2-Year Follow-Up and Changes in Reflective Functioning in Specialist and Nonspecialist Treatment Models for Personality Disorder

Marco Chiesa MD, FRCPsych (corresponding author) Research Department of Clinical, Educational & Health Psychology, University College London, London, United Kingdom m.chiesa@ucl.ac.uk

Patrick Luyten PhD, Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium and Research Department of Clinical, Educational & Health Psychology, University College London, London, United Kingdom patrick.luyten@kuleuven.be

Peter Fonagy PhD FBA, Head of the Division of Psychology and Language Sciences, University College London, London, United Kingdom and Anna Freud Centre, London, United Kingdom p.fonagy@ucl.ac.uk

Abstract

There is a growing body of studies linking impairments in mentalizing or reflective functioning (RF) with childhood adversity, the development of personality disorder (PD), and psychiatric morbidity. Fewer studies have investigated the purported role of changes in RF in relation to clinical outcome in treatments focusing on this capacity. Moreover, it is as yet unclear whether specialist and nonspecialist treatment models are equally effective in bringing about change in RF in conjunction with symptomatic improvement. This study aimed to investigate the association between changes in RF in three samples of individuals with PD treated in two specialist psychosocial programs (a step-down model; RT-CBP, and a long-term residential model; RT) and in an outpatient general psychiatric service (GP) over a 2-year period after intake into treatment. RF was assessed using the Reflective Functioning Scale scored on the Adult Attachment Interview and clinical outcome was assessed in terms of psychiatric distress, social adjustment, and global functioning. Changes in RF were most marked in RT-CBP compared with RT and GP. Changes in RF explained differences between treatment models for social and global adjustment outcomes, but not for disparities in psychiatric distress. A medium-intensity treatment approach to PD such as RT-CBP was more effective in improving RF and provided a balance between psychotherapy input and efforts at social integration, by bringing patients into closer contact with their social world.

Keywords: personality disorder, reflective functioning, psychotherapy, psychotherapy research, mechanisms of change

2-Year Follow-Up and Changes in Reflective Functioning in Specialist and Nonspecialist Treatment Models for Personality Disorder

Mentalizing or reflective functioning (RF) refers to the capacity to think coherently and consistently about self and others within an emotionally meaningful interpersonal context. A number of studies have shown that impairments in this capacity may play a role in the development and phenomenology of personality disorder (Antonsen, Johansen, Ro, Kvarstein, & Wilberg, 2016; Bateman & Fonagy, 2010; Fonagy, Campbell, & Bateman, 2016; Nazzaro et al., 2017). Systematic reviews of studies examining mentalizing have found consistent impairments in mental-state reasoning in individuals with PD (Nemeth et al., 2018; Richman & Unoka, 2015). Moreover, mentalizing has been suggested to be a psychological mechanism linking early adversity to later psychopathology, as low RF has been shown to link childhood adversity, PD, and other types of psychopathology (Chiesa & Fonagy, 2014; Cirasola, Hillman, Fonagy, & Chiesa, 2017; Fonagy et al., 2016).

Several studies have shown that day-hospital and outpatient specialist approaches based on mentalization-based treatment (MBT; Bateman & Fonagy, 2009; Gullestad et al., 2012; Laurenssen et al., 2018), transference-focused psychotherapy (Clarkin, Levy, Lenzenweger, & Kernberg, 2007; Doering et al., 2010), a modified form of cognitive-behavioral therapy (DeCou, Comtois, & Landes, 2019; McMain, Guimond, Streiner, Cardish, & Links, 2012; Mehlum et al., 2016; Neacsiu, Bohus, & Linehan, 2014), schema-focused therapy (Sempértegui, Karreman, Arntz, & Bekker, 2013), and community-based psychosocial treatment (CBP; Chiesa, Cirasola, & Fonagy, 2017; Chiesa, Fonagy, & Gordon, 2009) are promising treatment approaches in improving a number of clinical outcomes in a range of PDs. However, a recent meta-analysis (Cristea et al., 2017) of 33 randomized controlled trials found small to moderate differences in effectiveness between specialized psychotherapy and nonspecialized treatment in improving PD outcomes. Effect sizes were

further reduced when publication bias was taken into account, inviting caution when considering the results of these specialized treatment approaches. In addition, only a handful of studies have investigated the clinical efficacy of residential models versus less intensive approaches for PD (Bartak et al., 2011; Bartak et al., 2010; Chiesa et al., 2017). Although changes in mentalizing have been suggested to be a common factor underpinning symptomatic improvements and changes in functioning in all these contexts (Bateman, Campbell, Luyten, & Fonagy, 2018; Goodman, 2013; Montgomery-Graham, 2016), there is a dearth of research with regard to the evaluation of improvement in RF as a result of these different treatment approaches and the association between clinical changes and changes in RF in samples of patients with PD.

Psychotherapy in patients with PD and associated features has been shown to be typically associated with improvements in mentalizing, and changes in patients' mentalizing are associated with changes in their symptoms (Cologon, Schweitzer, King, & Nolte, 2017; De Meulemeester, Vansteelandt, Luyten, & Lowyck, 2018). However, changes in symptomatology are not invariably linked to improvement in mentalizing. Levy (2006), for instance, reported findings from a randomized controlled trial showing that only TFP, but not dialectical behavior therapy (DBT) or supportive therapy, was associated with changes in RF scored on the Adult Attachment Interview (AAI). Similarly, in another randomized controlled trial involving 104 patients with borderline PD (BPD) receiving either TFP or treatment by experienced community therapists, only patients in the TFP group showed significant improvements in RF assessed using the Reflective Functioning Scale (Fischer-Kern et al., 2015). Furthermore, these authors showed that improvements in RF were significantly correlated with improvements in personality organization, suggesting that changes in RF may be unique to psychodynamic treatments. Hence, more research is clearly needed in this area.

In this context, it is important to note that clinical approaches to the treatment of PD have gradually shifted over the past decades from high-intensity, low-volume, long-term residential programs to less intense, higher volume, day-hospital and outpatient programs. While services such as therapeutic communities were commonly offered to patients with PD in the past (Haigh, 2002; Norton & Hinshelwood, 1996), they have become increasingly less popular over the past three decades (Chiesa, 2005a; Rutter & Tyrer, 2003; Schimmel, 1997). One possible explanation for this change may lie in the association of intense attachment experiences and mentalizing (Zeegers, Colonnesi, Stams, & Meins, 2017). The excessive activation of attachment representations appears to compromise RF at both the behavioral and the neural level (Nolte et al., 2013; Nolte, Guiney, Fonagy, Mayes, & Luyten, 2011). It has been suggested that specialist residential treatment may be iatrogenic for many patients with PD, particularly for those with low levels of RF, because the constant activation of the attachment system within such intensive treatment programs may undermine patients' capacity for mentalizing and thus compromise the potential benefits of the program (Chiesa, Fonagy, & Holmes, 2003; Chiesa, Sharp, & Fonagy, 2011).

General psychiatric care may be a suboptimal context for bringing about changes in mentalizing for other reasons: It lacks a consistent focus on improving mentalizing capacity in patients with PD and it has been reported as having iatrogenic effects on PD patients' acute presentations, such as suicidality (Coyle, Shaver, & Linehan, 2018). Step-down programs and less intensive outpatient programs may provide a more optimal balance between a focus on reinforcing patients' capacity for mentalizing and a second focus on fostering social integration and patients' capacity to mentalize and function in their social world. In particular, these latter programs are thought to foster patients' capacity for epistemic trust and salutogenesis—that is, their capacity to benefit from positive social input in their everyday environment (Fonagy, Luyten, & Bateman, 2017).

The present investigation aimed to assess the degree of changes in RF in a sample of patients diagnosed with PD treated in specialist psychodynamic and nonspecialist settings. Patients were treated in (a) a mixed residential and community-based step-down program (RT-CBP) in which patients spent 6 months in inpatient psychotherapy treatment before stepping down to a psychosocial community-based approach; (b) a residential-only psychotherapy treatment program (RT) in which patients were admitted for a 12-month stay; and (c) a general psychiatric program (GP), which provided ongoing nonspecialist treatment for PD. Improving mentalizing was one of the aims of both the RT and RT-CBP services but was not considered pertinent to the GP intervention. Treatment outcome was assessed in three domains: psychiatric distress, social adjustment, and global functioning. RF was measured at intake and after 2 years. Treatment outcomes have been the subject of previous reports (Chiesa, Fonagy, & Holmes, 2006; Chiesa, Fonagy, Holmes, & Drahorad, 2004). Here we focused on the extent to which differences in the capacity for mentalizing may account for observed differences in the effectiveness of interventions in these settings. Based on findings of the potential iatrogenic effects of inpatient treatment, we expected that changes in RF would be greater in RT-CBP than in RT. Further, given the greater focus in GP on symptom management and medication rather than mentalizing, we expected to find only modest changes in RF associated with this intervention. Finally, we expected changes in RF to mediate improvements in clinical outcome in all three treatment models.

