Psychoanalytic Inquiry

A Topical Journal for Mental Health Professionals ISSN: 0735-1690 (Print) 1940-9133 (Online) Journal homepage: <http://www.tandfonline.com/loi/hpsi20>

Do the Processes of Psychoanalytic Work Lead to Benefit? Studies by the APS Research Group and the Psychoanalytic Research Consortium

Sherwood Waldron, Francesco Gazzillo & Karl Stukenberg

From the special issue of *Psychoanalytic Inquiry:* Volume 35, 2015 - Issue supp1ement:

 **Toward an Empirically Informed 21st-Century Psychoanalysis:**

**Challenges and Opportunities**

Journal homepage: https://doi.org/10.1080/07351690.2015.987602

**There is a need for careful study of the relationships between the psychoanalytic/psychodynamic processes and their outcomes, yet the raw data to accomplish this study (i.e., psychoanalyses and psychoanalytic psychotherapies fully recorded and transcribed, and empirical instruments for assessing from an analytic perspective both the processes and results of these treatments) have been limited. Two related strategies to solve this deúcit have been developed over the past 30 years by the Psychoanalytic Research Consortium (PRC): (1) the collection of an increasing database currently holding 27 fully recorded psychoanalyses and a number of long-term psychoanalytic psychotherapies safeguarded, confidentialized, and made available to the field; and (2) the development of new measures for assessing processes and outcomes of treatments from an analytic point of view, such as the Analytic Process Scales (APS), the Dynamic Interaction Scales (DIS), and the Personality Health Index (PHI) with RADIO categories. This article summarizes the history of the PRC and some of the findings of the research conducted with the APS, DIS, and PHI on the PRC cases.**

Substantial clinical and empirical evidence support the idea that, for many patients, psychoanalysis and long-term psychoanalytic therapy lead to substantial improvements in quality of life (see Bachrach et al., 1991, for review; and for more contemporary reviews and meta-analyses see also Abbass et al., 2014; Leichsenring and Rabung, 2008, 2011; Shedler, 2010; Leichsenring, 2010). What has remained uncertain is the role of specifically psychoanalytic techniques in contributing to positive change (Waldron, 1997; Ablon and Jones, 1998; Wampold, 2001; Høglend at al., 2011). The evidence has remained relatively thin and, to some extent, contradictory, about such important basic aspects as the ameliorating role of accurate interpretations, with a number of studies not supporting interpretation as contributing differentially to psychological change (Lambert, 2013). However, such studies often have substantial methodological limitations that limit their generalizability and ecological validity. In fact, other studies point to an important role of interpretation (Caston, Goldman, and McClure, 1986; Waldron, Scharf, Hurst, et al., 2004; Høglend et al., 2007; Waldron et al., 2011; Høglend et al., 2011).

The field requires demonstration of the efficacy of specifically psychoanalytic technical contributions to benefit, and the elucidation of when and to what degree the long-term benefit is greater than that derived from other therapies. It is helpful always to keep in mind the null hypothesis: that healers throughout history have apparently succeeded if they are well regarded by their community, have some kind of expertise also recognized by their community, and have a positive relationship with the ailing person, as so well described from a cross cultural and contextual perspective by Jerome Frank (1973). The conviction of its practitioners is not sufficient, hence the need to study the admittedly complex relationships between processes and outcomes of psychoanalyses. Studying these relationships, however, has been greatly limited by the lack of recorded cases available to researchers. A number of psychoanalytic thinkers have addressed this problem through the years by collecting recordings, most notably Dahl, Kächele, and Thomä (1988); Luborsky and colleagues (e.g., Luborsky et al., 2001; Roy et al., 2009); Gill and colleagues (Gill, 1982; Gill and Hoffman, 1982); and Waldron and colleagues (Waldron et al., 2011; Waldron et al., 2013; Waldron, Scharf, Crouse, et al., 2004; Waldron Scharf, Hurst, et al., 2004). Several others have tried but been less successful in making collections available for study. So, despite these efforts, the number of complete analyses available remains small.

A central difficulty in studying long-term psychodynamic psychotherapy and psychoanalysis is acquiring sufficient raw data. Over the course of many years spent attempting to recruit psychoanalysts to record their work, Waldron has found several factors contributing to this difficulty, including: fears about managing the technical aspects of the recording, the difficulty of gathering such data over a long and unpredictable period, and the fear that confidentiality will be compromised. But the fundamental problem appears to be a strong prejudice against the recording process itself. This can materialize as a fear that the recording will negatively impact the sessions; perhaps through the ongoing imagined presence of an anonymous all-hearing but judging third. Others have expressed concern that that those listening to the recordings will fail to appreciate the subjective experience of both the analyst and analysand. These prejudices have lessened, at least among those analysts who listen to the work; to clinical ears, the work of recorded analysts has not been distinguishable from that which occurs in the absence of recording. Jeremy Safran, who regularly listens to recorded psychotherapies as part of his research, noted something with which we concur; listening to the actual sessions brings out an empathic response on the part of researchers; they hear and understand the human dilemmas of treaters and patients, and this helps them in their practice have more latitude to feel permission to experiment with authentic ways of connecting (Safran, personal communication to Stukenberg, 2014). Although difficulties collecting the data have turned out to reùect severe hesitancy on the part of therapists, particularly psychoanalysts, to show their work unedited, once there is a will, the way is easy, particularly in today’s computerized world, where a web cam can be attached to any computer, inexpensive software can be obtained to audio-record sessions, and one external hard drive can store all the recordings of many long-term treatments. Finally, transcribing selected sessions and confidentializing both transcripts and audio files can be reliably accomplished for about $150 per session.

