Studies of psychotherapy with implications for psychoanalysis

Reanalysis of the NIMH Multicentre Trial of Affective Disorders

> Blatt, S. J., Quinlan, D. M., Pilkonis, P. A., & Shea, M. T. (1995). Impact of perfectionism and need for approval on the brief treatment of depression: The National Institute of Mental Health Treatment of Depression Collaborative Research Program revisited. *Journal of Consulting and Clinical Psychology, 63,* 125-132.

This benchmark study reported in a number of papers and summarised by Elkin (1994) examined the efficacy of cognitive-behavioural psychotherapy and interpersonal psychotherapy. Neither treatment is strictly psychodynamic, although interpersonal psychotherapy has its roots in the ideas of an interpersonalist psychoanalyst, Harry Stack Sullivan (1953). IPT focuses on current interpersonal problems rather than past relationships and is a brief rather than long-term therapy. Thus the findings have limited relevance overall to the efficacy of psychoanalysis and the study is considered here simply in terms of Blatt's reanalysis of the findings.

Sample

The research was carried out at three research sites in the United States. Overall 560 outpatients were screened and those meeting criteria for unipolar depression were included; 250 patients with moderate to severe depression were selected and 239 patients entered the trial. Sixty percent had been depressed for more than 6 months and only 36% did not have a previous episode.

Treatments

Patients were randomly assigned to one of four conditions: clinical management with imipramine; clinical management with placebo; CBT; and IPT. Clinical management consisted of weekly meetings of 20-30 minutes to discuss medication, side effects and the patient's clinical status. Both the medication conditions included some psychotherapeutic components including encouragement, direct advice, support, etc.

Measures

Patients were assessed before treatment and at 4, 8, 12 and 16 weeks and followed up at 6, 12 and 18 months. Throughout standardised measures of symptomatic status were used including the Beck Depression Inventory (BDI) and the Hamilton Rating Scale for Depression (HRSD). After discharge progress was assessed using a semi-structured interview designed to assess the longitudinal course of psychiatric disorders.

Results

There were no significant differences between the treated groups and all conditions resulted in a significant improvement between pre and post treatment, including the placebo group (Elkin et al., 1989). Using a strict criteria of clinically significant change (Jacobson & Truax, 1991) with the Hopkins Symptom Checklist, 78%, 93% and 87% of the CBT, IPT and imipramine groups respectively were designated as showing clinical improvement in contrast to 65% of the placebo group (Ogles, Lambert, & Sawyer, 1995). More severely depressed patients did somewhat better with imipramine than with CBT and there was a similar trend for IPT to be superior to CBT for this group (Elkin et al., 1995). Follow up of patients over 18 months revealed that only 20% of the original sample were free of episodes of depression of at least two weeks duration (Shea et al., 1992). The groups were not statistically significant in terms of relapse rates, but clearly the study unequivocally

demonstrated that the vast majority of patients who recover during brief treatment tend to relapse pretty soon afterwards.

The reanalysis of the data by Blatt et al (1995) indicated that self-critical introjective patients did not get better in any of the treatment conditions in this study. These patients were also more likely to make suicide attempts. These results are interesting, when taken together with the findings from the Menninger project, as well as the project reported by Blatt and Ford (1994), which showed that introjective perfectionistic patients do relatively well in psychoanalytic and psychoanalytically oriented intensive psychotherapeutic treatment - even though they do poorly in short-term psychotherapy. In the Blatt and Ford study, patients were treated four times a week for fifteen months and they were shown to do better than patients with a similar severity of illness, but with an anaclitic personality structure.

Evaluation

This interesting result could suggest an important role for long-term psychodynamic treatment with self-critical introjective patients.

The Hamburg Study: Psychoanalytic focal therapy versus client-centred therapy

Meyer, A. E., (ed). (1981). The Hamburg short-term psychotherapy comparison experiment. *Psychotherapy and Psychosomatics, 35*, 77-220.

The main aim of the study was to compare the efficacy of two forms of short-term psychotherapy. Such short-term procedures could be the solution to the well-known disparity between demand for and availability of psychotherapy in highly industrialised nations. This solution could prove to be illusory, however, if the efficacy of these short-term procedures is low or nil.

An additional motivation for this project stemmed from the fact that both Client Centred Therapy (CCT) and PT were derived from psychoanalysis but that each capitalised on different aspects of it (Meyer, 1981). Only in retrospect was it suggested (Meyer & Niemann, 1984) that CCT could be considered a focal therapy, albeit with an invariant focus identical for each and every patient. The focus of CCT was considered to be as follows: "Learn to perceive and express your emotional experiences (needs and reactions) and to accept your own self, and you will have learned how to live".

Method

Subjects and Selection

Beginning in 1972 and continuing over the next 1.5 years, all psychoneurotic and psychosomatic patients who applied for treatment intake at the outpatient department of the Hamburg University Clinic, who were between 20 and 40 years of age, and who reported receiving no previous psychotherapy were invited to undergo intake testing for the research program. Of 286 patients invited 72 did not show up. The remaining 214 patients started intake assessments but only 177 were entered into the study; the others either declined participation or were excluded because of contraindications (e.g.. psychoses or absence of motivation).

The case histories of these 177 patients were presented by the intake interviewer to the focus formulation seminar. This seminar formulated a therapeutic focus and a psychotherapy prognosis (suitability). The prognosis for psychotherapy was considered good for 88 patients and these patients were then randomly assigned to PT or CCT conditions. Not all patients accepted the offer of treatment but 34 of them started therapy in each therapy group.

Post hoc statistical examination indicated the success of the randomisation procedure. The patients refusing therapy differed in only two respects from those accepting therapy: they were less sick and less sophisticated than their peers who accepted psychotherapy assignment. As is often the case, young, attractive, verbal, intelligent and successful patients (Goldstein, 1971) were over-represented in the sample. This was not caused by therapist selection factors, however, because these characteristics also typified the intake sample (Kimm, Bolz, & Meyer, 1981).

Instruments

The investigators considered that psychotherapy outcome evaluation required the measurement of a number of variables representing different aspects of outcome. Therefore, evaluators were employed and a number of different measurement devices were used with repeated assessments (Burzig, Speidel, Bolz, & Meyer, 1981; Meyer, 1981).

The measurements included: sub-tests of a differential intelligence test; a personality inventory for personality factors such as neuroticism, extraversion, and psychosomatic disturbances (the Freiburg Personality Inventory - FIP); a personality questionnaire based on psychoanalytic theory; an anxiety questionnaire adapted form the Taylor Manifest Anxiety Scale; and a symptom checklist that contained 152 items incorporating neurotic and psychosomatic complaints.

Clinical assessment scales and ratings included the interviewer's assessment (symptom change, well-being and insight); the patient's self-ratings (of symptom change, well-being and interpersonal change); and blind ratings by a group of experts of the transcripts of the two follow-ups (for assessment of symptom change, wellbeing, aspects of insight, etc.) - adjusted to ensure confidentiality of the identity of the patient, blindness as to type of therapy and length of follow-up.

Process scales included the patient session questionnaire, the therapist session questionnaire, a content-analytic catalogue for interventions and interaction and Rogerian rating scales (of empathy. warmth. self-exploration. etc). Identification and rating were also carried out for crucial incidents (Rice & Greenberg, 1984).

Procedures

Comparison of the efficacy of short-term psychotherapy. All of the patient test variables were assessed five times: during the waiting period before assignment, immediately before therapy began, after therapy, at Follow-Up 1 (3 months) and at Follow-Up 2 (9 months). Thus, it was possible to analyse the test results in two different ways. For the within-group or own-group comparison, 13 PT patients and 12 CCT patients whose treatment had been delayed were used. In this analysis, the waiting period to the treatment period for each therapy was compared by means of one-way analyses of variance (ANOVAs). For the between-groups, or reference group, the two waiting groups were combined into one untreated control group (n = 25) and compared with 21 PT patients and 22 CCT patients who began their therapy immediately (bifactoral ANOVAs).