Method

Study Sample and Treatment Programs

The study sample consisted of patients who met PD criteria for at least one DSM-IV primary diagnosis, recruited from the [institution name] Personality Disorder Study (Chiesa & Fonagy, 2000a) and admitted to two specialist psychosocial treatments—step-down (RT-CBP) and long-term inpatient (RT)—at a PD specialist service, and to a general psychiatric

service (GP) located in a rural area in the south of England. Randomization to the three models was not practical, since patients referred to the RT model residing outside the Greater London area would have been unable to attend the outpatient phase of the step-down program if randomized to that condition. Therefore, we followed the hospital-established criteria of geographical accessibility to treatment, whereby the RT-CBP program was offered to patients admitted from within the Greater London area, and RT to patients referred from the rest of the Unite Kingdom. As discussed below, there were a few differences in demographic and clinical features among the three treatment groups, but none of these were related to RF.

Of the 160 patients who gave signed consent to study participation at intake into treatment (RT-CBP, *n* = 47; RT, *n* = 55; GP, *n* = 58), 17 were lost due to drop-out and consent withdrawal, leaving 143 patients by 24-month follow-up (RT-CBP, *n* = 45; RT, *n* = 49; GP, *n* = 49). Further participants were lost owing to a delay in introducing the AAI into the study (*n* = 18) and a number of patients not agreeing to either participate in the AAI (*n* = 9) or to be recorded as part of the interview (*n* = 5), leaving 111 patients with an AAI and a matching completed intake battery of measures to constitute the current study sample (RT-CBP, *n* = 32; RT, *n* =39; GP, *n* = 40). No differences in key demographic and risk factor variables were found between those who completed the AAI and those who did not.

All eligible patients were met by a senior researcher, who provided the study information sheet, discussed the study protocol, and sought written informed consent. Ethical approval was granted by the [institution name].

Residential program

The residential program (RT) is a multi-component psychosocial program that consists of a combination of formal psychodynamically oriented psychotherapy (twice-weekly individual psychotherapy and twice-weekly group therapy) and is run along the

principles of a therapeutic community (Hinshelwood & Skogstad, 1998). Therapeutic community treatment is run and managed by trained psychosocial nursing staff and consists of daily community meetings and a program of daily activities aimed at the acquisition of interpersonal skills, resocialization, and rehabilitation. Most patients admitted to this 16-bed unit are on psychotropic medication, which is monitored by psychiatric staff in daily multi-disciplinary meetings. The hospital has an “open door” policy and all patients are admitted to the hospital on a voluntary basis. After discharge from RT, patients are referred back to their mental health resource centers, which were involved in care program activity (CPA) meetings during the patients stay in RT.

Step-down program

Patients in the step-down program (RT-CBP) were treated for 6 months in a therapeutic community residential setting (as described above) before stepping down to a community-based program (CBP) for a further 2-year period of specialist outpatient treatment. A clinical program consisting of twice-weekly small-group psychodynamically oriented psychotherapy, twice-weekly outreach psychosocial nursing activities in the community, regular psychiatric reviews, family and couple therapy as required, social worker support, and psychotropic medication as required was provided by a dedicated multidisciplinary clinical team who met weekly for monitoring and management purposes (Chiesa et al., 2009). At the end of treatment, patients were referred back to their original general mental health teams.

General psychiatric program

Patients in the general psychiatric program (GP) received standard outpatient general psychiatric care, which included monitoring of psychotropic medication, supportive outpatient and community contact with one or more care workers (on average at intervals of 2–4 weeks), hospital acute admission or day hospital brief stay as required, and clinical

reviews on average once a month. This group would meaningfully reflect the typical outcomes expected from nonspecialist community-based treatment in the United Kingdom.

Measures

The *Structured Clinical Interview for DSM-IV* (First, Gibbon, Spitzer, Williams, & Benjamin, 1997; First, Spitzer, Gibbon, & Williams, 1997) were used to obtain diagnostic Axis-I and -II profiles based on the criteria of the DSM-IV (American Psychiatric Association, 1995). Axis-I diagnoses were grouped into four broad categories (mood disorder, anxiety disorder, eating disorder and substance misuse disorder), while the SCID-II yielded 11 different categories of PD diagnoses. The average number of PDs per patient and the total number of individual positive traits were then calculated. As previously reported, satisfactory inter-rater reliability was found in the application of the SCID interviews (Chiesa et al., 2004).

The *Symptom Checklist-90-R* (SCL-90-R; Derogatis, 1983), a four-point self-report clinical rating scale, was used to elicit symptoms in nine areas of the patient's functioning. The SCL-90-R general severity index (GSI) was the total score used in the study to report changes in the domains of subjective symptomatic distress.

The *Social Adjustment Scale* (SAS; Weissman, 1975) provides an assessment on a five-point scale of adjustment in the areas of work, family of origin, marriage, sexual functioning, and leisure activities. A total social adjustment score is derived from the means of the subcategories. An interclass correlation coefficient (ICC) of .78 obtained for the total score showed a satisfactory inter-rater agreement.

The *Global Assessment Scale* (GAS; Endicott, Spitzer, Fleiss, & Cohen, 1976) was used to evaluate patients' general outcome in accordance with their level of functioning assessed during the 4 weeks prior to the assessment. Good inter-rater reliability was found between research staff (ICC = .79).

The *Adult Attachment Interview* (AAI; George, Kaplan, & Main, 1985) is a semi-structured interview that focuses on the individual's childhood experiences and relationship with his or her primary caregivers (Hesse, 2016). The interviewer enquires about the subjects' relationship with the parents, possible occurrence of emotional upsets brought about by rejection during childhood, feelings of being frightened, separation from parents, experiences of abuse, significant losses through death or abandonment, and presence of possible traumatic experiences. The interviewee is then asked to evaluate the impact of these early experiences on his or her adult personality and to offer a possible explanation for the primary figures' behaviors in the past. Changes in the relationship with caregivers since childhood and (if relevant) the perceived impact of past experiences on the individual's relationship with his or her own child are also explored. The AAI has been widely used in psychotherapy research to study the association of early adversity with adult psychopathology such as depression, BPD, and posttraumatic stress disorder (Steele, 2003; Steele, Steele, & Murphy, 2009).

The *Reflective Functioning Scale* (RFS; Fonagy, Steele, Steele, & Target, 1998) was applied to AAI transcripts, which were coded by two independent raters of established reliability who were masked to the sample's treatment model allocation. The RFS is an 11-point scale that evaluates the quality of mentalizing in the context of attachment relationships as manifested in AAI transcripts. The RFS score ranges from –1 (negative RF, in which interviews appear explicitly antagonistic to thinking about mental states) through +1 to 3 (interviews show concrete or excessively detailed, low-level mentalizing, with distortions of the mental states of others), to 5 to 9 (interviews show increasingly complex, elaborate, or insightful reasoning about mental states).

The SCL-90-R, SAS, GAS, AAI, and RFS were applied at intake and at 24-month follow-up.

Information regarding demographics and risk factors (history of sexual and physical abuse) were also collected at intake into the study through a structured interview using the [institution name] *Baseline Questionnaire* (Chiesa & Fonagy, 2000b).

Statistical Analysis

All analyses were performed using SPSS for Windows version 26. Chi-squared tests for categorical variables and one-way ANOVAs for continuous variables were used to test differences between the samples in demographics, risk factors, and clinical variables. Linear and partial correlations were carried out to test for significant associations between pooled key variables in the study sample.

The SPSS general linear model for repeated measures (ANCOVA) was applied to the interval data from the three main clinical outcome measures (SAS, GAS, and GSI), with one repeated measure factor (Time), a between-subjects factor (Group), and significant intake between group variables (marital status, education, and PD dimensional traits) as covariates. Bonferroni tests of adjustment were used in pairwise contrasts to correct for risk of type I errors. As the outcome measures were highly correlated, a single integrated outcome variable was also used to evaluate changes over time and examine the association of RF and treatment outcomes in the three samples. Intake and 24-month follow-up scores were created by averaging the standardized scores from the SAS, GAS, and GSI. A GLM ANCOVA was then applied to the repeated measures data to test the significance of changes in the three groups.

A moderate effect size ($g > .35$) for the difference between the groups' mean change values was taken as an indication of treatment superiority.

Using G*Power software (Faul, Erdfelder, Buchner, & Lang, 2009) specifying a medium effect size of .35 with alpha set at .05 for three groups, two points of measurement,

and three covariates, our sample with 111 participants yielded 91% power for detecting between group interactions.

The extent of change in RF was computed by subtracting the 24-month scores from the scores at intake. One-way ANOVA was then applied to the computed "change in RF" variable to test significance between the three treatment models. The effect size of the difference in RF change between the three groups was calculated using Hedge's *g* formula ($mean_1 - mean_2 / SD_{pooled}$) (Ellis, 2010).The next step in the analyses was to compare differences in RF change between intake and 24-month follow-up values between the three treatment programs, using multilevel modeling (MLM) (Field, 2013). The best fit model included "RF" as the dependent variable, "Treatment Group" as independent variable, and "Time" and "PD Dimensional" as predictors, allowing both intercept and slopes to vary randomly for the effect of Time. Scaled Identity as covariance structure to the model was selected.

