Do Psychotherapists Improve With Time and Experience? A Longitudinal Analysis of Outcomes in a Clinical Setting

Simon B. Goldberg
University of Wisconsin—Madison

Tony Rousmaniere
University of Alaska—Fairbanks

Scott D. Miller
International Center for Clinical Excellence, Chicago, Illinois

Jason Whipple
University of Alaska—Fairbanks

Stevan Lars Nielsen
Brigham Young University

William T. Hoyt
University of Wisconsin—Madison

Bruce E. Wampold
University of Wisconsin—Madison and Modum Bad Psychiatric Center, Vikersund, Norway

Objective: Psychotherapy researchers have long questioned whether increased therapist experience is linked to improved outcomes. Despite numerous cross-sectional studies examining this question, no large-scale longitudinal study has assessed within-therapist changes in outcomes over time. ***Method:*** The present study examined changes in psychotherapists' outcomes over time using a large, longitudinal, naturalistic psychotherapy data set. The sample included 6,591 patients seen in individual psychotherapy by 170 therapists who had on average 4.73 years of data in the data set (range = 0.44 to 17.93 years). Patient-level outcomes were examined using the Outcome Questionnaire-45 and a standardized metric of change (prepost d). Two-level multilevel models (patients nested within therapist) were used to examine the relationship between therapist experience and patient prepost d and early termination. Experience was examined both as chronological time and cumulative patients seen. ***Results:*** Therapists achieved outcomes comparable with benchmarks from clinical trials. However, a very small but statistically significant change in outcome was detected indicating that on the whole, therapists' patient prepost d tended to diminish as experience (time or cases) increases. This small reduction remained when controlling for several patient-level, caseload-level, and therapist-level characteristics, as well as when excluding several types of outliers. Further, therapists were shown to vary significantly across time, with some therapists showing improvement despite the overall tendency for outcomes to decline. In contrast, therapists showed lower rates of early termination as experience increased. ***Conclusions:*** Implications of these findings for the development of expertise in psychotherapy are explored.

Keywords: expertise, therapist effects, therapist experience, psychotherapy training, clinical feedback

Supplemental materials: http://dx.doi.org/10.1037/cou0000131.supp

   The question of whether therapists' accrued experience in conducting psychotherapy over their professional careers leads to improved patient outcomes has been a topic of interest since the origins of psychotherapy research (Bergin, 1971; Beutler et al., 2004; Meltzoff & Kornreich, 1970; Myers & Auld, 1955). The importance of the question related to experience has varied over the years. As early as 1970, Meltzoff and Kornreich noted, "the fact that experience has been a subject of research seems to be a

   Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin—Madison; Tony Rousmaniere, Student Health and Counseling, University of Alaska—Fairbanks; Scott D. Miller, International Center for Clinical Excellence, Chicago, Illinois; Jason Whipple, Department of Psychology, University of Alaska—Fairbanks; Stevan Lars Nielsen, Department of Psychology, Brigham Young University; William T. Hoyt, Department of Counseling Psychology, University of Wisconsin—Madison; Bruce E. Wampold, Department of Counseling Psychology, University of Wisconsin—Madison, and Modum Bad Psychiatric Center, Vikersund, Norway.

   Correspondence concerning this article should be addressed to Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin—Madison, 335 Education Building, 1000 Bascom Mall, Madison, WI 53706. E-mail: sbgoldberg@wisc.edu

reflection of deep-seated doubts about psychotherapy" (p. 268), but their concern was that experienced therapists (those with training) were not more effective than novice therapists (e.g., Strupp & Hadley, 1979). Some of these "doubts" have arguably been dispelled with the advent of meta-analysis and the robust evidence for the effectiveness of psychotherapy (Smith & Glass, 1977; Wampold & Imel, 2015). The question is now not whether psychotherapy, as practiced by trained therapists, is effective, but whether therapist experience with patients over the course of time builds therapeutic competence and leads to better outcomes. Said simply, do therapists, as they practice their craft and see additional patients, improve their outcomes over time?

Studies examining therapist experience predicting outcomes have been discussed in several reviews (cf., Bickman, 1999; Stein & Lambert, 1995; Tracey, Wampold, Lichtenberg, & Goodyear, 2014). Stein and Lambert (1995) concluded in their meta-analytic review that more therapist experience is modestly linked with both lower rates of dropout and better outcomes in psychotherapy. Stein and Lambert (1995) discussed both between-study comparisons (using meta-analysis and study-level estimates of therapist training level; see, e.g., Smith & Glass, 1977) and within-study comparisons. Within-study comparisons involved comparing patient outcomes for groups of therapists with varying levels of training (e.g., trainees vs. professional staff; Myers & Auld, 1955). Importantly, the within-study comparisons included in Stein and Lambert's meta-analysis were cross-sectional, examining differences between groups of therapists at one point in time. More recent cross-sectional studies have generally failed to detect superior outcomes for more experienced clinicians relative to trainees or less experienced clinicians (Budge et al., 2013; Minami et al., 2009; Okiishi et al., 2006; Okiishi, Lambert, Nielsen, & Ogles, 2003; Wampold & Brown, 2005). Similarly, Brown, Lambert, Jones, and Minami (2005) found that therapists' prior experience did not explain differences between highly effective therapists and less effective therapists in managed care environments.

Cross-sectional studies provide only a limited test of whether accrued experience relates to improved outcomes for several reasons, the most important of which is that therapists measured at one point in time differ on many characteristics other than experience, creating multiple confounds. Such designs necessarily use some proxies of experience, such as years since degree, which ignores predegree experience and the amount of experience accumulated in each year (e.g., some therapists see more clients per year than others). A direct assessment of the effects of experience on outcomes requires outcomes of therapists observed over the course of their career, that is, a longitudinal design in which outcomes for a given therapist are measured over time as experience accumulates. To our knowledge, no studies to date have employed longitudinal methods in a large sample of patients and therapists. The closest approximation in the literature was provided by Leon, Martinovich, Lutz, and Lyons (2005), who used a matching procedure to examine differences in outcomes for demographically and clinically matched pairs of patients seen by a given therapist at two different times. Over the 83 pairs of patients examined, the patients seen second did not overall demonstrate a faster rate of improvement, except in instances where the gap in time between the two patients was short in duration (i.e., between 15 and 75 days). A second more recent study examined the changes in a variety of patient- and therapist-level outcomes over

the course of training (n = 23 therapy trainees), noting improvements in some domains (e.g., patient- and therapist-rated alliance, therapist ability to use helping skills) although not in terms of outcomes (viz., patient-rated symptoms; Hill et al., 2015). There appear to be no longitudinal studies of professional therapists outcomes over extended periods of time to appropriately determine whether or not therapists' outcomes improve over time.

