

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/281226287>

# The Role of Deliberate Practice in the Development of Highly Effective Psychotherapists

Article in *Psychotherapy Theory Research Practice Training* · September 2015

DOI: 10.1037/pst0000015 · Source: PubMed

CITATIONS

46

READS

2,124

6 authors, including:



**Daryl Chow**

International Center for Clinical Excellence; ...

13 PUBLICATIONS 162 CITATIONS

SEE PROFILE



**Scott D. Miller**

International Center for Clinical Excellence

98 PUBLICATIONS 3,789 CITATIONS

SEE PROFILE



**Jason Seidel**

Colorado Center for Clinical Excellence

12 PUBLICATIONS 147 CITATIONS

SEE PROFILE



**Robert Kane**

Curtin University

116 PUBLICATIONS 1,519 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



The Study of Supershrinks Project [View project](#)



Self-guided online treatment for young people with Obsessive Compulsive Disorder [View project](#)

All content following this page was uploaded by [Scott D. Miller](#) on 27 August 2015.

The user has requested enhancement of the downloaded file.

# The Role of Deliberate Practice in the Development of Highly Effective Psychotherapists

Daryl L. Chow

International Center for Clinical Excellence, Chicago, Illinois,  
and Curtin University

Scott D. Miller

International Center for Clinical Excellence, Chicago, Illinois

Jason A. Seidel

Colorado Center for Clinical Excellence, Denver, Colorado, and  
International Center for Clinical Excellence, Chicago, Illinois

Robert T. Kane and Jennifer A. Thornton  
Curtin University

William P. Andrews

Pragmatic Research Network, Rhosneigr Gwynedd, United Kingdom

Little empirical research exists about highly effective psychotherapists, and none about the factors that mediate the acquisition and maintenance of superior performance skills (e.g., Ericsson, 1996, 2006; Ericsson, Krampe, & Tesch-Romer, 1993). In the full sample, a 3-level multilevel modeling (Level 1: clients; Level 2: therapists; Level 3: organization types) of practitioner outcomes was used to examine the contribution of the therapist to treatment effectiveness. Consistent with prior research, in the full sample ( $n = 69$  therapists;  $n = 4,580$  clients) it was found that therapist effects explained 5.1% of the variance in outcome, after adjusting for initial severity. Therapist gender, caseload, and age were not found to be significant predictors. In a subsample of therapists, the relationship between outcome and therapist demographic variables, professional development activities, and work practices was analyzed ( $n = 17$  therapists,  $n = 1,632$  clients). Therapist characteristics (e.g., years of experience, gender, age, profession, highest qualification, caseload, degree of theoretical integration) did not significantly predict client-reported outcomes. Consistent with the literature on expertise and expert performance, the amount of time spent targeted at improving therapeutic skills was a significant predictor of client outcomes. Further, highly effective therapists indicated requiring more effort in reviewing therapy recordings alone than did the rest of the cohort. Caveats and implications for clinical practice, continuing professional development, and training are discussed.

**Keywords:** deliberate practice, expertise in psychotherapy, professional development, psychotherapy outcomes, therapist effects

**Supplemental materials:** <http://dx.doi.org/10.1037/pst0000015.supp>

With the exception of a few recent proposals (Miller & Hubble, 2011; Miller, Hubble, Chow, & Seidel, 2013; Tracey, Wampold, Lichtenberg, & Goodyear, 2014), the field has yet to systematically examine the development of superior performance in psy-

chotherapy. Studies document significant variability in outcome among therapists, showing that the clinician accounts for a greater percentage of effectiveness than does treatment modality (0%–1%; e.g., Wampold, 2001; Wampold & Brown, 2005), client-rated alliance (5%–15%; Crits-Christoph, Gibbons, Hamilton, Ring-Kurtz, & Gallop, 2011; Horvath, Del Re, Fluckiger, & Symonds, 2011), or use of an empirically supported treatment (0%–4%; Wampold, 2005). Although researchers have provided rich and detailed descriptions of therapist personality characteristics, professional development activities, and work practices believed to be related to outcomes, no studies have empirically examined the links between such traits and activities and therapy outcomes (Levitt & Williams, 2010; Orlinsky & Ronnestad, 2005; Skovholt & Jennings, 2005).

