

**Multimodal assessments of therapist characteristics are largely unrelated to patient outcomes: A preregistered analysis**

Simon B. Goldberg<sup>1,2</sup>, Scott A. Baldwin<sup>3</sup>, Anthony W. P. Flynn<sup>1</sup>, Robbie Babins-Wagner<sup>4</sup>, Derek D. Caperton<sup>4</sup>, Camille Y. Williams<sup>1,2</sup>, Emily H. Hamm<sup>1</sup>, Sin U Lam<sup>1,2</sup>, Jessica K. Cozart<sup>1</sup>, Geetanjali Deole<sup>1</sup>, Nili Solomonov<sup>5</sup>, S. Mark Kopta<sup>6</sup>, Timothy Anderson<sup>7</sup>, Bruce E. Wampold<sup>1,8</sup>, Jesse Owen<sup>9</sup>

<sup>1</sup>Department of Counseling Psychology, University of Wisconsin – Madison, Madison, WI, USA

<sup>2</sup>Center for Healthy Minds, University of Wisconsin – Madison, Madison, WI, USA

<sup>3</sup>Department of Psychology, Brigham Young University, Provo, UT, USA

<sup>4</sup>Calgary Counselling Centre, Calgary, Alberta, Canada

<sup>5</sup>Weill Cornell Medicine, New York, NY, USA

<sup>6</sup>Department of Psychology, University of Evansville, Evansville, IN, USA

<sup>7</sup>Department of Psychology, Ohio University, Athens, OH, USA

<sup>8</sup>Modum Bad Psychiatric Center, Vikersund, Norway

<sup>9</sup>Department of Counseling Psychology, University of Denver, Denver, CO, USA

**Author Note**

Simon B. Goldberg:  <https://orcid.org/0000-0002-6888-0126>

Corresponding author: Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin – Madison. Email: [sbgoldberg@wisc.edu](mailto:sbgoldberg@wisc.edu)

## PREDICTING THE THERAPIST EFFECT

### **Abstract**

While it is known that therapists vary in effectiveness, it is unclear what therapist-level characteristics predict this variation. We conducted a large-scale, preregistered study ( $n = 97$  therapists from the United States and Canada,  $n = 6,152$  patients) examining a multimodal set of 38 therapist-level predictors that have been empirically or theoretically linked with patient outcomes. We examined associations with pre-post change and rate of change in psychological distress, and likelihood of attending  $>1$  treatment session. We largely did not find associations between therapist-level characteristics and patient outcomes. Most predictors failed to replicate across sensitivity analyses and/or were non-significant following  $p$ -value correction. The most robust evidence suggested that interpersonal capacities assessed via a performance task are associated with likelihood of attending  $>1$  treatment session. A key limitation of the study is small therapist effects which may have reduced statistical power. Empirically, it remains uncertain what qualities characterize highly effective therapists.

*Keywords:* psychotherapy; expertise; therapist effects; multimodal assessment; behavioral measures

## PREDICTING THE THERAPIST EFFECT

The idea that some therapists are better than others dates back to the 1930s (e.g., Rosenzweig, 1936). Therapist differences in outcome have been studied empirically for over 40 years (Luborsky et al., 1985), with a rapid increase in work in this area in the past 20 years (Wampold & Owen, 2021). It is now well established that therapists indeed differ systematically in their patients' outcomes – that is, some therapists are consistently more effective than other therapists (Wampold & Owen, 2021). A meta-analysis of 20 studies that explicitly focused on the therapist effect (Johns et al., 2019) estimated that 5% of variance in treatment outcome is attributable to therapists (i.e., therapist-level intraclass correlation [ICC] = .050). This estimate is identical to that provided by an earlier meta-analysis of 46 studies that allowed the inclusion of studies that reported therapist-level ICCs, but were not specifically designed to study the therapist effect (Baldwin & Imel, 2013). While 5% would appear to be a relatively modest proportion of variance explained, it is similar in magnitude or larger than that associated with other primary contributors to treatment outcome (e.g., differences between treatments, therapists' adherence and competence, therapeutic alliance; Power et al., 2022; Wampold & Owen, 2021).

Whether or not therapists can improve their outcomes is an area of active debate (Wampold & Owen, 2021). Some have argued that the context of therapy is a difficult environment in which to learn, due to its unpredictable nature and therapists' lack of access to feedback necessary for improving performance (Tracey et al., 2014). Thus, differences between therapists may be due more to innate qualities (e.g., interpersonal skill; Anderson et al., 2009) than training or experience (Goldberg, Rousmaniere et al., 2016; Minami et al., 2009). Drawing from the science of expertise (e.g., Ericsson & Charness, 1994), there has been increasing emphasis on the use of deliberate practice (e.g., rehearsal of specific therapeutic skills paired with outcome monitoring and feedback; Miller et al., 2013; Rousmaniere, 2017) as a means for

## PREDICTING THE THERAPIST EFFECT

improving therapists' outcomes. Although empirical data are limited, there is evidence supporting both the notion that therapists do not typically improve (Germer et al., 2022; Goldberg, Rousmaniere, et al., 2016) and that they can improve in an adequately enriched learning context (e.g., Goldberg, Babins-Wagner, et al., 2016). Regardless of whether therapists' skills are inborn or trainable, there is general agreement that therapists vary in effectiveness.

Given therapists differ in their patients' outcomes, a key question naturally follows: what are the qualities of effective therapists? The answer to this question may lie in one (or both) of two types of qualities: (a) therapist actions during psychotherapy with a given patient (e.g., therapists' use of specific skills such as Socratic questioning or empathic reflections; Hill & Norcross, 2023), which can be measured as in-session behavior, and (b) therapist characteristics (e.g., training, relational skills; Heinonen & Nissen-Lie, 2019), measured outside of psychotherapy. Therapists' in-session behavior is certainly worthy of investigation. However, in-session behavior is a somewhat limited explanation for therapist effects given that therapists actions within psychotherapy are dependent to a large degree on the particular patient being treated (Boswell et al., 2013; Wampold & Imel, 2015). Assessing what the therapist brings to psychotherapy, including various characteristics (e.g., age, experience, personality) as well as skills (e.g., ability to express empathy with difficult patients) has self-evident implications for training, professional development, and personnel selection. Here we aim examine the association between therapist qualities assessed outside of therapy and treatment outcomes.

The most recent systematic review of the link between therapist characteristics and treatment outcomes (Heinonen & Nissen-Lie, 2019) identified 31 studies measuring professional characteristics (e.g., subjective efficacy, professional self-doubt), socio-emotional characteristics such as performance-based measures of interpersonal skills (e.g., Facilitative Interpersonal Skills

## PREDICTING THE THERAPIST EFFECT

[FIS] task; Anderson et al., 2009), multicultural counseling competencies, and personal characteristics (e.g., attachment style). Therapists' self-report of some constructs (e.g., professional self-doubt, attachment style, time spent engaging in deliberate practice) have shown links with outcomes. However, these constructs have largely been examined in only one or two studies that included relatively small numbers of patients and/or therapists (Heinonen & Nissen-Lie, 2019). Consistent with other reviews (e.g., Lingiardi et al., 2018; Wampold & Owen, 2021), the strongest evidence in Heinonen and Nissen-Lie's (2019) review emerged for therapists' interpersonal skills assessed via performance tasks such as the FIS. This task involves therapists recording responses to vignettes of patients (played by actors) displaying interpersonally challenging behavior in psychotherapy. Responses are then coded by raters.

It is difficult to draw firm conclusions from this literature. Dozens of candidate therapist characteristics have been examined across studies, with most studies reporting only a small number of potential characteristics (Heinonen & Nissen-Lie, 2019). Further, some theoretically important therapist characteristics (e.g., multicultural orientation; Davis et al., 2018) have not been thoroughly investigated. Thus, there is a need for large-scale evaluations of a range of therapist characteristics. Given evidence that non-self-report measures of therapist characteristics have often been most consistently predictive of patient outcomes (e.g., FIS skills), multimodal measurement may be particularly worthwhile. By multimodal evaluation, we are referring to assessment of therapist characteristics using multiple modes of data collection (i.e., self-report, performance tasks such as the FIS, behavioral measures).

### **Current Study**

We aimed to evaluate a diverse set of therapist characteristics that have been empirically or are theoretically linked with patient outcomes (Heinonen & Nissen-Lie, 2019). We focused on

## PREDICTING THE THERAPIST EFFECT

four categories of predictors: interpersonal skills and related domains, multicultural orientation and related domains, professional characteristics (e.g., use of deliberate practice, years of experience), and attachment-related measures. To evaluate these constructs, we gathered multimodal data through self-report, performance tasks, and computerized behavioral measures. Given a lack of certainty regarding characteristics most likely to predict patient outcomes, we intentionally gathered a large set of candidate variables. We conducted a preregistered set of analyses in a sample of therapists and patients drawn from clinics in the United States and Canada. We report results with and without false-discovery rate (FDR; Benjamini & Hochberg; 1995) *p*-value correction.

### **Transparency and Openness**

#### **Preregistration**

Study hypotheses and analytic plan were preregistered. This included preregistrations focused on interpersonal skills and related constructs (<https://osf.io/pfkdj/overview>), professional training and related constructs (<https://osf.io/ecv2z/overview>), multicultural orientation and related constructs (<https://osf.io/wp6kv/overview>), and hypothesized mediation of therapist effects via therapist attachment style (<https://osf.io/7xn48/overview>). Preregistrations were submitted when a portion of the therapist data was collected, but prior to any analyses. Linked patient data had not been received at the time of preregistration. We made ten deviations from our preregistration which are described in Method.

#### **Data, Materials, Code, and Online Resources**

Due to IRB restrictions, we are not able to share study data. However, analysis scripts are publicly available (<https://osf.io/am4x9/files/osfstorage>). All study materials except the FIS and Multicultural Orientation Task are publicly or commercially available (e.g., through Inquisit).

## PREDICTING THE THERAPIST EFFECT

### **Reporting**

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

### **Ethical Approval**

This research was approved by the University of Wisconsin – Madison Institutional Review Board ([IRB], 2019-0091).

## **Method**

### **Participants and Procedure**

Therapists were recruited from clinics in the United States and Canada that use routine outcome monitoring systems (e.g., Outcome Questionnaire – 45 [OQ-45; Lambert et al., 2004], Behavioral Health Monitor [BHM; Kopta & Lowry, 2002]). Study staff sent email invitations to directors of clinics known to be using the OQ-45 or BHM-20. Invitations were forwarded to clinical staff (licensed staff and clinical trainees) who were invited to complete an approximately 2-hour assessment battery. All assessments were delivered online. Questionnaire measures were delivered through Qualtrics. Performance tasks (FIS and Multicultural Orientation [MCO] task; Stewart et al., 2024) were completed using the Skillsetter platform ([www.Skillsetter.com](http://www.Skillsetter.com)). Computerized behavioral measures were completed using the Inquisit platform ([www.millisecond.com/](http://www.millisecond.com/)) and a web-app built by our team (for the Empathic Accuracy Task). Participants were paid \$200 for completing the assessment battery. Patient data provided by participating clinics were linked with therapist data. Therapists completed assessments between November 2019 and January 2021. Patients' first session occurred between November 2010 and October 2024. However, data used in all but one sensitivity analysis were restricted to patients whose first session occurred on or after the date the therapist assessment occurred.

## PREDICTING THE THERAPIST EFFECT

We included an estimated power table in our preregistration and aimed to collect data from between 100 and 200 therapists anticipating that each would have 20 patients. Using variance estimates from a similar naturalistic psychotherapy data set (Goldberg et al., 2023), we estimated that 100 therapists would provide approximately 80% power to detect a difference in patient outcomes associated with a given therapist-level predictor in standard deviation units (i.e., Cohen's  $d$ ) of  $d \geq 0.40$ . We estimated a sample of 200 therapists would provide  $\geq 80\%$  power to detect a difference of  $d \geq 0.30$ . We anticipated that we would likely be underpowered to detect small effects ( $ds \leq 0.20$ ).

### ***Therapists***

A total of 97 therapists provided data for one or more predictors and could be matched with patient data. Therapists were recruited from 13 clinics in the United States ( $n = 10$ ) and Canada ( $n = 3$ ). Most therapist (71.1%) were recruited from a large community mental health agency in Canada. The remaining therapists were recruited from university counseling centers, with 10 based in the United States and two based in Canada. Therapists were on average 35.77 years old (standard deviation [ $SD$ ] = 9.99); 81.4% identified as women, 15.4% as men (one identifying as a transman), 3.1% did not report a gender; 69.1% identified as non-Latinx White, 6.2% as Black, 20.6% as Asian, 1.0% as Latinx, 1.0% as "other" race/ethnicity, and 2.1% did not report a race/ethnicity. Therapists had an average of 63.42 ( $SD = 79.62$ , median = 27) patients in the data set.

### ***Patients***

A total of 6,152 patients were seen by the 97 therapists who provided therapist-level data. Patients attended individual psychotherapy for a total of 28,843 sessions. Data were analyzed from the first episode of care with gaps >120 days used to demarcate distinct episodes (Goldberg

## PREDICTING THE THERAPIST EFFECT

et al., 2016). Patients' initial session occurred on average 1.29 years ( $SD = 1.17$ ) after the therapists completed the assessment battery.

Most patients (86.8%) were from the community mental health agency in Canada, with the remaining patients coming from university counseling centers in the United States (10.3%) and Canada (3.0%). Patients were on average 35.05 years old ( $SD = 12.63$ ); 61.2% identified as women, 33.3% as men, 5.5% did not report a gender; 53.9% identified as non-Latinx White, 4.1% as "other" race/ethnicity, 9.0% as Asian, 2.3% as Black, 2.4% as Native American / Aboriginal / First Nations / Alaskan Native, 2.2% as Latinx, 1.3% as Middle Eastern, 0.1% as multiracial, and 24.8% did not report a race/ethnicity. Detailed clinical information was not available at the patient level. However, common reasons for seeking treatment at the Canadian community mental health agency included family/marital problems, depression, anxiety, stress, and eating disorders (Rousmaniere et al., 2016) and common reasons for seeking treatment at college counseling centers include depression, anxiety, stress, as well as family and academic concerns (Xiao et al., 2017).

### **Measures**

#### ***Patient Outcome Measures***

Two different outcome measures were used across the 13 clinics. The community mental health agency used the OQ-45 (Lambert et al., 2004). This is a 45-item self-report measure designed to index psychological distress in the context of psychotherapy. Scores  $\geq 63$  are considered clinically elevated (Lambert et al., 2004). Internal consistency for the total score was high ( $\alpha = .94$ ). The university counseling centers used the BHM-20 (Kopta & Lowry, 2002). This is a 20-item self-report measure that was also designed to index psychological distress in the context of psychotherapy. Scores  $\leq 2.78$  on the total score (i.e., General Mental Health scale)

## PREDICTING THE THERAPIST EFFECT

are considered clinically elevated (Kopta et al., 2015). Internal consistency for the total score was high ( $\alpha = .91$ ).

### *Therapist-Level Characteristics*

Therapists completed an assessment battery that included self-report questionnaires, performance tasks, and computerized behavioral tasks. Tasks were grouped in four primary content domains: FIS-related measures, multicultural-related measures, professional characteristics, and attachment-related measures (Table 1). Given the FIS has shown the most consistent associations with patient outcomes, many measures were selected to assess constructs broadly related to the eight FIS domains (described below). Given the growing literature highlighting the importance of multicultural capacities in therapy, we also included the MCO task (described below) and related measures. We assessed several professional characteristics that have been empirically or are theoretically linked with patient outcomes (Heinonen & Nissen-Lie, 2019). Finally, we assessed attachment style and a related construct (childhood trauma). As noted above, we intentionally gathered a large set of candidate characteristics based on the view that a more comprehensive (rather than narrow) evaluation would be most valuable at the current stage of work in this area.

#### **Performance Tasks. *Facilitative Interpersonal Skills task (FIS; Anderson et al., 2009).***

The FIS task involves responding to brief video vignettes of actors depicting interpersonally challenging moments in psychotherapy. To reduce participant burden, the current study used four vignettes. Participants are asked to record themselves responding as if they were the patients' therapist. Recorded responses are rated on eight standardized FIS items (Anderson et al., 2009): verbal fluency; hope and positive expectations; persuasiveness; emotional expression; warmth, acceptance, and understanding; empathy; alliance-bond capacity; and alliance rupture-repair

## PREDICTING THE THERAPIST EFFECT

responsiveness. Ratings range from 1 (skill deficit) to 5 (optimal presence of skill). In the current study, three raters who were graduate students in counseling psychology received 17 hours of training in FIS ratings over the course of 10 training sessions by licensed psychologists who were former students of the FIS developer. Initial training occurred using non-study videos until the raters achieved adequate inter-rater reliability ( $ICC_{2,3} \geq .60$ ; Cicchetti, 1994). Each video was rated by all three raters. Inter-rater reliability across all items was good ( $ICC = .67$ ), although it ranged from fair to excellent across FIS domains ( $ICCs = .75, .58, .62, .66, .70, .68, .68, .60$ , for verbal fluency; hope and positive expectations; persuasiveness; emotional expression; warmth, acceptance, and understanding; empathy; alliance-bond capacity; and alliance rupture-repair responsiveness). Item scores were computed by averaging across the three raters. The FIS total score was computed by averaging across the eight domains ( $\alpha = .94$ ).

***Multicultural Orientation Task (MCO; Stewart et al., 2024).*** The MCO also involves responding to brief video vignettes. The current study used four vignettes. Each vignette depicts a patient sharing content with culturally relevant elements (e.g., a Black man experiencing discrimination at work, a gay man struggling with dating). Raters evaluate the degree to which the therapists' response demonstrated cultural comfort (i.e., comfort having a conversation about culture, three items), cultural humility (i.e., non-superior stance toward patients, three items), and cultural opportunity (i.e., seizing the moment to ask about cultural identities, one item). A final item assesses the degree to which a therapist's response was generally therapeutic. Items are rated from 1 (poor performance) to 5 (high performance). A second set of three raters who were graduate students in counseling psychology received 10 hours of training in MCO ratings over the course of eight training sessions by one of the MCO developers. Training occurred using non-study videos until the raters achieved adequate inter-rater reliability ( $ICC_{2,3} \geq .60$ ;

## PREDICTING THE THERAPIST EFFECT

Cicchetti, 1994). Each video was rated by all three raters. Inter-rater reliability across all items was excellent ( $ICCs = .81$ ), although it ranged from fair to excellent across MCO domains ( $ICCs = .55, .62, .91, \text{ and } .64$ , for comfort, humility, opportunity, and generally therapeutic, respectively). Item scores were computed by averaging across the three raters. Cultural comfort and cultural opportunity items were averaged for each scale ( $\alpha s = .91 \text{ and } .89$ , respectively).

**Computerized Behavioral Measures and Self-Report Questionnaires.** A description of these measures is included in Supplemental Materials Table 1.

### Data Analysis

We made ten deviations from the preregistered plan. First, we restricted our primary analyses to patients whose course of therapy began on or after the day therapists completed their assessments. This was done based on the notion that these characteristics may not be stable over time and that it may be problematic to predict patient outcomes that occurred prior to this assessment. Second, we report results with an FDR-correction applied. Third, we do not report analyses examining FIS as a mediator between attachment security and patient outcomes as this was outside our focus on therapist-level predictors of patient outcomes. Fourth, we do not report analyses examining a face rating task as a measure of racial bias, therapist grade point average, or therapist graduate record exam scores. The face rating task could not be used due to a technical error in which two versions of the task were used over the course of the study. Grade point average and graduate record exam scores were reported in inconsistent ways and infrequently such that they were deemed unusable. Fifth, we included a dummy-coded variable indicating whether data was drawn from a clinic using the BHM-20 (or the OQ-45) based on exploratory data analyses demonstrating that changes on the OQ-45 were substantially larger than changes on the BHM-20 on average ( $d s = -0.54 \text{ and } -0.36$ , for OQ-45 and BHM-20,

## PREDICTING THE THERAPIST EFFECT

respectively,  $p < .001$ ). Sixth, we added three sensitivity analyses that examined results restricted to the OQ-45 and BHM-20 and a model covarying the time between a patient's first session and when the therapist completed the assessment battery. Seventh, to help with model convergence, we z-scored all therapist-level predictor variables. Eighth, we modeled clinic as a random intercept. Ninth, we modified the random effects for models that would not converge. Tenth, we conducted a set of analyses using all patient data, regardless of whether a patient's first session occurred prior to their therapist's assessment. This sensitivity analysis included 109 therapists who saw 15,550 patients.

In total, we examined 38 therapist-level predictors (Table 1) of three patient-level outcomes: pre-post change in psychological distress (on OQ-45 or BHM-20), rate of change in psychological distress, and early termination (i.e., whether patients completed >1 treatment session)<sup>1</sup>. Separate sets of models were conducted for each predictor. Data were analyzed using the 'lme4 package (Bates et al., 2015) in R (R Core Team, 2024). Maximum likelihood estimation which is robust to data missing at random (Graham, 2009) was used across models.

For pre-post change, we used two-level multilevel models (i.e., patients nested within therapists) with therapist-level characteristics entered as Level 2 predictors:

$$Y_{ij} = \beta_{00} + \beta_{01}(BHM - 20) + \beta_{02}(Predictor) + [U_{0j} + e_{ij}], \quad (\text{Equation 1})$$

where  $Y_{ij}$  reflects the pre-post changes in psychological distress on the OQ-45 or BHM-20 in standardized Cohen's  $d$  units for patient  $i$  seen by therapist  $j$ . This value was calculated by subtracting psychological distress at the first session from psychological distress at the final session and dividing the difference score by the first session  $SD$ . In order to allow inclusion in

---

<sup>1</sup> Of note, a wide variety of operationalizations of "early termination" have appeared in the psychotherapy literature (Smith & Greenberg, 2012). We used attending only a single session, one of the four operationalizations studied by Hatchett and Park (2003), as a definition of early termination that could be applied reliably across samples. However, results may have differed had another method been used to define early termination.

