Psychoanalytic Psychology

Mentalization as a Predictor of Psychoanalytic Outcome: An Empirical Study of Transcribed Psychoanalytic Sessions Through the Lenses of a Computerized Text Analysis Measure of Reflective Functioning

Tommaso Boldrini, Maria Paola Nazzaro, Rachele Damiani, Federica Genova, Francesco Gazzillo, and Vittorio Lingiardi

Online First Publication, July 24, 2017. http://dx.doi.org/10.1037/pap0000154

Psychoanalytic Psychology

0736-9735/17/$12.00 http://dx.doi.org/10.1037/pap0000154

Mentalization as a Predictor of Psychoanalytic Outcome: An Empirical Study of Transcribed Psychoanalytic Sessions Through the Lenses of a Computerized Text Analysis Measure of Reflective Functioning

Tommaso Boldrini, PhD Student, Maria Paola Nazzaro, PhD, Rachele Damiani, MSc, Federica Genova, PhD, Francesco Gazzillo, PhD, and Vittorio Lingiardi, MD
Sapienza University of Rome

The Reflective Functioning Scale (Fonagy, Target, Steele, & Steele, 1998) was developed to empirically assess the capacity to mentalize thoughts, intentions, feelings and beliefs of oneself and others in the context of attachment relationships (Jurist & Meehan, 2009). To overcome the complexity of the RF scale scoring, the Computerized Text Analysis measure of Reflective Functioning (CRF) was created by Fertuck, Mergenthaler, Target, Levy, and Clarkin (2012). We report the results of a preliminary study applying the CRF to a sample of 540 sessions comprising 27 psychoanalytic treatments. Results show that patients' reflective functioning (RF), as measured at the beginning of treatment, was positively correlated with two global measures of healthy personality functioning—the *Global Assessment of Functioning* (GAF; American Psychiatric Association, 2000) and the *Personality Health Index* (PHI; Waldron et al., 2011)—as measured at the end of treatment. Even when the PHI and GAF levels at the beginning of treatment, the length of the treatments, and the average number of sessions per week were controlled for, these correlations remained significant. At the same time, the RF of patients did not increase throughout treatment. The implications of these results and the validity of CRF as an outcome predictor of long-term psychoanalytic treatment are discussed.

Keywords: computerized text analysis, reflective functioning, psychoanalytic treatment, therapy outcome

Mentalization (Allen, 2003; Bateman & Fonagy, 2016; Fonagy, Gergely, Jurist, & Target, 2002) can be defined as a form of "imaginative" mental activity, mostly preconscious, that interprets human behaviors in terms of intentional mental states (e.g., needs, desires, feelings, beliefs, goals, intentions, and motivations). Mentalization theory is an important contribution to the field of clinical psychology and provides a guide to treatments integrating psychoanalysis and attachment theory (Jurist & Meehan, 2009). Moreover, mentalization theory is grounded in psychoanalysis, neuroscience, cognitive psychology and philosophy; thus, it is a truly interdisciplinary theory (Jurist, Slade, & Bergner, 2008).

The term "reflective function" (RF) is the empirical operationalization of the construct of mentalization. It is assessed by the Reflective Functioning Scale (RFS; Fonagy, Target, Steele, & Steele, 1998), which provides a measure of the psychological processes underlying the ability to imagine and think about one's and others' mental states in order to construct realistic models for interpreting behaviors, thoughts, and feelings. RF is a narrative-based measure of mentalization mostly assessed from Adult Attachment Interviews (AAI; George, Kaplan, & Main, 1996), in contrast with other aspects of mentalization from social–cognitive measures (such as theory of mind or metacognition).

Several studies have investigated how an inhibition in understanding one's and others' mental states relates to psychopathology and personality organization, in general. Deficits in mentalization have been found to characterize a broad range of mental disorders (for a review, see Katznelson, 2014), in particular personality disorders (e.g., Antonsen, Johnansen, Rø, Kvarstein, & Wilberg, 2016; Fonagy, Luyten, & Bateman, 2015). Consequently, the assessment of RF appears relevant to psychotherapeutic processes. For example, it is reasonable to predict that patients with different levels of mentalization will differ in their patterns of change and in their outcomes in response to therapy. Patients with low levels of RF can be assumed to be more challenging for their therapists; furthermore, the proclivity to act out, suggesting low levels of mentalization, may impair the ability to establish and maintain a good therapeutic alliance (e.g., Lingiardi, Tanzilli, & Colli, 2015; Tanzilli, Colli, Gualco, & Lingiardi, 2017). Bateman and Fonagy (2004a) hypothesized that, independently from the treatment model implemented, a good outcome therapy is accompanied by an increase in mentalization; therefore, mentalization can be considered a process factor that is common to different

Tommaso Boldrini, PhD Student, Maria Paola Nazzaro, PhD, Rachele Damiani, MSc, Federica Genova, PhD, Francesco Gazzillo, PhD, and Vittorio Lingiardi, MD, Department of Dynamic and Clinical Psychology, Sapienza University of Rome.

We gratefully acknowledge Sherwood Waldron and the Psychoanalytic Research Consortium (PRC, New York) to have shared the audio recordings of the psychotherapy sessions evaluated in this study. The findings reported in the manuscript have not been previously published and the manuscript is not being simultaneously submitted elsewhere. The original research procedures were consistent with the principles of research ethics, published by the American Psychological Association.

Correspondence concerning this article should be addressed to Tommaso Boldrini, PhD Student, Department of Dynamic and Clinical Psychology, Faculty of Medicine and Psychology, Sapienza University of Rome, Via degli Apuli, 1, 00185 Rome, Italy. E-mail: tommaso.boldrini@hotmail.it

treatment models. Indeed, in order to achieve substantial therapeutic change, many forms of psychotherapy target—either directly or indirectly—the promotion of mentalization in patients. The patient–therapist dyad, in fact, provides an interpersonal environment in which the reflective ability of the patient can flourish, thanks also to the therapist's help. According to Holmes (2001), an effective therapeutic treatment improves the patient's narrative competence and, according to Bateman and Fonagy (2004b), "the potential effectiveness of all treatments depends not so much on their frame, but on their ability to increase a patient's capacity to mentalize" (p. 46).

Actually, RF could be considered either (a) a mediator of change or (b) a moderator and/or predictor of outcome. With regards to the former role, which highlights RF as a change mediator, Levy and colleagues (2006) conducted a clinical trial on three types of manualized psychotherapy for the treatment of patients with borderline personality disorders (transference focused psychotherapy [TFP], psychodynamic supportive therapy, and dialectical behavior therapy), evaluating the importance of RF as an outcome measure. Their results showed that TFP significantly promotes an increase in RF (assessed on Adult Attachment Interview [AAI] transcripts) compared to the other two forms of psychotherapy studied after 1 year of treatment. The authors suggested that this finding may indirectly support the hypothesis that a change in mentalizing capacity is central to the efficacy of TFP. Also, in a recent 1-year randomized controlled trial in which TFP was compared with community therapists' treatment (Fischer-Kern et al., 2015), only patients in the TFP group showed a significant improvement in RF and reduced levels of impairment in personality organization as assessed with the Structured Interview of Personality Organization (STIPO; Clarkin, Caligor, Stern, & Kernberg, 2004). These changes imply a more balanced and stable representation of oneself and others, a lessening of the tendency to act out, and better capacity of affect regulation.

