Chapter 27
Ten Diverse Outcome Measures for Psychodynamic Psychotherapy Research

Caleb J. Siefert and Jared A. DeFife

Keywords Outcome measures • Outcome research • Process research • Psychodynamic psychotherapy • Psychotherapy outcome • Treatment outcome

Introduction

Understanding the changes arising from psychotherapy is an important area considered by psychotherapy researchers across treatment orientations. This type of research focusing on treatment effects is referred to as outcomes research. When assessing treatment outcomes, it is important to select measures that evaluate the type of changes a given therapy is expected to bring about. Thus, it is important that psychodynamic psychotherapy researchers have instruments that measure constructs that are expected to change as a function of psychodynamic therapy. A number of measures have been developed precisely for this purpose. This has fostered the growth of multiple programs of research dedicated to studying psychodynamic theory, techniques, process, and outcomes [1, 2]. Such studies have helped to establish the overall efficacy and benefits associated with psychodynamic forms of treatment [3]. Outcomes research compliments other forms of psychotherapy research associated with understanding therapy process, identifying what works best for whom, and determining what specific interventions are the active ingredients in facilitating specific psychological change [4].

Traditional outcome measures have historically focused on symptoms. Treatments are expected to reduce symptoms and suffering in individuals seeking therapy for psychopathology. There are numerous measures typically employed to study symptom changes. Many of these are multi-item, multi-scale measures tapping a broad range of psychopathological symptoms and difficulties such as the Personality Assessment Inventory [5]. While psychodynamic psychotherapy has been associated with symptom relief, it also seeks to facilitate other changes such as personality change, changes in interpersonal functioning, improvements in coping abilities, and improved self-understanding and sense of coherence [1–4, 6–8]. Such changes are not always easily measured. However, a number of

[AU1] C.J. Siefert, Ph.D. (✉)
Department of Psychology, University of Michigan—Dearborn, Dearborn, Michigan, USA
e-mail: csiefert@umd.umich.edu

J.A. DeFife, Ph.D.
Research Scientist, Department of Psychology, Emory University, Atlanta, Georgia, USA
e-mail: jdefife@emory.edu

R.A. Levy et al. (eds.), *Psychodynamic Psychotherapy Research: Evidence-Based Practice and Practice-Based Evidence*, Current Clinical Psychiatry,
DOI 10.1007/978-1-60761-792-1_27, © Springer Science+Business Media, LLC 2012

measures exist to capture such changes, and many of them have been developed by psychodynamic investigators for the purpose of outcomes research. In the remainder of this chapter, we provide a brief review of ten measures likely to be of use to psychodynamic psychotherapy researchers. We have included a range of measures that differ considerably with regard to method of measurement (e.g., self-report; expert coded), breadth of assessment (e.g., global character; specific construct), concept assessed, history of use in psychotherapy research, and requirements for training. For each measure reviewed, we have attempted to include a review of the purpose of the measure, discussion of the scales and scores the measure produces, information on administration (e.g., self-report; narrative based; interview based), details about each measure's psychometric properties, and, when applicable, how the measure has previously been used to study psychodynamic psychotherapy. This chapter is in no way comprehensive, ~~and there are~~ a number of appropriate measures ~~that~~ are not reviewed. This chapter serves as a companion to an earlier chapter [9] that describes process measures relevant to psychodynamic psychotherapy research.

The Social Cognitions and Object Relations Scale – Global Rating Method (SCORS-G)

The Social Cognitions and Object Relations Scale – Global Rating Method (SCORS-G) [10, 11] is an observer and expert-rated coding system applied to narrative data. It quantifies individuals' internal representations of self and others, as well as tapping the cognitive, affective, and motivational elements associated with object relations. Changing internal representations and individuals' abilities to conceptualize the self, others, and relationships in general is often a goal of psychodynamic psychotherapy. The SCORS-G provides a method for quantifying these representations and capacities.

The original SCORS-G system used a Q-sort approach to code for four areas of object relations [10]. Hilsenroth and colleagues, however, recently developed a training protocol for an expert scoring method that built on the original work by Westen [10]. The manual created by Hilsenroth et al. [11] reviews changes, provides scoring instructions, includes specifics on how to set up a program for training coders, has examples of narrative content to be coded, and introduces specific target criteria for coders to meet prior to using the SCORS-G. It also includes detailed instructions for examining inter-rater reliability statistically. This version also included a total of eight domains [11]. The domains are as follows:

- Complexity of Representations (COM)
- Affective Quality of Representations (AFF)
- Emotional Investment in Relationships (EIR)
- Emotional Investment in Values and Moral Standards (EIM)
- Understanding Social Causality (SC)
- Experience and Management of Aggressive Impulses (AGG)
- Self-esteem (SE)
- Identity and Coherence of Self (ICS)

While all domains are scored on 7-point scales, each has its own scale anchors. Lower scores are always indicative of less adaptive functioning. For example, for the Self-Esteem (SE) domain, a score of 1 indicates that the respondent views the self as loathsome, evil, rotten, contaminating, or globally bad. A score of 3 indicates that the respondent has low self-esteem or is unrealistically grandiose. A score of 5 indicates that a range of positive and negative feelings were used to describe the self. Finally, a score of 7 indicates that the respondent tends to have relatively positive feelings toward the self. Full descriptions of these domains, the global rating method, administration details,

scoring procedures, and training examples can be found in the manuals developed by Westen [10] and Hilsenroth et al. [11]. A copy of the SCORS coding form can be found in Appendix 27.1 of this chapter.

SCORS-G domain scores can be used at the domain level or at the global level [11]. Thus, if ten stories are used, the eight domains are coded for *each story*. Overall domain scores are then calculated by averaging across stories. Finally, these scores can then be aggregated to form an overall Global Rating Scale (GRS) score by combining scores across all domains. Originally developed for coding narratives produced in response to the evocative pictures, subsequent investigators have successfully extended the SCORS-G method to assess narratives produced for dream [12], psychotherapy interviews [13], and early memories [14]. The length of time required to collect and code narratives varies as a function of the data collection method used.

A number of independent investigators have demonstrated that once raters have been successfully trained, narratives can be reliably coded using the SCORS-G. Westen reported coefficient alphas ranging from 0.80 to 0.90 and strong inter-rater reliabilities for SCORS-G domains [10]. Additional investigations have ~~tended to~~ produce similar results. For example, Stein et al. [13] obtained SCORS-G ratings for early memory narratives in a clinical sample and reported Intra-class Correlation Coefficients (ICCs) for the eight respective SCORS-G items ranged from good (e.g., 0.60) to excellent (e.g., 0.84). Ackerman et al. [15, 16] reported highly similar inter-rater reliability coefficients for SCORS-G ratings of narrative stories in clinical samples. Inter-rater reliability estimates across studies have generally been in the good to excellent range despite considerable differences in sample level of distress, and type of narrative data used [17].

The construct validity of the SCORS-G has been demonstrated in several ways [17]. SCORS-G ratings have shown relationships with other scales designed to assess problematic interpersonal functioning. SCORS-G ratings show convergence with similar rating scales tapping object representations, defense mechanisms, and behavior for psychiatric inpatients [18]. SCORS-G ratings also show convergence with self-report ratings for Axis II psychopathology [19]. Stein et al. [13] found that SCORS-G profiles for early memory narratives were related to self-reported interpersonal problems and clinician-rated global relational functioning at time of initial assessment for a clinical sample of patients seeking psychotherapy. Thus, SCORS-G ratings are related to clinician ratings for global relational and social functioning [13, 18] and self-report ratings for personality problems [13, 19].

SCORS-G profiles also differentiate among types of psychopathology. What is notable is that they differentiate between pathologies believed to differ with regard to object relations. For example, patients with borderline personality disorder (BPD) achieved lower SCORS-G ratings for complexity of others, emotional investments in relationships, affective quality of relationships, and greater need gratification as compared to patients without BPD and non-clinical controls [20]. Similarly, Ackerman et al. [21] found that patterns of object relations, as assessed by the SCORS-G, differentiated among different types of Cluster B pathology in a sample of outpatients seeking psychotherapy.

The SCORS-G has been utilized directly in psychotherapy research. Ford et al. [22] found that low pre-treatment SCORS-G ratings for overall quality of object relations was associated with less of a treatment response (i.e., less symptom reduction and improved quality of life) in inpatient therapy. Ackerman et al. [16] found that patients who had lower SCORS-G ratings for affective quality of object relations, but higher ratings for capacity to emotionally invest in relationships tended to stay in psychodynamic psychotherapy longer. Regarding sensitivity to change as a function of treatment, Fowler et al. [23] found that psychiatric inpatients receiving intensive psychodynamic psychotherapy over 16 months evidenced significant changes in the overall SCORS-G ratings as well as in the specific SCORS-G domains (COM, SC, SE, and ICS). Porcerelli et al. [18] reported similar results for inpatients receiving 15 months of intensive psychodynamic therapy at an independent facility. Such studies suggest that the SCORS-G is sensitive to change.

The SCORS-G has a number of strengths that make it appealing to psychodynamic psychotherapy researchers. It clearly measures constructs associated with object relations, and validity data to date strongly supports the convergent and construct validity of the tool. It provides an overall score as well as scores for specific areas of object relations. Psychometrics are strong and several respective investigators have reported good to strong inter-rater agreement at the item level. SCORS-G ratings can be obtained using a number of methods for soliciting narratives, though it is possible that method used may impact scoring [24]. SCORS-G ratings are not based exclusively on patient report. Thus, it may be particularly appealing for use in samples that may have difficulty accurately describing themselves or reporting on their personalities. On the other hand, the SCORS-G does require an investment of time in training and practice for raters to become reliable. There are also no available norms in general, and no available norms based on method of narrative collection. We are unaware of any studies examining test–retest stability for SCORS-G ratings.

The Shedler–Westen Assessment Procedure-200 (SWAP-200)

The Shedler–Westen Assessment Procedure (SWAP-200) [8, 25] is an assessment instrument of personality psychopathology and health designed to provide clinicians of all theoretical orientations a standard "vocabulary" for case description [26]. It may be used to identify the long-standing and maladaptive cognitive, affective, relational, and behavioral patterns that treatment may seek to change. It also measures inner capacities and resources which promote healthy functioning. It is a measure that evaluates both broad and specific features of personality. Thus, it is likely to be of use to psychodynamic psychotherapy researchers interested in studying how character or structural capacities change from treatment.

