Contribution to the Measurement of Mode-Specific Effects in Long-Term Psychoanalytic Psychotherapy

Dorothea Huber and Guenther Klug

In psychotherapy research generally agreed upon standards have been developed for the assessment of outcome during the last few years. According to Lambert and Hill (1994) the state of the art requires a) assessment from different sources (self-reporting, trained observers, relevant others, therapist rating, institutional); b) application of different, non-reactive technologies (global ratings, specific symptom index, observer ratings, physiological measures, life records); c) atheoretic, pragmatic measures; d) bi-directional measures; e) multidimensional measures (intrapsychic, interpersonal, psychosocial); and f) follow-up measurements. Strupp, Horowitz and Lambert (1997) developed criteria for a core battery to measure outcome: a) clear and standardised procedures for administering and scoring the instrument; b) norms for patient and non-patient populations; c) demonstrated reliability; d) demonstrated validity; e) demonstrated sensitivity to change; f) demonstrated feasibility in clinical settings; g) systematic rater-training and calibration tapes available if administered by trained clinicians; h) atheoretic measures; i) multimodal measures; j) categorical and dimensional measures; and k) measurement before, during and after treatment.

Both approaches have to cope with the problem of grasping the mode-specific and therefore theory-bound effects of one specific therapeutic modality as well as applying theory-free, pragmatic measures not confined to the perspectives of one specific theory of change.

Schulte (1995) tried to elucidate the dilemma by distinguishing four levels within the general concept of disease: a) causes (biological, psychological, sociological); b) disease (pathological changes in the person); c) illness

(symptoms, complaints, medical findings); and d) consequences (sick-role and impairment of normal role behaviour). He argues that on level a) causes and level b) disease only school-specific, hence theory-bound measures are appropriate, while on level c) illness and level d) consequences theory-free measures are appropriate.

Reviewing the literature on the outcome of psychoanalysis and psychoanalytic psychotherapy, all authors (e.g. Roth and Fonagy, 1996; Vaughan, Marshall, MacKinnon, Vaughan, Mellman and Roose, 2000) agree that, although some evidence of the effectiveness of these treatments has been accumulated, there still is a regrettable lack of studies that meet the requirements of modern empirical research.

One of their main points of critique has been the absence of appropriate measures to encompass the more ambitious aims of psychoanalysis and psychoanalytic psychotherapy. The outcome studies of the 70¥s and 80¥s mostly applied global assessments of therapeutic benefit that were not able to capture the specific effects of psychoanalysis and psychoanalytic psychotherapy. The outcome measures, mostly expert ratings of interviews with the patient and the treating analyst, did not meet modern research standards such as, for example, reliability, validity and sensitivity to change and could only cover dimensions of change in a very restricted way only, using vague categories. Bachrach, Galatzer-Levy, Skolnikoff and Waldron (1991) therefore concluded in their critical survey of outcome studies: ì...the research methods, especially of the clinical-quantitative studies, reflect the state-of-the-art of the 1950¥s and 1960¥s more than currently available methods.î Very much in the same vein is Luborsky, Diguer, Luborsky, Singer, Dickter and Schmidt¥s (1993) critique of unsuitable outcome measures because they ìdo not make an adequate distinction between short-term and long-lasting improvement, nor do they make a distinction between the parallel related changes referred to as non-structural and structural changeî.

Another decisive point is that psychotherapy research shifted from the investigation of the outcome of psychotherapy to the connection between process and outcome. This brings about a new challenge for the conceptualisation of outcome measures because the link between outcome

and process has to be observed attentively. As a principle that tries to meet the needs of process-outcome research, Strupp, Schacht and Henry (1988) proposed the ìProblem-Treatment-Outcome Congruenceî (P-T-O Congruence method). Dahl (1988) elaborated on it thus: ìthis principle (says) that the description and representation, theoretically and operationally, of a patientís conflicts, of the patientís treatment, and of the assessment of the outcome, must be congruent, which is to say, must be represented in comparable, if not identical terms.î Therefore an outcome measure is needed that both grasps the specific effects of psychoanalysis and psychoanalytic psychotherapy, and is able to link the outcome, on a conceptual basis of a commonly agreed upon theory of change to the process, making ìpsychoanalytic changeî more conceivable that way.

Psychoanalysis and psychoanalytic psychotherapy used to understand Schulte¥s level b) disease (pathological changes in the person) as ìpsychic structureî and coined the term ìstructural changeî for changes on that level. Pulver (1991) defined psychic structure ìas any organisation of mental contents and processes which, in a systematic way, carried out the various tasks of the psycheî. According to Piaget (1970) its basic principles are wholeness, transformation and self-regulation. The ìOperationalized Psychodynamic Diagnosticsî group (OPD-Task Force, 2001) characterised ìpsychic structureî by six features: a) capacity for self-reflection; b) capacity for self-management; c) capacity for defense; d) capacity for object perception; e) capacity for communication and f) capacity for binding.

