Journal of Contemporary Psychotherapy (2019) 49:87–97
https://doi.org/10.1007/s10879-018-9405-z

ORIGINAL PAPER

Expanding the Conceptualization of Outcome and Clinical Effectiveness

Michael Helge Rønnestad[1] · Helene A. Nissen-Lie[1] · Hanne Weie Oddli[1] · Kirsten Benum[1] · Vidar Blokhus Ekroll[1] · Siri E. Gullestad[1] · Hanne Haavind[1] · Sissel Reichelt[1] · Marit Råbu[1] · Erik Stänicke[1] · Anna Louise von der Lippe[1] · Margrethe Seeger Halvorsen[1]

Published online: 10 September 2018

Abstract
There is a need for studies that advance our knowledge of therapist effectiveness, expanding the definition of what constitutes therapeutic success. In the present study, four aspects of clinical outcome were analyzed using a sample of highly experienced psychotherapists (mean years of experience as therapy practitioners = 30) who also serve as clinical teachers. The four aspects were: (1) overall change in different outcome domains, including long-term outcome; (2) outcome with clinically distressed clients (i.e. clients above the level of clinical disturbance at pretreatment); (3) level of drop-out; and (4) the degree to which clients re-entered treatment during the follow-up period. Participants were 48 clients treated by 18 therapists. Treatments were open-ended, 'therapy as usual', mostly within the setting of independent private practice. Outcome measurements included the OQ-45.2 and IIP-64, with repeated measurements during therapy and extended long-term follow up (up to 3.5 years). Cohen's *d* effect sizes and multilevel growth trajectories were estimated for patient change in symptomatic and interpersonal distress. The findings indicate that these therapists obtain very good outcomes overall, particularly with clients above cut-off for clinical disturbance. Moreover, there was a high level of maintenance and increase of treatment gains over time. No clients dropped out. Overall, the findings suggest that a sample of highly experienced therapists obtains particularly good outcomes as defined in more comprehensive terms than what is typical in psychotherapy research. Implications for the understanding of therapeutic success and clinical expertise are discussed.

Keywords Client outcomes · Clinical expertise · Client drop out

Psychotherapy research relies heavily on patient outcome as operational definition of therapeutic success, which in turn legitimizes practice. Outcome is typically conceptualized as patient change in symptoms from before to after treatment. Correspondingly, therapist effectiveness is also operationally defined on the basis of patient outcomes, and refers to the therapist's ability to produce positive client change (see Baldwin and Imel 2013). However, a definition based on patient pre to post change may be limited since it does not take into account the degree to which potential gains resulting from psychotherapy are maintained over time, the level of drop-out from psychotherapy and degree to which clients re-enter therapy after the end of treatment. Obviously, clients, therapists and health administrators, as well as policy makers, would expect that psychotherapy produces (at least relatively speaking) enduring effects so that the investment is worthwhile. Hence, we need to adopt a more comprehensive operational definition of what constitutes a good outcome in psychotherapy in order to study therapeutic success and evaluate the benefits of psychotherapy.

Furthermore, research on the relationship of therapist experience to therapy outcome has not provided a clear answer as to whether increased years of practice leads to therapists becoming more effective. On the contrary, this research area has produced somewhat counterintuitive findings; some studies find that therapist experience is not related to, or even negatively associated with, patient outcomes (e.g., Goldberg et al. 2016a; Tracey et al. 2014).

In order to expand the scope of clinical outcomes as well as shed light on the possible link between patient outcomes and the clinician's professional experience, in this study we

✉ Michael Helge Rønnestad
m.h.ronnestad@psykologi.uio.no

[1] Department of Psychology, University of Oslo, Pb. 1094 Blindern, 0317 Oslo, Norway

included a broader spectrum of clinical outcomes obtained by a select sample of highly experienced psychotherapists in private practice who, in addition, are also respected teachers of psychotherapy. Although there are some indications that more experienced therapists may obtain better outcomes and higher alliances with more challenging or more severely distressed clients (Tschuschke et al. 2015), the empirical support for these indications is limited. By studying clients presenting with varying levels of clinical disturbance, we will reflect on the potential relationship of therapist experience and variations in clinical disturbance.

Our study thus has several aims. First, the study will focus on different aspects of clinical outcomes (i.e., change for clients with high initial clinical disturbance; level of drop-out and re-entering therapy), ensuring a richer and more nuanced operationalization of outcome from psychotherapy. Second, the data collection is as long as 3.5 years post-therapy ensuring a possibility to assess maintenance of change over time. Third, the sample of therapists is characterized by varied and extensive experience as clinicians, teachers and clinical authors, which presumably provides them with an opportunity to reflect on their own clinical practice and enables us to move beyond the question of the impact of clinical experience merely in terms of years of practice. What follows next is a brief presentation of the empirical literature in these realms.

Psychotherapy with Clients Who Are Clinically Distressed

Patients do not only seek therapy to get relief from symptomatic distress, but could do so for other reasons such as to explore existential concerns, to process issues around self-concept and personal identity, or to get assistance in making important life-decisions. These concerns are certainly worthy of psychotherapeutic attention. There are however reasons to study clients who present with clinical disturbance or increased psychopathology, which is the emphasis in this article since there is some support to suggest that inexperienced and experienced therapists may differ in their effectiveness in treating clients with more distress (cf. Keefe et al. 2014).

Recent analyses show that therapist effects (the proportion of patient outcomes that can be attributed to individual therapists; see Baldwin and Imel 2013) increase with initial level of clinical distress in patients (Dinger et al. 2016; Saxon and Barkham 2012), as well as when treatment time increases (Goldberg et al. 2016b). This indicates that for less disturbed clients and in short-term treatments, therapists are more or less equally effective (i.e. therapist effects are small), whereas when therapy is more challenging and longer lasting, the therapist makes more of a difference to client change. Whether experience has an influence in this is an important question for further investigation. In their review, Horvath and Bedi (2002) suggested that inexperienced therapists appeared to have more difficulties with "attuning to relational difficulties with severely impaired clients than their more experienced colleagues" (p. 60). In the same vein, Skovholt et al. (1997) suggested that expertise in the alliance domain involved "the ability to establish, maintain, and creatively use a positive working relationship with highly distressed—angry, depressed, rebellious, disturbed—individuals" (p. 363). Correspondingly, in a discussion of therapist expertise in psychotherapy, Hill et al. (2017) suggest that "experienced therapists are better at dealing with the more complex aspects of the therapeutic relationship" (p. 13).