The SWAP-200 consists of 200 statements, each of which may describe a given patient very well, somewhat, or not at all. The clinician or expert rater ranks statements into eight categories from those that are most descriptive (assigned a value of 7) to those that are not descriptive (assigned a value of 0). Thus, the SWAP-200 yields a score from 0 to 7 for each of 200 personality-descriptive variables. The "standard vocabulary" of the SWAP-200 allows clinicians to provide in-depth psychological descriptions of patients in a systematic and quantifiable form and ensures that all clinicians attend to the same spectrum of clinical phenomena. SWAP-200 statements are written in a manner close to clinical observation (e.g., "Tends to get into power struggles," or "Is capable of sustaining meaningful relationships characterized by genuine intimacy and caring"), and statements that require inference about internal processes are written in clear, unambiguous language (e.g., "Tends to see own unacceptable feelings or impulses in other people instead of in him/herself"). Writing items in this jargon-free manner minimizes unreliable interpretive leaps and makes the item set useful to clinicians of all theoretical perspectives. A systematic Clinical Diagnostic Interview (CDI) [26–28] (see Chap. 28), which can be administered in approximately two and a half hours, yields sufficient patient information to score the SWAP-200 reliably and validly. The interview can be used in either clinical or research contexts. When the interview is not used, clinicians can generally score the SWAP-200 after 5 or more clinical contact hours with a patient.

SWAP-200 profiles can be diagnostically scored using a matching procedure in which the patients profile is compared to 11 Personality Disorder Factors. These factors were derived through a large expert survey of clinicians who created SWAP-200 profiles for ten Personality Disorder diagnoses [8, 25]. These clinicians also created SWAP-200 profiles for healthy and adaptive functioning. This profile has become the "healthy functioning" profile. Table 27.1 includes examples of SWAP-200 items comprising each of the Personality Disorder Factors. It should also be noted that there are alternative strategies for organizing and scoring the SWAP-200 to assess various personality dimensions as opposed to specific personality disorders [28].

Table 27.1 SWAP-200 personality disorder factors with example items

Personality disorder factors	Example items
Paranoid	Is suspicious; tends to assume others will harm, deceive, conspire against, or betray him/her
	Tends to hold grudges; may dwell on insults or slights for long periods
Schizoid	Appears to have little need for human company or contact; is emotionally detached or indifferent
	Appears to have a limited or constricted range of emotions
Schizotypal	When distressed, perception of reality can become grossly impaired (e.g., thinking may seem delusional)
	Lacks social skills; tends to be socially awkward or inappropriate
Antisocial	Tends to engage in unlawful or criminal behavior
	Tends to show reckless disregard for the rights, property, or safety of others
Borderline	Emotions tend to change rapidly and unpredictably
	Emotions tend to spiral out of control, leading to extremes of anxiety, sadness, rage, etc.
Histrionic	Expresses emotion in exaggerated and theatrical ways.
	Seeks to be the center of attention
Narcissistic	Has an exaggerated sense of self-importance (e.g., feels special, superior, grand, or envied)
	Appears to feel privileged and entitled; expects preferential treatment
Avoidant	Tends to feel he/she is inadequate, inferior, or a failure
	Tends to feel like an outcast or outsider
Dependent	Appears to fear being alone; may go to great lengths to avoid being alone
	Tends to be ingratiating or submissive (e.g., consents to things he/she does not want to do, in the hope of getting support or approval)
Obsessive-compulsive	Tends to become absorbed in details, often to the point that he/she misses what is significant.
	Is excessively devoted to work and productivity to the detriment of leisure and relationships
High functioning	Appreciates and responds to humor
	Is able to form close and lasting friendships characterized by mutual support and sharing of experiences

Because the SWAP-200 is jargon free and clinically comprehensive, it has the potential to serve as a language for describing personality pathology that can be used by any skilled clinical observer. Studies demonstrate that experienced clinicians of all theoretical orientations understand the items and score them reliably. For example, a nationwide sample of 797 experienced psychologists and psychiatrists of diverse theoretical orientations, who had an average of 18 years practice experience post-training, used the SWAP-200 to describe patients with personality pathology [8]. These experienced clinicians provided similar SWAP-200 descriptions of patients with specific PDs regardless of their theoretical background, and 72.7% agreed with the statement "I was able to express most of the things I consider important about this patient" (the highest rating category). In a subsequent study using a sample of 1,201 psychologists and psychiatrists, 84% "agreed" or "strongly agreed" with the statement: "The SWAP-200 allowed me to express the things I consider important about my patient's personality." Again, the ratings were unrelated to clinicians' theoretical orientation.

Numerous studies have demonstrated the reliability and validity of SWAP-200 data. In these studies, reliability of SWAP-200 personality descriptions has generally ranged from 0.75 to 0.98 [8, 29–31]. The SWAP-200 predicted a range of relevant external criteria, from those that are relatively objective to those that require greater inference. These include, for example, history of suicide attempts and psychiatric hospitalizations; adaptive functioning assessed by measures such as the Global Assessment of Functioning index; aggressive ward behavior; non-engagement in psychotherapy; family history variables such as psychosis in first- and second-degree relatives; and developmental variables, including being raised by a substance-abusing parent or guardian, childhood history of physical abuse; childhood history of sexual abuse, and problems with parental bonding and attachment [3, 8, 29–34]. Furthermore, the SWAP-200 does not simply measure personality psychopathology, but also includes an empirically derived Healthy Functioning Index that defines

and operationalizes mental health in a manner consensually understood by clinical practitioners across theoretical orientations [3, 8, 25].

While the SWAP-200 measures personality problems related to various forms of psychopathology and disruptive behavior as well as inner capacities and psychological strengths for sustaining mental health and well-being, few studies have utilized the SWAP-200's potential as a measure of change. Two studies, however, have yielded results which support the viability of the SWAP-200 as a therapeutic outcome measure. The first is a single-case study of a woman diagnosed with BPD who was assessed with the SWAP-200 by independent assessors (not the treating clinician) at the beginning of treatment and again after 2 years of psychodynamic therapy [29]. In addition to meaningful decreases in SWAP-200 scales that measure psychopathology, the patient's score on the SWAP-200 Healthy Functioning Index increased by approximately two standard deviations over the course of treatment and indicated that the patient showed a greater capacity for empathy and greater sensitivity to others' needs and feelings; increased ability to recognize alternative viewpoints, even when emotions ran high; greater ability to comfort and soothe herself; increased recognition and awareness of the consequences of her actions; and growth from the working through of painful past experiences.

The SWAP-200 has also been used to compare a small sample of patients beginning psychoanalysis with a matched group of patients at the completion of psychoanalysis [35]. The group ending psychoanalysis had significantly lower scores for items assessing depression, anxiety, guilt, shame, feelings of inadequacy, and fears of rejection but significantly higher scores for SWAP-200 items assessing adaptive strengths and well-being, including greater enjoyment of challenges and pleasure in accomplishments, ability to utilize talents and abilities, contentment in life's activities, empathy for others, and interpersonal assertiveness and effectiveness.

For a more technically thorough discussion of the psychometric strengths and weaknesses of the SWAP-200, see Westen and Shedler [2]. These authors elaborate on the Q-sort methodology, implications of the SWAP-200's fixed distribution ranking system, taxonomy, and scaling. Finally, the SWAP-200 has not yet been applied as an outcome measure in large sample longitudinal or controlled clinical trials, although the studies discussed previously provide supporting evidence indicating both temporal stability and sensitivity to change. Despite these limitations, the SWAP-200 is not only a useful tool for personality diagnosis and case conceptualization, but is also a viable option for studying meaningful changes in enduring patterns of functioning, both maladaptive and adaptive, over long-term treatment. It is likely to be particularly useful to researchers who are interested in how therapy can produce character change. Many of the features assessed by the SWAP-200 are unlikely to change easily and are likely to be fairly stable over time. Thus, it may be most appropriate for researchers interested in studying how character changes over the course of treatment and/or researchers interested in how longer-term treatments facilitate changes to basic personality traits. The format of the SWAP and the level of detail also make it an extremely useful measure for quantifying clinical description. While capable of being used in any type of outcome research design, it has notable advantages over traditional outcome measures for understanding and describing patient change over the course of therapy. Thus, it is also highly useful for increasing the scientific integrity of single-case designs and clinical case studies.

[AU2]

The Defense Mechanism Rating Scales (DMRS)

Psychological defenses have been a staple of psychodynamic theory from its outset and are a frequent target of exploration and change in treatments. The Defense Mechanism Rating Scales (DMRS) [36] is an observer-rated measure that can be scored by a clinician or external rater. The authors of the DMRS, Perry and his colleagues, were also involved in the development of the qualitative Provisional Defense Axis in Appendix B of the DSM-IV [37], and thus it is of little surprise

Table 27.2 Defensive levels and their corresponding defenses

Defense level	Defenses included on this level
High adaptive level	Affiliation, altruism, anticipation, humor, self-assertion, self-observation, sublimation, suppression
Obsessional	Isolation, intellectualizing, undoing
Other neurotic	Repression, dissociation, reaction formation, displacement
Minor image distorting	Omnipotence, idealization, devaluation
Disavowal	Negation, projection, rationalization, autistic fantasy
Major image distorting	Splitting of others' image, splitting of self-image, projective identification
Action	Acting-out, passive aggression, hypochondriasis

that there are strong similarities between the DMRS and this axis [38]. Like many of the measures discussed, the DMRS can be employed in a variety of research designs. It is simultaneously a measure of personality function and an outcome measure. It can be used for group-design studies or in more granular studies focusing on a single individual. (See also Chap. 22.)

The DMRS includes 28 defenses that are hierarchically ordered into seven levels ranging from least to most mature. Table 27.2 displays the seven defensive levels and includes a list of the defenses included in that level.

The organization of the DMRS was based on a review of the prior literature on psychological defenses followed by systematic investigations linking defensive levels to measures of adaptiveness, pathology, and/or psychological defense [38–41]. The DMRS can be scored from narratives and is most often scored using psychotherapy session material. It provides an overall score for defensive functioning (Overall Defensive Functioning, ODF) that ranges from a 1 (least adaptive) to 7 (most adaptive). In addition, a score is provided for each specific defense level. This allows investigators to study change at the overall level and in specific areas of functioning.