The present study examined whether therapists have better outcomes as they gain more experience. As both patient outcomes and therapy dropout have been shown to relate to therapist experience in previous meta-analyses (Stein & Lambert, 1995), both were considered in the current study. In this study, the therapists had ongoing, real time access to measures of patient progress, providing one of the conditions generally thought to be necessary for improvement (i.e., feedback about performance; Lambert, Hansen, & Finch, 2001; Tracey et al., 2014). The basic conjecture tested was that the outcomes of therapists would improve over time or with a greater number of cases treated and that they would have fewer dropouts.

Method

Participants and Procedures

Data were obtained from the treatment research archive at the counseling center of a large, U.S. university. Data were collected over the course of 18.43 years on therapists in practice during that period, although no therapists had data spanning this entire period. Psychotherapy at the counseling center was provided without session limits or extra fees beyond academic tuition. Patients completed the Outcome Questionnaire-45 (OQ-45; Lambert et al., 2004) prior to each session. Analyses of the available data were limited in several ways in keeping with studies examining naturalistic psychotherapy data (e.g., Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009). First, only outcome data from individual counseling sessions (excluding group and couples therapy) were included. Second, to avoid cross-classification of patients and therapists, data were limited to individuals who met with only one therapist at a time. Third, only the first episode of care with the therapist was included, considering an episode of care as ending if a period of 120 days had elapsed between sessions. Fourth, only patients who attended at least three sessions and completed OQ-45 measures for at least two sessions (two sessions were necessary for computing a prepost effect size described below) were included in the analyses. Fifth, the sample was limited to patients whose first OQ-45 total score was in the clinical range (i.e., 63 or above; Lambert et al., 2004). Last, given the focus on therapist-level variables, only patients whose therapist had 10 or more cases were included in the data set. Setting a minimum number of patients per therapist was intended to allow more reliable estimates of therapist-level outcomes (Baldwin & Imel, 2013).

Patients. Based on these requirements, sufficient data were available for 6,591 treated patients: 62.6% were women; average age at intake was 22.60 years, SD = 4.06. Reported ethnicities were 81.9% Caucasian; 6.0% Hispanic; 3.4% Asian; 1.4% Indigenous American; 1.3% Pacific Islander; 0.8% Black; 0.5% other; and 4.6% gave no report. Patients had agreed to use of these deidentified records in research, and the university's human subject review board approved use of these deidentified records.

The data set included OQ measurements from a total of 53,351 sessions. On average, patients attended on average 8.09 sessions (SD = 8.17, range = 3 to 153, Mdn = 6). The average time of treatment was 12.99 weeks (SD = 15.38, range = 0.29 to 237.14, Mdn = 8.15).

Therapists. Psychotherapy was provided by 170 therapists, 71 (41.8%) women and 99 (58.2%) men. Of the 170 psychotherapists, 36 (20.8%) worked first as therapists in training, then as licensed professionals, beginning as graduate students, predoctoral interns, or postdoctoral residents, then as licensed professionals. Of the psychotherapy sessions provided, 30.5% were provided by trainees, 38.7% were provided by licensed professionals, and 30.8% were provided by the therapists who straddled these two statuses. On average, therapists saw 38.77 patients in the data set (SD = 51.36, range = 10 to 360, Mdn = 19). The primary means to assign patients to therapist was based on available slots in the therapist schedules, although occasionally clients requested a therapist who was either a male or female and such requests were honored. Assignment was not based on patient severity, chronicity, or prognosis. Although assignment to therapist was not completely random, it could be described as quasi-random.

Over the course of their careers at the counseling center, therapists had ongoing, real-time access to measures of client progress in the form of OQ scores, which they could use as they saw fit. Licensed therapists were required by state law to complete at least 24 hr of approved continuing education every two years in order to maintain their professional licenses. The center also had in-service training, including approximately 1 hr of discussion and training each month that focused on accommodating client diversity. Therapists in training received from 1 to 3 hr of supervision per week, with the requirement that supervisors viewed video recording of selected sessions each week. State law required that for supervision to count toward licensure, supervisors must have at least two years of postlicensure professional experience. The majority of therapists described themselves as following an integrative or eclectic approach to treatment, adopting techniques, interventions, and styles as seemed to fit the therapeutic situation. Exceptions were one therapist who described himself as a dedicated practitioner of rational emotive behavior therapy, another who described herself as a psychodynamically oriented therapist, and two others who identified themselves as acceptance and commitment therapy therapists.

Variables

Outcome questionnaire. Progress in treatment was measured with the OQ-45 (Lambert et al., 2004). Patients completed the measure at intake and prior to each visit. Respondents rate the frequency at which each event or situation occurred on a five-point Likert-type scale ranging from *never* to *almost always*. The OQ-45 items were developed to assess three domains: symptom distress (25 items, e.g., "I feel no interest in things"), interpersonal relationships (11 items, e.g., "I have frequent arguments"), and social role functioning (nine items, e.g., "I feel stressed at work/school"). A total score is commonly computed and is supported by factor analytic work on the OQ (Bludworth, Tracey, & Glidden-Tracey, 2010). The measure has been widely used and shown to possess desirable psychometric properties, including high internal consistency reliability (α = .94 for the total scale in the current sample)

and adequate test–retest reliability over a 3-week range (from .78 to .84; Snell, Mallinckrodt, Hill, & Lambert, 2001).

Early termination. Available data did not provide a direct measure of treatment dropout (i.e., whether terminations were planned). Thus, very early termination (i.e., treatment durations of one or two sessions) was used as a best approximation. Patients received a code of 0 (early termination) or 1 (nonearly termination) on this dichotomous variable.

Therapist experience. Therapist experience was operationalized in two ways: time and number of patients seen.

Time. A metric of chronological time was computed as time from beginning of therapy with a particular patient relative to the time of the beginning of therapy with the therapist's first patient in the database. Each therapist's first session with their first patient in the data set was coded as time zero, and time for subsequent patients reflected the time between the start of therapy with this patient and time zero (start of therapy with first patient) in years. Therapists had, on average, seen patients for 4.73 years at the center (SD = 5.09, range = 0.44 to 17.93, Mdn = 2.56). In most cases, therapists had clinical experience prior to beginning therapy with the first patient in the data set (on average, therapists saw their first patient in this data set 5.15 years after starting graduate school, SD = 7.50, range = 0.05 to 40.10, Mdn = 2.91), but the primary interest in this analysis was the growth in effectiveness during the time covered in the data set.

Cases. It could be argued that the number of patients seen is a better indicator of experience than the passage of time, and consequently, therapist experience was also operationalized as the cumulative number of cases in the data set that a given therapist had seen prior to a given patient (i.e., equivalent to an ordered count of patients seen by a given therapist across time centered at the first patient seen). Thus, the first patient seen by a given therapist in the data set was coded as 0, the second as 1, and so on. The number of cases of the therapists ranged from 10 to 360, with a mean of 38.77 (SD = 51.36, Mdn = 19).