Structural change as ìthe generally accepted goal of psychoanalysisî (Moore & Fine, 1990) and also to some degree of psychoanalytic psychotherapy (e.g. Kernberg, 1991; Wallerstein, 1986), is an explanatory construct that tries to capture the specific effects of psychoanalysis and psychoanalytic psychotherapy relating them to the concept of psychic structure and their modifications by psychoanalytic treatment. Structural change signifies a type of change, beyond symptoms and manifest behaviour, rooted in the matrix of both. Kernberg (1991) defines it as a ìsignificant modification in the unconscious intrapsychic conflicts underlying symptom formation. Change in the underlying unconscious intrapsychic structures is usually revealed in shifts in the equilibrium of ego, superego, and id, with a significant expansion of the system ego and a corresponding reduction of the pressures

of the unconscious superego and id î.

The application of such global ego-psychological concepts in follow-up studies revealed, however, that they had to be more clearly and differentially operationalised and had to be assessed by experts in order to grasp the more subtle effects of psychoanalytic treatments (Wallerstein, 1986). As to the technology (Lambert & Hill, 1994) of the instrument and above all its non-reactivity we would like to point out that only expert clinicians will be able to assess adequately the transference- related cognitive distortions that regularly take place in follow-up interviews (Pfeffer, 1959).

In summary we conclude that an empirically-founded outcome measure that intends to grasp changes beyond symptoms and behaviour in a process-outcome context, must a) be based on expert judgement; b) be able to grasp different functions of psychic structure and their modifications; c) be able to rule out the influence of symptoms on psychic functioning; d) be able to assess the influence of transference-related cognitive distortions in the follow-up interviews; e) be based on concepts agreed upon by adherents of different psychoanalytic schools; f) have satisfactory psychometric qualities (reliability, validity, sensitivity to change, clinical significance), and g) must meet the requirements of the P-T-O principle.
We consider the Scales of Psychological Capacities, SPC (Wallerstein, 1991; Wallerstein, unpublished manuscript) as the measure that complies best with these standards and that has sufficiently proven its psychometric qualities (DeWitt, Hartley, Rosenberg, Zilberg, & Wallerstein, 1991; Dewitt, Milbrath, & Wallerstein, 1999).

Therefore we decided to apply the Scales of Psychological Capacities (SPC) to a comparative process-outcome study of psychoanalysis and psychoanalytic psychotherapies, the Munich Psychotherapy Study (MPS), which we present next.

The Munich Psychotherapy Study, MPS (Huber, Klug, & von Rad, 1997; Huber, & Klug, 1999; Huber, Klug, & von Rad, 2000) is a process-outcome comparison study in progress that tries to answer two questions:-

1. Are there any differences in effectiveness between psychoanalysis and

psychodynamic psychotherapy? And if so: are those changes brought about by psychoanalysis based on ìstructural changeî and as a result of this, are they more profound and more stable than those brought about by psychodynamic psychotherapy ?

2. Are there any links between therapeutic process and outcome? And if so: what are they?

In order to answer the first research question a randomised controlled design was chosen to compare the two experimental groups:

1. A group of patients treated with psychoanalysis (PA) taking place 3 times a week in a recumbent position with an average duration of 240 - 300 hours.
2. A group of patients treated with psychodynamic psychotherapy (PT) place once a week in a seated position with an average duration of 80 - 120 hours.

As already stated the effectiveness of the two treatments can only be evaluated correctly if the patients are assigned at random to the two experimental groups. Because of the relatively small number of patients in each group (N=30) a strict allocation at random would lead to an uneven distribution of important patient variables, which was one of the main issues of the NIMH depression study (Elkin et al., 1989); we therefore decided to stratify the patients with regard to severity of symptoms and age. We set high value on the therapies being assigned at random and not the therapists so as not to interfere with the important, individual patient-therapist match.

Each patient presenting at the outpatient department of the Institute of Psychosomatic Medicine, Psychotherapy and Medical Psychology of the Technical University of Munich who met the inclusion criteria received an extensive clinical intake interview that was audio-recorded. Based on this recorded interview a board of 3 experienced psychoanalysts (so called ìindication boardî) decided if the patient could be randomly assigned to one of the two experimental groups. This decision process was documented as precisely as possible.

The inclusion criteria were as follows: between 25 and 45 years of age; ICD-10 diagnosis of depressive episode or recurrent depressive disorder; Beck-Depression-Inventory, BDI (Beck, Ward, Mendelson, Mock, & Erbaugh,

1961) > 16; previous psychotherapy to be finished at least 2 years before entering the study; not on anti-depressant medication; living in Munich or nearby; adequate German language skills.