Detailed descriptions of each type of defense and examples of how the defense is conceptualized can be found in the DMRS manual to help raters clearly score each defense and help distinguish among defenses [36]. To date, investigations into the reliability of the DMRS scores have generally been positive with reliability coefficients for ODF ranging from good to excellent across studies [36–41]. As would be expected, reliability coefficients have been slightly lower for individual defense scores, but still within the accepted range [41, 42]. Ratings for overall level of defensive functioning show a fair degree of stability over time as ODF have obtained an interclass *R* (*IR*) of 0.48 for a 5-week period [43].

The validity of DMRS has been established through a number of studies (for reviews, see [36]). Lower levels of defensive functioning have been associated with high distress, personality pathology, and interpersonal difficulties [40, 41, 44]. A handful of studies have used the DMRS to study change in psychodynamic psychotherapies [42]. For example, Perry and Hoglend [39] found significant improvements in ODF for depressed individuals following psychotherapy. Similarly, in a study of 61 patients undergoing very brief psychodynamic investigations (i.e., four sessions), Drapeau and colleagues [45] found a significant increase in ODF and the use of defenses from the obsessional level. In a very recent study, significant changes in DMRS ODF scores were observed for a small sample of patients who completed psychoanalysis [46].

The DMRS has clearly defined defenses, can be learned fairly quickly, and provides scores at multiple levels (i.e., overall, defensive level, specific defense). It has been utilized in treatment research, and these studies serve as templates for future investigators. Though the DMRS has well-defined defenses, some familiarity with the concept and types of defenses is likely to be of benefit for investigators hoping to use this system. Some training is required to be able to use the measure effectively, and steps should be taken to ensure reliability of scoring prior to use. Future investigation examining how ratings of an individuals' defensive functioning vary as a result of method of

scoring (e.g., scoring interview vs scoring a narrated stories) would be of benefit in increasing the measures utility. In summary, the DMRS is likely to be appealing to investigators interested in learning how pre-treatment defensive functioning impacts treatment and outcome, as well as determining how psychotherapy changes defensive functioning over time.

The Inventory of Interpersonal Problems – Circumplex (IIP-C)

Changing interpersonal functioning is an important therapeutic target for a number of insight-oriented treatments. The Inventory of Interpersonal Problems (IIP) [47–49] is a self-report inventory for identifying problematic aspects of relational functioning. Thus, while related to measures of personality traits and indexes of interpersonal style, the IIP focuses on areas that are problematic. As such, it is similar to measures assessing symptoms and distress.

The original IIP contained 127 items comprising eight subscales forming octants along the interpersonal circumplex. However, the most widely used version of the test contains 64 items and is referred to as the Interpersonal Problems – Circumplex (IIP-C) [48, 49]. A number of additional short forms and derivative measures exist containing item sets designed to accomplish specific purposes (e.g., screen for personality disorders) or that map to a particular conceptual and/or data analytic approaches [50]. When selecting a version of the IIP to use in a research study, we strongly recommend reviewing the guidelines and advice put forth by Hughes and Barkham (pp. 491–492) [50] for selecting a version. We also suggest reviewing information by Gurtman ~~and Pincus~~ [51] for selecting methods for scoring the IIP. For the remainder of this section, we focus on the IIP-C. [AU3]

The 64 items for the IIP-C are organized into two item sets: things the respondent does too much (excesses; e.g., "It is hard for me to join in groups") and things the respondent finds it hard to do (inhibitions; e.g., "I fight with other people too much") [48]. Each item is rated on a scale ranging from 0 (not at all) to 4 (extremely). The IIP-C produces eight subscales. Table 27.3 presents the eight subscales as well as example items from each subscale.

t3.1 **Table 27.3** Example items from IIP-C subscales

Scale	Example items
Domineering	I am too aggressive toward other people
	I try to control other people too much
Vindictive	It is hard for me to trust other people
	I want to get revenge against people too much
Cold	It is hard for me to feel close to other people
	It is hard for me to get along with other people
Socially avoidant	It is hard for me to ask other people to get together
	I feel embarrassed in front of other people too much
Non-assertive	It is hard for me to tell a person to stop bothering me
	It is hard for me to be assertive with another person
Exploitable	I am too easily persuaded by other people
	I let other people take advantage of me too much
Overly nurturant	I try to please other people too much
	I put other people's needs before mine too much
Intrusive	I want to be noticed too much
	It is hard for me to stay out of other people's business

t3.19 Adapted with permission of Taylor and Francis Group from [48]

The psychometric adequacy of the IIP-C has been well established. Internal consistency alpha coefficients for the various subscales range from 0.76 to 0.88 and test–retest reliabilities for scales have ranged from 0.58 to 0.84 [49]. Alden and colleague [48] reported that data derived from large non-clinical samples supported the circumplex structure of the IIP-C. Scales of the IIP-C have been consistently shown to converge with measures tapping similar constructs in clinical [51, 52] and non-clinical samples [48, 49].

The IIP-C has been used previously to study various aspects of psychotherapy process and treatment response [51, 53]. In fact, of all the measures reviewed in this chapter, the IIP-C has been the most frequently employed for the purpose of outcomes research. Høglend et al. [54] examined changes in interpersonal problems in a diverse clinical sample that completed a year of psychodynamic psychotherapy. They found that the overall amount of interpersonal problems reported on the IIP-C declined following psychodynamic treatment. Similarly, using a naturalistic sample of patients undergoing psychodynamic-interpersonal psychotherapy, Paley et al. [55] used an abbreviated form of the IIP-C, the IIP-32, and found that long-term treatment was associated with clinically significant improvement in interpersonal functioning as measured by this scale. Using a sample of generalized anxiety disorder (GAD) patients completing 16 sessions of supportive-expressive psychodynamically focused psychotherapy, Crits-Christoph et al. [56] found that treatment improved overall interpersonal functioning on the IIP-C and also found significant pre- to post improvements in specific subscales: Non-Assertive, Exploitable, Overly Nurturant, Socially Avoidant, and Intrusive. Thus, across independent labs and patient population, the IIP-C has shown sensitivity to change.

Clearly, there are several reasons that the IIP-C is frequently employed. First, it has strong psychometric properties, and independent researchers have repeatedly found the IIP-C to be reliable over time, sensitive to change, convergent with similar measures, and predictive of therapeutic outcomes and course. The IIP-C has been effectively employed with general clinical samples, specific clinical samples, and non-clinical samples. It has normative reference scores for both clinical and non-clinical populations. It is easy to administer and requires minimal training to score. Finally, it provides an overall assessment of interpersonal functioning as well as specific assessment of relational domains allowing for investigation of more global changes in interpersonal relating, specific areas of change, or both. The measure has very few drawbacks. Of course, like many self-report measures, the items do have a high degree of face validity, and it is easy to determine the intent of the item, possibly increasing the measure's vulnerability to faking. Also, like all of the self-report measures discussed in this chapter, the measure does not include any validity scales capable of detecting inconsistent, odd, or faked responding. Overall, the drawbacks to this measure are quite limited and the strengths many. Improvement in social and relational functioning has been cited as an important outcome of psychodynamic psychotherapy, and it is highly likely that the IIP-C will continue to be the self-report measure of choice for studying how psychodynamic psychotherapies lead to improvements in this domain.

Adult Attachment Interview (AAI)

A fundamental proposition of Attachment Theory is that experiences with caregivers influence future relationships, expectations of others, and views of the self, via the formation of internal representational models. George ~~eta l.~~ et al. [57] developed an interview procedure, the Adult Attachment Interview (AAI), for assessing "state of mind with respect to attachment" in adults. The AAI can be used for multiple purposes, one of which is to aid in understanding how psychotherapy can change attachment representations [58].

The AAI is a 1-hour semi-structured interview-based measure. The interview focuses on early attachment experiences as well as adult relational experiences. Respondents are encouraged to explore and discuss thoughts, feelings, and understandings for how past relational experiences have impacted them. The interviewer probes for supporting evidence to back up statements provided by the respondent. The AAI involves 20 questions with some standardized probes. For example, respondents are asked to provide five adjectives to describe their relationship with each parent and are then asked to provide evidence and relational experiences that support these descriptors. Consistent with attachment theory, they are also asked about how their parents responded to them when they were in need and are asked about experiences of loss, separation, and rejection. To properly administer the AAI, the interview must be conducted by a trained interviewer and is transcribed verbatim for coding. The transcriptions must be scored by an AAI expert coder who first produces subscale ratings that culminate in an attachment classification. The AAI classifies individuals as secure/autonomous, dismissing, enmeshed/preoccupied, unresolved/disorganized, or cannot classify [59]. The unresolved/disorganized rating may be given in addition to another pattern code or can be given as the primary pattern code.

Training in the coding of the AAI is quite rigorous which likely accounts, in part, for the high quality of research that has been produced using this instrument. Coders are required to complete a minimum of 2 weeks of training and must obtain an 80% or greater agreement with 30 expert coded standard AAI transcripts [60]. Training is essential for this instrument as the unit of analysis is complex, analyzing structural features of language (e.g., coherence of the narrative) rather than the content itself. The actual coding of the AAI also involves notable time and resources often requiring 6–8 h to complete. Further, at this time, the AAI manual is only available to approved coders from the AAI Institute. Though the AAI is typically employed to code respondents into attachment groups and to assess some specific areas of functioning (e.g., coherence of narrative), there is also some evidence from taxometric analysis that attachment may be best represented dimensionally, and there are multiple means by which to score the measure [61].

The scale properties of the AAI have been substantially researched [60, 61] and are impressive, especially for a verbally elicited interview-based coding system. Over 10,000 AAI interviews have been conducted and coded, and there is extensive information on distributions and factors that impact attachment [62]. Test–retest reliabilities and inter-rater agreement have been found to be strong in several populations and across labs [63–65]. Further, attachment categorization via the AAI appears to be unrelated to non-attachment related memories, as well as other variables unrelated to attachment such as IQ, verbal abilities, and social desirability [63, 64]. AAI classification codes have been found to be stable over brief (several months) and long periods (several years) [62–66].

