiv zusammenhängen.

⁴⁵ Die letztlich unentscheidbare Frage, die hier angesprochen wird, ist, ob es bei dieser Grundgestimmtheit "bloss" um ein Response Set geht. Der Unterschied dieses Konzepts vom weiter oben benutzten Konzept Response Style besteht darin (siehe z.B. Rorer, 1965), dass es sich bei einem Response Style um einen inhaltsfreien Antwortstil handelt (z.B. Akquieszenz), während ein Response Set auf Inhalte bezogen ist (z.B. Social Desirability). Diese Kontamination mit (potentiell persönlichkeitspsychologisch relevanten) Inhalten erklärt, warum die Response Set-Forschung mit sehr viel grösseren Schwierigkeiten zu kämpfen hat als die Response Style-Forschung (vgl. z.B. die Diskussion um die Rolle der Social Desirability bzw. "Positivität" für die regelhaft auftretenden Interkorrelationen im Fünf-Faktoren-Modell).

hungen der drei Skalen zu Drittvariablen noch unterscheiden – oder anders ausgedrückt: können sie dann noch eine "differenzielle Validität" haben?⁴⁶

Diese Frage kann natürlich nur empirisch beantwortet werden. Im Folgenden werden beispielhaft drei Kriterien benutzt, bei denen sich nach dem heutigen Stand des Wissens leicht Hypothesen über die erwartete differenzielle Validität der drei PANAVA-KS formulieren lassen, nämlich *Depressivität*, *Arbeitsbelastung* und *Lebenszufriedenheit*: *Depressivität* ist primär ein Ausdruck fehlender Positiver Aktivierung (z.B. Watson et al., 1999) und sollte daher in erster Linie einen (negativen) Zusammenhang mit dem individuellen PA-Level aufweisen. Der individuelle NA-Level hingegen widerspiegelt vor allem das im Alltag erlebte Ausmass der Belastung einer Person. Mangels eines allgemeineren Kriteriums in den gegebenen Daten wird das Kriterium *Arbeitsbelastung* benutzt, das nur für einen Teil der Personen verfügbar ist ($N = 428$). Beim dritten Kriterium, *Lebenszufriedenheit*, liegt es nahe, zu vermuten, dass ein spezifischer Zusammenhang mit dem individualtypischen Level von VA besteht.

Die drei Konzepte wurden wie folgt operationalisiert: "Depressivität" wurde mit drei Items der Skala "Depressive Stimmung" des Berner Fragebogens zum Wohlbefinden (BFW; Grob, Lüthi, Kaiser et al., 1991) erfasst ("Ich habe keine Lust etwas zu tun"; "Ich finde mein Leben uninteressant"; "Nichts macht mir mehr richtig Freude"; Cronbach- $\alpha = .71$); "Arbeitsbelastung" mit vier Items des Kurzfragebogens zur Arbeitsanalyse (KFZA; Prümper, Hartmannsgruber & Frese, 1995) ("Bei meiner Arbeit gibt es Sachen, die zu kompliziert sind", "Bei meiner Arbeit werden zu hohe Anforderungen an meine Konzentrationsfähigkeit gestellt", "Ich stehe häufig unter Zeitdruck"; "Ich habe zuviel Arbeit"; Cronbach- $\alpha = .74$); "Lebenszufriedenheit" mit vier Items der Satsfaction with Life Scale (SWLS; Pavot, Diener, Colvin & Sandvik, 1991) ("Mein Leben entspricht in den meisten Hinsichten meinen Idealvorstellungen"; "Meine Lebensbedingungen sind optimal"; "Ich bin mit meinem Leben zufrieden"; "Auch wenn ich neu anfangen könnte, würde ich in meinem Leben kaum etwas ändern"; Cronbach- $\alpha = .80$).

In Tabelle 10 werden diese drei Kriterien mit den individuellen Durchschnittswerten der Personen in PA, NA und VA in Beziehung gesetzt, wobei dies wieder für verschiedene Aggregationsstufen – der Übersichtlichkeit halber sind es nur drei – geschieht.

Betrachtet man die einfachen Korrelationen in Tabelle 10, scheint sich die Vermutung zu bewahrheiten, dass die drei PANAVA-KS auf der Personenebene kaum (mehr) differentielle Validität aufweisen: Sämtliche Skalen korrelieren – wenn alle Zeitpunkte aggregiert werden – mit allen Kriterien signifikant. Interessanterweise gilt dies auch bereits dann, wenn nur 7 Zeitpunkte aggregiert werden – und meist sogar schon, wenn nur ein einzelner Zeitpunkt benutzt wird. Dies

⁴⁶ Watson und Tellegen (2002) verwenden für denselben Sachverhalt den Begriff "diskriminante Validität". Da dieser Begriff in der MTMM (Abschnitt 3.3.4) eine wohlbestimmte, etwas abweichende Bedeutung hat (Beziehung einer Skala zu solchen, die etwas anderes erfassen wollen, *im Kontrast zur* Beziehung zu Skalen, die dasselbe erfassen wollen [= konvergente Validität]), wird hier der Begriff "differenzielle Validität" vorgezogen. Gemeint ist damit, ob zwei Skalen je spezifische Beziehungen zu Drittvariablen haben.

Tabelle 10

Die Beziehungen der PANAVA-KS zu Depressivität, Arbeitsbelastung und Lebenszufriedenheit

Anzahl aggregierter Zeitpunkte	einfache Korrelationen mit dem Kriterium			multiple Regression des Kriteriums auf PA, NA und VA			
	PA	NA	VA	β_{PA}	β_{NA}	β_{VA}	R
Kriterium: Depressivität (N = 528)							
1	-.18*	.12*	-.17*	-.12*	-.01	-.11	.20*
7 (1 Tag)	-.23*	.11*	-.17*	-.19*	-.05	-.09	.24*
49 (7 Tage)	-.27*	.15*	-.23*	-.22*	-.11	-.17*	.28*
Kriterium: Arbeitsbelastung (N = 428)							
1	.02	.22*	-.11*	.11	.29*	.05	.24*
7 (1 Tag)	-.08*	.23*	-.14*	.00	.30*	.09	.24*
49 (7 Tage)	-.11*	.25*	-.16*	.00	.36*	.13	.26*
Kriterium: Lebenszufriedenheit (N = 529)							
1	.11*	-.25*	.26*	-.05	-.12	.19*	.27*
7 (1 Tag)	.24*	-.29*	.38*	.02	-.01	.35*	.37*
49 (7 Tage)	.23*	-.30*	.38*	-.03	.01	.41*	.38*

Anmerkungen: β = standardisierte Regressionskoeffizienten

* = $p < .05$ (es wird nicht nach verschiedenen Signifikanzniveaus unterschieden)

illustriert eindrücklich die Tatsache, dass in jedem ESM-Datenpunkt person- und zeitpunkt- (situations-) bedingte Varianz kontaminiert sind (und damit Auswertungen, die diese Mehrebenenstruktur von ESM-Daten nicht berücksichtigen, prinzipiell mehrdeutige Ergebnisse liefern). In praktisch allen Fällen steigt aber der Betrag der Korrelation mit steigendem Aggregationsniveau an, wie es gemäss der Abhängigkeit der Reliabilität vom Aggregationsniveau zu erwarten ist.

Differenzierter ist das Bild, wenn man die standardisierten Regressionskoeffizienten β aus den multiplen Regressionen betrachtet: Hier erweisen sich – mit einer Ausnahme – jeweils nur mehr die postulierten "Determinanten" als signifikant. Die Ausnahme findet sich bei Depressivität: Hier ist auf der höchsten Aggregationsstufe neben PA auch VA ein signifikanter Prädiktor – ein post hoc gut nachvollziehbarer Sachverhalt.

Generell fällt noch auf, dass der Regressionskoeffizient des jeweils als relevant postulierten Prädiktors über die drei, in Tabelle 10 unterschiedenen Aggregationsstufen sukzessive ansteigt. Dies ist wohl so zu erklären, dass durch die Aggregation der Zeitpunkte nicht nur (wie im vorherigen Abschnitt gezeigt) das Gemeinsame der drei Skalen – die so genannte "Grundgestimmtheit" – immer reliabler erfasst wird, sondern naturgemäss auch das Spezifische jeder einzelnen Skala. Für die stetige Zunahme von R mit steigender Aggregation dürften dann die Reliabilitätsgewinne in beiden Hinsichten zusammen verantwortlich sein. Eine entsprechende Varianzzerlegung zeigt nämlich, dass bei allen drei Kriterien, die von PA, NA und VA geteilte Varianz (oben als "Grundgestimmtheit" bezeichnet) neben dem spezifischen Beitrag der einzelnen Skalen ein

ganz wesentlicher Prädiktor darstellt: Bei der Lebenszufriedenheit macht der Erklärungswert der gemeinsamen Varianz (auf der höchsten Aggregationsstufe) 66 % des R^2 aus, bei der Depressivität 52 % und bei der Arbeitsbelastung 30 %.

In Bezug auf die Hauptfrage dieses Abschnitts kann zusammenfassend festgehalten werden, dass die differenzielle Validität der drei Skalen PA, NA und VA auf der Personenebene – trotz des durch den Aggregationsprozess entstandenen massiven Anteils gemeinsamer Varianz – erhalten bleibt. Um diese differenzielle Validität zu erkennen, ist es allerdings notwendig, die von den Skalen geteilte Varianz statistisch zu kontrollieren, wie es oben mittels der Regressionsanalyse geschah. Eine bloße Betrachtung einfacher Korrelationen (oder – was auf dasselbe hinausläuft – von Mittelwertsunterschieden) ist nicht in der Lage, zwischen PA-, NA- und VA-spezifischen Effekten zu unterscheiden.

4.9 Tages-, Wochen- und Jahresrhythmen in den PANAVA-KS

4.9.1 Einleitung

In der letzten, in diesem Papier präsentierten Untersuchung geht es um die Frage, ob es Hinweise darauf gibt, dass die mit den PANAVA-KS erfassten Sachverhalte zeitlichen Rhythmen bzw. Zyklen unterworfen sind. Diese Frage ist im gegebenen Zusammenhang aus zwei völlig unterschiedlichen Gründen von Interesse: Erstens handelt es sich bei solchen Zyklen um ein prominentes Beispiel eines Zeiteffekts, die – wie bereits mehrfach erwähnt – die Ergebnisse statistischer Analysen auf der Zeitpunktebene beeinflussen können. Zweitens handelt es sich dabei – folgt man der Argumentation von Watson et al. (1999) – um einen Validitätshinweis. Was damit gemeint ist, sei im Folgenden noch genauer erläutert.