Clinical assessment. Each patient was assessed at 3 and 9 months during follow-up. These follow-up evaluations were performed by the same clinician who had conducted the initial intake interview. Independent clinical experts were asked to make blind ratings of the transcribed interview. In addition, patient self-ratings were collected at these times.

Process research. Both randomly chosen therapy sessions and sessions selected on the questionnaire (e.g., "good hours") were subjected to a content analysis and associated ratings. These process data were used to assess differences between the two therapy groups and to investigate the interactions that occurred between patients and therapists in the sessions and were correlated with outcome variables.

Long term follow-up. After 12 years the patients in this study were invited for another interview which was conducted and coded by independent experts. The patient again answered the FPI, an anxiety questionnaire (SAL) and a social questionnaire concerning post-therapeutic developments (e.g. symptom change, further medical aid and psychotherapy, partnerships, work, and sexuality).

Results

Psychological tests. To ensure a conservative interpretation of psychological test findings, only those differences that were replicated in the own and reference control comparison and that were still observable at the 9-month follow-up were considered. Even using these rigorous criteria of change, significant gains in sociability for patients in both types of therapy were found. Additionally, significant reductions in psychosomatic disturbance, depression and anxiety among patients in CCT were observed.

In direct comparison of psychological test scores between CCT and PT groups, few significant differences emerged and none of those that did were replicated in subsequent analyses. Thus, although more significant differences were observed for CCT than for PT, the advantage that this suggests for CCT is small.

Clinical ratings. When the dependent variables produced by the blind ratings of clinicians were analysed (Meyer, Bolz, Stuhr, & Burzig, 1981), a somewhat different pattern of between-treatment differences emerged. First, PT patients developed more insight as related to the concepts described in and outside of the focus formulation than did CCT patients. Second, at Follow-Up 1, CCT patients showed a larger reduction of secondary symptoms than their PT counterparts. This difference vanished by Follow-Up 2, however, as a result of a slight additional reduction of symptoms in the PT

patients and a slight symptom increase among those in the CCT sample. Third, in the general domain defined as personality and human relations variables, patients in PT received higher ratings than those in CCT for introspection at Follow-Up 1 and patients in CCT received higher ratings for attitude toward self at Follow-Up 2. Fourth, there was a general indication that at Follow-Up 2 more CCT patients than PT patients felt that therapy had helped them.

Interactions among rating sources. When a factor analysis of the three different sources of ratings (blind clinical raters, intake interviewers and self reports) was undertaken four factors emerged:

(a) All three groups of raters agreed that treated patients felt better, had fewer symptoms and were better able to cope than they had been before treatment. This factor was labelled 'Gain in Subjective Symptom-Oriented Well-Being and in Coping Competence'.

(b) The ratings of independent (i.e., blind) clinicians emerged as a second factor reflecting a dimension of insight.

(c) The intake interviewers' ratings also emerged as a separate factor, unrelated to that defined by the independent clinicians, but as one that appeared to reflect a dimension of insight.

(d) The final factor described the patient's belief that others act differently towards and see changes in him or her.

Process differences. The results of the content analyses of therapist interventions corresponded with the theoretical concepts of each therapy school. The strongest differences found in this study indicated that CCT therapists relied on verbal repetitions, emphasised the verbalisation of emotions and supported patient defences, whereas PT therapists focused on working through, emphasised the importance of past experience and used the focus formulation to direct their interventions.

Process-outcome relationships. The process-outcome analyses based on Rogerian rating scales (e.g. empathy, warmth and self-exploration) were not closely related to outcome in CCT as had been assumed. Nonetheless, the emotionally warm, concrete and active CCT therapist was associated with changes reducing patients' introversion. In contrast to these findings, it was observed that the PT patient's process ratings (but not the PT therapist's ratings) of such dimensions as emotional involvement and acceptance of one's own feelings were positively related to reductions in neuroticism. An unusually active PT therapist, however, produced non-constructive effects on aggression and self-confidence in interaction with highly resistant patients.

Pattern of change. A cluster-analytic investigation of different clinical assessments suggested that it is possible to identify patterns of outcome that distinguished between the treatments. For example, one cluster comprised CCT patients who had been successfully treated as judged by a reduction of symptoms and improved ratings but who did not show corresponding changes or increases in insight. On the other hand, another cluster comprised PT patients who gained insight bud did not improve in symptom-oriented well-being. Two clusters were found comprising patients from both therapy groups, one with non-responders to therapy and the other with successful outcomes (particularly as indexed by changes in variables related to introspection and attitudes toward themselves). Finally, two patients did not fit into any of the foregoing clusters, suggesting that there existed some highly specific outcome configurations.

Long-term follow-up. The 12-year follow-up study, obtained further relevant information from many patients. The patients who had been offered psychotherapy but did not accept it were used as a reference control group. They showed small but at times significant gains (which can be interpreted as spontaneous remission). Compared with this control group, only changes indexed by the Extraversion scale were of sufficient magnitude to suggest the long-term advantages of both short-term psychotherapies. Additionally, long-term treatment advantages were observed on the Depression scale among patients in CCT. Among those in PT there was a tendency for changes to be observed somewhat later during the follow-up period. PT patients also reported having more subsequent treatment during the follow-up period than those in CCT. These findings suggest that psychoanalytic short-term psychotherapy may be a more mobilising experience than CCT.

Our global clinic results concerning symptomatic changes and coping skills showed no significant differences between the treatment groups and the non-treatment groups after 12 years.

Evaluation

The relevance of this study for psychoanalysis is limited by the short term nature of the original intervention. The limited observed efficacy of PT may be explicable in terms of the inadequate dose at which this treatment was administered. Of greater interest was the observation that long-term effects of PT included what may be interpreted as greater openness to change by the PT group as well as so-called sleeper effects (positive effects of therapy manifesting after termination of the treatment). Most evaluations of short-term PT do not have long follow-up periods. The Hamburg findings suggest that without such follow-up the full impact of psychodynamic treatment may not be observed.

The Norwegian Psychotherapy Study

Høglend, P. (1993). Personality disorders and long-term outcome after brief psychodynamic psychotherapy. *Journal of Personality Disorders, 7*, 168-181.

Høglend, P. (1993). Suitability for brief dynamic psychotherapy: Psychodynamic variables as predictors of outcome. *Acta Psychiatrica Scandinavica, 88*, 104-110.

Høglend, P., Sørlie, T., Heyerdahl, O., Sørbye, Ø., & Amlo, S. (1993). Brief dynamic psychotherapy: Patient suitability, treatment length and outcome. *Journal of Psychotherapy Practice and Research, 2*, 230-241.

This was a study of moderate length psychodynamic psychotherapy which is of relevance here because it emphasises the importance of long-term treatment if psychoanalytically oriented therapy is offered to personality disordered patients.

Sample

Forty five outpatients presenting in an outpatient psychotherapy clinic were offered psychotherapy. The patients were moderately disturbed and had an average GAF score of 62. DSM III diagnoses were established on the basis of a structured interview. There was a subsample of 15 personality disordered individuals whose outcome is contrasted with that of the rest of the sample.

Treatment

Patients were treated by experienced psychodynamic psychiatrists using a therapy manual based on the work of Malan and Sifneos. The length of the therapy varied from 9-53 sessions. The treatments were open-ended and the sessions were audio-taped. Treatment adherence was ensured by peer supervision.

