How to Assess Social Regulation of Stress and Emotions in Daily Family Life? A Computer-assisted Family Self-monitoring System (FASEM-C)

Meinrad Perrez*, Dominik Schoebi and Peter Wilhelm
Department of Psychology, University of Fribourg, Switzerland

A major problem for scientific and practical work in many branches of psychology is the reliable access to phenomena under study. Traditional assessment procedures are often reliant on retrospective recall, self report measures and/or laboratory simulations. All these have their strengths and weaknesses. Ideally observations of people's behaviour and interactions, and the monitoring of thoughts and feelings, should be of real events as they occur. This paper focuses on some methodological and reliability issues associated with development in hand-held computer self-recording systems and describes a new computer-assisted self-monitoring system for emotional processes in families with adolescents. The method is based on simultaneous self-observations of family members older than 13 years. The paper describes the procedures involved in using this approach and demonstrates how to obtain indicators for reactivity, reliability and some aspects of validity. Copyright © 2000 John Wiley & Sons, Ltd.

INTRODUCTION: AMBULATORY CLINICAL PSYCHOLOGICAL ASSESSMENT

The last 20 years has seen a growth of research on the role of early family environments in creating vulnerability to social relationship difficulties and psychopathology. This has been advanced with the development of self-report questionnaires requiring people to recall their experiences, often many years later. Many authors to this special edition outline findings using this methodology. However, both reliance on subjective evaluations and memory of distal events are not without their problems. Self-report data have the possible disadvantage of being influenced by 'mood-congruent memory processes' (Blaney, 1986), by 'processes of cognitive consistency' (Ross, 1989) and by processes of causal attributions for understanding one's own behaviour. These and other factors can impair validity and reliability of self-report data. Although retrospective recall is less susceptible to mood biases than previously thought (Brewin *et al.*, 1993), there is convincing evidence that memory disturbances increase with the time latency between the event and its retrieval processes from memory, if the

*Correspondence to: Professor M. Perrez, Department. of Psychology, University of Fribourg, 2 Rue de Faucigny, CH-1700 Fribourg, Switzerland. E-mail: meinrad.perrez@unifr.ch

event is not a clearly defined macro-event. Specific retrospection effects for self-monitoring of mood states and somatic and psychological complaints have been observed even for short latency times, such as days (Margraf and Schneider, 1990) or hours (Fahrenberg *et al.*, 1999; Käppler *et al.*, 2000).

Until recently the only alternative source of data was systematic observation in laboratory and field situations. These are appropriate for the observation of behavioural data but not for recording cognitive processes and experienced emotions in naturalistic setting as they happen. To address this problem new methods of data gathering have slowly been evolved, one of the most important being the Experience Sampling Method (EMS). This method emphasizes the representativeness of experiences from natural settings (Csikszentmihalyi and Larson, 1987; de Vries, 1992; Larson and Richards, 1994). An early approach asked subjects to monitor their thoughts and behaviours and complete booklets in the form of daily diaries. More recent developments have enabled people to record their thoughts and feelings to actual events as they occur or shortly afterwards with the use of pocket-computer or hand-held computers instead of booklets. Pocket computers have many advantages (Fahrenberg, 1997; Wilhelm and Perez, 2000).

- Subjects are reminded to record at the scheduled time by the signalling function. This reduces the missing data and improves the representativity of the sampled situations
- Technical functions of control and blocking prevent unanswered questions
- Previously recorded data cannot be revised and changed
- The data transfer to the personal computer is economic and coding errors are avoided
- The exact answering time is being recorded automatically. Compliance can therefore be precisely controlled
- Specific hierarchically-structured questions elicit information on a variety of person–environment interactions.

In Europe several computer-aided assessment systems for different types of disorders are under evaluation, especially in Germany (cf. Fahrenberg and Myrtek, 1996). We developed a systematic self-observation method based on the use of pocket computers. Our assessment procedure is built on the experience with the COMRES (COMputer REcording System), that Perez and Reicherts (1996)

developed and evaluated in earlier projects investigating *individual stress* experiences and coping with them. The pocket computer was used as an external memory for stress, that could be directly applied in stressful situations or other situations of interest. This allowed us to minimize subjective retrospective distortion and other memory problems.

A new development is the assessment of covert and overt behaviour in the context of *daily family life*. In a first attempt we developed a family self-monitoring system (FASEM), that was set up for booklet use and pocket computers (Perez *et al.*, 1998). With this system we gathered data from 96 families in Switzerland (Perez *et al.*, 1998; Perez and Wilhelm, 2000) and 26 families in Canada (Quebec) (Schumacher, 2000) with a total of 416 subjects. The present paper introduces the revised version of the Family Self-Monitoring system (FASEM-C). We first give an overview of the conceptual background and then describe the assessment procedure in more detail. We further demonstrate how to compute differentiated methodological indicators for this type of field data, especially for different concepts of reliability, reactivity and some aspects of validity.

THEORETICAL BACKGROUND OF THE FAMILY SELF-MONITORING

In the clinical context negative emotions and their regulation play a key role in psychopathology and psychotherapy of individuals. They are also of central importance for the functioning of intimate relations. Negative emotions are often triggered by stressful events. To analyse the stress experience and coping behaviour of fathers, mothers and adolescents living together in a family, we have proposed a *conceptual basis* (Perez, 1995). This involves assessing stress in the individual family members, in subgroups of the family and in the whole family (see Figure 1). We also focus on how each member of the family, a subgroup or the whole family reacts to emotional disequilibria or a stress reaction of an individual, a subgroup or the family as a whole. For the dyadic constellation, Bodenmann (1995) has proposed detailed conceptual and theoretical differentiations.

