Psychotherapy Research

ISSN: 1050-3307 (Print) 1468-4381 (Online) Journal homepage: http://www.tandfonline.com/loi/tpsr20

Predicting psychotherapy outcome based on therapist interpersonal skills: A five-year longitudinal study of a therapist assessment protocol

Henning Schöttke, Christoph Flückiger, Simon B. Goldberg, Julia Eversmann & Julia Lange

To cite this article: Henning Schöttke, Christoph Flückiger, Simon B. Goldberg, Julia Eversmann & Julia Lange (2017) Predicting psychotherapy outcome based on therapist interpersonal skills: A five-year longitudinal study of a therapist assessment protocol, Psychotherapy Research, 27:6, 642-652, DOI: 10.1080/10503307.2015.1125546

To link to this article: http://dx.doi.org/10.1080/10503307.2015.1125546

Psychotherapy Research, 2017
Vol. 27, No. 6, 642–652, http://dx.doi.org/10.1080/10503307.2015.1125546

EMPIRICAL PAPER

Predicting psychotherapy outcome based on therapist interpersonal skills: A five-year longitudinal study of a therapist assessment protocol

HENNING SCHÖTTKE[1], CHRISTOPH FLÜCKIGER[2], SIMON B. GOLDBERG[3], JULIA EVERSMANN[1], & JULIA LANGE[1]

[1]*Department of Clinical Psychology & Psychotherapy, Institute of Psychology, University of Osnabrück, Osnabrück, Germany;* [2]*Department of Clinical Psychology & Psychotherapy, Institute of Psychology, University of Bern, Bern, Switzerland &* [3]*Department of Counseling Psychology, School of Education, University of Wisconsin–Madison, Madison, WI, USA*

(Received 24 August 2015; revised 12 November 2015; accepted 19 November 2015)

Abstract

Objective: In the past decade, variation in outcomes between therapists (i.e., therapist effects) have become increasingly recognized as an important factor in psychotherapy. Less is known, however, about what accounts for differences between therapists. The present study investigates the possibility that therapists' basic therapy-related interpersonal skills may impact outcomes. **Method:** To examine this, psychotherapy postgraduate trainees completed both an observer- and an expert-rated behavioral assessment: the Therapy-Related Interpersonal Behaviors (TRIB). TRIB scores were used to predict trainees' outcomes over the course of the subsequent five years. **Results:** Results indicate that trainees' with more positively rated interpersonal behaviors assessed in the observer-rated group format but not in a single expert-rated format showed superior outcomes over the five-year period. This effect remained controlling for therapist characteristics (therapist gender, theoretical orientation [cognitive behavioral or psychodynamic], amount of supervision, patient's order within therapist's caseload), and patient characteristics (patient age, gender, number of comorbid diagnoses, global severity, and personality disorder diagnosis). **Conclusions:** These findings underscore the importance of therapists' interpersonal skills as a predictor of outcome and source of therapist effects. The potential utility of assessing therapists' and therapists-in-training interpersonal skills are discussed.

Keywords: psychotherapy; therapist effects; observer ratings; expert ratings

Clearly not everyone is a good fit for every occupation. It is therefore important to allocate training opportunities to candidates most likely to excel in a given field. Research on reliability and objectivity of personnel selection instruments has been conducted in the world of business since the early 1980s (Robinson, 1981). Unstructured employment methods (e.g., unstructured interviews) are commonly used for personnel selection, despite the fact that they show low validity for the prediction of either job training performance or overall job performance (Schmidt & Hunter, 1998). Meta-analytic evidence suggests that more objective measures of ability (e.g., general mental ability tests, work sample measures) are more valid predictors of future job performance (Schmidt & Hunter, 1998). More recently, several meta-analyses in managerial samples have highlighted the relevance of behavioral assessments in favor of psychometric tests of personality and cognitive ability (Dilchert & Ones, 2009 as a means to evaluate more subtle interpersonal and cognitive capacities such as communication (Arthur, Day, McNelly, & Edens, 2003), awareness of others (Arthur et al., 2003; Dilchert & Ones, 2009), influence over others (Arthur et al., 2003), organization and planning skills (Arthur et al., 2003; Dilchert & Ones, 2009), and stress tolerance (Meriac,

*Correspondence concerning this article should be addressed to Henning Schöttke, Department of Clinical Psychology & Psychotherapy, Institute of Psychology, University of Osnabrück, Knollstraße 15, Osnabrück 49069, Germany. Email: henning.schoettke@uni-osnabrueck.de

Hoffman, Woehr, & Fleisher, 2008; see Hermelin, Lievans, & Robertson, 2007 for a meta-analytic review).

Assessing an applicant's fitness for future work as a psychotherapist poses unique challenges, as the job requires a variety of cognitive, affective, and interpersonal skills that may be inherently difficult to measure (e.g., Ackerman & Hilsenroth, 2003). Historically, interviews with graduate program faculty (i.e., expert interviews) along with a review of applicants' academic credentials are used to select future psychotherapists. The few studies that have investigated the predictive validity of expert selection interviews have (consistent with Schmidt and Hunter (1998)) reported they show low or no predictive value for later patient outcomes or supervisor ratings of therapists' performance (Hackmann, Wiggins, & Bass, 1970; Stricker & Huber, 1967; Whiteley, Sprinthall, Mosher, & Donaghy, 1967). Based on this lack of predictive validity for traditional selection methods, there is a lasting need to test alternative empirical selection methods to predict applicants' future performance in psychotherapy training programs and later career outcomes (Bergin, 1997; Costanzo & Philpott, 1986; Kuhr, 1998).

As one of the very few examples of an attempt to develop predictive methods of psychotherapy trainee selection, Costanzo and Philpott (1986) assessed psychotherapeutic talent using the Group Assessment of Interpersonal Traits (GAIT; Goodman, 1972) in which therapy trainees were rated on the Rogerian constructs of empathy, acceptance, and openness, as well as on their rapport, insight, nonverbal attentiveness, emotional-affective orientation and intellectual-cognitive orientation. Costanzo and Philpott (1986) found that a combination of multiple measures of interpersonal intelligence was a strong predictor of psychotherapeutic talent in contrast to other candidate predictors (e.g., demographic information, personality orientation, and academic achievement). Other than the GAIT, however, and perhaps because of the lack of preliminary evidence to predict future therapeutic outcomes, few attempts have been made in the past decades to investigate scientific methods/assessments for the systematic selection of psychotherapists or students for psychotherapy training programs.