After transforming the clinical outcome variables at 24 months (SAS24 and GAS24) into standardized variables, we computed a single clinical outcome variable by combining the averaged SAS and GAS standardized values. In order to test the hypothesis that change in RF mediates the relationships between PD specialist (RT-CBP and RT) and nonspecialist (GP) treatment models, and clinical outcomes at 24-month follow-up, and since the correlational analysis showed that changes in RF were not significantly associated with psychiatric distress (GSI), three separate mediation analyses were carried out with SAS24, GAS24, and overall clinical outcome as dependent variables, treatment group (specialist vs. nonspecialist) as the independent variable, and change in RF as the mediating variable. PROCESS (Hayes & Matthes, 2009) was used as the computational tool to carry out the three mediation analyses. Following the estimation of the coefficients of the model, in

maximum likelihood regression, PROCESS generates direct and indirect effects, which indicate the level of significance of the mediating variable.

Results

Demographics, Risk Factors, and Reflective Functioning

Table 1 shows the demographic, diagnostic and clinical features of the PD samples. Patients were well matched on age, sex, race, history of abuse, PD diagnoses, intake levels of symptom distress, social maladjustment, global functioning, and RF. However, significant differences in marital status, $\chi^2_{(1)} = 27.09$, $p < .001$, education, $\chi^2_{(1)} = 27.09$, $p < .001$, and positive PD traits as scored on the SCID-II, $F_{(2, 108)} = 6.11$, $p = 0.003$, were present. These variables were not correlated with RF scores at intake: marital status, $r = .06$, $p = .517$; education, $r = .11$, $p = 0.262$; PD traits, $r = –0.13$, $p = .170$. The GP sample had a higher number of people who had ever been married, with a lower level of educational attainment and less positive individual PD traits compared with the two specialist samples. The typical patient was in the early thirties, female, and White, and one in two had experienced sexual and/or physical abuse before the age of 18. With regard to Axis-I psychiatric diagnoses, no significant differences between the three samples were found. Mood disorder (N = 59, 53%), anxiety disorder (N = 61, 55%), eating disorder (N = 17, 15.3 %) and substance abuse (N = 17, 15.3 %) were the DSM-IV psychiatric diagnoses found in the total sample. Comorbidity between Axis-II diagnoses was high, with a mean of 3.4 (*SD* = 1.7) diagnoses per patient. 73 (65%) patients met criteria for BPD, with 38 (35%) patients being diagnosed with a non-borderline PD. The most represented PD diagnoses in this latter group were avoidant PD, paranoid PD, obsessive–compulsive PD and dependent PD. We found that the most frequent co-occurrence between Axis-I and –II disorders involved borderline PD and mood disorder (N = 41, 36.9%) and borderline PD, and anxiety disorder and borderline PD (N = 39, 35.1%).

A bivariate correlational analysis revealed a significant association between PD dimensional traits, GSI, and SAS at baseline, $r = .37$, $p > .001$, and $r = .28$, $p = .003$, respectively, while a partial correlation controlling for intake scores showed a significant association between GSI24 and SAS24, $r = .63$, $p < .001$, between GAS24 and SAS24, $r = –.71$, $p < .001$, and between GSI24 and GAS24, $r = –.68$, $p < .001$.

Clinical Outcome at 24-Month Follow-Up

Table 2 outlines means, standard deviations, estimated marginal means, between-group mean differences, and effect sizes for the changes between intake and 24-month follow-up in SAS, GAS, GSI, and overall outcome.

We found a significant effect of Time, Wilks's lambda = .93, $F_{(1, 105)} = 8.51$, $p = .004$, and a Time × Group interaction, Wilks's lambda = .88, $F_{(2, 105)} = 7.12$, $p = .001$, for symptom distress (GSI). While a large effect size was found between RT-CBP and GP ($g = .89$), small to medium effect sizes were found between RT and GP ($g = .46$) and between RT-CBP and RT ($g = .46$).

A significant effect of Time, Wilks's lambda = .88, $F_{(1, 105)} = 14.89$, $p < .001$, and a Time × Group interaction, Wilks's lambda = .92, $F_{(2, 105)} = 4.57$, $p = .013$, was also found for global functioning. A medium effect size was found in GAS change between RT-CBP and GP ($g = .73$), a small to medium effect size between RT and GP ($g = .46$), and a small effect size between RT-CBP and RT ($g = .27$).

With regard to changes in social adjustment, a significant Time effect, Wilks's lambda = .96, $F_{(1, 105)} = 4.24$, $p = .042$, and a Time × Group interaction, Wilks's lambda = .87, $F_{(2, 105)} = 8.22$, $p < .001$, were also found. The size of the effect of SAS change was large between RT-CBP and GP ($g = .98$), medium between RT and GP ($g = .68$), and small between RT-CBP and RT ($g = .29$).

Across the three outcome dimensions, we found that RT-CBP and RT were superior to GP. Although effects for RT-CBP were greater than those for RT, RT-CBP was not found to be superior to RT, as shown by between-group comparisons and effect sizes (Table 2).

Multivariate statistics for the simple effect of Time for the standardized overall outcome variable showed that RT-CBP improved significantly between intake and 24-month follow-up, Wilks's lambda = .93, $F_{(1, 105)} = 7.68$, $p = .007$, while RT did not significantly improve, Wilks's lambda = .99, $F_{(1, 105)} = .30$, $p = .583$, and GP significantly deteriorated, Wilks's lambda = .94, $F_{(1, 105)} = 7.23$, $p = .008$ (Figure 1). The effect size was large between RT-CBP and GP ($g = .87$), medium between RT and GP ($g = .51$), and small between RT-CBP and RT ($g = .35$). The ANOVA repeated measures showed a significant Time × Group interaction, Wilks's lambda = .88, $F_{(1, 105)} = 6.85$, $p = .002$, but no main effect of Time, Wilks's lambda = .99, $F_{(1, 105)} = .22$, $p = .637$. Hence, for overall outcome, RT-CBP and RT were superior to GP, while RT-CBP was not superior to RT.

Change in Reflective Function

A one-way-ANOVA showed a significant difference between the three treatments in RF level of change between intake and 2-year follow-up, $F_{(2, 8.5)} = 20.47$, $p < .001$. A large effect size between RT-CBP and GP ($g = 1.54$), a medium to large effect size between RT and GP ($g = .81$), and a medium effect size between RT-CBP and RT ($g = .65$) were found. Hence, as predicted, RT-CBP was associated with the largest changes in RF.

The MLM analysis revealed no significant linear relationship between RF and Time, $F_{(1, 134.7)} = 1.07$, $p = .302$, indicating that the total sample did not improve significantly over the 2-year period. However, a significant interaction between Treatment Group and Time, $F_{(2, 106.3)} = 4.08$, $p = .020$, was found. The linear trend of the improvement in RF was significantly greater in RT-CBP compared with GP, $b = .68$, $SE = .24$, 95% CI [0.20, 1.16], $t_{(104.6)} = 2.81$, $p = .006$, while RT showed a nonsignificant improvement relative to GP, $b =$

.40, *SE* = .23, 95% CI [–0.06, 0.86], $t_{(105.9)} = 1.71$, $p = .090$ (Figure 2). PD dimensional was a significant covariate in the model, $F_{(1, 151.8)} = 4.47$, $p = .036$.

Relationship between Clinical Outcome and Changes in Reflective Function

Partial correlational analyses between changes in RF and clinical outcome at 24-month follow-up in SAS, GAS, and GSI, controlling for intake scores, showed that changes in RF were significantly associated with social adjustment, $r = –.36$, $p < .001$, and global functioning, $r = .29$, $p = .002$, but not with psychiatric distress, $r = –.14$, $p = .141$.

In the mediation analysis with Treatment Group (specialist vs nonspecialist) as predictor variable, SAS24 as outcome variable and RF change as mediator, Treatment Group was found to be a significant predictor of RF change, $\beta = 0.53$, *SE* = .09, $t = 5.70$, $p < 0.001$, and of social adjustment at 24-month follow-up, $\beta = –.42$, *SE* = .10, $t = –4.10$, $p < .001$. RF change also significantly predicted SAS24, $\beta = –.20$, *SE* = .09, $t = –2.10$, $p = .038$. RF change was found to be a significant mediator between Treatment Group and SAS24, $\beta = –.10$, 95% CI [–0.22, –0.01], *SE* = .05, $t = –2.00$, $p = .053$, although the direct effect remained significant $\beta = -.42$, 95% CI [-.62, -.22], *SE* = -.10, $t = -4.10$, $p < .001$, indicating that RF change was a significant partial mediator in the equation The size of the indirect effect of RF change was medium, d = .41.

The mediation analysis with treatment group as predictor variable, GAS24 as outcome variable and RF change as mediator, showed that the difference between the specialist and nonspecialist groups and RF change was significant, $\beta = –6.35$, *SE* = 2.77, $t = –2.30$, $p = .024$. RF change and Treatment Group also significantly predict GAS24 scores, $\beta = –5.09$, *SE* = 2.52, $t = –2.02$, $p = .046$, and $\beta = 6.35$, *SE* = 2.76, $t = 2.30$, $p = .024$. RF change was found to be a significant partial mediator between Treatment Group and GAS24, $\beta = 2.67$, 95% CI [0.46, 5.44], *SE* = 1.25, $t = 2.14$, $p = .040$. The direct effect was also

significant, β = 6.35, 95% CI [.87, 11.83], *SE* = 2.76, *t* = 2.30, *p* = .024. The size of the indirect effect of RF change was medium, d = .40.