Measures

There was a two year and a four year follow-up. Patients were rated on the GAF, on a target complaints inventory and a psychodynamic scale with adequate reliability but unknown validity. Ratings were made on the bases of interviews conducted and audio-taped by the clinicians who were not blind to pre-treatment evaluations or length of treatment. The group of clinicians listening to the interviews independently rated the patients' outcome.

Results

Patients with personality disorder diagnosis did less well than those without at the end of treatment and at two year follow-up. Four years after therapy there were no significant differences between the PD group and the rest of the sample. There were no differences between DSM cluster B and cluster C personality disorders.

For the PD group alone, there was a significant correlation between degree of insight and number of sessions of treatment two years after therapy and degree of dynamic change and number of sessions four years after termination. The results imply that long-term improvement only occurred after more than 30 sessions of treatment. Patients with PD who had less focused treatments, allowing for more free association and a less exclusive focus on pre-assigned dynamic goals, did better. Level of insight achieved after two years predicted dynamic change at four year follow-up.

Evaluation

There are major methodological weaknesses in this study. There were no self-report or standardised measures of outcome. The sample size was small. The findings are incidental observations in a study

designed to examine the effectiveness of focused psychotherapy in general. The raters were not blind as to treatment length. They were, however, blind to the hypothesis of the study with regard to the effect of personality disorder on treatment outcome. The results are suggestive rather than definitive in the absence of an experimental design with random assignments into treatments of different lengths. This is a hypothesis generating study which produced interesting results which now require experimental testing.

McLean's Follow-on study (MCFO)

Najavits, L.M., & Gunderson, J.G. (1995). Improvements in borderline personality disorder in a 3 year prospective outcome study. *Comprehensive Psychiatry, 36,* 296-302.

This was a naturalistic study of patients who entered psychotherapy following a period of hospitalisation.

Sample

Thirty seven female patients meeting criteria for a diagnosis of BPD were consecutively recruited.

Treatment

The patients were in psychodynamic psychotherapy once or twice a week.

Measures

Assessments were made at baseline (discharge from hospital) and at one, two and three years.

Results

Patients generally improved, although they showed a fluctuating course. Co-morbid anxiety-related disorders were associated with a relatively poor outcome.

Erica Process and Outcome Study (EPOS) of goal directed, time-limited child psychotherapy with parental counselling

Carlberg, G. (1999). *Vändpunkter i barnpsykoterapi. Psykoterapeuters erfarenheter av förändringsprocesser*. [Turning points in child psychotherapy. Psychotherapists' experiences of change processes]. Dissertation, Department of Education, University of Stockholm. Edsbruk: Akademitryck.

Boalt Boëthius, S., & Berggren, G. (2000). *Forskning om barn- och ungdomspsykoterapi. En kunskapsöversikt*. [Current research in child and adolescent psychotherapy. A comprehensive overview.]. Stockholm: The Erica Foundation.

Carlberg, G. (1997). Laughter opens the door: turning points in child psychotherapy. *Journal of Child Psychotherapy, 23*, 331-349.

Background

EPOS –The Erica Process and Outcome Study is based on earlier studies of change processes in child psychotherapy (Carlberg, 1999) and on a comprehensive overview of current research in child and adolescent psychotherapy (Boalt Boëthius & Berggren, 2000). These two projects were in turn founded on knowledge acquired from international research (Fonagy & Target, 1996b; Kazdin, 1995).

Aim

The aim of the project is to study a form of psychotherapy that can be performed in ordinary clinical settings in Sweden. The increased pressure on child psychiatry makes it necessary to develop time-limited and well-defined psychotherapeutic methods. In this project both outcome and process in such child psychotherapies will be studied. The aim is to deepen knowledge of the connection between process and outcome.

Sample & Treatment

Extensive data will be collected from 24 cases. Child guidance clinics from different parts of Sweden will be involved in the project. The form of psychotherapy studied is defined as "goal directed, time-limited child psychotherapy with parallel parental counselling". Therapy frequency will be 1-2 sessions a week with a duration of 1-2 years. The parents meet their counsellor once a week or at least every fortnight. Therapists and parents formulate goals and frames for the therapies as carefully as possible at the start of therapy. The children are between 5 and 10 years of age at the beginning of therapy. Each clinic makes a diagnostic assessment and decides about the choice of psychotherapy form.

In order to compare the children studied with a group of children receiving "treatment as usual", co-operation with other child guidance clinics has been initiated. The children in the control group will be assessed with SDQ, have been referred for the same reasons, and will be age and sex matched.

Measures

In this study the process of change will be followed with the help of various specific research instruments, questionnaires and interviews. These will be used at the start, during and after the treatment period. A follow-up after three years is planned. Processes in the child psychotherapies and the parallel parental counselling will be studied.

Besides routine psychological assessment at the *start of therapy* the following instruments will be used: DSM-IV, HCAM–The Hampstead Child Adaption Measure, and SDQ–Strength and Difficulties Questionnaire (parent and teacher versions). The same instruments will be used after therapy.

In connection with each session the child psychotherapist and the parental counsellor will make process notes and complete a form, FWC–Feeling Word Checklist, in order to follow the therapists' countertransference feelings and to facilitate studying sessions of special interest.

Every third month questionnaires will be distributed to the therapists and parents. The questionnaires focus on important themes and changes inside and outside therapies during the time period studied. The therapeutic alliance and changes in the goals of therapies will also be investigated.

Taped interviews with therapists, parents and, sometimes, children will be conducted in 12 of the 24 cases, twice during the treatment period.

Data will be analysed using both qualitative and quantitative methods.

Results

The project started in September 1999 with planning, selection and testing of instruments. Collection of data started in January 2001. Data collection will be completed 2003.

Evaluation

This is a small scale study from a new research team in an under-researched area, which may have quite a lot to contribute to the process-outcome literature.

Part 5 Summary and Conclusions

Limitations of the evidence

It is easy to be critical of psychoanalytic studies. There are no definitive studies which show psychoanalysis to be unequivocally effective relative to an active placebo or an alternative method of treatment. There are no methods available that might definitively indicate the existence of a psychoanalytic process. Most studies have major limitations which might lead critics of the discipline to discount their results. Others have limitations that are so grave that even a sympathetic reviewer might be inclined to discount the findings. For example, is the analyst in a position to judge the outcome of a treatment? Not only is there the issue of a self-serving bias, but also is the context of free association not totally incompatible with the systematic gathering of data concerning adjustment and the like? Amongst the most common problems are: the lack of use of standardised diagnoses, inadequate specification of the treatment procedures, lack of control for selection biases in sampling, the absence of intent to treat controls and the failure to follow up drop outs, the use of inexperienced therapists, the lack of homogeneity of the patient groups considered, heterogeneous methods of intervention and related to this the lack of a generally accepted manualised method of intervention, the lack of statistical power, the lack of random assignments to treatment groups, lack of independent assessment of outcome, lack of standardisation of measures of outcome, questionable validity of some outcome measures, poorly matched comparison groups, absence of control for the law of initial and of regression to the mean, failure to take adequate baseline measures, and related to this reliance on retrospectively collected data, inadequate detail on statistical analysis and inappropriate statistics reported, inadequate control for intercurrent treatments, and so on.