Wie früher erläutert betrachten Watson et al. (1999) PA und NA als subjektive Widerspiegelung der Aktiviertheit von zwei grundlegenden Motivationssystemen, und zwar soll PA die Aktiviertheit des Annäherungssystems, NA die Aktiviertheit des Vermeidungssystems signalisieren. Diese beiden Systeme haben völlig verschiedene psychologische Funktionen: Das Vermeidungssystem (und damit NA) ist ein Warnsystem, das auf aktuell oder potentiell bedrohliche Situationen reagiert. Das Annäherungssystem (und damit PA) hingegen steuert das proaktive Verhalten des Organismus in seiner Umwelt. Daraus ergeben sich unterschiedliche *Hypothesen* über endogene Rhythmen der beiden Systeme. Am klarsten ist die Situation in Bezug auf *circadiane Rhythmen*: Das *Vermeidungssystem (NA)* sollte keinem derartigen Rhythmus unterworfen sein, weil bedrohliche Situationen jederzeit auftreten können. Natürlich kann es im Tagesverlauf beobachtbare Rhythmen geben; diese wären aber exogen bedingt, d.h. durch mit der Zeit gekoppelte, mit einer gewissen Wahrscheinlichkeit regelhaft auftretende Bedrohungssituationen. Ganz anders ist die Situation beim *Annäherungssystem (PA)*: Aus evolutionspsychologischer Sicht wäre es sinnvoll, wenn dieses System mit der Wachzeit gekoppelt wäre und – insofern der Mensch ein "Tagwesen" ist – einen entsprechenden endogenen circadianen Rhythmus aufwei-

sen würde. Die analoge Hypothese äussern Watson et al. (1999) auch in Bezug auf die Jahreszeiten (vgl. unten, Abschnitt 4.9.4).

Nachstehend soll versucht werden, Indizien für die Existenz oder Nichtexistenz solcher Rhythmen zusammenzutragen. Das methodische Vorgehen orientiert sich dabei an jenem, das Watson et al. (1999) bei der Analyse circadianer Rhythmen anwenden: Es besteht darin, ungeachtet der Herkunft der Messungen von bestimmten Personen, die Gesamtheit der Messungen in Beziehung zum Messzeitpunkt (z.B. Tageszeit) zu setzen. Um dabei die personbedingte Varianz zu eliminieren, die – wenn bestimmte Personen bevorzugt zu bestimmten Zeiten erfasst worden sind – stören könnte, werden die einzelnen Messungen zuerst intraindividuell z-standardisiert.

Angemerkt sei noch, dass die vorliegenden Daten eine repräsentativere Grundlage zur Bestimmung von Rhythmen bieten als jene von Watson et al. (1999), vor allem deswegen, weil sie aus einer deutlich grösseren und heterogeneren Stichprobe stammen, die nicht nur Studierende umfasst. Weiter werden im Folgenden nicht nur circadiane Rhythmen untersucht, sondern – zum Teil allerdings mit schwächerer Aussagekraft – auch Wochen- und Jahresrhythmen. Generell ist aber festzuhalten, dass die folgenden Analysen – wie jene von Watson et al. (1999) – ebenfalls nur heuristischen Charakter haben. Insbesondere wird (noch) kein Versuch gemacht, die Rhythmen mehrebenenanalytisch zu modellieren und auf Signifikanz zu testen.

4.9.2 Circadiane Rhythmen

Wie oben ausgeführt, bestehen für PA und NA in Bezug auf circadiane Rhythmen, d.h. Rhythmen, die ungefähr dem Tagesablauf folgen, klar unterschiedliche Hypothesen: Ein endogener circadianer Rhythmus sollte nur bei PA auftreten, bei NA hingegen nicht. Es ist zwar möglich, dass bei NA ebenfalls ein circadianer Rhythmus festzustellen ist, dieser müsste aber exogen erklärbar sein. Für VA gibt es keine analog prägnante Hypothese. Die starke empirische Korrelation mit NA (und die früher beschriebene Auffassung von VA als eine Art Bilanz) lässt aber eher ein Verlaufsprofil vom Typus NA erwarten.

Um die Logik der Untersuchung dieser Hypothesen und die daraus folgende Bedeutung der Ergebnisse durchsichtig zu machen, wird im Folgenden etwas umständlicher vorgegangen als es an sich notwendig wäre.

Hinweise auf einen circadianen Rhythmus lassen sich im Prinzip bereits an einem durchschnittlichen Tagesverlauf ablesen. Die Abbildungen 3a und 3b stellen diesen Verlauf in unterschiedlichen Bezugssystemen dar. In beiden Fällen (und in späteren Darstellungen analog) repräsentiert der pro Zeitbereich angegebene Skalenwert den Mittelwert aller Zeitpunkte, die in den betreffenden Zeitabschnitt fallen. (In den Abbildungen 3 sind dies zwischen ca. 1'700 und 3'500 Zeitpunkte. Diese grossen N sind auch dafür verantwortlich, dass keine Standardfehler angegeben sind: Sie liegen ungefähr im Bereich der Grösse der Symbole.) In der Abbildung 3a handelt es sich um Skalenrohwerte, die – um zwischen den Skalen vergleichbar zu sein – in der Antwort-

skala der Items (1 = Minimum, 7 = Maximum, 4 = theoretische Skalenmitte) ausgedrückt sind. Der auffälligste Befund ist hier, dass die Durchschnittsverläufe aller Skalen ganz im jeweils positiven Bereich der Skalen verlaufen. Was – wie viele Befindlichkeitsstudien lehren – im Allgemeinen gilt (vgl. Abschnitt 4.4), gilt auch für die einzelnen Tageszeitabschnitte: Das Durchschnittsfinden liegt im positiven Bereich.

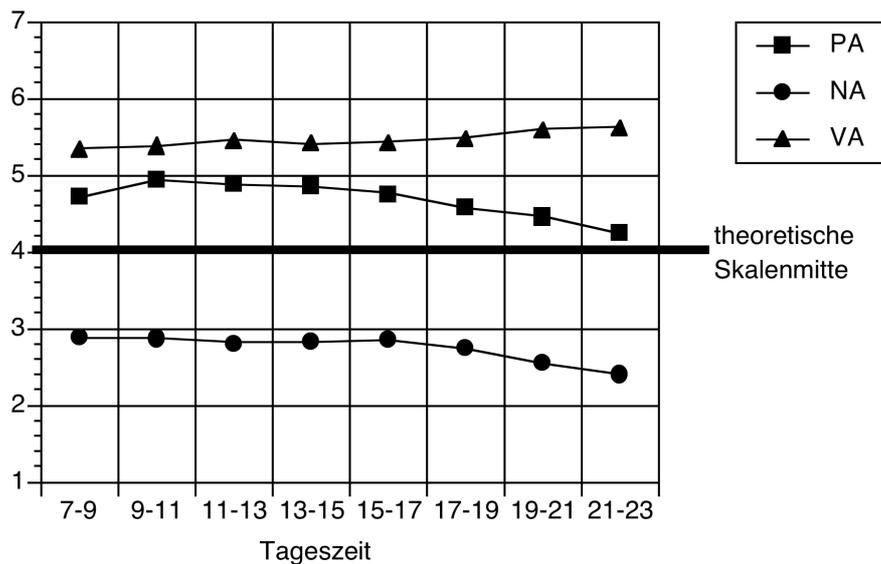


Abbildung 3a. Der durchschnittliche Tagesverlauf, ausgedrückt in Rohwerten

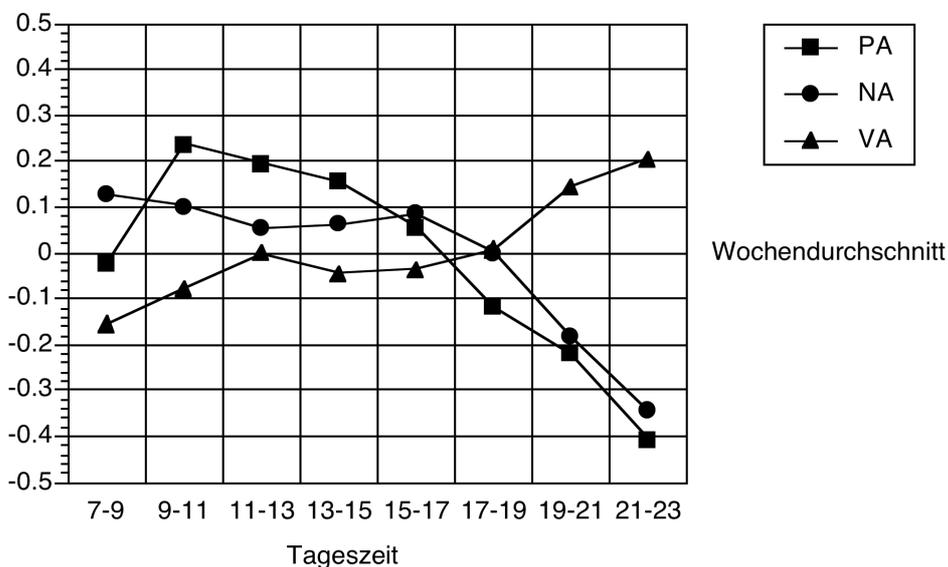


Abbildung 3b. Der durchschnittliche Tagesverlauf von PA, NA und VA, ausgedrückt in individuell z-standardisierten Skalenwerten

Obwohl sich bereits Trends andeuten, ist die Abbildung 3a zur Feststellung von Tageszyklen wenig geeignet. Der Grund liegt einerseits darin, dass die an sich beträchtlichen Streuungen der Skalenwerte auf der Zeitpunktebene (vgl. Abschnitt 4.4) im Bezugssystem der gesamten Antwortskala nicht zum Ausdruck kommen können. Andererseits sind in den Rohwerten person- und

zeitpunktbedingte Varianz konfundiert. Beide Probleme können durch eine intraindividuelle z-Standardisierung behoben werden. Die Ergebnisse, die sich in diesen z-Werten ergeben, sind in Abbildung 3b festgehalten. Der Wert 0 entspricht nun dem Wochendurchschnitt der Personen, der Wert 1 der Standardabweichung, welche die Personen in einer Woche zeigen. Diese Transformation bewirkt eine Spreizung, welche die Zyklen deutlicher erkennbar macht: Bei PA zeigt sich ein Anstieg am frühen Morgen, gefolgt von einer zuerst langsamen, dann immer schnelleren Abnahme gegen Abend. NA hingegen zeigt vom Morgen bis zum späteren Nachmittag eine gewisse Konstanz auf relativ hohem Niveau und beginnt dann am späten Nachmittag abzusinken. VA scheint – abgesehen von einem gewissen Einbruch am frühen Nachmittag – über den Tagesverlauf anzusteigen.

Eine wichtige Zusatzfrage ist nun, ob diese Rhythmen als endogen oder exogen bedingt zu verstehen sind. Dies kann natürlich rein statistisch nicht definitiv entschieden werden. Immerhin gibt es ein recht starkes, indirektes Argument: In unserem Kulturkreis geht man unter der Woche in der Regel einer Berufsarbeit nach, während das Wochenende (ebenfalls in der Regel) ganz anderen Tätigkeiten gewidmet ist. Damit sind natürlich auch massive Unterschiede in den typischen Tagesabläufen verbunden. Zeigt sich – trotz dieser psychologisch weit reichenden Unterschiede in den ökologischen Bedingungen – kein Unterschied im circadianen Zyklus zwischen diesen Tagestypen, kann dieser fast nur endogen gesteuert sein. Zeigen sich hingegen Unterschiede, muss der Zyklus exogen gesteuert sein: Es ist nicht vorstellbar, dass ein endogener Zyklus besteht, der am Wochenende sozusagen auf ein anderes Programm umgeschaltet wird.