Statistical Analyses

Estimation of treatment effects. To capture the effects of treatment, a prepost Cohen's d effect size was computed for each patient using her or his first OQ observation minus last OQ observation with these difference scores then divided by the pooled standard deviation of pre- and posttreatment OQ scores within the full sample, yielding a metric of change in standardized units. As lower scores on the OQ indicate better functioning, positive prepost d values indicate patient improvement.

Statistical models. Two-level multilevel models were constructed with patient d values or patients' early termination status (0 or 1) nested within therapists across either chronological time or number of cases seen. An unconditional model was fit initially with no additional predictors in order to assess the magnitude of therapist effects in the data and the need for multilevel modeling. Next, random intercept models were fit in order to assess the impact of either time or cumulative cases. These models included a fixed intercept, fixed slope, and a random intercept coefficient that allowed therapists to vary in their overall effectiveness. Finally, random slope models were fit in order to assess the impact of either time or experience, when the relationship between time (or cases) with patient outcome was also allowed to vary across

therapists. Formal model comparison was conducted to assess whether the random slope coefficient improved model fit. An example of the random slope models employed was as follows:

$$Y_{ij} = \beta_{00} + \beta_{10}(Time) + [U_{0_j} + U_{1_j}(Time) + e_{ij}]$$

where Y_{ij} reflects the outcome (e.g., change in OQ in standardized Cohen's d units) of a given patient (i) seen by a given therapist (j). The fixed intercept (β_{00}) reflects the overall mean effect size across the first patients of each therapist. The fixed slope (β_{10}) reflects the overall mean change in outcome for each year of therapist experience across all therapists. This coefficient would be positive for the prepost d outcome if therapists were achieving better outcomes over time (i.e., producing greater prepost reductions in OQ total scores over time), which is the critical test of the conjecture of this study. The parameters inside the brackets were random effects included in the model. Therapist variability around the fixed intercept was modeled with a random intercept coefficient (U_{0j}) indexing *therapist j*'s deviation from the overall mean outcome (β_{00}). Therapist variability around the fixed slope was modeled with a random slope coefficient (U_{1j}). This coefficient represents how each therapist j's trajectory (change in effectiveness over time) deviates from the fixed slope (β_{10}). Last, e_{ij} reflects the error of prediction or residual for patient i seen by therapist j. Equivalent models were fit using cases instead of time as a random intercept and slope parameter. For models predicting early termination, multilevel logistic regression was used.

Addressing potential confounds. Although, as discussed above, assignment of patients to therapists was quasi-random, it is possible that some patient-level variables may have been confounded with years of therapist experience. More experienced therapists may, for example, have a higher likelihood of engaging more severely distressed patients in therapy, which could have required a longer course of treatment. As baseline severity and length of treatment could impact metrics of effectiveness (and thus influence the association between experience and outcome), subsequent models were constructed that included patients' baseline OQ total scores and number of sessions as patient-level predictors. Further, as data were obtained from a center that includes clinicians-in-training, it was possible that models examining the change in outcomes over years of clinician experience could be unduly influenced by differences between trainees and other staff who were at the center for short periods of time (e.g., less than 1 year) and staff clinicians who remained at the clinic for longer periods of time. In order to rule out this possibility, models were also constructed excluding clinicians who had less than 1 year of data available.

A related possibility is that therapists who entered the data set with more experience may have reached the upper limit of their effectiveness at the outset of the study—thus, it would be unlikely for these therapists to show improvements in outcomes over time. In order to examine this possibility, we constructed models that included therapists' years since beginning their graduate studies or therapists' age (both measured relative to therapists' first case in the data set) first entered simply as control variables and then entered as interaction terms interacting with either time or cases. The interaction terms, in particular, test whether therapists' change trajectories are influenced by their prior experience.

We considered two additional potential confounds that may theoretically vary across time and could impact patient outcomes

(i.e., time-varying confounds). The first of these is the number of patients seen by a given therapist at a given time (i.e., size of current caseload). It has previously been reported that therapists with larger caseloads have poorer outcomes (Borkovec, Echemendia, Ragusea, & Ruiz, 2001). Importantly, if therapists with differing levels of experience carry differing sized caseloads, a spurious relationship may appear between experience and outcome, driven by caseload size. The second potential confound we examined was early termination (i.e., attending fewer than the three sessions needed to be included in the analysis). Just as with caseload, it is theoretically plausible that rates of early termination change as experience increases (e.g., with higher likelihood of early termination for beginning therapists). This loss of patients to early termination could bias estimates of effectiveness.

The primary complication in addressing the impact of caseload size and early termination is that both are, like experience, time-varying. Thus, to examine these variables aggregate estimates of cases initiating treatment (as number of cases) and cases terminating prior to session three (as proportion of a therapist's total cases) were computed across 3-month periods. Therapists' outcomes were likewise aggregated across 3-month periods. Two-level multilevel models were fit to these data (aggregate outcomes nested within therapists across time, 3-month periods). Separate models were constructed with either caseload or early termination for a given 3-month period entered as level one time-varying covariates.

Last, models were run excluding several kinds of outliers. Specifically, models were run excluding patients whose prepost effect size was more than three *SD*s from the overall mean, therapists whose change in outcome over time (derived from regression models fit within each therapist's caseload described below) was more than three *SD*s from the overall mean, and therapists who had a number of cases three *SD*s from the overall mean.

Statistical software. Data were analyzed in the R programming language (version 3.1.0, R Development Core Team, 2014) using the "nlme" multilevel modeling package (Pinheiro, Bates, DebRoy, Sarkar, and the R Development Core Team, 2013). The default covariance structure in "nlme" was used, which assumes that e_{ij} values are independent.

Results

Descriptive Data

The sample overall showed a significant drop in psychological symptoms rated on the OQ over the course of treatment. The average drop on the OQ was 17.17 points ($SD = 20.23$), with a corresponding prepost d of 0.94 ($SD = 1.10$). The average rate of change, in standardized units was $d = 0.16$ per session ($SD = 0.22$). Approximately half of patients showed a reliable (RCI; Jacobson & Truax, 1991) drop on the OQ ($n = 3,413$, 51.8%). Less than half of the sample fell within the nonclinical range at posttreatment ($n = 2,782$, 42.2%).