There is far too much validity research on the AAI to review here, and readers are referred to recent reviews [58, 60]. A brief review of some key findings follows. Parent AAI classifications, obtained prior to the birth of the infant, have been shown to predict the quality of subsequent infant–parent relationships [67]. In his meta-analysis of 18 studies, van IJzendoorn [68] found roughly 70% agreement between parent's AAI classifications and their children's attachment status. Waters and colleagues [65] have shown that infant ratings of attachment based on behavioral coding show strong agreement with adult attachment classifications made 20 years later using the AAI. An interesting note is that AAI ratings and classification are not always in strong agreement with self-report measures of adult attachment [61]. This may be partially the result of differences in measurement method as well as specific constructs tapped by the AAI.

Few studies have utilized the AAI to assess the effects of psychotherapy. However, two studies are notable. In a randomized control trial with patients diagnosed with BPD, Levy and colleagues [69] found that BPD patients in Transference-Focused Psychotherapy (TFP) were more likely to shift to a more secure attachment organization on the AAI following 12 months of treatment as compared to such patients receiving Dialectical Behavioral Therapy (DBT) or Modified Psychodynamic Psychotherapy. The AAI has also been used in single-case designs as well. For example, Gullestad [70] [AU4]

presents pre–post AAI data on a single patient who underwent psychoanalysis four times a week, resulting in changes in a number of specific elements in his discourse from time 1 (pre) to time 2 (post) completion of the AAI.

The AAI is a powerful research instrument that has tremendous promise for psychodynamic psychotherapy research. Its psychometric properties are well established, and the authors have wisely developed a rigorous training process to ensure fidelity of measurement. This has culminated in an impressive series of validity studies that consistently link the AAI classification and scores to key variables (e.g., parent status is highly related to child attachment style). While many have pointed out that the measure is likely to have a number of clinical uses [71], outcome studies making use of the AAI continue to be somewhat rare. In part, this may reflect the fact that the AAI does involve a commitment in terms of time, resources, and effort in order to administer and score. Still, given its particular focus on internal representations, the AAI should be considered among the most appropriate outcome measures available for studying change in insight-oriented therapies.

The Reflective Functioning Scale (RFS)

The concept of reflective functioning arose, in part, from an intermingling of research from attachment theory and clinical observations and research focusing on individuals with notable difficulties in affect regulation, coping, identity, and interpersonal functioning. It refers to an individual's ability to understand others' reactions and behaviors in terms of underlying mental states (e.g., intentions; motivations; affect). This construct, also referred to as mentalization, is viewed by many as among the central benefits that can result from psychodynamic psychotherapy [71]. The Reflective Functioning Scale (RFS) was developed by Fonagy and colleagues [72] to assess the capacity for engaging in mentalization. It is scored based on coding narratives typically obtained from AAIs.

The RFS is an additional scale that can be used in collaboration with the AAI. Using verbatim AAI transcripts, trained coders give the respondent a single score ranging from 1 (very low) to 9 (very high) for reflective functioning. Though the AAI is typically employed for scoring the RFS, other authors have used alternative approaches [73]. The scale characteristics of the RFS have ranged from adequate to strong. For example, in initial investigations of the rater agreement, rater pairs consistently produced correlations of >0.80 [72]. Subsequent independent researchers trained in coding the RFS have also produced reliable scores. For example, Bouchard et al. [74] reported rating pair correlations of >0.80 using 22 AAI transcripts. Rudden et al. [73] reported inter-rater reliabilities in the excellent range (i.e., ICC = 0.74) for this measure.

Evidence for the convergent validity of the RFS has been obtained from studies showing agreement with similar measures of mentalization or reflective abilities. For example, Bouchard et al. [74] showed that RFS scores were related to other narrative-based measures of mentalization. As would be predicted theoretically, RFS scores were negatively correlated with independent measures of less adaptive defenses and concreteness of narrative in this study. Regarding construct validity, Fonagy et al. [75] found that RFS scores mediated the relationship between parental attachment classification on the AAI and their children's attachment status. In this study, insecure parents who produced securely attached children tended to have higher RFS scores. Slade et al. [76] have provided data demonstrating that higher levels of reflective functioning in mothers are associated with more secure patterns of infant–mother attachments. Reflective functioning has also been linked to the development of pathology in theoretically predicted ways. For example, Fonagy et al. [77] found that patients with a trauma history and poor reflective functioning were at extremely high risk for BPD, while patients with a trauma history and higher levels of reflective functioning were at substantially lower risk.

Similar to the AAI, only a few studies have used the RFS to study treatment outcome. Levy and colleagues [69] found that, following a year of therapy, BPD patients in TFP were more likely to

evidence improved RFS scores as compared to such patients in supportive treatment or DBT. Rudden et al. [73] used the RFS to examine change in reflective functioning for a sample of patients randomly assigned to psychodynamic treatment for panic disorder or relaxation treatment. RFS ratings were not made based on AAI interviews, but were given in response to interview narratives derived from a shorter semi-structured interview that included the AAI demand questions. While changes in general RFS ratings were not found, changes in reflective functioning specific to panic symptoms were obtained.

Similar to the AAI, though the RFS has clear promise and application for the study of psychotherapy outcomes, its use has been somewhat limited. As such, it is difficult to assess this measure's ability to capture change. It was designed to be used with the AAI, and thus, as previously discussed, its limited use in outcome studies may reflect cost, time, and training necessary to employ this measure. Meehan et al. [78] have presented preliminary data for a recently developed 50-item expert-coding system based on the RFS, the Reflective Functioning Rating Scale (RFRS). The RFRS was developed to provide more granular assessment of factors contributing to reflective functioning and was also designed in a manner that would allow it to be used with a wider range of narrative-based data. A potential benefit of the RFRS is more items which will allow for better understanding of the measure's scale properties and over time may provide a more granular understanding of reflective functioning. Initial investigation into this measure appears very promising though further research is clearly necessary.

The Psychodynamic Functioning Scales (PFS)

The Psychodynamic Functioning Scales (PFS) is an expert-rated system for assessing important structural aspects of character. It was developed by Høglend et al. [7], who have also pioneered a number of research studies examining psychodynamic psychotherapy process and outcome. The PFS was developed, in part, in an attempt to create a measure that evaluated key clinical concepts in a manner that was relatively easy to score and that would be sensitive to detecting change from brief psychodynamic treatments. As its authors point out (p. 191), similar to other dynamically oriented scales, the PFS scales are intended to assess "internal predispositions, psychological resources, capacities, or aptitudes that can be mobilized by the individual to achieve adaptive functioning and life satisfaction."

The PFS contains six capacity scales that are rated by an interviewer or expert rater. The six capacity scales are listed here:

- Quality of Family Relationships
- Quality of Friendships
- Quality of Romantic Relationships
- Tolerance for Affects
- Insight
- Problem Solving Capacity

Similar to the Global Functioning scale of the DSM-IV [37], the PFS capacity scales are rated on a 1–100 scale, with higher scores being indicative of more adaptive functioning. Consistent with the goal of developing a scale sensitive enough for detecting changes as a function of "brief" dynamic therapy, ratings for each capacity are based on the individual's functioning for the past 3 months. In addition to specific domain scores, an overall score can be calculated by simply calculating the weighted average of the respective scale scores [4].

Thus far, studies using the PFS have suggested that the scales are psychometrically sound. Inter-rater agreement ~~among raters~~ has been ~~considered~~ in multiple studies. Høglend et al. [7, 79] reported

inter-rater reliability estimates (i.e., ICCs) of 0.90 for agreement in average scores across three expert raters. To date, the use of the PFS has been limited to Høglend et al.'s study focusing on comparing psychodynamic psychotherapy outcomes for treatments with and without transference interpretations. These studies demonstrate the PFS's sensitivity to detecting changes from various forms of psychodynamic psychotherapy. For example, Høglend et al. [7] found that PFS scores for patients completing a year of psychodynamic psychotherapy with transference interpretations or psychodynamic psychotherapy without transference interpretations improved significantly over the course of treatment.

The PFS has a number of strengths: it measures constructs of clear importance to psychodynamic researchers, scale formats are straightforward, and scale content has been clearly defined by the authors. Data to date indicate that scales can be reliably scored, and the PFS has shown sensitivity to detecting changes as a function of therapy. The measure also has the advantage of being created by researchers highly experienced in studying psychodynamic psychotherapy. To date, it has been primarily used by labs associated with the authors who created the measure. It will be important for additional labs to examine the utility and psychometric adequacy of the instrument to better understand if its scale properties remain strong even when utilized by investigators less familiar with the measures development. Given the relative newness of the PFS, this is to be expected, and initial data on the measure is very promising. Given the strengths of the PFS, particularly the clarity of the scales and ease of scoring procedures, we anticipate the measure will gain wider usage. The measure has potential ~~to prove highly useful as an~~ outcome tool for studying change in a variety of therapies, including psychodynamic psychotherapy.

Karolinska Psychodynamic Profile (KAPP)

Throughout the psychodynamic literature, conceptualizations of the building blocks of character can be found. The Karolinska Psychodynamic Profile (KAPP) [6] is a structured interview-based procedure for assessing relatively stable character traits and modes of functioning [80]. It was created by Weinryb et al. [80]. In describing their aims, the KAPP's authors [80] state: "We wanted an instrument that could be useful for (a) the assessment of relatively stable modes of mental functioning and character traits with a focus on structure not on specific conflicts, defenses, wishes, or fears…" The measure was developed following a thorough review of the psychodynamic literature on character and thus is built to tap constructs expected to change as a function of psychodynamic psychotherapies.

The KAPP is composed of 18 independent single-item scales. Seventeen scales assess modes of mental functioning: Intimacy and Reciprocity, Dependency and Separation; Controlling Personality traits; Frustration Tolerance; Impulse Control; Regression in Service of the Ego; Coping with Aggressive Affects; Alexithymia; Normopathy; Bodily Appearance; Bodily Function; Bodily Image; Sexual Functioning; Sexual Satisfaction; Sense of Belonging; Feeling of Being Needed; and Access to Advice for Help. The final scale provides a more global index of personality organization (i.e., neurotic, borderline, psychotic; for detailed description of these scales see [80]). Each scale is rated by the interviewer and is given a score ranging from 1 to 3. Midpoint scores (i.e., 1.5 and 2.5) are also included resulting in a 5-point scale. Lower scores are indicative of more "normal" or "normal-neurotic" functioning. In contrast, ratings of two or greater suggest pathology. Ratings are made following an interview of the patient specifically designed to assess these areas of functioning. KAPP interviews require roughly 90–120 min [81, 82] and should be administered by a trained interviewer familiar with the system. The scoring system has also been employed to code data obtained via projective methods [83].