Long Term Follow-Up and Maintenance of Therapeutic Gains

Outcome studies using long-term follow-up measurement are rare even if both clients and therapists likely hope that potential treatment gains of psychotherapy are lasting (Ekroll and Rønnestad 2018). The relationship between duration of therapist professional experience and truly long-term effects (defined as > 3 years post-therapy follow-up) in treatments within out-patient/independent private practice settings has received insufficient research interest. In the meta-analysis of 23 studies of long-term psychodynamic psychotherapy of patients with complex and comorbid mental disorders by Leichsenring and Rabung (2008), the mean number of weeks follow-up was 93.23 (approximately 21 months). It is noteworthy that in their extensive review, only two studies were identified with follow-up data beyond 2 years (Monsen et al. 1995; Sandell et al. 2000). In a recent meta-analysis comparing effects of 20 psychodynamic versus other bona fide treatments, Kivlighan III et al. (2015) identified only one study with results beyond 2 years follow-up, while the modal number was only 12 months. There is indeed a need for more studies on the long-term effects of psychotherapy beyond 1–2 years to examine if treatment effects last over time. It is possible that with more experience therapists become increasingly skilled at stimulating client agency or other psychological resources, which could enable further growth when the treatment is over (von der Lippe et al. 2017). The longitudinal design of the current naturalistic 'therapy-as-usual' study includes process, outcome, and follow-up data extending to 3.5 years post therapy.

It is plausible that the potential competencies acquired by highly experienced therapists may not so much be demonstrated as client change *at the end of therapy* or in the short run but as sustained or increased change in a more extended follow-up period. Maintenance of gains is an important issue

both from an individual client perspective as well as from a societal cost-utility perspective. Relapse of symptoms has long been recognized as a major problem in the field of mental health (Ekroll and Rønnestad 2018). It is possible that therapist experience plays a role in the maintenance of treatment gains. We will summarize briefly the empirical literature on therapist experience level and outcome below.

Psychotherapists' Experience Level, Patient Outcome and Drop-Out

An extensive literature review by Tracey et al. (2014) concludes that there is no established relationship of client outcome with therapist experience (usually defined as years of practice). In reaching this conclusion, the authors draw on a range of conceptual and empirical contributions (e.g., Baldwin and Imel 2013; Okiishi et al. 2003). Other research studies and reviews support this conclusion (e.g., Hubble et al. 2009). This negative conclusion was also supported by a recent analysis by Goldberg et al. (2016a), reporting a small but statistically significant *inverse* relationship between therapist experience and client outcome.

Even though the dominant voice disconfirms a positive experience-outcome relationship, some other voices can also be heard. In a review of therapist variables, Beutler et al. (2004) scrutinized meta-analyses which failed to find an experience-outcome relationship, but located studies which effectively differentiated therapist experience level from extent of training. These found a positive therapist experience-patient outcome relationship. In a recent naturalistic study (Tschuschke et al. 2015), a positive effect of therapists' increased experience was found, especially for clients with more severe problems. In the most comprehensive meta-analysis to date of controlled studies on psychodynamic therapy outcomes for anxiety disorder, Keefe et al. (2014) concluded: "More experienced therapists may be better able to recognize and maintain a competent therapeutic focus on typical dynamics underlying specific anxiety disorders" (p. 320).

As an indirect aspect of client outcome, drop-out is another parameter potentially relevant to the understanding of the therapist experience-outcome relationship. In a review by Roos and Werbart (2013), a weighted dropout rate of 35% was found. The authors noted that rates varied greatly and that "…clients often explain their discontinuation of treatment with dissatisfaction with the therapist or therapy" (p. 396). In this review, the authors also noted that higher levels of therapist experience predicted lower rates of drop-out. Based on this literature we would expect low rates of drop-out in the present study with a highly experienced therapist sample.

Aims and Research Questions

Using a broad definition of outcome, and a well-defined sample of highly experienced therapy practitioners, we hope to broaden the perspectives on clinical outcomes and therapeutic effectiveness in the current study. We investigated four dimensions of patient outcome: (a) overall outcome through treatment and follow-up; (b) outcome with clients above the level of clinical disturbance at pretreatment (c) level of drop-out, and (d) the degree to which clients have re-entered treatment during the follow-up period. The following five research questions were explored:

1. What are the overall outcomes, as measured by standard outcome measures (i.e., the OQ-45 and IIP-64) for clients in psychotherapies conducted by a highly experienced therapist sample?
2. What are the outcomes for clients with high levels of symptomatic distress (i.e., in the clinical range at pretreatment)?
3. To what degree do clients in these therapies maintain their gains at follow up?
4. What is the drop-out rate for clients in these therapies?
5. What proportion of clients re-entered therapy during the 3.5 years follow-up period?

Method

Setting

This paper presents the results from a longitudinal, naturalistic study of clients in the independent private practices of 18 psychotherapists who were highly experienced as therapists, teachers and supervisors. For a description of the study, see Rønnestad (2009). Six of the therapists had academic appointments in a university setting (including three professors), and consequently had a part-time practice as therapists. The others were full-time clinicians, also with experience as teachers at private and/or public training institutions. Of the 17 therapists in private practice, all but one had contracts with the county for partial, third-party reimbursement, which implies that therapy was generally provided at a low cost for the clients.

Treatments

Forty-six of the 48 clients who participated received treatment in the context of the psychotherapists' private practice, while two clients received therapy in a public out-patient clinic. All therapies were open-ended. Some of the therapies

were psychodynamic in the more pure form, while the majority of treatments were integrative as described in the section on Therapists below. Treatment duration, defined as the number of sessions, varied greatly, with a low of eight sessions and a high of 216 (see Table 1). The mean number of sessions for all clients was 64.9 (*Mdn* = 46). For clients above cut-off on OQ-45, the mean number of sessions was 76.3 (*Mdn* = 63), while the mean number of sessions for clients below cut-off was 40 (*Mdn* = 31).