The 10 therapists who participate in the study are experienced psychoanalysts and psychotherapists in private practice and have been working with patients for at least 5 years. They were trained at an approved institute and graduated there. They only apply therapies in which they are experienced and are not obliged to apply a therapeutic modality he or she does not consider as suitable for a particular patient.

The data come from three different sources of observation: the patient, the therapist and the researcher (ìexternal investigatorî). The test battery of outcome measures is adapted from the core battery suggested by the Society for Psychotherapy Research (SPR), published by Grawe and Braun (1994) and chosen to be comparable with other ongoing studies.

As already stated a major concern of the study is to measure not only symptoms and behaviour, but also mode-specific effects and therefore, special instruments to measure structural change and individual therapeutic goals are administered. Structural change is measured with the Scales of Psychological Capacities (SPC); individual goals are assessed by means of the Goal Attainment Scaling method, developed by Kiresuk and Sherman in 1968, which in the Heidelberg-Study of von Rad, Senf and Br‰utigam (1998) yielded an interesting discrimination between psychoanalysis and psychotherapy.

The procedural plan (schedule) of the study is shown in table 1:

Table 1

Procedural Plan of the MPS (see text for abbreviations of the instruments)

Pre-treatment Measurement External investigator 1 and patient: intake-interview, ICD-10 and DSM-IV diagnosis, GAF, BADO, BDI (>16), BSS, HAMD Board of three experienced analysts: decision on patientís inclusion in the study. External investigator 1 and patient: SPC- interview; informed consent

Patient: self-rating questionnaires: BDI, SCL-90-R, IIP, FKBS, INTREX, SOZU, BADO, FLZ, FPI-R. External investigator 1 and patient: assessment of individual goals (goal attainment scaling, GAS)

Referral to therapist: Therapist: documentation of diagnosis, psychodynamic hypothesis, level of personality organisation, treatment goals, prognosis, HAQ-T Process Measurement, Audio-recording of every session

Patient: self-rating questionnaires: BDI, SCL-90-R, IIP, GAS and HAQ-P every 6 months

Therapist: therapy accompanying card to be filled out after every session; periodical process rating scale with HAQ-T every 6 months Post-treatment

Measurement: External investigator 2 (ìblindî for applied therapy) and patient: post-treatment interview, SPC-interview, life-events checklist, ICD-10 and DSM-IV diagnosis, GAF, BSS, HAMD, BADO

Patient: self-rating questionnaires: BDI, SCL-90-R, IIP, FKBS, INTREX, SOZU, BADO, FLZ, FPI-R, GAS, VEV, HAQ-P

Therapist: periodical process rating scale and HAQ-T, assessment of termination of treatment. Follow-up Measurement

(annually) External investigator 2 and patient: follow-up interview, SPC-interview, life-events checklist, ICD-10 diagnosis, BSS, GAF, HAMD, BADO

Patient: self-rating questionnaires: BDI, SCL-90-R, IIP, FKBS, INTREX, SOZU, BADO, FLZ, FPI-R, GAS, VEV

At the end of the intake-interview with ICD-10 and DSM-IV diagnosis the external investigator completes the Global Assessment of Functioning Scale (GAF, DSM-IV axis 5; American Psychiatric Association, 1994), the Symptom Severity Score (BSS; Schepank, 1995), the Hamilton Rating Scale for Depression (HRSD, Hamilton, 1960) and the Basic Documentation of the German College of Psychosomatic Medicine (BADO, our version described by Huber, Henrich, & von Rad, 2000), including the rating of the psychic structure of the patient (axis 4: Structure of the Operationalized Psychodynamic Diagnostics, OPD; OPD-Task Force, 2001). After a positive decision by the ìindication boardî and the obtaining of ìinformed consentî by the patient the external investigator interviews the patient with a semi-structured SPC-interview to obtain the appropriate information to score the SPC-scales. In the third pre-treatment session the external investigator and the patient assess together the individual goals the patient wishes to achieve during the therapy. The patient is assigned to one of the experimental groups after this intake procedure, so that the external investigator is ìblindî to

therapeutic modality during the pre-treatment assessment.