1. BRIEF HISTORY AND DESCRIPTION OF

THE PSYCHOANALYTIC RESEARCH CONSORTIUM

Thirty years ago, Hartvig Dahl, a pioneering psychoanalytic researcher-clinician best known in connection with his studies of Mrs C. (e.g., Dahl et al., 1988, pp. vii-xvi, 15–28, 51–66, 109–116),

 came to a meeting of the Committee on Scientific Activities of the American Psychoanalytic Association. with a problem. “What should happen,” he said, “to recordings of psychoanalysis, made for research purposes, after the analyst-investigator is no longer active or dies?” Other members of the committee turned to Waldron because he was setting up a study of recorded psychoanalyses. Out of a subsequent subcommittee came the Psychoanalytic Research Consortium (PRC), an independent not-for-profit organization to collect, safeguard, confidentialize, and distribute recordings of psychoanalyses and psychodynamic therapies to qualified researchers and teachers. The Fund for Psychoanalytic Research of the American Psychoanalytic Association funded the legal costs to incorporate (completed in 1989) and to develop consent and confidentiality agreements. Leaders of the American Psychoanalytic Association wanted to avoid any legal liability they feared might occur from connection with the PRC, so it was set up independently, with a board of directors that included Robert Wallerstein as Vice President, and Stephen Firestein, Robert Galatzer-Levy, Alice Brand Bartlett, Marianne Goldberger, David Hurst, and Leonard Horwitz as directors. At that time, Merton Gill had already collected an extensive set of recordings of analyses and psychotherapies, reflected in a variety of publications (e.g., Gill, 1982; Gill and Hoffman 1982). He decided to contribute his collection to the newly incorporated PRC. One of the cases he contributed had been fully transcribed, and by scanning the 324 sessions of hard copy, the PRC has available the text of an entire analysis, which is searchable for specific words. Decades later, Lester Luborsky contributed his extensive collection, as well. Meanwhile, Waldron recorded his own psychoanalytic and psychotherapeutic work to add to the collection. Currently, the PRC has 27 fully recorded psychoanalyses and a variety of psychodynamic psychotherapies and partially recorded psychoanalyses. These analyses are conducted by six different analysts starting from 1968 to 2011, and the patients treated by them showed a variety of nonpsychotic clinical disorders (mood, anxiety, paraphilic and somatoform disorders, and so on), several personality disorders (narcissistic, borderline, histrionic, passive aggressive, and so on), and different levels of personality organization. Most of the analyses recorded had good outcomes (Waldron et al., 2013; Gazzillo et al., 2014; see also table 3), but there were also some poor outcome analyses. Terminations varied from mutual consent to drop-outs.

The original intention of the Board of Directors of the PRC was to collect a broad sample over years. The difficulty changing attitudes of analysts, even including some members of the Board of Directors, to recording their own work prevented the accomplishment of this goal. It is young analysts, many of whom had trained as psychologists and had recorded their own work, who have shown the liveliest interest, but the final step of actually becoming comfortable with introducing recording still represents a hurdle. At the PRC web site (www.psychoanalyticresearch.org) there is a detailed description of how one can accomplish the introduction of recording with minimal distress to analyst and patient.

In brief, a two-step process to recruit analysand participants works best. When a patient first arrives for a consultation, the analyst can say: “Before we begin, I would like to record the sessions, unless you would prefer me not to” or words to that effect. Waldron has found that 80% of his new patients are comfortable with this. Subsequently, if an agreement is reached about regular treatment, the analyst can mention that, “in addition to the utility of the recordings for me, they can serve a research purpose for those who are comfortable with it, and here is a description and consent form which you can look over at your convenience and let me know your preference.” (The model explanation and consent form is on the web site.)

The patient receives two copies of the form at that point, and does what s/he wants with it. About 90% of Waldron’s patients who had already consented to recording the initial session agreed to the research use of the recordings. Of course, the patient also has read in the consent form that they may withdraw their consent at any time.[[1]](#footnote-1)

Apart from the completely transcribed treatment of Merton Gill mentioned previously, most sessions of most of the treatments of the PRC database are not yet transcribed. As is described in the following, we are completing transcription of a sample of 20 sessions from early, middle, and late in the analyses of each of the 27 completely recorded analyses. As other researchers utilize the collection, their additional transcriptions will be added to those already available. Currently, the PRC has received some funding to digitize the extant collection, which is deteriorating with time. PRC activities also include providing confidentialized sessions used by teachers at several US and Italian psychoanalytic institutes (e.g., Karp et al., 1993; Firestein 2001).

THE APS RESEARCH GROUP—ASSESSMENT AND RESULTS

Development of the APS

In 1985, a small group of experienced analysts, led by Waldron,[[2]](#footnote-2)began meeting to develop scales aimed at the empirical assessment of the core dynamic features of patient and analyst contributions in psychoanalytic psychotherapy and psychoanalysis. In more than fifteen years of work and discussion, this group developed the Analytic Process Scales (APS: Waldron, Scharf, Crouse, et al., 2004; Waldron, Scharf, Hurst, et al., 2004). The APS are 32 five-point Likert scales aimed at the assessment of both patient and therapist contributions to the psychotherapeutic process. The group developed an 81-page coding manual with examples at the scale points to illustrate each dimension via clinical vignettes (Scharf et al., 1999/2010). As well, the group refined methods of segmenting the sessions into meaningful units that could be rated, to make detailed study of the immediate impact of analyst communications on the patient within particular sessions (Waldron, Scharf, Crouse, et al., 2004).

The APS group hoped to tease out those elements in the therapeutic communication that might indicate a useful therapeutic or analytic process, and thereby distinguish successful from unsuccessful therapeutic engagements. In assessing the analyst’s contribution, ve “core analytic activities” were dimensionalized, including clarifying, interpreting, addressing defenses, reactions to the analyst or analytic situation, and addressing conflict (Waldron, Scharf, Hurst, et al., 2004, p. 450). Two other analyst activities were considered important: the analyst addressing developmental issues and addressing the patient’s self-esteem problems. The analyst’s encouraging elaboration and providing support—two crucial activities in any psychotherapy—were also assessed. More interactional variables included how confrontational the analyst was and how

 much feeling the analyst expressed, whether friendly or hostile. As is clarified later, the most important dimension was the assessment of the *goodness* of the analyst’s communication, which was defined in terms of aptness of the communication, the potential usefulness of its content, and the skill of its presentation. The skill of the presentation is scored higher when the communication is more tactful; well timed; and its language is more clear, vivid, or likely to appeal to the patient.