Participants

Therapists

Twenty-two psychotherapists, who were all highly experienced and who also had experience as psychotherapy teachers, were invited to participate. Of these, 19 therapists agreed to participate initially, but one therapist dropped out early and was not included in the study. Of the 18 therapists who participated, 15 were licensed psychologists with a specialist status in the national association of psychologists (a status awarded after a minimum of 5 years post-licensure training and practice including extensive supervision and course-work). Three of the therapists were medical doctors with a specialty in psychiatry. The mean number of years of therapy practice was 30, varying between 15 and 45, while the mean number of years of experience in private practice was 17. Therapists' mean age when they started with their first clients in the project was 63. The majority had extensive experience as teachers of psychotherapy (*M* = 19 years), and all but one had also been sought out for personal therapy by several other therapists. This variety in professional roles is reflected in the data of clients seen by each therapist. Based on data from 17 therapists (missing data from one therapist on number of clients seen per week) the mean number of clients seen per week was 12.8. Eight therapists saw between four and eight clients per week, four therapists saw between 10 and 15 clients per week, while five therapists saw between 20 and 25 clients per week.

To assess theoretical orientations, the therapists were asked to rate the degree to which their work was influenced by each of the following on a 6-point scale from 0 ('not at all') to 5 ('very much'): analytic/psychodynamic, behavioral, cognitive, humanistic, systemic, and other (with space to specify which). Using a criterion of *moderately to strongly influenced* (defined as ratings of 3, 4 or 5) the most frequently endorsed theoretical orientations for the 15 therapists (missing data on theoretical orientations were from three therapists): Cognitive or Cognitive Behavioral (13), Analytic/Psychodynamic (12), Humanistic (11), and Systemic (11). Therapists varied in terms of how many theoretical orientations they rated as *moderately to strongly* influencing their practice. Only two said they were moderately or strongly influenced by a single orientation, six were influenced by three orientations at this level, and nine claimed to be moderately or strongly influenced by four or more orientations. Thus, all but two (both psychodynamically oriented) could be considered 'broad spectrum' integrative/eclectic in their theoretical orientation.

Table 1 Therapy duration defined as number of sessions

	All clients	OQ-45 ≥ 63	OQ-45 < 63
N	48	33	15
Mean	64.9	76.3	40.0
SD	52.1	56.8	28.0
Mdn	45.5	63.0	31
Range	8-216	8-216	12–111

Clients

Clients were recruited in these therapists' private practice (with the exception mentioned above) and assessed as typical of other clients in their practice. Exclusion criteria were: psychosis, a main drug diagnosis, known neuropsychological damage, or age below 18 years.

The 48 clients had an average age of 38 years, ranging from 24 to 71 years; 62% were females and 38% males. Educational levels varied greatly, with a modal level between 3 and 5 years of college education. Ten percent had not finished high school. One in six was either unemployed or on sick leave.

Measures of symptomatic distress and interpersonal functioning indicate that the clients seemed to be within the range of typical private practice/outpatient clients, with a mean score of 75.1 on OQ-45 and a mean score of 1.28 on the IIP-64, measured at the beginning of therapy.

The original intention-to-treat sample consisted of 55 clients. Early in the project, three clients discontinued participation in the study but continued in therapy with their psychotherapists, so these were not regarded as therapy drop-outs. Generally, the reasons why these clients did not continue in the research study were sensitivity issues and burdensome logistics. One client did not meet the inclusion criteria; another was excluded due to clearly invalid outcome data, and a third was excluded for having considerable language problems. One therapist with one client withdrew early from the project without providing a reason for doing so. A completer sample of 48 clients was thus included in the study. Thirty-three of these (68.75%) started out with symptoms in the clinical range (i.e., with a total OQ-45 score ≥ 63).

Instruments

The *Outcome Questionnaire 45.2* (OQ-45) of Lambert et al. (1996, 2004) was used to assess the level and change in clients` clinical distress. The OQ-45 is a self-report measure composed of 45 items with a Likert-type response format ranging from 0 (not at all) to 4 (almost always).The instrument includes three subscales: symptomatic distress (SD), interpersonal relations (IR), and social role functioning (SR). Summed scores are computed, with higher scores representing increasing levels of psychopathology.

The OQ-45 was administered in a paper and pencil format. A total score of 63 or more indicates symptoms of clinical significance while reliable change is defined as a change of 14 points or more on the total score (Lambert et al. 1996, 2004). Extensive analyses of the psychometric properties of the scale have been conducted in the USA and elsewhere (e.g., Lambert et al. 1996; Bludworth et al. 2010; Kim et al. 2010) indicating good test–retest reliability, high internal consistency and concurrent validity. Analyses of the OQ-45 have also been done in Norway suggesting that the translated scale is adequate for clinical and research use (Amble et al. 2014). Cronbach's alpha for a clinical Norwegian sample was .93 for the total score, .90 for *symptom distress*, .78 for *interpersonal relations*, and .77 for *social role functioning* (Amble et al. 2014, p. 509). However, confirmatory factor analyses showed that the correlations among the latent factor subscales (SD, IR and SR) ranged between .57 and .75, similar to that found in other countries, and higher than desirable for subscale analyses. Consequently, only the Total (sum) score was used in the present study.

A Norwegian translation of the *Inventory of Interpersonal Problems* or IIP-64 (Horowitz et al. 2000) was used to assess the level and change in clients' interpersonal problems. The IIP-64 consists of 64 items rated on a 5-point scale from 0 (*not at all)* to 4 (*extremely)*. The IIP-64 consists of 39 items following the phrase "It is hard for me to…" and 25 items that describe "Things that you do too much". A total interpersonal distress score (IIP global) was calculated from the mean of the IIP-64 at measurement points. This global score is considered an appropriate index of an individual's interpersonal functioning (Gurtman and Balakrishnan 1998). According to Horowitz et al. (2000), the test–retest reliability, internal consistency and construct validity are excellent. In a study of outpatient clients, the psychometric properties of the Norwegian translation of the instrument have been replicated (Monsen et al. 2006).

Procedures

Outcome Assessments

Data on the OQ-45 and IIP-64 were scheduled to be collected at the 1st session (T1); at the 3rd, 6th, 12th, and 20th session (if therapy lasted that long); and thereafter every 20 sessions, up to and before the last therapy session (T2). Follow-up data were collected at approximately 1.5 years (T3) and 3.5 years (T4) post-termination. However, in this naturalistic study slight modifications of data logistics inevitably occurred. Not all therapists felt comfortable collecting OQ-45 and IIP-64 data in the first session, and if so, data were collected during one of the two following sessions instead. As change on outcome measures may have taken place already after the first session, effect sizes may thus be underestimated in those cases.