Before the treatment starts the patient completes the following self-rating questionnaires: Symptom-Check-List (SCL-90-R, Derogatis, Lipman, & Covi, 1975), Beck-Depression-Inventory (BDI, Beck et al., 1961), Inventory of Interpersonal Problems, short version (IIP-C, Horowitz, Rosenberg, Bauer, Ureno, & Villasenor, 1988), Introject questionnaire (INTREX, Benjamin, 1974), Questionnaire for Coping Strategies (FKBS, Hentschel, 1998), Freiburg Personality Inventory, revised version (FPI-R, Fahrenberg, Hampel, & Selg, 1989), Life Satisfaction Questionnaire (FLZ, Huber, Henrich, & Herschbach, 1988), Basic Documentation of the German College of Psychosomatic Medicine (BADO, our version described by Huber, et al., 2000), Social Support Questionnaire, short version (F-SOZU-K-22, Sommer & Fydrich, 1991).
The therapist completes the Helping Alliance Questionnaire (HAQ-T Alexander & Luborsky, 1986) and a documentation form with psychodynamic diagnoses, main defences, level of personality organisation, motivation, main psychodynamic hypotheses, treatment goals and prognosis.
During the ongoing therapeutic process neither the patient nor the therapist is contacted personally so as not to interfere with the process too much; of course research itself inevitably influences the process. The process measures are sent to the patient and psychotherapist by mail every 6 month. Each therapy session is audio-recorded.

Measurement points for the outcome measures are at pre-treatment, at post-treatment and at follow-up each year after end of treatment. The external investigator 2 at post-treatment and follow-up is not the same as at pre-treatment and is ìblindî to the therapeutic modality ñ although the possibility of remaining ìblindî during a clinical interview has been questioned in the literature (Luborsky et al., 1999). At post-treatment and follow-up the pre-treatment instruments are applied again, along with a retrospective life-event checklist and a self-rating questionnaire: Change in Experiencing and Behaviour, VEV (Zielke & Kopf-Mehnert, 1978).

There follows a more detailed description of the Scales of Psychological Capacities (SPC), the measure of mode-specific effects of psychoanalysis and psychoanalytic psychotherapy (ìstructural changeî) applied in our study.

The Scales of Psychological Capacities, SPC are an expert-rating measure that, whilst theoretically-informed but not theory-specific, evaluates the level of psychic structure. They have been developed from the research methodology of the Psychotherapy Research Project (PRP) of the Menninger Foundation (Wallerstein, 1986) and try to operationalise the concept ìpsychic structureî and ìstructural changeî as independently as possible of the differing theoretical perspectives in psychoanalysis so that they are able to assess reliably the specific changes after psychoanalyses and psychoanalytic psychotherapies. Bound to an empirical research strategy these psychological capacities are designed to be as low-level (experience-near) constructs as possible, and readily inferable from observable behaviours and conscious states of mind so that underlying intrapsychic structures and their changes after treatment can be reliably captured.

The Psychological Capacities Scales consist of 17 dimensions, 14 of which are divided into 2 subdimensions and 2 of which into 3 subdimensions; 1 dimension is not divided. The assessment of all 36 subdimensions is based on a tape-recorded, one-hour clinical intake interview together with a one- to two-hour semi-structured SPC-interview with probe questions, developed by the test author and his group. The material gained this way is scored for each subdimension on a 7-point scale from 0 for ìnormalî or fully adaptive functioning to 3 for functioning seriously and obviously disturbed, with half points in between. The dimensions are constructed one subdimension to be designated for different degrees of inhibited and another for different degrees of exaggerated functioning. Both directions have to be assessed and both subdimensions can be scored simultaneously. The rating procedure requires an extensive manual with a detailed description of each subdimension together with one or more clinical vignettes to anchor each scale point.

Although inter-rater reliability (DeWitt et al., 1999), content validity (DeWitt et al., 1991), and convergent validity (DeWitt et al., 1999) of the SPC have already been examined there is a lack of studies that prove their feasibility for German research projects as well as a lack of discriminant validity studies.

METHOD

Any outcome measure needs to have sufficient evidence that reliability and validity are warranted as the basic psychometric properties. Developers of outcome measures must strike a balance between these two psychometric qualities to be able to offer instruments to researchers that meet their demands.
Comparing the SPC with instruments that measure interpersonal functioning and personality structure can evaluate convergent validity. Discriminant validity is evaluated in the present study by comparing the SPC with instruments measuring symptomatology.

We expect a zero or only weak correlation between the Scales of Psychological Capacities (SPC) and the construct-distant measures and a moderate but not very high correlation between the Scales of Psychological Capacities (SPC) and the construct-near measures.

Validity study I

As this pre-study has already been published (Huber & Klug, 1997; Huber, Klug, & von Rad, 2000), we provide only a short description of the procedure. The sample consisted of a homogenous group of 41 depressed patients of the MPS study (see above). Two judges (DH & GK) rated the 41 SPC-interviews from the audiotapes. Both are psychoanalysts with a completed analytic training and many years of professional experience. They trained themselves with the SPC-manual and rated the first three interviews together. Afterwards they had recalibration sessions after every fifth rating. The construct-distant instruments were the SCL-90-R, a self-rating symptom inventory constructed to assess the psychological and symptom status of psychiatric patients on 9 scales and a global severity index and the BSS, and the GAF; the BSS is an observer rating scale that evaluates the impact of psychic illness on three dimensions: physical, psychological and social. The GAF (recent and highest level of functioning in the last year) is rated by the external investigator as well and is an internationally used scale.