## PREDICTING THE THERAPIST EFFECT

the same model, scores on the BHM-20 (where higher scores reflect lower levels of psychological distress; Kopta et al., 2002) were reversed prior to calculating  $d$ . Thus, smaller (e.g., more negative)  $d$ s reflect greater reductions in psychological distress. The fixed intercept ( $\beta_{00}$ ) reflects the overall mean change in psychological distress and the fixed slope for the BHM-20 ( $\beta_{01}$ ) reflects the overall effect of whether the BHM-20 was used. The fixed slope for the therapist-level predictors ( $\beta_{02}$ ) reflects the relationship between each of the 38 predictors with pre-post change in psychological distress. Negative coefficients for  $\beta_{02}$  indicate that higher levels of the predictor (e.g., FIS total score) was associated with larger (i.e., more negative) pre-post reductions in psychological distress. Random effects (shown within brackets) included a random intercept at the therapist level ( $U_{0j}$ ) and an error term ( $e_{ij}$ ). Including a random intercept at the clinic level, which would have produced a three-level model (i.e., patients nested within therapists nested within clinics), resulted in a singular fit due to insufficient variance at the clinic level. Thus, a two-level model was used. Consistent with our preregistration, the pre-post model and all other models examining the Multitheoretical List of Therapeutic Interventions–30 (MULTI-30; Solomonov et al., 2019) included a linear and quadratic version of the entropy score. We examined interactions with the quadratic version. An empty model (i.e., Equation 1 with the therapist characteristic predictor [i.e.,  $\beta_{02}$ ] omitted) was used to calculate therapist-level ICCs. Specifically, we divided the variance in pre-post change at the therapist level by the total variance.

For rate of change, we used four-level multilevel models (i.e., sessions nested within patients nested within therapists nested within clinics). To determine how best to model rate of change, we compared models with linear only; linear and quadratic; linear, quadratic, and cubic; and log-linear session terms. The model with log-linear session number showed superior fit

## PREDICTING THE THERAPIST EFFECT

(Bayesian Information Criterion [ $BIC$ ] = 54114.25, 53573.07, 52862.28, 52307.04; for linear only; linear and quadratic; linear, quadratic, and cubic; and log-linear respectively). Adding a random slope for the log-linear session number coefficient to the log-linear model further improved the fit ( $BIC$  = 50559.04,  $\chi^2$  [4] = 1786.60,  $p < .001$ ). Therapist-level characteristics were entered as Level 3 predictors in interaction with the log session term:

$$Y_{ijkl} = \beta_{0000} + \beta_{0001}(BHM - 20) + \beta_{1000}(\text{Log Session \#}) + \beta_{0010}(\text{Predictor}) + \beta_{0020}(\text{Log Session \#} * \text{Predictor}) + [U_{0jkl} + U_{00kl} + U_{000l} + U_{1jkl} + U_{01kl} + e_{ijkl}],$$

(Equation 2)

where  $Y_{ijkl}$  reflects level of psychological distress on the OQ-45 or BHM-20 at session  $i$  for patient  $j$  seen by therapist  $k$  in clinic  $l$ . In order to allow inclusion in the same model, scores were standardized (i.e., z-scored) by subtracting the grand mean score (on OQ-45 or BHM-20) at the first session across the full sample from a given patient's score and dividing by the sample  $SD$  at the first session. As for the pre-post models, BHM-20 scores were reverse scored so that lower scores reflect lower psychological distress. Psychological distress scores were predicted by a fixed intercept ( $\beta_{0000}$ ), a fixed slope for the BHM-20 ( $\beta_{0001}$ ), a fixed slope for a natural log-transformed session number ( $\beta_{1000}$ ), and fixed slope for the therapist-level predictors ( $\beta_{0010}$ ). The fixed slope for the interaction between the log session number term and the therapist-level predictors ( $\beta_{0020}$ ) is the coefficient of interest for evaluating whether therapist-level predictors moderate rate of change in psychological distress. The model included random intercepts at the patient level ( $U_{0jkl}$ ), the therapist level ( $U_{00kl}$ ), the clinic level ( $U_{000l}$ ), and an error term ( $e_{ijkl}$ ). The model also included a random slope that allowed the relationship between log session number and psychological distress to vary across patients ( $U_{1jkl}$ ) and therapists ( $U_{01kl}$ ).

## PREDICTING THE THERAPIST EFFECT

Multilevel logistic regression was used to examine therapist-level predictors of whether a given patient attended >1 treatment session. This model was equivalent to Equation 1, except a dichotomous variable (coded as 1 when a given patient attended >1 treatment session and 0 when a given patient did not attend >1 treatment session) was entered as the dependent variable and a random intercept was included at the clinic level (i.e., producing a three-level model with patients nested within therapists nested within clinics). A logit link function was used to model the dichotomous outcome. Positive coefficients for the therapist-level predictors reflect a greater likelihood that a therapists' patients on average were more likely to attend >1 treatment session. As for the pre-post change models, an empty model (i.e., Equation 1 but with the therapist characteristic predictor [i.e.,  $\beta_{01}$ ] omitted and a clinic random intercept included) was used to calculate therapist-level ICCs. Specifically, we divided the variance in likelihood of attending >1 treatment session at the therapist and clinic levels by the sum of total variance combined with  $\pi^2 / 3$  (Rabe-Hesketh & Skrondal, 2012).

### *Sensitivity Analyses*

Models were rerun in eight sensitivity analyses. The decision to conduct a large set of sensitivity analyses was motivated by multiverse analysis within psychology that highlights the increased transparency these analyses can provide (Stegen et al., 2016). Our four preregistered sensitivity analyses included models with outliers on therapist-level predictors excluded (i.e., values  $\geq 3$  SDs from sample mean), restricting to therapist with  $\geq 5$  patients, restricting to patients with  $\geq 3$  sessions, and restricting to patients with clinically elevated symptoms at baseline<sup>2</sup>. In addition, we ran separate models for the OQ-45 and BHM-20, a model where we covaried the

---

<sup>2</sup> At the request of an anonymous reviewer, we examined the association between attending  $\geq 3$  sessions and reporting clinically elevated symptoms at baseline. Rates of clinically elevated symptoms were similar across the groups (72.5% versus 77.7%, for those attending <3 sessions and those attending  $\geq 3$  sessions, respectively). However, these characteristics were very modestly correlated ( $r = .06, p < .001$ ).

## PREDICTING THE THERAPIST EFFECT

time between the date of each patient's first session and when the therapist completed the assessment battery, and a model that used all patient data regardless of whether the assessment battery occurred later than a patient's first session. For pre-post change and early termination prediction sensitivity analyses restricted to the OQ-45 or BHM-20, we used the models in Equation 1 and 2, but with the BHM-20 term omitted. The clinic random effect was omitted for the OQ-45 models as only one clinic used the OQ-45. For the rate of change model, we followed the same model building steps to identify the best fitting model for the sensitivity analyses restricted to the OQ-45 or BHM-20. In all cases, models with log-linear session number and random slopes fit best. In instances where models did not converge (e.g., due to singular fit), random intercepts at the clinic level were removed.

### Results

Patients attended an average of 4.69 sessions ( $SD = 4.45$ ). Average pre-post change in psychological distress was  $d = -0.50$  ( $SD = 0.77$ ). Most patients (77.6%) had >1 sessions. Most patients (75.7%) had clinically elevated psychological distress at baseline. Descriptive statistics for therapist-level variables are reported in Supplemental Materials Table 2. Patient demographic characteristics across analytic samples are reported in Supplemental Materials Table 3. The therapist-level *ICC* for pre-post change in psychological distress was 1.63%. The therapist-level *ICC* when restricting to patients with >1 session was 1.56%. Therapist-level *ICCs* for the OQ-45 and BHM-20 were 1.59% and 5.27%, respectively. The therapist-level *ICC* for likelihood of attending >1 treatment session was 5.74% (*ICCs* = 4.96% and 18.5% for data drawn from clinics using the OQ-45 and BHM-20, respectively).

### Pre-Post Change

## PREDICTING THE THERAPIST EFFECT

Results of the pre-post change model are reported in Table 2 and Supplemental Materials Tables 4 to 11. In the primary model, only higher FIS verbal fluency was associated with larger pre-post reductions in distress ( $b = -0.0380$ ). A similar effect was observed in four sensitivity analyses (outliers removed, therapists with  $\geq 5$  patients, clinically elevated at baseline, covarying time to therapist assessment). Across the eight sensitivity analyses, there were eight additional instances of significant predictors. Higher therapist-level negative affect, higher FIS emotional expression, higher FIS empathy, more use of deliberate practice, and higher professional self-doubt were associated with smaller pre-post reductions in distress. More years of providing therapy and higher self-rated effectiveness were associated with larger pre-post reductions in distress. None of the associations in the pre-post change models survived  $p$ -value correction.

### **Rate of Change**

Results of the rate of change model are reported in Table 3 and Supplemental Materials Tables 12 to 19. In the primary model, negative affect and three FIS dimensions (emotional expression, empathy, alliance rupture-repair responsiveness) were the only therapist-level predictors of rate of change in psychological distress, with higher scores on each being associated with a slower (less steep) decrease in psychological distress over time (interaction  $bs = 0.028$  to  $0.039$ ,  $ps = .011$  and  $.041$ ). These effects did not survive  $p$ -value correction. Negative affect and various FIS scores were significant predictors in some (but not all) sensitivity analyses, always in the same direction as in the primary model (i.e., higher scores, slower decrease in psychological distress). Higher MCO scores (four models), loneliness (two models), verbal fluency, emotional Stroop interference, self-compassion, self-report empathy, agreeableness, and use of a feminist and/or multicultural theoretical orientation were associated with a slower decrease in psychological distress over time. Higher racial implicit bias was

## PREDICTING THE THERAPIST EFFECT

associated with a faster decrease in distress over time in one model. In total, 17 variables emerged as significant predictors when restricted to the BHM-20 data set including the FIS total score and seven FIS subscales, verbal fluency, emotional Stroop interference, self-compassion, self-reported empathy, MCO generally therapeutic response score and two MCO subscales, racial implicit bias, and use of a feminist and/or multicultural theoretical orientation. Fifteen of these effects survived *p*-value correction. However, these effects were based on a small number of therapists ( $ns \leq 28$ ) and most did not emerge as significant predictors in the primary model or the other sensitivity analyses, and thus should be interpreted very cautiously. No other effects survived *p*-value correction.

### **Early Termination**

Results of the early termination prediction model are reported in Table 4 and Supplemental Materials Tables 20 to 26. In the primary model, FIS verbal fluency, FIS persuasiveness, and mindfulness were all associated with greater likelihood of attending >1 treatment session. Deliberate practice was associated with lower likelihood of attending >1 treatment session. None of these effects survived *p*-value correction in the primary model.

All four variables that were significant predictors in the primary model (FIS verbal fluency, FIS persuasiveness, mindfulness, deliberate practice) showed associations in the same direction across multiple sensitivity analyses. The FIS total score was also associated with greater likelihood of attending >1 treatment session in two sensitivity analyses. Years of experience, self-rated effectiveness, professional self-doubt, and therapeutic technique diversity were associated with lower likelihood of attending >1 treatment session each in one model. Self-rated effectiveness showed the opposite effect in one model. Use of a feminist and/or multicultural theoretical orientation, MCO generally therapeutic response score, MCO

## PREDICTING THE THERAPIST EFFECT

opportunity, and MCO humility were associated with greater likelihood of attending >1 treatment session. Across sensitivity analyses, nine effects survived *p*-value correction, eight of which reflected the association between FIS verbal fluency or FIS persuasiveness with greater likelihood of attending >1 treatment session.

### **Discussion**

We conducted the first relatively large-scale, multimodal, preregistered assessment of the relationship between therapist-level characteristics and variation in patient outcomes (i.e., therapist effects). We examined 38 predictor variables across three types of outcomes (pre-post change, rate of change, early termination). We evaluated the stability of results across eight sensitivity analyses. Taken together, we found limited evidence suggesting that therapist characteristics impact treatment outcome. Despite selecting a large set of candidate predictors that have been empirically or theoretically linked with outcomes, results were largely null. The few statistically significant findings often failed to replicate across sensitivity analyses and/or failed to survive *p*-value correction. The most consistent evidence suggested that higher FIS verbal fluency and FIS persuasiveness are associated with greater likelihood of patients attending >1 session of therapy. This finding suggests that therapist interpersonal capacities assessed through brief responses to patient vignettes may contain signals that patients use when determining whether to return for a second therapy session.

Aside from demonstrating linkages between FIS verbal fluency and persuasiveness with early termination (which, to our knowledge, has not been reported before so cannot really be considered a replication), our results are largely disappointing for those hoping to identify reliable therapist-level predictors of treatment outcome. However, before drawing a firm conclusion that we simply cannot predict therapist effects, it is important to acknowledge

## PREDICTING THE THERAPIST EFFECT

methodological factors that may have influenced the current findings. The most obvious limitation is the likely low statistical power. Notably, our recruited sample size ( $n = 97$  therapists,  $n = 6,152$  patients) is larger than most even highly cited associations in the literature (e.g., Anderson et al., 2009; Chow et al., 2015; Nissen-Lie et al., 2013). However, the therapist effect in the current sample ( $\sim 1\text{-}2\%$ ) was considerably smaller than that reported in both meta-analyses as well as large-scale studies using individual patient data (i.e.,  $ICCs \sim 5\%$ ; Johns et al., 2019; Schiefele et al., 2017), as well as smaller than recent studies that have demonstrated linkages between therapists' interpersonal skills and outcomes (e.g., Schwartz et al., 2025). It is not possible to determine precisely why the therapist effect was smaller in the current study. One potential contributor is the relatively large number of therapists and patients per therapist, both of which tend to be associated with smaller therapist effects (Schiefele et al., 2016). In addition, the Canadian site implemented routine outcome monitoring and encouraged the use of deliberate practice; this may improve patient outcomes generally and reduce variability between therapists (de Jong et al., 2021; Goldberg, Babins-Wagner, et al., 2016). Our final therapist sample was also lower than our target sample size (100-200) and technical issues with some measures (especially the computerized behavioral measures) further reduced sample size and statistical power for models involving those predictors. Moreover, effect sizes observed for most models in the current study were modest. For example, in the primary pre-post model, the largest effect was  $b = -0.0380$  (for FIS verbal fluency), which indicates that a one  $SD$  increase in a therapist's FIS verbal fluency score was associated with a 0.038 lower pre-post  $d$  for their patients. Similarly, in the early termination model, the largest effect was  $b = 0.1886$  (also for FIS verbal fluency), which indicates that a one  $SD$  increase in a therapist's FIS verbal fluency score was associated with a 21% increase in the likelihood of their patients attending more than one session (i.e.,

## PREDICTING THE THERAPIST EFFECT

$e^{0.1886} = 1.208$ ). Both of these effect sizes are below commonly used benchmarks for small magnitude effects (Chen et al., 2010; Cohen, 1992).

We hope this study does not discourage efforts to identify therapist-level characteristics that are linked with patient outcomes. If anything, we hope this study highlights the importance of attending to methodological rigor in our efforts in this area, by preregistering our hypotheses prior to data analysis and by recruiting ideally much larger samples of therapists and patients. The many significant results that emerged when examining data from a relatively small sample of therapists ( $n_s \leq 28$  for the BHM-20 data set) that did not emerge in other models highlights the danger of drawing conclusions from small numbers of therapists. Unfortunately, likely due to logistical challenges with collecting linked therapist- and patient-level data, much of the existing literature on predictors of therapist differences is based on small numbers of therapists and patients (Heinonen & Nissen-Lie, 2019; Wampold & Owen, 2021, although see Schwartz et al., 2025). Had we selectively reported our results, we could have reported that therapists with higher FIS verbal fluency, higher self-rated effectiveness, more years of providing therapy, and lower negative affect have patients whose symptoms reduce more rapidly. However, this would have ignored the many unexpected associations that emerged in this sensitivity analysis. We just as easily could have concluded that those with higher FIS emotional expression and FIS empathy, higher professional self-doubt, and more use of deliberate practice have patients whose symptoms reduce more slowly.

Research on therapist-level predictors of outcomes that is truly definitive may well require thousands of therapists, especially in contexts where therapist differences are small. This may be feasible in health systems that routinely monitor outcomes (e.g., National Health Service in the United Kingdom). Given that therapist interpersonal skills have been most consistently

## PREDICTING THE THERAPIST EFFECT

associated with patient outcomes (Heinonen & Nissen-Lie, 2019), it may be crucial to include non-self-report measures of these capacities. While conducting thousands of FIS assessments would be extraordinarily expensive using human ratings, advances in artificial intelligence and machine learning suggest that automated scoring may soon be ready for use (Goldberg et al., 2021; Goldberg et al., in press). If coupled with text-based administration (Zech et al., 2023), a scalable FIS may not be too far in the future.

In addition to potentially modest statistical power, several other limitations are worth mentioning. First, though drawn from two countries and multiple clinics, our sample of therapists and patients were relatively homogenous in terms of race/ethnicity which limits generalizability to more racially/ethnically diverse populations. Second, there surely is variation in the degree to which the therapist characteristics we assessed are stable qualities. Some characteristics (e.g., negative affect, mindfulness) very likely change over time, even rapidly so. Other characteristics (e.g., FIS skills) may change systematically in the context of training (Anderson et al., 2020). We examined associations between these characteristics assessed at one time point and patient outcomes collected over years. This may have further reduced the likelihood that we were able to detect relevant signals. Although results were similar when covarying time between therapist assessment and a patient's first session and when expanding the window of acceptable patient data to include sessions that occurred prior to the therapists' assessment, ideally future studies will include multiple assessments of therapist characteristics to allow examination of their association with patient outcomes at times more proximal to when treatment actually occurred. Gathering such measures during therapist training and beyond may help clarify the degree which therapist expertise develops over the course of training and clinical practice (or not; Tracey et al., 2014). Third, although we examined several types of outcomes (i.e., pre-post change, rate of

## PREDICTING THE THERAPIST EFFECT

change, early termination), there were additional ways of defining outcome that may have been considered (e.g., reliable change index; defining early termination as attending only two sessions or not attending the last scheduled appointment; Jacobson & Truax, 1991; Smith & Greenberg, 2012) which may have produced different results. Fourth, we focused only on therapist predictors of outcomes and did not model patient factors or the interaction between therapist and patient factors. This may be an especially important limitation to address in future studies using large samples and analytic approaches particularly well suited to examining interactions (e.g., random forest machine learning methods; Denisko & Hoffman, 2018). Fifth, we examined only a subset of the universe of potentially relevant therapist characteristics. Further, some of the included characteristics and the measures we used to assess these constructs may be impacted by confounding factors that may obscure a meaningful relationship between these constructs and patient outcomes (e.g., acting ability may impact FIS performance but may or may not impact behavior in therapy). Finally, it would be valuable for future studies to employ statistical methods particularly well suited to handling multiple correlated predictors, such as many machine learning approaches. We considered employing machine learning in the current study. However, we opted to conduct individual multilevel models to mirror prior work in this area, to clearly characterize any associations with the individual characteristics examined, and to maintain consistency with our preregistered analytic plan. Moreover, the sample size of therapists ( $n = 97$ ) is woefully below the recommended sample size for machine learning (e.g.,  $n = 500-1000$ ; Zantvoort et al., 2024). Nonetheless, we are hopeful that future studies employing machine learning on adequately large (and ideally adequately rich, e.g., using session recordings; Kuo et al., 2024) datasets will greatly enhance our understanding of therapist contributors to patient outcome. Evaluating whether models trained using one data source (e.g., from a large

## PREDICTING THE THERAPIST EFFECT

Canadian mental health agency) generalize to another data source (e.g., center in the U.S.) could be a valuable future direction, given evidence that generalizability is often poor (Chekroud et al., 2024).

In conclusion, it remains an open question as to what characteristics make one therapist more effective than another. Therapist interpersonal skills (e.g., assessed via FIS) have been suggested as the most consistent predictor of patient outcomes, yet we largely failed to replicate these associations or the various other associations that have been reported in the literature. Some interpersonal capacities (i.e., verbal fluency, persuasiveness) do appear to be linked with likelihood of attending more than one therapy session, but these capacities are largely not associated with larger or faster reductions in symptoms over time. We hope this study encourages further large-scale, preregistered, multimodal assessment of therapist characteristics that can be used to ultimately improve treatment for the benefit of patients.

### **Author Contributions**

SBG: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project Administration; Supervision; Writing – original draft; Writing – review & editing. SAB: Conceptualization; Formal analysis; Methodology; Software; Writing – review & editing. AWP: Data curation; Investigation; Writing – review & editing. RB: Data curation; Writing – review & editing. DDC: Data curation; Writing – review & editing. CYW: Data curation; Writing – review & editing. EHH: Data curation; Writing – review & editing. SUL: Data curation; Writing – review & editing. JKC: Data curation; Writing – review & editing. GD: Data curation; Writing – review & editing. NS: Methodology; Writing – review & editing. SMK: Data curation; Writing – review & editing. TA: Methodology; Writing – review &

## PREDICTING THE THERAPIST EFFECT

editing. BEW: Conceptualization; Writing – review & editing. JO: Conceptualization; Methodology; Writing – review& editing.

### **Conflicts of Interest**

SMK is the CEO of CelestHealth which operates the Behavioral Health Monitor (BHM) measure. The remaining authors have no conflicts to disclose.