Die Abbildung 4a bis 4c enthalten derartige Vergleiche der sieben Wochentage. Sie wirken naturgemäß etwas wirr; die zentrale Information ist aber gleichwohl gut erkennbar. Sie besteht im Kontrast zwischen dem Tageszyklus an den (Normal-) Arbeitstagen Montag bis Freitag (durchgezogene Linien) und an den Wochenendtagen Samstag und Sonntag (punktierte Linien). Die Darstellungen für PA (Abbildung 4a), NA (Abbildung 4b) und VA (Abbildung 4c) liefern ein sehr klares Bild: Wie postuliert kommt einzig bei PA die Annahme eines endogenen Zyklus in Frage. Die Zyklen von NA und VA sind zwar an den verschiedenen Arbeitstagen (durchgezogen) sehr verwandt und wohl auf die "Normalstruktur" eines Arbeitstages zurückzuführen. Die Wochenendtage (punktiert) zeigen aber deutlich andere Verläufe, die sich sogar zwischen Samstag und Sonntag wesentlich unterscheiden.

Im Gegensatz zu solchen Variationen (die sich auf Grund von alltagspsychologischem Wissen meist gut nachvollziehen lassen) ist der Zyklus von PA im Vergleich der verschiedenen Tagestypen erstaunlich stabil. Die bedeutsamsten Unterschiede zwischen Arbeitstagen und Wochenendtagen besteht in leichten zeitlichen Verschiebungen, die mit dem an Wochenendtagen häufig zu beobachtenden anderen Schlaf-Wach-Rhythmus zusammenhängen dürften. Aus der Literatur zu circadianen Rhythmen ist bekannt, dass diese weniger an objektive Zeiten gebunden sind als an den Schlaf-Wach-Rhythmus einer Person. Ob das soweit geht, dass sich bei Nacharbeit der

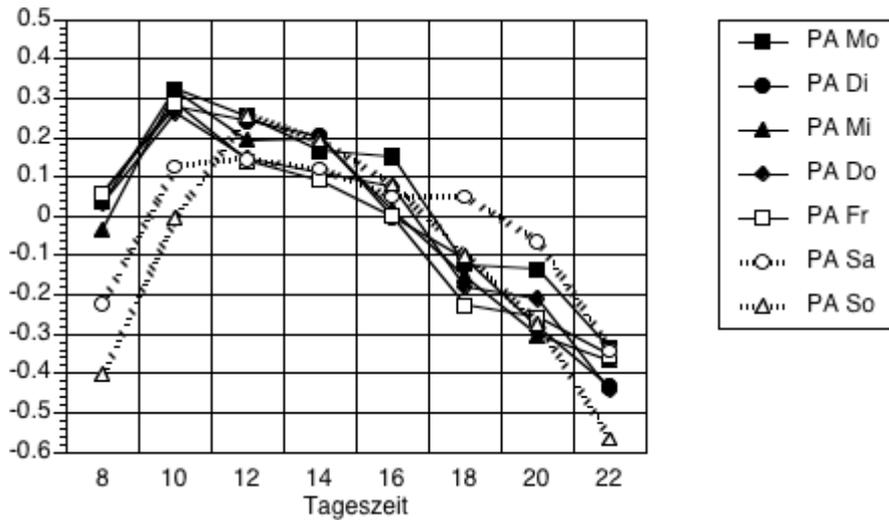


Abbildung 4a. Die PA-Tagesverläufe pro Wochentag (individuell z-standardisiert)

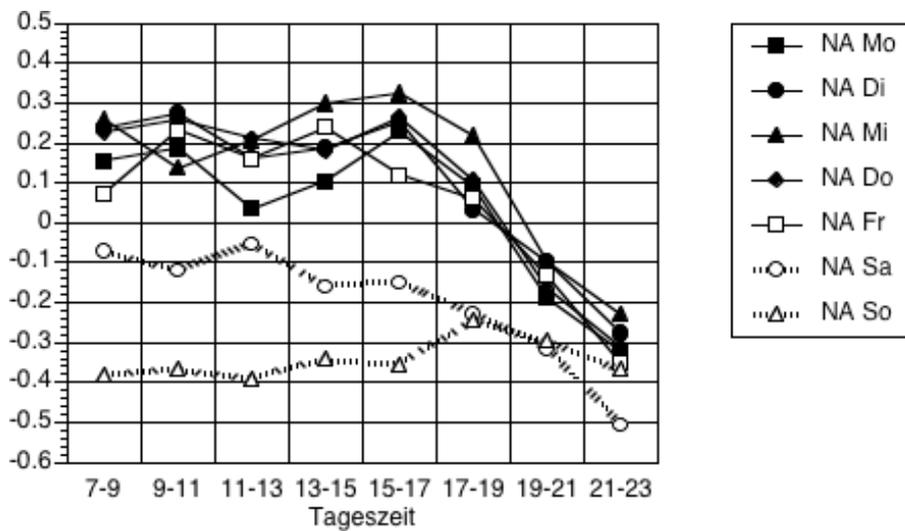


Abbildung 4b. Die NA-Tagesverläufe pro Wochentag (individuell z-standardisiert)

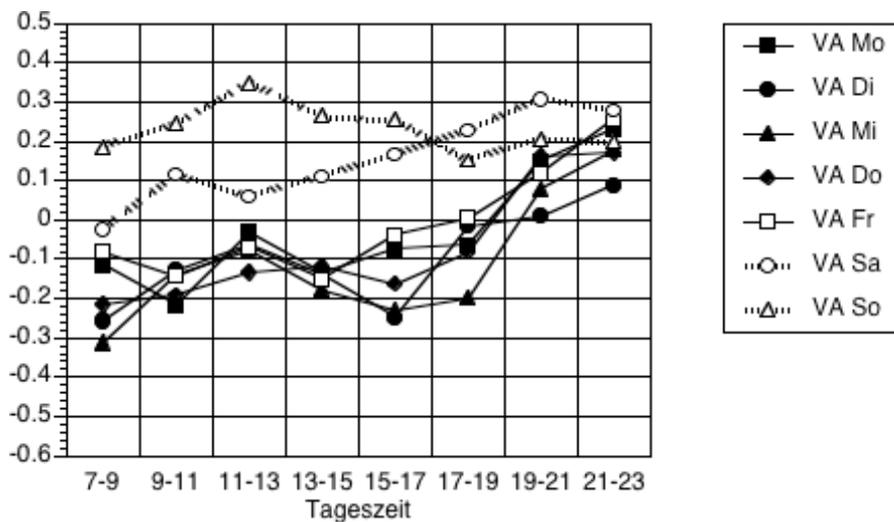


Abbildung 4c. Die VA-Tagesverläufe pro Wochentag (individuell z-standardisiert)

PA-Rhythmus vollständig verschiebt, könnte zwar in den zur Verfügung stehenden Daten abgeklärt werden (sie umfassen auch eine Gruppe von Schichtarbeitern), muss im gegebenen Rahmen aber offen bleiben.

Zusammenfassend ergeben diese Analysen, dass – wie postuliert – einzig PA einen höchstwahrscheinlich endogen bestimmten circadianen Rhythmus aufweist. Bei den beobachtbaren Tagesverläufen von NA und VA hingegen sprechen alle Indizien dafür, dass sie exogen erzeugt werden, d.h. durch regelhafte Abfolgen von Tätigkeiten und Umgebungsbedingungen im Alltagsleben.

4.9.3 Wochenrhythmen

Es wäre wohl riskant, anzunehmen, dass es endogen gesteuerte Rhythmen über den Verlauf einer Woche gibt: Bei der Woche handelt es sich um eine kulturelle Institution und nicht um ein biologisches Programm. Trotzdem ist es natürlich interessant, ob es einen (exogenen) Wochenzyklus in PA, NA und/oder VA gibt.

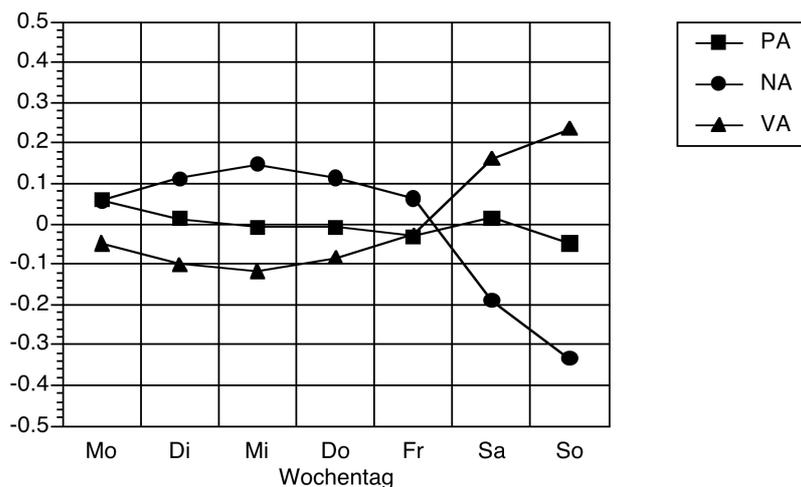


Abbildung 5. Der durchschnittliche Wochenverlauf von PA, NA und VA, ausgedrückt in individuell z-standardisierten Skalenwerten

Abbildung 5 ist analog konstruiert wie Abbildung 3b: Sie widerspiegelt die Mittelwerte sämtlicher (individuell z-standardisierter) Skalenwerte pro Wochentag. (Jeder Mittelwert basiert auf rund 3'200 Zeitpunkten.) Die geringsten Ausschläge zeigt PA – nicht ganz unerwartet, wenn man den endogenen Tageszyklus mitberücksichtigt. VA und NA erweisen sich aber wieder sozusagen "ökopsychologisch" besonders sensitiv: Die Wochenendtage zeigen ein vollständig anderes Bild (hohe VA, tiefe NA) als die Arbeitstage (tiefe VA, hohe NA). Aber auch die kleinen Ausschläge laden zu Interpretationen ein, z.B. das (relative) PA-Hoch am Montag (!) im Unterschied zu den (relativen) PA-Tiefs am Freitag und am Sonntag – oder die Kombination von NA-Hoch und VA-Tief am Mittwoch, etc. Sie seien jedoch dem Leser überlassen.

4.9.4 Jahresrhythmen

Über den Jahresrhythmus in PA und NA existieren wieder Hypothesen. Grund dafür ist, dass die Jahreszeiten (wie die Tageszeit) biologische Implikationen haben: So vermuten Watson et al. (1999, S. 832), dass der Sommer mit seinem reichen Angebot an Nahrungsmitteln aus endogenen Gründen mit einer grösseren Positiven Aktivierung einhergehen sollte als z.B. der Winter. NA hingegen ist (auch hier) auf eine derartige Vorprogrammierung nicht angewiesen, da sein Trigger in situativen Bedrohungen liegen. Aber es ist natürlich denkbar, dass die Häufigkeit solcher Bedrohungen zwischen den verschiedenen Jahreszeiten ebenfalls variiert.