Unconditional models were fit initially (see Table 1). The intraclass correlation coefficient (ICC) indicated that approximately 1% of variance in patients' prepost change was explained at the therapist level. It is important to note that this ICC is independent of the basic conjecture of the current study. Indeed, the ICC could theoretically be zero (i.e., all therapists were achieving comparable

Table 1
Therapist Experience Predicts Change in Patient Prepost D

Model type	Predictor	FE [95% CI]	SE	t	p	Int Var	Res Var	ICC	Slp Var
Unconditional						.012	1.21	.010	
Random intercept	Time	−.010 [−.017, −.0042]	.0031	−3.31	<.001	.013	1.20	.010	
Random slope	Time	−.012 [−.021, −.0038]	.0044	−2.84	.004	1.8×10^{-12}	1.20	1.5×10^{-12}	.00028
Random intercept	Cases	−.00077 [−.0013, −.00023]	.00027	−2.83	.005	.014	1.20	.011	
Random slope	Cases	−.0020 [−.003, −.0010]	.00052	−3.83	<.001	.0034	1.20	.0028	3.5×10^{-6}

Note. FE = fixed effect from multilevel models; CI = confidence interval; SE = standard error; t = t statistic; p = p value; Int Var = intercept variance; Res Var = residual variance; ICC = intraclass correlation coefficient; Slp Var = random slope variance. Outcome modeled as patient prepost *d* (computed as pretest minus posttest). $n = 6,591$ patients seen by $n = 170$ therapists.

outcomes with patients), and therapists could still uniformly show an increase or decrease in outcomes across time (i.e., the test of the fixed effect does not rely on variability of therapist outcomes at a point in time). Thus, this modest ICC does not detract from the hypothesis of interest.

Predicting Patient-Level Prepost Outcomes From Therapist Experience

Random intercept models were fit with either time (in years) or case number entered as predictors at level 2 predicting patient-level *d* values at level 1. These models included fixed intercept and random intercept parameters (allowing therapists to vary in their overall effectiveness), but the coefficient for predicting outcome from experience was treated as a fixed effect. We then compared these random intercept models with random slope models that allowed the effect of experience on outcome to vary by therapist. Formal model comparison was conducted between random intercept and random slope models. A significant improvement in fit for the random slope was found for models predicting both prepost *d*, $\chi^2(2) = 14.56$, 23.20 for time and cases, respectively, p values < .001. Given this improved fit, random slope models were interpreted.

A consistent pattern of findings appeared across these models with regard to change in therapists' outcomes across experience (both time and cases). As shown in Table 1, the fixed effect reflecting change in therapists' outcomes was significantly different from zero in all models with the direction of effects indicating that therapists tended on average to obtain slightly poorer outcomes as experience increased (indexed either by time or cumulative cases). The fixed effect in the better fitting random slope model indicated a drop of −0.012 (p = .004) in average outcomes per year (see Figure 1) and −0.0020 (p < .001) per additional patient seen.[1]

Predicting Early Termination From Therapist Experience

Multilevel logistic regression models were used to predict patient's early termination from either time or cumulative cases. An initial random intercept model showed a significant effect of time (estimate = −0.019, z = −2.27, p = .024, odds ratio, OR = 0.98), indicating that patients were less likely to terminate early as therapists accrued years of experience. A random slope model with the same predictors and outcome did not improve model fit, $\chi^2(2) = 1.95$, p = .377, indicating that therapists did not vary in

their change in early termination across time. A second set of models were fit using cumulative cases as the metric of experience. A marginally significant effect of cases was seen in the random intercept model (estimate = −0.0011, z = −1.78, p = .076, OR = .99), and the random slope model did not improve model fit, $\chi^2(2) = 3.29$, p = .193.

Addressing Potential Confounds

In order to assess the potential impact of baseline severity and length of treatment on the observed relationship between experience and outcome, the random slope models described above were reestimated with severity and length of treatment included as patient-level covariates. The significant fixed effect for both time and cases remained significant in all models (p values < .01, Table 2, see also supplemental materials).

In order to assess the potential impact of differences between staff and trainees, models were reestimated excluding therapists who had less than 1 year of data available in the data set. As before, the random slope models were reestimated. The significant fixed effect for both time and cases remained in all models (p values < .05, Table 2, see also supplemental materials). Further, models were rerun controlling for either years since beginning graduate studies or therapist's age at first patient in the data set. Increased experience (time and cases) remained associated with slightly poorer outcomes in all models (p < .05, Table 2, see also supplemental materials), with similarly small effect sizes reflecting change over time. In addition, no significant interactions were detected between therapist's years since beginning graduate studies or therapist's age at first patient in the data set with either time or cases (p > .10), providing evidence that experience prior to the beginning of data collection is not distorting the results.

In order to assess the potential impact of size of caseload and early termination, models were constructed as described above

---

[1] In addition to a prepost *d*, two additional continuous metrics of change and two dichotomous metrics of change were also examined. These included modeling posttest OQ scores controlling for pretest, examining rate of change (by dividing prepost change by patient's length of treatment), modeling whether a patient showed a clinically significant drop on the OQ (based on the reliable change index of Jacobson & Truax, 1991) and whether the patient ended treatment in the nonclinical range (i.e., posttreatment OQ score below 63). Results from these models are reported in the supplemental materials. Of note, evidence of small declines in therapists' outcomes across experience was found for all four additional outcome variables with and without controlling for various potential confounds.
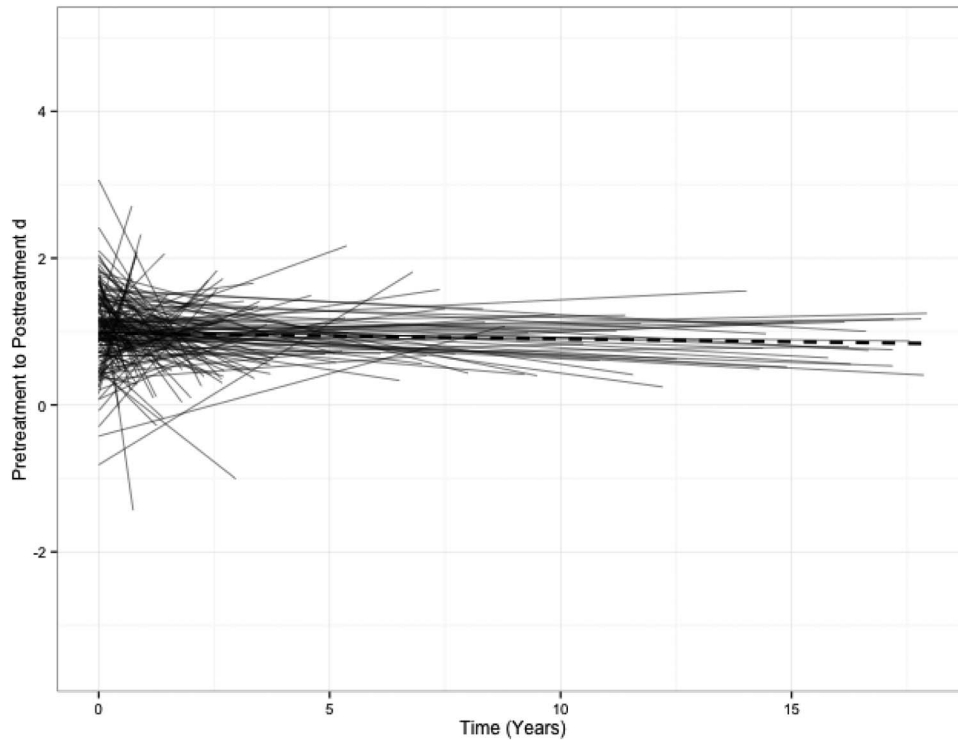
Figure 1. Individual regression lines fit within-therapist reflecting change in patient pretreatment to posttreatment effect size (*d*) as a function of time (in years since beginning of treatment with first patient). Dashed line represents fixed effect of time (overall).