The patients’ contributions were also assessed via several scales. In what could be considered a variation on the theme of experiencing scales, the group evaluated the degree to which the patient conveyed his or her experiences, reflected upon them, and expressed feelings so that the clinician rater could understand the patient better. Those assessments were each made in terms of communications about the analyst or analytic situation or about the rest of the patient’s life. There were also variables that appropriately matched the analyst variables, such as communications about the developmental years and about self-esteem issues. Last, the quality of the patient’s communications was assessed, both overall in the variable called *patient productivity*, and in response to the analyst’s communications. This patient productivity variable was defined in terms of depth or breadth of the patient’s or rater’s emotional understanding, in the intensity of the patient’s involvement and collaboration with the therapist, and in the quality of other emotional expressions (Scharf et al., 1999/2010).

The APS was originally intended to be applied to each segment (as appropriate), but also to the entire session (Waldron, Scharf, Crouse, et al., 2004; Waldron, Scharf, Hurst, et al., 2004). The division into segments is often at the point of a change in speaker, resulting in *therapist segments* and *patient segments*. When there is a rapid exchange between patient and analyst, a segment may include several changes of speaker and is considered a *joint segment*. In general, a single session contains from 8 to 30 segments. The APS ratings can be used for (a) *describing* the analytic work of sessions, periods of therapy, and entire therapies; (b) *comparing* different sessions, periods, and therapies along many dynamically relevant dimensions; and (c) *exploring* the relations among the patient’s and therapist’s contributions to the clinical process. The whole-session assessment method can be easily applied to larger numbers of sessions than the much more time-intensive segmental ratings, but is not as well suited for the micro-analytic exploration of the interaction between therapist and patient. We have studied the APS and found it be a reliable measure of process. The median intraclass correlation coefficient (ICC) of the APS *patient* scales applied by three independent raters to 120 sessions from six different treatments was .76; the median ICC for APS *therapist* scales was .77 (Gazzillo et al., 2014).

INITIAL RESEARCH RESULTS WITH THE APS

The APS research group initially applied the APS to segmented analytic sessions. We wanted to find which of the process variables best predicted patient productivity, one of the process variables we were measuring. The underlying assumption here is that productivity would, in the long run, lead to better outcomes. As the research matured, we moved to listening to whole sessions rather than segmented sessions, and we began to measure outcome more directly, in a variety of ways that are described in the following. The direct measurement of outcome allowed us both to observe the relationship between the quality of the processes of analyses and the outcomes, and also to test the notion that patient productivity was a good predictor of outcome at the end of treatment. Fortunately, that proved to be the case. The review following highlights the

 ways in which we have come to see the relationship between the process scales and the outcomes of treatment, both as measured by patient productivity in the session and by more conventional methods of psychotherapy outcome assessment.

When the APS research group applied the APS in its segmental version to nine sessions from three different analyses in the PRC collection, for an overall total of 150 patient segments and 128 therapist segments (Waldron, Scharf, Crouse, et al., 2004; Waldron, Scharf, Hurst, et al., 2004), they found a significant correlation between the *summed core analytic activities* of the therapist (clarification, interpretation, addressing defenses, transference and conùicts) and the *overall productivity* of the patient’s subsequent communication ( *r* = .25, a small to medium effect size). This result clarified that good psychoanalytic process predicted patient productivity, something that we certainly expected to be the case. In the next study of 1,173 segments from 81 sessions of three different psychoanalyses, other dimensions of the therapist communications, such as their supportive quality and their being centered on developmental issues, were significantly correlated with the overall productivity of the patient’s next communications (Waldron and Stukenberg, in preparation). Again, and consistent with prior research (Wallerstein, 1986), we were seeing that the quality and type of psychoanalytic intervention predicted the patient’s productivity; in our case on a moment to moment basis during the psychoanalytic hours.

In this and subsequent studies, the most global analyst variable, *the goodness of the therapeutic communication,* as rated by clinician-raters, seemed to be the single most powerful predictor of subsequent patient increase in productivity, irrespective of the particular type of therapist communication (Waldron, Scharf, Crouse, et al., 2004; Waldron, Scharf, Hurst, et al., 2004; Gazzillo and Lingiardi, 2007; Lingiardi, Gazzillo, and Waldron 2010). In other words, the capacity of a patient to better understand himself/herself, to better express his/her emotions, and to better collaborate in the therapeutic enterprise (i.e., the productivity of a patient’s communication as assessed by the APS patient scale) is substantially influenced by the capacity of the therapist to “say the right thing at the right time” (Waldron, Scharf, Hurst, et al., 2004), i.e., by the aptness of the kind, content, timing, and other characteristics of his/her interventions.

This was a robust result. We controlled for the role of the patient’s productivity in the previous segment by computing the partial correlation between therapist’s intervention goodness and the overall productivity of the following patient’s communications. The partial correlation in the first APS study was 0.44 (Waldron, Scharf, Hurst, et al., 2004); in subsequent studies it ranged from 0.19 to 0.45 (see Figure 1 for a graphical example of the association between the goodness of therapist’s intervention and overall therapeutic productivity of patient’s communications, correlating 0.38 for this patient). On the other hand, when controlling for the influence of the goodness of therapist’s communications, the core analytic activities show a modest or absent influence on the productivity of the following patients’ communications. That is, the *quality* of the intervention appears to be a better predictor of good analytic work than the *type* of intervention.

This suggests that, rather than a technical aspect of the analytic encounter (e.g., the use of interpretation or clarification) accounting for greater productivity, there was something at work that cut across techniques. Our hypothesis was that the clinician raters were responding, as they listened to the therapeutic communications, to relational aspects of the analysts’ communications to their analysands, and the relational attunement of these communications was reflected in higher ratings of their goodness. Our finding then was entirely consistent with the psychotherapy research data, which consistently stress a positive relationship between the quality of the therapeutic relationship and outcome (Norcross, 2011).

The goodness of therapist communication is, in turn, positively influenced by the overall productivity of the previous patient’s communication, and this last variable positively influences also the productivity of the next patient communication. But, as already stressed, even controlling for the influence of the productivity of the previous patient’s communications on the subsequent therapist’s intervention, and the influence of all the other therapist variables on the productivity of the following patient’s

FIGURE 1 Patient productivity and therapist good communication (“Penny”).

 communication, the correlation between the quality of analyst intervention and the productivity of the following patient’s communication remains positive and significant, a result obtained both in the US and in the Italian studies, and in the small comparative study of the core analytic activities and therapist communication quality of short-term psychodynamic and cognitive behavioral therapy cases (Waldron and Helm, 2004). Good patients produce good analytic work—but they produce better work when their work is accompanied by timely and apt analytic interventions.