Für die Überprüfung solcher Hypothesen bieten ESM-Daten weitaus schlechtere Voraussetzungen als in den beiden vorher diskutierten Fällen. ESM-Studien dauern in der Regel nicht jahrelang. Man kann aber unterschiedliche ESM-Studien, die in verschiedenen Jahreszeiten durchgeführt wurden, miteinander kombinieren. Allerdings ist dann die personbedingte und die allenfalls durch die Jahreszeit bedingte Varianz untrennbar konfundiert; der "Trick" mit der individuellen z-Standardisierung würde zugleich die Jahreszeitvarianz eliminieren. Es muss daher auf die Skalenrohwerter zurückgegriffen werden. Bei der Interpretation von Abbildung 6 ist dies zu berücksichtigen: Die Ausschläge sind mit jenen aus Abbildung 3a zu vergleichen – die anderen Abbildungen, die an der (individuellen) Standardabweichung orientiert sind, sind kein tauglicher Massstab. Ferner sei noch darauf hingewiesen, dass die (rund 3'000 bis 8'000) Zeitpunkte, die in der Datensammlung pro Jahreszeit zur Verfügung stehen, nicht gleichmässig über die Monate verteilt gesammelt wurden, d.h. nicht unbedingt perfekt repräsentativ für den jeweiligen Zeitraum sind.

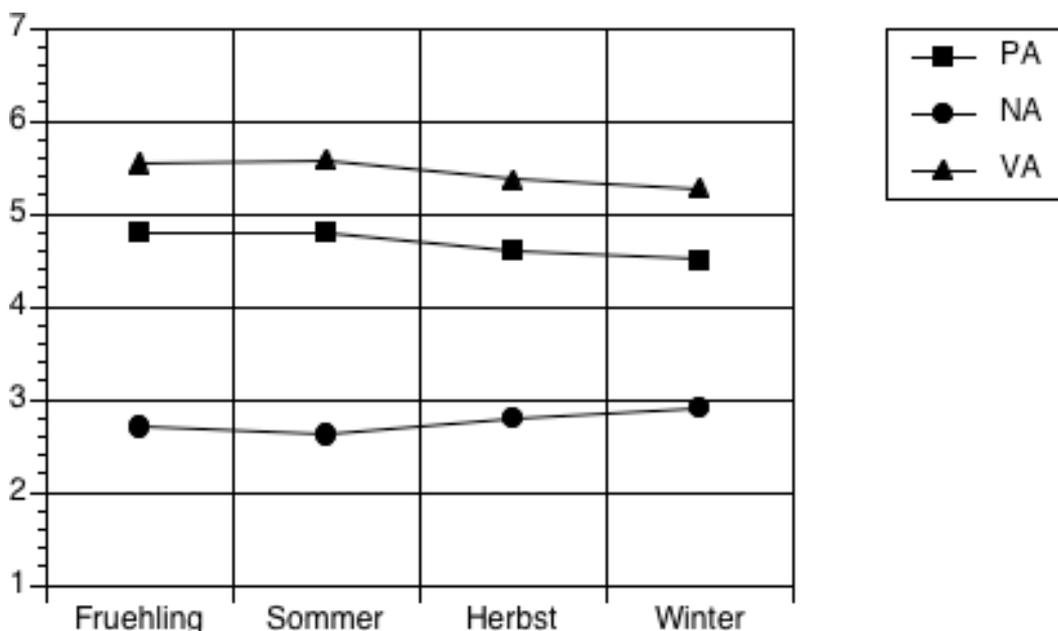


Abbildung 6. Der durchschnittliche Jahresverlauf von PA, NA und VA, ausgedrückt in Rohwerten

Frühling: März bis Mai; Sommer: Juni bis August; Herbst: September bis November; Winter: Dezember bis Februar

Betrachtet man Abbildung 6, so lassen sich – mit den genannten Vorbehalten – doch einige Schlüsse ziehen: Die Ausschläge sind zwar nicht ganz so gross wie im Tagesverlauf (Abbildung 3a), aber doch sichtbar: PA ist im Frühling und im Sommer am stärksten und sinkt über den Herbst zum Winter ab. Dieser Trend erinnert natürlich an das Phänomen der Winterdepression. VA verläuft praktisch parallel, NA gegenläufig. Die Verläufe sind mit den eingangs genannten Hypothesen verträglich. Für die Beantwortung der Frage, ob es sich dabei um endogen oder exogen bedingte Rhythmen handelt, können hier den Daten – im Unterschied zur Situation bei den circadianen Rhythmen – aber keine Hinweise entnommen werden.

4.9.5 Schlussbemerkungen

Die Frage nach zeitabhängigen Rhythmen, die in diesem Abschnitt 4.9 untersucht wurde, ist sowohl methodisch wie auch inhaltlich von Interesse. Es zeigte sich, dass in allen untersuchten Zeiträumen (Tag, Woche, Jahr) solche Rhythmen zu beobachten sind. Sie dürften in den meisten Fällen exogener Natur sein, was die Herausforderung impliziert, die relevanten exogenen Faktoren zu identifizieren. Hingegen ergaben sich starke Hinweise, dass PA einem *endogenen* circadianen Rhythmus unterworfen ist. Methodisch gesehen bedeutet dies, dass dieser Rhythmus bei Auswertungen, die sich auf PA beziehen, die Ergebnisse in unerwünschter Weise beeinflussen kann. Studiert man beispielsweise den Zusammenhang von Tätigkeiten und PA, ist damit zu rechnen, dass sich auch die Tageszeit, an der die Tätigkeit typischerweise ausgeübt wird, in den Ergebnissen manifestieren kann.

Die Ergebnisse sind natürlich auch in inhaltlicher Hinsicht von Bedeutung. Im gegebenen Zusammenhang sind sie vor allem als Hinweise auf die Validität der PANAVA-KS *auf der Zeitpunktebene* von Interesse. Auch wenn es hier nur um relativ grobe Fragen ging: Viele der berichteten Befunde kann man auch als Hinweise darauf verstehen, dass die PANAVA-KS sehr sensitiv alltägliche Fluktuationen des Befindens abzubilden in der Lage sind.

4.10 Zusammenfassung und Bilanz aus Studie 2

Ziel von Studie 2 war, zu untersuchen, wie sich die PANAVA-KS beim Einsatz in ESM-Studien bewähren. Wie in Abschnitt 4.1 erwähnt, gibt es für ein derartiges Unterfangen noch keine Standards. Vielmehr könne – so wurde gesagt – bereits die sinngemässe Übertragung elementarer konventioneller psychometrischer Konzepte auf den Fall von ESM-Daten rasch einmal zu einer Inflation psychometrischer Analysen führen. Die vorstehenden Seiten liefern wohl eine anschauliche Illustration dieser Behauptung. Ohne nochmals alle Einzelheiten zu referieren, lassen sich vier Haupteinsichten formulieren:

1.) Die ESM-Tauglichkeit der Items und Skalen, die bei der Entwicklung der PANAVA-KS angestrebt wurde, scheint weitgehend erreicht worden zu sein: Die PANAVA-KS werden von der grossen Mehrheit der UPn als valides Mittel empfunden, das Befinden im Alltag zu beschreiben, und geniessen – gemessen an der Compliance – eine gute Akzeptanz (Abschnitt 4.3).

Die Antwortverteilungen der Items sind – im Unterschied zu typischen Befindlichkeitsitems – befriedigend. Die Items erfassen nicht nur (zwar auch) Persönlichkeitsmerkmale, sondern sind sehr "Situations-sensitiv": Der grösste Teil ihrer Varianz entsteht auf der Zeitpunkt- und nicht auf der Person-Ebene. Dasselbe gilt für die Skalen (Abschnitt 4.4).

2.) Auch in psychometrischer Hinsicht sind die Skalen gut gelungen. Die Bildung der Skalen PA und NA lässt sich sowohl auf der Person- wie auf der Zeitpunkt-Ebene auch faktorenanalytisch rechtfertigen. VA nimmt – konsistent mit der referierten Literatur und Studie 1 – eine asymmetrische Zwischenposition ein, deren exakte Natur weiterer Abklärungen bedarf (Abschnitt 4.5; siehe auch unten). Die Reliabilitäten der Skalen auf der Personebene sind hervorragend, jene auf der Zeitpunktebene – bei deutlichen interindividuellen Differenzen – für Forschungszwecke befriedigend. Einige Indizien sprechen dafür, dass dies auf der Personebene nicht nur für Mittelwerte gilt, sondern auch für Streuungsmasse (Abschnitt 4.6), möglicherweise auch für intraindividuelle Konsistenzen und Korrelationen. Hier eröffnet sich ein interessanter Forschungsbereich (Abschnitt 4.6 und 4.7).

3.) Die bei konventionellen Skalen triviale Frage nach deren Interkorrelationen erwies sich in den ESM-Daten als äusserst vielschichtig, lieferte aber einige auch über die PANAVA-KS hinausreichende interessante Einsichten: Erstens zeigt sich jene Interkorrelationsstruktur, die auf Grund der Literatur und Studie 1 zu erwarten ist: eine schwache bis "moderate" (Watson et al., 1999) negative Beziehung von PA und NA und eine asymmetrische Beziehung dieser beiden Skalen zu VA. Zweitens fallen aber die einfachen Korrelationen auf der Personebene wesentlich stärker aus als auf der Zeitpunktebene – ein Befund, dem in dieser prägnanten Form in der Literatur einen gewissen Neuartigkeitscharakter zukommt (Abschnitt 4.7). Drittens konnte gezeigt werden, dass bei diesem Grössenunterschied der Aggregationsprozess eine wichtige Rolle spielt: Dieser Prozess erhöht einerseits die Skalenreliabilitäten und -validitäten, macht aber auch, dass die gemeinsame Varianz der drei Skalen immer grösser wird. Eine plausible Erklärung dafür ist, dass die Skalen durch zunehmende Aggregation von Zeitpunkten neben ihrem spezifischen Gehalt zunehmend auch ein übergeordnetes Personmerkmal, hier Grundgestimmtheit genannt, erfassen (Abschnitt 4.8.1). In diesem Prozess scheint aber die differentielle (oder – im Sinne von Watson und Tellegen, 2002 – "diskriminante") Validität der Skalen nicht verloren zu gehen, wenn diese personale Grundstimmung kontrolliert wird. Dieser Grundstimmung bzw. gemeinsamen Varianz kommt zudem – über die Einzelskalen hinaus – bei der Voraussage von Kriterien eine erhebliche inhaltliche Aussagekraft zu (Abschnitt 4.8.2). Auch in all diesen Fragen scheint aber noch einige Forschungsarbeit zu leisten zu sein, bevor die Situation völlig durchsichtig ist.