To date, a handful of published studies have examined the psychometrics of the KAPP. Inter-rater reliability for clinical and non-clinical groups has generally been in the acceptable range (i.e., ICCs > 0.70) or better [80, 83, 84] with agreement improving when raters are more experienced

and familiar with psychodynamic treatments and theory. Ratings across specific rating pairs have occasionally produced inconsistent agreement for some specific scales [80]. Haver et al. [85] failed to achieve reliability estimates within the accepted range for six scales. In reviewing these results, Weinryb et al. [80] provide explanations for these findings and suggestions for optimizing reliability among coders. Test–retest coefficients for a non-psychiatric sample of patients receiving surgery for ulcerative colitis who completed the KAPP on two occasions (mean interval of 22 months; range of 16–34 months) revealed good stability (i.e., median $r = 0.57$ for sample) [86].

A limited number of studies have attempted to examine the convergent validity of the KAPP with other measures. To some extent, this is to be expected given the relative dearth of measures designed to assess character in a multi-faceted manner. Instead, the validity of the KAPP has been established in other ways. For example, KAPP ratings consistently discriminated surgical patients without a psychiatric diagnosis from patients who had a psychiatric diagnosis [86]. A small number of investigators have also used the KAPP to study treatment outcome. For example, Wilczek et al. [87] examined change in personality structure over the course of long-term (average length of treatment = 3 years) psychodynamic psychotherapy. They found that treatment was associated with significant changes on eight KAPP subscales. In a fairly large sample of patients in treatment for personality disorders, Vinnars et al. [81] found significant changes in KAPP composite scales associated with object relations and ego functioning. Significant pre–post changes in KAPP scores have also been reported for patients undergoing treatment of eating disorders [88].

The developmental approach and focus of the KAPP is likely to make it of interest for psychodynamic investigators. Similar to the SWAP-~~200~~ discussed previously, its focus on overarching conceptualizations of character and personality is likely to increase its utility for studying the ways in which psychodynamic psychotherapy fosters character change. The KAPP is likely to be particularly appealing for researchers investigating how therapy over time produces change in areas of character that tend to be relatively stable. The scales of the KAPP assess character traits that are often ignored by traditional outcome measures. These traits, however, are not expected to change easily or quickly. In other words, they tend to be relatively stable over time. Nonetheless, longer-term treatments often produce changes precisely in these areas. Thus, the KAPP may be especially useful for researchers interested in understanding how longer-term therapies produce changes that may be unlikely to be obtained in brief treatment formats. Despite these strengths, continued research to demonstrate that the scales can be reliably coded when the measure is used by independent research labs would be of benefit. Overall, the KAPP remains an interesting and balanced measure of character that has much to offer psychodynamic investigators.

Self-Understanding of Interpersonal Patterns: Revised (SUIP-R)

The Self-Understanding of Interpersonal Patterns – Revised (SUIP-R) [89, 90] is a 28-item self-report measure that assesses individuals' awareness and understanding of problematic relationship patterns and conflicts. The SUIP-R was developed by investigators at the University of Pennsylvania Medical Center who have been extensively involved in psychotherapy process and outcome research for quite some time, and the tool was developed to be consistent with modern psychodynamic models of change [90]. Though initially developed, in part, to investigate factors that may mediate or explain how changes occur as a result of different interventions, the SUIP-R measures the development of insight which is a construct expected to improve in insight-oriented psychotherapies.

The SUIP-R is a revision of the Self-Understanding of Interpersonal Patterns (SUIP) [89]. The original SUIP included 19 items comprising two scales: recognition and self-understanding. Items for the SUIP were generated from patients' descriptions of common problems, contributions from expert therapists, and the Core Conflictual Relationship Theme (CCRT) standard category list.

The revision that produced the SUIP-R resulted in the inclusion of nine more items to better capture the breadth of relationship patterns. It also involved some structural changes to the scales (e.g., self-understanding was rated on a 6-point and opposed to 4-point scale). SUIP-R items are written as interpersonal patterns that may or may not be relevant to the respondent's relationship. The following is an example of some SUIP-R items:

- I feel the need to "save" others when I see them having a tough time and therefore try to solve their problems for them.
- I feel the need to keep someone close, and do whatever is necessary to keep him/her with me even when they need to leave me.
- I need to feel free of responsibility, and I distance myself from someone I care about because they are too dependent on me.
- I want another person to accept me as I am, but I bottle up my feelings and do whatever the other wants when I feel like he/she is putting me down.

All 28 items are rated in a two-step fashion. First, for each relationship pattern, respondents indicate how pervasive the problem is and how much they understand it by circling as many of the seven descriptors accurately describing how the pattern functions in their life and their understanding of it. Descriptors are letters (e.g., a, b, c) anchored to statements such as "I do not feel and act this way in my current relationships" (a), "I feel and act this way with multiple people in my life" (c), "I am in part responsible for continuing to feel and act this way with multiple people" (e), and "when I recognize that I am feeling and acting this way, I am able to consider other ways of viewing the situation in the moment" (g). Next, respondents indicate how important this experience is in their current relationships using a scale ranging from 1 (Not Important) to 10 (Very Important).

The SUIP-R produces a single self-understanding score. For items in which the letter "a" is circled, the pattern is not relevant for the current relationship, and thus the item is not used to calculate a score. For all items that are relevant (i.e., the respondent circled "b" or higher), the scale is treated like a Likert scale where a "b" = 1, a "c" = 2, and so on. Since respondents are allowed to circle more than one letter, the letter with the highest value is used in the final analysis. The self-understanding score is calculated by averaging all of the items not coded "a" together. Finally, the SUIP-R has both time 1 and time 2 forms to simplify in pre–post assessments.

Early examinations into the SUIP-R's psychometric properties suggest that the scale has excellent internal consistency (Coefficient Alpha = 0.92) and reasonable stability over a 1 month period (test–retest reliability of $r = 0.76$). It is also important to note that the psychometric properties for the original SUIP were systematically evaluated through multiple studies. Expert raters agreed on the clarity and coverage of the SUIP, scales were found to have adequate internal consistency (i.e., coefficient alphas ranging from 0.79 to 0.88), and corrected item-scale correlations were generally in the accepted ranges in both clinical and non-clinical samples.

Early investigations also support the construct validity for the SUIP-R. In a large clinical sample, baseline SUIP-R scores had correlations of less than 0.10 at baseline with measures of anxiety, depression, interpersonal problems, and overall quality of life [89]. Again, these results are quite similar to previous studies examining divergence and convergence of the SUIP which also found little correlation between SUIP scores and global symptom measures. Similarly, SUIP-R residual change scores show small, non-significant correlations with compensatory skill acquisition and changes in self-image, again suggesting that the measure is distinct. Such data suggest that the SUIP-R is assessing a unique construct (i.e., insight) that is separate from psychiatric distress, self-image, or skill acquisition.

The SUIP-R's construct validity of the tool and sensitivity to change has been investigated by its authors. In a pooled sample of patients undergoing one of five different forms of psychotherapy, there was significantly more change for SUIP-R scores for patients in dynamically focused psychotherapies (as compared to cognitive-oriented therapies) in which increasing awareness of interpersonal patterns

was a goal [90]. These results replicated and extended prior work with SUIP which found, as predicted by theory, that individuals undergoing psychodynamic-oriented treatment for ~~GAD~~ evidenced significantly greater changes in SUIP-R scores compared to controls receiving medication though both groups experienced significant reduction of symptoms [89].

The SUIP-R is a highly flexible measure capable of being used for many purposes in psychotherapy research: pre–post change, prediction of process, assessing stages of change, and understanding factors that moderate response to treatment. In addition, Connolly-Gibbons and colleagues' systematic early investigations of the original measures psychometric adequacy and cross-correlates with other measures serve to more clearly define the construct assessed by the instrument and highlight its distinctiveness from similar constructs (e.g., Openness to Emotions). The measure is also relatively easy to administer and score, and places minimal burden on research participants. Finally, psychodynamic psychotherapy researchers are likely to find the model on which the measure is built to be appealing and familiar. The measure also faces the same challenges of many self-report scales and also does not include scales to assess for validity of response. Still, given the overall strengths of the measure and relative ease of administration, it has much to offer as a psychotherapy research tool.

The Central Relationship Questionnaire (CRQ)

Conflicts that are central to the individual often manifest themselves in an interpersonal manner. These conflicts often involve patterns of wishes, fears, expectations, and behavioral responses. Psychodynamic psychotherapy often aims at generating insight into these patterns to facilitate change. Barber et al. [91] developed a self-report measure, the Central Relationship Questionnaire (CRQ), as a means to more systematically quantify and understand these types of problems and changes while placing minimal burden on the examiner. Similar to the SUIP-R just discussed, the CRQ is based on Luborsky's CCRT [92]. A unique and interesting feature of the CRQ is that it can be used to understand changes in patterns within specific relationships.

The original CRQ contained 139 items. Based on psychometric investigations (discussion further on in this chapter) and logistical factors, McCarthy et al. [93] recently developed a version of the CRQ with 101 items that maintains the same factor structure, has some enhanced scale properties, and reduces item redundancy. When completing the CRQ, respondents are asked to rate each item based on their relationship with a specific target relationship (e.g., spouse; parent; romantic partner), and respondents are often asked to complete multiple CRQs focused on several respective relationships. Prior to completing the core items for the measure, respondents rate the target relationship by responding to questions about key descriptors (e.g., intimacy; closeness; authority; importance). Respondents are also asked to rate the quality of the target relationship at its best and worst. Three sets of items comprise the CRQ and these are consistent with the CCRT: Wishes, Response from Others, and Response to Self. The Wishes component contains items tapping wishes, needs, and desires and produces seven subscales (e.g., The wish to be in conflict; The wish to be Independent). The Response from Others component contains items evaluating typical responses from others that either prevent or help individuals obtain wishes, wants, and needs and also includes seven subscales (e.g., Other Hurts Me; Other Loves Me). The Response to Self component has eight subscales revealing various self and other directed responses (e.g., Feel Valued; Feel Disliked; Am Domineering; for a detailed description of subscales [94]).