As a construct-near measure we used the Inventory of Interpersonal Problems (IIP) in its self-rating form, and additionally in an observer-rating form (Horowitz, personal communication) in order not to contaminate the results by changing the source of observation for the only measure used. The IIP is

an internationally established instrument for assessment of interpersonal problems and concerns.

Inter-rater reliability study and validity study II

We performed a second construct validity study of the SPC as a replication of the above presented convergent and discriminant validity studies to further examine validity three years later. This study consisted of another sample of the MPS, a different interviewer, different raters who were trained according to the formal method (Mercer & Loesch, 1979), and of additional construct-near and construct-distant instruments. As the validity of any diagnostic instrument presupposes a reasonable degree of inter-rater reliability, we started this time with an inter-rater reliability study. Both studies are not yet published and therefore will be presented in more detail here.

The sample consisted of a homogenous group of 47 depressed patients, between 25 and 45 years. The diagnosis was made by an experienced clinician (psychiatrist and psychotherapist) using the IDCL classification schema for ICD 10 and DSM IV diagnosis (Hiller, Zaudig, & Mombour, 1995) after having discussed the cases with two other experts. According to the intake criteria all patients received a diagnosis of a depressive disorder. A description of the socio-demographic data and the ICD 10 diagnosis is given in table 2.

Table 2
Description of the sample (N = 47)

age (mean, sd) 34,6 (6,5)
Romantic relationship 60%
gender children

employment
  full time 63%
  part time 11%
  unemployed 4%
  other 22%

female 66%
male 34%
marital status

single 66%
married 19%
divorced 15%

1st diagnosis (ICD-10)
F 32.1 depressive episode ñ moderate 43%
F 32.2 depressive episode ñ severe 17%
F 33.1 recurrent depressive disorder ñ moderate 23%
F 33.2 recurrent depressive disorder - severe 17%

2nd diagnosis (ICD-10)
F 34.1 dysthymia 47%
other 8%
no 45%

The 38 patients in the inter-rater reliability study were a subsample of the above described sample.

Procedure: The authors of this paper have attended a rater-training according to the formal method (Mercer & Loesch, 1979) with the PRP-II group in San Francisco and have afterwards trained three German raters. After every 5th patient there was a recalibration session where the three judges and the two trainers met to correct for judges¥ drift.

The two raters, who came out with the best reliability scores, rated the scales for the validity study II. If they disagreed more than one scale-point, a senior rater (one of the authors) rated the scale again ñ a method recommended by Jones, Cumming and Horowitz (1988).

Instruments: An instrument already used in the pre-study (validity study I) for construct-validity of the SPC, the IIP, short form (IIP-C) was used again. Additionally to the pre-study we compared the SPC with the FKBS measuring 5 defense mechanisms / coping strategies. A personality questionnaire (FPI-R) with 12 scales was added.

Besides these self-rating questionnaires, the external investigator rates the psychic structure of the patient (OPD, axis 4) on a 4-point scale (good, moderate, low integrated, disintegrated).
We again assessed discriminant validity with the SCL-90-R. Additionally the

Beck Depression Inventory (BDI) was used for assessment of the severity of depression.

Statistics: The inter-rater reliability between the three raters was calculated by means of Intra Class Correlation Coefficient (ICC, Shrout & Fleiss, 1979) for all subdimensions separately.
Correlations between the SPC and the tests for discriminant and convergent validity were calculated with the Pearson Correlation Coefficient. Because of the large amount of correlations computed, we decided to interpret only findings on a 0.1% level of significance.

RESULTS

Validity study I: At first content validity was examined by prospectively (that means before the respective subdimensions were scored on empirical material) assessing, from a clinical point of view, those subdimensions which would be expected to be most highly scored. This clinically expected ìprototypicî profile of depressive patients was compared with the empirically found profile. With one exception, the empirically found mean profile of depressive patients was correctly predicted by the clinicians, very clearly demonstrating a high consensus between clinical judgement and SPC ratings.

The results of the discriminative and convergent construct validity study showed that there were no significant correlations between the SPC and symptoms measured by the SCL-90-R, the BSS and the GAF. There were 10 medium correlations (rr: .49- .64) between the SPC and the IIP scales, when IIP was rated by the patient (self-rating). There were 20 significant (?) correlations between SPC and IIP scales when IIP was rated by the external investigator (judge rating). This can be readily understood, as it is well known, that there is a higher correlation between data from the same source of observation than between the content of the scales. These results confirm the hypothesis that the SPC measure is relatively independent of current symptoms and measures something similar but not identical to interpersonal functioning (Huber, Klug, & von Rad, 2000).