with outcomes, number of cases initiating, and cases terminating prior to the third session aggregated across 3-month periods. Two-level models were then constructed predicting aggregate outcomes from time (3-month period) and either number of cases initiated for a given period or proportion of cases terminating prior to Session 3 (i.e., as time-varying covariates). Therapist experience (indexed as chronological 3-month periods of time) remained a significant predictor of outcomes in all models (*p* values $< .05$, Table 2), indicating declining performance over time that was not explained by variations in either caseload or early termination (see supplemental materials for full multilevel model results).

Last, primary models for each of the five outcomes were run excluding patients who were outliers (three *SD*s above or below the overall mean) in their prepost *d* and therapists who were outliers in their change in prepost *d* over time or in their number of cases in the data set. Time and cases remained significant predictors of patient outcomes with these individuals excluded (*p* $< .05$, Table 2, see also supplemental materials).

Examining Between-Therapist Variation in Changes in Outcome in Response to Experience

The previous series of models provide robust support for the notion that therapists, on average, tend to show declines in their patient-level outcomes as experience was amassed. This result was seen whether experience was operationalized as chronological time or chronological cases and when controlling for several potential confounds.

The significant improvement in fit for random slope models when predicting prepost change implies that therapists vary significantly in their trajectory of change over the course of experience. That is, therapists differed in their rate of change in effectiveness as a function of experience: some had poorer outcomes over time and some had better outcomes over time. In order to estimate the magnitude of this variation, ICCs were computed as the ratio of random slope variance to overall variance (using variance estimates reported in Table 1). The ICC was close to zero (ICCs $= .00023$, 2.9×10^{-6} for time and cases, respectively) indicating that, although statistically significant, therapist variability in rate of change in outcomes across experience explained relatively little variability in patient outcomes.

To express this variation graphically, individual ordinary least squares (OLS) regression analyses were conducted within each therapist's caseload regressing prepost *d* values onto time (in years), providing an estimate of that particular therapist's change in outcome (*d*) by year of experience. Slopes estimated in individual regression models provide an alternative method for indexing an individual therapist's change in outcome across time that is not weighted for the number of patients the therapist saw within the data set (as is the case for the multilevel model; Pinheiro & Bates, 2000). Although the multilevel models employed above that account for size of caseload are most appropriate for testing our primary hypotheses, the individual slopes provide an interpretable means for examining the range of slopes indicating the experience-effectiveness relation across therapists. The mean slope derived

Table 2
Summary of Models Addressing Potential Confounds

Confound addressed	Model type	Predictor	Fixed effect	p
Outcome variable	Posttest controlling for pretest	Time	.0046	<.001
Outcome variable	Posttest controlling for pretest	Cases	.00035	.001
Outcome variable	Rate of change	Time	−.0015	.017
Outcome variable	Rate of change	Cases	−.00012	.024
Outcome variable	RCI drop	Time	−.021	<.001
Outcome variable	RCI drop	Cases	−.0017	<.001
Outcome variable	Clinical cutoff	Time	−.026	<.001
Outcome variable	Clinical cutoff	Cases	−.0019	<.001
Baseline severity, length of treatment	Change in OQ, random slope	Time	−.013	.002
Baseline severity, length of treatment	Change in OQ, random slope	Cases	−.0021	<.001
Excluding therapists with <1 year of data	Change in OQ, random slope	Time	−.012	.006
Excluding therapists with <1 year of data	Change in OQ, random slope	Cases	−.002	<.001
Therapist age	Change in OQ, random slope	Time	−.015	.001
Therapist age	Change in OQ, random slope	Cases	−.0023	<.001
Therapist years since beginning graduate school	Change in OQ, random slope	Time	−.015	.001
Therapist years since beginning graduate school	Change in OQ, random slope	Cases	−.0022	<.001
Cases initiating	Change in OQ, random slope	Time	−.0036	.003
Early termination	Change in OQ, random slope	Time	−.0037	.003
Patient-level prepost *d* outliers	Change in OQ, random slope	Time	−.013	.003
Patient-level prepost *d* outliers	Change in OQ, random slope	Cases	−.0021	<.001
Therapist change in prepost *d* across experience outliers	Change in OQ, random slope	Time	−.012	.004
Therapist change in prepost *d* across experience outliers	Change in OQ, random slope	Cases	−.002	<.001
Therapists with outlying number of cases	Change in OQ, random slope	Time	−.015	.002
Therapists with outlying number of cases	Change in OQ, random slope	Cases	−.0027	<.001

Note. Models run as random intercept unless specified otherwise. See supplemental materials for full model results. RCI = reliable change index; OQ = Outcome Questionnaire.

from the regression models (−0.15, SD = 0.99, range = −5.57 to 4.60, Mdn = −0.05) corroborated the significant fixed effect noted above, suggesting that overall therapists tend to achieve poorer outcomes over time. As seen in Figure 2, the distribution was quite peaked with most observations near the mean (25th percentile = −0.47, 75th percentile = 0.10), but also shows that some therapists did improve over time. Indeed 67 of the 170 therapists (39.41%) had individual regression slopes across their caseloads that were numerically larger than zero, reflecting improvements across time.