The theoretical framework focuses on the micro-analysis of the *social regulation of emotions* in family daily life. We assume that essential components of this regulation are appraisal and evaluative processes, and emotional states. For example: if the

Disturbance of equilibria	Adaptive reactions (dysfunctional or functional)		
	Individual	Subsystem	System
Individual	individual stress and coping		
Subsystem		e.g. dyadic stress and coping	
System			e.g. family stress and coping

Figure 1.   Matrix of equilibria and coping disturbances in families

father encounters stressors that disturb him, what are his typical cognitions and evaluations? Is it possible to identify styles of appraisal (e.g. external or internal causal attributions), which are typical for him in specific situations? As cognitive appraisal processes are important for stress experience and coping, we focus on *valence* (positive and negative aspects of stress), the perceived control, and causal attributions. We explore emotional states and emotions (positive and negative) and the appraisal characteristics which are related to emotional (and somatic) states. Generally, cognitive evaluations have been shown to be related to emotions in theoretically predictable ways (Scherer, 1988; Lazarus, 1991; Perrez and Reicherts, 1992). Finally, we study coping behaviours (both individual and social). We assume that functional and dysfunctional coping reactions can be distinguished for the individual (Reicherts and Perrez, 1992; Reicherts, 1999) as well as for the social level and their interaction. For example a functional coping reaction by the family or a subgroup should not hurt one's own or the other person's self-esteem (Perrez and Schoebi, 2000). Social coping reactions considered to be functional should lead, in the short term, to better emotional and social well-being and in the long run it should have a positive influence on the individual, dyads (partner satisfaction) and family (cohesion and family satisfaction). The opposite is assumed for dysfunctional coping tendencies). In the current project we are able to test some of these expectations.

THE ASSESSMENT OF EMOTIONAL AND SOMATIC STATES. ANTECEDENTS AND COPING UNDER NATURAL CONDITIONS

The microanalysis of social emotional regulation needs four types of psychological information, as

depicted in Figure 2. These are (i) information on antecedents of emotions such as the setting, the events and appraisal cognitions; (ii) information on the experienced emotional and somatic states; (iii) information on individual and social (adaptive and maladaptive) responses (what people actually did), and (iv) information on the individual and social short-term effects. Our intention was to cover the whole range of emotional states from negative stress emotions to positive states. Information on the somatic state is included in the form of somatic well-being and somatic complaints.

In order to avoid the impairing effects on validity of self-report data, especially of retrospective recall we needed assessment strategies which met the three following criteria: (i) the procedure should allow recording of experienced emotional and somatic states, recording of cognitive characteristics, and recording of the social and environmental conditions in which these states and events occur; (ii) the time lag between states and events and their recording should be as short as possible, in order to minimize memory distortions. The recording should take place when the subjects are still in a state of emotional arousal for storing 'hot emotions' and 'hot cognitions'; (iii) the method should assess psychologically relevant data, not using a diary free text self-description, but structured according to the theoretical framework.

OVERVIEW OVER THE FAMILY SELF-MONITORING SYSTEM (FASEM-C)

The method is devised for small intimate groups, who may include children older than 13 years. It can also be used for couples or adapted for individuals in the preventive or therapeutic context. Table 1 shows the different information types and item formats assessed by the family self-monitoring system.
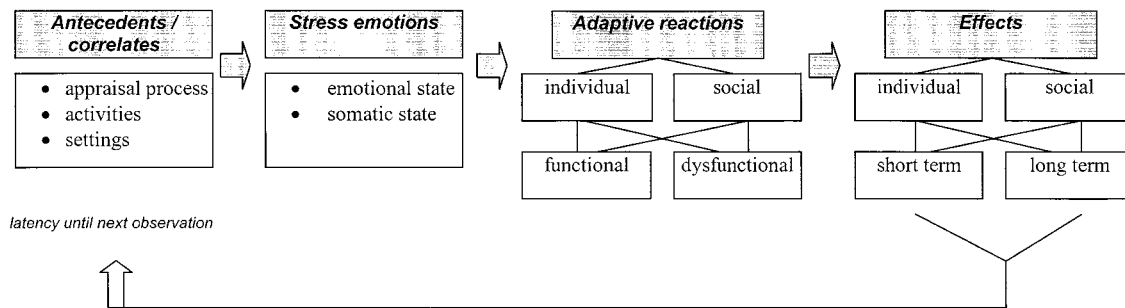
Figure 2.   Structure of stress episode

Table 1.   FASEM-C information types and item formats (questions, triggered by situational keys in italic)

	Information type	Item types
1.	Duration and quality of sleep (First observation of the day)	Five items (6- and 7-point scales)
2.	Emotional and somatic state	11 items (6-point scales)
3.	Causal attribution	
	• Internal or external attribution	Items series (3-point scales)
4.	*Control expectation*	*(triggered by 2.)*
	• *Internal/external*	*Item series (3-point scales)*
5.	Somatic complaints/pain	Yes/No, triggers:
	• *Type/attribution*	*Items series (categories)*
6.	Place	Six categories
7.	Social presence	Yes/No, triggers:
	• *Presence of other persons*	*Item series (categories)*
	• *Evaluation of pleasantness*	*Item series (6-point scale)*
8.	Evaluation of partner's affective state (only for parents)	Eight items (categories and 6-point scales)
9.	Actual activity	Item series (categories)
	• Valence of activity and duration	Item series (6-point scales, categories)
10.	Conflicts with others	Yes/No, triggers:
	• *Social coping*	*Item series (3-point scales)*
	• *Description of own behaviour*	*Item series (3-point scales)*
	• *Description of other's behaviour*	*Item series (3-point scales)*
11.	Individual stress	Two items (category and 3-point scale), triggers:
	• *Individual coping*	*Item series (3-point scales)*
12.	Social support	Yes/No, triggers:
	• *Details*	*Item series (categories, 6-point scales)*

Situation-Sampling

The system works with a time sampling method supplemented with an event sampling strategy. The event sampling covers events during the night, assessed at the first observation in the morning after waking up, and the events and behaviours between the actual and the last self-observation. For the time sampling, the computer alerts the subject acoustically according to a random time-point inside a time window of about 3 h six times per day over 1 week. At these measurement points all subjects had to record their current emotional and somatic state and the other information mentioned above. The signal-contingent assessment assures the simultaneous recording of all members of the family (older than 13 years).

One part of the information is recorded at every observation and another part depends on how a certain previous question has been answered. For example: if the question 'Do you feel any pain or physical complaints?' is answered with 'yes', one has to answer two supplementary menus: 'Which kind of pain or complaint do you feel, and how strongly?' and 'What do you think, is the reason for your complaints?'. If one answers 'no', the sup-

plemental questions are skipped. In the same way social support and coping reactions are assessed. Coping responses are only asked if an individual or a socially stressful situation has been previously observed. Questions on social coping depend on the involvement of other people in the situation (e.g. conflicts). For situations, requiring adaptation without social involvement, individual coping items are presented. Further the program only presents questions concerning control expectations, if a negative emotional state has been recorded.