Despite limited available methods for assessing applicants' potential for becoming effective therapists, there is ample evidence that therapists in practice vary considerably in their outcomes (see Baldwin & Imel, 2013 for a review). Differences between therapists in their overall effectiveness, broadly termed therapist effects (Baldwin & Imel, 2013), have been demonstrated in a variety of study designs including within randomized clinical trials (Kim, Wampold, & Bolt, 2006) and in naturalistic settings (Lutz, Leon, Martinovich, Lyons, & Stiles, 2007; Okiishi, Lambert, Nielsen, & Ogles, 2003; Wampold & Brown, 2005). Therapist effects have been present even in naturalistic settings employing manualized treatments (Laska, Smith, Wislocki, Minami, & Wampold, 2013). Therapist effects have also been detected on key therapy ingredients such as the therapeutic alliance (Del Re, Flückiger, Horvath, Symonds, & Wampold, 2012).

Some reports have begun to investigate therapist variables potential linked to patient outcomes as an attempt to explain variation between therapists. Perhaps the two therapist characteristics most consistently proposed as predictors of treatment outcomes include therapist competence, therapist adherence, and therapist interpersonal skills (Mallinckrodt, 2000; Rønnestad & Skovholt, 1997, 2005; Skovholt & Rønnestad, 1995). However, meta-analytic reviews have not always supported these proposals; in their meta-analytic review, Webb, DeRubeis, and Barber (2010) report no significant effects for therapist adherence ($r = 0.02$) or therapist competence ($r = 0.07$) on treatment outcome. In their programmatic study of the professional development of therapists, Orlinsky and Rønnestad (2005) report that effective therapists might possess a "natural talent" for excellent basic relational skills, potentially developed during the whole therapeutic career. Furthermore, relational skills might be explicitly prompted during psychotherapeutic training (Beidas & Kendall, 2010; Boswell & Castonguay, 2007; Dennhag & Armelius, 2012; Herschell, Kolko, Baumann, & Davis, 2010).

One key category of candidate therapist variables with obvious relevance to the actual practice of psychotherapy centers on therapists' interpersonal capacities. The following therapist interpersonal behaviors have been proposed as predictors of patient outcomes: displays empathy and communicative attunement (Bohart, Elliott, Greenberg, & Watson, 2002), is a good communicator (Henry, Schacht, Strupp, Butler, & Binder, 1993), has a warm and trusting interpersonal style (Mohl, Martinez, Ticknor, Huang, & Cordell, 1991; Nissen-Lie, Monsen, & Rønnestad, 2010), responds thoughtfully and appropriately to patient hostility (Binder & Strupp, 1997), is sensitive to and aware of the interpersonal therapeutic process (Hartley & Strupp, 1983), displays the ability to address difficulties/ruptures that occur during the therapeutic process (Castonguay, Boswell, Constantino, Goldfried, & Hill, 2010; Safran, Muran, Samstag, & Stevens, 2002; see also Ackerman & Hilsenroth, 2003).

Anderson, Ogles, Patterson, Lambert, and Vermeersch (2009) have found in their work, that treatment outcome is associated with the degree of therapists' facilitative skills. Furthermore, the results of their study imply an important methodological factor: Interpersonal skills appeared as a relevant predictor only when assessed with an objective performance task, independent from the course of therapy (and not when assessed via therapist self-report). This performance task measures therapists' interpersonal skills by rating therapists' responses to short standardized video simulations of problematic client-therapist interactions. Indeed, no relationship was observed between therapists' self-report interpersonal skill and treatment outcome. However, the results are agnostic whether these relational skills are personal competences that can be modified during training as there exist only a few studies that address this question (Crits-Cristoph & Gallop, 2006; Henry et al., 1993; Hilsenroth, Ackerman, Clemence, Strassle, & Handler, 2002; Muran, Safran, Samstag, & Winston, 2005; Raue, Goldfried, & Barkham, 1997; Safran, Muran, Samstag, & Winston, 2005). Overall, it appears that therapists' interpersonal skills may be a robust predictor of actual outcomes in clinical practice. However, there is little knowledge if and how interpersonal skills can be assessed before psychotherapy training and how predictive such an assessment would be.

The objective of this five-year longitudinal study was to develop a personnel pre-training assessment relying on therapy-related interpersonal skills that may or may not predict future therapeutic outcomes. More specifically, we developed an observer- and expert-rated behavioral assessment measure to systematically assess therapist qualities over the course of a one day protocol. We investigated (a) if therapists' interpersonal skills can be reliably assessed in a group format (observer ratings) or in an interview format (expert rating) and (b) if the group or interview assessments predict psychotherapy outcome over the subsequent five years of clinical work. We have considered relevant, generally discussed variables related to outcome for patients' and therapists' (global severity, number of comorbid diagnoses, therapeutic orientation, level of training) to assess the predictive validity of interpersonal skills at pre-training assessment when various potential confounds are modeled.

Method

Design

This longitudinal study was conducted between 2004 and 2014 at the outpatient clinic of the University of Osnabrück (Figure 1). At a first phase, the Therapy-Related Interpersonal Behavior (TRIB) assessments were developed in a group format (TRIB-G) and in a single structured expert interview format (TRIB-I) based on a sample of 60 and 152 therapists, respectively. In a second phase, 42 candidates were