4.) Am Beispiel der Tages-, Wochen- und Jahreszyklen konnte schliesslich gezeigt werden, dass die PANAVA-KS Konstrukte erfassen, die mit dem Zeitablauf kovariieren. Während die meisten anderen festgestellten Rhythmen exogen bedingt sein dürften, bestehen starke Hinweise auf einen endogenen circadianen Rhythmus in PA, wie er von Watson et al. (1999) auf Grund evo-

lutionspsychologischer Überlegungen postuliert und – allerdings mit weniger aussagekräftigen Daten als hier – empirisch belegt wurde (Abschnitt 4.9).

5. Rückblick und Ausblick

Am Ausgangspunkt dieses Papiers (Abschnitte 1 und 2.1) standen Argumente dafür, dass die zunehmende Beliebtheit der Experience Sampling Method (ESM) und verwandter Datensammungsverfahren in der psychologischen Forschung nach einer neuen Generation von Skalen verlange, die im Hinblick auf die spezifischen Anforderungen dieser Verfahrensfamilie konstruiert und überprüft sind. Bei der Überprüfung sei insbesondere der hierarchischen Datenstruktur, die solche Verfahren implizieren, Rechnung zu tragen. Und schliesslich sollte – so wurde herausgestrichen – bei solchen neuen ESM-Skalen auch abgeklärt werden, in welchem Verhältnis sie zu konventionellen Skalen derselben Konstrukte stehen, damit die Ergebnisse verschiedener Studientypen aufeinander bezogen werden können.

Gegenstand dieses Papiers war die Entwicklung und empirische Überprüfung eines solchen neuen Instruments, der PANAVA-Kurzskalen (PANAVA-KS). Dieses Instrument dient der Erfassung der drei Konstrukte Positive Aktivierung (PA), Negative Aktivierung (NA) und Valenz (VA) in ESM-Studien. In Abschnitt 2 wurde der theoretische Hintergrund dieser Konstrukte und die bei der Skalenkonstruktion massgeblichen methodischen Gesichtspunkte dargestellt. In Abschnitt 3 ging es um einen empirischen Vergleich der drei neuen Kurzskalen mit fünf konventionellen Skalen, die dieselben Konstrukte betreffen. In Abschnitt 4 wurden schliesslich ganz verschiedenartige psychometrische Eigenschaften der PANAVA-KS in ESM-Daten untersucht. Die in den Abschnitten 3 und 4 dargestellten Untersuchungen sind – gemessen an gewohnten Darstellungen der Entwicklung und Überprüfung neuer Skalen – ausserordentlich umfangreich. Dafür sind vor allem zwei Gründe verantwortlich: Erstens stammen die Konstrukte PA und NA einerseits und das Konstrukt VA aus zwei verschiedenen theoretischen Traditionen, wobei in der Literatur sehr verschiedenartige Modellvorstellungen darüber existieren, wie das gegenseitigen Verhältnis dieser Konstrukte zu denken ist (vgl. Abschnitt 2.2.2). In manchen Analysen war daher der Frage nach den Beziehungen zwischen PA, NA und VA vertieft nachzugehen. Zweitens bringt es die hierarchische Natur von ESM-Daten mit sich, dass schon einfachste psychometrische Fragen (wie zum Beispiel jene nach den Verteilungskennwerten von Skalen, nach ihren Interkorrelationen oder ihren Reliabilitätskoeffizienten) jeweils nicht nur *eine* Antwort haben, sondern – parallel zur Mehrschichtigkeit der Daten – mehrere. Im Rahmen dieses Papiers wurden die zwei wichtigsten Ebenen unterschieden, nämlich die Zeitpunktebene (d.h. die Ebene der *intraindividuellen* Differenzen) und die Personenebene (d.h. die Ebene der *interindividuellen* Differenzen). Dabei war zu berücksichtigen, dass auch mit interindividuellen Unterschieden in den intraindividuellen Kennwerten zu rechnen ist. Gemeinsamkeiten und Unterschiede der Ergebnisse all dieser verschiedenartigen Analysen warfen mehrere neue Fragen auf, denen in Zusatzuntersuchungen nachzugehen war.

Die Einzelergebnisse der durchgeführten Analysen sind schon früher zusammengefasst worden (Abschnitte 3.5 und 4.10). Im Rahmen der folgenden Schlussbetrachtungen sollen sie nun noch

im Hinblick auf zwei übergeordnete Fragen betrachtet werden. Die erste Frage ergibt sich aus dem Hauptziel aller präsentierten Untersuchungen, das in der Evaluation der PANAVA-KS bestand. Sie lautet, welches *Gesamtbild* sich aus dieser Evaluation ergibt. Die zweite Frage knüpft daran an, dass manche der gefundenen Ergebnisse auch über die PANAVA-KS hinaus von Bedeutung sind und *weiterführende Fragen für die ESM- bzw. Befindlichkeitsforschung* aufwerfen. Diese zweite Frage lautet entsprechend, welche Punkte sich dabei besonders in den Vordergrund drängen.

Zur ersten Frage, jene nach dem *Gesamtbild aus der Evaluation der PANAVA-KS*: Die Ziele, die bei der Entwicklung der PANAVA-KS verfolgt wurden, scheinen weit gehend erreicht worden zu sein, d.h. die PANAVA-KS scheinen tatsächlich dafür geeignet, zur Befindlichkeitsmessung in ESM-Studien (im in der Einleitung definierten, weiten Sinne des Wortes) eingesetzt zu werden. Für diese Beurteilung sind vor allem die drei folgenden Sachverhalte relevant:

1.) Die PANAVA-KS erwiesen sich aus forschungspraktischer Sicht als ausgesprochen ESM-tauglich. Die früher abgeleiteten Kriterien (Abschnitt 2.1) sind alle erfüllt: Die PANAVA-KS sind mit total nur 10 (bipolaren) Items und einer sehr kurzen Bearbeitungszeit (ca. 30 bis 40 Sekunden) sehr *ökonomisch* und lassen damit genügend Spielraum, um im Experience-Sampling-Fragebogen noch andere interessierende Sachverhalte zu erfragen, ohne dass daraus eine unzumutbare Belastung der Untersuchungspersonen resultiert. Die Items scheinen zudem *Au-genscheinvalidität* zu besitzen, kaum als *redundant* erlebt zu werden und – im Gegensatz zu den üblicheren, unipolaren Befindlichkeitsitems – auch bei oft wiederholten Messungen im normalen Alltag eine deutliche *Antwortstreuung* zuzulassen. Für die Erfüllung dieser drei Bedingungen sprechen so verschiedene Hinweise wie das subjektive Urteil der Untersuchungspersonen, die geringe Anzahl fehlender Werte und die empirischen Antwortverteilungen (Abschnitte 4.3 und 4.4). Und schliesslich scheint es mit der hinsichtlich Polung ausbalancierten Darbietung bipolarer Items gelungen zu sein, die Rolle des *Antwortstils* zu kontrollieren; im Unterschied zu den Verhältnissen bei PANAS-analogen unipolaren Items zeigen die PANAVA-KS selbst bei starker Aggregation keinen Akquieszenzeffekt (Abschnitt 4.8.1).

2.) Die – die hierarchische Struktur von ESM-Daten berücksichtigenden – psychometrischen Analysen der PANAVA-KS ergaben in der Regel sehr befriedigende Ergebnisse: Die Skalen erweisen sich als sehr *situationssensitiv*: Der grösste Teil der Varianz widerspiegelt intraindividuelle Unterschiede (vgl. Abschnitt 4.4). Die *Faktorenstruktur* der Item entspricht sowohl auf der Zeitpunkt- wie der Personenebene den Erwartungen, die sich aus den Verhältnissen bei den konventionellen Skalen ableiten lassen (Abschnitt 4.5). Die *Konsistenzkoeffizienten* (Cronbach- α) erfüllen auf beiden Ebenen die üblichen Anforderungen. Dasselbe gilt für die nur auf der Personenebene bestimmbaren *Retest-Koeffizienten*, und zwar nicht nur für die Personunterschiede im Niveau, sondern auch für jene in der intraindividuellen Streuung der Skalenwerte (Abschnitt 4.6). Obwohl hier nur ein Nebenziel, ergaben sich auch eine Reihe von Hinweisen auf die *kriterielle Validität* der Skalen. Weil sich in einer Interkorrelationsanalyse (Abschnitt 4.7) – konform

mit der Literatur, die von drei Befindlichkeitsdimensionen ausgeht – zeigte, dass die Skalen zum Teil stark korreliert sind, sind dabei die positiven Befunde zur *differentiellen Validität* der Skalen besonders von Bedeutung. Entsprechende Analysen wurden primär auf der Personenebene vorgenommen (Abschnitt 4.8.2). Auf der Zeitpunktebene liefert aber die Analyse der Beziehungen zwischen den drei Skalen (Exkurs in Abschnitt 4.7) sowie die Untersuchung von zeitlichen Zyklen (Abschnitt 4.9) ebenfalls Hinweise dieser Art.

3.) Die drei angezielten Dimensionen werden durch die PANAVA-KS auf sehr *konstruktvalide* Weise erfasst, wie die konfirmatorischen Analysen zeigen, in denen die PANAVA-KS, die PANAS- und die AD-ACL-Skalen sowie die Skala MDBF-GS einbezogen wurden (Abschnitt 3). Setzt man also in einer ESM-Studie – statt, wie meist üblich, ad-hoc-Skalen – die PANAVA-KS ein, ist der direkte Anschluss an die ausserordentlich umfangreiche konventionelle Forschung, die sich auf die genannten Vergleichsskalen abstützt, gewährleistet.

Die zweite Frage, die hier noch aufgegriffen werden soll, richtet sich auf Ergebnisse, die über die PANAVA-KS hinaus von Interesse sind und *allgemeine Problemfelder der ESM- bzw. Befindlichkeitsforschung* betreffen. Vier Punkte seien hier hervorgehoben, und zwar in Form eines Ausblicks auf Forschungsfragen, die in Zukunft vermehrte Beachtung verdienen:

1.) Das erste Problemfeld bezieht sich auf die *Zeiteffekte* (Zyklen, Trends, Autokorrelationen), die hier – wie in vielen ESM-Untersuchungen (Abschnitt 4.1) – weitgehend ausgeklammert blieben. Der recht starke Hinweis auf die Existenz eines endogenen circadianen Rhythmus in PA (Abschnitt 9) belegt aber klar, dass mit solchen Zeiteffekten bei mit hoher Frequenz wiederholten Befindlichkeitsmessungen zu rechnen ist. Die explizite Analyse aller Typen von Zeiteffekten konstituiert offensichtlich ein Gebiet der Befindlichkeitsforschung, das wichtige Beiträge für das Verständnis der Befindensfluktuationen im Alltag liefern könnte. Neben diesem inhaltlichen Aspekt drängt sich aber auch die methodische Frage auf, in welchem Sinn und in welchem Ausmass solche Zeiteffekte die Ergebnisse einer ESM-Studie zu anderen Fragen beeinflussen.