Notwithstanding the many limitations, however, the sheer number of studies available is encouraging, particularly the range of ongoing studies. This was by no means an exhaustive review. Limitations of time principally prevented us from reviewing a large number of investigations, some very well known, many with findings consistent with those which were included. The review is labelled "open" to underscore our intention to include further studies as time permits and as these are brought to our attention. The emphasis has been on some less well publicised studies and studies with challenging findings, not necessarily on investigations with the best methodology. Many of the conclusions should therefore be heavily qualified in the light of the questionable internal validity of the observations reported. In summarising these results, however, we will adopt a cautiously optimistic attitude in relation to the evidence presented. It is not that in this way we are turning a blind eye to the weakness of the evidence, but rather we wish to highlight what could be shown by these studies and which way the evidence currently points. Many of the ongoing studies are methodologically "state of the art" and this is of course encouraging from the point of view of persuading sceptics in the field. The present review, however, is intended for "internal" consumption. As psychoanalysts we all know that psychoanalysis works. Our own analytic experience is probably sufficient in most instances to persuade us of its effectiveness. The purpose of the review was to assist in making accessible studies which have systematically explored the patient groups which benefit from treatments administered by members of our organisation. In general, the findings underscore the effectiveness of our work and should encourage us to undertake further, even more rigorous, explorations of treatment outcome.

Appendix of Process Measures

A. Process Measures

The Psychoanalytic Process Rating Scale (PPRS)

Sandler, A.-M. (1993). Introduction to the one-day conference on the work of the Anna Freud Centre's Young Adults Scheme: Development Issues in Psychoanalytic work with Young Adults. *Bulletin of the Anna Freud Centre, 16*, 3.

Beenen, F., & Stoker, J. (2001). *Psychoanalytic Process Visualised, first version* (available only in Dutch, no English publication yet): Dutch Psychoanalytic Institute (NPI).

Aim

By means of the PPRS the course of treatment in psychoanalysis and long-term psychoanalytic psychotherapy, especially the form and content of the curative psychoanalytic process, can be captured, visualised and evaluated.

Description

The PPRS is a systematic clinical judgement scale of about 250 items that represent relevant aspects of the psychoanalytic process. It has been constructed and tested in clinical practice by Beenen and Stoker at the Dutch Psychoanalytic Institute (NPI). The scale had as its starting point the about 500 items of the Session Rating Scale of the Anna Freud Centre in London(Bulletin Anna Freud Centre,1993).

The PPRS items are subdivided in three chapters. Chapter I, General Aspects of the Treatment, representing significant form elements like general attitude of the patient, treatment commitment and quality of the sessions in the period rated. Also basic defence and resistance patterns of the patient are being checked, including his or her general mood states in the analysis.

The items in Chapter II, The Psychic Content, refer to the conscious and unconscious material that dominates the treatment period under consideration. Next to 'classical' areas like sexuality and aggression, the focus is also on issues like bodily sensations, types and vicissitudes of patient's object relationships and so on. The psychic content either is actual or was present in the past, and can be conscious or unconscious. Chapter III more or less takes up the issues of Chapter I again, but now the focus is on the (curative) interaction between the analyst and analysand. Transference themes, the analyst's style of work and the analysand's reactions to his attitude, interventions and interpretations as well as the analyst’s (countertransference) feelings and general feeling of (dis)satisfaction about the treatment are examples of the content of this chapter.

The PPRS can be used to judge one or more sessions (a period of treatment). It uses a 4-point scale to determine presence/absence or agreement/disagreement and/or the item was, yes or no, subject of intervention or interpretation.

Practical Issues

Applying the PPRS, which in principle should be done once a month, at the moment takes 1 to 1.5 hours for the ratings and ten minutes for a secretary to process the scores into the computer. After six or more filled in PPRS's it makes sense to produce an overview of the process, which can be done at the NPI by a special computer program (also applicable to other languages). Interpretation of the

output takes another 0.5 to 1 hour. At the moment a substantially shorter form of the PPRS (about 100 items) is under construction and will be ready in the course of 2001. This will make regular clinical application much more attractive. The PPRS and its manual can be obtained from the NPI (contact person Jan Stoker, email address: kc@npsai.nl).

Psychometric Properties

To test the inter-rater reliability of the list is problematic, as each time it concerns the clinical judgement of this analyst about one of his or her analyses/analysands. No third party is or can be involved directly. Moreover, themes like countertransference feelings and so on are by definition personal and subjective. An extensive manual was constructed in which every item is described as clearly and operationally as possible, in order to increase the chance for reliable judgements. In practice, that after some training and some experience with using the list, the majority of the clinicians interpret and score the items in a similar enough way.

The validity of the PPRS still has to be proved, i.e. how relevant are (differences in) PPRS - measured courses of process for success or failure of the treatment. And moreover, do analyses that during treatment have been 'corrected' or guided by the PPRS results on the whole produce better outcomes than those that have been not. So, in all cases systematic and reliable follow-up studies of the treatments are obligatory.

In the meantime at the NPI an extensive follow-up programme of psychoanalytic treatments has been implemented.

Clinical Utility

This is a core strength of the PPRS. At the NPI it has been used so far for 55 psychoanalyses to monitor the treatment and this resulted in a big step forward, away from clinical 'fairy tales' and judgement/decisions by hierarchy. The yearly 'objective' PPRS- picture of the analytic process furnishes the analyst with a self - constructed mirror of his view on and position in the treatment. It is also a systematised and standardized clinical recording of the process, which enables a systematic comparison of the same treatment over time as well as comparison of different treatments on the same base.

Referential Activity (RA): Scales and Computer Procedures

Bucci, W. (1995). The power of the narrative; A multiple code account. In J. Pennebaker (Ed.), *Emotion, Disclosure and Health*. Washington, D.C.: American Psychological Association.

Bucci, W. (1997). *Psychoanalysis and Cognitive Science: A Multiple Code Theory*. New York: Guilford Press.

Bucci, W., Kabasakalian, R., & al., (1992). *Instructions for scoring Referential Activity (RA) in transcripts of spoken narrative texts*. Ulmer Textbank, Ulm, Germany.

Bucci, W., & Miller, N. (1993). Primary process: A new formulation and an analogue measure. In N. Miller & L. Luborsky & J. Barber & J. Docherty (Eds.), *Psychodynamic Treatment Research* (pp. 381-406). New York: Basic Books.

De Coro, A., & Caviglia, G. (Eds.). (2000). *La valutazione dell' attivita' referenziale*. Roma: Edizioni Kappa.

Dubé, J. E., Roussos, A. J., & Bucci, W. (2001). *New procedures to ensure cross-cultural validation of the Computerized Referential Activity dictionaries*. Paper presented at the 32nd Annual Meeting of the Society for Psychotherapy Research, Montevideo, Uruguay.

Mergenthaler, E. (1985). *Textbank Systems. Computer science applied in the field of psychoanalysis*. Heidelberg & New York: Springer.

Mergenthaler, E., & Bucci, W. (1999). Linking verbal and non-verbal representations: computer analysis of referential activity. *Br J Med Psychol, 72*, 339-354.

Roussos, A., Acosta, S., Juárez, C., & Mergenthaler, E. (in press). Introducción a las Técnicas De Investigación sobre Procesos Terapéuticos Asistidas por Computadora, "Modelos de Ciclos Terapéuticos". *Revista Interamericana de Psicología*.

Solano, L. (2001). *Tra Mente e Corpo: come si costruisce la salute*. Roma: Raffaello Cortina.

Aim

The construct of the referential process, as defined in the theoretical context of the multiple code theory (Bucci, 1997), concerns the function of connecting nonverbal experience, including emotional experience, with language. The referential process plays a central role in psychotherapy; the patient needs to express subjective emotional experience in the shared verbal code, and to connect the words of the therapist back to these emotional representations to bring about change. Measures of RA have been developed to represent this process, to trace its fluctuations in therapeutic interactions, and to identify factors determining treatment effectiveness.