Both Treatment Group and RF change also significantly predicted the combined standardized clinical outcome at 24-month follow-up, β = .45, *SE* = .20, *t* = 2.23, *p* = .028; and β = –1.09, *SE* = .32, *t* = –3.36, *p* = 0.001, respectively. Treatment Group was a significant predictor for RF change, β = –.20, *SE* = .06, *t* = –3.56, *p* < .001.The indirect effect of Treatment Group and clinical outcome with RF change as mediator was significant, β = .22, 95% CI [0.07, 0.43], *SE* = .09, *t* = 2.38, *p* = .023, although the direct effect was also significant, β = .45, 95% CI [0.05, 0.85], *SE* = .20, *t* = 2.23, *p* = .028. The effect size for the indirect effect of RF change on the overall clinical outcome was medium-to-large, d = .66.

Figure 3 outlines the levels of significance between specialist vs. non-specialist treatment groups, change in reflective function, and the three clinical outcomes at 24-month follow-up.

Following the suggestion of one of the reviewers, we also investigated the impact of each individual treatment model in increasing RF levels by carrying out separate correlational and mediation analyses, with Bonferroni correction (.05 / 3 = .017). Significant associations were found in the RT-CBP group between RF change and SAS24, *r* = .47, *p* = .006, but not for overall clinical outcome, *r* = .35, *p* = .047, and for GAS24, *r* = .25, = .176. No significant associations in the RT group were found between RF change and SAS24, *r* = .24, *p* = .150, GAS24, *r* = .22, *p* = .178, and overall clinical outcome, *r* = .19, *p* = .260. Similarly, the GP group was found to have no significant associations between RF change and SAS24, *r* = .14, *p* = .375, GAS24, *r* = .03, *p* = .856, and overall clinical outcome, *r* = .131, *p* = .419.

First, when the GP group was excluded and the two specialist programs were compared, we found that RF change was a significant predictor of SAS24, β = –.36, *SE* = .13,

$t = –2.85$, $p = .006$. RT change was a significant mediator between treatment group (RT and RT-CBP) and SAS24, $\beta = .10$, 95% CI [.02, .21], $SE = .05$, $t = 2.11$, $p = .043$. The magnitude of the indirect effect was medium, d = .42. RF change was also revealed to markedly mediate between treatment group and GAS24, $\beta = 2.05$, 95% CI [.16, 4.67], $SE = 1.15$, $t = 1.78$, $p = .082$, with an indirect effect size of medium magnitude, d = .41. This suggests that the magnitude of RF change was an important factor distinguishing these types of intervention.

When we compared RT-CBT and GP, significant effects between group and RF change and between group and SAS24 were found, $\beta = .68$, 95% CI [.47, .89], $SE = .10$, $t = 6.56$, $p < .001$, and $\beta = .57$, 95% CI [.28, .85], $SE = .14$, $t = 3.95$, $p < .001$, respectively. However, RF change was not a significant mediator in this equation, $\beta = .09$, 95% CI [-.09, .27], $SE = .09$, $t = .96$, $p = .252$. We also found significant predictions between RF change and GAS24, $\beta = 6.73$, 95% CI [2.03, 11.43], $SE = 2.37$, $t = 2.84$, $p = .005$, and between group and RF change, $\beta = .20$, 95% CI [.09, .31], $SE = .06$, $t = 3.56$, $p < .001$. RF change was found to be a marked, albeit nonsignificant mediator between group and SAS24, $\beta = 1.35$, 95% CI [.36, 2.73], $SE = .61$, $t = 2.21$, $p < .034$.

No significant association between RF change and SAS24 or GAS 24 was found, when comparing RT and GP. RF Change was not a significant mediator between group and clinical outcome.

These results suggest that, when compared to RT and GP, the better clinical outcomes in RT-CBT was linked to positive changes in RF in that group alone.

Discussion

The results of this study provide further evidence that specialist psychosocial residential and step-down programs may be superior to a general psychiatric approach in improving symptom distress, social adjustment, global functioning, and overall clinical

outcome in a group of patients with PD (Chiesa et al., 2006; Chiesa et al., 2004). The difference between the three treatment programs with regard to psychiatric distress symptoms, social adjustment, global functioning, and overall outcome appear to be clinically meaningful: in GP there was no change or deterioration, in RT moderate changes were found, while RT-CBP achieved larger improvements as shown by the effect sizes. However, when we compared changes in RF we found a slight deterioration in GP, only a marginal improvement in RT, and significant improvement in RT-CBP, the last group being the closest to reaching the average score of 5 on the RFS, which represents the conventional cut-off for adaptive functioning.

We also found that treatment model was a significant predictor of changes in RF in the three groups, with patients assigned to specialist RT-CBP improving significantly more compared with those in the specialist RT or GP programs. In RT, levels of RF hardly improved, while in GP there was evidence for deterioration of RF. This finding is consistent with the notion that patients in GP may show RF deterioration as a result of their relative isolation, which may have a toxic effect on their capacity for RF that is not counteracted by the therapeutic input associated with general psychiatric care (Luyten, Campbell, Allison, & Fonagy, 2020). Specialist residential treatment, meanwhile, consisted of an intensive structured psychosocial therapeutic program aimed at improving RF, but the constant activation of the attachment system inherent in the intensity of the therapeutic community setting may actually undermine the potential benefits of the program, which may explain the marginal improvements in RF in this group (Chiesa et al., 2011). In contrast, RT-CBP may provide an optimal compromise between a situation of relative neglect (GP) and over-intensive therapeutic input (RT), both of which weaken rather than strengthen mentalizing. RT-CBP may thus provide a better balance between intensive psychotherapy input and external efforts at social integration, by bringing patients into closer contact with their social

world, enhancing their mentalizing (Fonagy & Allison, 2014; Fonagy, Luyten, Allison, & Campbell, 2019). The results of the mediation analysis strengthen this argument by pointing to a pathway in which the effect of treatment model on changes in social and global functioning was mediated by improvement in RF.

Three specific issues related to the observed pattern of results deserve further comment. First, the mediation analysis revealed that while the combined outcome measures pointed to a partial mediation through RF, there was a substantially stronger effect in relation to both global functioning and social adjustment than for general symptom severity. This observation is in line with a conceptualization of RF as rooted in social interaction and social collaboration (Fonagy et al., 2019; Fonagy et al., 2017; Luyten, Campbell, & Fonagy, 2020). The capacity for mentalizing emerges in the context of the family environment, and the interaction between caregiver and child generates increased sophistication about the state of mind of the other and, by inference, the state of mind of the self (Sroufe, 1983; Sroufe & Fleeson, 1986). Similarly, in a therapeutic context, interactions between therapists and patients—and indeed between patients—are likely to gradually enhance an individual patient's capacity for mentalizing (Luyten, Campbell, Allison, et al., 2020). Such opportunities are clearly greater in the context of RT and RT-CBP than in GP. Since improved RF is likely first and foremost to enhance the capacity for social adaptation and social collaboration, a greater impact of psychosocial interventions and RF on the social domain is to be expected and may have significant implications for our understanding of the therapeutic process.

Second, the finding that RF deteriorated in the GP condition can be understood in terms of the stigmatization and associated social isolation of individuals with a PD diagnosis (Chiesa, 2005b; Lewis & Appleby, 1988). It is not surprising that the wish to socially engage and maintain a propensity for social interaction declines against the background of frequent

misunderstandings and misattributions to which individuals with a tendency for self-harm and suicidality are frequently exposed in society and, all too frequently, within the medical context of general psychiatric services. It is perhaps more surprising that improvement in RF was marginal within the RT specialist setting. Our understanding of this finding is linked to the potential challenge to interpersonal interaction that arises in therapeutic community settings. Individuals with limited understanding and appreciation of the complex determinants of behavior in terms of mental states frequently find themselves in conflict with others in the close social environment of the inpatient therapeutic community setting. When these conflicts occur against a background of developing attachment relationships with patients and staff in the confined social environment of a small hospital unit, it is likely that interactions will often be highly charged with emotion, which compromises the potential for developing interpersonal understanding. The emotional dysregulation that occurs commonly in such environments directly inhibits the capacity for mentalizing (Chiesa, 2010; Chiesa et al., 2011). Thus, the potential for enhanced mentalizing is a key part of the therapeutic community, and the careful regulation of emotion-driven interactions must be a critical role for staff in these settings. Of course, intense emotionality is contagious, and experienced members of staff are just as vulnerable to the intense emotions of those around them as are other patients. In contrast, RT-CBP was uniquely associated with improvement in RF. A treatment environment that combines the advantages of kick-starting mentalizing through social interaction needs to be balanced by a loosening of emotional ties rather than enabling such links to become overly intense and entangled, as may occur in the course of a 12-month inpatient stay. The RT-CBP program may have offered a judicious compromise between the extremes of absent involvement and over-involvement. It is likely that other modes of treatment, such as intensive outpatient treatment, may represent similar opportunities for creating a platform for developing mentalizing at the same time as giving the individual

patient sufficient space to practice mentalizing according to his or her capacities and gradually improving both the capacity for social interaction and the patient's prognosis (Smits et al., 2020).