2.) Ein zweiter Problemkreis für betrifft die *Rolle der Aggregation* für die aus ESM-Studien gewonnenen Kennwerte auf der Personenebene. Es wird in der Literatur oft hervorgehoben, dass diese Kennwerte konventionellen Massen (z.B. Antworten auf die Frage: Wie glücklich sind Sie im Allgemeinen?) überlegen sind, weil bei den erstgenannten die Aggregation objektiv vorgenommen wird, während die konventionellen Masse auf subjektiven Zusammenfassungen beruhen, die auf vielfältige Weise verzerrt sind (z.B. Kahnemann, 1999). Watson und Tellegen (2002) haben aber gezeigt, dass bei PANAS-analogen Items und Skalen die "objektive" Aggregation dazu führt, dass sie in starkem Ausmass durch die Akquieszenz beeinflusst sind. Tatsächlich zeigte sich hier (Abschnitt 4.8.1), dass der Akquieszenzeffekt bei den (anders konstruierten) PANAVA-KS nicht auftritt. Dies bedeutet – verallgemeinert –, dass Befunde auf der Personenebene bei Items und Skalen, die nur einen Pol der interessierenden Dimension abfragen, prinzipiell problematisch sind. Es wäre nützlich, die vorhandene ESM-Literatur zu bestimmten

inhaltlichen Fragen jeweils auch unter diesem Aspekt zu evaluieren. Beispielsweise kommen Schmuckle et al. (2002) zum Schluss, dass die latenten Variablen, die den beiden PANAS-Skalen zu Grunde liegen, auf der Zeitpunktebene negativ, auf der Personenebene jedoch unkorreliert sind. In der vorliegenden Studie zeigte sich bei den PANAVA-KS auf der (durch Aggregationsprozesse unbeeinflussten) Zeitpunktebene ein analoges Ergebnis, auf der Personenebene jedoch – in Übereinstimmung mit einem Grossteil der einschlägigen Literatur – ein deutlicher negativer Zusammenhang. Der auf den ersten Blick irritierende Unterschied geht genau in die Richtung, die man auf Grund von Watson und Tellegen (2002) zwischen einem Instrument mit unkontrollierter Akquieszenz (PANAS) und einem solchen mit kontrollierter Akquieszenz (PANAVA-KS) erwartet. Das Beispiel vermag wohl zu illustrieren, dass es sich lohnen würde, Aggregationseffekten in der ESM-Forschung künftig vermehrt Aufmerksamkeit zu schenken.

3.) Der dritte Problemkreis bezieht sich auf die Frage nach der *Dimensionalität einer sparsamen Beschreibung der Befindlichkeit*. In der heutigen Forschung dominieren klar Modelle, die zwei Affekt- bzw. Aktivierungsdimensionen umfassen (vor allem PANAS, AD-ACL). Die Annahme ist dabei, dass diese beiden Dimensionen die klassische Valenzdimension ersetzen/aufklären. In Übereinstimmung mit anderen Untersuchungen, die sich mit diesem Problem beschäftigen (z.B. Schimmack & Grob, 2000; Steyer et al., 1997); Tellegen et al., 1999a) sprechen aber verschiedene Befunde der hier vorgelegten Untersuchungen (Abschnitte 3, 4.7ff.) dafür, dass Valenz eine dritte Dimension konstituiert, die mit den beiden anderen zwar zum Teil stark korreliert ist, aber doch einen eigenständigen Informationsgehalt besitzt. Nicht zuletzt wegen der beiden oben angesprochenen Probleme (Zeiteffekte, Aggregationseffekte) stellt aber die Frage nach den empirischen Beziehungen zwischen den drei Dimensionen ein weiterhin offenes Forschungsfeld dar – insbesondere wenn die verschiedenen theoretischen Modellvorstellungen über die Beziehung von PA, NA und VA einbezogen werden.

4.) Der letzte Fragenkomplex, der hier genannt sei, betrifft die *Analyse von interindividuellen Unterschieden* mit Hilfe von ESM-Daten. Das Hauptinteresse lag dabei in der bisherigen Forschung auf Mittelwerten von Zeitpunktdaten. Sie bilden – auch in der Mehrebenenanalyse – sozusagen die klassischen Kennwerte auf der Personenebene, die den dispositionellen Anteil der Zeitpunktdaten repräsentieren. Es mehren sich aber die Indizien, die dafür sprechen, dass auch die intraindividuelle Streuung ein interessanter Personkennwert ist. So zeigte sich in Abschnitt 4.6.4, dass diesen Streuungen eine beträchtliche Retest-Reliabilität zukommt – eine notwendige Bedingung für die persönlichkeitspsychologische Relevanz eines Kennwerts. Es ergaben sich aber auch Hinweise auf andere, potentiell persönlichkeitspsychologisch bedeutsame interindividuelle Unterschiede in intraindividuellen Kennwerten. Hauptkandidaten sind hier die (skalenspezifische) intraindividuelle Skalenkonsistenz, die eine enge Beziehung zur Skalenstreuung aufweist (vgl. Abschnitt 4.6.3) sowie die intraindividuellen Skaleninterkorrelationen (Abschnitt 4.7). An der heutigen Persönlichkeitsforschung wird oft die dominante Stellung von Self-Reports kritisiert. Die genannten ESM-Kennwerte basieren zwar auch auf Self-Reports. Sie ver-

werten aber nicht – wie üblich – nur deren Inhalte, sondern formale Charakteristika des Antwortverhaltens. Sie können also im Sinne Cattells als neue Kategorie "objektiver Tests" (T-Daten) verstanden werden, deren persönlichkeitspsychologische Ergiebigkeit aber noch der Abklärung bedarf.

Insgesamt ergeben sich also aus den hier präsentierten Untersuchungen Anregungen für eine Reihe weiterführender methodischer und inhaltlicher Forschungsfragen. Das Hauptergebnis ist aber natürlich das positive Gesamtbild, das sich bei der Evaluation der PANAVA-KS ergab.

Literaturverzeichnis

- Abele-Brehm, A. & Brehm, W. (1986). Zur Konzeptualisierung und Messung von Befindlichkeit. *Diagnostica*, 32, 209-228.
- Aellig, St. (2004). *Über den Sinn des Unsinn. Flow-Erleben und Wohlbefinden als Anreiz für autotelische Tätigkeiten*. Münster: Waxmann.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & F. A. Schumaker (Eds.), *Advanced structural equation modeling. Issues and techniques* (pp 243-277). Mahwah NJ: Lawrence Erlbaum Associates.
- Arbuckle, J. L. (2003). *Amos 5.0 update to the Amos User's Guide*. Chicago: Small Waters Corporation.
- Arbuckle, J. L. & Wothke, W. (1999). *Amos 4.0 User's Guide*. Chicago: Small Waters Corporation.
- Bolger, N., Davis, A. & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579-616.
- Brandstätter, H. (1994). Alltagsereignisse und Wohlbefinden. In A. Abele & P. Becker (Hrsg.), *Wohlbefinden. Theorie, Empirie, Diagnostik* (2. Auflage) (S. 191-225). Weinheim: Juventa.
- Brandstätter, H. & Elias, A. (eds.). (2001). *Persons, situations, and emotions. An ecological approach*. New York: Oxford University Press.
- Brockner, J & Higgins, E. T. (2001). Regulatory focus theory: Implications for the study of emotions at work. *Organizational Behavior and Human Decision Processes*, 86, 35-66.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS. Basic concepts, applications, and programming*. London: Lawrence Erlbaum.
- Cacioppo, J. T., Gardner, W. L. & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, 76, 839-855.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by The Multitrait-Multimethod Matrix. *Psychological Bulletin*, 56, 81-105.
- Carver, Ch. S. (2001). Affect and the functional bases of behavior: On the dimensional structure of affective experiences. *Personality and Social Psychology Bulletin*, 5, 345-356.
- Carver, Ch. S. (2003). Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition and Emotion*, 17, 241-261.
- Carver, Ch. S., Sutton, S. K., & Scheier, M. F. (2000). Action, Emotion, and Personality: Emerging conceptual integration. *Personality and Social Psychology Bulletin*, 26, 741-751.
- Collins, R. L., Kashdan, T. B. & Gollnisch, G. (2003). The feasibility of using cellular phones to collect ecological momentary assessment data: Application to alcohol consumption. *Experimental and Clinical Psychopharmacology*, 11, 73-78.
- Csikszentmihalyi, M., Larson, R. & Prescott, S. (1977). The ecology of adolescence activity and experience. *Journal of Youth and Adolescence*, 6, 281-294.
- Curran, P. J., West, St. G. & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29.
- Delespaul, P. (1995). *Assessing schizophrenia in daily life. The experience sampling method*. Maastricht: Universitaire Pers Maastricht.
- Egloff, B., Schmuckle, St. C., Burns, L. R., Kohlmann, C.-W. & Hock, M. (2003). Facets of dynamic positive affect: Differentiating joy, interest, and activation in the Positive and Negative Affect Schedule (PANAS). *Journal of Personality and Social Psychology*, 85, 528-540.
- Eid, M. & Diener, E. (1999). Intraindividual variability in affect: Reliability, Validity, and personal correlates. *Journal of Personality and Social Psychology*, 76, 662-676.
- Engeser, St. (2005). *Lernmotivation und volitionale Handlungssteuerung: eine Längsschnittuntersuchung beim Statistik Lernen im Psychologiestudium*. Dissertation, Institut für Psychologie, Universität Potsdam. (Verfügbar unter <http://opus.kobv.de/ubp/volltexte/2005/240/>)
- Fahrenberg, J. & Myrtek, M. (Eds.) (1996). *Ambulatory assessment. Computer-assisted psychological and psychophysiological methods in monitoring and field studies*. Seattle: Hogrefe.
- Fahrenberg, J. & Myrtek, M. (eds.). (2001). *Progress in ambulatory assessment. Computer-assisted psychological and psychophysiological methods in monitoring and field studies*. Göttingen: Hogrefe.
- Feldman Barrett, L. & Barrett, D. J. (2001). An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, 19, 175-185.
- Feldman Barrett, L. & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, 74, 967-984.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161-180.
- Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. (2004). *Applied longitudinal analysis*. Hoboken NJ: Wiley.