Interviews

After the completion of therapy, all clients and therapists were interviewed about their experiences in therapy. The post-therapy interview (soon after the end of treatment, T2) served several purposes for the project at large. For this article, it was used to answer research question 5 on drop-out, based on clients' and therapists' reflections on the nature of treatment termination and the extent to which the ending of treatment was agreed upon. Drop-out was defined as occurring when a client terminated therapy without discussing whether to do so with the therapist. This definition effectively matches Hatchett and Park's (2003, p. 228) definition as "termination by failure to attend the last scheduled appointment," which was one of the two definitions recommended by Wierzbicki and Pekarik (1993) for use in research. The incidence of drop-outs was found analyzing transcripts of post-therapy client-interviews with all clients about how their therapy ended. The whole completer sample in this study (48 clients) was interviewed after the end of their treatment.

Three-and-a-half years after completion of therapy, the clients were interviewed again. At the time these analyses were conducted, only 28 clients were interviewed at follow-up. Data from seven clients had not yet been analyzed, and for the remaining 13 clients, data were not available for various reasons (such as the research teams' inability to locate clients and practical limitations such as audiotape malfunction). Client narratives in the interviews were used to answer research question 5 about whether clients had re-entered therapy in the follow-up period.

Statistical Analyses

In order to study clinical outcomes of this sample we analyzed the data the following way: First, means and standard deviations were computed for OQ-45 and IIP-64 at four time points: during the first three sessions (T1); at the last session (T2), at approximately 1.5 years post therapy (T3), and at 3.5 years post therapy (T4). Magnitude of change in mean scores was computed using the Cohens's *d* effect

size measure (Cohen 1977) for the following time intervals: T1–T2, T2–T4, and T1–T4. The pooled SD was used as variation estimates for all comparisons. Supplementing the conventional definitions of effect size threshold as small $d >= .20$, medium $d >= .50$ and large $d >= .80$ (Rosenthal and Rosnow 1984), we added very large $d >= 1.30$ as introduced by Rosenthal (1996).

Descriptive analyses of the outcome data of the 48 outpatients also included calculations of the proportion of patients who obtained *reliable improvement, reliable deterioration* and *no change* (see Jacobson and Truax 1991) using the pre-determined cut-off limits for the total scale of the OQ-45 (Lambert et al. 1996) with a positive change ≥ 14 points indicating 'reliable improvement', and negative change ≥ 14 indicating 'reliable deterioration', while scores in between indicate 'no change'.

Inferential Statistics

In order to further study change in this client group, growth curve analyses were performed, using a multilevel modeling (MLM) approach (i.e. Snijders and Bosker 2012). This was appropriate because of the structure of the data, with repeated outcome measurements *nested* within patients. Unlike commonly used statistical procedures (such as repeated measures ANOVA), multilevel models account for data dependency due to covariance between levels (in this case, level 1 and 2) in the data, thus reducing the risk of committing Type I errors (the error of rejecting the null hypothesis when it is true) (Raudenbush and Bryk 2002). Another advantage is that multilevel models are robust in permitting nested models with unequal sample sizes on different levels, and in the way missing data are handled (Hox 2010; Tasca and Gallop 2009). In the following MLM analyses, the intercept and slope were treated as fixed effects. The random models included only the intercept due to the limited N. The models tested the starting point (intercept) and change (slope) in the sample as a whole over the measurement waves (T1–T4), in order to examine whether the group improved significantly in the outcome variables over time. Guided by a deviance statistic (AIC), we examined whether change was linear or non-linear (i.e. curvilinear). A linear slope demonstrated a slightly better fit to the data. The Time variable was coded 0–3 corresponding to session T1 (initial assessment); T2 (treatment termination); T3 (1.5 years follow-up) and T4 (3.5 years follow up). The longitudinal analyses were performed on two outcome variables (i.e. OQ total and IIP total). The analyses were done for the full sample and the clinically distressed subsample (i.e. with OQ-45 total of 63 or higher at pre-treatment). The analyses corrected for number of sessions which varied greatly in this naturalistic study. The models were estimated using restricted maximum likelihood estimation (REML) procedures. The Linear Mixed Models, IBM SPSS (version 25.0) program was used to conduct the multilevel growth curve analyses.

Results

Research Question 1: Overall Change

As shown in Table 2, the initial Total OQ-45 score for all clients at T1 was 75.6, which is similar to norms provided by Lambert et al. (2004) for university outpatient clients (normative M = 78.0, *SD* = 25.71) and also for clients treated for an emotional problem within the Employee Assistance Program or EAP (normative M = 73.02, *SD* = 73.02). For clients within the EAP, the Total OQ-45 normative was lower for ethnic Caucasians (M = 63.9, *SD* = 22.7) who would most resemble Norwegian clients. Initial mean Total OQ-45 scores in the present study were lower than reported in a study by Amble et al. (2014) in a Norwegian clinic where clients were treated in an in-patient as well as an out-patient setting (M = 92.3, *SD* = 20.2).

Table 2 Outcome for all clients

					Cohen's *d*			
	T1	T2	T3	T4	T1–T2	T2–T3	T2–T4	T1–T4
OQ-45 total								
N	48	46	26	28				
M	75.06	55.23	54.48	53.81	**.82**	.03	.06	**.98**
SD	22.54	25.56	22.46	20.74				
IIP-64 total								
N	48	46	26	29				
M	1.28	1.01	.91	.92	***.50***	.19	.09	***.74***
SD	.53	.54	.49	.44				

Large effect size (Cohen's *d*) in bold; medium effect size in bold/italics. T1 = session 1; T2 = last session; T3 = 1.5 years post termination; T4 = 3.5 years post termination

For the whole sample, the mean Total OQ-45 score dropped to 55.2 by the end of therapy (T2), which was maintained both 1.5 years post termination (T3) and at 3.5 years follow-up (T4). The T1–T2 difference in mean OQ-45 scores produced a large effect (Cohen's $d = .82$), and the gain was maintained at 3.5 year follow-up. The mean OQ-45 score at 3.5 years post therapy (T4) was 53.8, and thus the T1–T4 difference in OQ-45 mean scores (from beginning of therapy to 3.5 years post-therapy) showed a large effect (Cohen's $d = .98$).