Technical Aspects

The palmtop computer chosen for our studies was the HP 360 LX based on the Windows CE operating system. It has a pencil pointing device, a complete Qwerty and touch screen with a resolution of 640×240. Data entry was done almost exclusively using the pointing device with the keyboard being only of marginal utility. These characteristics make it possible to carry it during everyday activities and allow the immediate entry of information in a structured format. All subjects have been trained in self-observation before the observation period. To minimize the reactivity the questions are as brief as possible and every reply is immediately recorded. It is also impossible for a user to miss completely or incorrectly fill in a question as sometimes occur in the booklet version. The computer also allows the self-observation grid to be personalized using the individual first names of the family members.

Our program separates the questionnaire interpreter from the questionnaire specification. This allows for people to use the computer-offered questionnaire without any knowledge of computer programming. It also allows easy adaptation to different languages, which is important in Switzerland, with different official languages.

Description of the Sample

The data reported here, are gathered in the context of an interdisciplinary project supported by the Swiss National Foundation. For the selection of families we used the following criteria: each family should consist of both parents and at least one adolescent between 14 and 17 years, living together in the same household. We also made sure that the self-observation took place during a normal family week and not during holidays or while moving house. We recruited families through the German- and French-speaking school administrations of sec-

ondary schools, using the town population census (information freely available to the public). To each family (about 5200) a letter was written, to inform them about the purpose of our study and to invite the family to participate.

Interested families sent back a coupon, signed by each participating family member. A total of 101 two-parent families were willing to participate. Two families cancelled their participation before starting. Two stopped the participation because of computer problems. One family was from Kosovo and was dropped from data analysis because the mother had language problems. Another family just finished self-observation and the data were not yet ready for analysis. The following results are based on 95 families with 311 persons (95 mothers, 95 fathers, 56 female and 65 male adolescents); 39 families were French-speaking, the other 56 were German-speaking. The average parent was 46.13 years old (SD = 5.86 years), and the adolescents/young adults 15.45 years (SD = 1.12 years). Most families belonged to the middle class with a comparably high educational level.

Procedures

Data acquisition took place from November 1998 to April 2000. Interviewers visited the family at home. The participants were given a first introductory training in self-observation with the pocket computer. In addition, subjects completed several questionnaires including: the Relationship Assessment Scale (Hendrick, 1988; Sander and Böcker, 1993), measuring cohesion, adaptation, communication and family satisfaction, the Family Life Satisfaction Scale (Schneewind and Weiss, 1999) and others appropriate for this study. All scales were translated into French and then retranslated into the original version to check accuracy. To enhance motivation, the adolescent participants received a coupon for a CD with a value of about 30 sFr at the end of their self-monitoring task. All families took part in a raffle for a holiday prize with a value of about 3500 sFr, and they could ask for information about their results.

RESULTS

Acceptance of the Procedure

All subjects were questioned after the self-observation week about their experience with the assessment method: 76% felt that the duration of the

recording task was acceptable; for the remainder it was too long (17% said a little too long, 6% too long and 1% much too long). Nearly all participants (94%) judged the work with the pocket computer as positive. These results are in accordance with the experience of the first study and belies the expectation, that ambulatory computer-aided self-monitoring disturbs participants, as was suggested by Asendorpf and Wilpers (1999).

Accuracy and Duration of Recording

The palmtop computer automatically stores the real recording time and the duration of every assessment. This gives a measure of the subject's commitment. More than 60% of the recordings started within 5 min after the signal 75–80% of the subjects completed their self-monitoring task within a 30-min time limit. The *duration* of recording depends on whether there was a need for adaptation or not (coping questions). For the short version the mean needed time for the protocol was 4.65 min (SD = 2.89 min) and for the complete version 7.28 min per protocol (SD = 3.82 min).

Reliability

In the context of self-observation methods there are different possibilities for obtaining estimates of the reliability of measures (Buse and Pawlik, 1996; Wilhelm and Perrez, 2000): the local reliability, the total–aggregate reliability, the situation-specific reliability and for simultaneous observations, the inter-rater reliability. We give examples for each type of reliability.

Single Observation

The local or single observation reliability refers to the reliability of a measure at a single observation. We computed the local reliability for the emotional and physical state. These two correlated dimensions (median $r = 0.63$) were the result of repeated factor analysis (for every observation) over the bipolar items belonging to the question, 'How do you feel at the moment?' The emotional state is a composite of the following eight items: tensed–relaxed, unsatisfied–satisfied, stressed–easy, sad/depressed–happy, anxious/concerned–confident, angry–peaceful, secure–insecure, quiet–excited. Cronbach's alpha was computed for each of the 42 observations, yielding an alpha between 0.92 and 0.95 (median = 0.935). The somatic state consists of two items: bodily unwell–bodily well, without energy/tired–full of energy.

Cronbach's alpha for the somatic state was between 0.46 and 0.81 (median = 0.708).

Total–Aggregate Reliability

The total–aggregate reliability provides information on the reliability of a variable that was collapsed over all the observations. To compute the total–aggregate reliability the 42 observations were divided into two samples according to the odd–even method, with a slight modification, to ensure, that also the times of the day were equally distributed in each sample: the first sample consisted of the observations 1, 2 and 3 at the first, third, 5th and 7th day and observations 2, 4 and 6 of the second, fourth and 6th day. The second sample contained the remaining 21 observations.

For each half, the intrapersonal mean for metric variables and the frequency percentage for categorical variables was calculated. The means or frequencies of both halves were correlated with one another and evaluated according to the formula of Flanagan. The calculation of the split-half reliability according to Flanagan corresponds in the case of identical standard deviation of both test halves, to the results of Spearman-Brown and gets a more exact estimate of the actual reliability in the case of different standard deviations (Lienert and Raatz, 1994). To calculate the average split-half reliability (total–aggregate reliability) for the items belonging to a question, the coefficients were subjected to a Fischer's z-transformation.

The average total–aggregate reliabilities were quite high. For the average emotional state as well as for the average somatic state the total–aggregate reliability was $r = 0.98$, for bodily complaints (answering the question 'Do you feel any pain or do you have any physical complaints at the moment?' with 'yes' or 'no'), $r = 0.97$, for feeling hungry, $r = 0.91$ and for the evaluation of the activity as pleasant or unpleasant, $r = 0.89$.

Another measure of interest is the variability over the observations. So for each half the intraindividual standard deviation was computed and correlated as described above. For the variability of the emotional state the split-half coefficient was $r = 0.85$ and $r = 0.86$ for the variability of the somatic state, $r = 0.87$ for being hungry, and $r = 0.74$ for the evaluation of the activities.