RF has been found to change as a function of not only treatment type, but also patient characteristics. Vermote (2005), analyzing RF as determined by transcripts of sessions of psychodynamic psychotherapy, identified a different pattern of change in different types of patients (anaclitic vs. introjective; Blatt & Ford, 1994). In introjective patients, reflective capacity progressively increased during the treatment. Anaclitic patients, however, showed a decrease in reflective capacity in the early stages of treatment and an increase in RF at the end of therapy. According to Vermote, this decrease in RF corresponded to a regression to more primitive levels of functioning within the treatment (in Jurist & Meehan, 2009). In general, these findings seem to support the hypothesis that psychodynamic psychotherapies promote psychological health by positively affecting patients' understanding and reasoning on their own mental states in relation to their symptoms (Katznelson, 2014).

Nevertheless, other controversial results raise the question of whether RF increases during psychotherapy and whether changes in RF and changes in symptoms are associated. For example, in Karlsson and Kermott's (2006) study, RF ratings were evaluated by repeated measure during the treatment process (through session transcripts), and no increase in RF was found in either cognitive–behavioral therapy (CBT) or brief psychodynamic psychotherapy (BPDT). RF even decreased in patients in interpersonal psychotherapy (IPT). This lack of improvement in RF might be attributed to the shortness of these treatments (e.g., Fonagy et al., 2002; Jones, 2000), but it is still surprising, considering that one of the suggested objectives of BPTD is to generate skills that facilitate mentalization (Bernbach, Muran, Slade, & Tuber, 2000). Similarly, Möller, Karlgren, Sandell, Falkenström, and Philips (2016) found that RF, as rated from video recordings of mentalization based treatment sessions, did not improve from the 2nd to the 6th month of therapy in a small sample of borderline patients with comorbid addiction. In contrast, another change was found: in the 6th month, patients' RF intrasession scores become more stable from statement to statement, compared to the stability shown in the 2nd month.

Vermote and colleagues (2010) assessed changes in RF in 44 hospitalized patients with personality disorder during a 12-month period of psychodynamic treatment. RF was assessed from patients' narratives of the Object Relations Inventory (ORI; Blatt, 1998) and results indicated that a decrease in symptomatology correlated with a general increase in felt personal safety and with an improvement in the complexity of the representations of oneself and others. However, these improvements were not associated with an increase in RF, pointing to a nonlinear relationship between RF and the outcome, as measured at the follow-up. The authors suggested that the discrepancy between these results and those of previous studies may be due to differences in the assessment of RF, which they generally measured on the AAI transcripts. As Lowyck and colleagues (2009) pointed out, RF based on AAI coding may reflect a more stable trait-like characteristic, whereas RF based on ORI transcripts or transcriptions of psychotherapy sessions could tap into more state-sensitive qualities. Incidentally, it is difficult to understand if the symptomatic improvements of these patients, and the increase in their perceived safety, were connected to the fact that they were hospitalized or to their having participated in a psychodynamic psychotherapy. Furthermore, these findings may be understood considering the complexities of the relationship between RF and therapeutic change. Vermote et al. (2010) argue that, in an attachment relationship—such as a therapeutic relationship—mentalizing ability might show some deterioration; this could account for the lack of RF increase found in this study.

While the above mentioned studies focused on RF as a change mediator and outcome measure, other studies have investigated the role of RF in moderating and/or predicting therapy outcome. Müller, Kaufhold, Overbeck, and Grabhorn (2006) assessed RF at the beginning of treatment from AAI transcripts. Patient symptomatology and general psychological functioning were assessed in the initial phase and at the end of treatment with the Symptom Checklist-90-R (SCL-90–R; Derogatis, 1977) and the Operationalized Psychodynamic Diagnosis interview (OPD Task Force, 2001). Results showed that patients with pretreatment higher levels of RF and a higher personality functioning according to the OPD tended to improve in terms of overall mental health after 3 months of a psychodynamically oriented treatment. Moreover, RF's power in predicting a positive therapeutic outcome was independent from others variables assessed with OPD. This finding support the idea that RF is an independent variable of the therapeutic process.

Taubner, Kessler, Buchheim, Kächele, and Staun (2011) investigated the role of mentalization in long-term psychoanalytic treatments of chronic depression. The RF, as measured on AAI transcripts, predicted the general symptom of distress (SCL-90;

Derogatis, 1977) but not a decrease in depressive symptoms (Beck Depression Inventory, BDI; Hautzinger, Bailer, Worall, & Keller, 1994). Moreover, patients with higher RF were found to be more capable of establishing a strong therapeutic alliance (assessed through the Helping Alliance Questionnaire, HAQ; Alexander & Luborsky, 1986).

Finally, Ekeblad, Falkenström, and Holmqvist (2016) found that RF predicted therapist-rated alliance (on the Working Alliance Inventory, WAI-S; Tracey & Kokotovic, 1989) across treatments (cognitive–behavioral therapy and interpersonal psychotherapy) but not patient-rated alliance. RF was also found to predict better outcome on self-reported depression (on the Beck Depression Inventory-II [BDI–II; Beck & Steer, 1996]). These findings show that RF may explain treatment outcome by moderating the construction of therapeutic alliance.

However, the results of the abovementioned studies are hardly generalizable because of differences in outcome measures, samples, and time frames. Moreover, in most of these studies, the Reflective Functioning Scale (RFS) was applied to AAI transcripts (AAI-based RF), whereas in some studies it was applied to psychotherapy session transcripts (session-based RF). To that extent, Hörz-Sagstetter, Mertens, Isphording, Buchheim, and Taubner (2015) investigated the relationship between AAI-based RF and session-based RF in two patients involved in long-term psychoanalytic psychotherapy. Although the results were purely exploratory, they showed that the RFS had good external validity when applied to psychoanalytic sessions. Moreover, Möller et al. (2016) found that pretreatment (AAI-based) RF ratings strongly correlated with the average (session-based) RF in each psychotherapy session, demonstrating that "a quite strong component of RF" seems to remain stable over time (p. 8).