Discussion

The present study, which to our knowledge is the first large-scale longitudinal examination of therapist professional development over time using patient outcomes, involved 170 therapists treating over 6,500 patients over an extended period of time (on average, almost 5 years). Therapy was generally effective in this real-world setting, with pretreatment to posttreatment standardized effects of approximately one standard deviation, which is comparable to effects achieved in clinical trials (cf., Minami, Wampold, Serlin, Kircher, & Brown, 2007). At the same time, the present analyses show that, in the aggregate, therapists did not improve with more experience, operationalized as either time or number of cases. Indeed, results suggest that therapists on the whole became slightly less effective over time, although the magnitude of the deterioration was extremely small. The magnitude of the pretreatment to posttreatment effect size diminished 0.012 each year, which is a very small effect: each year, it would be expected that only 1 fewer out of 148 patients would have had a successful outcome, on average (Kraemer & Kupfer, 2006). Of note, this

slight deterioration is in line with previous cross-sectional accounts comparing outcomes for staff clinicians relative to trainees (e.g., Budge et al., 2013). The small decline over time should be considered in the context of the random effect indicating that there was significant variation in the therapists' trajectories over time. A sizable proportion of therapists (39.41%) improved over time (evidenced by slopes that were numerically larger than zero) and others deteriorated (60.59%), although therapist variability in this regard was also quite small. Finally, the very slight deterioration in effects that was detected remained after controlling for initial patient severity, length of treatment, therapist age and years since beginning graduate training, therapists' rates of early termination and caseload size, as well as when excluding patient- and therapist-level outliers and therapists who had less than 1 year of data available (i.e., trainees). Importantly, while some therapists improved, the present findings provide no evidence that therapists *on average* become more effective over time.

Curiously, the results of the present study contrast with clinician self-reported experience. In a large, 20-year, multinational study of over 4,000 therapists, Orlinsky and Rønnestad (2005) found that the majority of practitioners experience themselves as developing professionally over the course of their careers. In particular, therapists with 15 or more years in practice were significantly "more likely than their juniors to experience work with patients as an effective practice, were less likely to have a disengaged practice, and only rarely found themselves in a distressing practice" (p. 88). As well, it appears that therapists tend to overestimate their effectiveness (Walfish, McAlister, O'Donnell, & Lambert, 2012) and fail to recognize failing cases (Hannan et al., 2005; Hatfield, McCullough, Frantz, & Krieger, 2010). It is important to note,
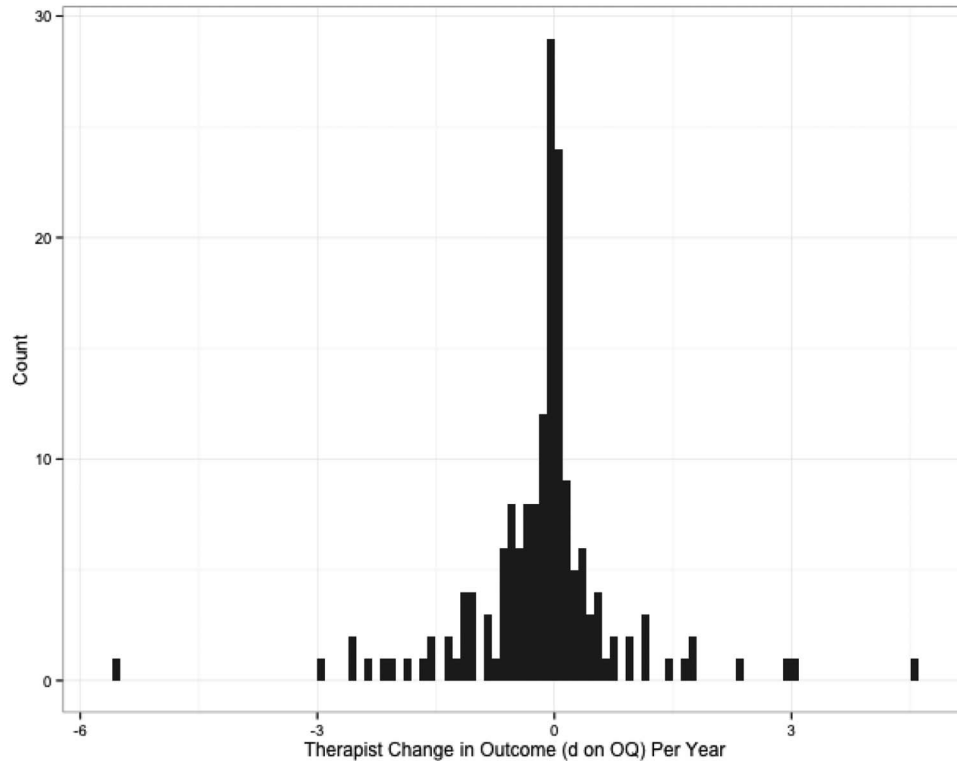
Figure 2. Distribution of therapists as a function of change in effects (prepost *d*) over time (years). OQ = Outcome Questionnaire.

however, that the therapists in the current sample may differ in meaningful ways from those in Orlinsky and Rønnestad's (2005) sample, which included highly experienced therapists working in more diverse settings (e.g., independent practice). It may well be that the therapists in Orlinsky and Rønnestad's sample may have had a very different experience than the therapists in the present sample, and may have used that experience to improve over time.

A contrasting finding was noted when examining rates of early termination. When experience was operationalized as time (although not for number of cases), a significant decrease in early termination was detected as experience accrued. This effect too was quite small (*OR* = 0.98), albeit statistically significant. It appears that although therapists do not appear to get better outcomes overall as experience accrues, they are better able to maintain patients in therapy beyond the second session. This finding is in keeping with Stein and Lambert's (1995) meta-analysis drawn from cross-sectional studies comparing trainees with experienced clinicians that showed decreased dropout for the more experienced therapists.

One reason why we may have failed to detect improvements in outcomes in our sample overall (despite indication that some therapists did improve across time) could be due to assessing only the *quantity* of experience, with no measure of the *quality* of experience. This confound was raised in some of the earlier discussion of this area, with Meltzoff and Kornreich (1970) noting that it "may be the *type* of experience that is important, not the amount" (p. 268). Research in learning science provides some indication of what might constitute high quality experience, which

would then lead to increases in effectiveness. Ericsson (2009) asserts that the key aspect of feedback is pushing performers to "seek out challenges that go beyond their current level of reliable achievement—ideally in a safe and optimal learning context that allows immediate feedback and gradual refinement by repetition" (p. 425). To be effective, efforts must be "focused, programmatic, carried out over extended periods of time, guided by analyses of level of expertise reached, identification of errors, and procedures directed at eliminating errors" (Horn & Masunaga, 2006, p. 601).

The conditions necessary for improvement are typically not present for therapists in practice settings such as the one in the present study (Tracey et al., 2014), but some therapists may engage in such practice (Chow et al., 2015). While there is no clear consensus on precisely how therapists can improve their outcomes, the training literature in psychotherapy and medicine offers a few potential directions. In the psychotherapy profession, individual practitioners may set small process and outcome goals based on patient-specific outcome information (e.g., at-risk cases), create social experiments in naturalistic settings to test, recalibrate, and improve empathic accuracy (Sripada et al., 2011), enhance environments for targeted learning of fundamental therapeutic skills, such as rehearsing difficult conversations (Bjork & Bjork, 2011; Storm, Bjork, & Bjork, 2008), use standardized patients' simulated case vignettes to improve interaction with patients (Issenberg et al., 2002; Issenberg, McGaghie, Petrusa, Lee Gordon, & Scalese, 2005; Issenberg et al., 1999; Ravitz et al., 2013), and set aside time to reflect and plan ahead individually and in clinical consultation

or supervision (Lemov, Woolway, & Yezzi, 2012; Miller & Hubble, 2011).