### **Acknowledgements**

The authors are grateful to William T. Hoyt for early discussions about this project.

### **Funding**

SBG was partially supported by the National Center for Complementary and Integrative Health (K23AT010879) and the National Institute of Mental Health (R01MH139512). Nili Solomonov was supported by a grant from the National Institute of Mental Health (K23MH123864). This project was funded in part by the 50<sup>th</sup> Anniversary Grant from the American Psychological Association Division 29 (Psychotherapy).

### **References**

- Anderson, T., Ogles, B.M., Patterson, C.L., Lambert, M.J., & Vermeersch, D.A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology, 65*(7), 755-768. doi: 10.1002/jclp.20583
- Anderson, T., Perlman, M. R., McCarrick, S. M., & McClintock, A. S. (2020). Modeling therapist responses with structured practice enhances facilitative interpersonal skills. *Journal of Clinical Psychology, 76*(4), 659-675.
- Baldwin, S.A., & Imel, Z.E. (2013). Therapist effects: Findings and methods. In M.J. Lambert (Ed.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change (6th ed.)* (p. 258-297). Hoboken, NJ: Wiley & Sons.

## PREDICTING THE THERAPIST EFFECT

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1), 289-300.

Boswell, J. F., Gallagher, M. W., Sauer-Zavala, S. E., Bullis, J., Gorman, J. M., Shear, M. K., Woods, S., & Barlow, D. H. (2013). Patient characteristics and variability in adherence and competence in cognitive-behavioral therapy for panic disorder. *Journal of Consulting and Clinical Psychology*, 81(3), 443-454. doi: 10.1037/a0031437

Chekroud, A. M., Hawrilenko, M., Loho, H., Bondar, J., Gueorguieva, R., Hasan, A., Kambeitz, J., Corlett, P. R., Koutsouleris, N., Krumholz, H. M., Krystal, J. H., & Paulus, M. (2024). Illusory generalizability of clinical prediction models. *Science*, 383(6679), 164-167.

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitude of odds ratios in epidemiological studies. *Communications in Statistics-Simulation and Computation*, 39, 860-864. doi: 10.1080/03610911003650383

Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.

Davis, D. E., DeBlaere, C., Owen, J., Hook, J. N., Rivera, D. P., Choe, E., ... & Placeres, V. (2018). The multicultural orientation framework: A narrative review. *Psychotherapy*, 55(1), 89-100.

de Jong, K., Conijn, J. M., Gallagher, R. A., Reshetnikova, A. S., Heij, M., & Lutz, M. C.

## PREDICTING THE THERAPIST EFFECT

- (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, 85, 102002.
- Denisko, D., & Hoffman, M. M. (2018). Classification and interaction in random forests. *Proceedings of the National Academy of Sciences*, 115(8), 1690-1692.
- Ericsson, K.A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, 49(8), 725-747.
- Germer, S., Weyrich, V., Bräscher, A.-K., Mütze, K., & Witthöft, M. (2022). Does practice really make perfect? A longitudinal analysis of the relationship between therapist experience and therapy outcome: A replication of Goldberg, Rousmaniere, et al. (2016). *Journal of Counseling Psychology*, 69(5), 745–754.
- Goldberg, S. B., Babins-Wagner, R., Imel, Z. E., Caperton, D. D., Weitzman, L. M., & Wampold, B. E. (2023). Threat alert: The effect of outliers on the alliance–outcome correlation. *Journal of Counseling Psychology*, 70(1), 81–89. doi: 10.1037/cou0000638
- Goldberg, S. B., Flemotomos, N., Martinez, V. R., Tanana, M., Kuo, P., Pace, B. T., Villatte, J. L., Georgiou, P. G., Van Epps, J., Imel, Z. E., & Atkins, D. C. (2020). Machine learning and natural language processing in psychotherapy research: Alliance as example use case. *Journal of Counseling Psychology*, 67(4), 438-448. doi: 10.1037/cou0000382
- Goldberg, S.B., Rousmaniere, T., Miller, S.D., Whipple, J., Nielsen, S.L., Hoyt, W.T., & Wampold, B.E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology*, 63(1), 1-11. doi: 10.1037/cou0000131
- Goldberg, S. B., Tanana, M., Imel, Z. E., Atkins, D. C., Hill, C. E., & Anderson, T. (2021). Can

## PREDICTING THE THERAPIST EFFECT

- a computer detect interpersonal skills? Using machine learning to scale up the Facilitative Interpersonal Skills task. *Psychotherapy Research*, 31(3), 281-288.
- Goldberg, S. B., Tanana, M., Stewart, S. H., Williams, C. Y., Atkins, D. C., Imel, Z. E., & Owen, J. (in press). Automating the assessment of multicultural orientation through machine learning and natural language processing. *Psychotherapy*.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549-576.
- Hatchett, G. T., & Park, H. L. (2003). Comparison of four operational definitions of premature termination. *Psychotherapy: Theory, Research, Practice, Training*, 40(3), 226-231.
- Heinonen, E., & Nissen-Lie, H. A. (2020). The professional and personal characteristics of effective psychotherapists: A systematic review. *Psychotherapy Research*, 30(4), 417-432. doi: 10.1080/10503307.2019.1620366
- Hill, C. E., & Norcross, J. C. (Eds.). (2023). *Psychotherapy skills and methods that work*. Oxford University Press.
- Jacobson, N. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19.
- Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to Baldwin and Imel's (2013) review. *Clinical Psychology Review*, 67, 78-93. doi:10.1016/j.cpr.2018.08.004
- Kopta, S.M., & Lowry, J.L. (2002). Psychometric evaluation of the Behavioral Health Questionnaire-20: A brief, comprehensive instrument for assessing mental health. *Psychotherapy Research*, 12, 413-426.

## PREDICTING THE THERAPIST EFFECT

- Kopta, M., Owen, J., & Budge, S. (2015). Measuring psychotherapy outcomes with the Behavioral Health Measure-20: Efficient and comprehensive. *Psychotherapy, 52*(4), 442-448. doi: 10.1037/pst0000035
- Kuo, P. B., Tanana, M. J., Goldberg, S. B., Caperton, D. D., Narayanan, S., Atkins, D. C., & Imel, Z. E. (2024). Machine-Learning-Based Prediction of Client Distress From Session Recordings. *Clinical Psychological Science, 12*(3), 435-446. doi: 10.1177/216770262311726
- Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour, 4*(4), 423-434.
- Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C.,...Burlingame, G. B. (2004). *Administration and scoring manual for the Outcome Questionnaire-45*. Orem, UT: American Professional Credentialing Services.
- Lingiardi, V., Muzi, L., Tanzilli, A., & Carone, N. (2018). Do therapists' subjective variables impact on psychodynamic psychotherapy outcomes? A systematic literature review. *Clinical Psychology & Psychotherapy, 25*(1), 85-101. doi: 10.1002/cpp.2131
- Luborsky, L., McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry, 42*(6), 602-611.
- Miller, S.D., Hubble, M.A., Chow, D.L., & Seidel, J.A. (2013). The outcome of psychotherapy: Yesterday, today, and tomorrow. *Psychotherapy, 50*(1), 88-97. doi: 10.1037/a0031097
- Minami, T., Davies, D.R., Tierney, S.C., Bettman, J.E., McAward, S.M., Averill, L.A.,...Wampold, B.E. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of*

## PREDICTING THE THERAPIST EFFECT

*Counseling Psychology*, 56(2), 309-320. doi: 10.1037/a0015398

Power, N., Noble, L. A., Simmonds-Buckley, M., Kellett, S., Stockton, C., Firth, N., &

Delgadillo, J. (2022). Associations between treatment adherence–competence–integrity (ACI) and adult psychotherapy outcomes: A systematic review and meta-analysis.

*Journal of Consulting and Clinical Psychology*, 90(5), 427–445.

R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata, Volume II: Categorical responses, counts, and survival*. Stata Press.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods in psychotherapy. *American Journal of Orthopsychiatry*, 6, 412-415.

Rousmaniere, T. (2017). *Deliberate practice for psychotherapists: A guide to improving clinical effectiveness*. New York: Routledge.

Rousmaniere, T.G., Swift, J.K., Babins-Wagner, R., Whipple, J.L., & Berzins, S. (2016). Supervisor variance in psychotherapy outcome in routine practice. *Psychotherapy Research*, 26(2), 1-10. doi: 10.1080/10503307.2014.963730

Schiefele, A. K., Lutz, W., Barkham, M., Rubel, J., Böhnke, J., Delgadillo, J., ... & Lambert, M. J. (2017). Reliability of therapist effects in practice-based psychotherapy research: A guide for the planning of future studies. *Administration and Policy in Mental Health and Mental Health Services Research*, 44, 598-613.

Schwartz, B., Hehlmann, M. I., Deisenhofer, A. K., Rubel, J. A., Fischer, L., Lutz, W., & Schöttke, H. (2025). Elucidating Therapist Differences: Therapists' Interpersonal Skills and Their Effect on Treatment Outcome. *Behaviour Research and Therapy*, 104689, 1-

## PREDICTING THE THERAPIST EFFECT

11.

Simmons, J.P., Nelson, L.D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant.

*Psychological Science*, 22(11), 1359-1366. doi: 10.1177/0956797611417632

Solomonov, N., McCarthy, K. S., Gorman, B. S., & Barber, J. P. (2019). The Multitheoretical List of Therapeutic Interventions–30 items (MULTI-30). *Psychotherapy Research*, 29(5), 565-580. doi:10.1080/10503307.2017.1422216

Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702-712.

Stewart, S. H., Drinane, J. M., Owen, J., & Dumas, D. (2024). Psychotherapy training via a task-based assessment of the multicultural orientation framework: A pilot study. *Counselling and Psychotherapy Research*, 24(1), 190-198.

Swift, J.K., & Greenberg, R.P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 80(4), 547-559.

doi: 10.1037/a0028226

Tracey, T.J.G., Wampold, B.E., Lichtenberg, J.W., & Goodyear, R.K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist*, 69(3), 218-229.

doi: 10.1037/a0035099

Wampold, B., & Imel, Z.E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd ed.)*. New York: Routledge.

Wampold, B. E., & Owen, J. (2021). Therapist effects: History, methods, magnitude and characteristics of effective therapists (pp. 297-326). In M. Barkham, W. Lutz, and L. G. Castonguay (Eds.), *Bergin and Garfield's Handbook of Psychotherapy (7th ed.)*. John

## PREDICTING THE THERAPIST EFFECT

Wiley & Sons, Inc.

Xiao, H., Carney, D. M., Youn, S. J., Janis, R. A., Castonguay, L. G., Hayes, J. A., & Locke, B.

D. (2017). Are we in crisis? National mental health and treatment trends in college counseling centers. *Psychological Services, 14*(4), 407-415.

Zantvoort, K., Nacke, B., Görlich, D., Hornstein, S., Jacobi, C., & Funk, B. (2024). Estimation of

minimal data sets sizes for machine learning predictions in digital mental health interventions. *npj Digital Medicine, 7*, 361.

Zech, J., Foley, V. K., Hull, T. D., & Anderson, T. (2023). Assessing the quality of digital

patient-therapist communication: the development and validation of a text-based facilitative interpersonal skills task. *Psychotherapy Research, 33*(6), 743-756.

## PREDICTING THE THERAPIST EFFECT

**Table 1**

*Study Measures Organized by Domain*

Measure	Domain	Modality
FIS total	FIS overall	Performance task
FIS verbal fluency	FIS verbal fluency	Performance task
Word fluency test	FIS verbal fluency	Computerized behavioral measure
FIS hope and positive expectations	FIS hope and positive expectations	Performance task
Depression, Anxiety, and Stress Scale	FIS hope and positive expectations	Self-report
Psychological Well-being	FIS hope and positive expectations	Self-report
FIS persuasiveness	FIS persuasiveness	Performance task
Ten Item Personality Inventory extraversion	FIS persuasiveness	Self-report
FIS emotional expression	FIS emotional expression	Performance task
Emotional Stroop task	FIS emotional expression	Computerized behavioral measure
FIS warmth, acceptance, and understanding	FIS warmth, acceptance, and understanding	Performance task
Ten Item Personality Inventory agreeableness	FIS warmth, acceptance, and understanding	Self-report
Five Facet Mindfulness Questionnaire Short Form total score	FIS warmth, acceptance, and understanding	Self-report
Self-Compassion Scale Short Form total score	FIS warmth, acceptance, and understanding	Self-report
Dispositional Contempt	FIS warmth, acceptance, and understanding	Self-report
FIS empathy	FIS empathy	Performance task
Interpersonal Reactivity Index	FIS empathy	Self-report
Empathic accuracy task	FIS empathy	Computerized behavioral measure
FIS alliance bond capacity	FIS alliance bond capacity	Performance task
FIS alliance rupture and repair	FIS alliance rupture and repair	Performance task

## PREDICTING THE THERAPIST EFFECT

NIH Toolbox Loneliness	FIS alliance bond capacity / rupture and repair	Self-report
NIH Toolbox Emotional Support	FIS alliance bond capacity / rupture and repair	Self-report
Multicultural Orientation task overall	Multicultural	Performance task
Multicultural Orientation task comfort	Multicultural	Performance task
Multicultural Orientation task humility	Multicultural	Performance task
Multicultural Orientation task opportunity	Multicultural	Performance task
Racial Implicit Association task	Multicultural	Computerized behavioral measure
Multicultural Knowledge, Awareness, and Skills Scale total score	Multicultural	Self-report
Use of feminist and/or multicultural theory	Multicultural	Self-report
Deliberate practice hours	Professional characteristics	Self-report
Difficulties in Therapeutic Practice professional self-doubt	Professional characteristics	Self-report
Years providing psychotherapy	Professional characteristics	Self-report
Perceived efficacy	Professional characteristics	Self-report
Years of personal therapy	Professional characteristics	Self-report
Multitheoretical List of Interventions therapeutic technique diversity	Professional characteristics	Self-report
Experiences in Close Relationships anxious	Attachment	Self-report
Experiences in Close Relationships avoidant	Attachment	Self-report
Childhood Trauma Questionnaire total	Attachment	Self-report

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health.

PREDICTING THE THERAPIST EFFECT

**Table 2**

*Pre-Post Model Results*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	4,728	84	NA	-0.0025	0.884	0.953
FIS verbal fluency	4,728	84	NA	-0.0380	0.037	0.953
Word fluency test	4,122	69	NA	0.0044	0.828	0.953
FIS hope and positive expectations	4,728	84	NA	-0.0005	0.976	0.976
Depression, Anxiety, and Stress Scale	4,964	88	NA	0.0290	0.077	0.953
Psychological Well-being	4,964	88	NA	-0.0212	0.214	0.953
FIS persuasiveness	4,728	84	NA	-0.0148	0.409	0.953
Ten Item Personality Inventory extraversion	4,964	88	NA	-0.0134	0.451	0.953
FIS emotional expression	4,728	84	NA	0.0159	0.380	0.953
Emotional Stroop task	3,941	69	NA	0.0069	0.728	0.953
FIS warmth, acceptance, and understanding	4,728	84	NA	0.0112	0.525	0.953
Ten Item Personality Inventory agreeableness	4,964	88	NA	-0.0035	0.837	0.953
Five Facet Mindfulness Questionnaire Short Form total score	4,964	88	NA	0.0038	0.826	0.953
Self-Compassion Scale Short Form total score	4,964	88	NA	0.0107	0.526	0.953
Dispositional Contempt	4,964	88	NA	0.0213	0.157	0.953
FIS empathy	4,728	84	NA	0.0097	0.581	0.953
Interpersonal Reactivity Index	4,964	88	NA	0.0050	0.772	0.953
Empathic accuracy task	4,566	82	NA	-0.0019	0.912	0.953
FIS alliance bond capacity	4,728	84	NA	0.0026	0.870	0.953
FIS alliance rupture and repair	4,728	84	NA	0.0021	0.903	0.953
NIH Toolbox Loneliness	4,964	88	NA	0.0177	0.283	0.953
NIH Toolbox Emotional Support	4,964	88	NA	0.0016	0.928	0.953
Multicultural Orientation task overall	4,826	85	NA	0.0138	0.462	0.953
Multicultural Orientation task comfort	4,826	85	NA	0.0052	0.790	0.953

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	4,826	85	NA	0.0121	0.527	0.953
Multicultural Orientation task opportunity	4,826	85	NA	0.0029	0.862	0.953
Racial Implicit Association task	4,122	69	NA	-0.0091	0.617	0.953
Multicultural Knowledge, Awareness, and Skills Scale total score	4,964	88	NA	-0.0156	0.347	0.953
Use of feminist and/or multicultural theory	5,181	96	NA	0.0124	0.473	0.953
Deliberate practice hours	4,773	82	NA	0.0247	0.134	0.953
Difficulties in Therapeutic Practice professional self-doubt	4,964	88	NA	-0.0022	0.900	0.953
Years providing psychotherapy	5,171	95	NA	-0.0142	0.565	0.953
Perceived efficacy	5,171	95	NA	-0.0117	0.512	0.953
Years of personal therapy	5,047	93	NA	0.0103	0.516	0.953
Multitheoretical List of Interventions therapeutic technique diversity	4,964	88	NA	-0.006	0.660	0.953
Experiences in Close Relationships anxious	4,964	88	NA	0.0068	0.691	0.953
Experiences in Close Relationships avoidant	4,964	88	NA	0.0041	0.812	0.953
Childhood Trauma Questionnaire total	4,704	80	NA	0.0307	0.078	0.953

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting pre-post changes in psychological distress; Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with larger pre-post reductions in psychological distress.

PREDICTING THE THERAPIST EFFECT

**Table 3**

*Rate of Change Model Results*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Level 4 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	23,719	5,426	84	12	0.0287	0.052	0.363
FIS verbal fluency	23,719	5,426	84	12	0.0040	0.797	0.864
Word fluency test	21,968	4,763	69	8	0.0060	0.719	0.854
FIS hope and positive expectations	23,719	5,426	84	12	0.0225	0.123	0.389
Depression, Anxiety, and Stress Scale	24,978	5,712	88	12	0.0279	0.041	0.363
Psychological Well-being	24,978	5,712	88	12	-0.0157	0.270	0.489
FIS persuasiveness	23,719	5,426	84	12	0.0269	0.072	0.363
Ten Item Personality Inventory extraversion	24,978	5,712	88	12	-0.012	0.403	0.589
FIS emotional expression	23,719	5,426	84	12	0.0387	0.013	0.247
Emotional Stroop task	21,044	4,605	69	8	0.0194	0.221	0.489
FIS warmth, acceptance, and understanding	23,719	5,426	84	12	0.0172	0.243	0.489
Ten Item Personality Inventory agreeableness	24,978	5,712	88	12	0.0154	0.258	0.489
Five Facet Mindfulness Questionnaire Short Form total score	24,978	5,712	88	12	0.0042	0.762	0.864
Self-Compassion Scale Short Form total score	24,978	5,712	88	12	0.0156	0.253	0.489
Dispositional Contempt	24,978	5,712	88	12	0.0119	0.346	0.564
FIS empathy	23,719	5,426	84	12	0.0375	0.011	0.247
Interpersonal Reactivity Index	24,978	5,712	88	12	0.0242	0.074	0.363
Empathic accuracy task	22,922	5,235	82	13	0.0018	0.903	0.903
FIS alliance bond capacity	23,719	5,426	84	12	0.0206	0.137	0.400
FIS alliance rupture and repair	23,719	5,426	84	12	0.0316	0.029	0.363
NIH Toolbox Loneliness	24,978	5,712	88	12	0.0172	0.208	0.489
NIH Toolbox Emotional Support	24,978	5,712	88	12	-0.0037	0.798	0.864
Multicultural Orientation task overall	24,115	5,541	85	12	0.0277	0.078	0.363
Multicultural Orientation task comfort	24,115	5,541	85	12	0.0260	0.116	0.389

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	24,115	5,541	85	12	0.0243	0.117	0.389
Multicultural Orientation task opportunity	24,115	5,541	85	12	0.0229	0.086	0.363
Racial Implicit Association task	21,968	4,763	69	8	-0.0131	0.387	0.588
Multicultural Knowledge, Awareness, and Skills Scale total score	24,978	5,712	88	12	0.0058	0.685	0.840
Use of feminist and/or multicultural theory	25,957	5,963	96	13	0.0134	0.326	0.563
Deliberate practice hours	24,759	5,543	82	12	-0.0127	0.356	0.564
Difficulties in Therapeutic Practice professional self-doubt	24,978	5,712	88	12	-0.0090	0.533	0.675
Years providing psychotherapy	25,904	5,948	95	13	-0.0045	0.819	0.864
Perceived efficacy	25,904	5,948	95	13	-0.0028	0.841	0.864
Years of personal therapy	25,386	5,807	93	13	0.0178	0.149	0.404
Multitheoretical List of Interventions therapeutic technique diversity	24,978	5,712	88	12	-0.0157	0.217	0.489
Experiences in Close Relationships anxious	24,978	5,712	88	12	0.0109	0.436	0.614
Experiences in Close Relationships avoidant	24,978	5,712	88	12	0.0092	0.518	0.675
Childhood Trauma Questionnaire total	23,547	5,431	80	11	0.0104	0.456	0.619

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = interaction between each predictor variable and log session number from multilevel models predicting psychological distress; Level 1 = session level; Level 2 = patient level; Level 3 = therapist level; Level 4 = clinic level. Negative estimates reflect associations with faster reductions in psychological distress over time.