To summarize, the results of empirical research on the relationship between RF and therapeutic outcomes are complex and inconclusive. Research focusing on the mediating value of RF (e.g., Levy et al., 2006; Rudden, Milrod, Target, Ackerman, & Graf, 2006; Vermote et al., 2010) has unveiled that improvements in RF may be related to good outcomes of psychotherapy, particularly in psychodynamic treatments. This finding has been especially observed when RF has been assessed on data capturing more trait-like and stable characteristics of personality (e.g., data collected from AAI transcripts) rather than data capturing more state-dependent psychological variables with a greater sensitivity to change in relation to patient's environment (e.g., data collected from psychotherapy session transcripts or ORI). Conversely, research focusing on the predictive value of RF (e.g., Müller et al., 2006; Taubner et al., 2011) suggests that "having an understanding of, and thinking in terms of, mental states and how they affect behavior does have an impact on how patients are able to make use of psychotherapy, especially in the beginning" (Katznelson, 2014, p. 114). Following the principle that the "rich get richer," it seems that patients with high pretreatment capacity for mentalization tend to improve more than patients with lower pretreatment levels of mentalization.

In spite of the abovementioned promising and clinically relevant studies, limitations in the assessment of mentalization using the RF scale endure. The RF Scale is primarily rated from AAI or similar costly and time-consuming interviews, requires specifically trained independent raters, and a range of different rating procedures are used when it is applied to psychotherapy sessions, making it more difficult to establish a standard method of assessment. To overcome such limitations, which may hinder research efforts, alternative and more cost-effective measures to assess RF were developed—such as shorter interviews (Rudden et al., 2006; Rutimann & Meehan, 2012) or questionnaires (Fonagy et al., 2016). Another alternative approach to measure reflective functioning, which could be particularly suitable for psychotherapy research, imply a computer assisted content text analysis. Although it was validated on AAI transcripts, the CRF (Fertuck et al., 2012) provides a simple and strongly standardized RF measure potentially applicable to psychotherapy session transcripts. It doesn't require any raters training, and so it can be used in more extensive research programs.

The aim of this study was to apply the CRF to psychoanalytic sessions in order to explore the relationship between RF ratings and psychotherapy outcome. In accordance with the literature, we followed the hypothesis that a good outcome of psychoanalytic treatment implies an increase in RF (as a mechanism of change in psychotherapy). In addition, another hypothesis tested was that RF evaluated at the beginning of the treatment could predict treatment outcome.

Instruments and Method

Measures

Computerized Reflective Functioning (CRF). The Computerized Text Analysis measure of Reflective Functioning (CRF) was developed by Fertuck and colleagues (2012). The CRF was built through the marker approach (Mergenthaler, 1996; Mergenthaler & Bucci, 1999), which is a procedure used to transform a consolidated and manualized coding system for data records into a computerized text scoring method. Both in order to establish the CRF scoring method and to check the criterion validity of the CRF, the authors used 113 AAI transcripts gathered from one nonclinical and one clinical sample. The nonclinical sample consisted of 40 adult residents in the United Kingdom, none of whom had ever received psychiatric or psychotherapeutic treatment, and none of whom met the criteria for Axis I and II disorders of the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision DSM-IV-TR;* American Psychiatric Association, 2000). The clinical sample consisted of 73 participants who had been diagnosed with BPD and were residents in the metropolitan area of New York. All participants met the criteria for BPD and had had in their lifetimes at least one of the following disorders: a mood disorder according to Axis I of the *DSM–IV* (77%), an anxiety disorder (48%), a behavioral disorder (44%), or a substance abuse disorder (39%). The AAIs, having been transcribed verbatim, were formatted according to the standards for computerized text analysis (Mergenthaler & Stinson, 1992) and evaluated using the RF scale. The AAIs were divided into two groups: AAIs with RF scores over 5 (from average to extraordinary RF) and AAIs with RF scores below 3 (from naïve or simplistic to absent or negative RF). Nine AAIs were selected from each group, resulting in nine interviews with higher RF scores and nine interviews with lower scores. The linguistic indicators of RF were derived from those groups of AAIs (extreme cases of RF score distribution) by applying a computer-assisted content analysis. This procedure permitted the sys-

tematic identification of content or stylistic features from language samples. The most diffused form of computer-assisted content analysis utilizes "dictionaries" of words that have been sorted into categories, or "tags." In this specific case, the most characteristic linguistic terms employed in these AAIs groups were identified (high and low RF respectively), and the CRF High and CRF Low dictionaries were built. The CRF High dictionary consists of just 54 linguistic markers that are highly frequent and relatively common words, such as "because," "think," "was," and "and." These verbal indicators are identified by an inductive method, and thus can at times seem counterintuitive in their association with mentalization construct. In fact, the marker approach methodology focuses on the *stylistic* aspect of speech regardless of themes or contents within speech. Word markers describe this stylistic aspect and are considered as operational indicators of a psychological state or capacity (Mergenthaler, 1996; Mergenthaler & Bucci, 1999).

Once the CRF High and CRF Low dictionaries were identified, frequencies of the linguistic markers included in the dictionaries were calculated for the remaining 95 interview texts. The CRF High dictionary significantly correlated with human rated RF scores in both the clinical (Spearman $\rho = .57$, $p < .0001$) and nonclinical samples (Spearman $\rho = .57$, $p = .002$), indicating a preliminary criterion validity of the CRF measure. When the CRF Low dictionary was added, the correlation between CRF and RF scoring did not significantly increase.

CRF represents a tool to assess RF in a less labor-intensive and more efficient manner, and it could also broaden the understanding of the verbal aspects of RF speech. Nevertheless, the CRF criterion validity must be considered cautiously due to the relatively small validation sample. Despite this also being true for RF human ratings, CRF measure (based on lexical items) may be confounded by individual differences in verbal IQ or verbal fluency.

Global Assessment of Functioning Scale. This measure is the fifth axis of the *DSM–IV-TR* (American Psychiatric Association, 2000). The GAF level of the patients involved in this study was assessed by two independent raters on the basis of transcriptions and audio-recordings of the first 8 sessions (the very first 4 sessions of the treatment + 4 sessions after one month of psychotherapy = beginning period) and the last 8 sessions (4 sessions from one month before the termination of the treatment and the very last 4 sessions = termination period). The two independent raters were two doctoral-level students in dynamic psychology. In a sample of 10 cases that had been assessed at the beginning and end of their analyses ($N = 20$)—cases that are part of this study—the intraclass correlation coefficient (ICC) of the assessments of the two independent raters was .65.

The Shedler-Westen Assessment Procedure-200 and the Personality Health Index. The SWAP-200, a widely researched personality assessment instrument (Shedler & Westen, 2007; Shedler, Westen, & Lingiardi, 2014; Westen & Shedler, 1999a, 1999b), is composed of 200 jargon-free items that describe both healthy and pathological personality features. It is completed by a clinician with knowledge of the patient. The SWAP-200 uses a Q-Sort method, which requires the rater to assign the 200 items so that a fixed distribution is attained (Block, 1978; Westen & Shedler, 1999a); this helps to improve rating reliability. The rater assigns a score from 0 to 7 for each item, according to how strongly the item describes that particular patient's functioning (from 0 [*not descriptive*] to 7 [*most descriptive*]). The SWAP-200 provides a common vocabulary that organizes clinical observations and inferences about a patient's personality and provides a "snapshot" of a patient's psychological functioning (Lingiardi, Gazzillo, & Waldron, 2010; Lingiardi, Shedler, & Gazzillo, 2006; Shedler, 2001). Moreover, 24 of the 200 items express aspects of positive mental health, and thus correct a focus on pathology. One of the outputs of the SWAP-200 software is a profile of 12 personality prototypes that closely parallel *DSM* personality diagnoses (the personality disorder [PD] scales). The correspondence between the patient profile and the PD scales is standardized as a *T* score (average = 50, standard deviation = 10), in comparison with a nationwide sample of personality disordered patients. When the *T* score for a given prototype is ≥ 60, the patient is considered to have that personality disorder; if the score is 55 to 59, the patient is considered to have significant traits of that disorder.