Description

Variation in activity of the referential process is shown in features of language style. Two methods for assessing RA variation are in use: scales rated by judges and a computerized procedure. The rating scales measure the Concreteness, Imagery, Specificity, and Clarity of speech. Concreteness is based on degree of perceptual or sensory quality, including references to all sensory modalities, action and bodily experience. Imagery refers to the degree to which language evokes corresponding experience in the reader or hearer. Specificity refers to amount of detail; a highly specific text involves explicit descriptions of persons, objects, places or events. Clarity refers to clarity of an image as seen through the language; how well-focused is the linguistic image. For all scales, the rating is done on an 11-point range, from 0 to 10. Scores on the four scales may be averaged to yield an overall RA score. Scale descriptions and rating procedures are provided in the manual for scoring RA (Bucci et al., 1992).

To facilitate the application of this measurement in large sample and longitudinal studies and in psychotherapy research, a computerized referential activity measure (CRA) was developed by

modeling the overall RA score as rated by trained judges in a large and varied set of texts. (Mergenthaler & Bucci, 1999). Essentially, the CRA consists of two list of highly frequent words: one most closely and uniquely associated with texts rated as High in RA, and the other with texts scored Low in RA. To apply the CRA, these word lists are compared to a text or text segment, the number of matches (tokens) are counted for the High and Low lists, and the difference between the matches are computed as a proportion of total word count in the text. (Also see Roussos et al., In press.) The CRA differs from traditional computer assisted content analysis measures in that the items in the list are indicators of style rather than representative elements of particular content areas. The procedures followed by Mergenthaler and Bucci yielded a very small list of less than 200 exceedingly frequent words (types) which together account for approximately half of all words (tokens) in most spoken and written texts.

Practical Issues

For the rating scales, scorers are trained by following instructions and scoring sets of practice excerpts provided in the scoring manual, then discussing deviant scoring, and scoring additional practice items until an acceptable interrater reliability is achieved. In addition to the English version of the manual (Bucci et al., 1992), an approved Italian version is available (De Coro & Caviglia, 2000) and French and Spanish translations are in preparation. A program for application of the CRA is available through the University of Ulm. New PC versions and other foreign language versions, including Spanish, Dutch, Norwegian, Italian and French are in preparation (Dubé et al., 2001).

Psychometric Properties

a) Inter-rater reliability of RA scales: Roussos et al. (Roussos, Boffi Lissin, & Bucci, in preparation) found current scorers after brief training achieved high reliability in a pairwise comparison (ICC mean R square = 0.75) on a set of text excerpts. In a correspondance analysis, considering the whole group rather than pair-wise comparison, all scorers were considered to evaluate the excerpts in a similar way (matrix trace > .05). The rating procedure shows excellent stability; an ICC mean R square = 0.70 was found comparing current scorers with ratings carried out more than ten years previously, following the same manualized instructions. RA is an easily scored measure that taps natural linguistic intuition, and is also an implicit part of the clinical listening process. Psychology undergraduate and graduate students and experienced clinicians, scoring the same excerpts after reading the manualized instructions, without training, did not show significant differences from the trained scorers in a correspondance analysis (matrix trace > .05). The trained group, however, showed a substantially lower range of dispersion among judges (-.02 to .25 on axis 1) compared to the nontrained groups (-.48 to .36 on axis 1).

b) Reliability of CRA: The broad coverage, accounting for approximately half of all words spoken, makes CRA an exceedingly powerful dictionary, requiring a minimal text size of only 15 words to achieve adequate reliability (with an error estimate of 5%) according to the power analysis procedure as applied by Mergenthaler (1985). (The issue of inter-rater reliability does not apply for computerized procedures).

c) Validity of RA scales: The construct validity of RA has been developed in the theoretical context of the multiple code theory. Connections to emotional experience are expressed in the specific and concrete stories people tell about their lives; RA has been found to be significantly higher in texts identified as narratives than in non-narrative texts (Moore, 1992). Studies in the areas of object relations, attachment and physical health have validated the role of the RA function in bodily and emotional self regulation and interpersonal interaction; some of these are summarized in Bucci (1995; 1997), and Mergenthaler & Bucci (Mergenthaler & Bucci, 1999).[1] In a meta-analysis of 23 RA studies, Samstag (1996) found a significant relationship, with moderate to strong effect size, between RA measures and the capacity to synthesize cognitive, linguistic and emotional experience. Discriminant as well as convergent validation has been developed. RA has been shown by Solano

[1]More recent studies are summarized in Bucci, W., "State and trait features of the Referential Activity Dimension," working paper.

(2001) and his colleagues to be associated with levels of physical health, but distinct from alexithymia and affective tone. RA also varies independently of standard verbal intelligence, and Experiencing and Insight measures applied to the same text material, as summarized in Bucci and Miller (Bucci & Miller, 1993), Bucci (1997), and Mergenthaler and Bucci (Mergenthaler & Bucci, 1999).

d) Validity of CRA: The criterion and construct validity of CRA have been outlined by Mergenthaler and Bucci (Mergenthaler & Bucci, 1999). CRA showed a pooled correlation of R' = .47 with RA scale ratings, across 6 large text samples, Several studies summarized by Mergenthaler and Bucci have shown significantly higher correlation of CRA with narrative than non-narrative speech, providing independent support for the construct validity of this measure. The findings suggest that high CRA language is dominated by the special narrative features – the setting of time and place and the introduction of other persons, usually in relation to oneself – that are associated with descriptions of episodes.

Clinical Utility

The RA measures have been applied to many types of texts, including brief monologues, early memories and Thematic Apperception Test (TAT) protocols, as well as transcripts of therapy sessions, and have been used with a wide range of populations, including children, adolescents, and adults, in clinical and nonclinical contexts. The CRA has been extensively applied in psychoanalytic and psychotherapy process research; fluctuations in CRA (along with associated computerized procedures and rating scales) reflect the patient's movement through the phases of the referential process, and enable tracking of change from dissociation to connection within emotion schemas.

The TCA Rating Methodology for Psychotherapy Process

Foelsch, P.A., Normandin, L. & Clarkin, J.F. (1997). *Transference and Countertransference themes as Indicators of Structural Change*. Paper presented at the Annual Conference of the Society for Psychotherapy Research, Geilo, Norway.

Foelsch, P.A. (1998). *Transference and Countertransference Themes as Indicators of Structural Change.* Paper presented in D. Diamond, Ph.D. (Chair), *Attachment style, transference-countertransference dynamics and structural change in borderline patients.* Symposium at the 18th annual spring meeting of the Division 39, American Psychological Association, Boston, MA, April 25th, 1998.

Normandin, L. & Foelsch, P.A. (1999). *Transference Counter Transference Analysis (TCA): A System to Assess Structural Change in Severe Personality Disorders Throughout Treatment*. Paper presented at the Annual Conference of the Society for Psychotherapy Research, Braga, Portugal.]

Normandin, L. (1991). *Le contre-transfert comme élément constitutif du travail et de l'espace thérapeutiques*. Thèse de doctorat inédite. Université de Montréal.

Normandin, L. & Bouchard, M.A. (1993). The effects of theoretical orientation and experience on rational, reactive, and reflective countertransference. *Psychotherapy Research, 3*(2), 77-94.

Aim

The TCA is designed to assess the moment-to-moment changes in psychotherapy process by rating the Modes of Mental Activity of both the patient and therapist. It also provides an index of structural change for the patient in psychotherapy.

Description

The TCA (derived from "Transference Countertransference Analysis") is a rating system applied to videotapes of psychotherapy sessions. Raters are trained to code utterances (defined as turns of speech) in the session for the Modes of Mental Activity (Reactive, Rational, Reflective), Dominance (which of these three modes is most salient in the utterance), Theme (the content of what is said), Agreement of Focus (codes the extent to which the therapist and patient are communicating about the same topic and its priority within the session), and Affect (codes for the intensity and type of affect expressed in the utterance).