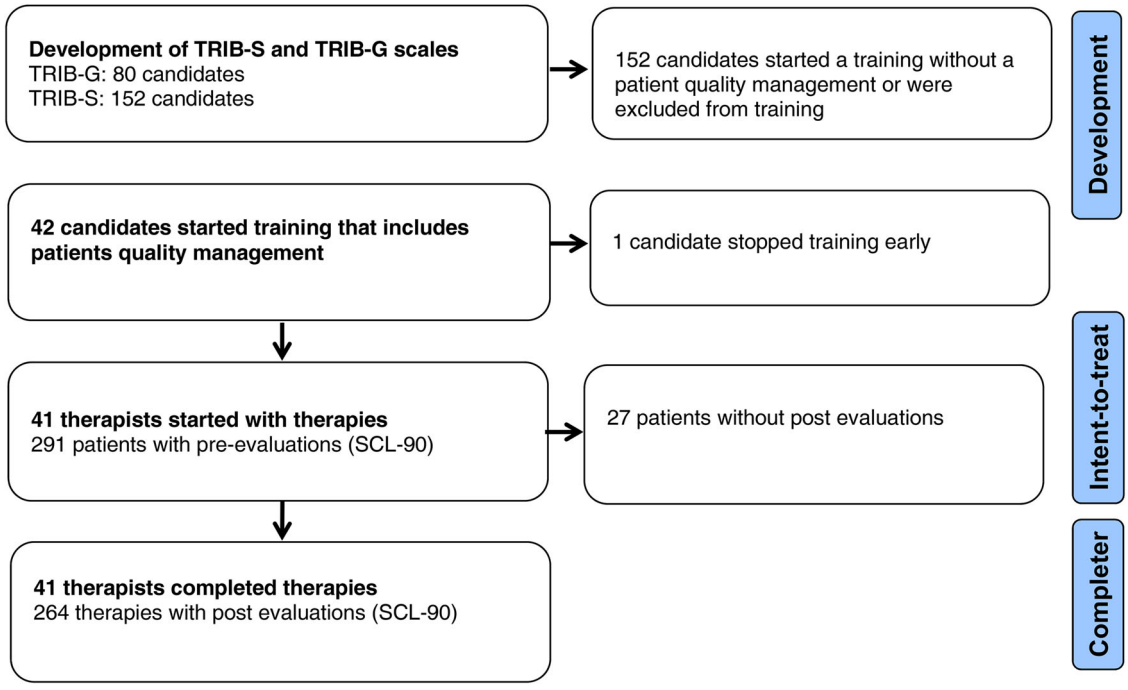

Figure 1. Flowchart.

selected for their postgraduate training, were screened by both TRIB assessments and started with the training that included patients' outcome assessment based on a newly-created clinic quality management. From these 42 candidates, 41 therapists started with their therapies and treated an intent-to-treat (ITT) sample of 291 patients (mean of patients by each therapist = 4.74 [*SD* = 2.83] range from 1 to 13 therapies) and a completer sample of 264 that participated in the post-treatment assessment (mean of patients by each therapist = 4.67 [*SD* = 2.80] range from 1 to 13 therapies).

Participants

Therapist trainees. All candidates had previously attained a European master's degree in psychology (five -year combined bachelor's and master's degree in psychology) and committed for a further self-selected five-year postgraduate training either in cognitive behavioral therapy (CBT) or psychodynamic therapy (PDT) (Table I). Each therapist of the University of Osnabrück who had successfully completed the first two and a half years of the training and had at least one completed course of therapy was included in this study.

The training followed the standard requirements for psychologists to obtain licensure in Germany. After having completed the behavioral assessment, therapist trainees began a training program in either CBT or PDT. The choice of the therapeutic training orientation (CBT or PDT) was decided by the trainee based on his or her therapeutic preference. Both training programs (CBT, PDT) started simultaneously and were comprised a total of 4200 training hours, including 600 hr of theoretical education, 930 hr of self-study, 120 hr of personal therapy and at least 150 hr of monthly video-supported supervision. During the first two and a half years, trainees worked as psychiatric interns (total of 1200 hr) and completed another 600 hr in an accredited inpatient medical clinic with either a psychosomatic or rehabilitation focus.

The therapists practiced a manual-oriented, evidence-based approach of CBT (Margraf & Schneider, 2008) or PDT (Wöller & Kruse, 2010) based on individualized case formulations that were completed within the first five sessions. Within the context of these case formulations, CBT therapists preferred problem- and symptom-focused as well as well-being goals, while PDT therapists showed a strong preference for personal growth and interpersonal goals (Schöttke, Trame, & Sembill, 2014). The number of sessions was limited *a priori* with the possibility of requesting further sessions if deemed appropriate. On average, every fourth session was supervised by licensed experts who supported the trainees in therapy planning and treatment. Individual therapy sessions for CBT and PDT typically lasted 50 min.

Table I. Sample characteristics at intake (*N* = 291 patients/41 therapists).

Patients	
Gender (f/m)	178/113
Age (*SD*)	40.5 (12.0)
No. of patients (%) with primary diagnosis	
Depression	110 (38)
Dysthymia	9 (3)
Agoraphobia/panic disorder	18 (6)
Social phobia	24 (8)
Specific phobia	4 (1)
Generalized anxiety	4 (1)
Obsessive compulsive disorder (OCD)	5 (2)
Posttraumatic stress disorder (PTSD)	13 (4)
Somatoform disorders	17 (6)
Eating disorders	7 (2)
Dyssomnia	3 (1)
Adjustment disorder	36 (12)
Personality disorders	25 (9)
Other diagnoses	16 (6)
No diagnosis	0
No. of patients (%) with comorbidities	
Second diagnosis	145 (50)
Third diagnosis	54 (19)
Personality disorder comorbidity	35 (12)
Therapists	
Gender (f/m)	30/11
Theoretical orientation (CBT/PDT)	25/16
No. of supervision sessions (*SD*)	13.6 (7.2)

Patients. To be included in the study, all patients had to be age 21 or older, have the presence of a psychological disorder, and receiving no simultaneous psychotherapeutic treatment. Patients were excluded if they displayed or presented with acute suicidality, physical violence, psychosis, or current severe substance abuse. Each participant completed an initial five treatment sessions, referred to as probationary sessions in the German healthcare system. During these sessions, individual targets complaints, diagnoses (determined using the Structured Clinical Interview for DSM-IV [SCID-I; Wittchen, Zaudig, & Fydrich, 1997]) and prognoses were established, and treatment goals were specified and agreed upon. The average age in the sample was 40.5 years (*SD* = 12.0), and the majority of the patients were female (61%). The most common primary diagnosis was depression (38%, see Table I). Patients completed on average 53.8 sessions of therapy (*SD* = 31.73) over the course of 572 days (*SD* = 253).

Measures

Standardized assessments were used to assess both therapist capacities and patient baseline characteristics and patient as outcomes:

Therapists—TRIB. This study employed a novel behavioral assessment of therapists' interpersonal behavior in both a group format (TRIB-G) and a single structured expert interview format (TRIB-I).

TRIB-G. The contributions of each candidate to a group discussion on a provoking film were rated using the TRIB-G by trained raters. Ratings were based on nine behaviors selected from a larger pool of 17 items linked to aspects of potential group dynamics a prior pilot study of $n = 82$ candidates (Eversmann, 2008). Items were selected based on their relevance to therapist features theoretically linked to building a pantheoretical successful therapeutic relationship (e.g., Ackerman & Hilsenroth, 2001; Hermer & Röhrle, 2008). Hence, the rating scale entailed behavioral patterns from the following domains: (1) "clear and positive communication" (two items); (2) "empathy and communicative attunement" (two items); (3) "respect and warmth" (one item); (4) "managing of criticism" (three items); and (5) "willingness to cooperate" (one item). A high raw score indicates high interpersonal competence. Each item was rated on a five-point Likert-type scale ranging from zero to four. The end-points of the scale were behaviorally anchored to ensure a uniform understanding of the terms. A total score was computed by taking the sum of all nine items. The factor analysis using varimax-rotation indicated that all items loaded on one main factor (78% of explained variance) (Eversmann, Schöttke, Wiedl, & Rogner, 2011). The measure has shown some evidence of interrater reliability (e.g., estimated intraclass correlation [*ICC*] varied between 0.61 and 0.71 depending on training experience and the group of raters (Eversmann et al., 2011, see Supplemental Materials). In a prior study, that described the construction of the TRIB-G, results of the factor analysis and the *ICC*'s were reported. The sample involved $n = 82$ therapist trainees. In the present sample ($n = 42$), the average score of the TRIB-G was 27.0 ($SD = 4.85$; Min. = 12, Max. = 35).