Situation Specific Reliability

Generally more interesting than the total–aggregates are setting- or situation-specific reliability scores.
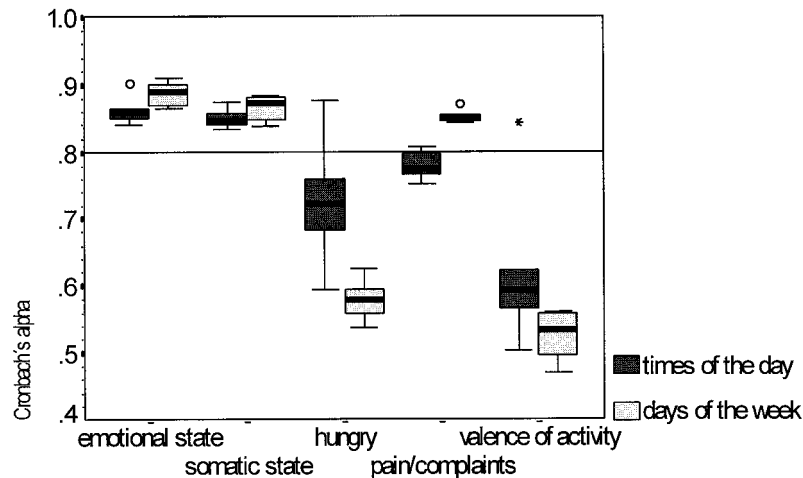
Figure 3 shows the Cronbach's alpha coefficients

Figure 3. Time-specific reliability coefficients. Range of Cronbach's alpha coefficients computed over every single day of a week and over times of a day. For this analysis *n* varies between 180 and 248 persons, because subjects with missing observations had to be dropped

for scores computed over the six observations of each day of the week and the seven observations of each day. For the emotional state as well as for the physical state and the bodily complaints the coefficients are quite high for the days of the week as well as for the times of the day. Generally lower coefficients resulted for the single items feeling hungry–full, and the evaluation of the activity 'at the moment'. These items indicate fast changing states and are therefore less stable. Especially in these items, the consistency over the times of the days was higher than the consistency over the days of the week.

For differentiated situation-specific analysis one often has to take into account more than one criterion to control the effects of other variables. The consequence is that the number of observations often decreases dramatically. Therefore the central question is: how many observations have to be aggregated to obtain scores with satisfactory reliability?

To estimate the reliability as a function of the number of observations the Pearson correlation coefficients were computed for different sample sizes, according to Epstein (1977). To demonstrate reliability of the different variables under specific conditions, we selected all the observations, when another family member was present. Then two single observations (s1 and s2) were correlated with each other. Then two situations (s1, s3) were averaged and correlated with the average of two other

observations (s2, s4) and so on until each mean based on 10 different observations was achieved.

As can be seen in Figure 4, it depends on the variable, how many situations one has to aggregate to obtain a satisfactory reliability of e.g. $r = 0.80$. To estimate the general emotional or somatic state in situations being together with other family members four situations are sufficient, six are needed for complaints and seven for the evaluation of the presence of the other family members as pleasant or unpleasant. To achieve reliable estimates for feeling hungry, or the general evaluation of the current activity in this kind of situation, one needs to have more then 10 situations. This is because the evaluation of the activity depends more on the type of activity than on the presence of family members. Being hungry or not has more to do with the time of the day, than with the social contacts (cf. Figure 3 and Figure 4).

If one does not care about a specific situation and just takes an unsystematic sample of observations, the reliability based on one or two observations is slightly lower, but then increases almost in the same way as the coefficients of the specific situation shown in Figure 4. The only substantial difference one can find is that the reliability of feeling hungry grows faster and reaches a coefficient of $r = 0.79$ after 10 situations. In summary, the last results show that it is possible to generalize the growth in reliability to other situations. As a rule, aggregations based on at least five or six

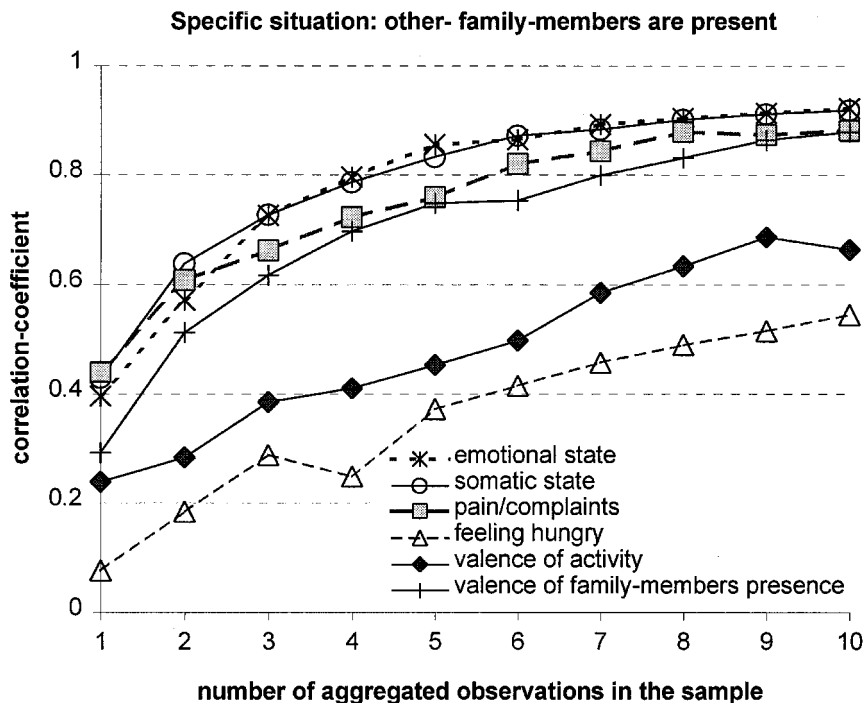**Specific situation: other- family-members are present**

Figure 4. Reliability coefficients as a function of the number of observations included. Only situations were selected, when other family members were present. For the coefficients that were based on one to five aggregated observations the number of persons was held constant ($n = 253$). The other coefficients are based on decreasing number of persons (for 10 aggregated observations $n = 128$ was the lowest)

observations show a high internal consistence for the most important variables.