All this research is based on the assumption that if the patient is productive, then the therapy will have a good outcome. We decided that we needed to test that assumption. We did this by developing a measure of outcome that we could derive through listening to the sessions as we had no measures in our existing database that measured this. We also wanted to test the clinical significance of our hypothesis that highly rated analyst communications, leading to greater patient productivity, were characterized by a good relational attunement. So we developed the scales outlined in the next two sections.

Development of the Personality Health Index (PHI) and Subscales: Reality Functioning, Affective Regulation and Tolerance, Defensive Organization, Identity Integration and-Object Relations (RADIO) Scales

After the APS group studied current outcome measures, and found them limited, the decision was made to attempt to adapt a well-established procedure, the Shedler-Westen Assessment Procedure-200 or SWAP-200 (Westen and Shedler, 1999a, 1999b; Shedler, 2002), as a means by which to provide a more precise measure of mental health throughout the entirety of treatment. The SWAP utilizes a Q-Sort method, which requires the rater to assign 200 items to eight

 piles according to their descriptiveness of patient’s personality and following a fixed distribution (Block, 1978, 2008; Shedler and Westen, 2007; Westen and Shedler, 1999a), which helps to improve the reliability of ratings. Each of the SWAP’s items provides a description of an aspect of the personality functioning of a patient. The rater assigns a score from zero to seven to each, according to how descriptive the item is in that particular patient’s functioning (0 = *not descriptive*,7 = *most descriptive*). Only eight items can be given the highest score of 7, 10 are scored 6, 12 are scored 5, and so on; 100 items have to be scored 0. This procedure results in the 100 chosen items being distributed approximately as the right half of a bell-shaped curve. The items are straightforward, jargon-free clinical statements about the individual being rated. Training for raters is minimal, as long as they are experienced therapists. The SWAP items are most conveniently sorted on a computer program (Shedler, 2002). The software provides a profile of twelve personality prototypes that closely parallel DSM-IV Axis II diagnoses. This procedure was originally developed from a nationwide survey of clinicians experienced in the treatment of patients with personality disorders who were asked to respond to sample item sets. Its reliability and validity have been reported (Westen and Shedler, 1999a, 1999b) and extended over the past ten years (Lingiardi, Shedler, and Gazzillo, 2006; Westen and Muderrisoglu, 2006; Shedler and Westen, 2006). The SWAP provides a common vocabulary that organizes clinical observations and inferences about a patient’s personality and provides a snapshot of a patient’s psychological functioning; although consisting of only 200 items, the SWAP provides nearly innumerable permutations that permit the capture of complex patterns (Shedler, 2002). Further studies (e.g., Lingiardi et al., 2006) have shown that the thirty items assigned to the three most salient categories provide a useful summary of patient functioning. Twenty-four of the 200 items, moreover, express aspects of positive mental health, providing a corrective to the pathology-oriented bias mentioned earlier.

For all these reasons, the SWAP is well suited to evaluate the changes occurring over the course of therapy, but does not spell out very fully the degree of change and it is not specifically aimed at the assessment of personality changes of patients in psychoanalyses. Consequently the APS group developed two indices for these purposes. The first index, the Personality Health Index (PHI), provides a global assessment of personality health based upon the 200 clinical judgments comprising the SWAP. The index is based upon the percentile standing of the patient’s PHI at that point in time, which is referenced against a national sample of 308 SWAPs from different points in the treatments of 70 psychoanalytic patients (Cogan and Porcerelli, 2004, 2005). The percentile standing provides a ready and easily understandable reference point. Details of its development have been fully described, along with evidence for its validity (Waldron et al., 2011).[[3]](#footnote-3)The PHI provides an overall measure of personality functioning, similar to the HealthSickness Rating Scale (Luborsky 1962) or the derived Global Assessment of Functioning (GAF: American Psychiatric Association, 2000), but with a number of advantages. The PHI percentile score is based upon a normative sample of patients in psychoanalysis; it is based upon specific clinical judgments across 200 psychological variables (items), which provide greater precision than a single global assessment; it describes aspects of personality functioning rather than symptomatic distress or behavioral functioning; and, finally, the PHI provides a statistical comparison

 of a given patient’s profile with a nationwide sample of psychoanalytic patients’ profiles, making this instrument much more useful for purposes of psychoanalytic and psychotherapeutic research.

The second measure derived from the SWAP delineates an individual’s particular strengths and difficulties within five central domains of personality functioning: Reality testing, Affect regulation and tolerance, Defensive organization, Identity integration, and Object relations (hence the acronym RADIO). The APS modification of the SWAP software derives these indices automatically whenever a set of SWAP scores is entered, and automatically lists the most prominent strengths and weaknesses in each of these RADIO areas. Although these indices have obvious applications for psychotherapeutic and psychoanalytic research, they comprise a useful resource for clinical work and training as well (Waldron et al., 2011).

In a study of three cases (Waldron and Stukenberg, 2010, unpublished study)—early, middle, and late, totaling 81 sessions and 1,173 segments—the APS group demonstrated that the assessment of one week’s treatment at a given point in an analysis using the PHI and RADIO appears to give a reliable measure of his/her level of functioning (PHI) and descriptive diagnosis (RADIO); that is, scores from adjacent weeks are quite similar to each other. That said, Waldron also found that there is enough variation between the weeks that two weeks’ work at any given phase of treatment is more precise than one week to characterize the patient’s level of functioning. It is also desirable to rate two weeks, rather than one, to achieve a more reliable representation of the process (APS).

The Dynamic Interaction Scales (DIS)

There has been an increasing shift from a more classical theory of technique to a more relational one in understanding the therapeutic situation, and because we suspected that part of what was driving increased patient productivity were relational components of the interaction, the APS group also decided to develop an instrument that would more directly assess aspects of the relationship than the APS was designed to do. This would hopefully serve to unpack the quality of the analyst’s communication as judged by other clinicians. The results of this effort are the Dynamic Interaction Scales (DIS; Waldron et al., 2013). The DIS provides a more holistic, interactional set of measures than the APS, as described in the following.