The study of expertise in other fields provides a potential model for understanding the key mediating factors involved in the development of top-level performers in psychotherapy. Across a variety of domains—including music (Ericsson et al., 1993; Krampe & Ericsson, 1996), chess (Gobet & Charness, 2006), sports (Côté, Ericsson, & Law, 2005), business (Sonntag & Kleine, 2000),

---

Daryl L. Chow, International Center for Clinical Excellence, Chicago, Illinois, and Department of Psychology and Speech Pathology, Curtin University; Scott D. Miller, International Center for Clinical Excellence; Jason A. Seidel, Colorado Center for Clinical Excellence, Denver, Colorado, and International Center for Clinical Excellence; Robert T. Kane and Jenny Thornton, School of Psychology and Speech Pathology, Curtin University; William P. Andrews, Pragmatic Research Network, Rhosneigr Gwynedd, United Kingdom.

The article is based on a doctoral dissertation by Daryl L. Chow at Curtin University, funded by the Department of Innovation, Industry, Science and Research (DIISR), Australia. We express our deepest gratitude to K. Anders Ericsson for his help and guidance in conducting this study.

Correspondence concerning this article should be addressed to Daryl L. Chow, Henry Street Centre, 17 Henry Street, Fremantle, Western Australia 6160. E-mail: [daryl@darylchow.com](mailto:daryl@darylchow.com)

and medicine and surgery (Norman, Eva, Brooks, & Hamstra, 2006)—researchers have found that engagement in extended, deliberate practice facilitates incremental development resulting in superior performance.

*Deliberate practice* (DP) is a highly specialized process. According to Ericsson and Lehmann (1996), it is defined as “individualized training activities especially designed . . . to improve specific aspects of an individual’s performance through repetition and successive refinement” (pp. 278–279). Available evidence makes this clear: Experience involving the routine and proficient execution of skills associated with a particular performance domain is not enough to lead to improvement. To be effective, DP must be focused on achieving specific targets just beyond a performer’s current abilities, guided by the conscious monitoring of outcomes, and carried out over extended periods of time. Studies show that engagement in DP not only facilitates the development of superior performance but is also crucial for its maintenance. For example, not only is engaging in DP at the early skills-acquisition phase important, but it continues to be essential for older expert pianists to maintain their level of performance (Krampe & Ericsson, 1996). However, less is known about DP in psychotherapy.

Accordingly, the present study sought to determine whether DP accounted for the development of superior performance by examining the link between the nature and amount of time spent in efforts to improve performance and individual clinician outcomes aggregated over time. First, therapist effectiveness on the basis of actual client outcomes was determined after adjusting for initial severity. Next, the relationship between outcomes and time spent by therapists in DP (e.g., solitary practice aimed at improving therapeutic engagement) was determined. Finally, the specific nature of DP activities was explored, as was their relationship to differences in outcomes among more effective therapists.

## Method

### Participants

**Therapists.** Participants were psychotherapists, psychologists, social workers, marriage and family therapists, and counselors practicing independently within the Human Givens Institute Practice Research Network (HGIPRN)<sup>1</sup> in the United Kingdom. To be eligible for inclusion, practitioners had to have a caseload of 10 or more clients. Similar to other naturalistic outcome studies (e.g., Clark et al., 2009), each client attended a minimum of two sessions.

On the basis of inclusion criteria, the final sample consisted of 4,580 clients working with 69 therapists from 45 organizations. To ensure independence at the organizational level (several practiced at more than one organization), organizations were partitioned into six conceptually distinct types. Nine out of the 69 therapists were working in more than one type of organization. For the purposes of analysis, each of these nine therapists was coded with a primary organization type on the basis of where he or she had treated the largest number of clients. The mean caseload of each therapist (treating clients with at least two sessions) was 66.38 clients ( $SD = 70.03$ ,  $Mdn = 40.00$ , minimum = 10, maximum = 335). There were 28 (40.6%) male therapists and 38 (55.1%) female therapists (three therapists were unspecified). The majority (52.2%) were between the ages of 41 and 55 years. In terms of organization, the

majority worked in the voluntary sector (42%), followed by independent practice (39.1%), primary care (National Health Service [NHS]; 8.7%), secondary care (NHS; 4.3%), insurance based (2.9%), and occupational health and counseling settings (2.9%).