Inter-rater Reliability

For observable facts like the place, where somebody is located at a particular time, or the presence or absence of other persons, we can look for agreement between the protocols of the family members and obtain measures for inter-rater reliability. So if the mother tells us that the father is present, does the father also report that the mother is present at that time? To answer this question, we computed the percentage agreement over all the 42 observations of each dyad: father–mother, father–adolescent, mother–adolescent.[1]

Because the percentage agreement score counts

agreement in the presence as well as in the absence of the event of interest, it has been criticized in overestimating the concordance, if the event has a low frequency (see Rosenthal, 1987). Therefore we computed a second percentage score that ignores absence and only counts for agreement, when the presence of the other person was indicated in accordance (see Mees, 1977). This score is a much more conservative estimate of the agreement. Kappa coefficients computed on the situation level (42 observations by 95 couples − missings = 3395 to 3545 valid situations) correspond very well to the mean of those scores. With the minicomputer it can be checked whether the agreement becomes better when the time difference between the protocol of the two family members decreases. To do this we exclude successively all the observations, that span a time limit of 1 h, 30, 15 and 5 min.

Figure 5 shows the results for each dyad. The average percentage agreement scores, that were based on presence and absence (light symbols), are around 80 and did not change much with the time-

---

[1] There were 26 families in which two adolescents participated in the study. For this analysis we took only one adolescent for each family. Selection criteria were sex (to ensure equal distribution of boys and girls), and age (the older one was taken).
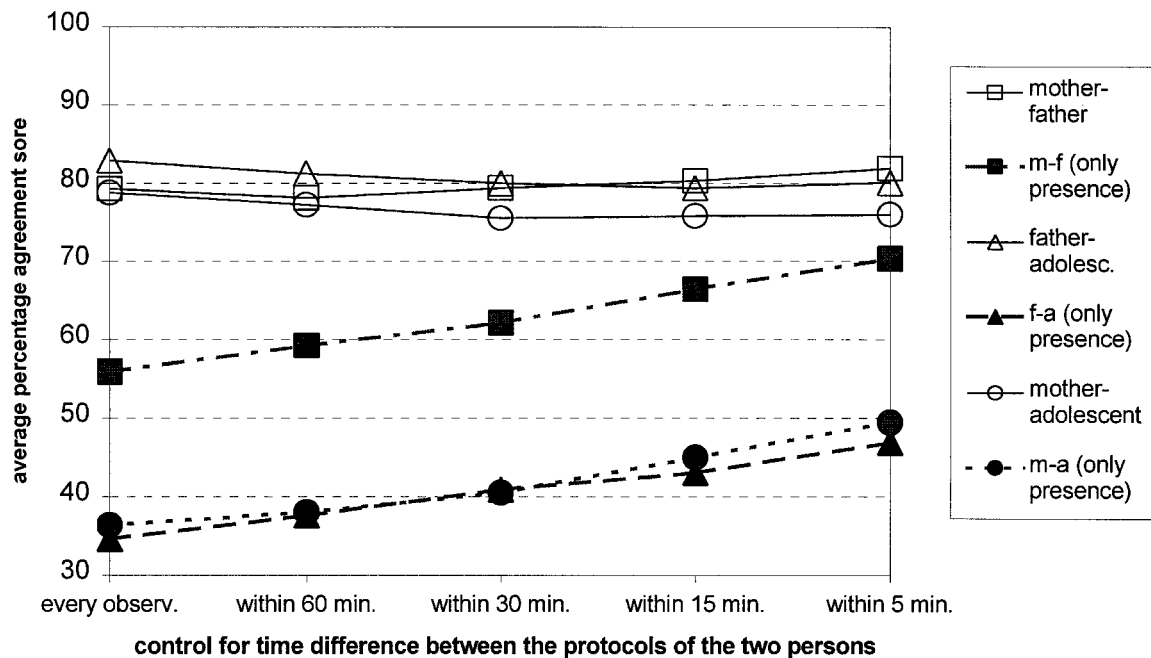
Figure 5. Average percentage agreement scores between family members, as a function of the time difference between protocols within a dyad. Light symbols represent the averaged percentage agreement scores for presence as well as for absence of the other person. Dark symbols with broken lines represent the averaged agreement scores that ignore absence and only count presence

lag controlled. The conservative agreement estimates are very low when only presence is counted, especially for the dyads with the adolescents. These agreement scores increase linearly if the time difference becomes smaller. Further examination of the data at the situation level, found that the discordance was more than 50% due to the fact that one family member indicated that the other was there, while the other reported that they were alone. This was most often the case when the mother or the father said that the adolescent was present (65% and 62%). With regard to agreement about being in the same place, without controlling for the time difference, the mean percentage score was 91% for the parents, 86% for the father–adolescent and 90% for the mother–adolescent dyads and 94, 95 and 95% when a time difference of only 5 min was allowed.

Reactivity

If the same questions are asked six times a day over the course of a week, it is probable that the way the questions are answered changes with time, due to learning-, fatigue- or sensitivity effects. Our

main expectation was that the family members learn how to use the instrument in a more economic way. To explore this we looked at 'key' questions, that were followed by more detailed questions. Another possibility, to examine economy effects, was to look for changes in the way the persons indicated their emotional state over the course of the 7 days. Do they differentiate less between the single items to indicate their emotional state at each observation? Do they vary less in their emotional state over different observations? Is there an increase in the level of the emotional state over time? The same questions could also be asked for the estimation of the partner's emotional state.

In order to analyse the hypothesis we computed for every person the following scores for each of the seven observation days:

(1) the frequency of missing observations (as a percentage);
(2) the frequency of agreement to the above mentioned key questions (as a percentage);
(3) the mean of the variation between the single emotional state items (computed as the stan-

dard deviation between the items creating the score), as an indicator for item stereotypy;
(4) the standard deviation of the emotional state score as an indicator for situation stereotypy;
(5) the mean of the emotional state score.

The last three scores (3–5) were also computed for the estimation of the partner's emotional state.

Every score was then treated as a dependent variable in a multilevel analysis with the MLn Program (The Multilevel Models Project, Department of Mathematics, Statistics and Computing, Institute of Education, University of London). Multilevel models (or random coefficient models, see Longford (1995) or for an example Hedeker *et al.* (1996)) allow the hierarchical structure of the family self-observation data to be modelled and controlled where there are at least three different levels: within families there are individuals, and within individuals we have multiple observations. Therefore a three-level model was chosen, allowing the intercept to vary at each level. We postulated a linear decrease over the course of the week for most variables except frequency of missing and emotional state, where we expected an increase. The variable indicating the observation days was therefore effect-coded. To control for the effects of important other variables we put further predictor variables into the analysis: weekend (weekend versus weekdays), generation (parent versus adolescent), sex and the interaction between generation and sex. Furthermore we allowed for variation in the weekend effects at the individual level (weekend as an additional random variable at level 2). Because the distribution of the frequency variables was skewed, the residuals at the individual level and the family level were not equally distributed around zero. For these variables additional Friedman tests for non-parametric data were carried out and showed essentially the same results.