TRIB-I. The structured interview component of the TRIB also contained a pantheoretical assessment of applicants interpersonally related competencies and personal strengths/capabilities developed within a group of training-experts of both traditions at the University of Osnabrück (CBT and/or PDT). The structured interview lasted between 30 and 45 min and consisted of six questions that were asked to applicants (Schöttke, 2014, see Supplemental Materials). Answers to the questions were evaluated by experts from both orientations (i.e., CBT or PDT) conducting the interviews using eight items. Psychometric properties of TRIB-I are reported in a previous study that describes the construction of this instrument on the basis of data from $n = 194$ therapists (Schöttke, 2014). Therapists' data of the current study are a subsample of the original study group. The results of a confirmatory factor analysis (CFA) based on the overall sample ($n = 194$) indicated that a two-dimensional model with three items (factor I1 "motivation") and five items (factor I2 "personal strengths") resulted in the best model fit (RMSEA = 0.08; CFI = 0.98; Schöttke, 2014). The factor "motivation" included allegiance/enthusiasm in the selected tradition, interest in patients, experience in and motivation for personal-reflection. The factor "personal strengths" included quality of self-perception, quality of self-communication, quality of perception of others, quality of communication, quality, and stability in personal relationships. High raw scores of TRIB-I1 "motivation" indicates high level of motivation and a low raw score of TRIB-I2 "personal strengths" indicates high level of personal strength. Within a subsample of14 applicants interrater agreement was acceptable ($ICC = .79$) by two independent expert raters of both traditions (Schöttke, 2014). In the present sample ($n = 42$), the average score of the TRIB-I1 was $M = 0.18$ ($SD = .34$; Min. = 0, Max. = 1.6) and of TRIB-I2 $M = 3.24$ ($SD = .59$; Min. = 1.7, Max. = 4). Within the sample of 42 therapists that were screened by all TRIB observer ratings, the three scales correlated significantly to each other in the expected directions (TRIB-I1/I2 = –.52; TRIB-I1/G = .25; TRIB-I2/G = –.36, $p < .001$).

Procedure. Both observer- and expert-rated assessments of trainees' interpersonal skills occurred over the course of a one day standardized assessment protocol. The day opened with a welcome reception, an explanation of the routine, and the introductions of the rater and the moderator of the group discussions. The participants then participated in structured group discussions in groups ranging in size from 6 to 12 participants. The group discussion started with a provocative 15 min film clip. The film clip included a demonstration of an intervention that was not part of the therapeutic training offered at the institute. Previous work has established that this clip incites debate among therapy trainees (Reckewerth, 2006). The participants were asked to

monitor their feelings and thoughts during the viewing of the clip and to discuss them afterwards. After watching the video clip, the moderator repeated the standardized instructions and emphasized that the focus of the discussion was the emotional, physiological, and behavioral reactions experienced by the trainees. A scientific debate of therapeutic techniques was not requested. Trained raters observed the behavior of the candidates and evaluated their contributions to the discussion using the TRIB-G. The raters were physically present during the discussions.

After a 30 min break, senior therapists conducted selection interviews with each candidate in a single assessment using the TRIB-I. Each therapist trainee was rated by at least two experts from a pool of three female and eight male senior therapists who were working as lecturers ($n = 3$) or supervisors ($n = 8$) at the training institute. Each rater held a university degree in psychology, had successfully completed training either for CBT or PDT, and was state licensed to practice as a psychotherapist. Their professional expertise was based on more than 10 years of therapeutic practice in either CBT or PDT.

Every three months, in addition to reporting the number of sessions provided to patients, therapists reported the number of supervision sessions the therapists had received. The director of the training center compared these data with recorded information from the supervisors.

Patients. A number of patient demographic and potentially clinically relevant variables were assessed. These included patient's age, gender, number of comorbid DSM-IV diagnoses (number of diagnoses), presence of a personality disorder, global overall score (GSI) score at pre-assessment, and position in therapist's order of patients (order within case load, i.e., first patient seen by therapist, second patient seen by therapist, etc.).

Two standardized measures were used:

Structured Clinical Interview for DSM–IV Disorders (SCID; Wittchen et al., 1997). An expert-rated diagnostic interview was conducted at baseline on all patients using the SCID and yielding DSM-IV diagnoses. Standardized administration was supervised regularly by the first author. SCID interviews were conducted by experienced psychotherapists during initial appointments.

The Symptom Checklist Revised (SCL-90 R; Franke, 1995). The well-established German version of this widely used self-report inventory by Franke (1995; German translation of the SCL-90 R from L. R. Derogatis) consists of 90 items that measures patients' subjective physical and psychological distress within the past seven days. For the purposes of this study, the GSI was computed at pre- and post-treatment. The average score on the GSI at pre-assessment was $M = 1.19$ ($SD = .67$; Min. = 0, Max. = 3.5) and at post-assessment $M = .63$ ($SD = .56$; Min. = 0, Max. = 2.6; pre-post effect size $[M_{pre} - post/SD_{post}]$ was ES = 1.0, $n = 264$). Cronbach's alpha for the GSI was .93 in present sample.