PREDICTING THE THERAPIST EFFECT

**Table 4**

*Results of Multilevel Logistic Regression Models Predicting Attendance at >1 Treatment Session*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	5,584	85	12	0.1124	0.078	0.494
FIS verbal fluency	5,584	85	12	0.1886	0.003	0.076
Word fluency test	4,905	70	8	-0.0212	0.766	0.845
FIS hope and positive expectations	5,584	85	12	0.0857	0.185	0.684
Depression, Anxiety, and Stress Scale	5,878	89	12	-0.0281	0.651	0.825
Psychological Well-being	5,878	89	12	0.0183	0.774	0.845
FIS persuasiveness	5,584	85	12	0.1785	0.004	0.076
Ten Item Personality Inventory extraversion	5,878	89	12	0.0132	0.839	0.883
FIS emotional expression	5,584	85	12	0.0746	0.273	0.684
Emotional Stroop task	4,768	70	8	0.0937	0.210	0.684
FIS warmth, acceptance, and understanding	5,584	85	12	0.0522	0.425	0.734
Ten Item Personality Inventory agreeableness	5,878	89	12	0.0719	0.227	0.684
Five Facet Mindfulness Questionnaire Short Form total score	5,878	89	12	0.1315	0.022	0.209
Self-Compassion Scale Short Form total score	5,878	89	12	0.0424	0.486	0.769
Dispositional Contempt	5,878	89	12	0.0161	0.775	0.845
FIS empathy	5,584	85	12	0.0731	0.267	0.684
Interpersonal Reactivity Index	5,878	89	12	-0.0111	0.860	0.883
Empathic accuracy task	5,393	83	13	-0.0376	0.570	0.809
FIS alliance bond capacity	5,584	85	12	0.0286	0.640	0.825
FIS alliance rupture and repair	5,584	85	12	0.0643	0.311	0.684
NIH Toolbox Loneliness	5,878	89	12	0.1134	0.064	0.486
NIH Toolbox Emotional Support	5,878	89	12	-0.1049	0.105	0.570
Multicultural Orientation task overall	5,701	86	12	0.0343	0.627	0.825
Multicultural Orientation task comfort	5,701	86	12	0.0862	0.226	0.684

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	5,701	86	12	0.0207	0.778	0.845
Multicultural Orientation task opportunity	5,701	86	12	0.0621	0.312	0.684
Racial Implicit Association task	4,905	70	8	0.0651	0.344	0.684
Multicultural Knowledge, Awareness, and Skills Scale total score	5,878	89	12	0.0336	0.575	0.809
Use of feminist and/or multicultural theory	6,152	97	13	-0.0062	0.930	0.930
Deliberate practice hours	5,723	83	12	-0.1541	0.006	0.076
Difficulties in Therapeutic Practice professional self-doubt	5,878	89	12	-0.0601	0.360	0.684
Years providing psychotherapy	6,137	96	13	-0.0696	0.444	0.734
Perceived efficacy	6,137	96	13	-0.0250	0.736	0.845
Years of personal therapy	5,994	94	13	0.0617	0.356	0.684
Multitheoretical List of Interventions therapeutic technique diversity	5,878	89	12	-0.0306	0.563	0.809
Experiences in Close Relationships anxious	5,878	89	12	0.0492	0.425	0.734
Experiences in Close Relationships avoidant	5,878	89	12	0.0748	0.281	0.684
Childhood Trauma Questionnaire total	5,592	81	11	-0.0676	0.293	0.684

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting likelihood of attending >1 treatment session (in log units); Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level. Positive estimates reflect associations with greater likelihood of attending >1 treatment session.

**Supplemental Materials Table 1**

*Description of Computerized Behavioral Measures and Self-Report Questionnaires*

**Computerized Behavioral Measures. *Word Fluency Test.*** This task is based on the Controlled Oral Word Association Test (Benton, 1967). It requires participants to type as many unique words as they can beginning with specific letters (A, F, S). They have two minutes to provide responses to each letter. The score represents the number of unique English words provided across the three letter types. Word fluency tasks are widely used to assess verbal fluency (Ruff et al., 1996).

***Emotional Stroop Task (Smith & Waterman, 2003).*** This task is an adaptation of the classic Stroop task (Stroop, 1935). Participants are asked to identify the color of words (blue, red, yellow, green). In the emotional version, words are drawn from five categories (aggression, neutral, positive emotion, negative emotion, color word). An interference score was computed reflecting the degree to which mean latency for responses to negative emotion words differed from responses to neutral words. Larger positive values reflect greater interference associated with the negative emotion words. Prior meta-analytic work has demonstrated greater interference with negative emotion words for individuals with depressive or anxiety disorders relative to healthy controls (Joyal et al., 2019).

***Racial Implicit Association Test (Nosek et al., 2007).*** This task is a version of the Implicit Association Test (IAT; Greenwald et al., 1998). The IAT is designed to assess the strength of automatic associations for various categories. During the racial IAT, participants are asked to sort target items (images of Black and White individuals) and attributes (positive and negatively valenced words) into categories by pressing specific keys (i.e., left key [E] when items belong to the category displayed on the left side of the screen, right key [I] when items

## PREDICTING THE THERAPIST EFFECT

belong to the category on the right side). During the test trials, participants sort items into combined categories (e.g., “White OR Good” and “Black OR Bad”, “White OR Bad”, “Black OR Good”). A measure of bias ( $D$ ) is calculated by comparing response latencies from trials where White and Good / Black and Bad are paired versus those where Black and Good / White and Bad are paired, divided by the standard deviation of response latencies in the two blocks (Greenwald et al., 2003). The racial IAT is a widely used, if imperfect, implicit measure of racial bias as it does not appear to strongly relate to most criterion measures of racial bias and is not necessarily a stronger predictor than explicit measures (Oswald et al., 2013).

***Empathic Accuracy Task (EAT; Kral et al., 2017; Zaki et al., 2009).*** This task involved viewing video recordings of individuals (targets) sharing emotional experiences (e.g., death of a grandparent, winning a sporting event). The targets in these videos had previously watched their own videos and provided ratings of the emotions they were experiencing in the videos on a scale ranging from 1 (very negative) to 9 (very positive). Ratings were collected continuously throughout the video. Participants in the current study viewed six videos (three positive and three negative emotional experiences) and made continuous ratings of the emotions they believed the target was experiencing in the videos. Correlations were computed for each participant reflecting the correspondence between their ratings of the targets’ emotions and those of the targets themselves. Higher correlations represent closer correspondence with target ratings. Performance on this task has previously been shown to correspond with neural activation in brain regions subserving empathy (e.g., superior temporal sulcus, medial prefrontal cortex; Zaki et al., 2009).

**Self-Report Questionnaires.** A variety of self-report measures theoretically linked with FIS or MCO and/or empirically or theoretically linked with therapist effects were collected. These measures are described in the order in which they appear in Table 1.

## PREDICTING THE THERAPIST EFFECT

***Depression, Anxiety, and Stress Scale - 21 (DASS-21; Henry & Crawford, 2005)*** is a 21-item measure of depression, anxiety, and stress symptoms. It has shown strong convergent validity with legacy measures of negative affect (Henry & Crawford, 2005). Internal consistency reliability for the total score was high ( $\alpha = .85$ ).

***Psychological Well-being (PWB; Ryff, 1989)*** is an 18-item measure of well-being. It assesses well-being across six domains: autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance. The measure has shown predictable changes across the lifespan and associations with measures within the nomological network of well-being (Ryff & Keyes, 1995). Internal consistency reliability for the total score was high ( $\alpha = .80$ ).

***Ten-Item Personality Inventory (TIPI; Gosling et al., 2003)*** is an ultra-brief assessment of the big five personality dimensions. The measure has shown acceptable test-retest reliability and convergent validity with longer measures and observer ratings (Gosling et al., 2003). We used the items corresponding to extraversion and agreeableness. Inter-item correlations were above the recommended cut-off ( $r_s = .59$  and  $.16$ , for extraversion and agreeableness, respectively; Clark & Watson, 1995).

***Five Facet Mindfulness Questionnaire Short Form (FFMQ; Bohlmeijer et al., 2011)*** is a 24-item measure designed to assess five dimensions of mindfulness: acting with awareness, describe, non-reactivity, non-judgment, and observe (Baer et al., 2006). The short-form FFMQ has shown factorial and convergent validity (Bohlmeijer et al., 2011). Internal consistency reliability for the total score was high ( $\alpha = .85$ ).

***Self-Compassion Scale Short Form (SCS; Raes et al., 2011)*** is a 12-item measure designed to assess six dimensions of self-compassion: over-identification, self-kindness,

## PREDICTING THE THERAPIST EFFECT

mindfulness, isolation, common humanity, and self-judgment. The short-form SCS is nearly perfectly correlated with the long-form version and has demonstrated factorial validity (Raes et al., 2011). Internal consistency reliability for the total score was high ( $\alpha = .87$ ).

***Dispositional Contempt (Schriber et al., 2017)*** is a 10-item measure designed to assess contempt (i.e., the tendency to look down on and derogate others) as a personality characteristic. The measure has shown convergent validity through associations with measures within the nomological network of dispositional contempt (Schriber et al., 2017). Internal consistency reliability for the total score was high ( $\alpha = .79$ ).

***Interpersonal Reactivity Index (IRI; Davis, 1983)*** is a 28-item measure designed to assess individual differences in empathy. It assesses four dimensions: perspective taking, fantasy, empathic concern, and personal distress. It has shown strong convergent and discriminant validity (Davis, 1983). Internal consistency reliability for the total score was high ( $\alpha = .77$ ).

***National Institutes of Health (NIH) Toolbox Loneliness (Cyranowski et al., 2013)*** is a five-item measure designed to assess loneliness. It has shown strong convergent validity with legacy measures of loneliness and differentiation from other aspects of social functioning (Cyranowski et al., 2013). Internal consistency reliability for the total score was high ( $\alpha = .87$ ).

***NIH Toolbox Emotional Support (Cyranowski et al., 2013)*** is an eight-item measure designed to assess the degree to which an individual perceives themselves to have access to emotional support. This construct showed differentiation from other forms of social support (e.g., instrumental support) as well as convergent validity with other measures of social functioning (Cyranowski et al., 2013). Internal consistency reliability for the total score was high ( $\alpha = .93$ ).

***Multicultural Knowledge, Awareness, and Skills Survey Counselor Edition Revised (MAKSS-CE-R; Kim et al., 2003)*** is a 33-item measure designed to assess therapists' perception

## PREDICTING THE THERAPIST EFFECT

of their multicultural competence across three domains: knowledge, awareness, and skills. The measure has shown evidence for convergent validity with legacy measures and criterion-related validity with higher scores associated with more training and experience with multicultural counseling (Kim et al., 2003). Internal consistency reliability for the total score was high ( $\alpha = .82$ ).

***Feminist and/or Multicultural Theoretical Orientation.*** Participants indicated their use of various theoretical orientations by selecting orientations from a list that included feminist and multicultural orientations. Participants indicated all that applied to them. A dichotomous variable was computed reflecting their use of a feminist and/or multicultural orientation.

***Use of Deliberate Practice.*** We used a single item that has been previously used to assess engagement in deliberate practice (“How many hours per week (on average) to you spend alone seriously engaging in activities related to improving your therapy skills”; Chow et al., 2015). More hours spent in deliberate practice has previously been linked with better patient outcomes (Chow et al., 2015).

***Professional Self-Doubt*** was assessed using the 5-item subscale of the Difficulties in Therapeutic Practice measure (Orlinsky et al., 1999). Items assess therapists’ doubt about their ability to effectively treat their patients. Higher professional self-doubt has previously been associated with better patient outcomes (Nissen-Lie et al., 2013). Internal consistency reliability for the subscale score was high ( $\alpha = .74$ ).

***Years Providing Psychotherapy.*** Participants indicated their years providing psychotherapy in a single item: “How many years have you been providing psychotherapy?”

## PREDICTING THE THERAPIST EFFECT

***Perceived Self-Efficacy.*** Participants indicated their perceived self-efficacy using a single item, “Relative to other therapists in general, how would you rate your effectiveness with clients?” Response options ranged from 0 (much less effective) to 100 (much more effective).

***Use of Personal Therapy.*** Participants indicated whether they have ever engaged in personal psychotherapy.

***Multitheoretical List of Interventions (MULTI-30; Solomonov et al., 2019)*** is a 30-item measure assessing therapists’ use of therapeutic techniques drawn from diverse theoretical orientations (e.g., cognitive behavioral, psychodynamic, interpersonal). Following Chen et al. (2020), we calculated an index of therapeutic technique diversity to assess therapists’ use of a range of therapeutic interventions from different therapeutic orientations. We applied Shannon’s (1948) entropy index:

$$\text{Therapeutic Technique Diversity} = -\sum_{j=1}^m p_j (\ln p_j),$$

where  $m$  is the number of techniques used (i.e., MULTI item scores above 1),  $p_j$  is the proportional use of a given technique ( $j$ ). Scores range from 0 (i.e.,  $\ln [1]$ ) to  $\ln (m)$ . Higher scores indicate greater therapeutic technique diversity.

***Experiences in Close Relationship Scale Short Form (ECR-S; Wei et al., 2007)*** is a 12-item measure designed to assess attachment patterns in close relationships. We calculated subscales corresponding to anxious (six items) and avoidant (six items) attachment styles. This measure has shown strong factorial validity and convergent validity with measures drawn from the nomological network of insecure attachment styles (Wei et al., 2007). Internal consistency reliability for both subscales was acceptable ( $\alpha_s = .72$  and  $.85$ , for anxious and avoidant, respectively).

## PREDICTING THE THERAPIST EFFECT

*Childhood Trauma Questionnaire Short Form (Bernstein et al., 2003)* is a 28-item measure designed to assess adverse childhood experiences across five domains: emotional abuse, physical abuse, sexual abuse, emotional neglect, physical neglect. A global maltreatment scale was calculated by averaging across the 25 items which compose the five domains (three items are validity items; Gerdner & Allgulander, 2009). The measure has shown factorial validity and convergent validity with gold-standard assessments (Bernstein et al., 2003). Internal consistency reliability for the global maltreatment total score was high ( $\alpha = .89$ ).

### References

- Baer, R.A., Smith, G.T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment, 13*, 27–45.
- Benton, A. L. (1967). Problems of test construction in the field of aphasia. *Cortex, 3*, 32-58.
- Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., ... & Zule, W. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse & Neglect, 27*(2), 169-190.
- Bohlmeijer, E., ten Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. *Assessment, 18*(3), 308-320.
- Chen, R., Rafaeli, E., Ziv-Beiman, S., Bar-Kalifa, E., Solomonov, N., Barber, J. P., ... & Atzil-Slonim, D. (2020). Therapeutic technique diversity is linked to quality of working alliance and client functioning following alliance ruptures. *Journal of Consulting and Clinical Psychology, 88*(9), 844-858.
- Chow, D., Miller, S. D., Seidel, J., Kane, R.T., & Thornton, J.A., & Andrews, W.P. (2015). The role of deliberate practice in the development of highly effective psychotherapists.

## PREDICTING THE THERAPIST EFFECT

*Psychotherapy*, 52(3), 337-345. doi: 10.1037/pst0000015

- Cyranowski, J. M., Zill, N., Bode, R., Butt, Z., Kelly, M. A., Pilkonis, P. A., ... & Cella, D. (2013). Assessing social support, companionship, and distress: National Institute of Health (NIH) Toolbox Adult Social Relationship Scales. *Health Psychology*, 32(3), 293-301. doi: 10.1037/a0028586
- Davis, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113-126.
- Gerdner, A., & Allgulander, C. (2009). Psychometric properties of the swedish version of the childhood trauma Questionnaire-Short form (CTQ-SF). *Nordic Journal of Psychiatry*, 63(2), 160-170.
- Gosling, S.D., Rentfrow, P.J., & Swann, W.B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504-528.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197-216.
- Henry, J.D. & Crawford, J.R. (2005). The short-form version of the Depression Anxiety and Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 44, 227-239.
- Joyal, M., Wensing, T., Levasseur-Moreau, J., Leblond, J., T. Sack, A., & Fecteau, S. (2019).

## PREDICTING THE THERAPIST EFFECT

- Characterizing emotional Stroop interference in posttraumatic stress disorder, major depression and anxiety disorders: A systematic review and meta-analysis. *PLOS ONE*, *14*(4), e0214998.
- Kim, B. S., Cartwright, B. Y., Asay, P. A., & D'Andrea, M. J. (2003). A revision of the multicultural awareness, knowledge, and skills survey-counselor edition. *Measurement and Evaluation in Counseling and Development*, *36*(3), 161-180.
- Kral, T. R., Solis, E., Mumford, J. A., Schuyler, B. S., Flook, L., Rifken, K., ... & Davidson, R. J. (2017). Neural correlates of empathic accuracy in adolescence. *Social Cognitive and Affective Neuroscience*, *12*(11), 1701-1710. doi: 10.1093/scan/nsx099
- Nissen-Lie, H.A., Monsen, J.T., Ulleberg, P., & Rønnestad, M.H. (2013). Psychotherapists' self-reports of their interpersonal functioning and difficulties in practice as predictors of patient outcome. *Psychotherapy Research*, *23*(1), 86-104.  
doi:10.1080/10503307.2012.735775
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., ... & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36-88.
- Orlinsky, D., Ambühl, H., Rønnestad, M., Davis, J., Gerin, P., Davis, M., ... & SPR Collaborative Research Network. (1999). Development of psychotherapists: Concepts, questions, and methods of a collaborative international study. *Psychotherapy Research*, *9*(2), 127-153.
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171-192. doi: 10.1037/a0032734

## PREDICTING THE THERAPIST EFFECT

- Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the self-compassion scale. *Clinical Psychology & Psychotherapy, 18*(3), 250-255.
- Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1996). Benton controlled oral word association test: Reliability and updated norms. *Archives of Clinical Neuropsychology, 11*(4), 329-338.
- Ryff, C.D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*(6), 1069-1081.
- Ryff, C.D., & Keyes, C.L.M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*(4), 719-727.
- Schriber, R. A., Chung, J. M., Sorensen, K. S., & Robins, R. W. (2017). Dispositional contempt: A first look at the contemptuous person. *Journal of Personality and Social Psychology, 113*(2), 280-309.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379–423. doi: 10.1002/j .1538-7305.1948.tb01338.x
- Solomonov, N., McCarthy, K. S., Gorman, B. S., & Barber, J. P. (2019). The Multitheoretical List of Therapeutic Interventions–30 items (MULTI-30). *Psychotherapy Research, 29*(5), 565-580. doi:10.1080/10503307.2017.1422216
- Smith, P., & Waterman, M. (2003). Processing bias for aggression words in forensic and nonforensic samples. *Cognition & Emotion, 17*(5), 681-701.
- Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

## PREDICTING THE THERAPIST EFFECT

Wei, M., Russell, D. W., Mallinckrodt, B., & Vogel, D. L. (2007). The Experiences in Close Relationship Scale (ECR)-short form: Reliability, validity, and factor structure. *Journal of Personality Assessment, 88*(2), 187-204.

Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences, 106*(27), 11382-11387.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 2**

*Therapist-Level Predictors Descriptive Statistics*

Measure	<i>n</i>	Mean	<i>SD</i>	Min	Max
FIS total	85	3.36	0.32	2.51	4.18
FIS verbal fluency	85	3.23	0.48	1.70	4.26
Word fluency test	70	23.58	8.04	1.00	42.33
FIS hope and positive expectations	85	3.30	0.29	2.48	3.96
Depression, Anxiety, and Stress Scale	89	1.51	0.28	1.00	2.33
Psychological Well-being	89	5.28	0.61	3.61	6.33
FIS persuasiveness	85	3.40	0.40	2.10	4.32
Ten Item Personality Inventory extraversion	89	4.25	1.61	1.00	7.00
FIS emotional expression	85	3.42	0.36	2.43	4.27
Emotional Stroop task	70	16.94	108.01	-275.29	253.77
FIS warmth, acceptance, and understanding	85	3.38	0.39	2.26	4.15
Ten Item Personality Inventory agreeableness	89	5.76	0.91	2.50	7.00
Five Facet Mindfulness Questionnaire Short Form total score	89	3.63	0.40	2.17	4.50
Self-Compassion Scale Short Form total score	89	3.57	0.61	2.08	4.75
Dispositional Contempt	89	2.31	0.49	1.60	4.40
FIS empathy	85	3.32	0.39	2.39	4.14
Interpersonal Reactivity Index	89	3.50	0.38	2.68	4.54
Empathic accuracy task	83	0.48	0.12	-0.02	0.69
FIS alliance bond capacity	85	3.42	0.35	2.63	4.27
FIS alliance rupture and repair	85	3.45	0.38	2.37	4.40
NIH Toolbox Loneliness	89	1.97	0.70	1.00	4.60

PREDICTING THE THERAPIST EFFECT

NIH Toolbox Emotional Support	89	4.43	0.57	2.00	5.00
Multicultural Orientation task overall	86	3.41	0.41	2.25	4.23
Multicultural Orientation task comfort	86	3.65	0.34	2.25	4.28
Multicultural Orientation task humility	86	3.60	0.38	2.50	4.41
Multicultural Orientation task opportunity	86	1.94	0.67	1.08	3.62
Racial Implicit Association task	70	0.34	0.43	-0.70	1.10
Multicultural Knowledge, Awareness, and Skills Scale total score	89	2.81	0.26	2.15	3.52
Use of feminist and/or multicultural theory	97	0.40	0.49	0.00	1.00
Deliberate practice hours	83	5.75	7.21	0.00	50.00
Difficulties in Therapeutic Practice professional self-doubt	89	2.95	0.72	1.60	4.80
Years providing psychotherapy	96	4.26	6.51	0.00	40.00
Perceived efficacy	96	63.92	13.10	29.00	90.00
Years of personal therapy	94	0.85	0.36	0.00	1.00
Multitheoretical List of Interventions therapeutic technique diversity	89	2.18	0.04	2.07	2.26
Experiences in Close Relationships anxious	89	3.34	1.07	1.17	6.00
Experiences in Close Relationships avoidant	89	2.26	1.11	1.00	6.33
Childhood Trauma Questionnaire total	81	1.59	0.63	1.00	3.80

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health; *SD* = standard deviation; Min = minimum; Max = maximum.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 3**

*Patient Demographics Across Analytic Samples*

Demographic	Primary Sample	Outliers Removed	Therapists with $\geq 5$ Patients	Patients with $\geq 3$ Sessions	Clinically elevated at baseline	OQ-45	BHM-20	No Restricted Based on Time
<b>Age</b>	35.05 (SD = 12.63)	35.05 (SD = 12.63)	35.06 (SD = 12.63)	35.11 (SD = 12.73)	34.27 (SD = 12.31)	35.05 (SD = 12.63)	NA	33.97 (SD = 12.20)
<b>Gender</b>								
Woman	61.2	61.2	61.17	62.74	62.9	62.83	50.49	68.74
Man	33.31	33.31	33.4	33.67	30.76	34.69	24.2	26.37
Not reported	5.49	5.49	5.44	3.58	6.34	2.47	25.31	4.89
<b>Race/ethnicity</b>								
African	2.31	2.31	2.3	1.83	1.91	2.29	2.46	2.99
Asian	9.01	9.01	8.96	8.53	9.09	9.5	5.77	8.42
European	53.85	53.85	53.95	54.95	54.41	57.27	31.45	57.2
Indigenous	2.39	2.39	2.38	2.33	2.11	2.74	0.12	2.84
Latin American	2.19	2.19	2.2	2.15	2.04	2	3.44	2.65
Middle Eastern	1.25	1.25	1.26	1.23	1.08	1.39	0.37	1.22
Multiracial	0.13	0.13	0.13	0.05	0.14	NA	0.98	0.44
Other	4.1	4.1	4.1	3.82	4.09	3	11.3	17.72
Not reported	24.77	24.77	24.71	25.12	25.14	21.82	44.1	6.51

Note. OQ-45 = Outcome Questionnaire – 45; BHM-20 = Behavioral Health Monitor – 20.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 4**

*Pre-Post Model Results with Outliers Removed*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	4,728	84	NA	-0.0025	0.884	0.968
FIS verbal fluency	4,710	83	NA	-0.0470	0.014	0.532
Word fluency test	4,122	69	NA	0.0044	0.828	0.968
FIS hope and positive expectations	4,728	84	NA	-0.0005	0.976	0.976
Depression, Anxiety, and Stress Scale	4,964	88	NA	0.0290	0.077	0.968
Psychological Well-being	4,964	88	NA	-0.0212	0.214	0.968
FIS persuasiveness	4,710	83	NA	-0.0216	0.254	0.968
Ten Item Personality Inventory extraversion	4,964	88	NA	-0.0134	0.451	0.968
FIS emotional expression	4,728	84	NA	0.0159	0.380	0.968
Emotional Stroop task	3,941	69	NA	0.0069	0.728	0.968
FIS warmth, acceptance, and understanding	4,728	84	NA	0.0112	0.525	0.968
Ten Item Personality Inventory agreeableness	4,946	87	NA	-0.0087	0.627	0.968
Five Facet Mindfulness Questionnaire Short Form total score	4,943	87	NA	0.0009	0.962	0.976
Self-Compassion Scale Short Form total score	4,964	88	NA	0.0107	0.526	0.968
Dispositional Contempt	4,700	86	NA	0.0062	0.787	0.968
FIS empathy	4,728	84	NA	0.0097	0.581	0.968
Interpersonal Reactivity Index	4,964	88	NA	0.0050	0.772	0.968
Empathic accuracy task	4,352	80	NA	0.0023	0.917	0.968
FIS alliance bond capacity	4,728	84	NA	0.0026	0.870	0.968
FIS alliance rupture and repair	4,728	84	NA	0.0021	0.903	0.968
NIH Toolbox Loneliness	4,899	87	NA	0.0122	0.515	0.968
NIH Toolbox Emotional Support	4,943	87	NA	-0.0063	0.748	0.968
Multicultural Orientation task overall	4,826	85	NA	0.0138	0.462	0.968
Multicultural Orientation task comfort <sup>a</sup>	4,790	83	NA	0.0029	0.902	0.968

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	4,826	85	NA	0.0121	0.527	0.968
Multicultural Orientation task opportunity	4,826	85	NA	0.0029	0.862	0.968
Racial Implicit Association task	4,122	69	NA	-0.0091	0.617	0.968
Multicultural Knowledge, Awareness, and Skills Scale total score	4,964	88	NA	-0.0156	0.347	0.968
Use of feminist and/or multicultural theory	5,181	96	NA	0.0124	0.473	0.968
Deliberate practice hours	4,681	81	NA	0.0154	0.587	0.968
Difficulties in Therapeutic Practice professional self-doubt	4,964	88	NA	-0.0022	0.900	0.968
Years providing psychotherapy	5,144	93	NA	-0.0167	0.629	0.968
Perceived efficacy	5,171	95	NA	-0.0117	0.512	0.968
Years of personal therapy	5,047	93	NA	0.0103	0.516	0.968
Multitheoretical List of Interventions therapeutic technique diversity	4,964	88	NA	-0.0060	0.660	0.968
Experiences in Close Relationships anxious	4,964	88	NA	0.0068	0.691	0.968
Experiences in Close Relationships avoidant	4,933	87	NA	0.0082	0.666	0.968
Childhood Trauma Questionnaire total	4,611	78	NA	0.0367	0.072	0.968

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting pre-post changes in psychological distress; Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with larger pre-post reductions in psychological distress.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 5**

*Pre-Post Model Results Restricted to Therapists with  $\geq 5$  Patients*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	4,710	78	NA	-0.0009	0.959	0.985
FIS verbal fluency	4,710	78	NA	-0.0373	0.042	0.950
Word fluency test	4,108	64	NA	0.0057	0.781	0.950
FIS hope and positive expectations	4,710	78	NA	0.0003	0.987	0.987
Depression, Anxiety, and Stress Scale	4,946	82	NA	0.0293	0.075	0.950
Psychological Well-being	4,946	82	NA	-0.0221	0.197	0.950
FIS persuasiveness	4,710	78	NA	-0.013	0.470	0.950
Ten Item Personality Inventory extraversion	4,946	82	NA	-0.0109	0.539	0.950
FIS emotional expression	4,710	78	NA	0.0163	0.368	0.950
Emotional Stroop task	3,927	64	NA	0.0061	0.759	0.950
FIS warmth, acceptance, and understanding	4,710	78	NA	0.0122	0.492	0.950
Ten Item Personality Inventory agreeableness	4,946	82	NA	-0.0046	0.790	0.950
Five Facet Mindfulness Questionnaire Short Form total score	4,946	82	NA	0.0054	0.756	0.950
Self-Compassion Scale Short Form total score	4,946	82	NA	0.0111	0.512	0.950
Dispositional Contempt	4,946	82	NA	0.0217	0.152	0.950
FIS empathy	4,710	78	NA	0.0116	0.513	0.950
Interpersonal Reactivity Index	4,946	82	NA	0.0063	0.714	0.950
Empathic accuracy task	4,548	76	NA	-0.0033	0.850	0.950
FIS alliance bond capacity	4,710	78	NA	0.0045	0.783	0.950
FIS alliance rupture and repair	4,710	78	NA	0.0044	0.798	0.950
NIH Toolbox Loneliness	4,946	82	NA	0.0173	0.294	0.950
NIH Toolbox Emotional Support	4,946	82	NA	0.0012	0.946	0.985
Multicultural Orientation task overall	4,808	79	NA	0.0160	0.399	0.950
Multicultural Orientation task comfort	4,808	79	NA	0.0043	0.829	0.950

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	4,808	79	NA	0.0127	0.511	0.950
Multicultural Orientation task opportunity	4,808	79	NA	0.0035	0.831	0.950
Racial Implicit Association task	4,108	64	NA	-0.0093	0.609	0.950
Multicultural Knowledge, Awareness, and Skills Scale total score	4,946	82	NA	-0.0168	0.313	0.950
Use of feminist and/or multicultural theory	5,159	89	NA	0.0127	0.464	0.950
Deliberate practice hours	4,755	76	NA	0.0244	0.138	0.950
Difficulties in Therapeutic Practice professional self-doubt	4,946	82	NA	-0.0023	0.899	0.976
Years providing psychotherapy	5,149	88	NA	-0.0142	0.567	0.950
Perceived efficacy	5,149	88	NA	-0.0115	0.522	0.950
Years of personal therapy	5,025	86	NA	0.0114	0.472	0.950
Multitheoretical List of Interventions therapeutic technique diversity	4,946	82	NA	-0.0066	0.631	0.950
Experiences in Close Relationships anxious	4,946	82	NA	0.0066	0.700	0.950
Experiences in Close Relationships avoidant	4,946	82	NA	0.0036	0.839	0.950
Childhood Trauma Questionnaire total	4,687	75	NA	0.0291	0.097	0.950

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting pre-post changes in psychological distress; Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with larger pre-post reductions in psychological distress.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 6**

*Pre-Post Model Results Restricted to Patients with  $\geq 3$  Sessions*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	2,903	76	NA	0.0453	0.085	0.471
FIS verbal fluency	2,903	76	NA	-0.0034	0.902	0.926
Word fluency test	2,723	63	NA	0.0023	0.930	0.930
FIS hope and positive expectations	2,903	76	NA	0.0395	0.128	0.471
Depression, Anxiety, and Stress Scale	3,044	80	NA	0.0519	0.029	0.471
Psychological Well-being	3,044	80	NA	-0.0166	0.502	0.698
FIS persuasiveness	2,903	76	NA	0.0418	0.117	0.471
Ten Item Personality Inventory extraversion	3,044	80	NA	-0.0059	0.816	0.912
FIS emotional expression	2,903	76	NA	0.0691	0.013	0.471
Emotional Stroop task	2,574	63	NA	0.0298	0.226	0.477
FIS warmth, acceptance, and understanding	2,903	76	NA	0.0400	0.132	0.471
Ten Item Personality Inventory agreeableness	3,044	80	NA	0.0230	0.338	0.602
Five Facet Mindfulness Questionnaire Short Form total score	3,044	80	NA	0.0143	0.551	0.698
Self-Compassion Scale Short Form total score	3,044	80	NA	0.0172	0.465	0.698
Dispositional Contempt	3,044	80	NA	0.0294	0.164	0.471
FIS empathy	2,903	76	NA	0.0527	0.05	0.471
Interpersonal Reactivity Index	3,044	80	NA	0.0219	0.365	0.602
Empathic accuracy task	2,793	73	NA	0.0041	0.871	0.919
FIS alliance bond capacity	2,903	76	NA	0.0343	0.162	0.471
FIS alliance rupture and repair	2,903	76	NA	0.0453	0.083	0.471
NIH Toolbox Loneliness	3,044	80	NA	0.0318	0.174	0.471
NIH Toolbox Emotional Support	3,044	80	NA	-0.0073	0.772	0.912
Multicultural Orientation task overall	2,956	77	NA	0.0468	0.096	0.471
Multicultural Orientation task comfort	2,956	77	NA	0.0402	0.186	0.471

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	2,956	77	NA	0.0472	0.098	0.471
Multicultural Orientation task opportunity	2,956	77	NA	0.0291	0.212	0.474
Racial Implicit Association task	2,723	63	NA	-0.0251	0.300	0.600
Multicultural Knowledge, Awareness, and Skills Scale total score	3,044	80	NA	-0.0152	0.536	0.698
Use of feminist and/or multicultural theory	3,160	87	NA	0.0117	0.646	0.792
Deliberate practice hours	3,073	74	NA	0.0041	0.856	0.919
Difficulties in Therapeutic Practice professional self-doubt	3,044	80	NA	-0.0162	0.526	0.698
Years providing psychotherapy	3,153	86	NA	-0.0564	0.179	0.471
Perceived efficacy	3,153	86	NA	-0.0231	0.38	0.602
Years of personal therapy	3,085	84	NA	0.0282	0.203	0.474
Multitheoretical List of Interventions therapeutic technique diversity	3,044	80	NA	-0.0208	0.354	0.602
Experiences in Close Relationships anxious	3,044	80	NA	0.0224	0.35	0.602
Experiences in Close Relationships avoidant	3,044	80	NA	-0.0062	0.806	0.912
Childhood Trauma Questionnaire total	2,877	73	NA	0.0169	0.482	0.698

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting pre-post changes in psychological distress; Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with larger pre-post reductions in psychological distress.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 7**

*Pre-Post Model Results Restricted to Patients with Clinically Elevated Symptoms at Baseline*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	3,582	84	NA	-0.0089	0.66	0.968
FIS verbal fluency	3,582	84	NA	-0.0482	0.023	0.646
Word fluency test	3,119	69	NA	0.0104	0.654	0.968
FIS hope and positive expectations	3,582	84	NA	-0.0016	0.938	0.968
Depression, Anxiety, and Stress Scale	3,771	88	NA	0.0342	0.070	0.665
Psychological Well-being	3,771	88	NA	-0.0204	0.296	0.968
FIS persuasiveness	3,582	84	NA	-0.0275	0.185	0.926
Ten Item Personality Inventory extraversion	3,771	88	NA	-0.0263	0.195	0.926
FIS emotional expression	3,582	84	NA	0.0083	0.694	0.968
Emotional Stroop task	2,986	69	NA	-0.0037	0.870	0.968
FIS warmth, acceptance, and understanding	3,582	84	NA	0.0147	0.475	0.968
Ten Item Personality Inventory agreeableness	3,771	88	NA	-0.0114	0.562	0.968
Five Facet Mindfulness Questionnaire Short Form total score	3,771	88	NA	0.0014	0.945	0.968
Self-Compassion Scale Short Form total score	3,771	88	NA	0.0060	0.758	0.968
Dispositional Contempt	3,771	88	NA	0.0257	0.133	0.926
FIS empathy	3,582	84	NA	0.0045	0.826	0.968
Interpersonal Reactivity Index	3,771	88	NA	0.0050	0.799	0.968
Empathic accuracy task	3,462	82	NA	0.0050	0.800	0.968
FIS alliance bond capacity	3,582	84	NA	0.0014	0.941	0.968
FIS alliance rupture and repair	3,582	84	NA	-0.0039	0.841	0.968
NIH Toolbox Loneliness	3,771	88	NA	0.0269	0.153	0.926
NIH Toolbox Emotional Support	3,771	88	NA	-0.0012	0.955	0.968
Multicultural Orientation task overall	3,663	85	NA	0.0105	0.628	0.968
Multicultural Orientation task comfort	3,663	85	NA	0.0022	0.923	0.968

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	3,663	85	NA	0.0085	0.701	0.968
Multicultural Orientation task opportunity	3,663	85	NA	-0.0016	0.933	0.968
Racial Implicit Association task	3,119	69	NA	-0.0055	0.793	0.968
Multicultural Knowledge, Awareness, and Skills Scale total score	3,771	88	NA	-0.0226	0.235	0.968
Use of feminist and/or multicultural theory	3,936	96	NA	0.0092	0.647	0.968
Deliberate practice hours	3,616	82	NA	0.0395	0.034	0.646
Difficulties in Therapeutic Practice professional self-doubt	3,771	88	NA	0.0008	0.968	0.968
Years providing psychotherapy	3,928	95	NA	-0.0144	0.614	0.968
Perceived efficacy	3,928	95	NA	-0.0127	0.536	0.968
Years of personal therapy	3,829	93	NA	0.0084	0.646	0.968
Multitheoretical List of Interventions therapeutic technique diversity	3,771	88	NA	-0.0118	0.467	0.968
Experiences in Close Relationships anxious	3,771	88	NA	0.0058	0.764	0.968
Experiences in Close Relationships avoidant	3,771	88	NA	0.0032	0.874	0.968
Childhood Trauma Questionnaire total	3,578	80	NA	0.0382	0.054	0.665

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting pre-post changes in psychological distress; Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with larger pre-post reductions in psychological distress.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 8**

*Pre-Post Model Results Restricted to Outcome Questionnaire - 45*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	3,979	60	NA	-0.0084	0.689	0.971
FIS verbal fluency	3,979	60	NA	-0.0347	0.110	0.971
Word fluency test	3,830	60	NA	-0.0009	0.968	0.971
FIS hope and positive expectations	3,979	60	NA	-0.0044	0.830	0.971
Depression, Anxiety, and Stress Scale	4,215	64	NA	0.0339	0.086	0.971
Psychological Well-being	4,215	64	NA	-0.0158	0.444	0.971
FIS persuasiveness	3,979	60	NA	-0.0208	0.328	0.971
Ten Item Personality Inventory extraversion	4,215	64	NA	-0.0119	0.567	0.971
FIS emotional expression	3,979	60	NA	0.0160	0.495	0.971
Emotional Stroop task	3,649	60	NA	0.0061	0.773	0.971
FIS warmth, acceptance, and understanding	3,979	60	NA	0.0044	0.832	0.971
Ten Item Personality Inventory agreeableness	4,215	64	NA	0.0030	0.876	0.971
Five Facet Mindfulness Questionnaire Short Form total score	4,215	64	NA	0.0123	0.539	0.971
Self-Compassion Scale Short Form total score	4,215	64	NA	0.0145	0.469	0.971
Dispositional Contempt	4,215	64	NA	0.0236	0.167	0.971
FIS empathy	3,979	60	NA	0.0013	0.953	0.971
Interpersonal Reactivity Index	4,215	64	NA	0.0062	0.757	0.971
Empathic accuracy task	3,816	57	NA	0.0011	0.955	0.971
FIS alliance bond capacity	3,979	60	NA	-0.0048	0.807	0.971
FIS alliance rupture and repair	3,979	60	NA	-0.0040	0.850	0.971
NIH Toolbox Loneliness	4,215	64	NA	0.0227	0.240	0.971
NIH Toolbox Emotional Support	4,215	64	NA	-0.0033	0.873	0.971
Multicultural Orientation task overall	4,077	61	NA	0.0207	0.357	0.971
Multicultural Orientation task comfort	4,077	61	NA	0.0100	0.679	0.971

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	4,077	61	NA	0.0127	0.579	0.971
Multicultural Orientation task opportunity	4,077	61	NA	0.0060	0.755	0.971
Racial Implicit Association task	3,830	60	NA	-0.0039	0.845	0.971
Multicultural Knowledge, Awareness, and Skills Scale total score	4,215	64	NA	-0.0163	0.434	0.971
Use of feminist and/or multicultural theory	4,402	68	NA	0.0134	0.507	0.971
Deliberate practice hours	4,388	66	NA	0.0238	0.162	0.971
Difficulties in Therapeutic Practice professional self-doubt	4,215	64	NA	-0.0147	0.470	0.971
Years providing psychotherapy	4,392	67	NA	0.0364	0.282	0.971
Perceived efficacy	4,392	67	NA	0.0051	0.801	0.971
Years of personal therapy	4,268	65	NA	0.0124	0.456	0.971
Multitheoretical List of Interventions therapeutic technique diversity	4,215	64	NA	-0.0157	0.411	0.971
Experiences in Close Relationships anxious	4,215	64	NA	0.0007	0.971	0.971
Experiences in Close Relationships avoidant	4,215	64	NA	-0.0073	0.734	0.971
Childhood Trauma Questionnaire total	4,059	59	NA	0.0313	0.103	0.971

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting pre-post changes in psychological distress; Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to data coming from a single clinic. Negative estimates reflect associations with larger pre-post reductions in psychological distress. As the Outcome Questionnaire – 45 was only collected in one clinic, a clinic random intercept was not estimated.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 9**

*Pre-Post Model Results Restricted to Behavioral Health Monitor - 20*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	749	24	NA	0.0258	0.341	0.907
FIS verbal fluency	749	24	NA	-0.0521	0.085	0.701
Word fluency test	NA	NA	NA	NA	NA	NA
FIS hope and positive expectations	749	24	NA	0.0195	0.500	0.907
Depression, Anxiety, and Stress Scale	749	24	NA	0.0118	0.657	0.907
Psychological Well-being	NA	NA	NA	NA	NA	NA
FIS persuasiveness	749	24	NA	0.0114	0.704	0.907
Ten Item Personality Inventory extraversion	749	24	NA	-0.0194	0.550	0.907
FIS emotional expression	749	24	NA	0.0192	0.419	0.907
Emotional Stroop task	292	9	NA	0.0204	0.742	0.907
FIS warmth, acceptance, and understanding	NA	NA	NA	NA	NA	NA
Ten Item Personality Inventory agreeableness	749	24	NA	-0.0717	0.114	0.752
Five Facet Mindfulness Questionnaire Short Form total score	749	24	NA	-0.0338	0.279	0.907
Self-Compassion Scale Short Form total score	749	24	NA	-0.0029	0.921	0.961
Dispositional Contempt	749	24	NA	0.0029	0.932	0.961
FIS empathy	NA	NA	NA	NA	NA	NA
Interpersonal Reactivity Index	749	24	NA	0.0003	0.992	0.992
Empathic accuracy task	750	25	NA	-0.0224	0.540	0.907
FIS alliance bond capacity	NA	NA	NA	NA	NA	NA
FIS alliance rupture and repair	749	24	NA	0.0276	0.254	0.907
NIH Toolbox Loneliness	749	24	NA	-0.006	0.841	0.957
NIH Toolbox Emotional Support	749	24	NA	0.0318	0.38	0.907
Multicultural Orientation task overall	749	24	NA	-0.0114	0.719	0.907
Multicultural Orientation task comfort	749	24	NA	-0.0076	0.800	0.943