The Modes of Mental Activity have been conceptualized by Normandin (1991) and are based on an integration of the classical Freudian, Kleinian, and Lacanian models, and the contemporary schools of object relations' theory on the subject of countertransference. Each Mode of Mental Activity (MMA) characterizes a type of communication used by both the patient and therapist, expressed via the patient's transference and therapist's countertransference manifest in the session. Briefly, the Reactive mode is characterized by the patient/therapist as an unaware, unconscious participant in the process. Rational mode is evident when the communication is primarily cognitively based without insight or integration of affect. The Reflective mode is characterized by the patient/therapist being at least partially aware of his or her experience as a participating subject. When the Reflective Mode is coded, the rater then rates the level of reflectivity, "emergence", "immersion," and "elaboration." Each level indicated a deepening awareness and integration, moving from preconscious to conscious. While the MMA are rated independently, the "Dominance" code allows the rater to select the MMA that is most salient, or predominant in the utterance. This rating allows for a "clinical" or "subjective experience" code by the rater.

The Theme, Agreement of Focus, and Affect ratings (in process) in the rating system will provide additional ways to monitor points of intervention, assess alliance, and the use of affect within a session.

Practical Issues

The TCA requires access to all channels of communication, verbal and non-verbal and thus requires videotapes of psychotherapy sessions. These are not always practical to acquire outside of research protocols. Thus far, these researchers have had the most success training raters who have had some clinical experience, and who have the ability to access their own countertransference reactions. Training consists of a didactic review of the manual, coding a video tape with an "expert rater" to see the definitions operationalized, then coding a series of videotapes independently (usually a minimum of five) to compare with expert ratings. Interrater reliability is assessed to determine when training is completed.

Psychometric Properties

Inter-rater reliability was established for the TCA method following 20 hours of training. After training, two judges independently rated 8 randomly chosen psychotherapy sessions recorded on videotape. The goal was to categorize the modes of mental activity reactions inferred from therapist and patient's verbal and non-verbal reactions. Kappa coefficients (Cohen, 1988) were calculated at each step of the TCA scoring. Rational mode showed moderately good agreement ($k = .53$). Reactive mode results suggested good agreement ($k = .62$). Reflective mode showed judges reached high agreement ($k = .71$). Thus good Interrater reliability was established for the TCA ratings. Individual ratings that differed following this initial rating were compared between raters and discussed. Consensus was reached for those items. Pilot data looking at the relationship between changes in TCA ratings and changes in patient symptoms and functioning indicate that the TCA correlates with patient changes in the hypothesized direction.

Clinical Utility

This rating system provides a measure of therapist activity in the psychotherapy session and its relation to patient activity. This allows therapists and supervisors to identify areas of potential difficulty in the therapy process and allows for focused review and analysis. This ability to target both the areas of the process and the specific modes of mental activity leads to direct interventions to improve the psychotherapy process. The TCA has been used as a training tool for both beginning and advanced therapists.

B. Process-Outcome Measures

The Psychotherapy Process Q-Set (PQS): Studying How Patients Change

Albani, C., Blaser, G., Jones, E.E., Thomae, H., & Kaechele, H. (2001). Amalia X in lichte des "Psychotherapie-Prozess Q-sort" von E. Jones". In U. Stuhr, M. Leuzinger-Bohleber, & M. Beutel, (Eds.). Langzeit-Psychotherapie: Perspektiven fuer therapeuten und wissenshaftler. Frankfurt: Kohlhammer Verlag (pp. 213-221).

Ablon, S. & Jones, E. E. (1999). Psychotherapy process in the National Institutes of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 67*, 64-75. .

Jones, E. E. (2000). *Therapeutic action: A guide to psychoanalytic research*: Northvale, N.J.: Jason Aronson.

Jones, E. E., Hall, S.A., & Parke, L. A. (1991). The process of change: the Berkeley Psychotherapy Research Group. In L. Beutler and M. Crago (Eds.), *Psychotherapy research: An international review of programmatic studies.* Washington, D.C. American Psychological Association, pp. 99-106.

Jones, E. E. & Pulos, S.M. (1993). Comparing the process in psychodynamic and cognitive-behavioral therapies. *Journal of Consulting and Clinical Psychology, 61*, 306-316.

Jones, E. E. & Windholz, M. (1990). The psychoanalytic case study: Toward a method for systematic inquiry. *Journal of the American Psychoanalytic Association, 38*, 985-1015.

Aim

A central difficulty for psychoanalytic process research lies in designing quantitative methods that both preserve the depth and complexity of clinical material while conforming to the requirements of empirical science. One method that meets these criteria is the *Psychotherapy Process Q-set* (PQS; Jones, 2000). The PQS is a 100-item rating instrument designed to provide a basic language for the description and classification of treatment processes in a form suitable for quantitative analysis. It forms the basis for an observationally grounded research. In order to use clinical (observational) data to test psychoanalytic constructs, clinical phenomena must be intersubjectively observable, which is to say that different judges can independently agree about whether they occur and their characteristics. Disagreements about the interpretation or meaning of the same case material are commonplace in clinical work and constitute important grounds for criticism about the scientific status of psychoanalytic methods for acquiring knowledge. It is crucial that any research methodology establish the extent of consensus, or reliability, among judges about the presence and nature of a clinical phenomenon. The PQS provides a language and rating system that helps clinical judges achieve reliable descriptions of complex treatment processes.

Description

The PQS yields a score from 1 to 9 for each of 100 descriptive variables or items describing patient attitudes, behaviors or experience; the therapist's actions and attitudes; and the nature of their interaction. Each statement can be rated from extremely characteristic to extremely uncharacteristic for a given treatment hour. The items are conceptualized at clinically meaningful levels and anchored, as far as possible, to specific, concrete behavioral and verbal cues that can be identified in recordings of therapy hours. Accurate records of the therapist's and patient's speech are essential in attaining rater reliability, especially for those processes that require inference. A coding manual (Enrico E. Jones, 2000) provides the items and their definitions along with examples of their application to minimize potentially varying interpretations. It also specifies the rules governing the use of inference in making Q-ratings. Since the items are not closely bound to particular theoretical concepts, but rather to

notions of analytic and therapeutic *process*, the influence of observers' theory on their descriptions of the process is subdued within the framework provided by the Q-set (Jones et al., 1991).

The initial item set was drawn from a wide variety of sources. The set was revised and refined through an iterative revision process in which the item set was repeatedly piloted on new samples of therapy transcripts and tapes. Almost all process rating systems rely on recordings of brief segments of therapy sessions, forcing judges to rate a dimension of presumed relevance on the basis of relatively brief impressions. In contrast, with the Q-technique an entire hour (audiotaped or videotaped) rather than a small segment is the time frame rated, allowing a greater opportunity to capture important events. The rating procedure permits judges to form hypotheses and study the material for confirmation or alternative conceptualizations. The ratings conform to a normal distribution, which requires judges to make multiple evaluations among items, thereby avoiding either positive or negative halo effects.

The PQS is used in research involving group comparison designs, in which Q-ratings of groups of cases (or hours) selected on some dimension of interest are compared (Ablon & Jones, 1999; Jones & Pulos, 1993) as well as in N=1 designs (Jones & Windholz, 1990). The special value of the Q-method is that it provides a way of quantifying the qualities of the analytic or therapeutic process. It can capture the uniqueness of each treatment hour while also permitting the assessment of the similarities or dissimilarities between hours and patients.