Inter-rater reliability study: Because the judges could not discriminate reliably enough between the subdimensions Drudgery and Apathy, these two subdimensions were condensed into one which we called Apathy, thus reducing the number of subdimensions to 35.

The mean ICC was 0.82 within a range from 0.54 to 0.89. Using as a standard cut-off score a correlation level of .70, according to the recommendations of Lambert & Hill (1994) for scales that do not have a low level of inference, only 4 of the 35 subdimensions had reliabilities below .70. All of the 35 subdimensions reached Cohen¥s cut-off point of .50 (Cohen, 1988).These remarkable results for inter-rater reliability allowed the continuation of the psychometric investigation and we conducted the extended validity replication study with a new sample, different interviewers and different raters as well as some additional instruments.

Validity study II: Figure 1 shows the mean profile of our group of depressed patients on the 35 SPC subdimensions. This profile proves construct validity of the SPC because it operates in the expected way for depressed patients. The highest mean is for the subdimension Self-Depreciation, followed by Overinvolvement in Relationship, Internalisation, Surrender of Self, and Pessimism (all means are above 1.5).

Insert figure 1 here

Discriminant validity study: The data clearly show no significant correlations between either the
SCL-90-R scales, the BDI, the Impairment-Severity Scores or the two GAF scales and the SPC subdimensions plus the SPC total score (see table 3).

Table 3

Significant correlations (between the 35 SPC subdimensions and the construct-distant / construct-near instruments; $p < 0.001$)

SCL (10 scales)
BDI (1 scale)
BSS (4 scales)

GAF (2 scales)
construct-distant instruments
(r†: .01 - .38)
(r†: .02 - .41)
(r†: .01 - .37)
(r†: .01 - .32)
FPI (12 scales)
IIP (9 scales)
FKBS (5 scales)
OPD (1 scale)
construct-near instruments
12 (r: .46 - .62)
16 (r: .47 - .59)
1 (r = .52)
2 (r = .48 / .49)

Convergent validity study:
As is shown in table 3 there are 12 significant correlations between the 12 FPI scales and the SPC (rr: .46 - .62, $p < 0.001$), and 16 correlations between the IIP scales and the SPC (rr: .47 - .59, $p < 0.001$). Out of the 5 FKBS scales there was only one significant correlation with the scale Turning Against Self (rr = .52, $p < 0.001$). All other defence mechanisms did not correlate. The OPD rating correlated significantly with the SPC total score and the subdimension Coherence (rr = .48 / .49, $p < 0.001$)

CONCLUSIONS

We consider the results of the inter-rater reliability study to be highly satisfactory especially when taking into account that homogenous samples tend to show considerably lower inter-rater reliabilities and that reliability is normally higher when instruments are applied by their developers (Zimmermann, 1994). Seen from an economic point of view, our reliability study suggests that raters do not need to have psychoanalytic training and are doing sufficiently well with a medium range rater training.
The results of the discriminant validity studies clearly reveal that there is no correlation between the SPC and the construct-distant measures, thus confirming our first assumption that the SPC measures beyond

symptomatology.
The convergent validity studies with their considerable correlations between the SPC and the construct-near instruments offer convincing evidence that the SPC measures something similar, but not identical to interpersonal functioning and personality, thus confirming our second assumption.

We do not want to go into detail by interpreting single correlations, but nevertheless want to state that all inter-scale relationships were conceptually consistent and meaningful. We would like to point out that the significant correlation between the SPC and the scale Turning Against Self of the FKBS can be interpreted as the Ñpathognomonic introjectionì (Fenichel, 1945) of the depressive patient. This finding can be understood as another test of construct validity. The absence of any other correlations between the SPC and the remaining four scales of the FKBS suggests that the SPC does not measure just defence mechanisms which is in accordance with our working model of ìstructural changeî (Kernberg, 1991, Wallerstein, 1991).

We conclude that from our psychometric studies that the SPC appears to be a reliable and valid instrument. The findings of validity study I were replicated in validity study II, although the interviewers and raters differed considerably between study I and study II in clinical experience, therapeutic education and rater-training.

Although there is still much work to be done to establish other psychometric qualities, above all sensitivity to change, there is already substantial evidence that researchers have available an instrument that measures psychic structure, which is a first step towards the measurement of structural change, the mode-specific effect of psychoanalysis and psychoanalytic psychotherapy.

REFERENCES

Alexander L. B., & Luborsky L. (1986). The Penn Helping Alliance Scales. In: L. S. Greenberg, W. Pinsof (Eds.), The psychotherapeutic process: a research handbook. New York: Guilford Press, 325-366.

American Psychiatric Association: Diagnostic and Statistical Manual of

Mental Disorders. Fourth Edition, Washington DC: APA, 1994.