Third, the surprising finding that changes in RF were not associated with improvement in psychiatric symptom distress requires further explanation. Perhaps the simplest explanation rests with possible floor or ceiling effects, in which change in RF is not large enough to show its impact on psychiatric distress, or is large because of regression to the mean, or some process going both ways (e.g., patients with better RF show more improvement but are also more aware of their symptoms). Another explanation may be that a focus on symptomatic change is not subtle enough to pick up experiential differences of pathology that RF marks. However, changes in RF predicted changes in social and global functioning. This can be accounted for by the role RF plays in ensuring connection to the social environment and collaboration with others, including a better understanding of their mental states (Fonagy et al., 2019). This may also indicate that improvement in symptom distress may be an indirect effect of changes in RF. Symptoms may be improved by the development of social connections and improved relationships as well as by the therapeutic experience. So, improvements in RF may lead to improvements in social functioning and global functioning, which in turn improves symptoms. Further research is needed to investigate this assumption.

This study has several advantages over previous reports. It is a relatively large-scale and adequately powered study. RF was measured in the context of AAIs, and the interviews were independently coded by raters outside the clinical settings and masked to the treatment program allocation. The diagnostic characterization of the samples was based on structured interviews with good inter-rater reliability. Above all, the longitudinal nature of the study allowed us to assess the potential mediator variable in advance of the dependent variables for

a more accurate evaluation of the pathways leading to clinical outcomes. The psychiatric sample was drawn from a range of clinical settings and is likely to be representative of individuals with severe PD presenting to clinical services.

However, the conclusions should be qualified by substantial limitations. The study design did not entail a randomization of the subjects to the three treatment conditions, which raises the possibility of selection bias undermining the comparability of the groups. In particular, the GP group was drawn from a PD prevalence community sample, while the specialist groups consisted of patients selected from mental health services and referred for tertiary care. It can be argued that these patients were selected on the basis of their potential for responsiveness to these specialist clinical programs, although it could be equally argued that only the most difficult and severe cases are indicated for a long-term and very expensive treatment. To mitigate the potential threat to the internal validity of the study, the groups were relatively well matched on most demographic, diagnostic, and other clinical (including severity) criteria that were measured. Moreover, in the statistical analyses we controlled for the three variables that were significantly different, which did not affect the observed differences in outcome. In addition, the lack of a comparison with an untreated control group, which would have strengthened the case for the observed changes in RF being the result of the treatments evaluated in this study. Comparison with other specialist outpatient programs such as MBT, TFP, or DBT would have yielded clearer indications regarding the relative effectiveness of the RT-CBP and RT specialist programs in bringing about positive change in RF. A further limitation is in the choice of the measures used to assess severity of presentation and longitudinal change. Although psychiatric morbidity, social adjustment and global functioning represent important dimensions for PD evaluation, in the light of recent trends in diagnosing and conceptualizing PD (Hopwood et al., 2018; Skodol, 2012; Tyrer, Crawford, & Mulder, 2011) it would have been desirable to have added a measure that more

specifically and accurately evaluated core aspects of change in personality functioning. The employment of a personality measure may have given greater sensitivity and specificity in revealing significant associations with changes in RF.

Despite these limitations, this study is the first to suggest that the extent to which treatments may foster changes in RF may explain some of the observed differences in the effectiveness of different specialist and nonspecialist treatment programs for patients with PD. Moreover, this study also suggests that treatment programs that provide an optimal balance between creating a platform for developing mentalizing and providing opportunities for the individual to practice mentalizing in real-world interactions may be the most effective.

References

American Psychiatric Association. (1995). *Diagnostic and Statistical Manual for Mental Disorders (DSM-IV)* (IV ed.). Washington: American Psychiatric Press.

Antonsen, B. T., Johansen, M. S., Ro, F. G., Kvarstein, E. H., & Wilberg, T. (2016). Is reflective functioning associated with clinical symptoms and long-term course in patients with personality disorders? *Compr Psychiatry, 64*, 46-58. doi: 10.1016/j.comppsych.2015.05.016

Bartak, A., Andrea, H., Spreeuwenberg, M. D., Ziegler, U. M., Dekker, J., Rossum, B. V., . . . Emmelkamp, P. M. (2011). Effectiveness of Outpatient, Day Hospital, and Inpatient Psychotherapeutic Treatment for Patients with Cluster B Personality Disorders. *Psychothererapy and Psychosomatics, 80*, 28-38. doi: 10.1159/000321999

Bartak, A., Spreeuwenberg, M. D., Andrea, H., Holleman, L., Rijnierse, P., Rossum, B. V., . . . Emmelkamp, P. M. G. (2010). Effectiveness of Different Modalities of Psychotherapeutic Treatment for Patients with Cluster C Personality Disorders: Results of a Large Prospective Multicentre Study. *Psychotherapy and Psychosomatics, 79*, 20-30. doi: 10.1159/000254902

Bateman, A., Campbell, C., Luyten, P., & Fonagy, P. (2018). A mentalization-based approach to common factors in the treatment of borderline personality disorder. *Curr Opin Psychol, 21*, 44-49. doi: 10.1016/j.copsyc.2017.09.005

Bateman, A., & Fonagy, P. (2009). Randomized Controlled Trial of Outpatient Mentalization-Based Treatment Versus Structured Clinical Management for Borderline Personality Disorder. *American Journal of Psychiatry, 166*, 1355-1364. doi: 10.1176/appi.ajp.2009.09040539

Bateman, A., & Fonagy, P. (2010). Mentalization based treatment for borderline personality disorder. *World Psychiatry, 9*, 11-15.

Chiesa, M. (2005a). Hospital-based psychosocial treatment of borderline personality disorder: A systematic review of rationale and evidence-base *Clinical Neuropsychiatry, 2*, 292-301.

Chiesa, M. (2005b). Stigma and personality disorder. *Noos Aggiornamenti in Psichiatria, 11*, 185-196.

Chiesa, M. (2010). Anti-therapeutic factors in the long-term residential treatment of personality disorder. *Psichiatria di Comunitá, 9*, 113-119.

Chiesa, M., Cirasola, A., & Fonagy, P. (2017). Four years comparative follow-up evaluation of community-based, step-down, and residential specialist psychodynamic programmes for personality disorders. *Clinical Psychology & Psychotherapy, 24*, 1331-1342. doi: 10.1002/cpp.2109

Chiesa, M., & Fonagy, P. (2000a). Cassel Personality Disorder Study. Methodology and treatment effects. *British Journal of Psychiatry, 176*, 485-491. doi: 10.1192/bjp.176.5.485

Chiesa, M., & Fonagy, P. (2000b). Cassel personality disorder study: methodology and treatment effects. *British Journal of Psychiatry, 176*, 485-491.

Chiesa, M., & Fonagy, P. (2014). Reflective function as a mediator between childhood adversity, personality disorder and symptom distress. *Personality and Mental Health, 8*, 52-66. doi: 10.1002/pmh.1245

Chiesa, M., Fonagy, P., & Gordon, J. (2009). Community-based psychodynamic treatment program for severe personality disorders: Clinical description and naturalistic evaluation. *Journal of Psychiatric Practice, 15*, 12-24. doi: 10.1097/01.pra.0000344915.61706.d4

Chiesa, M., Fonagy, P., & Holmes, J. (2003). When less is more: An exploration of psychoanalytically oriented hospital-based treatment for severe personality disorder. *International Journal of Psycho-Analysis, 84*, 637-650.

Chiesa, M., Fonagy, P., & Holmes, J. (2006). Six-year follow-up of three treatment programs to personality disorder. *Journal of Personality Disorders, 25*, 493-509.

Chiesa, M., Fonagy, P., Holmes, J., & Drahorad, C. (2004). Residential versus community treatment of personality disorder: A comparative study of three treatment programs. *American Journal of Psychiatry, 161*, 1463-1470.

Chiesa, M., Sharp, R., & Fonagy, P. (2011). Clinical associations of deliberate self-injury and its impact on the outcome of community-based and long-term inpatient treatment for personality disorder. *Psychotherapy and Psychosomatics, 80*, 100-109. doi: 10.1159/000320975

Cirasola, A., Hillman, S., Fonagy, P., & Chiesa, M. (2017). Mapping the road from childhood adversity to personality disorder: The role of unresolved states of mind. *Personal Ment Health, 11*, 77-90. doi: 10.1002/pmh.1365

Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2007). Evaluating Three Treatments for Borderline Personality Disorder: A Multiwave Study. *American Journal of Psychiatry, 164*, 922-928. doi: 10.1176/appi.ajp.164.6.922

Cologon, J., Schweitzer, R. D., King, R., & Nolte, T. (2017). Therapist Reflective Functioning, Therapist Attachment Style and Therapist Effectiveness. *Adm Policy Ment Health, 44*, 614-625. doi: 10.1007/s10488-017-0790-5

Coyle, T. N., Shaver, J. A., & Linehan, M. M. (2018). On the potential for iatrogenic effects of psychiatric crisis services: The example of dialectical behavior therapy for adult women with borderline personality disorder. *Journal of Consulting and Clinical Psychology, 86*, 116-124. doi: 10.1037/ccp0000275

Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of Psychotherapies for Borderline Personality Disorder: A Systematic Review and Meta-analysis. *JAMA Psychiatry, 74*, 319-328. doi: 10.1001/jamapsychiatry.2016.4287

De Meulemeester, C., Vansteelandt, K., Luyten, P., & Lowyck, B. (2018). Mentalizing as a mechanism of change in the treatment of patients with borderline personality disorder: A parallel process growth modeling approach. *Personal Disord, 9*, 22-29. doi: 10.1037/per0000256

DeCou, C. R., Comtois, K. A., & Landes, S. J. (2019). Dialectical Behavior Therapy Is Effective for the Treatment of Suicidal Behavior: A Meta-Analysis. *Behavior Therapy, 50*, 60-72. doi: https://doi.org/10.1016/j.beth.2018.03.009

Derogatis, L. R. (1983). *SCL-90-R. Administration, scoring and procedures. Manual-II*. Towson (MD): Clinical Psychometric Research.