Practical Issues

Q-sorting. The 100 Q-items are printed on 2 X 3 _ inch cards to permit easy arrangement and rearrangement. Clinical judges watch a videotape or read the verbatim transcripts of an entire therapy hour and then sort the 100 items in the Q-set on a continuum from least characteristic or negatively (category 1) to most characteristic or salient (category 9). The middle pile (category 5) is used for items deemed either neutral or irrelevant to the particular hour being rated. Each item contains a description of the two opposite ends of the continuum along which the items are to be rated. It is important to note that placement in the uncharacteristic direction does not signal that a particular behavior or experience is irrelevant. On the contrary, an uncharacteristic ranking signals that the absence of the item is meaningful and important to capture in the Q-sort the description. Most items have specific instructions that provide examples of the distinction between uncharacteristic and neutral ratings. For example, Q-item 9 describes the therapist as "distant or aloof" when rated in the characteristic range. However, when rated in the uncharacteristic range, the item indicates that the therapist was "genuinely responsive or affectively involved" (the opposite of "distant or aloof"). Only if the item were irrelevant to the description of the hour would it be placed in the neutral range. The number of cards sorted into each category of the Q-sort (from 5 at the extremes to 18 in the middle or neutral category) conforms to a normal distribution, requiring judges to make multiple evaluations among items thereby avoiding halo effects and response sets. The items are tied to specific actions, behaviors, and statements. A detailed coding manual provides the Q-items and their descriptions as well as operational examples. When rating, judges are asked to take the position of a "generalized other" i.e. an observer who stands mid-way between patient and therapist and who views the interaction from the outside. In placing each item, judges are instructed to ask themselves: Is this attitude, behavior, or experience clearly present (or absent)? If the evidence is not compelling, the judge is asked to search for specific evidence of the extent to which it present or absent. A transcript or videotape of a therapy hour typically requires 1 _ to 2 hours to Q-sort.

Training. Raters can be trained to Q-sort to acceptable levels of reliability relatively quickly if novice raters have had exposure to the fundamentals of therapy and some minimal clinical experience. Usually one year of supervised psychotherapy is sufficient. Training can be accomplished most efficiently in small groups in which trainees, after having studied the Q-sort Manual (Enrico E. Jones, 2000), independently sort a videotape or transcript of a treatment hour. Trainees then compare and discuss their item placements. The mechanics of Q-sorting, varying interpretations of the clinical material, and rules regarding the use of inference can all be taken up, using the manual as a guide. Calculating inter-rater reliabilities and presenting them to raters as a form of feedback is helpful. Three or four group sessions are usually sufficient to train raters to reliability, especially if they are clinically experienced. Thereafter, occasional calibration sessions are useful in preventing rater ‘drift’,

i.e. the tendency of some raters to use items in stereotypic or idiosyncratic ways after Q-sorting a number of hours.

The Q-sort is available in German and Spanish translation (contact author).

Psychometric Properties

The PQS is an observational instrument used for the description of interaction, and is not closely tied to a particular theory of therapy. It has been used to rate treatment sessions of psychoanalysis and psychodynamic therapy, as well as cognitive-behavioral, client-centered, Gestalt, and rational-emotive therapies. The PQS has consistently demonstrated high levels of inter-rater reliability, item reliability, and concurrent and predictive validity (Jones & Pulos, 1993) across a range of studies and treatment samples. Interrater reliability, which is calculated by correlating the Q-sorts of multiple raters across all 100 items of the PQS, has been consistently satisfactory, with alpha coefficients ranging from .83 to .89 for two raters. If inter-rater reliability for two judges fall below .50, a third rater should be added. Reliability calculated at the individual Q-item level has also been consistently satisfactory, ranging from .50 to .95 across several different samples.

Tools of the Saarbrücken Research Group (SASU)

FACS (Facial Action Coding System)

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the Human Face*. New York, N.Y.: Pergamon Press.

Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System (FACS): A technique for the measurement of facial action.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Rosenberg, S. E. (Eds.). (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. New York: Oxford University Press.

FACS is a method for *descriptively* registering every facial movement that is anatomically feasible on the basis of innervations of the facial muscles. Each movement is assigned to a so-called action unit (AU). Training time with a self-training guide is approximately 80 to 100 hours.

Reliability

Comparison of experts (Ekman and Friesen) and six coders just having learned FACS:

Mean ratio across all six coders compared to experts: .82

Intercoder agreement of the six coders: .76

Mean ratio across all six coders when considering differences in intensity scoring: .78

Validity

Inferences can be made from the following affective states: happiness, fear, contempt, disgust, anger, surprise and sadness according to a dictionary based on action unit combinations. Reliability of the affect- inferences is as high as the one for the action units since the dictionary is defined completely through the combination of action units. In addition, blends (simultaneous activation of two affects), masking (one affect masking another one), and degree of experiencing the emotion happiness can be measured reliably. The dictionary is rather conservative, pushing reliability, leaving usually forty to sixty percent of the action unit combinations uninterpreted. Combinations that a clinician would often use for his interventions can have different meanings related to the core conflict the patients are suffering from. The context variables mentioned in the next section may are usually helpful to make inferences on the nature of the conflict. (Intrapsychic or interpersonal in the sense of transference). Additional information can be extracted from semantics (especially metaphors, polysemia, etc.).

EMFACS (Emotional Facial Action Coding System)

Friesen, W. V., & Ekman, P. (1984). *EMFACS-7. Unpublished manual.*

EMFACS is a method for objectively scoring *emotional* facial action. It is based on FACS, but is a short, modified version of it and differs in scoring rules and measurement procedure. EMFACS is a selective instrument: only emotional or affective expressions are scored within the stream of behaviour. AUs (action units) are specified that are used to express emotions that have universal facial patterns. A computerized emotion dictionary is based on cross-cultural studies, laboratory studies and experiments. Anger, contempt, disgust, fear, sadness, surprise, happiness are registered, also blends and masks; felt and unfelt and controlled and uncontrolled events are differentiated.

Reliability

1. Comparison of FACS and EMFACS: 60 samples, scored by two coders. EMFACS does not detect as many events as FACS, but is capable of detecting differences between behaviour samples in a way comparable to FACS _rank-orders of behavioural samples are quite similar

2. Inter-coder reliability of EMFACS is satisfactory (comparing the way in which two coders ordered the samples; Kendall Taus range from .560 up to .873)

Advantages

EMFACS is a useful, economical tool when one is interested solely in the frequencies of emotional facial expressions (not, for example in the exact course and duration of the facial action or in the succession in which certain action units occur in an emotional expression). Depending on the interest of the research, the limited set of AUs can easily be expanded. It is a conservative method that avoids over-interpretation. EMFACS enforces strict separation between the objective scoring and following interpretation.

Disadvantages

The emotion dictionary identifies only some emotions; other expressions that may be of importance in clinical context are not included in the dictionary (e.g., crying, charming behaviour and so on). Only fully expressed emotional patters receive an interpretation in the dictionary; rudiments that may nevertheless be indicative for emotional activity receive none. There is a danger of attributing emotional facial signs to experienced feelings of a person. EMFACS is less precise and less sensitive than FACS.

QuickFACS

Benecke, C. Mimischer Affektausdruck und Sprachinhalt im psychotherapeutischen Prozess. Dissertation in der philosophischen Fakultät III der Universität des Saarlandes.

Merten, J., & Krause, R. (2000). What makes good therapists fail . In: Philipot , P., Coats, E. J. , Feldmann, R.S. (Eds.). Nonverbal behaviour in clinical settings. Accepted for publication.

Schulz, S. (2001). Affektive Indikatoren struktureller Störungen. Dissertation an der philosophischen Fakultät III der Universität des Saarlandes.

Merten, J. (2001b). Context analysis of facial affective behaviour in clinical populations. In Katzekitis, M. (Ed.) Human facial measurement and meaning. Clover Academic Publishers.