The Personality Health Index (PHI; Waldron et al., 2011) provides a global assessment of personality health using the SWAP-200 item scores. The index is based on pooled clinical judgments of the health significance of each SWAP-200 item, multiplied by the patient's actual score (0 to 7), as determined by the rating clinician. This total raw score is then automatically converted into a percentile score that references a national sample of 307 SWAP-200s from different points in the treatment of 70 psychoanalytic patients (Cogan & Porcerelli, 2004, 2005). The percentile standing provides a readily understandable metric of overall psychological health/illness, which was shown to have good validity (Waldron et al., 2011).

The SWAP/PHI level of the patients involved in this study was assessed by two independent raters on the basis of transcriptions and audio-recordings of the first 8 sessions (the very first 4 sessions of the treatment + 4 sessions after one month of psychotherapy = beginning period) and the last 8 sessions (4 sessions from one month before the termination of the treatment and the very last 4 sessions = termination period). The two independent raters were a researcher in dynamic psychology and psychoanalytic psychotherapist with more than 5 years of clinical experience after licensing and a doctoral-level student in dynamic psychology. In this study, the ICC of the SWAP-200 assessment, based on the assessment made by the two independent raters of 20 groups of eight contiguous sessions of 10 patients involved in this study, was .79.

Sample

Our sample consisted of 540 sessions from 27 psychoanalytic treatments delivered in the U.S., in both academic and private settings, by eight U.S. male experienced analysts (who had been in practice for more than 5 years after having completed their training in U.S. analytic institutions affiliated with the International Psychoanalytic Association). These analyses were conducted from the end of the 1960s to 2011, were completely audiotaped, and were on loan from the Psychoanalytic Research Consortium in New York (www.psychoanalyticresearch.org). The average frequency of sessions per week was 3.4 (ranging from 1 session every other week at the end of a long analysis to 5 sessions per week). The average duration of treatment was 366 sessions, ranging from 120 to 2,836 sessions. In several analyses, the frequency of sessions changed over the course of the treatment.

We selected and analyzed 20 sessions from each treatment: the first four consecutive sessions, four sessions from the 4th to 6th week (representative of the first month/phase of treatment), four consecutive sessions from the middle of treatment (the middle phase), four consecutive sessions from the 4th to 6th week before termination of treatment, and the last four consecutive sessions (with these last eight sessions representing the treatment's termination phase).

Of these 540 sessions, 383 (70%) were held on the couch, 138 (26%) were delivered face-to-face, and 19 (4%) were conducted via telephone or Skype; all of the sessions conducted via telephone or Skype were delivered in the final years of long analyses that had previously been conducted in person. In several cases, the patient passed from the couch to the chair, or vice versa.

Thirteen patients were women and 14 were men. The average age of patients at the beginning of their analysis was 33 years, ranging from 20 to 70. At the outset, 11 patients received a clinical disorder diagnosis, with depressive disorder and sexual disorders occurring most frequently (three patients were diagnosed as depressed and two were diagnosed with a sexual disorder). At the beginning of their treatment, 12 were diagnosed with a personality disorder, according to their SWAP-200 assessment; the most frequent personality disorders were schizoid personality disorder (5), avoidant personality disorder (5) and borderline personality disorder (4).[1] Patients' average GAF at the beginning of their treatment was 64 (ranging from 52 to 84, $SD = 7.1$). Their average level of PHI was 38 (ranging from 4 to 86; $SD = 25.9$). Given that PHI is a percentile score of personality health based on a U.S. sample of patients in analysis, we can say that the patients in our sample were, on average, more severely afflicted than 62% of patients in analysis.

In order to define a treatment outcome as "good," we established a priori three criteria: (a) at the end of treatment, the patient must not have a clinical or personality disorder; (b) at the end of treatment, the patient's GAF must be ≥ 65; and (c) at the end of the treatment, the patient's PHI must be 10 or more points higher than their beginning score. In order to establish whether the outcome of the treatment is "good," all of these three criteria must be fulfilled. The rationale for these three criteria was that we needed to take into account the possibility that a patient could have had neither a clinical nor a personality disorder at the beginning of his or her treatment, so the absence of clinical and personality diagnoses at the end of treatment could not be considered the sole criterion. We also needed to take into account the possibility that a patient's GAF could be quite high at the beginning of the analysis; thus, we could not use a high level of GAF as the only criterion of good treatment outcome. Finally, we assumed that a good outcome should always increase a patient's personality health (PHI), but we could not expect too large an increase of PHI because of the possibility that some patients may have had a high PHI since the beginning of their treatment, creating a ceiling effect. These three criteria, taken together, appeared to provide a plausible set of criteria for classifying "good" versus "poor" outcome treatments.

Overall, there was an average increase of 15.4 PHI points (ranging from -11 to $+75$; $t(26) = 5.25$; $p < .001$, $d = 2.06$) and an average increase of 7 GAF points (ranging from -2 to $+28$; $t(26) = 4.97$; $p < .000$, $d = 1.95$) in the patients. The correlation between patients' change in PHI and change in GAF during the treatments was $r = .68$ ($p < .001$).

According to the criteria previously described, we identified 17 good outcome treatments and 10 poor outcome treatments. There were no statistically significant differences between good and poor outcome cases in terms of either the age of patient at the beginning of treatment (30.8 vs. 30.4 years; $t[25] = .25$; $p = .80$) or the duration of treatment (372.4 vs. 337.6 sessions; $t[1] = .74$; $p = .46$). Finally, at the beginning of treatment there was no difference between patients with good and poor outcome treatments in terms of gender (four females had a poor outcome and nine had a good outcome; six males had a poor outcome and eight had a good outcome; χ^2 [1] = .42; $p = .52$), personality disorders and personality functioning ([SWAP-200 PD scales], see Table 1), and presence/absence of clinical disorders (nine good outcome patients had at least one clinical disorder diagnosis, nine did not; six poor outcome patients had a clinical disorder, three did not [χ^2 (1) = .3.46; $p = .06$]). However, there was a difference in terms of the average frequency of sessions per week (3.1 vs. 3.8; $t[1] = -2.10$; $p = .04$), with poor outcome patients having more sessions per week than good outcome patients. There was no difference in PHI at the beginning between patients with good and poor outcome treatments: The average PHI of patients with good outcome treatments, at the beginning of their analyses, was 38.7, compared to 36.5 for patients with poor outcome treatments ($t[26] = 0.21$; $p = .85$). There were also no differences in GAF scores, which were 66.1 and 63.2, respectively ($t[26] = .90$; $p = .38$).