Similarly, the mean IIP-64 score at T1 for all clients was 1.28, dropping to 1.01 at T2, and then stabilizing at T3 and T4 with IIP-64 scores of .91 and .92 respectively. The T1–T2 difference in mean IIP-64 scores yields a medium effect size of .50, a gain that was maintained at 3.5 years follow-up. The difference in mean IIP-64 scores between T1 and T4 yielded a medium effect size of.74.

In sum, analysis of improvements in OQ-45 scores across time (T1–T2; T1–T4) for all clients produced a Large effect, and improvement in IIP-64 scores for all clients showed a Medium effect size. According to the definitions of reliable improvement, no change and negative change from start to termination (T1–T2) as provided by Lambert et al. (2004), 30 clients (66.6%) of the 45 clients for which we have data showed reliable change/improvement, 14 clients (31.1%) demonstrated no change, and one client (2.2%) deteriorated.

Multilevel Growth Curve Analyses—Overall Sample

The growth curve analyses assessing change in OQ-45 and IIP-64 scores, correcting for treatment length, yielded highly significant change (i.e. $p < .001$) over the four measurement waves in both outcome measures: OQ total (slope = – 5.9, $p < .001$); and IIP total score (slope = – .11, $p < .001$). See Table 3 below.

Research Question 2: Clients with Initial Clinical Disturbance

For clients who started treatment above the cut-off for clinical disturbance (N = 33), the OQ-45 mean score at T1was 87.21, dropping to 65.91 at T2, constituting thus a Very large effect size (Cohen's $d = 1.34$) (See Table 4). The gain was maintained at T4, with a mean OQ-score of 62.28, just below the cut-off for clinical disturbance. The T1–T4 change represented a Very large effect size (Cohen's $d = 1.44$).

The initial IIP-64 Total mean score for clients above cut-off for clinical disturbance was 1.50, which dropped to 1.19 at T2, producing a Medium effect size (Cohen's $d = .65$). These clients continued to improve in the follow-up period with an IIP-64 mean score of 1.04 at 3.5 years post therapy. This yielded a Large (T1–T4) effect size (Cohen's $d = 1.01$).

Table 3 Results of multilevel growth curve analyses: Fixed effects models (unstandardized coefficients and standard errors)

Models	OQ_total Full sample/clin. distressed	IIP_total Full sample/clin. distressed
	Estim. (S.E)/	Estim. (S.E)/
Fixed effects		
Intercept	61**(3.29)/ 77.65**(5.05)	1.10**(.07)/ 1.41**(.13)
Slope	– 5.88**(1.12) – 8.36**(1.35)	– .11***(.002)/ .16**(.03)
N of sessions	.14*(.06)/ .05 (.05)	.002 (.001)/ .0003 (.001)

N = 48/N = 33

OQ total total score of Outcome Questionnaire 45.2 (Lambert et al. 1996), *IIP total* total score of the IIP-64 (Horowitz et al. 2000), *Estim.* coefficient estimate, *S.E.* standard error, *estimation method* restricted maximum likelihood (REML)

**p ≤ 001, *p ≤ .01

Analyses of reliable improvement on the OQ-45, no change and negative change for clients with clinical disturbance as defined above showed the following results from T1 to T2 for the 30 clients for whom we have data: 22 clients (73.3%) changed reliably, 7 clients (23.3%) did not change reliably, while no clients deteriorated. From T1 to T4, we have available data from 18 clients above level of clinical disturbance, of these 13 clients (72%) changed reliably, while 5 (28%) demonstrated no change. No clients deteriorated from the beginning of therapy to 3.5 years post therapy.

Multilevel Growth Curve Analyses—Clinically Distressed Sample

Examining patients who started out above the level of clinical disturbance (i.e. OQ total ≥ 63), and correcting for number of therapy sessions, the MLM analyses showed that a reduction in distress as measured by both outcome measures was highly significant (i.e. $p < .001$): OQ total (slope = – 8.36. $p < .001$); and IIP total score (slope = – .16, $p < .001$). The growth coefficients (the reduction in distress as one moves from T1–T2 and so on) in both measures were larger for clinically distressed clients compared to the full sample. See Table 3 for details.

Research Question 4: Incidence of Drop-Outs

The dropout rate for this sample was found by analyzing transcripts of post-therapy client-interviews about how their therapy ended. We had interview data to answer this question for all clients (at the end of treatment). Based on the definition of drop-out used in this study, there were no drop-outs. In all therapies one of the participants initiated some

Table 4 Outcome for clients above cut off for clinical disturbance (i.e. OQ-45 total ≥ 63)

					Cohen's *d*			
	T1	T2	T3	T4	T1–T2	T2–T3	T2–T4	T1–T4
OQ-45 total								
N	33	31	16	17				
M	87.21	65.91	60.84	62.28	**1.34**	*.23*	.17	**1.44**
SD	14.36	22.34	21.29	19.88				
IIP-64 total								
N	33	32	16	17				
M	1.50	1.19	.99	1.04	***.65***	*.38*	*.30*	**1.01**
SD	.49	.42	.51	.37				

T1 = session 1; T2 = last session; T3 = 1.5 years post termination; T4 = 3.5 years post termination. Large effect size (Cohen's *d*) in bold; medium effect size in bold/italics; small effect size in italics

discussion about whether they were moving towards the end. The therapist tended to define this as an issue to explore, and they worked together to reach a concerted decision. In the few cases where either the therapist had preferred that the client had continued therapy in order to deal with some "unfinished business", or the client would have preferred to go further, they seemed to recognize both the professional capacities of the therapists and the capacities of the clients to decide for themselves. Our analyses thus yielded that all therapies ended as a result of a mutual agreement.

Research Question 5: Further Therapy in Follow-Up Period?

Analyses of 28 client interviews at the 3.5 (T4) years follow-up indicated that none of the clients had reentered therapy with their therapist in the study. Note that there was no instruction to not restart therapy or seek other therapists during the extended follow up period. Of the 28 clients, three (10.7%) reported in the interviews that they had been in contact with other therapists: One client had reentered therapy with another psychotherapist, a second had seen a psychiatrist for medication, and a third had recently had two sessions with a new therapist, possibly starting a new therapy. It should be noted that as the response rate is only 58 per cent as we only had 3.5 year follow-up interviews with 28 clients, the results should be interpreted with caution.