Schwab, F. (2001). *Affektchoreographien. Eine evolutionspsychologische Analyse von Grundformen mimisch-affektiver Interaktionsmuster*. Berlin: Dissertation de Verlag im Internet.

Steimer-Krause, E. (1994). Übertragung, Affekt und Beziehung. Theorie und Analyse nonverbaler Informationen schizophrener Patienten. Bern. Peter- Lang- Verlag.

QuickFACS a method developed by the Saarbrücken Research Group to tackle the problem of sampling, essential within clinical research. With structurally disturbed people the problem is easier to handle because they are more homogeneous. With hysteric as well as some borderline patients the patterns change remarkably with changing identifications, the duration of which might be very short.

QuickFACS allows a description of the raw distribution patterns of affects. It is done in the following way: two coders look at real-time up to three times through the video using the AU Table for Affects trying to find an affect pattern according to the list. If both detect one it is registered as an affective event. A real content related coding has to be done using EMFACS or FACS.

Additionally the following indices where developed by the Saarbrücken Group:

a) Idiosyncrasy level (percentage of uninterpretable action units related to the affect: (Highly relevant for alexithymia as well as low level psychosomatic personality structure)

b) Repertoire affect describes sequential and nomothetical variability as well as lead-affect (the affect with the highest percentage) This is very relevant for inferences on the structural level (disgust as lead affect for the lowest, and psychosomatic structure, contempt for low structure and psychotic) Description of the relation of affect according to hedonic, information processing unhedonic affect. Dyadic indices defined as synchronisation reactions. Very relevant for the identification of schizophrenics and the prediction of psychotherapeutic outcome (Steimer-Krause, 1994).

c) Another dyadic Index is called attunement using gaze and speech- patterns in dyadic interaction. Inferences ca be made on different structural level of personality organisation, especially the absence respectively presence of the capacity to mentalize. The methodology can be found in Merten (2001b). A software tool for graphical context analysis is described there.

In addition a pattern recognition technique by Magnuson has been successfully used on dyadic facial and context patterns to described different forms of successful and unsuccessful short psychotherapy treatments as well as different forms of interaction patterns of healthy and mentally disturbed people (Schwab, 2001). This tool can be downloaded on the web from http://www.dissertation.de/html/schwab_frank.htm. Methodology for pattern detection within the psychotherapy-process- research can be found in detail in Merten (2001a).

A brief English summary can be found in Merten (2001b). In general caution has to be taken because the validity often mistakenly takes for granted inferences on feelings or internal states. This is usually not the case. The only inference which can be made and which is validated is in the field of attribution of emotion to other people. This involves seeing the action unit patterns in a very high percentage of different people, disregarding race, sex and status. Inferences on other sub-systems of the emotion-like feeling of physiology need additional context information like the above-mentioned case and speech patterns as well as the content of speech. It can be shown that the algorithm connecting he different modules of the emotion system differ according to psychiatric disturbance. By the end of a successful psychotherapy, one can observe a change in the way that facial affect is attributed to mental objects. In unsuccessful treatments, facial affects remain interactive.

C. Other Measures Referenced in This Review

Some further instruments that have been used in the studies reviewed below are listed here:

Structured instruments

Goal Attainment – Individual Treatment Goals

Heuft, G. and colleagues (Heuft et al., 1996) Langzeitoutcome ambulanter psychoanalytischer Psychotherapien und Psychoanalysen. *Forum Psychoanal., 12*, 342-355.

Therapist's Ratings of Clinical Findings

Rudolf, G. (Rudolf, 1981). Untersuchung und Befund bei Neurosen und Psychosomatischen Erkrankungen. Materialien zum *Psychischen und Sozial-Kommunikativen Befund* (PSKB). Beltz, Basel.

Operational Psychodynamic Diagnostics

Cierpka, M. and colleagues (Cierpka, 1995). Die erste Version einer Operationalisierten Psychodynamischen Diagnostik (OPD-1). *Psychotherapeut, 40*, 69-78.

Rudolf, G. and colleagues (Rudolf, 1995). Struktur und strukturelle Störung. Zsch. *Psychosom.Med. 41*, 197-212.

Arbeitskreis OPD (Hrsg). (1996). *Operationalisierte Psychodynamische Diagnostik*. Grundlagen und Manual. Hans Huber, Bern-Stuttgart.

CHAP : Change after psychotherapy

Sandell, R. (1987a). Assessing the effects of psychotherapy II. A procedure for direct rating of psychotherapeutic change. *Psychotherapy and Psychosomatics, 47*, 37-43.

Sandell, R. (1987b). Assessing the effects of psychotherapy III. Reliability and validity of "Change after psychotherapy". *Psychotherapy and Psychosomatics, 47*, 44-52.

The Hampstead Child Adaptation Measure

Psychometric Instruments:

Patient focussed self-report instruments

- Symptom-Check-List SCL 90-R (Derogatis et al., 1974)
- Beck-Depression Inventory BDI (Beck et al., 1961)
- Inventory of Interpersonal Problems IIP (Horowitz, Rosenbery et al., 1988)
- Introject Questionnaire INTREX (Benjamin, 1974; Tress, Benjamin, 1991)
- Questionnaire of Coping Strategies FKS (Hentschel, 1995)
- Questionnaire of Social Satisfaction SOZU (Sommer, Fydrich, 1991)
- Basic Documentation (Broda, Dahlbender, Schmidt, von Rad, & Schors, 1993)

- Freiburg Personality Inventory FPI-R (Fahrenberg et al., 1985)
- Narcissism Inventory (Deneke & Hilgenstock, 1988)
- Symlog: social interaction in small groups (Bales and Cohen, 1982)
- Inventory of Quality of Life (Huber et al., 1988)
- Inventory of Change in Experience and Behaviour VEV (Zielke, Kopf-Mehnert, 1978)
- Helping Alliance Questionnaire HAQ-P (Bassler et al., 1995)
- AIR questionnaire (Roose et al., 1994)

Therapist focussed psychometric instruments

- Short documentation of the initial interview (diagnosis, psychodynamic hypotheses, aims, assessment of level of personality organisation, of basic conflicts, of main defences).
- Helping Alliance Questionnaire: HAQ-T (Luborsky et al., 1996)
- Process-Rating-Scales (with questions concerning transference, resistance, analytic work, technique, setting, relevant hours, counter-transference and main transference themes)
- Global Assessment Scale (Endicott, Spitzer, Heiss, & Cohen, 1976; Luborsky & Bachrach, 1974)
- Level of Functioning Scale (Carter and Newman, 1980)
- Life Functioning Scales (Howard et al., 1993)
- Therapeutic Assets Questionnaire (Daskovsky, 1988)
- Personal Style - Therapist Form (Howard et al., 1988)
- Therapeutic Contract Questionnaire – Session Form (Howard et al., 1988)
- Therapeutic Procedures Inventory – R (Orlinsky, 1987)

Independent researcher based assessments:

Clinical diagnostic assessment interview (audio taped) as a basis for:

- ICD-I0 Diagnosis check list IDCL (Hiller, Zaudig, Mombour, & Bronisch, 1993)
- Impairment Severity Scales BSS (Schepank, 1995)
- Global Assessment of Functioning Scale GAF (DSM III-R) (Luborsky & Bachrach, 1974)
- Scales of Psychological Capacities SPC (Wallerstein, 1992)
- Goal Attainment Scaling GAS (Kiresuk and Sherman, 1968; Kiresuk et al., 1994)
- OPD-Assessment of level of personality organisation, basic conflicts, main defenses and psychodynamic hypotheses
- Computer-based quality evaluation in routine practice (System AQUASI, Kordy, 1997; Scheidt & Wirsching, 1998).