The average PHI at the end of the analysis, however, was 76.2 ($SD = 19.8$) in good outcome cases and 39.2 ($SD = 30.6$) in poor outcome cases, $t[26] = 3.8$, $p = .001$. The average GAF score at the end of the treatment was 75.7 ($SD = 6.8$) for good outcome cases and 65.5 ($SD = 4.8$) for poor outcome cases ($t[25] = 4.5$; $p < .001$). Therefore, in good outcome cases there was an average increase of 37.5 points in PHI and 9.6 points in GAF; in poor outcome treatments, there was an average increase of 2.7 points in PHI and 1.4 points in GAF.

Procedures

All of the 540 sessions, which were audio recorded and transcribed verbatim, were arranged for computer text analysis. The computer program CRF Yoshikoder calculated four indices (raw data): "CRF," frequency of words included in the vocabularies High and Low; "CRF High," frequency markers of High; "CRF Low," frequency markers of Low; "TOT," total number of words in the transcript. We then calculated two indices that allowed us to compare the various sessions: CRF High/TOT (H%: frequency of the markers High weighted on the total number of words in the transcript) and CRF Low/TOT (L%: Low frequency markers weighted on the total number of words in the transcript). Since Fertuck et al.'s (2012) study showed that the CRF Low vocabulary had a small correlation with the RF scores, we chose to use only the CRF High index, which showed stronger criterion validity. Two independent raters assessed the first and last eight sessions of each treatment with the SWAP-200, and two independent raters assessed the same sessions also with the GAF.

In order to test our hypotheses, we performed a series of generalized estimating equations and partial correlations. All of the analyses were conducted with SPSS.20.

[1] The number of personality disorders is greater than the number of personality disorder diagnoses (11) due to comorbidity.

Table 1
Comparison Between the Average Scores of PD Scales in Good and Poor Outcome Patients at the Beginning of Their Analyses

PD scales	Good outcome	Poor outcome	t (df = 25)	p
Paranoid	40.82	44.27	−1.116	.82
Schizoid	49.56	48.50	.352	.47
Schizotypical	45.13	45.18	−.018	.72
Antisocial	42.63	45.84	−1.511	.76
Borderline	44.62	45.64	−.251	.38
Histrionic	45.62	47.71	−.487	.99
Narcissistic	43.06	46.92	−1.284	.76
Avoidant	52.27	49.43	1.019	.74
Dependent	53.19	50.68	.994	.99
Obsessive	52.13	51.24	.295	.70
High-Functioning	62.79	59.99	.840	.18

Results

In order to verify the hypothesis of an increase in the CRF High index in patients with good outcome treatments, we performed a series of generalized estimating equations (GEE) with an exchangeable matrix of correlations. In fact, given that our sessions were "nested" within the treatments (each treatment comprised 20 sessions), general linear models, based on the assumption of independence among observations, would not have been appropriate. In this model of GEE, we used patients' CRF High index as a dependent variable and treatment phase as a factor. We did not control the therapist effect, that is, the fact that several patients were treated by the same analyst, because we have only eight analysts, an N which is too low for having reliable results with hierarchical models of data analysis (Maas & Hox, 2005).

The results showed that there was no difference, and thus no increase, in good outcome therapy patients' level of reflective functioning ($\chi^2 = 2.34$; $df = 2$, $p = .311$) from the beginning to the end of treatment. The same was also true for patients with poor outcome treatments ($\chi^2 = 2.15$; $df = 2$; $p = .341$; see Table 2 for the CRF values of the different treatments in the different phases). On the basis of these data, the hypothesis that the goodness of the outcome of a psychoanalytic treatment goes hand in hand with an increase in RF was not confirmed.

Our second aim was to verify the extent to which the CRF assessed at the beginning of the treatment predicted the levels of PHI and GAF reached at the end of therapy and the increase in PHI and GAF levels during the treatment. In order to test these hypotheses, we performed a GEE, using patients' PHI at the end of treatment as a dependent variable and patients' PHI in the early period and the average level of patients' CRF from the early period as covariates. In this way, we were able to control the effect of the early CRF on the late PHI by excluding the influence of the early PHI. The same procedure was followed to verify the predictive value of CRF on patients' GAF level at the end of treatment.

GEE showed that patients' CRF in the early period predicted their PHI at the end of treatment ($\chi^2 = 11.90$; $df = 1$; B = 1.49; $p = .001$). The partial correlation between early CRF and late PHI, controlling for the influence of early PHI, length of treatment, and the average number of sessions per week, was .47. Moreover, the partial correlation between early CRF and the change in patients' PHI (PHI late−PHI early) was $r = .55$, $p = .005$. GEE also showed that CRF in the early period predicted patients' GAF at the end of treatment ($\chi^2 = 15.77$; $df = 1$; B = .277; $p = .001$). The partial correlation between early CRF and late GAF, controlling for the influence of early GAF, length of treatment, and the average number of sessions per week, was .45. Moreover, the partial correlation between CRF early and the change in patients' GAF (GAF late−GAF early) was also $r = .51$, $p = .011$.

Discussion

This study formed the first investigation into the relationship between RF and psychotherapy outcome in long-term *psychoanalytic* treatments, using multiple and complex outcome measures and evaluating RF through a computerized text analysis of psychotherapy session transcripts. The most relevant finding is that patients' RF, assessed in the first month of long-term treatment, was able to predict changes in two global measure of personality functioning (PHI and GAF). These results are consistent with those

Table 2
Means (and Standard Deviations) of CRF High Scores in Different Phases of Treatment

	Early phase[a]	Middle phase[b]	Late phase[c]
Good outcome treatments (N = 17)	30.58 (6.15)	29.26 (7.35)	28.64 (7.91)
Poor outcome treatments (N = 10)	26.56 (9.41)	26.25 (9.38)	24.88 (10.22)
All treatments (N = 27)	29.09 (7.75)	28.14 (8.25)	27.25 (9)

Note. CRF = Computerized Text Analysis measure of Reflective Functioning.
[a] First four consecutive sessions of treatment + four sessions from the 4th to 6th week of treatment. [b] Four consecutive sessions from the middle of treatment. [c] Four consecutive sessions from the 4th to 6th week before treatment termination + last four consecutive sessions of treatment.

of previous studies investigating the predictive value of RF on therapy outcome (Ekeblad et al., 2016; Müller et al., 2006; Taubner et al., 2011). The findings seem to support the theoretical hypothesis that patients with better mentalizing capacity are more disposed to engage in a therapeutic relationship, establish a therapeutic alliance (as shown by both Taubner et al., 2011 and Ekeblad et al., 2016) and gain benefit from therapy (as shown by Gullestad, Johansen, Høglend, Karterud, & Wilberg, 2013). Moreover, a predictive value of constructs theoretically linked to mentalization—namely psychological mindedness (Conte et al., 1990; McCallum, Piper, Ogrodniczuk, & Joyce, 2003), insight (Høglend, Engelstad, Sørbye, Heyerdahl, & Amlo, 1994) and alexithymia (Leweke, Bausch, Leichsenring, Walter, & Stingl, 2009)—has already been found.