Doering, S., Horz, S., Rentrop, M., Fischer-Kern, M., Schuster, P., Benecke, C., . . . Buchheim, P. (2010). Transference-focused psychotherapy v. treatment by community psychotherapists for borderline personality disorder: randomised controlled trial. *Br J Psychiatry, 196*, 389-395. doi: 10.1192/bjp.bp.109.070177

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, UK: Cambridge University Press.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale. *Archives of General Psychiatry, 33*, 766-771.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods, 41*, 1149-1160. doi: 10.3758/BRM.41.4.1149

Field, A. (2013). Multilevel linear models. In A. Field (Ed.), *Discovering Statistics Using IBM SPSS Statistics* (4th ed., pp. 814-866). London: SAGE Publications.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B., & Benjamin, L. S. (1997). *Structured Clinical Interview for DSM-IV Axis-II Personality Disorders*. Washington DC: American Psychiatric Press.

First, M. B., Spitzer, R., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for the DSM-IV*. Washington DC: American Psychiatric Press.

Fischer-Kern, M., Doering, S., Taubner, S., Horz, S., Zimmermann, J., Rentrop, M., . . . Buchheim, A. (2015). Transference-focused psychotherapy for borderline personality disorder: change in reflective function. *Br J Psychiatry, 207*, 173-174. doi: 10.1192/bjp.bp.113.143842

Fonagy, P., & Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy, 51*, 372-380. doi: 10.1037/a0036505

Fonagy, P., Campbell, C., & Bateman, A. (2016). An update on diagnostic issues for borderline personality disorder. *Psychiatric Times, (in press)*

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2019). Mentalizing, Epistemic Trust and the Phenomenology of Psychotherapy. *Psychopathology, 52*, 94-103. doi: 10.1159/000501526

Fonagy, P., Luyten, P., & Bateman, A. (2017). Treating Borderline Personality Disorder With Psychotherapy: Where Do We Go From Here? *JAMA Psychiatry, 74*, 316-317. doi: 10.1001/jamapsychiatry.2016.4302

Fonagy, P., Steele, M., Steele, H., & Target, M. (1998). *Reflective-function manual: Version 5.0 for application to the adult attachment interview*. Research Department of Clinical, Educational and Health Psychology. University College London. Unpublished manuscript.

George, C., Kaplan, N., & Main, M. (1985). *The Adult Attachment Interview*, Department of Psychology, University of California at Berkeley.

Goodman, G. (2013). Is mentalization a common process factor in transference-focused psychotherapy and dialectical behavior therapy sessions? *Journal of Psychotherapy Integration, 23*, 179.

Gullestad, F. S., Wilberg, T., Klungsoyr, O., Johansen, M. S., Urnes, O., & Karterud, S. (2012). Is treatment in a day hospital step-down program superior to outpatient individual psychotherapy for patients with personality disorders? 36 months follow-up of a randomized clinical trial comparing different treatment modalities. *Psychother Res, 22*, 426-441. doi: 10.1080/10503307.2012.662608

Haigh, R. (2002). Therapeutic community research: Past, present and future. *Psychiatric Bulletin, 26*, 65-68.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods, 41*, 924-936.

Hesse, E. (2016). The Adult Attachment Interview Protocol, Method of Analysis, and Selected Empirical Studies: 1985–2015. In J. Cassidy & P. R. Shaver (Eds.), *Hanbook of Attachment* (3rd ed., pp. 553-597). London: Guilford Press.

Hinshelwood, R. D., & Skogstad, W. (1998). The hospital in the mind: In-patient psychotherapy at the Cassel Hospital. In J. Pestalozzi, S. Frisch, R. D. Hinshelwood & D. Houzel (Eds.), *Psychoanalytic Psychotherapy in Institutional Settings* (pp. 59-73). London: Karnac Books.

Hopwood, C. J., Kotov, R., Krueger, R. F., Watson, D., Widiger, T. A., Althoff, R. R., . . . Zimmermann, J. (2018). The time has come for dimensional personality disorder diagnosis. *Personal Ment Health, 12*, 82-86. doi: 10.1002/pmh.1408

Laurenssen, E. M. P., Luyten, P., Kikkert, M. J., Westra, D., Peen, J., Soons, M. B. J., . . . Dekker, J. J. M. (2018). Day hospital mentalization-based treatment v. specialist treatment as usual in patients with borderline personality disorder: randomized controlled trial. *Psychological Medicine, 48,* 2522-2529. doi: 10.1017/S0033291718000132

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.

Lewis, G., & Appleby, L. (1988). Personality disorder: the patient psychiatrists dislike. *British Journal of Psychiatry, 153*, 44-49.

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). The Mentalizing Approach to Psychopathology: State of the Art and Future Directions. *Annu Rev Clin Psychol, 16*, 297-325. doi: 10.1146/annurev-clinpsy-071919-015355

Luyten, P., Campbell, C., & Fonagy, P. (2020). Borderline personality disorder, complex trauma, and problems with self and identity: A social-communicative approach. *J Pers, 88,* 88-105. doi: 10.1111/jopy.12483

McMain, S. F., Guimond, T., Streiner, D. L., Cardish, R. J., & Links, P. S. (2012). Dialectical behavior therapy compared with general psychiatric management for borderline personality disorder: clinical outcomes and functioning over a 2-year follow-up. *Am J Psychiatry, 169,* 650-661. doi: 10.1176/appi.ajp.2012.11091416

Mehlum, L., Ramberg, M., Tørmoen, A. J., Haga, E., Diep, L. M., Stanley, B. H., . . . Grøholt, B. (2016). Dialectical Behavior Therapy Compared With Enhanced Usual Care for Adolescents With Repeated Suicidal and Self-Harming Behavior: Outcomes

Over a One-Year Follow-Up. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*, 295-300. doi: https://doi.org/10.1016/j.jaac.2016.01.005

Montgomery-Graham, S. (2016). DBT and schema therapy for borderline personality disorder: Mentalization as a common factor. *Journal of Contemporary Psychotherapy, 46*, 53-60.

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. doi: https://doi.org/10.1016/j.psychres.2017.06.045

Neacsiu, A. D., Bohus, M., & Linehan, M. M. (2014). Dialectical behavior therapy: An intervention for emotion dysregulation. In *Handbook of emotion regulation, 2nd ed.* (pp. 491-507). New York, NY, US: Guilford Press.

Nemeth, N., Matrai, P., Hegyi, P., Czeh, B., Czopf, L., Hussain, A., . . . Simon, M. (2018). Theory of mind disturbances in borderline personality disorder: A meta-analysis. *Psychiatry Res, 270*, 143-153. doi: 10.1016/j.psychres.2018.08.049

Nolte, T., Bolling, D. Z., Hudac, C. M., Fonagy, P., Mayes, L., & Pelphrey, K. A. (2013). Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Front Hum Neurosci, 7*, 816. doi: 10.3389/fnhum.2013.00816

Nolte, T., Guiney, J., Fonagy, P., Mayes, L. C., & Luyten, P. (2011). Interpersonal stress regulation and the development of anxiety disorders: an attachment-based developmental framework. *Front Behav Neurosci, 5*, 55. doi: 10.3389/fnbeh.2011.00055

Norton, K., & Hinshelwood, R. D. (1996). Severe Personality Disorder: Treatment issues and selection for inpatient psychotherapy. *British Journal of Psychiatry, 168*, 723-731.

Richman, M. J., & Unoka, Z. (2015). Mental state decoding impairment in major depression and borderline personality disorder: meta-analysis. *Br J Psychiatry, 207*, 483-489. doi: 10.1192/bjp.bp.114.152108

Rutter, D., & Tyrer, P. (2003). The value of therapeutic communities in the treatment of personality disorder: A suitable place for treatment? *Journal of Psychiatric Practice, 9*, 291-302.

Schimmel, P. (1997). Swimming against the tide? A review of the therapeutic community. *Australian and New Zealand Journal of Psychiatry, 31*, 120-127.

Sempértegui, G. A., Karreman, A., Arntz, A., & Bekker, M. H. J. (2013). Schema therapy for borderline personality disorder: A comprehensive review of its empirical foundations, effectiveness and implementation possibilities. *Clinical Psychology Review, 33*, 426-447. doi: https://doi.org/10.1016/j.cpr.2012.11.006

Skodol, A. E. (2012). Personality disorders in DSM-5. *Annu Rev Clin Psychol, 8*, 317-344. doi: 10.1146/annurev-clinpsy-032511-143131

Smits, M. L., Feenstra, D. J., Bales, D. L., Blankers, M., Dekker, J. J. M., Lucas, Z., . . . Luyten, P. (2020). Day hospital versus intensive outpatient mentalization-based treatment: 3-year follow-up of patients treated for borderline personality disorder in a multicentre randomized clinical trial. *Psychol Med*, 1-11. doi: 10.1017/S0033291720002123

Sroufe, L. A. (1983). Infant caregiver attachment and patterns of adaptation in preschool: The roots of maladaption and competence. In M.Perimutter (Ed.), *Minnesota symposium in child psychology* (pp. 41-81).