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	749	24	NA	0.0123	0.702	0.907
Multicultural Orientation task opportunity	749	24	NA	-0.0119	0.693	0.907
Racial Implicit Association task	292	9	NA	-0.0369	0.414	0.907
Multicultural Knowledge, Awareness, and Skills Scale total score	749	24	NA	-0.0166	0.478	0.907
Use of feminist and/or multicultural theory	779	28	NA	-0.0066	0.873	0.960
Deliberate practice hours	385	16	NA	0.1183	0.430	0.907
Difficulties in Therapeutic Practice professional self-doubt	749	24	NA	0.0717	0.041	0.451
Years providing psychotherapy	779	28	NA	-0.0802	0.019	0.451
Perceived efficacy	779	28	NA	-0.0945	0.030	0.451
Years of personal therapy	779	28	NA	-0.0362	0.639	0.907
Multitheoretical List of Interventions therapeutic technique diversity	749	24	NA	-0.0185	0.468	0.907
Experiences in Close Relationships anxious	749	24	NA	0.0483	0.153	0.792
Experiences in Close Relationships avoidant	749	24	NA	0.0349	0.168	0.792
Childhood Trauma Questionnaire total	645	21	NA	0.0214	0.636	0.907

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting pre-post changes in psychological distress; Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence, rows with all NAs did not converge with therapist level so were not reported. Negative estimates reflect associations with larger pre-post reductions in psychological distress.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 10**

*Pre-Post Model Results Covarying Time Since Therapist Assessment*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	4,728	84	NA	-0.0054	0.756	0.999
FIS verbal fluency	4,728	84	NA	-0.0403	0.028	0.988
Word fluency test	4,122	69	NA	0.0054	0.791	0.999
FIS hope and positive expectations	4,728	84	NA	-0.0027	0.877	0.999
Depression, Anxiety, and Stress Scale	4,964	88	NA	0.0273	0.101	0.999
Psychological Well-being	4,964	88	NA	-0.0224	0.192	0.999
FIS persuasiveness	4,728	84	NA	-0.0202	0.264	0.999
Ten Item Personality Inventory extraversion	4,964	88	NA	-0.0142	0.427	0.999
FIS emotional expression	4,728	84	NA	0.0141	0.436	0.999
Emotional Stroop task	3,941	69	NA	0.0023	0.910	0.999
FIS warmth, acceptance, and understanding	4,728	84	NA	0.0111	0.532	0.999
Ten Item Personality Inventory agreeableness	4,964	88	NA	-0.0035	0.839	0.999
Five Facet Mindfulness Questionnaire Short Form total score	4,964	88	NA	0.0016	0.928	0.999
Self-Compassion Scale Short Form total score	4,964	88	NA	0.0066	0.701	0.999
Dispositional Contempt	4,964	88	NA	0.0241	0.114	0.999
FIS empathy	4,728	84	NA	0.0066	0.710	0.999
Interpersonal Reactivity Index	4,964	88	NA	0.0072	0.680	0.999
Empathic accuracy task	4,566	82	NA	-0.0017	0.922	0.999
FIS alliance bond capacity	4,728	84	NA	0.0002	0.989	0.999
FIS alliance rupture and repair	4,728	84	NA	-0.0002	0.992	0.999
NIH Toolbox Loneliness	4,964	88	NA	0.0197	0.236	0.999
NIH Toolbox Emotional Support	4,964	88	NA	-0.0017	0.926	0.999
Multicultural Orientation task overall	4,826	85	NA	0.0109	0.567	0.999
Multicultural Orientation task comfort	4,826	85	NA	0.0055	0.780	0.999

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	4,826	85	NA	0.0085	0.661	0.999
Multicultural Orientation task opportunity	4,826	85	NA	-0.0016	0.922	0.999
Racial Implicit Association task	4,122	69	NA	-0.0144	0.436	0.999
Multicultural Knowledge, Awareness, and Skills Scale total score	4,964	88	NA	-0.016	0.339	0.999
Use of feminist and/or multicultural theory	5,181	96	NA	0.0119	0.494	0.999
Deliberate practice hours	4,773	82	NA	0.0249	0.134	0.999
Difficulties in Therapeutic Practice professional self-doubt	4,964	88	NA	0.0000	0.999	0.999
Years providing psychotherapy	5,171	95	NA	-0.0162	0.514	0.999
Perceived efficacy	5,171	95	NA	-0.0096	0.592	0.999
Years of personal therapy	5,047	93	NA	0.0128	0.420	0.999
Multitheoretical List of Interventions therapeutic technique diversity	4,964	88	NA	-0.0061	0.656	0.999
Experiences in Close Relationships anxious	4,964	88	NA	0.0070	0.684	0.999
Experiences in Close Relationships avoidant	4,964	88	NA	0.0071	0.685	0.999
Childhood Trauma Questionnaire total	4,704	80	NA	0.0342	0.052	0.988

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting pre-post changes in psychological distress; Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with larger pre-post reductions in psychological distress.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 11**

*Pre-Post Model Results Not Restricting Based on Time Since Therapist Assessment*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	12,218	94	12	-0.0056	0.690	0.984
FIS verbal fluency	12,218	94	12	-0.0226	0.105	0.708
Word fluency test	9,614	80	9	0.0080	0.596	0.984
FIS hope and positive expectations	12,218	94	12	-0.0060	0.669	0.984
Depression, Anxiety, and Stress Scale	12,905	98	12	0.0191	0.146	0.708
Psychological Well-being	12,905	98	12	0.0005	0.973	0.997
FIS persuasiveness	12,218	94	12	-0.0120	0.398	0.984
Ten Item Personality Inventory extraversion	12,905	98	12	-0.0034	0.812	0.984
FIS emotional expression	12,218	94	12	0.0081	0.566	0.984
Emotional Stroop task	9,482	80	NA	-0.0067	0.655	0.984
FIS warmth, acceptance, and understanding	12,218	94	12	0.0052	0.698	0.984
Ten Item Personality Inventory agreeableness	12,905	98	12	0.0089	0.529	0.984
Five Facet Mindfulness Questionnaire Short Form total score	12,905	98	12	-0.0023	0.867	0.984
Self-Compassion Scale Short Form total score	12,905	98	12	0.0048	0.718	0.984
Dispositional Contempt	12,905	98	12	-0.0001	0.996	0.997
FIS empathy	12,218	94	12	0.0027	0.847	0.984
Interpersonal Reactivity Index	12,905	98	12	0.0119	0.387	0.984
Empathic accuracy task	12,082	90	13	0.0024	0.866	0.984
FIS alliance bond capacity	12,218	94	12	-0.0004	0.978	0.997
FIS alliance rupture and repair	12,218	94	12	-0.0088	0.530	0.984
NIH Toolbox Loneliness	12,905	98	12	0.0001	0.997	0.997
NIH Toolbox Emotional Support	12,905	98	12	0.0225	0.108	0.708
Multicultural Orientation task overall	12,334	95	12	-0.0240	0.089	0.708
Multicultural Orientation task comfort	12,334	95	12	-0.0193	0.178	0.752

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	12,334	95	12	-0.0179	0.216	0.821
Multicultural Orientation task opportunity	12,334	95	12	-0.0218	0.101	0.708
Racial Implicit Association task	9,614	80	9	-0.0111	0.426	0.984
Multicultural Knowledge, Awareness, and Skills Scale total score	12,905	98	12	-0.0035	0.794	0.984
Use of feminist and/or multicultural theory	14,123	109	13	0.0021	0.880	0.984
Deliberate practice hours	11,759	94	12	0.0137	0.335	0.984
Difficulties in Therapeutic Practice professional self-doubt	12,905	98	12	0.0135	0.339	0.984
Years providing psychotherapy	14,054	107	13	-0.0230	0.139	0.708
Perceived efficacy	14,054	107	13	-0.0332	0.022	0.708
Years of personal therapy	13,880	105	13	-0.0070	0.603	0.984
Multitheoretical List of Interventions therapeutic technique diversity	12,905	98	12	-0.0019	0.780	0.984
Experiences in Close Relationships anxious	12,905	98	12	0.0144	0.302	0.984
Experiences in Close Relationships avoidant	12,905	98	12	-0.0200	0.149	0.708
Childhood Trauma Questionnaire total	11,758	88	11	0.0043	0.763	0.984

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting pre-post changes in psychological distress; Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with larger pre-post reductions in psychological distress.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 12**

*Rate of Change Model Results with Outliers Removed*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Level 4 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	23,719	5,426	84	12	0.0287	0.052	0.395
FIS verbal fluency	23,614	5,404	83	12	-0.0025	0.880	0.880
Word fluency test	21,968	4,763	69	8	0.0060	0.719	0.781
FIS hope and positive expectations	23,719	5,426	84	12	0.0225	0.123	0.400
Depression, Anxiety, and Stress Scale	24,978	5,712	88	12	0.0279	0.041	0.390
Psychological Well-being	24,978	5,712	88	12	-0.0157	0.270	0.515
FIS persuasiveness	23,614	5,404	83	12	0.0244	0.130	0.400
Ten Item Personality Inventory extraversion	24,978	5,712	88	12	-0.0120	0.403	0.635
FIS emotional expression	23,719	5,426	84	12	0.0387	0.013	0.247
Emotional Stroop task	21,044	4,605	69	8	0.0194	0.221	0.515
FIS warmth, acceptance, and understanding	23,719	5,426	84	12	0.0172	0.243	0.515
Ten Item Personality Inventory agreeableness	24,873	5,690	87	12	0.0115	0.426	0.635
Five Facet Mindfulness Questionnaire Short Form total score	24,841	5,691	87	12	0.0077	0.611	0.726
Self-Compassion Scale Short Form total score	24,978	5,712	88	12	0.0156	0.253	0.515
Dispositional Contempt	23,481	5,398	86	12	-0.0084	0.662	0.762
FIS empathy	23,719	5,426	84	12	0.0375	0.011	0.247
Interpersonal Reactivity Index	24,978	5,712	88	12	0.0242	0.074	0.400
Empathic accuracy task	21,720	4,927	80	13	0.0050	0.800	0.844
FIS alliance bond capacity	23,719	5,426	84	12	0.0206	0.137	0.400
FIS alliance rupture and repair	23,719	5,426	84	12	0.0316	0.029	0.367
NIH Toolbox Loneliness	24,509	5,645	87	12	0.0110	0.486	0.660
NIH Toolbox Emotional Support	24,876	5,689	87	12	-0.0085	0.599	0.726
Multicultural Orientation task overall	24,115	5,541	85	12	0.0277	0.078	0.400
Multicultural Orientation task comfort	23,984	5,501	83	12	0.0218	0.271	0.515

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	24,115	5,541	85	12	0.0243	0.117	0.400
Multicultural Orientation task opportunity	24,115	5,541	85	12	0.0229	0.086	0.400
Racial Implicit Association task	21,968	4,763	69	8	-0.0131	0.387	0.635
Multicultural Knowledge, Awareness, and Skills Scale total score	24,978	5,712	88	12	0.0058	0.685	0.766
Use of feminist and/or multicultural theory	25,957	5,963	96	13	0.0134	0.326	0.590
Deliberate practice hours	24,415	5,427	81	12	-0.0187	0.408	0.635
Difficulties in Therapeutic Practice professional self-doubt	24,978	5,712	88	12	-0.0090	0.533	0.684
Years providing psychotherapy	25,833	5,917	93	12	0.0389	0.109	0.400
Perceived efficacy	25,904	5,948	95	13	-0.0028	0.841	0.864
Years of personal therapy	25,386	5,807	93	13	0.0178	0.149	0.404
Multitheoretical List of Interventions therapeutic technique diversity	24,978	5,712	88	12	-0.0157	0.217	0.515
Experiences in Close Relationships anxious	24,978	5,712	88	12	0.0109	0.436	0.635
Experiences in Close Relationships avoidant	24,805	5,679	87	12	0.0120	0.451	0.635
Childhood Trauma Questionnaire total	23,071	5,336	78	11	0.0102	0.540	0.684

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = interaction between each predictor variable and log session number from multilevel models predicting psychological distress; Level 1 = session level; Level 2 = patient level; Level 3 = therapist level; Level 4 = clinic level. Negative estimates reflect associations with faster reductions in psychological distress over time.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 13**

*Rate of Change Model Results Restricted to Therapists with  $\geq 5$  Patients*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Level 4 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	23,688	5,407	78	11	0.0281	0.057	0.370
FIS verbal fluency	23,688	5,407	78	11	0.0036	0.813	0.847
Word fluency test	21,941	4,748	64	7	0.0065	0.698	0.829
FIS hope and positive expectations	23,688	5,407	78	11	0.0218	0.134	0.383
Depression, Anxiety, and Stress Scale	24,947	5,693	82	11	0.0283	0.038	0.361
Psychological Well-being	24,947	5,693	82	11	-0.0164	0.250	0.494
FIS persuasiveness	23,688	5,407	78	11	0.0264	0.077	0.370
Ten Item Personality Inventory extraversion	24,947	5,693	82	11	-0.0118	0.409	0.598
FIS emotional expression	23,688	5,407	78	11	0.0378	0.016	0.304
Emotional Stroop task	21,017	4,590	64	7	0.0195	0.218	0.487
FIS warmth, acceptance, and understanding	23,688	5,407	78	11	0.0159	0.280	0.507
Ten Item Personality Inventory agreeableness	24,947	5,693	82	11	0.0153	0.260	0.494
Five Facet Mindfulness Questionnaire Short Form total score	24,947	5,693	82	11	0.0040	0.771	0.847
Self-Compassion Scale Short Form total score	24,947	5,693	82	11	0.0156	0.256	0.494
Dispositional Contempt	24,947	5,693	82	11	0.0121	0.341	0.543
FIS empathy	23,688	5,407	78	11	0.0370	0.012	0.304
Interpersonal Reactivity Index	24,947	5,693	82	11	0.0241	0.076	0.370
Empathic accuracy task	22,891	5,216	76	12	0.0016	0.915	0.915
FIS alliance bond capacity	23,688	5,407	78	11	0.0204	0.141	0.383
FIS alliance rupture and repair	23,688	5,407	78	11	0.0316	0.029	0.361
NIH Toolbox Loneliness	24,947	5,693	82	11	0.0171	0.210	0.487
NIH Toolbox Emotional Support	24,947	5,693	82	11	-0.0043	0.765	0.847
Multicultural Orientation task overall	24,084	5,522	79	11	0.0277	0.078	0.37
Multicultural Orientation task comfort	24,084	5,522	79	11	0.0250	0.130	0.383

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	24,084	5,522	79	11	0.0241	0.121	0.383
Multicultural Orientation task opportunity	24,084	5,522	79	11	0.0222	0.096	0.383
Racial Implicit Association task	21,941	4,748	64	7	-0.0134	0.377	0.573
Multicultural Knowledge, Awareness, and Skills Scale total score	24,947	5,693	82	11	0.0062	0.665	0.815
Use of feminist and/or multicultural theory	25,920	5,940	89	12	0.0131	0.335	0.543
Deliberate practice hours	24,726	5,524	76	10	-0.0130	0.343	0.543
Difficulties in Therapeutic Practice professional self-doubt	24,947	5,693	82	11	-0.0084	0.561	0.711
Years providing psychotherapy	25,867	5,925	88	12	-0.0047	0.812	0.847
Perceived efficacy	25,867	5,925	88	12	-0.0031	0.825	0.847
Years of personal therapy	25,349	5,784	86	12	0.0181	0.141	0.383
Multitheoretical List of Interventions therapeutic technique diversity	24,947	5,693	82	11	-0.0163	0.201	0.487
Experiences in Close Relationships anxious	24,947	5,693	82	11	0.0101	0.473	0.666
Experiences in Close Relationships avoidant	24,947	5,693	82	11	0.0091	0.524	0.687
Childhood Trauma Questionnaire total	23,518	5,413	75	10	0.0096	0.493	0.669

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = interaction between each predictor variable and log session number from multilevel models predicting psychological distress; Level 1 = session level; Level 2 = patient level; Level 3 = therapist level; Level 4 = clinic level. Negative estimates reflect associations with faster reductions in psychological distress over time.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 14**

*Rate of Change Model Results Restricted to Patients with  $\geq 3$  Sessions*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Level 4 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	21,063	3,469	78	10	0.0354	0.024	0.228
FIS verbal fluency	21,063	3,469	78	10	0.0087	0.592	0.750
Word fluency test	19,877	3,246	64	6	0.0053	0.757	0.846
FIS hope and positive expectations	21,063	3,469	78	10	0.0280	0.071	0.245
Depression, Anxiety, and Stress Scale	22,171	3,653	82	10	0.0284	0.048	0.228
Psychological Well-being	22,171	3,653	82	10	-0.0082	0.582	0.750
FIS persuasiveness	21,063	3,469	78	10	0.0324	0.042	0.228
Ten Item Personality Inventory extraversion	22,171	3,653	82	10	-0.0092	0.537	0.750
FIS emotional expression	21,063	3,469	78	10	0.0445	0.008	0.152
Emotional Stroop task	18,995	3,106	64	6	0.0204	0.205	0.458
FIS warmth, acceptance, and understanding	21,063	3,469	78	10	0.0239	0.126	0.342
Ten Item Personality Inventory agreeableness	22,171	3,653	82	10	0.0195	0.170	0.404
Five Facet Mindfulness Questionnaire Short Form total score	22,171	3,653	82	10	0.0030	0.836	0.869
Self-Compassion Scale Short Form total score	22,171	3,653	82	10	0.0160	0.264	0.528
Dispositional Contempt	22,171	3,653	82	10	0.0116	0.378	0.613
FIS empathy	21,063	3,469	78	10	0.0443	0.005	0.152
Interpersonal Reactivity Index	22,171	3,653	82	10	0.0239	0.091	0.266
Empathic accuracy task	20,351	3,336	75	10	0.0031	0.846	0.869
FIS alliance bond capacity	21,063	3,469	78	10	0.0273	0.067	0.245
FIS alliance rupture and repair	21,063	3,469	78	10	0.0369	0.017	0.215
NIH Toolbox Loneliness	22,171	3,653	82	10	0.0122	0.399	0.613
NIH Toolbox Emotional Support	22,171	3,653	82	10	0.0003	0.985	0.985
Multicultural Orientation task overall	21,392	3,539	79	10	0.0331	0.047	0.228
Multicultural Orientation task comfort	21,392	3,539	79	10	0.0304	0.084	0.266

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	21,392	3,539	79	10	0.0320	0.051	0.228
Multicultural Orientation task opportunity	21,392	3,539	79	10	0.0271	0.054	0.228
Racial Implicit Association task	19,877	3,246	64	6	-0.0185	0.236	0.498
Multicultural Knowledge, Awareness, and Skills Scale total score	22,171	3,653	82	10	0.0089	0.552	0.750
Use of feminist and/or multicultural theory	23,014	3,790	89	10	0.0119	0.403	0.613
Deliberate practice hours	22,228	3,695	76	9	-0.0122	0.393	0.613
Difficulties in Therapeutic Practice professional self-doubt	22,171	3,653	82	10	-0.0131	0.387	0.613
Years providing psychotherapy	22,966	3,779	88	10	-0.0059	0.779	0.846
Perceived efficacy	22,966	3,779	88	10	-0.0050	0.735	0.846
Years of personal therapy	22,531	3,694	86	10	0.0179	0.167	0.404
Multitheoretical List of Interventions therapeutic technique diversity	22,171	3,653	82	10	-0.0131	0.340	0.613
Experiences in Close Relationships anxious	22,171	3,653	82	10	0.0080	0.586	0.750
Experiences in Close Relationships avoidant	22,171	3,653	82	10	0.0042	0.778	0.846
Childhood Trauma Questionnaire total	20,872	3,467	75	9	0.0063	0.665	0.815

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = interaction between each predictor variable and log session number from multilevel models predicting psychological distress; Level 1 = session level; Level 2 = patient level; Level 3 = therapist level; Level 4 = clinic level. Negative estimates reflect associations with faster reductions in psychological distress over time.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 15**

*Rate of Change Model Results Restricted to Patients with Clinically Elevated Symptoms at Baseline*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Level 4 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	18,295	3,956	84	12	0.0249	0.079	0.390
FIS verbal fluency	18,295	3,956	84	12	0.0022	0.884	0.960
Word fluency test	16,912	3,457	69	NA	0.0064	0.697	0.829
FIS hope and positive expectations	18,295	3,956	84	12	0.0205	0.141	0.390
Depression, Anxiety, and Stress Scale	19,354	4,175	88	12	0.0327	0.011	0.314
Psychological Well-being	19,354	4,175	88	12	-0.0139	0.311	0.532
FIS persuasiveness	18,295	3,956	84	12	0.0219	0.129	0.390
Ten Item Personality Inventory extraversion	19,354	4,175	88	NA	-0.0129	0.311	0.532
FIS emotional expression	18,295	3,956	84	NA	0.0301	0.051	0.386
Emotional Stroop task	16,302	3,337	69	8	0.0152	0.322	0.532
FIS warmth, acceptance, and understanding	18,295	3,956	84	12	0.0171	0.223	0.498
Ten Item Personality Inventory agreeableness	19,354	4,175	88	NA	0.0061	0.645	0.817
Five Facet Mindfulness Questionnaire Short Form total score	19,354	4,175	88	NA	0.0064	0.637	0.817
Self-Compassion Scale Short Form total score	19,354	4,175	88	NA	0.0134	0.320	0.532
Dispositional Contempt	19,354	4,175	88	NA	0.0135	0.252	0.504
FIS empathy	18,295	3,956	84	NA	0.0334	0.020	0.314
Interpersonal Reactivity Index	19,354	4,175	88	NA	0.0213	0.106	0.390
Empathic accuracy task	17,630	3,812	82	NA	0.0051	0.720	0.829
FIS alliance bond capacity	18,295	3,956	84	12	0.0196	0.136	0.390
FIS alliance rupture and repair	18,295	3,956	84	12	0.0305	0.027	0.314
NIH Toolbox Loneliness	19,354	4,175	88	NA	0.0274	0.033	0.314
NIH Toolbox Emotional Support	19,354	4,175	88	12	-0.0066	0.644	0.817
Multicultural Orientation task overall	18,623	4,047	85	12	0.0236	0.119	0.390
Multicultural Orientation task comfort	18,623	4,047	85	NA	0.0192	0.239	0.504