As Blatt and Auerbach (2003) discussed in relation to the nature of therapeutic change, "all psychoanalytic orientations (e.g., Freudian, ego-psychological, self-psychological, object-relational, interpersonal, etc.) agree that sustained symptom remission, while essential to any successful treatment outcome, is secondary to and dependent on more basic changes in the personality structure" (pp. 268–269). To that extent, high pretreatment RF may be necessary for changing the deeper and more structural aspects of a patient's personality—such as internalizing object relations into intrapsychic representational structures characterized by a clearer sense of boundaries and separateness, as well as a greater degree of empathic relatedness (Diamond, Kaslow, Coonerty, & Blatt, 1990).

In this study it was also shown that, when RF was investigated as an outcome/mediator variable that might increase in good outcome treatments, no significant results were obtained. Patients' RF ratings seem to have remained relatively stable during the therapeutic process, even in good outcome treatments. These results are similar to those found by Karlsson and Kermott (2006) and Vermote et al. (2010), who pointed to the nonlinear relationship between RF and treatment outcome. In order to understand these findings, an important issue could be the method used to evaluate RF. For instance, Levy et al. (2006) found an increase in mentalization associated with clinical improvement during TFP when RF was evaluated on AAI transcripts. As some authors have highlighted (Hörz-Sagstetter et al., 2015; Lowyck et al., 2009), RF assessed on AAI transcripts may reflect a more stable, trait-like characteristic, whereas RF based on psychotherapy sessions could tap into a "state-like" manifestation that fluctuates during the analytic process and within sessions, and in relation to the therapist's interventions. In this study, RF was assessed from psychotherapy sessions in which the patient was involved in an intense emotional context and was in an attachment relationship; these conditions could interfere with the patient's mentalization processes. However, it cannot be excluded that, in their daily life as in reflecting on their past experiences, patients were more capable of expressing their mentalization capacity; therefore, patient RF might increase during good outcome treatments even after the conclusion of the therapy. These data are consistent with Möller et al.'s (2016) study, in which the RF rated on video recordings of psychotherapy sessions was lower than the RF rated on interviews similar to the AAI (the Brief Reflective Functioning Interview, BRFI; Rudden, Milrod, Aronson, & Target, 2008). Consequently, the authors pointed out that "it seems like patients mentalized below their capacity much of the time during the psychotherapy sessions" (p. 13). Among their different explanations for these data, the authors underlined the importance of the different relationship that the patients developed with their therapist (an attachment relationship), relative to the relationship they developed with the pretreatment interviewer (a nonattachment relationship). More research is needed to understand AAI transcripts and session transcripts as different sources of information regarding RF. Another explanation of these results could rely on the different kinds of treatments evaluated in the various studies. For example, it is plausible to argue that TFP explicitly aims at promoting RF in the context of the affect-laden attachment relationship with the therapist. In fact, the emotional states of the patient are modulated by the therapist through clarifications, confrontations, and interpretations in the context of the systematic analysis of the therapeutic relationship in the "here and now." Additionally, when treatments that specifically involve borderline (BPD) patients are considered, a "floor effect" may be present. It may be possible that, starting from lower RF scores, BPD patients show a major change in mentalization in comparison with the average psychoanalytic outpatients, who instead may start their therapy with a better clinical condition, and consequently may show less significant changes in their reflective capacities during psychotherapy. Moreover, it must be better clarified if BPD patients are really characterized by a general global impairment or if deficits in RF have a minor clinical specificity and characterize a broader range of personality disorders (Nazzaro et al., 2017).

Taken together, the results of this study seem to suggest that RF is predictive of patient's functioning changes although it does not measure these changes. To sum up, it is possible that, in a psychoanalytic treatment, RF could be a necessary condition for the patient to better make use of communications and relationship with the therapist.

It is worth noting that this study has some limitations. First, this was the first attempt to use CRF in psychotherapy research. Although the CRF has shown good criterion validity (Fertuck et al., 2012), it has only been validated on the AAI transcript, while in this study it was used on psychotherapy sessions. Therefore, our results should be considered explorative and replication studies are necessary. Although marker approach instruments are very useful and allow for extensive research projects, they suffer from some limitations: they provide quantitative rather than qualitative measurements of underlying constructs and they are unable to consider contextual characteristics. Another limitation regards the impossibility to control one of the potentially confounder variables in the data analysis. In fact, although several patients were treated by the same analyst, the number of analysts ($N = 8$) is too low to have reliable results when using hierarchical models of data analysis (Maas & Hox, 2005). Moreover, all the clinical data have been derived from the audio recordings of the analytic sessions, and there were not any independent clinicians or raters who independently assessed the patients in another context (i.e., in a separate interview or assessment procedure). Finally, the generalizability of this study could be limited by the specific historical period when these psychoanalytic treatments were conducted. Considering that 20 years ago, before the so-called "relational turn" (Greenberg & Mitchell, 1983; Lingiardi, Holmqvist, & Safran, 2016), some of our psychoanalytic treatments were technically and theoretically different from their current form, some bias may be present in this study. Driven by the huge amount of literature on mentalization theory and research, it is possible that contemporary psychoanalysts tend to use, explicitly or implicitly, a specific set of thera-

peutic strategies, such as assuming a not-knowing stance, stimulating mentalization in psychotherapy by "demand questioning," or keeping the focus on affects "here and now" and in relation to an interpersonal event. To that extent, it is possible that these newer techniques promote mentalization more consistently than the "traditional" ones (Bateman & Fonagy, 2016; Ekeblad et al., 2016).

References

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

Allen, J. G. (2003). Mentalizing. *Bulletin of the Menninger Clinic, 67,* 91–112. http://dx.doi.org/10.1521/bumc.67.2.91.23440

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, 4th edition, text revision (DSM-IV-TR)*. Washington, DC: Author.

Antonsen, B. T., Johansen, M. S., Rø, F. G., Kvarstein, E. H., & Wilberg, T. (2016). Is reflective functioning associated with clinical symptoms and long-term course in patients with personality disorders? *Comprehensive Psychiatry, 64,* 46–58.

Bateman, A. W., & Fonagy, P. (2004a). *Mentalization based treatment for personality disorder: A practical guide*. New York, NY: Oxford University Press.

Bateman, A. W., & Fonagy, P. (2004b). Mentalization-based treatment of BPD. *Journal of Personality Dispersore, 18,* 36–51.