Discussion

The aim of the present study was to investigate various aspects of clinical outcomes and possibly to nuance our understanding of the therapist experience-client outcome relationship and clinical effectiveness. Using a sample of highly experienced therapists, the present study analyzed several important aspects of client outcomes; documenting overall good outcome—particularly with clinically distressed clients, high maintenance of treatment gains, no drop out, and a general tendency not to seek new treatment in the lengthy follow-up period.

What stand out as the most important findings from this study are the good outcomes measured at the end of therapy for clients with high levels of clinical distress, and, more importantly, the good results at follow-up, indicating sustained client change. It is noteworthy that the change in interpersonal functioning (as measured by the IIP-64) seemed to continue after treatment termination. In symptom distress, as measured by the OQ-45, the effect sizes of changes from the beginning of treatment to treatment termination, as well as to follow-up, were very large.

There were no drop-outs in this study, that is, all 48 therapies ended according to a mutual agreement, which is rare as far as we know from the literature referred to above (e.g. Roos and Werbart 2013). Three clients reported to having had some form of contact with other therapists in the follow-up period. The results on additional contact with therapists are important for at least two reasons. Firstly, they render the possibility that the good long-term effects of this study can be attributed to client's participation in effective therapies during the follow-up period unlikely. Secondly, these results suggest that for the majority of clients, with the exception of the three returning clients, therapies have ended without much "unfinished business".

Comparatively, how good are the results of this study, and how can we understand them? Measured at treatment termination, the Cohen' *d* effect sizes based on responses to the OQ-45, varied between .86 and 1.34, which place these results in the upper part of what can be expected given the commonly reported effect size estimate of psychotherapy at or around a Cohens *d* of .80 (Wampold and Imel 2015). For the IIP-64, the effect sizes are as expected lower than for symptom distress at treatment termination, but in fact it increased to a large effect at 3.5 years follow-up.

Since so few studies have been conducted with follow-up results beyond two years, we have insufficient comparison

of typical effect sizes for such an extended follow-up period. Even if we looked at studies with shorter follow-up terms, comparison would still be difficult to make for various reasons as treatment context/setting, therapists backgrounds and qualifications, clients characteristics and outcome measures (with differential sensitivities to change) may have differed. Also, as studies typically do not report if clients have had additional therapy during the follow-up period, one does not know if follow-up results are due to the therapies studied. In the present study that is not a problem, as described above.

Broadening the Scopes of Clinical Outcomes and Clinical Expertise

Compared to most psychotherapy studies, a unique feature of the therapists in this study is that they are highly experienced as therapists in addition to being experienced psychotherapy teachers. This combination of roles likely provides a context where a reciprocal influence of the teacher and therapist roles is possible. Typical senior roles like teaching, supervision and consultation have been found to influence therapists' career development across many parameters such as nationality, profession (psychology, medicine) as well as therapists' theoretical orientation (Orlinsky et al. 2001). In their roles as psychotherapy teachers, in combination with their practitioner experience, therapists in this study have had rich opportunity to acquire a comprehensive and nuanced body of conceptual knowledge, which has been found to characterize optimally developing therapists (e.g., Rønnestad and Skovholt 2013); master therapists (e.g., Jennings and Skovholt 2016; Jennings et al. 2013), and those with demonstrated expertise (e.g., Chi 2006). That said, we cannot know if the therapists in this study actually had acquired a rich body of conceptual knowledge fostering optimal development. However, as they have been selected as teachers and one may assume that the teacher role supports an articulation of a nuanced and rich body of conceptual knowledge, we may infer that they do possess a high degree of conceptualizing abilities, which might have supported their therapeutic skills.

Research on the professional development of psychotherapists (Orlinsky and Rønnestad 2005) supports the positive valence of theoretical breadth, which has been found to predict the role performance of 'healing involvement' (defined by items reflecting skillfulness, infrequent difficulties, use of constructive coping strategies when difficulties arise, being invested in affirmative relationships with clients, in addition to the experience of a sense of efficacy and flow in therapeutic work; see Orlinsky and Rønnestad 2005). However, it may be that the underlying element explaining the results above is not theoretical breadth per se, as defined by the number of theoretical orientations endorsed. Rather, it may be the *richness* of conceptual ideas that the therapists have available when understanding the clients and the therapy process, that guides their choice of concrete interventions, a perspective consistent with the findings of Ekroll and Rønnestad (2016). Taken together these results substantiate a claim that therapist expertise is something more than merely years of experience or even number of clients seen. It seems to be characterized by a certain way of organizing and using one's knowledge and skills as well as a way of staying attuned with and responsive to clients. It involves the ability to build and sustain a sufficiently strong working alliance with highly distressed clients, as well as being able to foster lasting changes, an idea that is in line with previous writings (e.g. Hill et al. 2017).

We do not have data that document the therapists' experience with deliberate practice, such as systematic monitoring of processes, ongoing consultation and attending to additional clinical courses, which is suggested as key to the development of expertise (e.g., Goldberg et al. 2016c; Rousmaniere 2017; Tracey et al. 2014). However, it seems reasonable to assume that the therapists' varied experiences in terms of clinical practice and teaching have provided opportunities to rehearse and reflect upon their own practices in ways that are associated with favorable therapist development and clinical expertise.

Strength and Limitations of the Study

There are three important features of this study. First, the background of the therapists is unique, with therapists not only being highly experienced psychotherapists, but also experienced in the psychotherapy teacher role. Secondly, follow-up data on standard measurement instruments are collected as long as 3.5 years post therapy. Thirdly, client interview-data were also collected on client experiences during the follow-up period, making it possible to assess topics such as whether clients had seen other therapists in the follow-up period.

Along with this, there are several limitations to be reported. First and foremost, we do not have a design that allows us to make causal claims about the impact of therapists' experience level on client outcomes. In order to make such claims, one should be able to contrast the results of these therapists with those of other therapists with low or lower levels of clinical experience, or study therapists of varying levels of experience within a longitudinal study design. This was not the scope for the present study. Moreover, we do not have comparative results from similarly experienced therapists who are not psychotherapy teachers. Thus, we cannot know if the addition of the therapy teachers experience has boosted the results. Adding to this, is missing data at follow-up, as well as the incompleteness of the data on whether clients had reassumed therapy in the follow-up period (i.e. we only had data on 58% of the sample on this).