The Wishes component item set begins with instructions orienting respondents to the nature of the items (i.e., "Below is a list of different wishes, needs, or desires that people often have of other people."). Respondents are then asked to rate each Wish item based on how typical they apply to the target relationship when the relationship is/was at its worst. It is important to recall that items are rated in response to a target relationship. Thus, they may be rated to describe the relationship with a

romantic partner, a parent, a sibling, or a close friend. For the purpose of this chapter, we have worded example items as if the target relationship to be rated was a romantic partner. Here are example items from the Wishes component:

- I wish for my partner to know that I am loyal.
- I wish to confide in my partner.
- I would like my partner to feel proud of her/his accomplishments.
- I wish to dominate my partner.

The Response from Other component item set begins with instructions asking respondents to consider how the person in the target relationship typically "responds to you." Respondents are oriented to this section with the statement: "We often see people as responding to us in a way that either prevents us from getting what we want, or helps us to get what we want. Here is a list of possible ways that a romantic partner can respond to you." Respondents are then asked to rate items for how typical they are when the relationship is/was at its worst. Response from Others component example items are presented here:

- My partner withdraws.
- My partner feels I am a special person.
- My partner controls me.
- My partner is emotionally close to me.

Finally, prior to completing the Response to Self component items, participants are oriented to this section with the statement: "Other people can deny your desires or meet your desires in responding to you. Below is a list of different ways that you might react when your romantic partner denies or meets your desires." Respondents are then asked to rate items for how well they reflect their typical reactions when their partner denies or meets their desires when the relationship is/was at its worst. Following are example items from the Response to Self component worded as if the target relationship to be rated was a romantic

No italics here.

- I accomplish my goals.
- I feel disliked.
- I *avoid* difficulties with my partner.
- I am confused by my relationship with my partner.

All items across scales are rated on a 7-point scale ranging from 1 (never true or typical of me) to 7 (always true or typical of me). Higher ratings indicate a greater presence of that Wish, Response from Other, or Response to Self within the target relationship. Roughly half of the CRQ subscales are positively valenced (i.e., affiliation; respect for self; respect for others; positive affect) and half negatively valenced (e.g., antagonism; lack of self-respect; failure to respect others; negative affect). The CRQ takes some time for a respondent to fill out, though the total time will depend on the number of relationships to be rated.

To date, a small number of detailed studies have examined the scale properties of the CRQ. The most extensive investigation of the scale made use of three samples: a large student sample; a clinical sample; and a retest sample [91]. Across samples, subscale internal consistency ratings were adequate (Alphas ranged from 0.78 to 0.95 for the Wishes component, 0.82 to 0.95 for the Response from Other component, and 0.71 to 0.94 for the Response to Self). Nineteen of twenty-two scales had alphas of >0.80. Corrected item-scale correlations were also in the acceptable range ranging from 0.44 to 0.90 for the Wishes component, 0.58 to 0.85 for the Response from Other component, and 0.42 to 0.87 for the Response from Self component. Test–retest reliability was established by comparing ratings of 54 non-clinical respondents over a 1-year period. Average test–retest reliability was >0.60 for all subscales. Subsequent scale analyses [94], conducted with a Swedish student sample, Swedish clinical sample, and a North American student sample, have revealed a pattern

of subscale correlations that was highly consistent with those originally reported by Barber and colleagues [91]. A more recent study also found that a revised CRQ with fewer items maintained a similar factor structure to the original and possessed similar, and in some cases stronger, psychometric properties [93].

Validity of the CRQ has been established by demonstrating that subscale ratings are associated with other measures of interpersonal problems, neuroticism, depression, and psychiatric symptoms. Additionally, as would be expected, clinical populations in distress produce higher scores on negatively valenced items (i.e., items tapping problematic relational patterns) than do non-clinical samples [91]. These findings have been replicated across two cultures [94]. Subscale interrelations provide some support for the three-component structure of the measure. To date, the CRQ has yet to be utilized as an outcome measure for psychotherapy. However, studies are currently underway to establish the measure's utility as a psychotherapy research tool and to evaluate sensitivity to change.

The CRQ provides a means for studying changes in relational patterns. The measure is still relatively new, and a revised version has recently been created. As such, further research regarding the scale properties, particularly for the revised measure, would be of benefit. The authors of the CRQ are currently in the process of assessing the utility of the instrument for psychotherapy research. The measure also contains many items requiring a fair amount of the respondents' time to complete. It is possible that this may limit the use of the tool. Nonetheless, changes in CCRTs are often a central part of what occurs in effective psychodynamic psychotherapy, as such the CRQ provides a means for assessing this key area of change. It is also one of the few tools of its kind available for assessing specific relationships.

Conclusions

When we initially set out to write this chapter, we had a handful of outcome measures in mind that focused on key psychodynamic constructs. However, we speculated that it might be necessary to borrow some measures from the field of social psychology to complete the mix. We quickly discovered that there are currently far more measures suitable for understanding change from psychodynamic psychotherapy than may have been anticipated. In fact, the topic may be more worthy of a book as opposed to a book chapter at this time. Thus, given limits of space and pages for this chapter, investigators may wish to examine similar reviews focusing on measures of attachment (e.g. [95]), measures of object relations (e.g.[17]), measures conceptualizing change in psychodynamic constructs (e.g. [96]), and measures of structural capacities (e.g. [97]). Koelen et al. [98] have also recently concluded a paper that reviews a number of measures that can be used in psychodynamic psychotherapy research as outcome measures or to predict psychotherapy process and/or treatment outcomes.

In addition to establishing treatment efficacy, outcome measures, such as the measures reviewed here, can also be employed in other types of research. The measurement of outcomes often compliments the study of psychotherapy process. Assessing outcome is also essential for actively testing theories underlying treatment effects, the role of specific interventions, and the process of change in treatment [4]. After all, if an underlying presumption of a particular therapeutic approach or intervention is that it facilitates a particular type of change, then it will be necessary to determine if change in that domain occurs as a result of the treatment or intervention. In short, understanding the outcome of treatment can also aid in understanding the process of treatment and stages of change.

The range and sheer number of such measures says something about the health of the field overall. Given the number of options, psychodynamic investigators are now in a position where they can pick and choose instruments to suit their research purposes. In addition, they are now capable of

assessing constructs at multiple ~~lefts~~ levels by making use of combinations of measurement methods to study explicit and implicit mechanisms of change. For example, observer-rated measures, self-report measures, and projective-narrative-based measures may all assess slightly different aspects (e.g., implicit; explicit) of a construct (e.g., object relations) allowing for investigations of some of the smaller nuances of change. It is an exciting period of time for psychodynamic researchers, and the growth of theory specific outcome measures continues to be an important part of the field's growth.

[AU5] ## References

1. Wallerstein RS. Psychoanalytic therapy research: it's coming of age. Psychoanal Inq. 2003;23:375–404.
2. Wallerstein RS, Fonagy P. Psychoanalytic research and the IPA: history, present status and future potential. Int J Psychoanal. 1999;80:91–109.
3. Shedler J. The efficacy of psychodynamic psychotherapy. Am Psychol. 2010;65(2):98–109.
4. Høglend P, Bøgwald KP, Amlo S, Marble A, Ulberg R, Sjaastad MC, et al. Transference interpretations in dynamic psychotherapy: do they really yield sustained effects? Am J Psychiatry. 2008;165:763–71.
5. Morey C. Personality assessment inventory professional manual. 2nd ed. Odessa: Psychological Assessment Resources; 2007.
6. Weinryb RM, Rössel RJ. Karolinska psychodynamic profile – KAPP. Acta Psychiatr Scand. 1991;83:1–23.
7. Høglend P, Bøgwald KP, Amlo S, Heyerdahl O, Sørbye O, Marble A, et al. Assessment of change in dynamic psychotherapy. J Psychother Pract Res. 2000;9:190–9.
8. Westen D, Shedler J. Revising and assessing axis II. Part 1: Developing a clinically and empirically valid assessment method. Am J Psychiatry. 1999;156:258–72.
9. Siefert CJ, ~~Deifie~~ DeFife JA, Baity MR. Process measures for psychodynamic psychotherapy. In: Levy RA, Ablon JS, editors. Handbook of evidence-based psychodynamic psychotherapy: bridging the gap between science and practice. Totowa: Humana Press; 2009. p. 157–78.
10. Westen D. Social cognition and object relations scale: Q-sort for projective stories (SCORS-G-Q). Unpublished manuscript. Cambridge, MA: Cambridge Hospital and Harvard Medical School; 1995.
11. Hilsenroth M, Stein M, Pinsker J. Social cognition and object relations scale: global rating method (SCORS-G-G). Unpublished manuscript. Garden City: The Derner Institute of Advanced Psychological Studies, Adelphi University; 2004.
12. Eudell-Simmons EM, Stein M, DeFife JA, Hilsenroth M. Reliability and validity of the social cognition and object relations scale (SCORS-G) in the assessment of dream narratives. J Pers Assess. 2005;85:325–33.
13. Stein M, Hilsenroth M, Pinsker-Aspen J, Primavera L. Validity of DSM-IV axis V global assessment of relational functioning scale. J Nerv Ment Dis. 2009;197:50–5.
14. Slaven JM, Stein M, Pinsker-Aspen J, Hilsenroth M. Early memories from outpatients with and without a history of childhood sexual abuse. J Loss Trauma. 2007;12:435–51.
15. Ackerman SJ, Hilsenroth M, Clemence AJ, Weatherill R, Folwer JC. The effects of social cognition and object representation on psychotherapy continuation. Bull Menninger Clin. 1999;64:386–408.
16. Ackerman SJ, Hilsenroth M, Clemence AJ, Weatherill R, Fowler JC. Convergent validity of Rorschach and TAT scales of object relations. J Pers Assess. 2001;77:295–306.
17. Huprich SK, Greenberg RP. Advances in the assessment of object relations in the 1990s. Clin Psychol Rev. 2003;23:665–98.
18. Porcerelli JH, Shahar G, Blatt SJ, Ford RQ, Mezza JA, Greenlee LM. Social cognition and object relations scale: convergent validity and changes following intensive inpatient treatment. Pers Individ Dif. 2004;41:407–17.
19. Porcerelli JH, Cogan R, Hibbard S. Cognitive and affective representations of people and the MCMI-II personality psychopathology. J Pers Assess. 1998;70:535–40.
20. Westen D, Lohr N, Silk KR, Gold L, Kerber K. Object relations and social cognition in borderlines, major depressives, and normals: a thematic apperception test analysis. Psychol Assess. 1990;2:355–64.
21. Ackerman SJ, Clemence AJ, Weatherill R, Hilsenroth MJ. Use of the TAT in the assessment of DSM-IV cluster B personality disorders. J Pers Assess. 1999;73(3):422–42.
22. Ford JD, Fisher P, Larson L. Object relations as a predictor of treatment outcome with chronic posttraumatic stress disorder. J Consult Clin Psychol. 1997;65:547–59.
23. Fowler JC, Ackerman SJ, Speanburg S, Bailey A, Blagys M, Conklin AC. Personality and symptom change in treatment-refractory inpatients: evaluation of the phase model of change using Rorschach, TAT, and DSM-IV axis V. J Pers Assess. 2004;83:306–22.