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	18,623	4,047	85	NA	0.0191	0.204	0.484
Multicultural Orientation task opportunity	18,623	4,047	85	12	0.0179	0.153	0.390
Racial Implicit Association task	16,912	3,457	69	8	0.0004	0.976	0.994
Multicultural Knowledge, Awareness, and Skills Scale total score	19,354	4,175	88	NA	0.0001	0.994	0.994
Use of feminist and/or multicultural theory	20,113	4,356	96	13	0.0085	0.529	0.761
Deliberate practice hours	19,143	4,032	82	12	-0.0051	0.702	0.829
Difficulties in Therapeutic Practice professional self-doubt	19,354	4,175	88	NA	-0.0009	0.950	0.994
Years providing psychotherapy	20,076	4,345	95	13	-0.0137	0.508	0.761
Perceived efficacy	20,076	4,345	95	13	-0.0032	0.818	0.914
Years of personal therapy	19,658	4,236	93	13	0.0165	0.154	0.390
Multitheoretical List of Interventions therapeutic technique diversity	19,354	4,175	88	12	-0.0229	0.061	0.386
Experiences in Close Relationships anxious	19,354	4,175	88	NA	0.0110	0.416	0.659
Experiences in Close Relationships avoidant	19,354	4,175	88	NA	0.0088	0.541	0.761
Childhood Trauma Questionnaire total	18,232	3,971	80	11	0.0202	0.146	0.390

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = interaction between each predictor variable and log session number from multilevel models predicting psychological distress; Level 1 = session level; Level 2 = patient level; Level 3 = therapist level; Level 4 = clinic level; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with faster reductions in psychological distress over time.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 16**

*Rate of Change Model Results Restricted to Outcome Questionnaire - 45*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Level 4 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	21,659	4,661	60	NA	0.0144	0.292	0.686
FIS verbal fluency	21,659	4,661	60	NA	0.0097	0.500	0.709
Word fluency test	20,991	4,462	60	NA	-0.0036	0.800	0.857
FIS hope and positive expectations	21,659	4,661	60	NA	0.0122	0.364	0.686
Depression, Anxiety, and Stress Scale	22,918	4,947	64	NA	0.0263	0.045	0.686
Psychological Well-being	22,918	4,947	64	NA	-0.0085	0.531	0.709
FIS persuasiveness	21,659	4,661	60	NA	0.0156	0.264	0.686
Ten Item Personality Inventory extraversion	22,918	4,947	64	NA	-0.0110	0.426	0.686
FIS emotional expression	21,659	4,661	60	NA	0.0197	0.194	0.686
Emotional Stroop task	20,067	4,304	60	NA	0.0162	0.243	0.686
FIS warmth, acceptance, and understanding	21,659	4,661	60	NA	0.0032	0.812	0.857
Ten Item Personality Inventory agreeableness	22,918	4,947	64	NA	0.0129	0.302	0.686
Five Facet Mindfulness Questionnaire Short Form total score	22,918	4,947	64	NA	0.0122	0.354	0.686
Self-Compassion Scale Short Form total score	22,918	4,947	64	NA	0.0111	0.401	0.686
Dispositional Contempt	22,918	4,947	64	NA	0.0133	0.236	0.686
FIS empathy	21,659	4,661	60	NA	0.0223	0.123	0.686
Interpersonal Reactivity Index	22,918	4,947	64	NA	0.0163	0.200	0.686
Empathic accuracy task	20,823	4,469	57	NA	0.0002	0.986	0.986
FIS alliance bond capacity	21,659	4,661	60	NA	0.0050	0.694	0.824
FIS alliance rupture and repair	21,659	4,661	60	NA	0.0157	0.255	0.686
NIH Toolbox Loneliness	22,918	4,947	64	NA	0.0248	0.047	0.686
NIH Toolbox Emotional Support	22,918	4,947	64	NA	-0.0078	0.560	0.709
Multicultural Orientation task overall	22,055	4,776	61	NA	0.0151	0.303	0.686
Multicultural Orientation task comfort	22,055	4,776	61	NA	0.0181	0.247	0.686

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	22,055	4,776	61	NA	0.0115	0.436	0.686
Multicultural Orientation task opportunity	22,055	4,776	61	NA	0.0117	0.339	0.686
Racial Implicit Association task	20,991	4,462	60	NA	0.0081	0.542	0.709
Multicultural Knowledge, Awareness, and Skills Scale total score	22,918	4,947	64	NA	0.0082	0.560	0.709
Use of feminist and/or multicultural theory	23,644	5,168	68	NA	0.0021	0.874	0.898
Deliberate practice hours	23,570	5,146	66	NA	-0.0092	0.451	0.686
Difficulties in Therapeutic Practice professional self-doubt	22,918	4,947	64	NA	-0.0127	0.354	0.686
Years providing psychotherapy	23,591	5,153	67	NA	-0.0068	0.793	0.857
Perceived efficacy	23,591	5,153	67	NA	-0.0074	0.590	0.723
Years of personal therapy	23,073	5,012	65	NA	0.0192	0.081	0.686
Multitheoretical List of Interventions therapeutic technique diversity	22,918	4,947	64	NA	-0.0155	0.213	0.686
Experiences in Close Relationships anxious	22,918	4,947	64	NA	0.0113	0.381	0.686
Experiences in Close Relationships avoidant	22,918	4,947	64	NA	0.0038	0.784	0.857
Childhood Trauma Questionnaire total	21,891	4,773	59	NA	0.0139	0.297	0.686

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = interaction between each predictor variable and log session number from multilevel models predicting psychological distress; Level 1 = session level; Level 2 = patient level; Level 3 = therapist level; Level 4 = clinic level; NA = random intercept omitted due to data coming from a single clinic. Negative estimates reflect associations with faster reductions in psychological distress over time. As the Outcome Questionnaire – 45 was only collected in one clinic, a clinic random intercept was not estimated.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 17**

*Rate of Change Model Results Restricted to Behavioral Health Monitor - 20*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Level 4 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	2,060	765	24	NA	0.1254	<.001	<.001
FIS verbal fluency	2,060	765	24	NA	-0.0654	0.135	0.233
Word fluency test	977	301	9	7	0.3524	<.001	<.001
FIS hope and positive expectations	2,060	765	24	NA	0.1049	0.005	0.014
Depression, Anxiety, and Stress Scale	2,060	765	24	NA	0.0213	0.426	0.540
Psychological Well-being	2,060	765	24	NA	-0.0407	0.216	0.316
FIS persuasiveness	2,060	765	24	NA	0.1298	0.001	0.005
Ten Item Personality Inventory extraversion	2,060	765	24	NA	0.0398	0.312	0.439
FIS emotional expression	2,060	765	24	NA	0.1307	<.001	<.001
Emotional Stroop task	977	301	9	NA	0.1024	0.048	0.107
FIS warmth, acceptance, and understanding	2,060	765	24	NA	0.1112	0.005	0.014
Ten Item Personality Inventory agreeableness	2,060	765	24	NA	0.0059	0.928	0.953
Five Facet Mindfulness Questionnaire Short Form total score	2,060	765	24	NA	-0.0229	0.482	0.583
Self-Compassion Scale Short Form total score	2,060	765	24	NA	0.0593	0.022	0.052
Dispositional Contempt	2,060	765	24	NA	-0.0203	0.662	0.719
FIS empathy	2,060	765	24	NA	0.0954	0.001	0.005
Interpersonal Reactivity Index	2,060	765	24	NA	0.1017	0.015	0.038
Empathic accuracy task	2,099	766	25	12	0.0286	0.491	0.583
FIS alliance Bond capacity	2,060	765	24	NA	0.0903	0.002	0.008
FIS alliance rupture and repair	2,060	765	24	NA	0.1131	<.001	<.001
NIH Toolbox Loneliness	2,060	765	24	NA	-0.0563	0.106	0.192
NIH Toolbox Emotional Support	2,060	765	24	NA	0.0492	0.329	0.447
Multicultural Orientation task overall	2,060	765	24	NA	0.1566	<.001	<.001
Multicultural Orientation task comfort	2,060	765	24	NA	0.0774	0.104	0.192

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	2,060	765	24	NA	0.1343	0.003	0.010
Multicultural Orientation task opportunity	2,060	765	24	NA	0.1140	0.001	0.005
Racial Implicit Association task	977	301	9	NA	-0.0923	0.003	0.010
Multicultural Knowledge, Awareness, and Skills Scale total score	2,060	765	24	NA	-0.0438	0.150	0.236
Use of feminist and/or multicultural theory	2,313	795	28	NA	0.0955	0.004	0.013
Deliberate practice hours	1,189	397	16	NA	-0.3928	0.076	0.160
Difficulties in Therapeutic Practice professional self-doubt	2,060	765	24	NA	0.0917	0.098	0.192
Years providing psychotherapy	2,313	795	28	NA	-0.0095	0.773	0.816
Perceived efficacy	2,313	795	28	NA	0.0429	0.362	0.474
Years of personal therapy	2,313	795	28	NA	-0.1559	0.151	0.236
Multitheoretical List of Interventions therapeutic technique diversity	2,060	765	24	NA	-0.0144	0.591	0.681
Experiences in Close Relationships anxious	2,060	765	24	NA	0.0010	0.980	0.980
Experiences in Close Relationships avoidant	2,060	765	24	NA	0.0140	0.610	0.682
Childhood Trauma Questionnaire total	1,656	658	21	NA	-0.0704	0.155	0.236

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = interaction between each predictor variable and log session number from multilevel models predicting psychological distress; Level 1 = session level; Level 2 = patient level; Level 3 = therapist level; Level 4 = clinic level; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with faster reductions in psychological distress over time.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 18**

*Rate of Change Model Results Covarying Time Since Therapist Assessment*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Level 4 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	23,719	5,426	84	12	0.0287	0.051	0.363
FIS verbal fluency	23,719	5,426	84	12	0.0037	0.812	0.864
Word fluency test	21,968	4,763	69	8	0.0059	0.723	0.859
FIS hope and positive expectations	23,719	5,426	84	12	0.0227	0.120	0.380
Depression, Anxiety, and Stress Scale	24,978	5,712	88	12	0.0279	0.041	0.363
Psychological Well-being	24,978	5,712	88	12	-0.0153	0.283	0.512
FIS persuasiveness	23,719	5,426	84	12	0.0269	0.072	0.363
Ten Item Personality Inventory extraversion	24,978	5,712	88	12	-0.0117	0.414	0.605
FIS emotional expression	23,719	5,426	84	12	0.0388	0.013	0.247
Emotional Stroop task	21,044	4,605	69	8	0.0193	0.223	0.488
FIS warmth, acceptance, and understanding	23,719	5,426	84	12	0.0174	0.236	0.488
Ten Item Personality Inventory agreeableness	24,978	5,712	88	12	0.0154	0.257	0.488
Five Facet Mindfulness Questionnaire Short Form total score	24,978	5,712	88	12	0.0043	0.753	0.864
Self-Compassion Scale Short Form total score	24,978	5,712	88	12	0.0157	0.251	0.488
Dispositional Contempt	24,978	5,712	88	12	0.0118	0.352	0.559
FIS empathy	23,719	5,426	84	12	0.0377	0.01	0.247
Interpersonal Reactivity Index	24,978	5,712	88	12	0.0239	0.077	0.363
Empathic accuracy task	22,922	5,235	82	13	0.0017	0.909	0.909
FIS alliance bond capacity	23,719	5,426	84	12	0.0208	0.134	0.392
FIS alliance rupture and repair	23,719	5,426	84	12	0.0317	0.028	0.355
NIH Toolbox Loneliness	24,978	5,712	88	12	0.0169	0.216	0.488
NIH Toolbox Emotional Support	24,978	5,712	88	12	-0.0033	0.821	0.864
Multicultural Orientation task overall	24,115	5,541	85	12	0.0277	0.078	0.363
Multicultural Orientation task comfort	24,115	5,541	85	12	0.0257	0.120	0.380

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	24,115	5,541	85	12	0.0243	0.117	0.380
Multicultural Orientation task opportunity	24,115	5,541	85	12	0.0229	0.086	0.363
Racial Implicit Association task	21,968	4,763	69	8	-0.0132	0.384	0.584
Multicultural Knowledge, Awareness, and Skills Scale total score	24,978	5,712	88	12	0.0057	0.688	0.843
Use of feminist and/or multicultural theory	25,957	5,963	96	13	0.0131	0.334	0.559
Deliberate practice hours	24,759	5,543	82	12	-0.0127	0.353	0.559
Difficulties in Therapeutic Practice professional self-doubt	24,978	5,712	88	12	-0.0089	0.540	0.684
Years providing psychotherapy	25,904	5,948	95	13	-0.0046	0.817	0.864
Perceived efficacy	25,904	5,948	95	13	-0.0028	0.841	0.864
Years of personal therapy	25,386	5,807	93	13	0.0176	0.152	0.413
Multitheoretical List of Interventions therapeutic technique diversity	24,978	5,712	88	NA	-0.0143	0.256	0.488
Experiences in Close Relationships anxious	24,978	5,712	88	12	0.0109	0.436	0.614
Experiences in Close Relationships avoidant	24,978	5,712	88	12	0.0089	0.531	0.684
Childhood Trauma Questionnaire total	23,547	5,431	80	11	0.0104	0.456	0.619

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = interaction between each predictor variable and log session number from multilevel models predicting psychological distress; Level 1 = session level; Level 2 = patient level; Level 3 = therapist level; Level 4 = clinic level. Negative estimates reflect associations with faster reductions in psychological distress over time.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 19**

*Rate of Change Model Results Not Restricting Based on Time Since Therapist Assessment*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Level 4 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	59,137	13,226	94	12	0.0182	0.074	0.456
FIS verbal fluency	59,137	13,226	94	12	0.0047	0.650	0.852
Word fluency test	48,487	10,520	80	9	0.0193	0.086	0.456
FIS hope and positive expectations	59,137	13,226	94	12	0.0144	0.163	0.497
Depression, Anxiety, and Stress Scale	62,659	13,985	98	12	0.0146	0.123	0.497
Psychological Well-being	62,659	13,985	98	12	0.0043	0.663	0.852
FIS persuasiveness	59,137	13,226	94	12	0.0192	0.062	0.456
Ten Item Personality Inventory extraversion	62,659	13,985	98	12	-0.0067	0.504	0.833
FIS emotional expression	59,137	13,226	94	12	0.0238	0.020	0.253
Emotional Stroop task	47,788	10,419	80	NA	0.0004	0.974	0.999
FIS warmth, acceptance, and understanding	59,137	13,226	94	12	0.0118	0.232	0.630
Ten Item Personality Inventory agreeableness	62,659	13,985	98	12	0.0233	0.019	0.253
Five Facet Mindfulness Questionnaire Short Form total score	62,659	13,985	98	12	0.0091	0.355	0.794
Self-Compassion Scale Short Form total score	62,659	13,985	98	12	0.0085	0.376	0.794
Dispositional Contempt	62,659	13,985	98	NA	0.0037	0.681	0.852
FIS empathy	59,137	13,226	94	12	0.0235	0.020	0.253
Interpersonal Reactivity Index	62,659	13,985	98	12	0.0186	0.056	0.456
Empathic accuracy task	58,066	13,050	90	13	0.0045	0.667	0.852
FIS alliance bond capacity	59,137	13,226	94	12	0.0137	0.162	0.497
FIS alliance rupture and repair	59,137	13,226	94	12	0.0139	0.170	0.497
NIH Toolbox Loneliness	62,659	13,985	98	12	0.0014	0.886	0.935
NIH Toolbox Emotional Support	62,659	13,985	98	12	0.0098	0.328	0.794
Multicultural Orientation task overall	59,594	13,363	95	12	0.0036	0.723	0.852
Multicultural Orientation task comfort	59,594	13,363	95	12	0.0054	0.604	0.852

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	59,594	13,363	95	12	0.0084	0.408	0.816
Multicultural Orientation task opportunity	59,594	13,363	95	12	0.0090	0.339	0.794
Racial Implicit Association task	48,487	10,520	80	9	-0.0076	0.476	0.833
Multicultural Knowledge, Awareness, and Skills Scale total score	62,659	13,985	98	12	0.0068	0.493	0.833
Use of feminist and/or multicultural theory	68,550	15,279	109	13	0.0031	0.740	0.852
Deliberate practice hours	56,713	12,848	94		-0.0150	0.167	0.497
Difficulties in Therapeutic Practice professional self-doubt	62,659	13,985	98	12	-0.0033	0.735	0.852
Years providing psychotherapy	68,230	15,195	107	13	0.0044	0.609	0.852
Perceived efficacy	68,230	15,195	107	13	-0.0018	0.847	0.920
Years of personal therapy	67,482	15,000	105	13	0.0145	0.096	0.456
Multitheoretical List of Interventions therapeutic technique diversity	62,659	13,985	98	12	0.0012	0.771	0.862
Experiences in Close Relationships anxious	62,659	13,985	98	12	0.0060	0.545	0.852
Experiences in Close Relationships avoidant	62,659	13,985	98	12	-0.0068	0.489	0.833
Childhood Trauma Questionnaire total	56,452	12,769	88	11	0.0000	0.999	0.999

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = interaction between each predictor variable and log session number from multilevel models predicting psychological distress; Level 1 = session level; Level 2 = patient level; Level 3 = therapist level; Level 4 = clinic level; NA = random intercept omitted due to model non-convergence. Negative estimates reflect associations with faster reductions in psychological distress over time.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 20**

*Results of Multilevel Logistic Regression Models Predicting Attendance at >1 Treatment Session with Outliers Removed*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	5,584	85	12	0.1124	0.078	0.674
FIS verbal fluency	5,562	84	12	0.2136	0.002	0.038
Word fluency test	4,905	70	8	-0.0212	0.766	0.870
FIS hope and positive expectations	5,584	85	12	0.0857	0.185	0.674
Depression, Anxiety, and Stress Scale	5,878	89	12	-0.0281	0.651	0.825
Psychological Well-being	5,878	89	12	0.0183	0.774	0.870
FIS persuasiveness	5,562	84	12	0.2018	0.002	0.038
Ten Item Personality Inventory extraversion	5,878	89	12	0.0132	0.839	0.908
FIS emotional expression	5,584	85	12	0.0746	0.273	0.684
Emotional Stroop task	4,768	70	8	0.0937	0.210	0.674
FIS warmth, acceptance, and understanding	5,584	85	12	0.0522	0.425	0.734
Ten Item Personality Inventory agreeableness	5,856	88	12	0.0794	0.210	0.674
Five Facet Mindfulness Questionnaire Short Form total score	5,857	88	12	0.1568	0.010	0.127
Self-Compassion Scale Short Form total score	5,878	89	12	0.0424	0.486	0.774
Dispositional Contempt	5,559	87	12	-0.0536	0.509	0.774
FIS empathy	5,584	85	12	0.0731	0.267	0.684
Interpersonal Reactivity Index	5,878	89	12	-0.0111	0.860	0.908
Empathic accuracy task	5,061	81	13	0.0038	0.965	0.965
FIS alliance bond capacity	5,584	85	12	0.0286	0.640	0.825
FIS alliance rupture and repair	5,584	85	12	0.0643	0.311	0.684
NIH Toolbox Loneliness	5,811	88	12	0.0918	0.175	0.674
NIH Toolbox Emotional Support	5,854	88	12	-0.0897	0.198	0.674
Multicultural Orientation task overall	5,701	86	12	0.0343	0.627	0.825
Multicultural Orientation task comfort	5,661	84	12	0.1266	0.158	0.674

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task humility	5,701	86	12	0.0207	0.778	0.870
Multicultural Orientation task opportunity	5,701	86	12	0.0621	0.312	0.684
Racial Implicit Association task	4,905	70	8	0.0651	0.344	0.684
Multicultural Knowledge, Awareness, and Skills Scale total score	5,878	89	12	0.0336	0.575	0.809
Use of feminist and/or multicultural theory	6,152	97	13	-0.0062	0.930	0.955
Deliberate practice hours	5,587	82	12	-0.1271	0.198	0.674
Difficulties in Therapeutic Practice professional self-doubt	5,878	89	12	-0.0601	0.360	0.684
Years providing psychotherapy	6,105	94	12	0.1695	0.213	0.674
Perceived efficacy	6,137	96	13	-0.0250	0.736	0.870
Years of personal therapy	5,994	94	13	0.0617	0.356	0.684
Multitheoretical List of Interventions therapeutic technique diversity	5,878	89	12	-0.0306	0.563	0.809
Experiences in Close Relationships anxious	5,878	89	12	0.0492	0.425	0.734
Experiences in Close Relationships avoidant	5,845	88	12	0.0505	0.503	0.774
Childhood Trauma Questionnaire total	5,496	79	11	-0.0765	0.317	0.684

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting likelihood of attending >1 treatment session (in log units); Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level. Positive estimates reflect associations with greater likelihood of attending >1 treatment session.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 21**

*Results of Multilevel Logistic Regression Models Predicting Attendance at >1 Treatment Session Restricted to Therapists with  $\geq 5$*