Data on client characteristics is limited by relying largely on the OQ-45 and IIP-64 to measure degree of symptomatic distress and degree of impairment. We do know that patients evaluate the outcome of psychotherapy in more complex ways than these measurements may capture. Although it would have been possible in this project to assess outcome by other methods such as qualitative analyses of transcripts of sessions, and of client interviews both immediately after therapy ended and at 3.5 year follow-up, this was not the scope for this article. Since the initial data-collection was done during the first three sessions of therapy, and not pre-therapy, we have not been able to assess change that had already taken place before initial assessment. Effect sizes may thus be slightly underestimated.

These limitations notwithstanding, in order to further study clinical outcomes and the role of increased clinical experience, we propose that one broadens the scope of what constitutes a good outcome and move to study in-depth how high performing therapists conduct their work and create productive, collaborative relationships with their clients, leading to sustainable therapeutic gains.

Funding This study was financed by the Norwegian Research Council and the University of Oslo. We want to thank David E. Orlinsky and John McLeod for their valuable contributions to the project.

References

Amble, I., Gude, T., Stubdal, S., Øktedalen, T., Skjorten, A. M., Andersen, B. J., et al. (2014). Psychometric properties of the Outcome Questionnaire-45.2: The Norwegian version in an international context. *Psychotherapy Research, 24*(4), 504–513. https://doi.org/10.1080/10503307.2013.849016.

Baldwin, S. A., & Imel, Z. (2013). Therapist effects. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 258–297). New York: Wiley.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., & Wong, E. (2004). Therapist variables. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 227–306). New York: Wiley.

Bludworth, J. L., Tracey, T. J. G., & Glidden-Tracey, C. (2010). The bilevel structure of the Outcome Questionnaire-45. *Psychological Assessment, 22*(2), 350–355. https://doi.org/10.1037/a0019187.

Burlingame, G., Paul, S., Fuhriman, A., & Ogles, B. M. (1989). Implementing a time-limited therapy program: Differential effects of training and experience. *Psychotherapy: Theory/Research/Practice/Training, 26*, 303–313.

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, R. R. Hoffman & P. J. Feltovich (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 21–30). New York: Cambridge University Press.

Cohen, J. (1977). *Statistical power analysis for behavioural sciences*. New York: Academic Press.

Dinger, U., Zimmermann, J., Masuhr, O., & Spitzer, C. (2016). Therapist effects on outcome and alliance in inpatient psychotherapy: The contribution of patients' symptom severity. *Psychotherapy*. https://doi.org/10.1037/pst0000059.

Ekroll, V. B., & Rønnestad, M. H. (2016). Processes and changes experienced by clients during and after naturalistic good outcome therapies conducted by experienced psychotherapists. *Psychotherapy Research, 27*, 450–468.

Ekroll, V. B., & Rønnestad, M. H. (2018). Exploring associations between therapy factors and post- therapy development after naturalistic psychotherapies. *Journal of Contemporary Psychotherapy, 48*(1), 115–125. https://doi.org/10.1007/s10879-017-9366-7.

Goldberg, S. B., Babins-Wagner, R., Rousmaniere, T., Berzins, S., Hoyt, W. T., Whipples, J. L., et al. (2016c). Creating a climate for therapist inprovement: A case study of an agency focused on outcome and deliberate practice. *Psychotherapy, 53*, 367–375. https://doi.org/10.1037/pst0000060.

Goldberg, S. B., Hoyt, W. T., Nissen-Lie, H. A., Nielsen, S. L., & Wampold, B. (2016b). Unpacking the therapist effect: Impact of treatment length differs for high- and low-performing therapists. *Psychotherapy Research*. https://doi.org/10.1080/10503307.2016.1216625.

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016a). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology, 63*(1), 1–11. https://doi.org/10.1037/cou0000131.

Gurtman, M. B., & Balakrishnan, J. D. (1998). Circular measurement redux: The analysis and interpretation of Interpersonal Circle Profiles. *Clinical psychology: Science and Practice, 5*(3), 344–360. https://doi.org/10.1111/j.1468-2850.1998.tb00154.x.

Hatchett, G. T., & Park, H. L. (2003). Comparisons of four operational definitions of premature termination. *Psychotherapy: Theory, Research, Practice, Training, 40*, 226–231. https://doi.org/10.1037/0033-3204.40.3.226.

Hill, C. E., Spiegel, S. B., Hoffman, M. A., Kivlighan, D. M., Jr., & Gelso, C. J. (2017). Therapist expertise in psychotherapy revisited. *The Counseling Psychologist, 45*, 7–53. https://doi.org/10.1177/0011000016641192.

Horowitz, L. M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *IIP-64/IIP-32 professional manual*. San Antonio: The Psychological Corporation.

Horvath, A. O., & Bedi, R. P. (2002). The alliance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work* (pp. 37–69). New York: Oxford University Press.

Hox, J. (2010). *Multilevel analysis. Techniques and applications* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.

Hubble, M. A., Duncan, B. L., Miller, S. D., & Wampold, B. E. (2009). Introduction. In B. L. Duncan, S. D. Miller, B. E. Wampold & M. A. Hubble (Eds.), *The heart and soul of change: Delivering what works in therapy* (pp. 23–46). Washington, DC: American Psychological Association.

Jacobson, N. S., & 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. https://doi.org/10.1037/0022-006X.59.1.12.

Jennings, L., & Skovholt, T. M. (2016). *Expertise in counseling and psychotherapy: Master therapist studies from around the world*. New York: Oxford University Press.

Jennings, L., Skovholt, T. M., Goh, M., & Lian, F. (2013). Master therapists: Explorations of expertise. In M. H. Rønnestad & T. M. Skovholt (Eds.), *The developing practitioner: Growth and stagnation of therapists and counselors* (pp. 213–264). New York: Routledge.

Keefe, J. R., McCarthy, K. S., Dinger, U., Zilcha-Mano, S., & Barber, J. P. (2014). A meta-analytic review of psychodynamic therapies for anxiety disorders. *Clinical Psychology Review, 34*, 309–323. https://doi.org/10.1016/j.cpr.2014.03.004.