Twelve aspects of the ongoing psychoanalytic or psychotherapeutic process were identified by the APS study group as reflecting dimensions of how the two individuals were working together and relating to one another: Therapist variables include to what degree the therapist is straightforward, warmly responsive with the patient, responds moment to moment to patient’s feelings, conveys aspects of his subjective experience in response to the patient’s material, and addresses the patient’s typical patterns of relating and feeling. Variables focused on the patient include: to what degree the patient flexibly shifts between experiencing and reflecting, to what degree is there a flexible interplay on the part of the patient between conscious waking life and dreams and how well is the patient working on his/her typical patterns of relating and emotions. Variables reflecting both therapist and patient include: to what degree does the patient experience the therapist as empathic, to what degree does the therapist’s contribution lead to the further development of the patient’s awareness of his or her own feelings, to what degree is there an integration of understanding of the relationship with the therapist to other relationships, and to what degree is the engagement in the therapeutic relationship by the two parties brought forward or experienced in an emotionally meaningful way. In developing these new variables, the group followed their

 experience in developing the APS variables, defining them on the basis of accumulated clinical experience (averaging more than 35 years for each clinician), then iteratively applying them to clinical sessions and discussing disagreements, then modifying the language of the items or of their explanations. These explanations were collected in a coding manual.

The Roman Study of 27 Completely Recorded Analyses

Francesco Gazzillo and Vittorio Lingiardi, of the Department of Dynamic and Clinical Psychology, Sapienza University, Rome studied 20 sessions from the first two years of an Italian recorded psychoanalysis using the APS, the SWAP, and the Defense Mechanisms Rating Scale (Perry, 1990). In this study (Lingiardi et al., 2010), the authors intertwined a classical clinical case presentation with an empirically based description of the process and outcome of a psychotherapy, and demonstrated that the APS can help to identify the more useful interventions for a patient in the different phases of her/his therapy.

Gazzillo then assumed the leadership of a team of psychologists[[4]](#footnote-4)for the systematic study of all 27of the fully recorded cases in the PRC collection and results from 19 of those analyses have been reported to date. From each of the treatments, 20 sessions are evaluated: the first four sessions, four sessions from the sixth week, four from the middle of the treatment, four sessions from the sixth week before the termination, and the last four sessions. The first eight sessions (the first four sessions + the four sessions from the sixth week) and the last eight sessions (the four sessions from the sixth week before the termination + the last four sessions) are listened to as whole sessions and rated with the APS and DIS by three independent raters. SWAP, PHI and RADIO and GAF are rated by four different raters (two for the SWAP assessment and two for the GAF assessment) after listening to the first eight sessions, and a separate evaluation based on material in the last eight sessions is performed. Two independent raters assess the first 4 sessions also with the *Helping Alliance Questionnaire* (Luborsky et al., 1983).

The first of these cases was studied intensively (Colli et al., 2011; De Bei et al., 2011;Di Giuseppe et al., 2011; Gazzillo et al., 2011). The DIS reliabilities were excellent among the Roman raters, and in a subsequent study, the DIS Spearman’s Rho value based on forty sessions per item averaged .93, with a range of .83 to .98 (Waldron et al., 2013). After evaluating another eighteen cases, mostly with good outcomes but with a few cases that did poorly, the researchers reported that most of the APS therapist variables and all the DIS variables distinguished between the two outcome groups (Waldron et al., 2013;see Figure 2).

Preliminary results (Waldron et al., 2013) from the comparison between sixteen good outcome psychoanalyses and six poor outcome psychoanalyses (assessed over 380 sessions) demonstrate that patient productivity is indeed greater over the course of treatment in good rather than poor outcome cases (mean rank 202.1 vs 165.3; *U* =12574.5; *p* < .001), even though the two groups do not differ on this measure initially. The same is true for therapist communication goodness (mean rank 209.1 vs 150.2; *U* = 10770; *p* < .001). Patient productivity from the beginning to the termination of good outcome psychoanalyses is consistent, but it decreases in poor outcome treatments. As well, the goodness of therapist’s interventions increases in good outcome therapies, but decreases in poor outcome treatments as treatment progresses. This supports the hypothesis that continuing good quality of interventions affects the final outcome of therapy,

FIGURE 2 Early relational differences between good and poor outcome cases.[[5]](#footnote-5) creating a virtuous cycle, but it also seems to suggest that a vicious cycle of interactions can begin to take place and that the work can deteriorate. In subjectively listening to our bad outcome cases (admittedly a very small number of cases), it is apparent that both the analyst and the analysand remain connected to each other, that they are each working to effect psychoanalytic change, but that they are involved in what appears from the outside to be enactments that they seem unaware of—or at least are not articulating—in their work together. This has the quality of a Greek tragedy—one in which both participants continue to be drawn by hope for something that seems increasingly and sadly elusive in the lived relationship.

In addition to evaluating the relationship between patient productivity and outcome, we were also curious about the impact of the relational quality on the treatment. The goodness of the therapist’s contribution, the variable that was the best predictor of patient productivity, correlated significantly with all the DIS scales, and in particular with the therapist’s capacity to follow the moment-to-moment shifts in patient’s feelings and with the therapist addressing and helping the patient to address his or her typical patterns of relating and feeling (respectively, Pearson

1. = 0.63 and 0.65, *p* < .001). This seems to support our idea that the goodness of therapist contribution variable is, not surprisingly, made up of a number of constituent elements and that our clinician raters were responding to a wide range of both relational and technical components in rating the quality of therapists’ contributions.

Taking into account all the APS therapist scales and all the DIS therapist and interaction scales, the assessment of the first 19 psychoanalyses of the PRC collection suggest that the best APS intrasession indicator of personality health is the patient’s ability to oscillate between experiencing and reflecting on his or her experience (PHI-DIS Pearson *r* = 0.41). This capacity, in turn, is best fostered in the treatment by the analytic couple integrating the understanding of the transference relationship to other relationships, past and present, and their being fully engaged in the therapeutic relationship (these two scales, taken together, explain 18% of the variance of the capacity of the patient to oscillate between experiencing and reflecting). These data, if confirmed

 by further studies, support the therapeutic relevance of both relational factors (the engagement) and more classical interventions (the analysis of transference), and the idea of the analysis as a relational learning experience.