- Gable, S. L., Reis, H. T. & Elliot, A. J. (2003). Evidence for bivariate systems: An empirical test of appetition and aversion across domains. *Journal of Research in Personality*, 37, 349-372.
- Gershuny, J. (2004). Costs and benefits of time sampling methodologies. *Social Indicators Research*, 67, 247-252.
- Ghisletta, P., Nesselroad, J. R., Featherman, D. L. & Rowe, J. R. (2002). Structure and predictive power of intraindividual variability in health and activity measures. *Swiss Journal of Psychology*, 61, 73-83.
- Gray, J. A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. Oxford: University Press.
- Grob, A., Lüthi, R., Kaiser, F. G., Flammer, A., Mackinnon, A. & Wearing, A. J. (1991). Berner Fragebogen zum Wohlbefinden Jugendlicher (BFW). *Diagnostica*, 37, 66-75.
- Hamm, A. O. & Vaitl, D. (1993). Emotionsinduktion durch visuelle Reize: Validierung einer Stimulationsmethode auf drei Reaktionsebenen. *Psychologische Rundschau*, 44, 143-161.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. London: Lawrence Erlbaum.
- Hu, L.-T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.
- Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park: Sage.
- Imhof, M. (1998). Erprobung der deutschen Version der Adjectiv-Checkliste nach Thayer (1989) zur Erfassung der aktuellen Aktiviertheit. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 19, 179-186.
- Kahnemann, D. (1999). Objective Happiness. In D. Kahnemann, E. Diener & N. Schwarz (Eds.), *Well-Being: The foundations of hedonic psychology* (pp. 3-25). New York: Russell Sage Foundation.
- Krohne, H. W. & Egloff, B. (2003). PANAS. Positive and Negative Affect Schedule (deutsche Form). In J. Schumacher, A. Klaiberg & E. Brähler (Hrsg.), *Diagnostische Verfahren zu Lebensqualität und Wohlbefinden* (S. 252-255). Göttingen: Hogrefe.
- Krohne, H. W., Egloff, B., Kohlmann, C.-W. & Tausch, A. (1996). Untersuchungen mit einer deutschen Form der Positive and Negative Affect Schedule (PANAS). *Diagnostica*, 42, 139-156.
- Larsen, R. J. & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Review of Personality and Social Psychology*, Vol. 13 (pp. 25-59). Newbury Park, CA: Sage.
- Marsh, H. W. & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issues, and applications* (pp. 177-198). Thousand Oaks, CA: Sage.
- Matthews, G., Jones, D. M. & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology*, 81, 17-42.
- Mayring, Ph. (1991). *Psychologie des Glücks*. Stuttgart: Kohlhammer.
- McConvill, C. & Cooper, C. (1999). Personality correlates of variable moods. *Personality and Individual Differences*, 26, 65-78.
- Mehrabian, A. & Russell, J. A. (1974). *An approach to environmental psychology*. Cambridge, MA: MIT Press.
- Nesselroade, J. R. & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp.223-255). Thousand Oaks, CA: Sage.
- Ostendorf, F. (1990). *Sprache und Persönlichkeitsstruktur. Zur Validität des Fünf-Faktoren-Modells der Persönlichkeit*. Regensburg: Roderer.
- Pavot, W., Diener, E., Colvin, C. R. & Sandvik, E. (1991). Further validation of the satisfaction with life scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment*, 57, 149-161.
- Pawlik, K. (1988). "Naturalistische" Daten für Psychodiagnostik: Zur Methodik psychodiagnostischer Felderhebungen. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 9, 169-181.
- Penner, L. A., Shiffman, S., Paty, J. A. & Fritzsche, B. A. (1994). Individual differences in intraperson variability in mood. *Journal of Personality and Social Psychology*, 66, 712-721.
- Perrez, M. & Reicherts, M. (1992). *Stress, coping, and health. A situation-behavior approach. Theory, methods, applications*. Seattle: Hogrefe & Huber.
- Pfister, R. (2002). *Flow im Alltag. Untersuchungen zum Quadrantenmodell des Flow-Erlebens und zum Konzept der autotelischen Persönlichkeit mit der Experience Sampling Method (ESM)*. Bern: Lang.
- Prümper, J., Hartmannsgruber, K. & Frese, M. (1995). KFZA. Kurz-Fragebogen zur Arbeitsanalyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, 39, 125-132.
- Radbash, J., Steele, F., Browne, W. & Prosser, B. (2003). *A user's guide to MLwiN*. London: Centre for Multi-level Modelling, University of London.

- Rau, R. & Riedel, St. (2004). Besteht ein Zusammenhang zwischen dem Auftreten von positivem Arbeitserleben unter Flow-Bedingungen und Merkmalen der Arbeitstätigkeit? *Zeitschrift für Arbeits- und Organisationspsychologie*, 48, 55-66.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, St. W., Rowan, B. & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM Algorithm and application to U.S. high-school data. *Journal of Educational Statistics*, 16, 295-330.
- Reich, J. W., Zautra, A. J. & Davis, M. (2003). Dimensions of affect relationships: Models and their integrative implications. *Review of General Psychology*, 7, 66-83.
- Rheinberg, F., Manig, Y. & Vollmeyer, R. (in Vorb.). *Flow bei der Arbeit, doch Glück in der Freizeit. Klären Ziele das "Paradoxon der Arbeit"?* (Manuskript)
- Roduner, K., Venetz, M., Aellig, St., Pfister, R. & Schallberger, U. (2001). *Anleitung zum Arbeiten mit der Experience Sampling Method*. Arbeitsberichte aus dem Projekt "Qualität des Erlebens in Arbeit und Freizeit", Nr. 4. Zürich: Abteilung für Angewandte Psychologie des Psychologischen Instituts der Universität.
- Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, 63, 129-156.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161-1178.
- Russell, J. A. (2003). Introduction: The return of pleasure. *Cognition and Emotion*, 17, 161-165.
- Russell, J. A. & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, 125, 3-30.
- Schafer, J. L. & Graham, J. W. (2002). Missing Data: Our view of the state of the art. *Psychological Methods*, 7, 147-177.
- Schallberger, U. (1999). *Theoretische Rahmenüberlegungen zum Erlebens-Stichproben-Fragebogen und zu den Operationalisierungen*. Arbeitsberichte aus dem Projekt „Qualität des Erlebens in Arbeit und Freizeit“, Nr. 1. Zürich: Abteilung Angewandte Psychologie des Psychologischen Instituts der Universität.
- Schallberger, U., Langemann, D., Sperisen, S. & Spruit, R. (1995). *Flow in Arbeit und Freizeit. Eine Pilotstudie*. Poster, präsentiert am 9. Symposium Arbeitspsychologie, Zürich, 28. - 30.8.1995.
- Schallberger, U. & Pfister, R. (2001). Flow-Erleben in Arbeit und Freizeit. Eine Untersuchung zum "Paradox der Arbeit" mit der Experience Sampling Method" (ESM). *Zeitschrift für Arbeits- und Organisationspsychologie*, 45, 176-187.
- Scherer, K. R. & Wallbot, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66, 310-328.
- Schimmack, U. (1997). Das Berliner-Alltagssprachliche-Stimmungs-Inventar (BASTI). Ein Vorschlag zur kontextvaliden Erfassung von Stimmungen. *Diagnostica*, 43, 150-173.
- Schimmack, U. (1999). Strukturmodelle der Stimmungen: Rückschau, Rundschau und Ausschau. *Psychologische Rundschau*, 50, 90-97.
- Schimmack, U. (2003). Affect measurement in experience sampling research. *Journal of Happiness Studies*, 4, 79-106.
- Schimmack, U., Böckenholt, U. & Reizenzein, R. (2002). Response styles in affect ratings: Making a mountain out of a molehill. *Journal of Personality Assessment*, 78, 461-483.
- Schimmack, U. & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, 14, 325-345.
- Schimmack, U. & Reizenzein, R. (2002). Experience activation: Energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion*, 2, 412-417.
- Schmidt-Atzert, L. & Hüppe, M. (1996). Emotionsskalen EMO 16. Ein Fragebogen zur Selbstbeschreibung des aktuellen emotionalen Gefühlszustandes. *Diagnostica*, 42, 242-267.
- Schmitz, B. (2000). Auf der Suche nach dem verlorenenen Individuum: Vier Theoreme zur Aggregation von Prozessen. *Psychologische Rundschau*, 51, 83-92.
- Schmuckle, St. C., Egloff, B., & Burns, L. R. (2002). The relationship between positive and negative affect in the Positive and Negative Affect Schedule. *Journal of Research in Personality*, 36, 463-475.
- Schumacher, J., Klaiberg, A. & Brähler, E. (Hrsg.) (2003). *Diagnostische Verfahren zu Lebensqualität und Wohlbefinden*. Göttingen: Hogrefe.
- Scollon, Ch. N., Kim-Prieto, Ch. & Diener, Ed. (2003). Experience Sampling: Promises and pitfalls, strengths and weakness. *Journal of Happiness Studies*, 4, 5-34.
- Staudinger, U.M. (2000). Viele Gründe sprechen dagegen, und trotzdem geht es vielen Menschen gut: Das Paradox des subjektiven Wohlbefindens. *Psychologische Rundschau*, 51, 185-197.
- Steyer, R., Schwenkmezger, P., Notz, P. & Eid, M. (1997). *Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF)*. Göttingen: Hogrefe.

- Stone, A. A., Shiffman, S. S. & DeVries, M. W. (1999). Ecological Momentary Assessment. In D. Kahnemann, E. Diener & N. Schwarz (Eds.), *Well-Being: The foundations of hedonic psychology* (pp. 26-39). New York: Russell Sage Foundation.
- Tellegen, A., Watson, D. & Clark, L. A. (1999a). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*, 297-303.
- Tellegen, A., Watson, D. & Clark, L. A. (1999b). Further support for a hierarchical model of Affect: Reply to Green and Salovey. *Psychological Science, 10*, 297-303.
- Tennen, H., Affleck, G. & Armeli., St. (2005). Personality and daily experience revisited. *Journal of Personality, 73*, 1465-1483.
- Thayer, R. E. (1989). *The biopsychology of mood and arousal*. New York: Oxford University Press.
- Thayer, R. E. (1996). *The origin of everyday moods. Managing energy, tension, and stress*. New York: Oxford University Press.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington DC: American Psychological Association.
- Triemer, A. (2003). *Ambulantes psychophysiologisches 24-Stunden-Monitoring zur Erfassung von arbeitsbezogenen Stimmungen und Emotionen*. Frankfurt a. M.: Lang.
- Triemer, A. & Rau, R. (2001a). Stimmungskurven im Arbeitsalltag - eine Feldstudie. *Zeitschrift für Differentielle und Diagnostische Psychologie, 22*, 42-55.
- Triemer, A. & Rau, R. (2001b). *Positives Arbeitserleben: Psychophysiologische Untersuchungen zum Einfluss kognitiv-emotionaler Bewertung der Arbeitssituation auf Wohlbefinden und Gesundheit*. Forschungsberichte der Bundesanstalt für Arbeitsmedizin (FB 907). Bremerhaven: Wirtschaftsverlag.
- Venez, M., Schmid, J. & Schallberger, U. (2001). Sportliche Aktivität und Erlebnisqualität im Alltag. In R. Seiler, D. Birrer, J. Schmid & St. Valkanover (Hrsg.), *Sportpsychologie. Anforderungen, Anwendungen, Auswirkungen* (S. 63-65). Köln: bsp-Verlag.
- Watson, D. & Clark, L. A. (1994). The PANAS-X. *Manual for the Positive and Negative Affect Schedule - Expanded Form* (updated 8/99). Unpublished Manuscript, University of Iowa, Iowa City, IA.
- Watson, D. & Clark, L. A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of Personality Assessment, 86*, 267-296.
- Watson, D., Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scale. *Journal of Personality and Social Psychology, 54*, 1063-1070.
- Watson, D. & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*, 219-235.
- Watson, D. & Tellegen, A. (2002). Aggregation, acquiescence, and the assessment of trait affectivity. *Journal of Research in Personality, 36*, 589-597.
- Watson, D., Wiese, D., Vaidya, J. & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*, 820-838.
- West, St. G. & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality, 59*, 609-662.
- Williams, K. J., Suls, J., Alliger, G. M., Learner, S. M. & Wan, Ch. K. (1991). Multiple role juggling and daily mood states in working mothers: An experience sampling study. *Journal of Applied Psychology, 76*, 664-674.
- Wothke, W. (1996). Models for multitrait-multimethod matrix analysis. In G. A. Marcoulides & F. A. Schumaker (Eds.), *Advanced structural equation modeling. Issues and techniques* (pp 7-56). Mahwah NJ: Lawrence Erlbaum Associates.
- Yik, M. S., Russell, J. A. & Feldman Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology, 77*, 600-619.