*Patients*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	5,562	78	11	0.0935	0.140	0.665
FIS verbal fluency	5,562	78	11	0.1627	0.011	0.139
Word fluency test	4,887	64	7	-0.0247	0.729	0.849
FIS hope and positive expectations	5,562	78	11	0.0715	0.263	0.754
Depression, Anxiety, and Stress Scale	5,856	82	11	-0.0317	0.606	0.849
Psychological Well-being	5,856	82	11	0.0299	0.636	0.849
FIS persuasiveness	5,562	78	11	0.1604	0.009	0.139
Ten Item Personality Inventory extraversion	5,856	82	11	0.0150	0.817	0.887
FIS emotional expression	5,562	78	11	0.0644	0.341	0.754
Emotional Stroop task	4,750	64	7	0.0985	0.185	0.754
FIS warmth, acceptance, and understanding	5,562	78	11	0.0275	0.672	0.849
Ten Item Personality Inventory agreeableness	5,856	82	11	0.0894	0.133	0.665
Five Facet Mindfulness Questionnaire Short Form total score	5,856	82	11	0.1367	0.017	0.162
Self-Compassion Scale Short Form total score	5,856	82	11	0.0562	0.357	0.754
Dispositional Contempt	5,856	82	11	0.0048	0.931	0.939
FIS empathy	5,562	78	11	0.0567	0.385	0.770
Interpersonal Reactivity Index	5,856	82	11	-0.0174	0.781	0.873
Empathic accuracy task	5,371	76	12	-0.0248	0.706	0.849
FIS alliance bond capacity	5,562	78	11	0.0204	0.737	0.849
FIS alliance rupture and repair	5,562	78	11	0.0602	0.342	0.754
NIH Toolbox Loneliness	5,856	82	11	0.1000	0.099	0.665
NIH Toolbox Emotional Support	5,856	82	11	-0.0991	0.123	0.665
Multicultural Orientation task overall	5,679	79	11	0.0056	0.935	0.939

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task comfort	5,679	79	11	0.072	0.310	0.754
Multicultural Orientation task humility	5,679	79	11	-0.0055	0.939	0.939
Multicultural Orientation task opportunity	5,679	79	11	0.0448	0.46	0.832
Racial Implicit Association task	4,887	64	7	0.0734	0.285	0.754
Multicultural Knowledge, Awareness, and Skills Scale total score	5,856	82	11	0.0325	0.586	0.849
Use of feminist and/or multicultural theory	6,126	89	12	-0.0275	0.700	0.849
Deliberate practice hours	5,701	76	10	-0.1582	0.004	0.139
Difficulties in Therapeutic Practice professional self-doubt	5,856	82	11	-0.0637	0.334	0.754
Years providing psychotherapy	6,111	88	12	-0.0572	0.538	0.849
Perceived efficacy	6,111	88	12	-0.0367	0.631	0.849
Years of personal therapy	5,968	86	NA	0.0581	0.415	0.788
Multitheoretical List of Interventions therapeutic technique diversity	5,856	82	11	-0.0303	0.571	0.849
Experiences in Close Relationships anxious	5,856	82	11	0.0354	0.564	0.849
Experiences in Close Relationships avoidant	5,856	82	11	0.0738	0.284	0.754
Childhood Trauma Questionnaire total	5,572	75	10	-0.0677	0.292	0.754

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting likelihood of attending >1 treatment session (in log units); Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Positive estimates reflect associations with greater likelihood of attending >1 treatment session.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 22**

*Results of Multilevel Logistic Regression Models Predicting Attendance at >1 Treatment Session Restricted to Patients with Clinically Elevated Symptoms at Baseline*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	5,562	78	11	0.0935	0.140	0.665
FIS verbal fluency	5,562	78	11	0.1627	0.011	0.139
Word fluency test	4,887	64	7	-0.0247	0.729	0.849
FIS hope and positive expectations	5,562	78	11	0.0715	0.263	0.754
Depression, Anxiety, and Stress Scale	5,856	82	11	-0.0317	0.606	0.849
Psychological Well-being	5,856	82	11	0.0299	0.636	0.849
FIS persuasiveness	5,562	78	11	0.1604	0.009	0.139
Ten Item Personality Inventory extraversion	5,856	82	11	0.0150	0.817	0.887
FIS emotional expression	5,562	78	11	0.0644	0.341	0.754
Emotional Stroop task	4,750	64	7	0.0985	0.185	0.754
FIS warmth, acceptance, and understanding	5,562	78	11	0.0275	0.672	0.849
Ten Item Personality Inventory agreeableness	5,856	82	11	0.0894	0.133	0.665
Five Facet Mindfulness Questionnaire Short Form total score	5,856	82	11	0.1367	0.017	0.162
Self-Compassion Scale Short Form total score	5,856	82	11	0.0562	0.357	0.754
Dispositional Contempt	5,856	82	11	0.0048	0.931	0.939
FIS empathy	5,562	78	11	0.0567	0.385	0.770
Interpersonal Reactivity Index	5,856	82	11	-0.0174	0.781	0.873
Empathic accuracy task	5,371	76	12	-0.0248	0.706	0.849
FIS alliance bond capacity	5,562	78	11	0.0204	0.737	0.849
FIS alliance rupture and repair	5,562	78	11	0.0602	0.342	0.754
NIH Toolbox Loneliness	5,856	82	11	0.1000	0.099	0.665
NIH Toolbox Emotional Support	5,856	82	11	-0.0991	0.123	0.665
Multicultural Orientation task overall	5,679	79	11	0.0056	0.935	0.939

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task comfort	5,679	79	11	0.0720	0.310	0.754
Multicultural Orientation task humility	5,679	79	11	-0.0055	0.939	0.939
Multicultural Orientation task opportunity	5,679	79	11	0.0448	0.46	0.832
Racial Implicit Association task	4,887	64	7	0.0734	0.285	0.754
Multicultural Knowledge, Awareness, and Skills Scale total score	5,856	82	11	0.0325	0.586	0.849
Use of feminist and/or multicultural theory	6,126	89	12	-0.0275	0.700	0.849
Deliberate practice hours	5,701	76	10	-0.1582	0.004	0.139
Difficulties in Therapeutic Practice professional self-doubt	5,856	82	11	-0.0637	0.334	0.754
Years providing psychotherapy	6,111	88	12	-0.0572	0.538	0.849
Perceived efficacy	6,111	88	12	-0.0367	0.631	0.849
Years of personal therapy	5,968	86	NA	0.0581	0.415	0.788
Multitheoretical List of Interventions therapeutic technique diversity	5,856	82	11	-0.0303	0.571	0.849
Experiences in Close Relationships anxious	5,856	82	11	0.0354	0.564	0.849
Experiences in Close Relationships avoidant	5,856	82	11	0.0738	0.284	0.754
Childhood Trauma Questionnaire total	5,572	75	10	-0.0677	0.292	0.754

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting likelihood of attending >1 treatment session (in log units); Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Positive estimates reflect associations with greater likelihood of attending >1 treatment session.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 23**

*Results of Multilevel Logistic Regression Models Predicting Attendance at >1 Treatment Session Restricted to Outcome*

*Questionnaire - 45*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	4,801	61	NA	0.1060	0.104	0.494
FIS verbal fluency	4,801	61	NA	0.1765	0.006	0.076
Word fluency test	4,597	61	NA	-0.0153	0.828	0.874
FIS hope and positive expectations	4,801	61	NA	0.0859	0.182	0.576
Depression, Anxiety, and Stress Scale	5,095	65	NA	-0.0173	0.790	0.874
Psychological Well-being	5,095	65	NA	0.0007	0.991	0.992
FIS persuasiveness	4,801	61	NA	0.1770	0.004	0.076
Ten Item Personality Inventory extraversion	5,095	65	NA	-0.0256	0.700	0.860
FIS emotional expression	4,801	61	NA	0.0463	0.534	0.846
Emotional Stroop task	4,460	61	NA	0.0895	0.230	0.606
FIS warmth, acceptance, and understanding	4,801	61	NA	0.0300	0.652	0.860
Ten Item Personality Inventory agreeableness	5,095	65	NA	0.0520	0.375	0.711
Five Facet Mindfulness Questionnaire Short Form total score	5,095	65	NA	0.1112	0.062	0.393
Self-Compassion Scale Short Form total score	5,095	65	NA	0.0220	0.724	0.860
Dispositional Contempt	5,095	65	NA	0.0202	0.718	0.860
FIS empathy	4,801	61	NA	0.0667	0.338	0.676
Interpersonal Reactivity Index	5,095	65	NA	-0.0213	0.736	0.860
Empathic accuracy task	4,608	58	NA	-0.0546	0.393	0.711
FIS alliance bond capacity	4,801	61	NA	0.0360	0.570	0.860
FIS alliance rupture and repair	4,801	61	NA	0.0761	0.250	0.606
NIH Toolbox Loneliness	5,095	65	NA	0.1109	0.091	0.494
NIH Toolbox Emotional Support	5,095	65	NA	-0.1036	0.127	0.536
Multicultural Orientation task overall	4,918	62	NA	-0.0170	0.813	0.874

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task comfort	4,918	62	NA	0.0857	0.256	0.606
Multicultural Orientation task humility	4,918	62	NA	-0.0372	0.605	0.860
Multicultural Orientation task opportunity	4,918	62	NA	0.0199	0.747	0.860
Racial Implicit Association task	4,597	61	NA	0.0742	0.272	0.606
Multicultural Knowledge, Awareness, and Skills Scale total score	5,095	65	NA	0.0430	0.513	0.846
Use of feminist and/or multicultural theory	5,338	69	NA	-0.1039	0.141	0.536
Deliberate practice hours	5,316	67	NA	-0.1540	0.004	0.076
Difficulties in Therapeutic Practice professional self-doubt	5,095	65	NA	0.0007	0.992	0.992
Years providing psychotherapy	5,323	68	NA	-0.2526	0.012	0.114
Perceived efficacy	5,323	68	NA	-0.1481	0.038	0.289
Years of personal therapy	5,180	66	NA	0.0408	0.517	0.846
Multitheoretical List of Interventions therapeutic technique diversity	5,095	65	NA	0.0232	0.693	0.860
Experiences in Close Relationships anxious	5,095	65	NA	0.0786	0.201	0.588
Experiences in Close Relationships avoidant	5,095	65	NA	0.0981	0.174	0.576
Childhood Trauma Questionnaire total	4,917	60	NA	-0.0678	0.287	0.606

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting likelihood of attending >1 treatment session (in log units); Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to data coming from a single clinic. Positive estimates reflect associations with greater likelihood of attending >1 treatment session. As the Outcome Questionnaire – 45 was only collected in one clinic, a clinic random intercept was not estimated.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 24**

*Results of Multilevel Logistic Regression Models Predicting Attendance at >1 Treatment Session Restricted to Behavioral Health*

*Monitor - 20*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	783	24	NA	0.2120	0.239	0.605
FIS verbal fluency	783	24	11	0.2755	0.214	0.605
Word fluency test	308	9	NA	0.3457	0.560	0.722
FIS hope and positive expectations	783	24	NA	0.2149	0.231	0.605
Depression, Anxiety, and Stress Scale	783	24	NA	-0.0136	0.938	0.938
Psychological Well-being	783	24	NA	0.0547	0.775	0.796
FIS persuasiveness	783	24	NA	0.2739	0.153	0.529
Ten Item Personality Inventory extraversion	783	24	NA	0.3142	0.093	0.401
FIS emotional expression	783	24	NA	0.1451	0.341	0.617
Emotional Stroop task	308	9	NA	-0.2523	0.627	0.722
FIS warmth, acceptance, and understanding	783	24	NA	0.1919	0.311	0.617
Ten Item Personality Inventory agreeableness	783	24	NA	0.3213	0.207	0.605
Five Facet Mindfulness Questionnaire Short Form total score	783	24	11	0.3028	0.077	0.401
Self-Compassion Scale Short Form total score	783	24	NA	0.1030	0.592	0.722
Dispositional Contempt	783	24	NA	0.1206	0.552	0.722
FIS empathy	783	24	NA	0.1827	0.300	0.617
Interpersonal Reactivity Index	783	24	NA	0.1120	0.582	0.722
Empathic accuracy task	785	25	NA	0.1770	0.445	0.669
FIS alliance bond capacity	783	24	NA	0.1250	0.458	0.669
FIS alliance rupture and repair	783	24	NA	0.1452	0.404	0.667
NIH Toolbox Loneliness	783	24	NA	0.0802	0.675	0.754
NIH Toolbox Emotional Support	783	24	NA	-0.1049	0.619	0.722
Multicultural Orientation task overall	783	24	NA	0.3098	0.093	0.401

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task comfort	783	24	NA	0.1627	0.352	0.617
Multicultural Orientation task humility	783	24	NA	0.3772	0.041	0.312
Multicultural Orientation task opportunity	783	24	NA	0.3061	0.095	0.401
Racial Implicit Association task	308	9	NA	-0.5245	0.131	0.498
Multicultural Knowledge, Awareness, and Skills Scale total score	783	24	NA	0.0484	0.758	0.796
Use of feminist and/or multicultural theory	814	28	NA	0.3871	0.033	0.312
Deliberate practice hours	407	16	NA	-0.8239	0.347	0.617
Difficulties in Therapeutic Practice professional self-doubt	783	24	11	-0.5021	0.006	0.114
Years providing psychotherapy	814	28	NA	0.1497	0.355	0.617
Perceived efficacy	814	28	NA	0.6143	0.001	0.038
Years of personal therapy	814	28	NA	0.2539	0.480	0.676
Multitheoretical List of Interventions therapeutic technique diversity	783	24	NA	-0.3612	0.021	0.266
Experiences in Close Relationships anxious	783	24	NA	-0.1654	0.454	0.669
Experiences in Close Relationships avoidant	783	24	NA	-0.0566	0.756	0.796
Childhood Trauma Questionnaire total	675	21	NA	-0.2503	0.357	0.617

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting likelihood of attending >1 treatment session (in log units); Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Positive estimates reflect associations with greater likelihood of attending >1 treatment session.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 25**

*Results of Multilevel Logistic Regression Models Predicting Attendance at >1 Treatment Session Covarying Time Since Therapist*

*Assessment*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	5,584	85	12	0.1344	0.052	0.395
FIS verbal fluency	5,584	85	12	0.2098	0.002	0.038
Word fluency test	4,905	70	8	-0.0222	0.770	0.839
FIS hope and positive expectations	5,584	85	12	0.1065	0.135	0.603
Depression, Anxiety, and Stress Scale	5,878	89	12	-0.0113	0.867	0.915
Psychological Well-being	5,878	89	12	0.0238	0.731	0.839
FIS persuasiveness	5,584	85	12	0.2160	0.001	0.038
Ten Item Personality Inventory extraversion	5,878	89	12	0.0254	0.722	0.839
FIS emotional expression	5,584	85	12	0.0891	0.222	0.603
Emotional Stroop task	4,768	70	8	0.1235	0.109	0.592
FIS warmth, acceptance, and understanding	5,584	85	12	0.0562	0.429	0.696
Ten Item Personality Inventory agreeableness	5,878	89	12	0.0704	0.284	0.646
Five Facet Mindfulness Questionnaire Short Form total score	5,878	89	12	0.1428	0.022	0.209
Self-Compassion Scale Short Form total score	5,878	89	12	0.0626	0.344	0.688
Dispositional Contempt	5,878	89	12	0.0084	0.892	0.916
FIS empathy	5,584	85	12	0.0956	0.178	0.603
Interpersonal Reactivity Index	5,878	89	12	-0.0197	0.773	0.839
Empathic accuracy task	5,393	83	13	-0.0351	0.628	0.795
FIS alliance bond capacity	5,584	85	12	0.0417	0.531	0.727
FIS alliance rupture and repair	5,584	85	12	0.0795	0.245	0.621
NIH Toolbox Loneliness	5,878	89	12	0.1150	0.084	0.532
NIH Toolbox Emotional Support	5,878	89	12	-0.0959	0.172	0.603
Multicultural Orientation task overall	5,701	86	12	0.0698	0.371	0.696

## PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task comfort	5,701	86	12	0.1010	0.191	0.603
Multicultural Orientation task humility	5,701	86	12	0.0617	0.449	0.696
Multicultural Orientation task opportunity	5,701	86	12	0.0974	0.148	0.603
Racial Implicit Association task	4,905	70	8	0.0902	0.217	0.603
Multicultural Knowledge, Awareness, and Skills Scale total score	5,878	89	12	0.0319	0.626	0.795
Use of feminist and/or multicultural theory	6,152	97	13	0.0027	0.971	0.971
Deliberate practice hours	5,723	83	12	-0.1519	0.011	0.139
Difficulties in Therapeutic Practice professional self-doubt	5,878	89	12	-0.0713	0.321	0.678
Years providing psychotherapy	6,137	96	13	-0.0589	0.536	0.727
Perceived efficacy	6,137	96	13	-0.0348	0.659	0.808
Years of personal therapy	5,994	94	13	0.0463	0.516	0.727
Multitheoretical List of Interventions therapeutic technique diversity	5,878	89	12	-0.0435	0.458	0.696
Experiences in Close Relationships anxious	5,878	89	12	0.0536	0.426	0.696
Experiences in Close Relationships avoidant	5,878	89	12	0.0559	0.454	0.696
Childhood Trauma Questionnaire total	5,592	81	NA	-0.0798	0.289	0.646

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting likelihood of attending >1 treatment session (in log units); Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level; NA = random intercept omitted due to model non-convergence. Positive estimates reflect associations with greater likelihood of attending >1 treatment session.

PREDICTING THE THERAPIST EFFECT

**Supplemental Materials Table 26**

*Results of Multilevel Logistic Regression Models Predicting Attendance at >1 Treatment Session Not Restricting Based on Time Since*

*Therapist Assessment*

Predictor	Level 1 <i>n</i>	Level 2 <i>n</i>	Level 3 <i>n</i>	Estimate	<i>p</i>	<i>p</i> <sub>FDR</sub>
FIS total	13,447	94	12	0.1027	0.021	0.158
FIS verbal fluency	13,447	94	12	0.1434	0.001	0.019
Word fluency test	10,716	80	9	0.0318	0.492	0.748
FIS hope and positive expectations	13,447	94	12	0.0758	0.096	0.304
Depression, Anxiety, and Stress Scale	14,226	98	12	0.0030	0.946	0.972
Psychological Well-being	14,226	98	12	0.0404	0.359	0.680
FIS persuasiveness	13,447	94	12	0.1400	0.001	0.019
Ten Item Personality Inventory extraversion	14,226	98	12	0.0100	0.830	0.906
FIS emotional expression	13,447	94	12	0.0816	0.077	0.269
Emotional Stroop task	10,640	80	9	0.0604	0.303	0.640
FIS warmth, acceptance, and understanding	13,447	94	12	0.0420	0.349	0.680
Ten Item Personality Inventory agreeableness	14,226	98	12	0.0507	0.260	0.581
Five Facet Mindfulness Questionnaire Short Form total score	14,226	98	12	0.1037	0.016	0.152
Self-Compassion Scale Short Form total score	14,226	98	12	0.0267	0.543	0.794
Dispositional Contempt	14,226	98	12	0.0132	0.751	0.865
FIS empathy	13,447	94	12	0.0668	0.147	0.399
Interpersonal Reactivity Index	14,226	98	12	-0.0011	0.981	0.981
Empathic accuracy task	13,278	90	13	-0.0229	0.615	0.823
FIS alliance bond capacity	13,447	94	12	0.0336	0.447	0.708
FIS alliance rupture and repair	13,447	94	12	0.0790	0.078	0.269
NIH Toolbox Loneliness	14,226	98	12	0.0377	0.376	0.680
NIH Toolbox Emotional Support	14,226	98	12	-0.0717	0.109	0.319
Multicultural Orientation task overall	13,586	95	12	0.0997	0.025	0.158

PREDICTING THE THERAPIST EFFECT

Multicultural Orientation task comfort	13,586	95	12	0.0836	0.071	0.269
Multicultural Orientation task humility	13,586	95	12	0.0907	0.052	0.247
Multicultural Orientation task opportunity	13,586	95	12	0.0918	0.032	0.174
Racial Implicit Association task	10,716	80	9	0.0250	0.587	0.823
Multicultural Knowledge, Awareness, and Skills Scale total score	14,226	98	12	0.0568	0.198	0.470
Use of feminist and/or multicultural theory	15,550	109	13	-0.0223	0.628	0.823
Deliberate practice hours	13,095	94	12	-0.1047	0.009	0.114
Difficulties in Therapeutic Practice professional self-doubt	14,226	98	12	-0.0154	0.738	0.865
Years providing psychotherapy	15,466	107	13	0.0106	0.858	0.906
Perceived efficacy	15,466	107	13	0.0381	0.441	0.708
Years of personal therapy	15,269	105	13	0.0605	0.169	0.428
Multitheoretical List of Interventions therapeutic technique diversity	14,226	98	12	0.0095	0.664	0.841
Experiences in Close Relationships anxious	14,226	98	12	0.0084	0.854	0.906
Experiences in Close Relationships avoidant	14,226	98	12	0.0359	0.441	0.708
Childhood Trauma Questionnaire total	12,997	88	11	-0.0152	0.743	0.865

*Note.* FIS = Facilitative Interpersonal Skills; NIH = National Institutes of Health  $p = p$ -value;  $p_{FDR}$  = false-discovery rate corrected  $p$ -values; Estimate = fixed effect of each predictor variable from multilevel model predicting likelihood of attending >1 treatment session (in log units); Level 1 = patient level; Level 2 = therapist level; Level 3 = clinic level. Positive estimates reflect associations with greater likelihood of attending >1 treatment session.