The early and late level of personality health (PHI) of 19 patients assessed so far is shown in Figure 3 (we have been referring to these data earlier in the article, but here they are being shown directly for the first time). Most of the patients benefitted substantially from their analyses, with an average PHI increase of 28 points. It is worth noting the lack of a significant correlation between early PHI and ultimate gain from the analysis: some patients who were judged as quite impaired had very considerable gains, although all but one of those who were functioning reasonably well at the beginning of treatment improved. Therefore, initial PHI, in the range found among our patients, is not predictive of the value they may derive from an analysis.[[6]](#footnote-6)

FIGURE 3 Early and late level of personality health: 19 patients of the PRC collection.

 As a final note, the APS, as a measure of the analytic or therapeutic process, can be applied to any type of therapy. In a pilot study (Waldron and Helm, 2004), we observed that two cognitive behavioral therapies had high levels of core analytic activities, although the degree to which this happened was only about half the level of the activity of the matched short-term treatment by a trained psychoanalyst. We mention this here as a kind of introduction to the next section. The APS as a scale, along with the other scales cited in this review, can be used to observe and quantify what is actually taking place, from a psychoanalytic perspective, within a wide range of psychoanalytic and therapeutic interactions.

FUTURE RESEARCH AND IMPLICATIONS

Expansion of the PRC collection is needed to make more detailed studies of the relationships among the variables. A study is in the planning stage of distance analyses carried out by American psychoanalysts of Chinese mental health professionals who are sufficiently fluent in English, using the same methodology and sampling as this Roman study. This will provide an opportunity to compare the group of analyses collected thus far—all between US analysts and analysands with both in the same room—and the analyses that are taking place between US analysts and Chinese analysands all at a distance, while expanding the PRC collection.

This expansion should partially address the limitations of the studies conducted so far with the APS and the DIS. First, they were based on a limited number of sessions, and on a limited number of treatments delivered by only a few analysts. Second, with such a small number of cases, we cannot definitively disentangle the complex relationships among process factors, changes in symptoms and relevant life-events. But, once completed, the ongoing study of the 27 psychoanalytic treatments of the PRC will be the most extensive research on process and outcome of psychoanalyses based on audiotaped and transcribed analytic sessions assessed with analytically based and empirically validated instruments. To our knowledge, in fact, no empirical research so far has been based on the careful assessment of more than 600 sessions from more than 30 analyses conducted by different experienced therapists. Moreover, most of the raters involved in this study, and all the authors of the APS and the DIS, are practicing psychoanalysts and researchers in psychotherapy, so that the instruments used in this study are well informed by the clinical experience of their creators and by their knowledge of research in psychotherapy. For these reasons, the results of this study may have relevant implications both for future research and for clinical practice.

If the empirical assessment of the remaining twelve cases will support the results briefly outlined in this and previous articles, we will be able to suggest that analytic training institutes should help their candidates to learn how and when to speak to their patients, and not only what to say to them. Further, we could suggest that working on the core transference patterns and on the integration of their understanding with the understanding of the relational patterns they show outside the therapeutic room is as important as fostering a full involvement in the therapeutic relationship. In other words, our research seems to support the idea that a good psychoanalysis can substantially help people to live better lives, but this happens when therapists and patients relate in an emotionally meaningful way, reflect on their relationship, learn from it, and the analysand applies it to his/her life. This result bridges the relevance of the therapeutic relationship, clearly supported by psychotherapy research in general (Horvath et al., 2011; Norcross, 2011), with the therapeutic value of transference interpretation, one of the main components of analytic technique (Strachey, 1934; Gill, 1982; Joseph, 1989).

Finally, if data from the last twelve treatments confirm the findings with the first nineteen, this study will have identified at least some of the therapeutic factors that predict early in the treatment whether a good outcome is likely. A desirable next step would be finding ways to apply this knowledge to improving outcomes while minimizing or altering the nature of treatment efforts which have low likelihood to provide much benefit to the patients.

REFERENCES

Abbass, A. A., S. R. Kisely, J. M. Town, F. Leichsenring, E. Driessen, S. De Maat, A. Gerber, J. Dekker, S. Rabung, Rusalovska, & E. Crowe. (2014), Short-term psychodynamic psychotherapies for common mental disorders. *Cochrane Database of Systematic Rev*., 2014(7): Art. No.: CD004687. doi:10.1002/14651858CD004687.pub4.

Ablon, J. S., & E. E. Jones. (1998), How expert clinicians prototypes of an ideal treatment correlate with outcome in psychodynamic and cognitive-behavioral therapy. *Psychother. Res*., 8: 71–83. doi: 10.1080/10503309812331332207.

American Psychiatric Association. (2000), *Diagnostic and Statistical Manual of Mental Diseases*, 4th edition, text revision. Washington, DC: American Psychiatric Publishing.

Bachrach, H., R. Galatzer-Levy, A. Skolnikoff, & S. Waldron. (1991), On the efficacy of psychoanalysis. *J. Amer.Psychoanal. Assn*., 39: 871–916. doi: 10.1177/000306519103900402.

Block, J. (1978), *The Q-sort Method in Personality and Psychiatric Research*. Palo Alto, CA: Consulting Psychologists Press.

—————. (2008), *The Q-sort in Character Appraisal: Encoding Subjective Impressions of Persons Quantitatively*. Washington, DC: American Psychological Association,.

Caston, J., R. K. Goldman, & M. M. McClure. (1986), The immediate effects of psychoanalytic interventions. In: *The Psychoanalytic Process: Theory, Clinical Observations, and Empirical Research*, ed. J. Weiss, H. Sampson, & Mt. Zion). New York, NY: Psychotherapy Research Group, Guilford, pp. 277–298, 377–379, 387–391.