Sroufe, L. A., & Fleeson, J. (1986). Attachment and the construction of relationships. In W.Hartup & Z.Rubin (Eds.), *Relationships and Development* (pp. 51-71). Hillsdale, NJ: Erlbaum.

Steele, H. (2003). Unrelenting catastrophic trauma within the family: when every secure base is abusive. *Attach Hum Dev, 5*, 353-366; discussion 409-314. doi: 10.1080/14616730310001633438

Steele, H., Steele, M., & Murphy, A. (2009). Use of the Adult Attachment Interview to measure process and change in psychotherapy. *Psychother Res, 19*, 633-643. doi: 10.1080/10503300802609698

Tyrer, P., Crawford, M., & Mulder, R. (2011). Reclassifying personality disorders. *Lancet, 377*, 1814-1815. doi: 10.1016/S0140-6736(10)61926-5

Weissman, M. M. (1975). The assessment of social adjustment. *Archives of General Psychiatry, 32*, 357-365.

Zeegers, M. A. J., Colonnesi, C., Stams, G. J. M., & Meins, E. (2017). Mind matters: A meta-analysis on parental mentalization and sensitivity as predictors of infant-parent attachment. *Psychol Bull, 143*, 1245-1272. doi: 10.1037/bul0000114

Table 1.

Demographic, diagnostic and risk profile features of the step-down (RT-CBP), residential (RT), and general psychiatric (GP) samples

	RT-CBP (*n* = 32)		RT (*n* = 39)		GP (*n* = 40)		Test of significance
	n	*%*	*n*	*%*	*n*	*%*	
Female	26	81.3	29	74.4	25	62.5	$\chi^2_{(2)}$=3.26, *p*=.196
White (any)	32	100	39	100	40	100	N/A
Single	21	65.5	29	74.4	14	35.0	$\chi^2_{(2)}$=13.70, *p*=.001
College education	21	65.6	29	74.4	9	22.5	$\chi^2_{(2)}$=24.14, *p*=.000
Any sexual abuse	18	56.3	17	43.6	25	62.5	$\chi^2_{(2)}$=2.93, *p*=.231
Physical abuse	12	37.5	17	43.6	20	50.0	$\chi^2_{(2)}$=1.13. *p*=.567
Mood Disorder	15	46.9	18	46.2	26	65.0	$\chi^2_{(2)}$=3.53. *p*=.171
Anxiety Disorder	18	56.3	20	51.3	23	57.5	$\chi^2_{(2)}$=1.13. *p*=.339
Eating Disorder	7	21.9	7	17.9	3	7.5	$\chi^2_{(2)}$=3.15. *p*=.207
Substance Misuse	4	12.5	5	12.8	8	20.0	$\chi^2_{(2)}$=1.06. *p*=.589
Avoidant PD	17	53.1	19	48.7	25	62.5	$\chi^2_{(2)}$=1.58, *p*=.455
Paranoid PD	10	33.3	22	56.4	18	45.0	$\chi^2_{(2)}$=4.50, *p*=.106
Borderline PD	20	62.5	28	71.8	26	65.0	$\chi^2_{(2)}$=0.76, *p*=.683
Schizotypal PD	6	18.8	6	16.4	10	25.0	$\chi^2_{(2)}$=0.18, *p*=.554
Narcissistic PD	3	9.4	4	10.3	2	5.0	$\chi^2_{(2)}$=0.76, *p*=.683
Obsessive-Compulsive PD	6	18.8	11	28.2	10	25.0	$\chi^2_{(2)}$=0.87, *p*=.648
Dependent PD	14	43.8	11	28.2	17	42.5	$\chi^2_{(2)}$=2.38, *p*=.304
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	

Age	33.19	8.93	31.18	7.62	34.55	7.84	$F_{(2)}$=1.73, p=.183
PD diagnoses	3.28	1.73	3.41	1.50	3.35	1.76	$F_{(2)}$=.05, p=.949
PD dimensional [a]	32.97	10.81	36.18	10.10	27.75	11.46	$F_{(2)}$=6.11, p=.003
GSI intake	1.99	0.83	1.97	0.57	1.88	.75	$F_{(2)}$=.20, p=.822
SAS intake	2.62	.54	2.69	0.44	2.68	.37	$F_{(2)}$=.25, p=.776
GAS intake	47.81	6.43	47.31	7.07	45.08	6.51	$F_{(2)}$=1.78, p=.173

Note. PD = personality disorder; GSI = General Severity Index; SAS = Social Adjustment Scale; GAS = Global Assessment Scale. [a] Number of positive PD criteria met.

Table 2.

Means, standard deviations (SD), estimated marginal means (EMM), and between-group mean differences and effect sizes for Symptom Checklist-90-R global severity index (GSI), Social Adjustment Scale (SAS), Global Adjustment Scale (GAS), and overall standardized outcome scores in the three PD samples.

	RT-CBP (*n* = 32)			RT (*n* = 39)			GP (*n* = 40)					
	Mean	*SD*	EMM	Mean	*SD*	EMM	Mean	*SD*	EMM	Mean difference	95% CI	Hedges' *g*
GSI										RT-CBP – GP = –0.74	–1.27, –0.22	RT-CBP vs. GP ***g* = .89**
Intake	1.99	.83	1.98	1.97	.60	1.89	1.89	.75	1.87	RT – GP = –0.46	–1.01, –0.08	RT vs. GP *g* = .46
24 months	1.21	1.03	1.19	1.55	.70	1.47	1.84	.82	1.93	RT-CBP – RT = –0.28	–0.75, 0.20	RT-CBP vs. RT *g* = .46
SAS										RT-CBP – GP = –0.67	–0.96, –0.38	RT-CBP vs. GP ***g* = .98**
Intake	2.62	.54	2.63	2.69	.44	2.67	2.68	.37	2.71	RT – GP = –0.46	–0.76, –0.15	RT vs. GP ***g* = .68**
24 months	2.01	.55	2.02	2.25	.41	2.23	2.67	.41	2.69	RT-CBP – RT = –0.21	–0.48, 0.05	RT-CBP vs. RT *g* = .29
GAS										RT-CBP – GP = 12.93	4.96, 20.88	RT-CBP vs. GP ***g* = .73**
Intake	47.81	6.43	47.90	47.31	7.07	47.68	45.08	6.51	44.64	RT – GP = 9.31	1.09, 17.53	RT vs. GP *g* = .46
24 months	62.16	15.89	62.40	57.72	12.54	58.78	50.70	8.37	49.47	RT-CBP – RT = 3.61	–3.56, 10.79	RT-CBP vs. RT *g* = .27
Overall outcome										RT-CBP – GP = –1.04	–1.55, –0.52	RT-CBP vs. GP ***g* = .87**
Intake	-.07	.79	-.08	-.00	.68	-.08	.06	.78	.14	RT – GP = –0.70	–1.23, –0.17	RT vs. GP ***g* = .51**
24 months	-.48	1.04	-.16	-.09	.76	-.16	.47	.65	.55	RT-CBP – RT = –0.33	–0.80, 0.13	RT-CBP vs. RT *g* = .35

Note. Effect size (*g*) in bold indicates clinical superiority margin has been exceeded