Anhang A: Mehrebenenanalytische Modelle

A1 Aufteilung der Varianzen, Kovarianzen und Korrelationen auf die Person- und die Zeitpunktebene

Die Korrelation von zwei Variablen ergibt sich bekanntlich aus der Kovarianz der beiden Variablen geteilt durch die Wurzel aus dem Produkt der Varianzen der beiden Variablen. Das zentrale Problem bei ebenenspezifischen Korrelationsberechnungen ist somit die Aufteilung der Varianzen und Kovarianzen auf die verschiedenen Ebenen. Geht es um ganze Korrelationsmatrizen besteht die effizienteste Methode in folgendem Vorgehen (z.B. Hox, 2002; die notwendigen programmtechnischen Spezifikationen in MLwiN sind beschrieben in: Radbash, Steele, Browne & Prosser, 2003, S. 178ff.):

Es wird ein multivariates Modell definiert, in dem die zu korrelierenden Variablen die abhängigen Variablen bilden. Jede der abhängigen Variablen wird dann durch ein Intercept-only-Modell modelliert:

$$Y_{ijk} = \beta_i + v_{ij} + u_{ijk} \quad (i = \text{Variable}; j = \text{Person}, k = \text{Zeitpunkt})$$

Ein Messwert Y_{ijk} einer Variablen i einer Person j in einem Zeitpunkt k wird damit in drei Komponenten zerlegt: in den Gesamtmittelwert der Variablen i ($= \beta_i$), die Abweichung des Mittelwerts der Person j vom Gesamtmittelwert ($= v_{ij}$) und das verbleibende zeitpunktspezifische Residuum ($= u_{ijk}$). Während β_i einen festen ("fixed") Effekt darstellt, also (für eine bestimmte Variable) eine Konstante darstellt, handelt es sich bei den zwei andern Termen um Random-Effekte (Zufallsvariablen), die je nach Person (v_{ij}) bzw. Person und Zeitpunkt (u_{ijk}) variieren. Die entsprechenden Varianzen liefern eine Aufteilung der Gesamtvarianz der Variablen in ebenenspezifische Varianzanteile:

$\text{Var}(v_{ij}) = \text{Varianz auf der Personenebene}$

$\text{Var}(u_{ijk}) = \text{(gepoolte) Varianz auf der Zeitpunktebene}$

Analog können die Kovarianzen zwischen zwei Variablen auf die Anteile, die beiden Ebenen entstammen, aufgeteilt werden, z.B. bei den Variablen 1 und 2:

$\text{Cov}(v_{1j}, v_{2j}) = \text{Kovarianz auf der Personenebene und}$

$\text{Cov}(u_{1jk}, u_{2jk}) = \text{(gepoolte) Kovarianz auf der Zeitpunktebene.}$

Damit sind alle Parameter gegeben, die es braucht, um die *Korrelationen* auf den beiden Ebenen zu berechnen. In MLwiN können die entsprechenden Koeffizienten in den "Estimate tables" (Model Menü) verlangt werden.

Die geschätzten Varianzanteile können auch dazu verwendet werden, um die *Intraclass-Korrelationen* zu bestimmen ($\rho_{ii} = \text{Var}(v_{ij}) / (\text{Var}(v_{ij}) + \text{Var}(u_{ijk}))$).

A2 Das Modell von Raudenbush et al. (1991) zur Bestimmung ebenenspezifischer Konsistenzkoeffizienten

Das Modell zur Schätzung ebenenspezifischer Konsistenzkoeffizienten wird in Raudenbush et al. (1991) ausführlich begründet, beschrieben und anhand von Daten zum Thema Schulklima

illustriert. Es ist auch bei Hox (2002, Kap. 9.3) dargestellt, wo allerdings – v. a. im Datenbeispiel – verschiedene Druckfehler vorhanden sind (siehe dazu Homepage von Joop Hox).

Ausgangspunkt ist ein multivariates Modell, indem die Items einer Skala die abhängigen Variablen bilden. Die Items werden (durch ein Set von Dummyvariablen) simultan betrachtet. Das Modell ist ein Intercept-only Modell:

$$Y_{ijk} = \beta_0 + v_{0k} + u_{0jk} + e_{0ijk} \quad (k = \text{Person}; j = \text{Zeitpunkt}; i = \text{Item})$$

Y_{ijk} = steht für die Antwort der Person k zum Zeitpunkt j im Item i und wird in einen festen Effekt und drei Zufallsvariablen zerlegt: Der feste Effekt β_0 ist der Mittelwert aller Items über alle Zeitpunkte und Personen, also des Skalenwerts (ausgedrückt als Itemmittelwert). Die drei Random-Effekte sind: v_{0k} = die Abweichung der Person k (über alle Zeitpunkte) von diesem Mittelwert; u_{0jk} = die Abweichung der Person k im Zeitpunkt j von der Summe der zwei bisher genannten Komponenten; e_{0ijk} = die Abweichung der Person k im Zeitpunkt j im Item i von der Summe der drei vorstehenden Komponenten. Die Varianzen dieser Randeffekte lassen sich somit wie folgt verstehen:

$$\text{Var}(v_{0k}) = \text{Varianz des Skalenwerts auf der Ebene der Personen} = V_{\text{person}}$$

$$\text{Var}(u_{0jk}) = \text{Varianz des Skalenwerts auf der Ebene der Zeitpunkte} = V_{\text{zeit}}$$

$$\text{Var}(e_{0ijk}) = \text{Varianz zwischen den Items}$$

Der letztgenannte Term kann zur Bestimmung der Fehlervarianz des Skalenwerts (= Mittelwert der Items), die durch die Inkonsistenz der Itemantworten verursacht wird, benutzt werden:

$$\text{Var}(\text{Fehler}) = V_{\text{Fehler}} = (\text{Var}(e_{0ijk}) / p) \quad \text{wo } p \text{ für die Anzahl Items steht}$$

Die ebenenspezifischen *Konsistenzkoeffizienten* kann man dann wie folgt (von Hand) berechnen:

$$\alpha_{\text{zeitpunkt}} = V_{\text{zeit}} / (V_{\text{zeit}} + V_{\text{Fehler}})$$

$$\alpha_{\text{person}} = V_{\text{person}} / (V_{\text{person}} + (V_{\text{zeit}} / n_k) + (V_{\text{Fehler}} * 1/n_k)) \quad \text{wo } n_k \text{ für die Anzahl}$$

Zeitpunkte der Person k steht – oder (um eine durchschnittliche
Konsistenzschätzung auf der Personenebene zu erhalten) die über die
Personen gemittelte Anzahl Zeitpunkte

Die obige Zerlegung kann auch zur Schätzung der *Intraclasskorrelation* der mit den Items erfassten latenten Variablen benutzt werden:

$$\rho_1(\text{latent}) = V_{\text{person}} / (V_{\text{person}} + V_{\text{zeit}})$$

Anhang B: Zusatztabelle

Tabelle B1

Verteilungskennwerte der individuellen Mittelwerte, Standardabweichungen, Trennschärfekoeffizienten (r_{it}) bzw. Konsistenzkoeffizienten (Cronbach- α) der PANAVA-Items bzw. Skalen

Skala bzw. Item	individuelle Mittelwerte		individuelle Standardabweichungen		individuelle r_{it}		individuelle Cronbach- α	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Positive Aktivierung (PA)	18.8	2.2	4.2	1.1			.77	.11
energieelos-energiegeladen	4.8	0.6	1.3	0.4	.66	.15		
müde-hellwach	4.3	0.8	1.6	0.4	.60	.16		
lustlos-hoch motiviert	4.9	0.6	1.2	0.3	.69	.14		
gelangweilt-begeistert	4.8	0.5	1.1	0.3	.50	.18		
Negative Aktivierung (NA)	11.0	2.7	4.1	1.2			.76	.14
entspannt-gestresst	3.1	0.8	1.5	0.4	.65	.17		
friedlich-verärgert	2.3	0.7	1.2	0.4	.61	.18		
ruhig-nervös	2.6	0.8	1.2	0.4	.63	.18		
sorgenfrei-besorgt	3.0	0.9	1.2	0.4	.52	.21		
Valenz (VA)	10.9	1.2	1.9	0.6			.70	.18
unzufrieden-zufrieden	5.6	0.6	1.2	0.4	.58	.19		
unglücklich-glücklich	5.3	0.7	1.0	0.3	.58	.19		

Anmerkungen: $N = 23'509$ Zeitpunkte von 530 Personen mit $M = 44.4$ Zeitpunkten ($SD = 4.0$). r_{it} = part-whole korrigierte Trennschärfe; α = interne Konsistenz. Angegeben sind jeweils die Mittelwerte M und Standardabweichungen SD der Ergebnisse aus den 530 intraindividuellen Item- und Konsistenzanalysen. Die (kleinen) Differenzen zu entsprechenden Angaben in Tabelle 4 erklären sich durch das unterschiedliche Vorgehen (Mehrebenenanalyse vs. individuelle Analysen).

Tabelle B2

Interkorrelationen der PANAVA-Items auf der Personenebene (oberhalb der Diagonalen) und auf der Zeitpunktebene (unterhalb der Diagonalen)

	energiegeladen	hellwach	motiviert	begeistert	gestresst	verärgert	nervös	besorgt	zufrieden	glücklich
energiegeladen	-	.69	.83	.66	-.34	-.46	-.37	-.32	.58	.49
hellwach	.60	-	.64	.46	-.52	-.39	-.41	-.40	.47	.40
motiviert	.59	.53	-	.80	-.44	-.58	-.46	-.43	.67	.63
begeistert	.41	.35	.59	-	-.36	-.59	-.42	-.46	.70	.68
gestresst	.06	.02	-.12	-.19	-	.64	.76	.61	-.62	-.54
verärgert	-.06	-.05	-.27	-.31	.53	-	.77	.63	-.84	-.75
nervös	.09	.06	-.08	-.14	.62	.52	-	.62	-.70	-.58
besorgt	-.08	-.08	-.23	-.28	.45	.48	.44	-	-.64	-.70
zufrieden	.18	.12	.36	.39	-.51	-.68	-.46	-.49	-	.80
glücklich	.21	.18	.39	.41	-.43	-.57	-.39	-.52	.61	-

Anmerkungen: $N = 23'509$ Zeitpunkte von 530 Personen mit $M = 44.4$ Zeitpunkten ($SD = 4.0$). Die Korrelationen beruhen auf einer mehrebenenanalytischen Zerlegung der Varianzen und Kovarianzen, die in Anhang A1 beschrieben ist.