Cogan, R., & J. H. Porcerelli. (2004), Personality pathology, adaptive functioning, and strengths at the beginning and end of psychoanalysis. *J. Amer. Psychoanal, Assn*., 52: 1229–1230.

—————, & J. H. Porcerelli. (2005), Clinician reports of personality pathology of patients beginning and patients ending psychoanalysis. *Psych. and Psychother.: Theory, Res. and Prac*., 78: 235–248. doi: 10.1348/147608305X28727

Colli, A., V. Condino, D. Gentile, & V. Lingiardi. (2011), Assessing changes in alliance ruptures with the Collaborative Interventions Rating Scale (CIS). Paper presented at the 42nd Annual Meeting of the Society for Psychotherapy Research (SPR), Bern, Switzerland.

Dahl, H., H. Kachele, & H. Thoma. (1988), *Psychoanalytic Process Research Strategies*. New York, NY: Springer-Verlag. De Bei, F., A. Tanzilli, N. Aioub, D. Giovannetti, D. Miccoli, & N. Dazzi. (2011), Assessing therapeutic process and changes in patient–therapist attachment with PQS and PTA Q-sort. Paper presented at the 42nd Annual Meeting of the Society for Psychotherapy Research (SPR), Bern, Switzerland.

Di Giuseppe, M., A. Codazzi, A. M. Stolfa, & A. R. Marseglia. (2011), Cambiamenti del funzionamento difensivo in una psicoanalisi di Merton Gill. Paper presented at the 42nd Annual Meeting of the Society for Psychotherapy Research (SPR), Bern, Switzerland.

Firestein, S. K. (2001), Teaching with tape-recorded psychoanalysis. *J. Clin. Psychoanal*., 11: 619–625. Frank, J. (1973), *Persuasion and Healing*. Baltimore, MD: Johns Hopkins University Press.

Gazzillo, F., & V. Lingiardi. (2007), The virtuous circle of a good analysis. Paper presented at the 38th International Meeting of the Society for Psychotherapy Research (SPR), Madison, Wisconsin.

—————, Waldron, C. Bandieri, & F. Genova. (2011), The Merton Gill Single Case Study: Assessing psychoanalytic process with the Analytic Process Scales (APS) and the Dynamic Interaction Scales (DIS). International Conference of the Society for Psychotherapy Research, Bern, Switzerland, June 29 July 2, 2011.

—————, Waldron, F. Genova, F. Angeloni, C. Ristucci, & V. Lingiardi (2014), An empirical investigation of

analytic process: Contrasting a good and poor outcome case. *Psychotherapy* [online first publication at http://psycnet. apa.org/psycinfo/2014-11299-001/].

Gill, M. (1982), *Analysis of Transference: Theory & Technique (Vol. I)*. New York, NY: International Universities Press. —————, & I. Hoffman. (1982), *Analysis of Transference: Studies of Nine Audio-Recorded Psychoanalytic Sessions (Vol. II)*. New York, NY: International Universities Press.

Høglend, P., A. G. Hersoug, K. P. Bøgwald, S. Amlo, A. Marble, Ø. Sørbey, ... P. Crits-Christoph. (2011), Effects of transference work in the context of therapeutic alliance and quality of object relations. *J. Consult. and Clin. Psych*., 79: 697–706. doi: 10.1037/a0024863.

—————, P. Johansson, A. Marble, K. P. Bøgwald, & S. Amlo. (2007), Moderators of the effects of transference interpretations in brief dynamic psychotherapy. *Psychother. Res*., 17: 160–171. doi: 10.1080/10503300701194206

Horvath, A. O., A. C. Del Re, C. Flückiger, & D. Symonds. (2011), Alliance in individual psychotherapy. *Psychotherapy*, 48: 9–16.

Joseph, B. (1989), *Psychic Equilibrium and Psychic Change*. New York: Tavistock/Routledge.

Karp, J. G., I. Hyler, M. Wald, L. Whitman, S. Herschkowitz, & M. Goldberger. (1993), The use of an audio-tapedanalysis in a continuous case seminar. *Psychoanal. Quart*., 62: 263–269.

Lambert, M. J. (2013), *Bergin and Garfield’s Handbook of Psychotherapy and Behavior Change* . Edison, NJ: Wiley and Sons.

Leichsenring, F. (2010), Evidence for psychodynamic psychotherapy in personality disorders: A review. In: *Psychodynamic Psychotherapy for Personality Disorders: A Clinical Handbook*, ed. J. F. Clarkin, P. Fonagy, & G. O. Gabbard. Arlington, VA: American Psychiatric Publishing, Inc., pp. 421–437.

—————, & S. Rabung. (2008), Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis. *J. Amer. Med. Assn*., 300 (13): 1551–1565. doi: 10.1001/jama.300.13.1551

—————, & S. Rabung. (2011), Long-term psychodynamic psychotherapy in complex mental disorders: Update of a meta-analysis. *Brit. J. Psych*., 199: 15–22. doi: 10.1192/bjp.bp.110.082776.

Lingiardi, V., F. Gazzillo, & S. Waldron. (2010), An empirically supported psychoanalysis: The case of Giovanna. *Psychoanal. Psych*., 27: 190–218. doi: 10.1037/a0019418.

—————, J. Shedler, & F. Gazzillo. (2006), Assessing personality change in psychotherapy with the SWAP–200: A case study. *J. Personal. Assess*., 86: 23–32. doi: 10.1207/s15327752jpa8601\_04

Luborsky, L. (1962), Clinicians’ judgments of mental health: A proposed scale. *Arch. of Gen. Psych*., 7: 407–417.doi:10.1001/archpsyc.1962.01720060019002.

—————, P. Crits-Christoph, L. Alexiander, M. Margolis, & M. Cohen. (1983), Two helping alliance methods for predicting outcomes of psychotherapy: Counting signs vs. a global rating. *J. Nervous and Mental Dis*., 171: 480–491. doi:10.1097/00005053-198308000-00005

—————, J. Stuart, S. Friedman, L. Diguer, D. A. Seligman, W. Bucci,... E. Mergenthaler. (2001), The Penn Psychoanalytic Treatment Collection: A set of complete and recorded psychoanalyses as a research resource. *J. Amer. Psychoanal. Assn*., 49: 217–233. doi: 10.1177/00030651010490011401

Norcross, J. C. (2011), *Psychotherapy Relationships That Work*, 2nd edition. New York, NY: Oxford University Press.