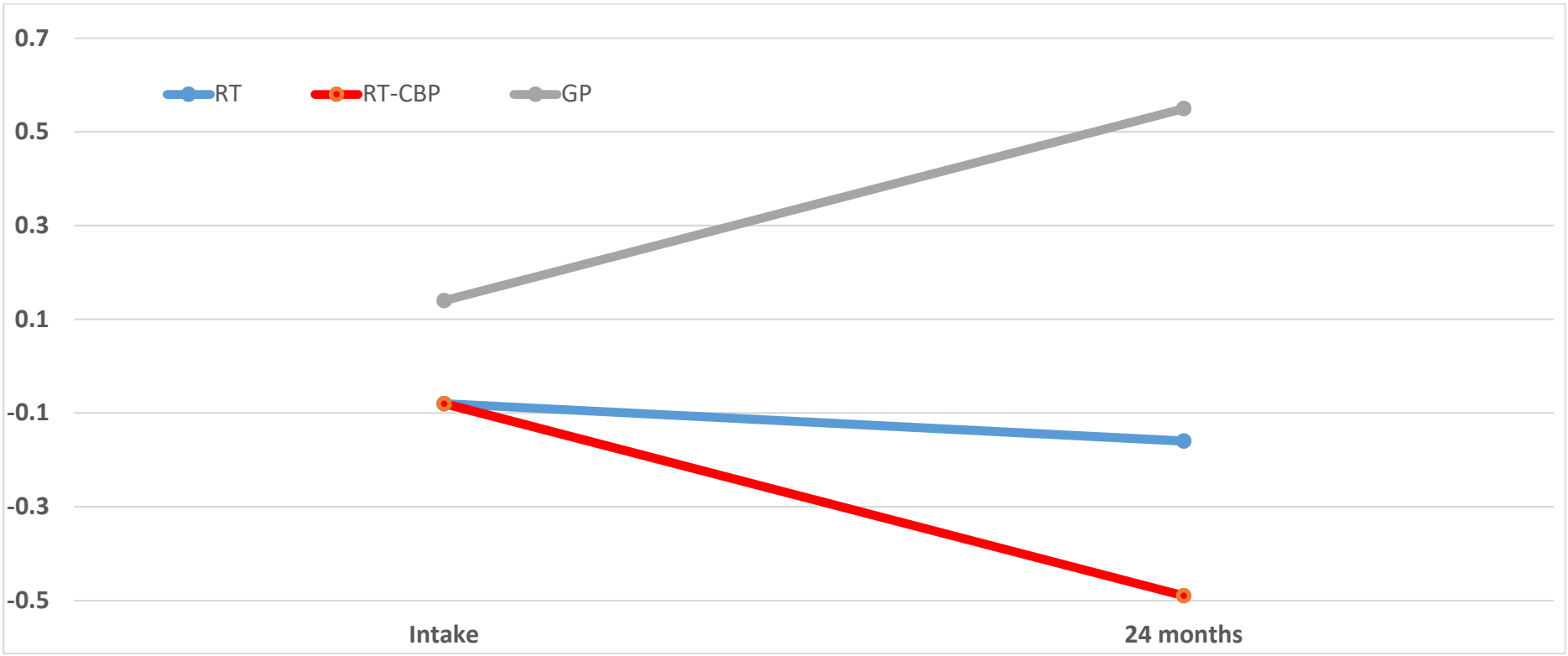

Figure 1. Significant change over the 24-month period in overall clinical outcome in the three personality disorder (PD) groups (Group × Time interaction $F_{(1, 105)} = 6.85$, $p = .002$). RT-CBP is the only PD group to show significant simple time changes, $F_{(1, 105)} = 7.68$, $p = .007$).

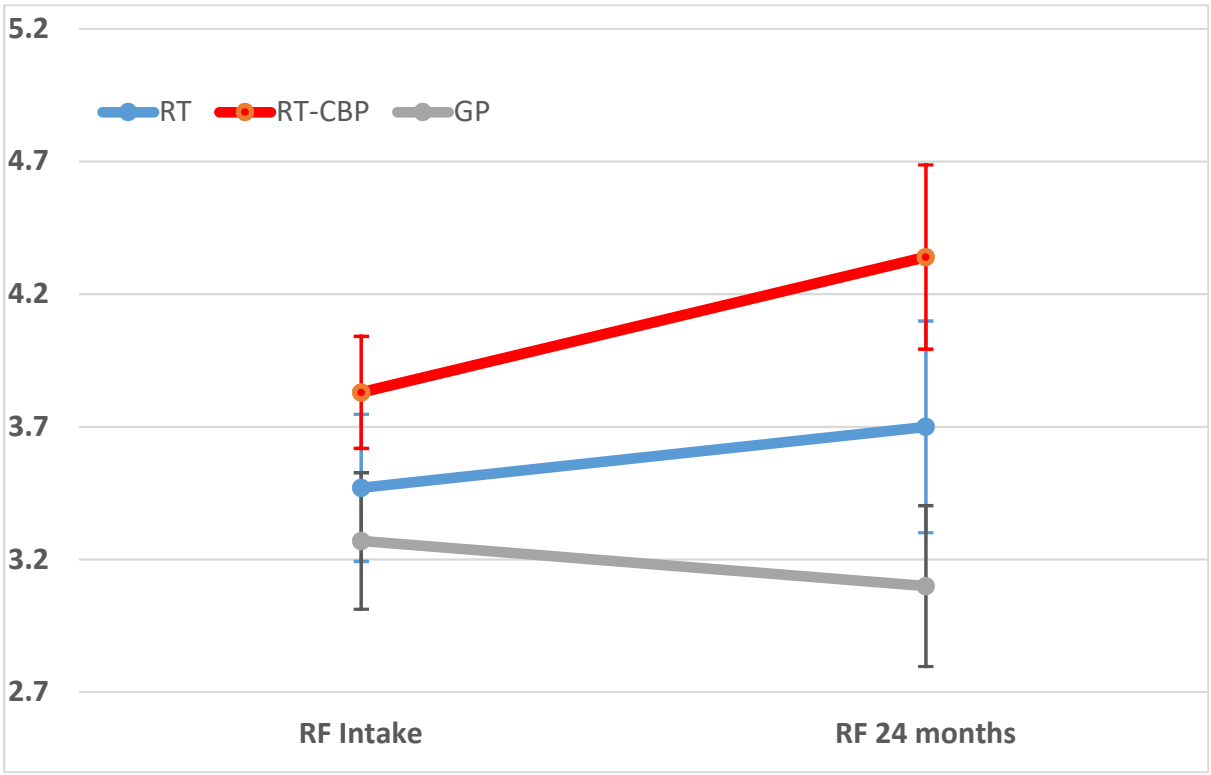

Figure 2. Change in reflective function (RF) scores in the three personality disorder (PD) groups between intake and 24-month follow-up shows a significant Group × Time interaction, $F_{(1, 105)}$ =4.08, p = .020, with RT-CBP being the only PD group to show significant simple time changes (t = 2.81, p = 0.006). Error bars represent the *SE*.

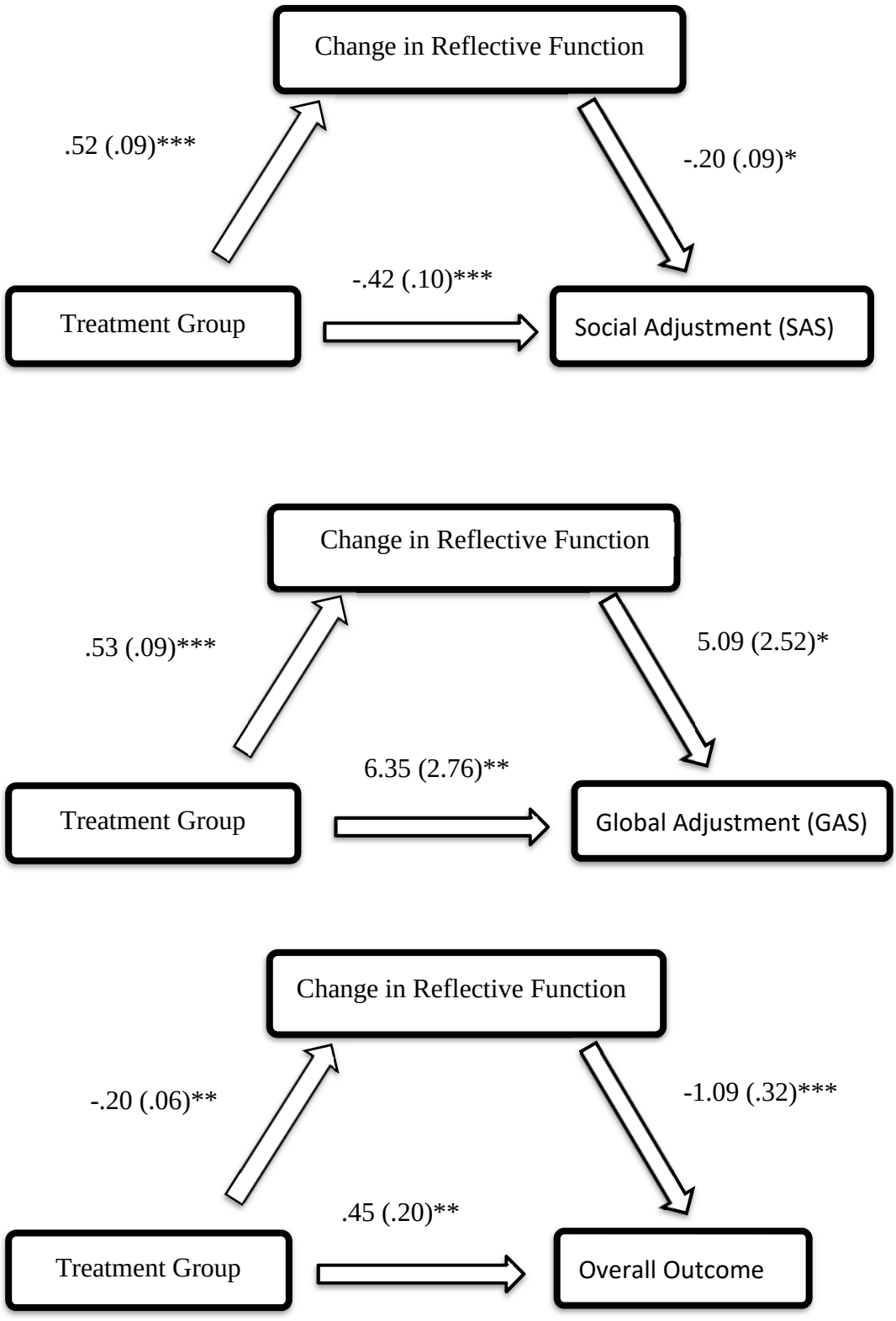

Figure 3. Mediation analyses between treatment groups (specialist vs. non-specialist), change in reflective function, and clinical outcomes at 24-month follow-up.

* p = .000. ** p = .028. *** p = .001.