Future work would clearly need to evaluate these and other efforts to improve outcomes, ideally in adequately large samples of both patients and therapists. Further, it may be fruitful to examine what personal, professional, or caseload differences differentiate the therapists who do show improvements over time from those who fail to improve (our results indicated that therapists indeed vary in their trajectories over time, with some showing improvements despite an average tendency to show slight decreases across time). One therapist variable to examine in a more fine grained way than was possible in the current study is therapists' level of prior experience—it may be that the trajectory of change in outcomes across time varies depending on experience level (although the nonsignificant interactions we report between time or cases and proxies for prior experience would suggest this is not the case). Likewise, it may be valuable to more fully understand why some clinicians show decreased outcomes over time. Professional burnout, a long noted liability in the helping professions (Raquepaw & Miller, 1989; Skovholt & Trotter-Mathison, 2011) may be worth examining.

There are a number of limitations to the present study. First, the sample of therapists was heterogeneous, including practicum students, interns, postdoctoral therapists, and licensed therapists. The more novice therapists received supervision, had reduced caseloads, and may have received other support. As time progressed, these therapists may have received fewer training experiences and increased caseloads that included more difficult patients, resulting in poorer outcomes even if their skill level was improving. However, controlling for initial severity (i.e., patient difficulties), removing novice therapists (viz., those with less than 1 year of data, primarily practicum and predoctoral interns), and examining the interaction between proxies for therapists' prior experience and time or cases did not change the results. A second limitation is that even though this is the longest longitudinal study of experience, the range of experience (viz., from 0.44 years to 17.93 years, with a mean of 4.73 years) was restricted. Skovholt, Rønnestad, and Jennings (1997) asserted that it takes 15 years on average to develop an internalized style, which according to some is an aspect of expertise. Third, outcome was the only indicator of skill development, and one could claim that particular skill domains should be the focus instead (see Shanteau & Weiss, 2014). However, the attempt to establish that rated competence, for instance, is related to outcome has been difficult (e.g., Branson, Shafran, & Myles, 2015; Webb, Derubeis, & Barber, 2010), and therefore we chose to focus on outcomes. Relatedly, no single standardized treatment was provided to patients, and thus it is not clear how therapist skill (as it relates to the delivery of a specific intervention) could be operationalized. Fourth, the amount of effort that therapists used to improve outcomes, including training, supervision, and continuing education, was largely unknown. Indeed, it may be that the *quality* of experience (that is, experience marked by training more likely to impact outcomes, perhaps through the inclusion of deliberate practice of specific therapy skills) proves to be a better predictor of outcomes than the mere *quantity* of experience measured in the present study. Fifth, while patient diagnosis was largely unknown, the setting from which these data were drawn (i.e., university counseling center) rarely includes patients with more severe mental illnesses (these illnesses can interfere with gaining admission to

or maintaining enrollment at the university), although patients with considerable distress are nonetheless increasingly found in counseling center samples (Benton, Robertson, Tseng, Newton, & Benton, 2003; Erdur-Baker, Aberson, Barrow, & Draper, 2006). As part of expertise in psychotherapy is dealing with a range of patient severity, it is not possible to evaluate the development of this kind of expertise in the current sample nor could the development of this kind of expertise be reflected in the results. Relatedly, patients' average age (i.e., 22.60 years) and the relatively brief courses of therapy on average, while typical of counseling center populations, may not generalize to other settings (e.g., community clinics). A replication of the current findings in a noncounseling center sample (perhaps even a sample using standardized treatments targeted to a specific disorder) would be worthwhile. Sixth, although early termination was used as a proxy for dropout (and, indeed, therapists were shown to improve as years of experience accumulated), it is likely that considerable dropout was not captured in this way. A future study would do well to examine whether rates of mutual termination increase as therapists become more experienced, regardless of when in the course of therapy termination occurs. Last, therapist effects in these data were small overall (explaining only approximately 1% of variance in patient outcomes) and considerably smaller than average (see Baldwin & Imel, 2013), suggesting that other factors (e.g., patient variables, relationship factors; Bohart & Wade, 2013; Norcross, 2011; Orlinsky, Rønnestad, & Willutzki, 2004) are likely stronger contributors to outcome than therapist experience. The small observed effects should thus be understood in the broader context of known therapy ingredients.

References

Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology, 77,* 203–211. http://dx.doi.org/10.1037/a0015235

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

Benton, S. A., Robertson, J. M., Tseng, W. C., Newton, F. B., & Benton, S. L. (2003). Changes in counseling center client problems across 13 years. *Professional Psychology: Research and Practice, 34,* 66–72. http://dx.doi.org/10.1037/0735-7028.34.1.66

Bergin, A. E. (1971). The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 217–270). New York, NY: 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* (5th ed., pp. 227–306). New York, NY: Wiley.

Bickman, L. (1999). Practice makes perfect and other myths about mental health services. *American Psychologist, 54,* 965–978. http://dx.doi.org/10.1037/h0088206

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY: Worth Publishers.

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

Bohart, A. C., & Wade, A. G. (2013). The client in psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 219–257). Hoboken, NJ: Wiley.

Borkovec, T. D., Echemendia, R. J., Ragusea, S. A., & Ruiz, M. (2001). The Pennsylvania Practice Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clinical Psychology: Science and Practice, 8,* 155–167. http://dx.doi.org/10.1093/clipsy.8.2.155

Branson, A., Shafran, R., & Myles, P. (2015). Investigating the relationship between competence and patient outcome with CBT. *Behaviour Research and Therapy, 68,* 19–26. http://dx.doi.org/10.1016/j.brat.2015.03.002

Brown, G. S., Lambert, M. J., Jones, E. R., & Minami, T. (2005). Identifying highly effective psychotherapists in a managed care environment. *The American Journal of Managed Care, 11,* 513–520.

Budge, S. L., Owen, J. J., Kopta, S. M., Minami, T., Hanson, M. R., & Hirsch, G. (2013). Differences among trainees in client outcomes associated with the phase model of change. *Psychotherapy, 50,* 150–157. http://dx.doi.org/10.1037/a0029565

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,* 337–345. http://dx.doi.org/10.1037/pst0000015

Erdur-Baker, O., Aberson, C. L., Barrow, J. C., & Draper, M. R. (2006). Nature and severity of college students' psychological concerns: A comparison of clinical and nonclinical national samples. *Professional Psychology: Research and Practice, 37,* 317–323. http://dx.doi.org/10.1037/0735-7028.37.3.317

Ericsson, K. A. (Ed.). (2009). *Development of professional expertise*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511609817

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61,* 155–163. http://dx.doi.org/10.1002/jclp.20108

Hatfield, D., McCullough, L., Frantz, S. H., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy, 17,* 25–32.