The results in Table 2 show that all the variables have a highly significant linear change over the course of the observation in the predicted direction, with only one exception. That means that over the course of the observation people learn how to use the observational system in a more economic way: they tend to avoid answering questions that are followed by more detailed questions. They also tend to document their emotional state in a more stereotypic way. This results in a slight increase of the emotional state score. Essentially the same was true for the estimation of the partner's emotions. It should be noted that the high z- and p-values

Table 2. Reactivity analysis: linear relationship between observation over the course of the week and different indicators of economic reponse behaviour (results of the multilevel analysis)

Variable	Fixed parameter estimates	z-Value
Missing	1.21	8.67*
Somatic complaints	−0.69	3.05*
Contact to other persons	−0.34	1.54
Consuming something	−1.05	5.16*
Social conflict	−0.91	7.20*
Individual stress	−1.34	9.72*
Social support	−0.37	3.01*
Item stereotypy (low value = high stereotypy)	−0.019	10.64*
Situation stereotypy (low value = high stereotypy)	−0.018	6.43*
Mean level of emotions	0.014	3.60*

Effect coded variable: (first day = −3; second day = −2 … to seventh day = 3)
* $p \leqslant 0.001$.

are due to the high power of the multilevel analysis (311 persons × 7 days − 39 missing = 2138 cases), and that the effects are rather medium or small; e.g. an estimated increase in the emotional state between the first and last observation day of 0.09 on a 6-point scale. The question regarding the presence of other people was not affected by a reactivity effect. If anybody else is present is an objective fact. All the other questions require a judgement of an inner state which is much less precise and therefore more delicate for reactivity effects.

Another indicator for reactivity is the subjective estimation of the subjects concerning the influence of the assessment procedure that they experienced. The participants were asked after their self-observation period to what extent the self-monitoring influenced family life, professional activity and their affective states: 57% answered that the self-observation had no impact on the family life aspect; 74% reported no impact on professional activity; 36% estimated a little influence. With respect to the effects on their affective/emotional state, 62% could not observe any effect, and 33% a little effect, and for the remaining 5% the effect was experienced as stronger.

Results on Validity—Some Preliminary Findings on Coping and Gender

For our method, we assume a strong *ecological validity* because, in contrast to laboratory or ques-

tionnaire methods, behavioural and mood data are gathered in the context and conditions under which they usually occur. The results concerning *social coping* indicate satisfactory validity. For analysing social coping responses we explored data on parent dyads where both partners report being in conflict with each other. Calculations have been made on the situation level, controlling for the frequency of events per dyad.

The data recorded with the instrument give information about the functionality of social coping responses (cf. Perrez and Schoebi, 2000). We found negative correlations between dysfunctional coping responses (e.g. to blame, to threaten) shown by women with their emotional state during or after an episode of conflict with their partner ($r = -0.55$; $p < 0.01$) and with their emotional state at the next observation time ($r = 0.60$; $p < 0.01$). Dysfunctional coping of men however was weakly positively correlated with their emotional state after conflicts with their partner ($r = 0.16$; $p < 0.05$). Moreover, the quality of one's coping responses seems to be a moderate predictor of the partner's emotional state after the conflict episode. We found negative correlations between dysfunctional coping responses and the partner's emotional state after the conflict episode ($r = -0.14$; $p < 0.05$ for men's coping; $r = -0.33$; $p < 0.01$ for women's coping).

Another way to measure the validity of the coping measures is to correlate measures on functional and dysfunctional social coping responses with indicators for the quality of social systems, namely indicators for dyadic and family functioning. Dyadic satisfaction was measured with the Relationship Assessment Scale (Hendrick, 1988; adapted and validated German version by Sander and Böcker, 1993), and family life satisfaction with a scale by Olson and Stewart (1990), adapted by Schneewind and Weiss (1999), which we moderately modified for language reasons. Results show constantly negative correlations of dysfunctional social coping responses with dyadic satisfaction, ranging from $r = -0.33$; $p < 0.05$ to $r = -0.42$; $p < 0.01$).

DISCUSSION

As many authors to this issue attest understanding the impact of family life on child development is an important research and clinical issue. To date though researchers have been reliant on laboratory methods or self-report. The computer-aided Family Self-Monitoring System (FASEM-C) offers a potentially powerful new research tool for studying family interactions. Before this can be more widely adopted and developed however key questions on reliability and validity need to be addressed and our work seeks to do that.

The evaluation of the computer-aided Family Self-Monitoring System (FASEM-C) had to consider several factors such as its acceptability by participants, its reactivity and the usual methodological criteria such as reliability and validity. The *acceptability* for those, who decide to participate, is surprisingly high. We counted only a few drop outs in both studies. In the first study with the first version of FASEM, only two of 99 Swiss families stopped their participation, another family had computer problems (Perrez *et al.*, 1998). In the second study three families cancelled their participation for personal reasons and one stopped because of computer problems. Other indicators of good acceptance are the direct answers to the questions related to the satisfaction with the method, and the objective parameters concerning the accuracy of the recording task. The satisfactory short latency time between prescribed and realized recording for a majority of the observations can be interpreted as an indication of compliance. This positive result may be an effect of the sampling selection and the motivating coaching by the project members. However, on-going work has proved the feasibility of using FASEM-C for normal families with adolescents, for case studies of families with patients with eating disorders and for individual self-monitoring of selected psychiatric outpatients.

The results on *reliability* are on the whole satisfactory. For some predominantly situation-specific variables, we found—in accordance with theoretical assumptions—lower odd−even coefficients; less situation-dependent variables gave higher coefficients. The local reliability analysis showed a satisfactory reliability for the 2-item scale of somatic state and a very high reliability for the emotional state scale. This indicates that the single items of the emotional state scale were used in a very consistent and stable pattern. For a revised version of the FASEM items could be dropped without much loss of information.