Bateman, A. W., & Fonagy, P. (2016). *Psychotherapy for borderline personality disorder: Mentalization based treatment*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/med:psych/9780199680375.001.0001

Beck, A. T., & Steer, R. A. (1996). *Beck Depression Inventory: Manual (Swedish version*). Stockholm, Sweden: Psykologiförlaget.

Bernbach, E. J., Muran, J. C., Slade, A., & Tuber, S. (2000, June). *Reflective-functioning and the therapeutic relationship: Understanding change in a brief relational psychotherapy*. Paper presented at the annual meeting of the Society for Psychotherapy Research, Chicago, IL.

Blatt, S. J. (1998). Contributions of psychoanalysis to the understanding and treatment of depression. *Journal of the American Psychoanalytic Association, 46,* 722–752.

Blatt, S. J., & Auerbach, J. S. (2003). Psychodynamic measures of therapeutic change. *Psychoanalytic Inquiry: A Topical Journal for Mental Health Professionals, 23,* 268–307.

Blatt, S. J., & Ford, R. Q. (1994). *Therapeutic change: An object relations perspective*. New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4899-1010-3

Block, J. (1978). *The Q–Sort method in personality assessment and psychiatric research*. Palo Alto, CA: Consulting Psychologists Press.

Clarkin, J. F., Caligor, E., Stern, B., & Kernberg, O. F. (2004). *The structured interview of personality organization*. Unpublished manuscript, Weill Medical College of Cornell University, New York, NY.

Cogan, R., & Porcerelli, J. H. (2004). Personality pathology, adaptive functioning, and strengths at the beginning and end of psychoanalysis. *Journal of the American Psychoanalytic Association, 52,* 1230–1231.

Cogan, R., & Porcerelli, J. H. (2005). Clinician reports of personality pathology of patients beginning and patients ending psychoanalysis. *Psychology and Psychotherapy: Theory, Research and Practice, 78,* 235–248. http://dx.doi.org/10.1348/147608305X28727

Conte, H. R., Plutchik, R., Jung, B. B., Picard, S., Karasu, T. B., & Lotterman, A. (1990). Psychological mindedness as a predictor of psychotherapy outcome: A preliminary report. *Comprehensive Psychiatry, 31,* 426–431. http://dx.doi.org/10.1016/0010-440X(90)90027-P

Derogatis, L. R. (1977). *SCL-90-R, administration, scoring and procedures manual—I for the revised version*. Baltimore, MD: Johns Hopkins School of Medicine.

Diamond, D., Kaslow, N., Coonerty, S., & Blatt, S. J. (1990). Changes in separation-individuation and intersubjectivity in long-term treatment. *Psychoanalytic Psychology, 7,* 363–397.

Ekeblad, A., Falkenström, F., & Holmqvist, R. (2016). Reflective functioning as predictor of working alliance and outcome in the treatment of depression. *Journal of Consulting and Clinical Psychology, 84,* 67–78. http://dx.doi.org/10.1037/ccp0000055

Fertuck, E. A., Mergenthaler, E., Target, M., Levy, K. N., & Clarkin, J. F. (2012). Development and criterion validity of a computerized text analysis measure of reflective functioning. *Psychotherapy Research, 22,* 298–305. http://dx.doi.org/10.1080/10503307.2011.650654

Fischer-Kern, M., Doering, S., Taubner, S., Hörz, S., Zimmermann, J., Rentrop, M., . . . Buchheim, A. (2015). Transference-focused psychotherapy for borderline personality disorder: Change in reflective function. *British Journal of Psychiatry, 207,* 173–174.

Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2002). *Affect regulation, mentalization and the development of the self*. New York, NY: Other Press.

Fonagy, P., Luyten, P., & Bateman, A. (2015). Translation: Mentalizing as treatment target in borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment, 6,* 380–392. http://dx.doi.org/10.1037/per0000113

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y., Warren, F., Howard, S., . . . Lowyck, B. (2016). Development and Validation of a Self-Report Measure of Mentalizing: The Reflective Functioning Questionnaire. *PLOS ONE, 11,* e0158678. http://dx.doi.org/10.1371/journal.pone.0158678

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-functioning manual: Version 5 for application to adult attachment interviews*. Unpublished manual. London, England: University College.

George, C., Kaplan, N., & Main, M. (1996). *Adult attachment interview* (3rd ed.). Unpublished manuscript, Department of Psychology, University of California at Berkeley

Greenberg, J., & Mitchell, S. (1983). *Object relations in psychoanalytic theory*. Cambridge, MA: Harvard University Press.

Gullestad, F. S., Johansen, M. S., Høglend, P., Karterud, S., & Wilberg, T. (2013). Mentalization as a moderator of treatment effects: Findings from a randomized clinical trial for personality disorders. *Psychotherapy Research, 23,* 674–689. http://dx.doi.org/10.1080/10503307.2012.684103

Hautzinger, M., Bailer, M., Worall, H., & Keller, F. (1994). *Beck Depression Inventory (BDI)*. Bern, Switzerland: Huber.

Høglend, P., Engelstad, V., Sørbye, O., Heyerdahl, O., & Amlo, S. (1994). The role of insight in exploratory psychodynamic psychotherapy. *British Journal of Medical Psychology, 67,* 305–316. http://dx.doi.org/10.1111/j.2044-8341.1994.tb01799.x

Holmes, J. A. (2001). *The search for the secure base: Attachment theory and psychotherapy*. London, England: Routledge.

Hörz-Sagstetter, S., Mertens, W., Isphording, S., Buchheim, A., & Taubner, S. (2015). Change in reflective functioning during psychoanalytic psychotherapies. *Journal of the American Psychoanalytic Association, 63,* 481–509. http://dx.doi.org/10.1177/0003065115591977

Jones, E. E. (2000). *Therapeutic action. A guide to psychoanalytic therapy*. Northvale, NJ: Aronsson.

Jurist, E., & Meehan, K. (2009). Attachment, mentalization and reflective functioning. In J. Obegi & E. Berant (Eds.), *Attachment theory and research in clinical work with adults* (pp. 71–93). New York, NY: Guilford Press.

Jurist, E. L., Slade, A., & Bergner, S. (2008). *Mind to mind: Infant research, neuroscience, and psychoanalysis*. New York, NY: Other Press.