In the subsample, 17 of the 69 original therapists voluntarily completed an online questionnaire about their professional development and work practices (Chow, 2014). Mean therapist caseload in this group was 94.24 ( $SD = 97.40$ ,  $Mdn = 46.00$ ; minimum = 10, maximum = 335), and they had an average of 8.45 ( $SD = 5.24$ ) years of experience. Briefly, 52.9% of the therapists were male, the majority of them (64.7%) were between the ages of 41 and 55 years, 58.8% were working in a private setting, and 52.9% were practicing as professional psychotherapists.

**Clients.** Only data from clients who were at least 18 years of age were included. In the full sample, the mean age for the 4,580 clients was 40.04 years ( $Mdn = 40.00$ ,  $SD = 12.86$ ); 2,999 (65.5%) were female, and 1,580 (34.5%) were male (one client’s gender was left unspecified), and the majority self-identified as White (66.2%). In terms of ethnicity, the majority was White (90.1%), followed by Asian (1.6%), Black (African, Caribbean, other Black, or Black British; 0.6%), other (0.6%), and mixed ethnicity (0.1%); 20.8% did not state an ethnicity. Consistent with most treatment settings, the majority of clients presented with concerns relating to anxiety ( $n = 3,670$ ; 74.90%) and depression ( $n = 2,690$ ; 59.58%). The average number of sessions attended was 4.72 ( $SD = 3.83$ ). A total of 2,503 clients (54.7%) had a planned termination (i.e., mutual agreement between client and therapist) with their therapist, 947 clients (20.7%) indicated an unplanned termination (i.e., client stopping therapy abruptly), and 1,130 clients (24.7%) did not indicate a planned or unplanned termination. A total of 3,632 out of 4,580 clients (79.3%) provided information about the use of medication, and 1,659 (36.2%) of these were prescribed a psychotropic medication.

In the subsample, data from 1,632 clients were included. The mean age for these clients was 40.19 years ( $Mdn = 40.00$ ,  $SD = 13.20$ ), of which 985 (60.4%) were female and 646 (39.6%) male, and the majority (83.5%) were White. Gender was not specified for one of the clients. Anxiety ( $n = 1,357$ ; 57.41%) was the major presenting complaint, followed by depression ( $n = 1,234$ ; 49.69%).

The average number of sessions attended was 3.89 ( $SD = 2.57$ ). A total of 1,087 clients (66.6%) had a planned termination with their therapist, 265 clients (16.2%) indicated an unplanned termination, and 207 clients (12.7%) did not indicate a planned or unplanned termination.

### Measures

**Clinical Outcomes in Routine Evaluation.** Clients completed the 10-item version of the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Connell & Barkham, 2007; Evans et al., 2000). The CORE-10 is a 10-item questionnaire that assesses symptoms (anxiety, depression, physical problems, trauma), functioning (general functioning, close relationships, social relationships), and risk (risk to self, risk to others). Items are scored using a 0–4 Likert-type response format (ranging from *not*

<sup>1</sup> For further details about HGIPRN, see Andrews, Wislocki, Short, Chow, and Minami (2013).

at all to most or all of the time) for the past week. The recommended cutoff between clinical and nonclinical populations is 10, with higher scores indicating more distress (Connell & Barkham, 2007). Internal consistency ( $\alpha = .94$ ) and 1-week test-retest reliability for the measure (Spearman's  $\rho = .90$  [Evans et al., 2002]) are both high. Total scores for the longer version (34 items) and this shorter form (10 items) are highly correlated ( $r = .94$ ) and, when used to assess outcome, have been shown to result in similar overall effect sizes (Andrews et al., 2013). The current study's average effect size ( $d = 1.22$ ) is equivalent to those found in past naturalistic outcome studies based on the same outcome measure, CORE-OM (e.g., Clark et al., 2009;  $d = 1.09$  for two treatment sites combined). Consistent with its length, internal consistency for this sample was somewhat lower than that for the longer version ( $\alpha = .83$ ) but still in the acceptable range ( $\alpha = .82$ ).