The results for *inter-rater reliability* concerning the parent's recordings are strongly dependent on the set criteria: for the mutual *presence* of the partners, agreement is distinctly lower than for agreement concerning the *location*. This difference is to be attributed to the subjective interpretation of

what 'presence of a family member or another person' actually means. We conclude from this result the necessity for a more precise instruction concerning this question.

A third problem concerns the question of *reactivity*, which could be crucial for this method. Analogous to memory distortions, which can reduce the validity of data from retrospective methods, the important question is how far the observed behaviour is changed by the observation method. Reactivity could prevent the collection of reliable and useful data. For the method developed by Pawlik and Buse (1982), Stern (1986) found that reactivity can be powerful when one isolated behaviour is observed rather than several, and when people find themselves in unfamiliar surroundings rather than in familiar settings. Our observation settings included the home, the professional situation and leisure time, and several behavioural types are simultaneously observed; therefore, the reactivity should be limited. In spite of this, the FASEM-C data did show some reactivity; the effects were small to medium, statistically significant and in the expected direction.

In the follow-up questionnaire most subjects answered that they did not notice any major influence of the self-monitoring on their behaviour. The future development of FASEM-C will take into account this result. The observation task will be devised as less time-consuming and less mentally demanding.

The relevance of FASEM-C as an assessment procedure for characteristics of the *interpersonal regulation of emotions* in families is supported by most of the results on validity. Inter-individual processes can be analysed with respect to both interactants' subjective view. The observed negative emotional consequences, contingent on dysfunctional social coping, are substantial within the person, especially for women, and to a lesser extent between the subjects. The dysfunctional social coping response of partner A in a conflict situation predicts the emotional state of partner B moderately, especially for women's coping. The self-monitoring parameters are furthermore—in accordance with the theoretical expectations—significantly correlated with indicators for the quality of dyadic and family functioning (for a more detailed discussion of the possibilities and problems in estimating the psychometric qualities, see Pihet, 2000; Buse and Pawlik, 1996; Wilhelm and Perrez, 2000).

With the application of the FASEM for the analysis of social conflict, we found some relevant information on social coping in the family environment. Subjects showed significantly more dysfunctional coping behaviour in the family setting than in other settings. These findings have been found in a study with FASEM I and could be confirmed with FASEM-C. There seems to exist something similar to 'display rules' for social coping in analogy to the display rules for emotion expression (cf. Perrez and Schoebi, 2000).

The *clinical importance* of this method of data gathering, is that it offers direct observations of behaviour, emotional states and cognitions and enables researchers to explore their interactions. By the use of FASEM, dysfunctional social patterns and their intrapersonal cognitive antecedents can be discovered. The utility is not restricted to the functional analysis, it also concerns the assessment of base-line data. This new procedure for the assessment of individual and social regulation of stress and emotion may be helpful in the future for planning and evaluating therapeutic interventions for individuals, couples and families. In general the future development of clinical assessment may benefit from the positive aspects inherent in the possibilities associated with modern palmtop computers. One special development, which extends the pure assessment function, focuses on ambulatory *expert systems*. They assist the patient to improve his/her control of covert and overt behaviour. Ager (1991) describes the new tutorial possibilities of microcomputers in the delivery of psychological interventions. In the context of anxious and eating disorders such systems already exist (Taylor *et al.*, 1991; Newman *et al.*, 1997, 1999). From these new possibilities we expect dynamic future developments for clinical, social and developmental psychology.

ACKNOWLEDGEMENTS

REFERENCES

Ager A. 1991. The role of microcomputers in clinical psychology. In *Microcomputers and Clinical Psychology*, Ager A, Bendall S (eds). John Wiley & Sons: Chichester.

Asendorpf JB, Wilpers S. 1999. Kontrolliertes interaktions-tagebuch zur Erfassung sozialer Interaktionen, Beziehungen und Persoenlichkeitseigenschaften. *Diagnostica* **45**: 82–94.

Blaney PH. 1986. Affect and memory: a review. *Psychological Bulletin* **99**: 229–246.

Bodenmann G. 1995. *Bewältigung von Stress in Partnerschaften. Der Einfluss von Belastungen auf die Qualität und Stabilität von Paarbeziehungen*. Freiburger Beiträge zur Familienforschung. Bd. 2. Universitätsverlag Fribourg Huber: Bern.

Bodenmann G, Perez M. 1991. Dyadisches Coping—eine systemische Betrachtungsweise der Belastungsbewältigung in Partnerschaften. *Zeitschrift für Familienforschung* **3**: 4–25.

Brandstätter M. 1983. Emotional responses to other persons in every day life situations. *Journal of Personality and Social Psychology* **45**: 871–883.

Brewin CR, Andrews B, Gottlib IH. 1993. Psychopathology and early experience: reappraisal of retrospective reports. *Psychological Bulletin* **113**: 82–98.

Buse L, Pawlik K. 1996. Ambulatory behavioral assessment and in-field performance testing. In *Ambulatory Assessment Computer-assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies*, Fahrenberg J, Myrtek M (eds). Hogrefe & Huber Publishers: Seattle, WA; 29–50.

Csikszentmihalyi M, Larson R. 1987. Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease* **175**: 526–536.

de Vries MW (ed.). 1992. *The Experience of Psychopathology. Investigating Mental Disorders in Their Natural Setting*. Cambridge University Press: Cambridge.

Epstein S. 1977. Traits are alive and well. In *Personality at the Crossroads: Current Issues in Interactional Psychology*, Magnusson, D, Endler NS (eds). Lawrence Erlbaum Associates Inc: Hillsdale, NJ.

Fahrenberg J. 1997. Ambulantes psychophysiologisches Monitoring in der Klinischen Psychologie. In *Impulse für die Klinische Psychologie*, Rockstruh B, Elbert T, Watzel H (eds). Hogrefe: Göttingen.

Fahrenberg J, Brügner G, Foerster F, Käppler C. 1999. Ambulatory assessment of diurnal changes with a hand-held computer: mood, attention, and morningness–eveningness. *Personality and Individual Differences* **26**: 641–656.

Fahrenberg J, Myrtek M (eds). 1996. *Ambulatory Assessment: Computer-assisted, Psychological and Psychophysiological Methods in Monitoring and Field Studies*. Hogrefe & Huber Publishers: Seattle, WA.

Hedeker D, Flay BR, Petraitis J. 1996. Estimating individual influences of behavioral intentions: an application of random-effects modeling to the theory of reasoned action. *Journal of Consulting and Clinical Psychology* **64**: 109–120.