24. Fowler C, Hilsenroth MJ, Handler L. Two methods of early memories data collection: an empirical comparison of the projective yield. Assessment. 1996;3:63–71.
25. Westen D, Shedler J. Rev 2: Toward an empirically based and clinically useful classification of personalit 999;156:273–85.
26. Shedler J, Westen D. The rocedure (SWAP-200): making personality diagnosis clinically meaningful. J Pe
27. DeFife J, Westen D. Empirically-informed clinical interviewing for personality disorders. In: Levy R, Ablon J, editors. Handbook of evidence-based psychodynamic psychotherapy. Vol. 2 (this volume).
28. Westen D, Muderrisoglu S. Reliability and validity of personality disorder assessment using a systematic clinical interview: evaluating an alternative to structured interviews. J Pers Disord. 2003;17:350–68.
29. Lingiardi V, Shedler J, Gazzillo F. Assessing personality change in psychotherapy with the SWAP-200: a case study. J Pers Assess. 2006;86(1):23–32.
30. Marin-Avellan L, McGauley G, Campbell C, Fonagy P. Using the SWAP-200 in a personality-disordered forensic population; is it valid, reliable and useful? J Crim Behav Ment Health. 2005;15:28–45.
31. Bradley R, Hilsenroth M, Guarnaccia C, Westen D. Relationship between clinician assessment and self-assessment of personality disorders using the SWAP-200 and PAI. Psychol Assess. 2007;19(2):225–9.
32. Russ E, Heim A, Westen D. Parental bonding and personality pathology assessed by clinician report. J Pers Disord. 2003;17:522–36.
33. Martin-Avellan LE, McGauley G, Campbell C, Fonagy P. Using the SWAP-200 in a personality-disordered forensic population: is it valid, reliable and useful? Crim Behav Ment Health. 2005;15(1):28–45.
34. Shedler J, Westen D. Refining personality disorder diagnoses: integrating science and practice. Am J Psychiatry. 2004;161:1350–65.
35. Cogan R, Porcerelli J. Clinician reports of personality pathology of patients beginning and patients ending psychoanalysis. Psychol Psychother Theor Res Pract. 2005;78(2):235–48.
36. Perry JC. Defense mechanism rating scales (DMRS). 5th ed. Cambridge, MA; 1990. [AU6]
37. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994.
38. Perry JC, Beck SM, Constantinides P, Foley JE. Studying change in defensive functioning in psychotherapy using the defense mechanism rating scales: four hypotheses, four cases. In: Levy RA, Ablon JS, editors. Handbook of evidence-based psychodynamic psychotherapy. Totowa: Humana Press; 2009.
39. Perry JC, Høglend P. Convergent and discriminant validity of overall defensive functioning. J Nerv Ment Dis. 1998;186(9):529–35.
40. Blais MA, Conboy CA, Wilcox N, Norman DK. An empirical study of the DSM-IV defensive functioning scale in personality disordered patients. Compr Psychiatry. 1996;37:435–40.
41. Hilsenroth MJ, Callahan KL, Eudell EM. Further reliability, convergent and discriminant validity of overall defensive functioning. J Nerv Ment Dis. 2003;191:730–7.
42. Perry JC, Henry M. Studying defense mechanisms in psychotherapy using the defense mechanism rating scales. In: Hentschel U, Smith G, Draguns J, Ehlers W, editors. Defense mechanisms: theoretical, research, and clinical perspectives. Amsterdam: Elsevier; 2004. p. 165–92.
43. Perry JC. A pilot study of defenses in adults with personality disorders entering psychotherapy. J Nerv Ment Dis. 2001;189:651–60.
44. DeFife JA, Hilsenroth MJ. Clinical utility of the defensive functioning scale in the assessment of depression. J Nerv Ment Dis. 2005;193(3):176–82.
45. Drapeau M, De Roten Y, Perry JC, Despland J-N. A study of stability and change in defense mechanisms during a brief psychodynamic investigation. J Nerv Ment Dis. 2003;191(8):496–502.
46. Roy CA, Perry JC, Luborsky L, Banon E. Changes in defensive functioning in completed psychoanalyses: the Penn psychoanalytic treatment collection. J Am Psychoanal Assoc. 2009;57(2):399–415.
47. Horowitz LM, Rosenberg SE, Baer BA, Ureño G, Villaseñor VS. Inventory of interpersonal problems: psychometric properties and clinical applications. J Consult Clin Psychol. 1988;56(6):885–92.
48. Alden LE, Wiggins JS, Pincus AL. Construction of circumplex scales for the inventory of interpersonal problems. J Pers Assess. 1990;55(3–4):521–36.
49. Horowitz LM, Alden LE, Wiggins JS, Pincus AL. Inventory of interpersonal problems. San Antonio: Psychological Corporations; 2000.
50. Hughes J, Barkham M. Scoping the inventory of interpersonal problems, its derivatives and short forms: 1988–2004. Clin Psychol Psychother. 2005;12(6):475–96.
51. Gurtman MB. Interpersonal problems and the psychotherapy context: the construct validity of the inventory of interpersonal problems. Psychol Assess. 2006;8(3):241–55.
52. Haggerty G, Hilsenroth MJ, Vala-Stewart R. Attachment and interpersonal distress: examining the relationship between attachment styles and interpersonal problems in a clinical population. Clin Psychol Psychother. 2009; 16(1):1–9.

Please update this reference to Chapter 28.

53. Ruiz MA, Pincus AL, Borkovec TD, Echemendia RJ, Castonguay LG, Ragusea SA. Validity of the inventory of interpersonal problems for predicting treatment outcome: an investigation with the Pennsylvania Practice Research Network. J Pers Assess. 2004;83(3):213–22.
54. Høglend P, Bøgwald K-P, Amlo S, Marble A, Ulberg R, Sjaastad MC, et al. Transference interpretations in dynamic psychotherapy: do they really yield sustained effects? Am J Psychiatry. 2008;165(6):763–71.
55. Paley G, Cahill J, Barkham M, Shapiro D, Jones J, Patrick S, et al. The effectiveness of psychodynamic-interpersonal therapy (PIT) in routine clinical practice: a benchmarking comparison. Psychol Psychother Theor Res Pract. 2008;81(2):157–75.
56. Crits-Christoph P, Gibbons MBC, Narducci J, Schamberger M, Gallop R. Interpersonal problems and the outcome of interpersonally oriented psychodynamic treatment of GAD. Psychother Theor Res Pract Train. 2005; 42(2):211–24.
57. George C, Kaplan N, Main M. Adult attachment interview protocol. 3rd ed. Unpublished manuscript. Berkeley: University of California; 2006.
58. Steele H, Steele M. Clinical applications of the adult attachment interview. New York: Guilford Press; 2008.
59. Hesse E. The adult attachment interview: historical and current perspectives. In: Cassidy J, Shaver PR, editors. Handbook of attachment: theory, research, and clinical applications. New York: Guilford Press; 1999. p. 395–433.
60. Hesse E. The adult attachment interview: protocol, method of analysis, and empirical studies. In: Cassidy J, Shaver PR, editors. Handbook of attachment: theory, research, and clinical applications. 2nd ed. New York: Guilford Press; 2008. p. 552–98.
61. Roisman GI, Fraley RC, Belsky J. A taxometric study of the adult attachment interview. Dev Psychol. 2007; 43(3):675–86.
62. Bakermans-Kranenburg MJ, van Ijzendoorn MH. The first 10,000 adult attachment interviews: distributions of adult attachment representations in clinical and non-clinical groups. Attach Hum Dev. 2009;11(3):223–63.
63. Bakermans-Kranenburg MJ, Van Ijzendoorn MH. A psychometric study of the adult attachment interview: reliability and discriminant validity. Dev Psychol. 1993;29(5):870–9.
64. Sagi A, van Ijzendoorn MH, Scharf M, Koren-Karie N, Joels T, Mayseless O. Stability and discriminant validity of the adult attachment interview: a psychometric study in young Israeli adults. Dev Psychol. 1994;30(5):771–7.
65. Waters E, Hamilton CE, Weinfield NS. The stability of attachment security from infancy to adolescence and early adulthood: general introduction. Child Dev. 2000;71(3):678–83.
66. Allen JP, McElhaney KB, Kuperminc GP, Jodl KM. Stability and change in attachment security across adolescence. Child Dev. 2004;75(6):1792–805.
67. Main M, Hesse E, Kaplan N. Predictability of attachment behavior and representational processes at 1, 6, and 19 years of age: the Berkeley longitudinal study. In: Grossmann KE, Grossmann K, Waters E, editors. Attachment from infancy to adulthood: the major longitudinal studies. New York: Guilford Publications; 2005. p. 245–304.
68. van Ijzendoorn M. Adult attachment representations, parental responsiveness, and infant attachment: a meta-analysis on the predictive validity of the adult attachment interview. Psychol Bull. 1995;117(3):387–403.
69. Levy KN, Meehan KB, Kelly KM, Reynoso JS, Weber M, Clarkin JF, et al. Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. J Consult Clin Psychol. 2006;74(6):1027–40.
70. Gullestad SE. The adult attachment interview and psychoanalytic outcome studies. Int J Psychoanal. 2003;84(3):651–68.
71. Jurist EL. Mentalized affectivity. Psychoanal Psychol. 2005;22(3):426–44.
72. Fonagy P, Steele M, Steele H, Target M. Reflective function manual for application to adult attachment interviews. London: University College London; 1997.
73. Rudden M, Milrod B, Target M, Ackerman S, Graf E. Reflective functioning in panic disorder patients: a pilot study. J Am Psychoanal Assoc. 2006;54(4):1339–43.
74. Bouchard MA, Target M, Lecours S, Fonagy P, Tremblay LM, Schachter A. Mentalization in adult attachment narratives: reflective functioning, mental states, and affect elaboration compared. Psychoanal Psychol. 2008;25(1):47–66.
75. Fonagy P, Steele M, Steele H, Moran GS. The capacity for understanding mental states: the reflective self in parent and child and its significance for security of attachment. Infant Ment Health J. 1991;12(3):201–18.
76. Slade A, Grienenberger J, Bernbach E, Levy D, Locker A. Maternal reflective functioning, attachment, and the transmission gap: a preliminary study. Attach Hum Dev. 2005;7(3):283–98.
77. Fonagy P, Leigh T, Steele M, Steele H, Kennedy R, Mattoon G, et al. The relation of attachment status, psychiatric classification, and response to psychotherapy. J Consult Clin Psychol. 1996;64(1):22–31.
78. Meehan KB, Levy KN, Reynoso JS, Hill LL, Clarkin JF. Measuring reflective function with a multidimensional rating scale: comparison with scoring reflective function on the AAI. J Am Psychoanal Assoc. 2009;57(1): 208–13.
79. Bøgwald K-P, Dahlbender RW. Procedures for testing some aspects of the content validity of the psychodynamic functioning scales and the global assessment of functioning scale. Psychother Res. 2004;14(4):453–68.