Statistical Analyses

The aim of this study was to test the predictive validity of both TRIB assessments using multilevel modeling. Given the nested nature of the data (i.e., patients nested within therapists) two-level random effects, restricted-maximum likelihood (REML) models were fit (with patients modeled as level 1 and therapists level 2). Within these models we tested three therapist characteristics (therapist gender, theoretical orientation [i.e., CBT or PDT], amount of supervision), the TRIB observer ratings (TRIB_motivation [TRIB-I1], TRIB_personal strengths [TRIB-I2], TRIB-G scores) as well as 6 patient characteristics (patients gender, age, number of comorbid diagnoses, presence of a personality disorder, GSI at intake, and order within therapist caseload) each in individual-predictor models (i.e., each patient or therapist predictor was entered individually). To further test the predictive value of the TRIB observer ratings when adjusted for the therapist and patient characteristics, we ran multipredictor models where the above-mentioned six therapist predictors and six patient predictors were included simultaneously. At Level 1,

$$\begin{aligned} \mathrm{GSI}_j = {} & \beta_{j0} + \beta_{j1}(\text{patient gender}) + \beta_{j2}(\text{patient age}) \\ & + \beta_{j3}(\text{number of diagnoses}) \\ & + \beta_{j4}(\text{existence of personality disorder}) \\ & + \beta_{j5}(\mathrm{GSI} - \text{pre}) \\ & + \beta_{j6}(\text{order within case load}) + e_j, \end{aligned}$$

where GSI_j is the estimated post-treatment GSI-value within each patient$_j$, β_{j0} represents the grand-mean centered intercept of the repeated GSI-values by each therapist, and β_{j1}–β_{j6} represent the specific estimates of the predictors on intercept GSI_j. Further, e_j represents the corresponding error term.

At Level 2:

$$\beta_{j0} = \gamma_{00} + \gamma_{01}(\text{TRIB} - \text{I1}) + \gamma_{02}(\text{TRIB} - \text{I2}) + \gamma_{03}(\text{TRIB} - \text{G}) + \gamma_{04}(\text{therapist gender}) + \gamma_{05}(\text{amount of supervision}) + \gamma_{06}(\text{theoretical orientation}) + u_{j0},$$

where β_{j0} is the estimate of the true population overall outcome, γ_{00} is the estimated overall grand mean, γ_{j1}–γ_{y6} represent the specific estimates of the predictors on intercept β_{j0}, and u_{j0} is the error at the patient level. Multilevel models were estimated using the HLM software (Raudenbush & Bryk, 2002).[1]

Results

Prediction of Therapy Outcomes

The random variance component of the two-level null model indicated that the integration of fixed predictors into the equation is justified (null model: $X^2(40) = 59.2$, $p = .03$; deviance [2] = 436.8). Therapist-level *ICC* indicated 6.6% variability at Level 2 ($u_{j0}/[e_j + u_{j0}] = 0.02/[0.28 + 0.02] = 0.066$). The fixed effects of the individual and multipredictor analyses are presented in Table II. Observer-rated TRIB of the therapists assessed in the group format (TRIB-G) but not the expert-rated interview format (TRIB-I) was associated with better outcomes (lower post-treatment GSI scores) in both models. With regard to patient variables, only the number of comorbid diagnosis was found to influence outcomes in both analysis. The model comparison likelihood ratio test of the unconditional model with the individual-predictor TRIB-G model indicated better fit for the individual-predictor model (deviance [4] = 429.5; $X^2[2] = 7.3$, $p = .025$). Furthermore, this test indicated better fit for the multipredictor model in comparison to the individual-predictor model (deviance [15] = 343.8; $X^2[13] = 85.7$, $p < .001$) as well as the unconditional model ($X^2[13] = 93.0$, $p < .001$). The random variance component of the multipredictor model reached non-significance, indicating that there was no remaining random variance in this model ($X^2[34] = 30.8$, $p > .5$).

Discussion

The main objective of the current study was to develop and evaluate a standardized behavioral assessment protocol for psychotherapy training programs. This assessment tool was based on extant literature that has emphasized therapists' interpersonal capacities as a key predictor of outcome (Ackerman & Hilsenroth, 2003). This endeavor was theoretically based on the idea of pre-training psychotherapeutic talent (Orlinsky & Rønnestad, 2005) and is set against the backdrop of increasing acknowledgement of and interest in between-therapist differences in outcomes (i.e., therapist effects; Baldwin & Imel, 2013; Stangier, 2015; Wampold & Imel, 2015). Although there were several studies on predictors of a psychotherapeutic talent conducted in the 1970s and 1980s, there is still no consensus regarding appropriate means to assess these capacities. Theoretically therapeutic talent is in part characterized by interpersonal competencies that include clear and positive communication, empathy, and communicative attunement in social relationships, respectful and warm interpersonal behavior, the respectful management of criticism, and a willingness to cooperate (e.g., Ackerman & Hilsenroth, 2003). In addition, general mental abilities such as motivation and individual strengths were assessed in the present structured interview.

The development of the TRIB relied heavily on item face validity, as the selection of the items was based on empirical findings of pantheoretical characteristics of successful therapists (Orlinsky & Rønnestad, 2005). The good *ICC*s of the experts of both traditions (CBT and PDT; *ICC* > .61) indicated, the items of both TRIB assessments might be evaluated independently of the underlying treatment tradition. Nonetheless, as this type of research is quite time consuming due in part to small training classes, future studies may benefit from multisite designs through the establishment of collaborative relations among several training institutes.

The major aim of the present study was to examine the relative importance of therapist factors on therapy outcome (effect of the second level against null model) and assess the predictive value of a rating scale used to assess TRIB of trainee therapists using observer- and expert-rated assessments. The proportion of variance due to therapists (*ICC*, 6.6%) is consistent to results of prior research in this area (Baldwin & Imel, 2013; Chow et al., 2015).

No therapist variables (including TRIB scores) were significant predictors of early termination. With regard to patient variables, only the presence of a personality disorder has the tendency to be associated with premature termination or missing data post-treatment. However, the analyses of patient symptom change indicated that those therapists who demonstrated strong interpersonal skills, as rated by trained observers within a group assessment format, achieved better outcomes with their patients. Indeed, the relationship between trainees'

Table II. Multilevel model results predicting post-treatment psychological symptoms (GSI).

	Individual-predictor models			Multipredictor model		
	Coefficient	SE	$t(261)$	Coefficient	SE	$t(425)$
Intercept (β_{j0})				.65	.026	24.6***
Patient gender (β_{j1})	.012	.075	0.2	.039	.061	.6
Patient age (β_{j2})	.007	.003	2.5*	.003	.002	1.2
Numbers of diagnoses (β_{j3})	.099	.033	3.0**	.120	.034	3.4**
Personality disorder (β_{j4})	.079	.081	1.0	−.063	.084	−.7
GSI intake (β_{j5})	.408	.031	13.1***	.352	.039	9.1***
Order within caseload (β_{j6})	−.004	.008	−.1	.002	.009	1.7
Therapist gender (γ_{01})	−.080	.082	−1.0	−.049	.069	−.7
Theoretical orientation (γ_{02})	.203	.090	2.3*	.105	.079	1.3
Amount of supervision (γ_{03})	.005	.008	.6	.001	.009	1.0
TRIB-I1 (γ_{04})	.072	.115	.6	.039	.081	−.4
TRIB-I2 (γ_{05})	.063	.076	.8	.035	.051	.7
TRIB-G (γ_{06})	−.017	.004	−2.6*	−.012	.006	−2.1*

*$p < .05$.
**$p < .01$.
***$p < .001$.

interpersonal capacities (assessed prior to formal training in the provision of therapy) remained a predictor of patient outcomes when simultaneously controlling for a host of patient and therapist variables (e.g., theoretical orientation, patient personality disorder diagnosis, amount of supervision).