Bachrach, H. M., Galatzer-Levy, R., Skolnikoff, A., & Waldron, S. (1991). On the efficacy of psychoanalysis. Journal of the American Psychoanalytic Association, 39: 871 ñ 916.

Beck, A. T., Ward, C. H., Mendelsohn, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4: 561-571.

Benjamin, L.S. (1974). Structural Analysis of Social Behaviour. Psychological Review, 81: 392-425.

Cohen, J. (1988). Statistical Power Analysis for Behavioral Sciences. Hillsdale, New York: Lawrence Erlbaum Assoc.

Dahl, H. (1988). Introduction. In H. Dahl, H. Kaechele, & H. Thomae (Eds.), Psychoanalytic Process Research Strategies. Berlin: Springer.

Derogatis, L. R., Lipman, R.S., & Covi, L. (1973). SCL 90. An outpatient psychiatric rating scale. Psychopharmacology Bulletin 9: 13-28.

DeWitt, K. N., Hartley, D., Rosenberg, S. E., Zilberg, N. J., & Wallerstein, R. S. (1991). Scales of psychological capacities: Development of an assessment approach. Psychoanalysis and Contemporary Thought, 14: 334 - 343.

DeWitt, K. N., Milbrath, C., & Wallerstein, R. S. (1999). Scales of Psychological Capacities: Support for a measure of structural change. Psychoanalysis and Contemporary Thought, 22: 453-480.

Elkin, I., Shea, M.T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S. J., & Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. Archives of General Psychiatry, 46: 971-982.

Fahrenberg, J., Selg, H., & Hampel, P. (1989). Das Freiburger Persˆnlichkeits-Inventar (FPI-R). Gˆttingen: Hogrefe.

Fenichel, O. (1945). The Psychoanalytic Theory of Neurosis. New York: Norton.

Grawe, K., & Braun, U. (1994). Qualit‰tskontrolle in der Psychotherapiepraxis. Zeitschrift f¸r Klinische Psychologie, 23: 242-267.

Hamilton, M. (1960). A rating scale for depression. Journal of Neurology, Neurosurgery and Psychiatry, 23: 56-62.

Hentschel, U. (1998). Fragebogen zu Konfliktbew‰ltigungsstrategien. Gˆttingen: Hogrefe.

Hiller, W., Zaudig, M., & Mombour, W. (1995). Internationale Diagnosen-Checkliste f¸r ICD 10 und DSM IV (IDCL). Gˆttingen: Hogrefe.

Horowitz, L. M., Rosenberg, S. E., Bauer, B. A., & Ureno, G., & Villasenor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. Journal of Clinical and Consulting Psychology, 56: 885-892.

Huber, D., Henrich, G., & Herschbach, P. (1988). Measuring the quality of life: A comparison between physically and mentally chronically ill patients and healthy persons. Pharmacopsychiatry 21: 453-455.

Huber, D., & Klug, G. (1997). How to measure structural change. Paper presented at the 28th annual meeting of the Society for Psychotherapy Research, Geilo, 25.-29.6.1997.

Huber, D., Klug, G., & von Rad, M. (1997). M¸nchner Psychotherapie-Studie (MPS). In: M. Leuzinger-Bohleber & U. Stuhr (Eds.), Psychoanalysen im R¸ckblick. Gieflen: Psychosozial.

Huber, D., & Klug, G. (1999). The Munich Psychotherapy Study (MPS) ñ Comparing the effects of psychoanalysis and psychotherapy. In P. Fonagy, H. K‰chele, R. Krause, E. Jones, & E. Perron, (Eds.), An Open Door

Review of Outcome Studies in Psychoanalyis. Report prepared by the Research Committee of the IPA at the request of the President. International Psychoanalytic Association, London: http://www.ipa.org.uk.

Huber, D., Klug, G., & von Rad, M. (2000). Die M¸nchner Psychotherapiestudie. Ein Vergleich zwischen Psychoanalysen und psychodynamischen Psychotherapien unter besonderer Ber¸cksichtigung therapiespezifischer Ergebnisse. In U. Stuhr, M. Leuzinger-Bohleber, & M. Beutel (Eds.), Psychoanalytische Langzeittherapien. Stuttgart: Kohlhammer.

Huber, D., Henrich, G., & von Rad, M. (2000). ‹ber den Nutzen von Beratungsgespr‰chen an einer psychosomatisch-psychotherapeutischen Ambulanz. Psychotherapie Psychosomatik Medizinische Psychologie, 50: 147-156.

Jones, E. E., Cumming, J. D., & Horowitz, M. J. (1988). Another look at the nonspecific hypothesis of therapeutic effectiveness. Journal of Consulting and Clinical Psychology, 56: 48-55.