Karlsson, R., & Kermott, A. (2006). Reflective-functioning during the process in brief psychotherapies. *Psychotherapy: Theory, Research, Practice, Training, 43,* 65–84. http://dx.doi.org/10.1037/0033-3204.43.1.65

Katznelson, H. (2014). Reflective functioning: A review. *Clinical Psychology Review, 34,* 107–117. http://dx.doi.org/10.1016/j.cpr.2013.12.003

Levy, K. N., Meehan, K. B., Kelly, K. M., Reynoso, J. S., Weber, M., Clarkin, J. F., & Kernberg, O. F. (2006). Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. *Journal of Consulting and Clinical Psychology, 74,* 1027–1040. http://dx.doi.org/10.1037/0022-006X.74.6.1027

Leweke, F., Bausch, S., Leichsenring, F., Walter, B., & Stingl, M. (2009). Alexithymia as a predictor of outcome of psychodynamically oriented inpatient treatment. *Psychotherapy Research, 19,* 323–331. http://dx.doi.org/10.1080/10503300902870554

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

Lingiardi, V., Holmqvist, R., & Safran, J. (2016). Relational turn and psychotherapy research. *Contemporary Psychoanalysis, 52,* 275–312. http://dx.doi.org/10.1080/00107530.2015.1137177

Lingiardi, V., Shedler, J., & Gazzillo, F. (2006). Assessing personality change in psychotherapy with the SWAP-200: A case study. *Journal of Personality Assessment, 86,* 23–32. http://dx.doi.org/10.1207/s15327752jpa8601_04

Lingiardi, V., Tanzilli, A., & Colli, A. (2015). Does the severity of psychopathological symptoms mediate the relationship between patient personality and therapist response? *Psychotherapy, 52,* 228–237. http://dx.doi.org/10.1037/a0037919

Lowyck, B., Vermote, R., Luyten, P., Franssen, M., Verhaest, Y., Vertommen, H., . . .Peuskens, J. (2009). *Comparisons of reflective functioning as measured on the Adult Attachment Interview (AAI) and the Object Relations Inventory (ORI) in patients with a personality disorder: A preliminary study*. Poster presented at the winter meeting of the American Psychoanalytic Association, New York, NY.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1,* 85–91. http://dx.doi.org/10.1027/1614-2241.1.3.85

McCallum, M., Piper, W. E., Ogrodniczuk, J. S., & Joyce, A. S. (2003). Relationships among psychological mindedness, alexithymia and outcome in four forms of short-term psychotherapy. *Psychology and Psychotherapy: Theory, Research and Practice, 76,* 133–144. http://dx.doi.org/10.1348/147608303765951177

Mergenthaler, E. (1996). Emotion-abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology, 64,* 1306–1315.

Mergenthaler, E., & Bucci, W. (1999). Linking verbal and non-verbal representations: Computer analysis of referential activity. *British Journal of Medical Psychology, 72,* 339–354. http://dx.doi.org/10.1348/000711299160040

Mergenthaler, E., & Stinson, C. (1992). Psychotherapy transcription standards. *Psychotherapy Research, 2,* 125–142. http://dx.doi.org/10.1080/10503309212331332904

Möller, C., Karlgren, L., Sandell, A., Falkenström, F., & Philips, B. (2016). Mentalization-based therapy adherence and competence stimulates in-session mentalization in psychotherapy for borderline personality disorder with co-morbid substance dependence. *Psychotherapy Research*. Advance online publication. http://dx.doi.org/10.1080/10503307.2016.1158433

Müller, C., Kaufhold, J., Overbeck, G., & Grabhorn, R. (2006). The importance of reflective functioning to the diagnosis of psychic structure. *Psychology and Psychotherapy: Theory, Research and Practice, 79,* 485–494. http://dx.doi.org/10.1348/147608305X68048

Nazzaro, M. P., Boldrini, T., Tanzilli, A., Muzi, L., Giovanardi, G., & Lingiardi, V. (2017). Does reflective functioning mediate the relationship between attachment and personality? *Psychiatry Research, 256,* 169–175. http://dx.doi.org/10.1016/j.psychres.2017.0

OPD Task Force. (2001). *Operationalized psychodynamic diagnostics: Foundations and manual*. Seattle, WA: Hogrefe and Huber.

Rudden, M. G., Milrod, B., Aronson, A., & Target, M. (2008). Reflective functioning in panic disorder patients: Clinical observations and research design. In F. N. Busch (Ed.), *Mentalization: Theoretical considerations, research findings, and clinical implications* (pp. 185–205). New York, NY: Analytic Press.

Rudden, M., Milrod, B., Target, M., Ackerman, S., & Graf, E. (2006). Reflective functioning in panic disorder patients: A pilot study. *Journal of the American Psychoanalytic Association, 54,* 1339–1343. http://dx.doi.org/10.1177/00030651060540040109

Rutimann, D. D., & Meehan, K. B. (2012). Validity of a brief interview for assessing reflective function. *Journal of the American Psychoanalytic Association, 60,* 577–589. http://dx.doi.org/10.1177/0003065112445616

Shedler, J. (2001). A new language for psychoanalytic diagnosis. *Journal of the American Psychoanalytic Association, 50,* 429–456. http://dx.doi.org/10.1177/00030651020500022201

Shedler, J., & Westen, D. (2007). The Shedler-Westen Assessment Procedure (SWAP): Making personality diagnosis clinically meaningful. *Journal of Personality Assessment, 89,* 41–55. http://dx.doi.org/10.1080/00223890701357092

Shedler, J., Westen, D., & Lingiardi, V. (2014). *La valutazione della personalità con la SWAP-200* [The evaluation of personality with the SWAP-200]. Milan, Italy: Raffaello Cortina.

Tanzilli, A., Colli, A., Gualco, I., & Lingiardi, V. (2017). Patient personality and relational patterns in psychotherapy: Factor structure, reliability, and validity of the Psychotherapy Relationship Questionnaire. *Journal of Personality Assessment*. Advance online publication. http://dx.doi.org/10.1080/00223891.2016.1272050

Taubner, S., Kessler, H., Buchheim, A., Kächele, H., & Staun, L. (2011). The role of mentalization in the psychoanalytic treatment of chronic depression. *Psychiatry: Interpersonal and Biological Processes, 74,* 49–57. http://dx.doi.org/10.1521/psyc.2011.74.1.49

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 207–210. http://dx.doi.org/10.1037/1040-3590.1.3.207

Vermote, R. (2005). *Touching inner change: Psychoanalytically informed hospitalization-based treatment of personality disorders: A process-outcome study* (Doctoral dissertation). Leuven, Belgium: Catholic University Press.

Vermote, R., Lowyck, B., Luyten, P., Vertommen, H., Corveleyn, J., Verhaest, Y., . . . Peuskens, J. (2010). Process and outcome in psychodynamic hospitalization-based treatment for patients with a personality disorder. *Journal of Nervous and Mental Disease, 198,* 110–115. http://dx.doi.org/10.1097/NMD.0b013e3181cc0d59

Waldron, S., Moscovitz, S., Lundin, J., Helm, F., Jemerin, J., & Gorman, B. (2011). Evaluating the outcomes of psychotherapies: The Personality Health Index. *Psychoanalytic Psychology, 28,* 363–388. http://dx.doi.org/10.1037/a0024559

Westen, D., & Shedler, J. (1999a). Revising and assessing axis II, Part I: Developing a clinically and empirically valid assessment method. *The American Journal of Psychiatry, 156,* 258–272.

Westen, D., & Shedler, J. (1999b). Revising and assessing axis II, Part II: Toward an empirically based and clinically useful classification of personality disorders. *The American Journal of Psychiatry, 156,* 273–285.