Hendrick SS. 1988. A generic measure of relationship satisfaction. *Journal of Marriage and the Family* **50**: 93–98.

Käppler C, Brügner G, Fahrenberg J. 2000. Pocketcomputer-unterstütztes Assessment mit MONITOR: Befindlichkeit im Alltag, Methodenakzeptanz und die Replikation des Retrospektionseffektes. *Zeitschrift für Differentielle und Diagnostische Psychologie*, in press.

Larson R, Richards M. 1994. *Divergent Realities. The Emotional Lives of Mothers, Fathers, and Adolescents*. Basic Books: New York.

Lazarus RS. 1991. *Emotion and Adaptation*. Oxford University Press: New York.

Lienert GA, Raatz U. 1994. *Testaufbau und Testanalyse*. Autl. Beltz. PVU: Weinheim.

Longford NT. 1995. Random coefficient models. In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Arminge G, Clogg CC, Sobel ME (eds). Plenum Press: New York; 519–577.

Margraf J, Schneider S. 1990. *Panik, Angstanfälle, und ihre Behandlung*. 2nd überarb. Aufl. Springer: Berlin.

Mees U. 1977. Methodologische Probleme der Verhaltensbeobachtung in der natürlichen Umgebung: Zuverlässigkeit und generalisierbarkeit von Beobachtungsdaten. In *Verhaltensbeobachtung und Verhaltensmodifikation*, Mees U, Selg H (eds). Klett-Cotta: Stuttgart; S. 43–65.

Newman MG, Consoli A, Taylor CB. 1999. A palmtop computer program for the treatment of generalized anxiety disorder. *Behavior Modification* **23**: 597–619.

Newman MG, Kenardy J, Herman S, Taylor CB. 1997. comparison of palmtop-computer-assisted brief cognitive–behavioral treatment to cognitive–behavioral treatment for panic disorder. *Journal of Consulting and Clinical Psychology* **65**: 178–183.

Olson DH, Stewart KL. 1990. *Health and Stress Profile Questionnaire*. Minneapolis, Minnesota: Profiles on Health Systems.

Pawlik K, Buse L. 1982. Rechnergestützte Verhaltensregistrierung: Beschreibung und erste psychometrische Ueberprüfung einer neuen Erhebungsmethode. *Zeitschrift für Differentielle und Diagnostische Psychologie* **3**: 101–118.

Perez M. 1995. Wie kann 'Familienstress' definiert werden? In *Familie und Recht/Famille et Droit*, Gauch P, Schmid J, Steinauer PH, Tercier P, Werro F (eds). Universitätsverlag: Fribourg; S.519–528.

Perez M, Reicherts M. 1992. *Stress, Coping, and Health: A Situational–Bheavior Approach. Theory, Methods, Applications*. Hogrefe & Huber: Göttingen.

Perez M, Reicherts M. 1996. A computer-assisted self-monitoring procedure for assessing stress-related behavior under real life conditions. In *Ambulatory Assessment. Computer-assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies*, Fahrenberg J, Myrtek M (eds). Hogrefe & Huber: Göttingen; 51–71.

Perez M, Berger K, Wilhelm P. 1998. Die Erfassung von Belastungserleben und Belastungsverarbeitung in der Familie: Self-monitoring als neuer Ansatz. *Psychologie in Erziehung und Unterricht*.

Perez M, Schoebi D. 2000. Soziales Coping in der Selbst- und in der Fremdperspektive. In *Familie und Entwick-*

lung: Perspektiven der Familienpsychologie, Walper S, Pekrun R (eds). Hogrefe: Göttingen.

Perrez M, Wilhelm P. 2000. Control psychology under the control of questionnaires. The search for an alternative assessment procedure. In *Control of Human Behaviour, Mental Processes and Consciousness*, Perrig WJ, Grob A (eds). Lawrence Erlbaum: Mahwah, NJ: 245–261.

Perrez M, Wilhelm P, Berger R, Horner M, Law I, Schöbi D, Zbinden M. 1998. *Belastungserleben und Belastungsverarbeitung in Familien mit Adoleszenten. Forschungsbericht.* Nr. 133. Psychologisches Institut, Universität Freiburg: Freiburg, CH.

Pihet S. 2000. Etudes des qualités psychométriques de l'auto-observation systématique au milieu naturel. Dissertation, University, Faculty of Philosophy: Fribourg.

Reicherts M. 1999. *Règles Cognitivo-compotementales. Bases Theoriques et Méthodologiques*. Presse Universitaire: Fribourg.

Reicherts M, Perrez M. 1992. Adequate coping behavior: the behavior rule approach. In *Stress, Coping, and Health. A Situation–Behavior Approach*, Perrez M, Reicherts M (eds). Hogrefe & Huber Publishers: Seattle: 161–182.

Rosenthal R. 1987. *Judgement Studies. Design, Analysis and Meta-analysis*. Cambridge University Press: Cambridge.

Ross M. 1989. Relation of implicit theories to the construction of personal histories. *Psychological Review* **96**: 341–357.

Sander J, Böcker S. 1993. Die Deutsche Form der Relationship Assessment Scale (RAS): Eine kurze Skala zur Messung der Partnerschaftszufriedenheit. *Diagnostica* **39**: 55–62.

Scherer KR. 1988. Criteria for emotion-antecedent appraisal: a review. In *Cognitive Perspective on Emotion and Motivation*, Hamilton V, Bower GH, Frijda NH (eds). Nijkoff: Dodrecht; 89–126.

Schneewind KA, Weiss J. 1999. *Gesundheit- und Stress-Fragebogen*. Hans Huber: Bern.

Schumacher J. 2000. Das Ritual der Familienmahlzeit. Ein interkultureller Vergleich zwischen der Schweiz und Canada. Unpublished master-paper. Fribourg (CH): University Department of Psychology.

Stern E. 1986. *Reaktivitätseffekte in Untersuchungen zur Selbstprotokollierung des Verhaltens im Feld*. PhD Dissertation, Universität Hamburg. Lang: Frankfurt.

Taylor CB, Agras WS, Losch M, Plante TG, Burnett K. 1991. Improving the effectiveness of computer-assisted weight loss. *Behavior Therapy* **22**: 229–236.

Wilhelm P, Perrez M. 2000. Felddiagnostik. In *Psychodiagnostik psychischer Störungen* (2nd edn), Stiegliz RD, Baumann U (eds). Enke: Stuttgart; in press.