Kernberg, O. F. (1991). Psychic structure and structural change: An ego psychology-object relations theory viewpoint. In T. Shapiro (Ed.), The Concept of Structure. Madison, Connecticut: International Universities Press, Inc.

Kiresuk, T. J., & Sherman,R. E. (1968). GAS: a general method for evaluating comprehensive community mental health programs. Community Mental Health Journal, 4: 443-453.

Lambert, M. J., & Hill, C. E. (1994). Assessing Psychotherapy Outcomes and Processes. In A. E. Bergin & S. L. Garfield (Eds.), Handbook of Psychotherapy and Behaviour Change. New York: J. Wiley & Sons.

Luborsky, L., Diguer, L., Luborsy, E., Singer, B., Dickter, D., & Schmidt, K. A. (1993). The Efficacy of Dynamic Psychotherapies: Is it true that Ñ Everyone has won and all must have prizes?ì In N. E. Miller, L. Luborsky, J. P. Barber, & J. P. Docherty (Eds.), Psychodynamic Treatment Research. New York: Basic Books.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., Halperin, G., Bishop, M., Berman, J. S., & Schweizer, E. (1999). The researcher¥s own therapy allegiances: A Ñwild cardì in comparisons of treatment efficacy. Clinical Psychology: Science and Practice, 6: 96-132.

Mercer, R. C., & Loesch, L. C. (1979). Audiotape ratings: Comments and guidelines. Psychotherapy: Theory, Research and Practice, 16: 79 - 85.

Moore, B. E., & Fine, B. D. (1990). Psychoanalytic Terms & Concepts. New Haven: Yale University Press.

OPD-Task Force (Eds.) (2001) Operationalized Psychodynamic Diagnostics. Foundations and Manual. Seattle, Toronto, Gˆttingen, Hogrefe & Huber.

Pfeffer, A. Z. (1959). A procedure for evaluating the results of psychoanalysis: a preliminary report. Journal of the American Psychoanalytic Association, 7: 418 - 444.

Piaget, J. (1970). Structuralism. New York: Basic Books.

Pulver, S. E. (1991). Psychic structure, function, process, and content: toward a definition. In T. Shapiro (Ed.), The Concept of Structure. Madison, Connecticut: International Universities Press, Inc.

Rad, M. von, Senf,W., & Br‰utigam, W. (1998) Psychotherapie und Psychoanalyse in der Krankenversorgung. Ergebnisse des Heidelberger Katamnese-Projektes. Psychotherapie Psychosomatik Medizinische Psychologie, 48: 88 - 100.

Roth, A., & Fonagy, P. (1996). What works for whom? New York: Guilford Press.

Schepank, H. (1995). Der Beeintr‰chtigungs-Schwere-Score (BSS). Gˆttingen: Beltz.

Schulte, D. (1995). How treatment success could be assessed. Psychotherapy

Research, 5: 281-296.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater-reliability. Psychological Bulletin, 86: 420 - 428.

Sommer, G., & Fydrich, T. (1991). Entwicklung und ‹berpr¸fung eines Fragebogens zur Sozialen Unterst¸tzung (F-SOZU). Diagnostica, 37: 160-178.

Strupp, H. H., Schacht, T. E., & Henry, W. P. (1988). Problem ñ Treatment ñ Outcome Congruence: A principle whose time has come. In H. Dahl, H. Kaechele, & H. Thomae (Eds.), Psychoanalytic Process Research Strategies. Berlin: Springer.

Strupp, H. H., Horowitz, L. M., & Lambert, M. J. (1997). Measuring Patient Changes. Washington DC: American Psychological Association.

Vaughan, S. C., Marshall, R. D., MacKinnon, R. A., Vaughan, R., Mellman, L., & Roose, S. P. (2000). Can we do psychoanalytic outcome research? A feasibility study. International Journal of Psychoanalysis, 81: 513-527.

Wallerstein, R. S. (1986). Forty-two Lives in Treatment. New York: Guilford Press.

.

Wallerstein, R. S. (1991). Assessment of Structural Change in Psychoanalytic Therapy and Research. In: Shapiro, T. (Ed.), The Concept of Structure in Psychoanalysis. Madison: International Universities Press.

Wallerstein, R. S. (O. J.). The Scales of Psychological Capacities. Version 1. unpublished manuscript.

Zielke, M., & Kopf-Mehnert, C. (1978). Ver‰nderungsfragebogen des Erlebens und Verhaltens, VEV. Weinheim: Beltz.

Zimmermann, M. (1994). Diagnosing Personality Disorders. Archives of General Psychiatry, 51: 225-245.

Acknowledgements:

The rater-trainig of the authors (DH and GK) with the PRP-II group in San Francisco at the California Pacific Medical Center has been supported by the German Research Foundation DFG (444 USA 111/3/98).

The inter-rater reliability and the validity study with the SPC are funded by a grant from the IPA.