Kim, S., Beretvas, S., & Sherry, A. R. (2010). A validation of the factor structure of OQ-45 scores using factor mixture modeling.

Measurement and Evaluation in Counseling and Development, 42, 275–295. https://doi.org/10.1177/0748175609354616.

Kivlighan, D. M. III, Goldberg, S. B., Abbas, M., Pace, B. T., Yulish, N. E., Thomas, J. G., et al. (2015). The enduring effects of psychodynamic treatments vis-à-vis alternative treatments: A multilevel longitudinal meta-analysis.. *Clinical Psychology Review, 40*, 1–14. https://doi.org/10.1016/j.cpr.2015.05.003.

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology & Psychotherapy, 3*, 249–258. https://doi.org/10.1002/(SICI)1099-0879(199612)3.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., et al. (2004). *Administration and scoring manual for the OQ®-45.2*. OQ-Measures: Brigham Young University.

Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 139–193). New York: Wiley. https://doi.org/10.1002/(SICI)1099-0879(199612)3:4%3C249::AID-CPP106%3E3.0.CO;2-S

Laska, K. M., Smith, T. L., Wislocki, A. P., & 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*, 31–41. https://doi.org/10.1037/a0031294.

Leichsenring, F., & Rabung, S. (2008). Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis. *JAMA, 300*(13), 1551–1565. https://doi.org/10.1001/jama.300.13.1551.

Liberman, B. L. (1978). The role of mastery in psychotherapy: Maintenance of improvement and prescriptive change. In J. D. Frank, R. S. D. Hoehn-Saric, B. L. Imber, A. Liberman & R. Stone (Eds.), *Effective ingredients of successful psychotherapy* (pp. 35–72). New York: Brunner/Mazel.

Luborsky, L., McLellan, A. T., Diguer, L., Woody, G., & Seligman, D. A. (1997). The psychotherapist matters: Comparison of outcomes across twenty-two therapists and seven patient samples. *Clinical Psychology, 4*(1), 53–65. https://doi.org/10.1111/j.1468-2850.1997.tb00099.x.

Monsen, J., Odland, T., Faugli, A., Daae, E., & Eilertsen, D. E. (1995). Personality disorders: Changes and stability after intensive psychotherapy focusing on affect consciousness. *Psychotherapy Research, 5*(1), 33–48.

Monsen, J. T., Hagtvet, K. A., Havik, O. E., & Eilertsen, D. E. (2006). Circumplex Structure and personality disorder correlates of the Interpersonal Problems Model (IIP-C): Construct validity and clinical implications. *Psychological Assessment, 18*(2), 165–173.

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. https://doi.org/10.1002/cpp.383.

Orlinsky, D. E., Botermans, J.-F., & Rønnestad, M. H. (2001). Towards an empirically, grounded model of psychotherapy training: Four thousand therapists rate influences on their development. *Australian Psychologist, 36*, 139–148.

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

Orlinsky, D. E., Rønnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. Lambert (Ed.), *Bergin and Garfield's hand- book of psychotherapy and behavior change* (5th ed., pp. 307–389). New York: Wiley.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks: Sage Publications.

Rønnestad, M. H. (2009). *An intensive process-outcome study of the interpersonal aspects of psychotherapy*. Research protocol, Department of Psychology, University of Oslo, Norway.

Rønnestad, M. H. (2016). Is expertise in psychotherapy a useful construct? *Psychotherapy Bulletin, 51*(1), 11–13.

Rønnestad, M. H., & Skovholt, T. M. (2013). *The developing practitioner: Growth and stagnation of therapists and counselors*. New York: Routledge.

Roos, J., & Werbart, A. (2013). Therapist and relationship factors influencing dropout from individual psychotherapy: A literature review. *Psychotherapy Research, 23*(4), 394–418. https://doi.org/10.1080/10503307.2013.775528.

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21*, 37–59.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Rousmaniere, T. (2017). *Deliberate practice for psychotherapists: A Guide to Improving Clinical Effectiveness*. New York: Routledge.

Sandell, R., Blomberg, J., Lazar, A., Carlsson, J., Broberg, J., & Schubert, J. (2000). Varieties of long-term outcome among patients in psychoanalysis and long-term psycho- therapy: A review of findings in the Stockholm Outcome of Psychoanalysis and Psychotherapy Project (STOPPP). *International Journal of Psychoanalysis, 81*, 921–942.

Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology, 80*(4), 535–546. https://doi.org/10.1037/a0028898.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252–266. https://doi.org/10.1016/0749-5978(92)90064-E.

Skovholt, T. M., Rønnestad, M. H., & Jennings, L. (1997). Searching for expertise in counseling, psychotherapy, and professional psychology. *Educational Psychology Review, 9*(4), 361–369. https://doi.org/10.1023/A:1024798723295.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752–760. https://doi.org/10.1037/0003-066X.32.9.752.

Snijders, T. A., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Stein, D. M., & Lambert, M. J. (1984). On the relationship between therapist experience and psychotherapy outcome. *Clinical Psychology Review, 4*, 127–142. https://doi.org/10.1016/0272-7358(84)90025-4.

Tasca, G. A., & Gallop, R. (2009). Multilevel modeling of longitudinal data for psychotherapy researchers: I. The basics. *Psychotherapy Research, 19*(4–5), 429–437.

Tracey, J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: *An elusive goal?. American Psychologist, 69*, 218–229. https://doi.org/10.1037/a0035099.

Tschuschke, V., Crameri, A., Koehler, M., Berglar, J., Muth, K., Staczan, P., et al. (2015). The role of therapists' treatment adherence, professional experience, therapeutic alliance, and clients' severity of psychological problems: Prediction of treatment outcome in eight different psychotherapy approaches. Preliminary results of a naturalistic study. *Psychotherapy Research, 25*, 420–434. https://doi.org/10.1080/10503307.2014.896055.

von der Lippe, A., Oddli, H. W., & Halvorsen, M. S. (2017). Therapist strategies early in therapy associated with good or poor outcomes among clients with low proactive agency. *Psychotherapy Research*. https://doi.org/10.1080/10503307.2017.1373205.

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

Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice, 24*(2), 190–195.