Despite promising results for the TRIB-G, it remains an open question whether interpersonal skills measured pre-training shows a direct or indirect influence on treatment outcome. In case of direct prediction, it is theoretically plausible that therapists' interpersonal skill itself facilitates change. Indeed, the interpersonal posture of the therapist is highlighted as a causal mechanism of change in various theories of psychotherapy (e.g., person-center therapy, interpersonal process; Rogers, 1961; Teyber & McClure, 2011). The influence could likewise be indirect; therapist interpersonal skills could be associated with other important variables in psychotherapy process. For example, it could be that therapist trainees with high interpersonal skills also show an extended degree of deliberate practice in the training which therefore indirectly accounts for superior therapy outcomes (Chow et al., 2015). Regardless of whether direct and/or indirect causality is present, one can safely assume within the longitudinal design that the causal direction is not reversed (i.e., treatment success does not result in superior interpersonal skills of therapy trainees).

Given the relatively small sample size of both patients and therapists as well as the fairly long time-frame over which patient outcomes were assessed (i.e., five years), the fact that a brief assessment of trainees' interpersonal capacities predicted outcomes may speak to the promise of the TRIB. The performance of the TRIB in predicting outcomes in this sample are on par with those reported in the industrial psychology literature when examining similar assessments and their links with job performance as well as the feedback from executives.

The inclusion of already selected candidates to examine the predictive validity may have reduced the variance in the TRIB assessments, which may or may not impact study results. Of note, limiting the sample to trainees who were selected for training theoretically results in a more conservative test of the TRIB-I assessments (again due to a presumably restricted range). Limiting the sample may, however, has limited our ability to detect non-linear relationships. For example, it may be well true that in some TRIB items may show, for example, a U-shaped relationship with patient outcomes (e.g., limited facial expression). Unfortunately, even in future studies it would not be possible to collect therapeutic outcomes from a population that was excluded from the training.

Several other limitations are still important to acknowledge. Other therapy-related therapist characteristics, such as the amount of previous therapy experience of the candidates, were not assessed. While psychotherapy experience in naturalistic conditions may not correlate with psychotherapy expertise (e.g., Goldberg et al., in press; Tracey, Wampold, Lichtenberg, & Goodyear, 2014), prior therapeutic experience before postgraduate training may still be an indicator of a trainee's engagement in and motivation for developing therapy-related interpersonal skills. Furthermore, future investigations should take a more precise look at other trainees characteristics such as academic performance

and personality traits (e.g., agreeableness, openness to experience), which are often used in industrial and organizational psychology assessments (e.g., Hough & Oswald, 2008).

It may be valuable, in future research, to investigate populations of trainees with less prior knowledge of psychotherapy and communication (e.g., master's degree applicants). This would help assess the generalizability of findings to more therapy naïve populations. Furthermore, future research could examine links between TRIB scores and outcomes in disorder-populations of outpatients as well as within more severely burdened inpatient settings. Along therapeutic outcomes, assessments of process characteristics (e.g., alliance) may provide a more precise picture regarding potential moderators or mediators of the relationship between the TRIB scores and patient outcomes. Along these lines, other pre-training therapists' characteristics such as academic grade or prior experience/engagement in psychotherapy practice might give a more precise picture about potential correlates of the TRIB assessment.

The current research may have implications for future methodological and training directions. Methodologically, the current work supports the development of new methods for studying psychotherapy process, and is in keeping with other recent studies employing non-self-report designs (e.g., peer ratings; Goldberg & Hoyt, 2015; observer ratings; Flückiger, Zinbarg, Znoj, & Ackert, 2014). Future work may do well to examine the training potential implied by our results. For example, based on the current study, TRIB-G scores could be used to develop specialized interpersonal skills training as a compliment to traditional psychotherapy training. Work in this direction could increase our knowledge regarding effective methods for the training of therapists.

Keeping both this study's limitations and the potential for future research in mind, the present promising but preliminary evidence of the outcome-relevance of therapists' pre-training characteristics may stimulate future research in the area of therapist characteristics, therapist effects and therapist training (Wampold & Imel, 2015).

Funding

This grant partly was sponsored by the Swiss National Science Foundation (SNSF) [grant number PZ00P1_136937] awarded to Christoph Flückiger.

Note

[1] The intercorrelations of the all predictors were $r < .52$ and there was no indication for multicollinearity (VIF < 1.5). Furthermore, we predicted missing GSI-data at post-treatment (= intent-to-treat sample [$n = 291$] vs. completer sample [$n = 264$]). None of the above-mentioned predictors indicated significance on multilevel binary logistic models ($p > .10$) except the trend of *existence of personality disorder* that was associated with higher missing data (OR = 2.1, 95-CI = 0.9/5.0, $p = .084$). Within the MLM for GSI at post-treatment, we also tested a random intercept and random slope model, where a linear slope of patients within therapist was tested. The random effect of the slope indicated not enough variability (variance component = .0001; $X^2[36] = 33.7$, $p > .5$) and therefore we did not include predictors in this longitudinal model (deviance [4] = 429.5; model comparison: $X^2[2] = 3.1$, $p = .21$). For comparable results, see Goldberg et al. (in press).

Supplemental data

Supplemental data for this article can be accessed at http://dx.doi.org/10.1080/10503307.2015.1125546.

ORCID

HENNING SCHÖTTKE http://orcid.org/0000-0003-0049-0217

References

Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of therapist characteristics and technique positively impacting the therapeutic alliance. *Clinical Psychology Review*, *23*, 1–33. doi:10.1016/S0272-7358(02)00146-0

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

Arthur, W., Day, A. E., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, *56*, 125–154. doi:10.1111/j.1744-6570.2003.tb00146.x

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* (pp. 258–297). New York, NY: Wiley.

Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice*, *17*(1), 1–30. doi:10.1111/j.1468-2850.2009.01187.x

Bergin, A. E. (1997). Neglect of the therapist and the human dimensions of change: A commentary. *Clinical Psychology*, *4* (1), 83–89. doi:10.1111/j.1468-2850.1997.tb00102.x

Binder, J. L., & Strupp, H. H. (1997). 'Negative process': A recurrently discovered and underestimated facet of the therapeutic process and outcome in the individual psychotherapy of adults. *Clinical Psychology: Science and Practice*, *4*(2), 121–139. doi:10.1111/j.1468-2850.1997.tb00105.x

Bohart, A. C., Elliott, R., Greenberg, L. S., & Watson, J. C. (2002). Empathy. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 89–108). New York, NY: Oxford University Press.

Boswell, J. F., & Castonguay, L. G. (2007). Psychotherapy training: Suggestions for core ingredients and future research.

Psychotherapy: Theory, Research, Practice, Training, *44*(4), 378–383. doi:10.1037/0033-3204.44.4.378

Castonguay, L. G., Boswell, J. F., Constantino, M. J., Goldfried, M. R., & Hill, C. E. (2010). Training implications of harmful effects of psychological treatments. *American Psychologist*, *65* (1), 34–49. doi:10.1037/a0017330

Chow, D. L., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J. A., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy*, *52*(3), 337–345. doi:10.1037/pst0000015.supp

Costanzo, M., & Philpott, J. (1986). Predictors of therapeutic talent in aspiring clinicians: A multivariate analysis. *Psychotherapy*, *23*(3), 363–369. doi:10.1037/h0085624

Crits-Cristoph, P., & Gallop, R. (2006). Therapist effects in the national institute of mental health treatment of depression collaborative research program and other psychotherapy studies. *Psychotherapy Research*, *16*(2), 178–181. doi:10.1080/10503300500265025

Del Re, A. C., Flückiger, C., Horvath, A. O., Symonds, D., & Wampold, B. E. (2012). Therapist effects in the therapeutic alliance-outcome relationship: A restricted-maximum likelihood meta-analysis. *Clinical Psychology Review*, *32*(7), 642–649. doi:10.1016/j.cpr.2012.07.002

Dennhag, I., & Armelius, B.-A. (2012). Baseline training in cognitive and psychodynamic psychotherapy during a psychologist training program. Exploring client outcomes in therapies of one or two semesters. *Psychotherapy Research*, *22*(5), 515–526. doi:10.1080/10503307.2012.677332

Dilchert, S., & Ones, D. S. (2009). Assessment center dimensions: Individual differences correlates and meta-analytic incremental validity. *International Journal of Selection and Assessment*, *17*(3), 254–270. doi:10.1111/j.1468-2389.2009.00468.x

Eversmann, J. (2008). *Psychometrische Überprüfung eines Auswahlverfahrens psychotherapeutischer Ausbildungsteilnehmer* [Psychometric evaluation of an assessment procedure for postgraduate student therapists] (Unpublished doctoral dissertation). Universität Osnabrück, Germany.

Eversmann, J., Schöttke, H., Wiedl, K. H., & Rogner, J. (2011). Die Beobachtungsskala "Therapie-relevante interpersonelle Verhaltensweisen" (TRIB) in der Auswahl von Teilnehmern der Psychotherapieausbildung [The observational rating scale Therapy-Related Interpersonal Behaviors (TRIB) in the selection of trainee psychotherapists]. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *40*(1), 11–21. doi:10.1026/1616-3443/a000062

Flückiger, C., Zinbarg, R. E., Znoj, H. J., & Ackert, M. (2014). Resource activation in generalized anxiety disorder: An observer-based microprocess analysis of in-session outcomes. *Psychotherapy*, *51*, 535–545. doi:0.1037/a0034119

Franke, G. H. (2002). *SCL-90 R. Die Symptom Checkliste von L. R. Derogatis—Deutsche Version. 2., vollst. überarb. und neu normierte Aufl* [SCL-90-R. Symptom Checklist from L. R. Derogatis—German version, 2nd rev. ed.]. Göttingen, Germany: Beltz Test.

Goldberg, S. B., & Hoyt, W. T. (2015). Group as social microcosm: Within-group interpersonal style is congruent with outside group relational tendencies. *Psychotherapy*, *52*(2), 195–204. doi:10.1037/a0038808

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (in press). Do psychotherapists improve with time and experience? A longitudinal analysis of real world outcome data.

Goodman, G. (1972). *Companionship therapy: Studies in structured intimacy*. San Francisco, CA: Jossey-Bass.

Hackmann, J. R., Wiggins, N., & Bass, A. R. (1970). Prediction of long-term success in doctoral work in psychology. *Educational and Psychological Measurement*, *30*, 365–374.

Hartley, D., & Strupp, H. (1983). The therapeutic alliance: Its relationship to outcome in brief psychotherapy. In F. Masling (Ed.), *Empirical studies of psychoanalytic theories* (Vol. 1, pp. 1–27). Hillsdale, NJ: Erlbaum.

Henry, W. P., Schacht, T. E., Strupp, H. H., Butler, S. F., & Binder, J. L. (1993). Effects on training in time-limited dynamic psychotherapy: Mediators of therapists responses to training. *Journal of Counseling and Clinical Psychology*, *61*(3), 441–447. doi:10.1037/0022-006X.61.3.434

Hermelin, E., Lievans, F., & Robertson, I. T. (2007). The validity of assessment centres for the prediction of supervisory performance ratings: A meta-analysis. *International Journal of Selection and Assessment*, *15*(4), 405–411. doi:10.1111/j.1468-2389.2007.00399.x

Hermer, M., & Röhrle, B. (2008). Therapeutische Beziehungen: Geschichte, Entwicklungen und Befunde [Therapeutic relationship: History, development and results]. In M. Hermer & B. Röhrle (Eds.), *Handbuch der therapeutischen Beziehung [Handbook of therapeutic relationship]* (pp. 15–105). Tübingen, Germany: DGVT.

Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review*, *30*, 448–466. doi:10.1016/j.cpr.2010.02.005

Hilsenroth, M. J., Ackerman, S. J., Clemence, A. J., Strassle, C. G., & Handler, L. (2002). Effects of structured clinician training on patient and therapist perspectives of alliance early in psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, *39*(4), 309–323. doi:10.1037/0033-3204.39.4.309

Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial-organizational psychology: Reflections, progress and prospects. *Industrial and Organizational Psychology*, *1*, 272–290.

Kim, D., Wampold, B., & Bolt, D. (2006). Therapist effects in psychotherapy: A random-effects modeling of the national institute of mental health treatment of depression collaborative research program data. *Psychotherapy Research*, *16*(2), 161–172. doi:10.1080/10503300500264911

Kuhr, A. (1998). Strukturqualität in der Verhaltenstherapieausbildung und ihre Sicherung [Structural quality in cognitive-behavioral therapy training program]. In A. R. Laireiter & U. Willutzki (Eds.), *Ausbildung in Verhaltenstherapie* [Training in cognitive-behavioral therapy] (pp. 362–382). Göttingen, Germany: Hogrefe.