Perry, J. C. (1990), *Defence Mechanism Rating Scale (DMRS)*, 5th edition. Cambridge, MA.

Roy, C. A., J. C. Perry, L. Luborsky, & E. Banon. (2009), Changes in defensive functioning in completed psychoanalyses: The Penn Psychoanalytic Treatment Collection. *J. Amer. Psychoanal. Assn*., 57: 399–415. doi:10.1177/0003065109333357

Scharf, R. D., S. Waldron, S. K. Firestein, M. Goldberger, & A. Burton. (1999), The Analytic Process Scales (APS) coding manual. (Unpublished, latest version 2010).

Shedler, J. (2002), A new language for psychoanalytic diagnosis. *J. Amer. Psychoanal. Assn*., 50: 429–456.doi:10.1177/00030651020500022201.

—————. (2010), The efficacy of psychodynamic psychotherapy. *Amer. Psychol*., 65: 98–109. doi: 10.1037/a0018378

—————, & D. Westen. (2006), Personality diagnosis with the Shedler–Westen Assessment Procedure (SWAP): Bridging the gulf between science and practice. In: *Psychodynamic Diagnostic Manual (PDM)*, ed. Alliance Task Force. Silver Spring, MD: Alliance of Psychoanalytic Organizations, pp. 573–613.

—————, & D. Westen. (2007), The Shedler-Westen Assessment Procedure (SWAP): Making personality diagnosis clinically meaningful. *J. Personal. Assess*., 89: 41–55. doi: 10.1080/00223890701357092

Strachey, J. (1934), The nature of the therapeutic action of psycho-analysis. *Internat. J. Psycho-Anal*., 15: 127–159.

Waldron, S. (1997), How can we study the efficacy of psychoanalysis?. *Psychoanal. Quart*., 66: 283–324.

—————. (2010), Examination of Findings from Main Study 2. Unpublished manuscript.

—————, F. Gazzillo, F. Genova, & V. Lingiardi. (2013), Relational and classical elements in psychoanalysis: An empirical study with case illustrations. *Psychoanal. Psych*., 30: 567–600.

—————, & F. Helm. (2004), Psychodynamic features of two cognitive-behavioural and one psychodynamic treatment compared using the Analytic Process Scales. *Can. J. Psychoanal*., 12: 346–368.

—————, S. Moskovitz, J. Lundin, F. Helm, J. Jeremin, & B. Gorman. (2011), Evaluating the outcomes of psychotherapies: The personality health index. *Psychoanal. Psych*., 28: 363–388.

—————, R. D. Scharf, J. Crouse, S. K. Firestein, A. Burton, & D. Hurst. (2004), Saying the right thing at the right time: A view through the lens of the Analytic Process Scales (APS). *Psychoanal. Quart*., 73: 1079–1125.

—————, R. D. Scharf, D. Hurst, S. K. Firestein, & A. Burton. (2004), What happens in a psychoanalysis: A view through the lens of the Analytic Process Scales (APS). *Internat. J. Psychoanal*., 85: 443–466.

—————, & K. Stukenberg. (in preparation), Analyst activities leading to greater patient productivity: Results from patients using the Analytic Process Scales (APS).

Wallerstein, R. (1986), *Forty-Two Lives in Treatment: A Study of Psychoanalysis and Psychotherapy*. New York: Guilford Press.

Wampold, B. E. (2001), *The Great Psychotherapy Debate: Models, Methods, and Findings*. Mahwah, NJ: Lawrence Erlbaum Associates.

Westen, D., & S. Muderissoglu. (2006), Clinical assessment of pathological personality traits. *Amer. J. Psych*., 163: 1285–1287.

—————, & Shedler, J. (1999a), Revising and assessing Axis II: Developing a clinically and empirically valid assessment method. *Amer. J. Psych*., 156(2): 258–272.

—————, & Shedler, J. (1999b), Revising and assessing Axis II: Toward an empirically based and clinically useful classification of personality disorders. *Amer. J. Psych*., 156(2): 273–285.

*Sherwood Waldron, M.D. 40 East 94th Street, Apt. 11B New York, NY 10128* *woodywald@earthlink.net*

To cite this article: Sherwood Waldron, Francesco Gazzillo & Karl Stukenberg (2015) Do the Processes of Psychoanalytic Work Lead to Benefit? Studies by the APS Research Group and the Psychoanalytic Research Consortium, Psychoanalytic Inquiry, 35:sup1, 169-184, DOI: [10.1080/07351690.2015.987602](http://www.tandfonline.com/action/showCitFormats?doi=10.1080/07351690.2015.987602)

1. Patients who give such permission are intending to make a contribution to the field, and usually are aware of the insecure empirical link between technique used and benefit. A member of an Institutional Review Board has pointed out clinicians’ ethical obligation to respect the intent of such permissions and not let patients’ contribution go to waste. [↑](#footnote-ref-1)
2. The group includes or has included Anna Burton, James Crouse, Stephen Firestein, Marianne Goldberger, Fonya Helm, David Hurst, John Jemerin, John Lundin, Seymour Moscovitz, Robert Scharf, Jonathan Shedler, and Kenneth Winarick, as well as the authors. [↑](#footnote-ref-2)
3. The underlying rationale stemmed from finding a high degree of agreement in clinicians’ judgments about how healthy or ill a given patient was, based on the highest scoring items in their SWAP, and also a high degree of agreement among clinicians about the likely overall health significance of each of the SWAP items. [↑](#footnote-ref-3)
4. Federica Genova, Federica Angeloni, Chiara Ristucci, Valentina Mellone, and others. [↑](#footnote-ref-4)
5. Scores are averages of the early sessions (first 4 then 4 six weeks later) in each case. [↑](#footnote-ref-5)
6. This finding is of course tentative, based on a small number of cases, and there are other aspects that need to be taken into account, such as the “ceiling effect” of the initial PHI values. For instance, one patient had an initial PHI of 88, and a final PHI of 100, hence could not improve more than 12 points. [↑](#footnote-ref-6)