Hill, C. E., Baumann, E., Shafran, N., Gupta, S., Morrison, A., Rojas, A. E., . . . Gelso, C. J. (2015). Is training effective? A study of counseling psychology doctoral trainees in a psychodynamic/interpersonal training clinic. *Journal of Counseling Psychology, 62,* 184–201. http://dx.doi.org/10.1037/cou0000053

Horn, J., & Masunaga, H. (2006). A merging theory of expertise and intelligence (pp. 587–612). In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance*. New York, NY: Cambridge University Press.

Issenberg, S. B., McGaghie, W. C., Gordon, D. L., Symes, S., Petrusa, E. R., Hart, I. R., & Harden, R. M. (2002). Effectiveness of a cardiology review course for internal medicine residents using simulation technology and deliberate practice. *Teaching and Learning in Medicine, 14,* 223–228. http://dx.doi.org/10.1207/S15328015TLM1404_4

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher, 27,* 10–28. http://dx.doi.org/10.1080/01421590500046924

Issenberg, S. B., Petrusa, E. R., McGaghie, W. C., Felner, J. M., Waugh, R. A., Nash, I. S., & Hart, I. R. (1999). Effectiveness of a computer-based system to teach bedside cardiology. *Academic Medicine, 74,* S93–S95. http://dx.doi.org/10.1097/00001888-199910000-00051

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,* 12–19. http://dx.doi.org/10.1037/0022-006X.59.1.12

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry, 59,* 990–996. http://dx.doi.org/10.1016/j.biopsych.2005.09.014

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69,* 159–172. http://dx.doi.org/10.1037/0022-006X.69.2.159

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., . . . Burlingame, G. B. (2004). *Administration and scoring manual for the Outcome Questionnaire-45*. Orem, UT: American Professional Credentialing Services.

Lemov, D., Woolway, E., & Yezzi, K. (2012). *Practice perfect: 42 rules for getting better at getting better*. San Francisco, CA: Jossey-Bass.

Leon, S. C., Martinovich, Z., Lutz, W., & Lyons, J. S. (2005). The effect of therapist experience on psychotherapy outcomes. *Clinical Psychology & Psychotherapy, 12,* 417–426. http://dx.doi.org/10.1002/cpp.473

Meltzoff, J., & Kornreich, M. (1970). *Research in psychotherapy*. Chicago, IL: Adline.

Miller, S. D., & Hubble, M. (2011). The road to mastery. *Psychotherapy Networker, 35,* 22–31.

Minami, T., Davies, D. R., Tierney, S. C., Bettman, J. E., McAward, S. M., Averill, L. A., & Wampold, B. E. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of Counseling Psychology, 56,* 309–320. http://dx.doi.org/10.1037/a0015398

Minami, T., Wampold, B. E., Serlin, R. C., Kircher, J. C., & Brown, G. S. J. (2007). Benchmarks for psychotherapy efficacy in adult major depression. *Journal of Consulting and Clinical Psychology, 75,* 232–243. http://dx.doi.org/10.1037/0022-006X.75.2.232

Myers, J. K., & Auld, F., Jr. (1955). Some variables related to outcome of psychotherapy. *Journal of Clinical Psychology, 11,* 51–54. http://dx.doi.org/10.1002/1097-4679(195501)11:1<51::AID-JCLP2270110112>3.0.CO;2-0

Norcross, J. C. (2011). *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed.). New York, NY: Oxford.

Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology, 62,* 1157–1172. http://dx.doi.org/10.1002/jclp.20272

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

Orlinsky, D. E., & Rønnestad, M. H. (2005). Work practice patterns. In D. E. Orlinsky & M. H. Rønnestad (Eds.), *How psychotherapists develop: A study of therapeutic work and professional growth* (pp. 81–99). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/11157-006

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

Pinheiro, J. S., & Bates, D. M. (2000). *Mixed-effect models in S and S-PLUS*. New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4419-0318-1

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & the R Development Core Team. (2013). *nlme: Linear and nonlinear mixed effects models* [Computer software manual]. Retrieved from http://cran.r-project.org/

Raquepaw, J. M., & Miller, R. S. (1989). Psychotherapist burnout: A componential analysis. *Professional Psychology: Research and Practice, 20,* 32–36. http://dx.doi.org/10.1037/0735-7028.20.1.32

Ravitz, P., Lancee, W. J., Lawson, A., Maunder, R., Hunter, J. J., Leszcz, M., . . . Pain, C. (2013). Improving physician-patient communication through coaching of simulated encounters. *Academic Psychiatry, 37,* 87–93. http://dx.doi.org/10.1176/appi.ap.11070138

R Development Core Team. (2014). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Shanteau, J., & Weiss, D. J. (2014). Individual expertise versus domain expertise. *American Psychologist, 69,* 711–712. http://dx.doi.org/10.1037/a0037874

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

Skovholt, T. M., & Trotter-Mathison, M. (2011). *The resilient practitioner: Burnout prevention and self-care strategies for counselors, therapists, teachers, and health professionals* (2nd ed.). New York, NY: Routledge.

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

Snell, M. N., Mallinckrodt, B., Hill, R. D., & Lambert, M. J. (2001). Predicting counseling center clients' response to counseling: A 1-year follow-up. *Journal of Counseling Psychology, 48,* 463–473.

Sripada, B. N., Henry, D. B., Jobe, T. H., Winer, J. A., Schoeny, M. E., & Gibbons, R. D. (2011). A randomized controlled trial of a feedback method for improving empathic accuracy in psychotherapy. *Psychology and Psychotherapy, 84,* 113–127. http://dx.doi.org/10.1348/147608310X495110

Stein, D. M., & Lambert, M. J. (1995). Graduate training in psychotherapy: Are therapy outcomes enhanced? *Journal of Consulting and Clinical Psychology, 63,* 182–196. http://dx.doi.org/10.1037/0022-006X.63.2.182

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2008). Accelerated relearning after retrieval-induced forgetting: The benefit of being forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 230–236. http://dx.doi.org/10.1037/0278-7393.34.1.230

Strupp, H. H., & Hadley, S. W. (1979). Specific vs nonspecific factors in psychotherapy. A controlled study of outcome. *Archives of General Psychiatry, 36,* 1125–1136. http://dx.doi.org/10.1001/archpsyc.1979.01780100095009

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

Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110,* 639–644. http://dx.doi.org/10.2466/02.07.17.PR0.110.2.639-644

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

Wampold, B., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). 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,* 200–211. http://dx.doi.org/10.1037/a0018912