80. Weinryb RM, Rössel RJ, Gustavsson JP, Åsberg M, Barber JP. The Karolinska psychodynamic profile (KAPP): studies of character and well-being. Psychoanal Psychol. 1997;14(4):495–515.
81. Vinnars B, Thormählen B, Gallop R, Norén K, Barber J. Do personality problems improve during psychodynamic supportive–expressive psychotherapy? Secondary outcome results from a randomized controlled trial for psychiatric outpatients with personality disorders. Psychother Theor Res Pract Training. 2009;46(3):362–75.
82. Heikkilä J, Karlsson H, Taiminen T, Lauerma H, Ilonen T, Leinonen KM, et al. Psychodynamic personality profile in first-episode severe mental disorders. Acta Psychiatr Scand. 2004;109(3):187–93.
83. Weinryb RM, Rössel RJ, Åsberg M. The Karolinska psychodynamic profile: II. Interdisciplinary and cross-cultural reliability. Acta Psychiatr Scand. 1991;83(1):73–6.
84. Weinryb RM, Busch M, Gustavsson JP, Saxon L, Skarbrandt E. Reliability of the Karolinska psychodynamic profile (KAPP) among patients with and without psychoactive substance abuse disorders. Psychother Psychosom. 1998;67(1):10–6.
85. Haver B, Svanborg P, Lindberg S. Improving the usefulness of the Karolinska psychodynamic profile in research: proposals from a reliability study. Acta Psychiatr Scand. 1995;92(2):132–7.
86. Weinryb RM, Gustavsson JP, Åsberg M, Rössel RJ. Stability over time of character assessment using a psychodynamic instrument and personality inventories. Acta Psychiatr Scand. 1992;86(2):179–84.
87. Wilczek A, Barber JP, Gustavsson JP, Åsberg M, Weinryb RM. Change after long-term psychoanalytic psychotherapy. J Am Psychoanal Assoc. 2004;52(4):1163–84.
88. Fassino S, Daga GA, Delsedime N, Busso F, Pierò A, Rovera GG. Baseline personality characteristics of responders to 6-month psychotherapy in eating disorders: preliminary data. Eat Weight Disord. 2005;10(1):40–50.
89. Connolly Gibbons MB, Crits-Christoph P, Shelton RC, Hollon S, Kurtz J, Barber JP. The reliability and validity of a measure of self-understanding of interpersonal patterns. J Couns Psychol. 1999;4:472–82.
90. Connolly Gibbons MB, Crits-Christoph P, Barber JP, Wiltsey Stirman S, Gallop R, Goldstein LA, et al. Unique and common mechanisms of change across cognitive and dynamic psychotherapies. J Consult Clin Psychol. 2009;77:801–13.
91. Barber JP, Foltz C, Weinryb RM. The Central Relationship Questionnaire: initial report. J Couns Psychol. 1998;45(2):131–42.
92. Luborsky L. Principles of psychoanalytic psychotherapy: a manual for supportive–expressive (SE) treatment. New York: Basic Books; 1984.
93. McCarthy KS, Gibbons MBC, Barber JP. The relation of rigidity across relationships with symptoms and functioning: an investigation with the revised Central Relationship Questionnaire. J Couns Psychol. 2008;55(3):346–58.
94. Weinryb RM, Barber JP, Foltz C, Göransson SGM, Gustavsson JP. The Central Relationship Questionnaire (CRQ): psychometric properties in a Swedish sample and cross-cultural studies. J Psychother Pract Res. 2000;9(4):201–12.
95. Bartholomew K, Shaver PR. Methods of assessing adult attachment: do they converge? In: Simpson JA, Rholes WS, editors. Attachment theory and close relationships. New York: Guilford Press; 1998. p. 25–45.
96. Blatt SJ, Auerbach JS. Psychodynamic measures of therapeutic change. Psychoanal Inq. 2003;23(2):268–307.
97. Huber D, Henrich G, Klug G. The scales of psychological capacities: measuring change in psychic structure. Psychother Res. 2005;15(4):445–56.
98. Koelen JA, Luyten P, Eurelings-Bontekoe EHM, Diguer L, Vermote R, Lowyck B, et al. The impact of personality organization on treatment response and the therapeutic process. Unpublished manuscript.

Appendix 27.1 The Social Cognition and Object Relations: Global Rating Scale

Complexity of representation of people: 1 = is egocentric, or sometimes confuses thoughts, feelings, or attributes of the self and others; 3 = tends to describe people's personalities and internal states in minimally elaborated, relatively simplistic ways, or splits representations into good and bad; 5 = representations of the self and others are stereotypical or conventional, is able to integrate both good and bad characteristics of self and others, has awareness of impact on others; 7 = is psychologically minded, insight into self and others, differentiated and shows considerable complexity

1 2 3 4 5 6 7

Affective quality of representations: (i.e., what the person expects from relationships, and how he/she tends to experience significant others and describe significant relationships): 1 = malevolent, abusive, caustic; 3 = largely negative or unpleasant, but not abusive; 5 = mixed, neither primarily positive nor primarily negative (needs to have some positive to be scored 5); 7 = generally positive expectations of relationships (but not pollyannaish), a favorable and affirmative view of relationships Note: where affective quality is absent, bland, or limited, code 4

1 2 3 4 5 6 7

Emotional investment in relationships: 1 = tends to focus primarily on his/her own needs in relationships, has tumultuous relationships, or has few if any relationships; 3 = somewhat shallow relationships, or only alludes to others; 5 = demonstrates conventional sentiments of friendship, caring, love, and empathy; 7 = tends to have deep, committed relationships with mutual sharing, emotional intimacy, interdependence, and respect, positive connectedness and appreciation of others Note: where only one character is described and no relationship is depicted, code 2

1 2 3 4 5 6 7

Emotional investment in values and moral standards: 1 = behaves in selfish, inconsiderate, self-indulgent or aggressive ways without any sense of remorse or guilt; 3 = shows signs of some internalization of standards (e.g., avoids doing "bad" things because knows will be punished for them, thinks in relatively childlike ways about right and wrong, etc.), or is morally harsh and rigid toward self or others; 5 = is invested in moral values and tries to live up to them; 7 = thinks about moral questions in a way that combines abstract thought, a willingness to challenge or question convention, and genuine compassion and thoughtfulness in actions (i.e., not just intellectualized) Note: where no moral concerns are raised in a particular story, code 4

1 2 3 4 5 6 7

Understanding of social causality: 1 = narrative accounts of interpersonal experiences are confused, distorted, extremely sparse, or difficult to follow, limited awareness and coherence; 3 = understands people in relatively simple, but sensible ways, or describes interpersonal events in ways that largely make sense but may have a few gaps or incongruities; 5 = tends to provide straightforward narrative accounts of interpersonal events in which people's actions result from the way they experience or interpret situations; 7 = tends to provide particularly coherent narrative accounts of interpersonal events, and to understand people very well, understands the impact of their behavior on others and others behavior on them Note: where subject describes interpersonal events as if they just happen, with little sense of why people behave the way they do (i.e., alogical rather than illogical stories that seem to lack any causal understanding), code 2

1 2 3 4 5 6 7

Experience and management of aggressive impulses: 1 = physically assaultive, destructive, sadistic, or in poor control of aggression, impulsive; 3 = angry, passive-aggressive, denigrating, or physically abusive to self (or fails to protect self from abuse); 5 = avoids dealing with anger by denying it, defending against it, or avoiding confrontations; 7 = can express anger and aggression and assert self appropriately Note: if no anger content in the story, code 4

1 2 3 4 5 6 7

Self-esteem: 1 = views self as loathsome, evil, rotten, contaminating, or globally bad; 3 = has low self-esteem (e.g., feels inadequate, inferior, self-critical, etc.) or is unrealistically grandiose; 5 = displays a range of positive and negative feelings toward the self; 7 = tends to have realistically positive feelings about him/herself

Note: needs to have some positive to be scored a 5 or above

1 2 3 4 5 6 7

Identity and coherence of self: 1 = fragmented sense of self, has multiple personalities; 3 = views of, or feelings about, the self fluctuate widely and unpredictably; unstable sense of self; 5 = identity and self-definition are not a major concern or preoccupation; 7 = feels like an integrated person with long-term ambitions and goals

Note: ambiguity about a goal is still considered a goal and may be scored in the higher range

1 2 3 4 5 6 7

Reprinted with permission from [11]. Originally adapted from [10].

Author Queries

Chapter No.: 27 0001331407

Queries	Details Required	Author's Response
AU1	First author has been treated as corresponding author, please check.	Correct
AU2	The author name "Westen and Shedler" given along with Ref. [27] does not match with the list. Please check.	See line 213
AU3	The author name "Gurtman and Pincus" given along with Ref. [51] does not match with the list. Please check.	See line 291
AU4	Please check the edit made in the sentence starting "For example, Gullestad…" and correct if necessary.	OK
AU5	Refs. [6] & [80] are identical. Hence Ref. [80] has been deleted and the subsequent references are renumbered both in text as well as in the list. Please check.	OK
AU6	Please provide publisher name for Ref. [36].	

"published by the author"