Laska, K. M., Smith, T. L., Wislocki, A. P., Minami, T., & Wampold, B. E. (2013). Uniformity of evidence-based treatments in practice? Therapist effects in the delivery of cognitive processing therapy for PTSD. *Journal of Counseling Psychology*, *60*(1), 31–41. doi:10.1037/a0031294

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects and therapist differences in a large mental health provider network. *Journal of Counseling Psychology*, *54*, 32–39. doi:10.1037/0022-0167.54.1.32

Mallinckrodt, B. (2000). Attachment, social competencies, social support and interpersonal processes in psychotherapy. *Psychotherapy Research*, *10*(3), 239–266. doi:10.1093/ptr/10.3.239

Margraf, J., & Schneider, S. (2008). *Lehrbuch der verhaltenstherapie* [Textbook to cognitive-behavioral therapy]. Berlin: Springer.

Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, *93*(5), 1042–1052. doi:10.1037/0021-9010.93.5.1042

Mohl, P. C., Martinez, D., Ticknor, C., Huang, M., & Cordell, L. (1991). Early dropouts from psychotherapy. *Journal of Nervous and Mental Disease*, *179*, 478–481.

Muran, J. C., Safran, J. D., Samstag, L. W., & Winston, A. (2005). Evaluating an alliance-focused treatment for personality disorders. *Psychotherapy: Theory, Research, Practice, Training*, *42* (4), 532–545. doi:10.1037/0033-3204.42.4.532

Nissen-Lie, H. A., Monsen, J. T., & Rønnestad, M. H. (2010). Therapist predictors of early patient-rated working alliance: A multilevel approach. *Psychotherapy Research*, *20*(6), 627–646. doi:10.1080/10503307.2010.497633

Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology and Psychotherapy*, *10*, 361–373. doi:10.1002/cpp.383

Orlinsky, D. E., & Rønnestad, M. H. (2005). *How psychotherapists develop. A study of therapeutic work and professional growth*. Washington, DC: American Psychological Association.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear modeling*. Thousand Oaks, CA: SAGE.

Raue, P. J., Goldfried, M. R., & Barkham, M. (1997). The therapeutic alliance in psychodynamic-interpersonal and cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology*, *65*(4), 582–587. doi:10.1037/0022-006X.65.4.582

Reckewerth, R. (2006). *Validierung der "Therapy-related interpersonal behavior scale" und die Entwicklung eines Beurteilertrainings* [Validation of the "Therapy-related interpersonal behavior scale" and the development of an inter-rater training] (Unpublished diploma thesis). Universität Osnabrück, Germany.

Robinson, D. D. (1981). Content-oriented personnel selection in a small business setting. *Personnel Psychology*, *34*, 77–87. doi:10.1111/j.1744-6570.1981.tb02179.x

Rogers, C. (1961). *On becoming a person: A therapist's view of psychotherapy*. New York, NY: Houghton Mifflin.

Rønnestad, M. H., & Skovholt, T. M. (1997). Berufliche Entwicklung und Supervision von Psychotherapeuten [The professional development and supervision of psychotherapists]. *Psychotherapeut*, *42*, 299–306.

Rønnestad, M. H., & Skovholt, T. M. (2005). Die professionelle Entwicklung von Psychotherapeuten während der Ausbildung [The professional development of psychotherapists during their training]. In A.-R. Laireiter & U. Willutzki (Eds.), *Ausbildung in Verhaltenstherapie* [Training in cognitive-behavioral therapy] (pp. 102–120). Göttingen, Germany: Hogrefe.

Safran, J. D., Muran, J. C., Samstag, L. W., & Stevens, C. (2002). Repairing alliance ruptures. In J. C. Norcross (Ed.), *Psychotherapy relationships that work* (pp. 235–254). New York, NY: Oxford University Press.

Safran, J. D., Muran, J. C., Samstag, L. W., & Winston, A. (2005). Evaluating alliance-focused intervention for potential treatment failures: A feasibility study and descriptive analysis. *Psychotherapy: Theory, Research, Practice, Training*, *42*(4), 512–531. doi:10.1037/0033-3204.42.4.512

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. doi:10.1037/0033-2909.124.2.262

Schöttke, H. (2014). *Zur Vorhersage des Therapie- und Ausbildungserfolges durch Therapeutenmerkmale* [Prediction of therapy and training success by the use of therapist variables]. Symposium conducted at the 32th meeting of the group for clinical psychology and psychotherapy of the German Psychological Society, Braunschweig, Germany.

Schöttke, H., Trame, L., & Sembill, A. (2014). Relevance of therapy goals in outpatient cognitive-behavioral and psychodynamic psychotherapy. *Psychotherapy Research*, *24*(6), 711–723. doi:10.1080/10503307.2014.881578

Skovholt, T. M., & Rønnestad, M. H. (1995). *The evolving professional self: Stages and themes in therapist and counselor development*. Chichester: Wiley.

Stangier, U. (2015). Psychotherapeutische Kompetenzen in der kognitiven Verhaltenstherapie [Psychotherapist competencies in cognitive behavioral therapy]. *Psychotherapeut*, *60*, 193–198. doi:10.1007/s00278-015-0022-y

Stricker, G., & Huber, J. T. (1967). The GRE and undergraduate grades as predictors of success in graduate school. *Journal of Educational Research*, *60*, 466–468.

Teyber, E., & McClure, F. H. (2011). *Interpersonal process in therapy: An integrative model* (6th ed.). Belmont, CA: Brooks/Cole.

Tracey, T. J., 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. E., & Brown, G. S. J. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, *73*(5), 914–923. doi:10.1037/0022-006X.73.5.914

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

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *78* (2), 200–211. doi:10.1037/a0018912

Whiteley, J. M., Sprinthall, N. A., Mosher, R. L., & Donaghy, R. T. (1967). Selection and evaluation of counselor effectiveness. *Journal of Counseling Psychology*, *14*(3), 226–234.

Wittchen, H.-U., Zaudig, M., & Fydrich, T. (1997). *SKID. Strukturiertes Klinisches Interview für DSM-IV (Achse I und II)* [Structured clinical interview for DSM-IV (Axis 1 and 2)]. Göttingen: Hogrefe.

Wöller, W., & Kruse, J. (2010). *Tiefenpsychologisch fundierte Psychotherapie* [Psychodynamic psychotherapy]. Stuttgart: Schattauer.