**Retrospective Analysis of Psychotherapists' Involvement in Deliberate Practice.** Retrospective Analysis of Psychotherapists' Involvement in Deliberate Practice (RAPIDPractice) is a survey instrument specifically designed to assess the amount of time a clinician spends in activities aimed at improving therapeutic performance. Items were developed from both a systematic review of the literature on DP and its research methodologies (e.g., Côte et al., 2005; Ericsson et al., 1993) and consultation with K. Anders Ericsson, a leading researcher on expert performance (Ericsson, 1996, 2004, 2006; Ericsson et al., 1993). RAPIDpractice contains 32 items, seven of which seek information regarding professional background, including gender identification, age, discipline, training, certification, years of clinical experience, and theoretical orientation. The remaining items capture the amount of time therapists spend in practice outside of work aimed at improving therapeutic skills. In this section, respondents were asked to rate (a) the frequency with which they engaged in 25 activities (the amount of time spent in the last typical work month), (b) the confidence they had in their frequency rating from 0 (*not at all confident in my time estimate*) to 10 (*highly confident in my time estimate*), (c) the relevance of the particular activity to their improving clinical skills from 0 (*not at all relevant*) to 10 (*highly relevant*), and (d) the cognitive effort required for engaging in the activity from 0 (*no effort exerted at all*) to 10 (*highest possible effort exerted*; Chow & Miller, 2012).

## Procedure

Over a 4-year period, data were gathered from practitioners who routinely administered a standardized outcome measure as part of their treatment services (Andrews et al., 2013). Therapists from the full sample were invited to participate in assessing the relationship between practitioner characteristics, work activities, and engagement in professional development and DP. In this instance, data were gathered via an online questionnaire (cf. Chow, 2014).

## Data Analyses

In the full sample, multilevel modeling (MLM) was implemented through SPSS's linear mixed model procedure (SPSS Version 19). Model parameters were estimated with maximum likelihood. All noncategorical explanatory variables were grand-mean centered to facilitate interpretation of the intercept (Raudenbush & Bryk, 2002).

Data from the subsample on DP were analyzed with a series of generalized linear mixed models (GLMMs). The GLMM is an attempt to overcome the limitations of classic least squares regression procedures (McCulloch, Searle, & Neuhaus, 2008). The GLMM can handle several types of nonnormal outcome variables and accommodate a mixture of fixed and random effects. After controlling for intratherapist and intraorganization-type dependencies, relationships between severity-adjusted client outcome and each of the conceptually distinct groups of therapist explanatory variables were examined. The fixed effects were the initial (pretest) CORE score and the therapist variables described later. The final (posttest) CORE score provided the outcome variable. Following the GLMM analyses, descriptive statistics were computed using standard statistical procedures (analysis of variance, one-sample  $t$  test, correlation) to further investigate the relationships between therapist explanatory variables and client outcomes.

In all, there were seven therapist variables: specifically, three therapist characteristics (demographics, caseload, and degree of theoretical integration), time spent on DP, and time spent in three DP subdomain activities (see Table 1). Prior to the primary analyses, the relationship between each of the seven predictors and the adjusted client outcome was examined with a separate GLMM. The aim of these analyses was to identify any predictors not related to the adjusted client outcome, and these were dropped from all subsequent analyses to avoid suppressor effects (MacKinnon, Krull, & Lockwood, 2000). The remaining therapist variables and adjusted client outcomes were analyzed with one or more GLMMs. Bonferroni adjustments were made to test alpha levels within each of the analytical groups in accordance with the number of GLMMs that were tested within a given group.

## Results

### Full Sample

To analyze the between-therapist variability in client outcome for the first study, an unconditional model (i.e., no predictors) was first introduced in the analysis. The results from the three-level multilevel model revealed that the intraclass correlation of therapist effects for the CORE was 5.35% ( $3.193/[3.193 + 56.514] = 0.0535$ ; Wald  $Z = 3.84$ ,  $p < .001$ ). Next, an adjustment of clients' initial severity of functioning (i.e., pretest score added as a covariate) was conducted, which served as a base model for comparison with subsequent models. This resulted in the therapist effects for the CORE accounting for 5.10% of the total variance in client outcomes ( $2.4262/[2.4262 + 45.206] = 0.0510$ ; Wald  $Z = 3.84$ ,  $p < .001$ ). A significant percentage of therapist variability was still unexplained, so the three-level model was retained for all subsequent analyses. There were no significant differences in adjusted client outcomes across the treatment organization types,  $F(3, 12) = 0.49$ ,  $p = .695$ . In addition, no other client or therapist variables reduced the percentage of variance in outcome attributed to therapists (the results are available as an appendix in the online supplemental materials).

### Comparing Therapists' Performance

Therapists were grouped into quartiles on the basis of their outcomes to facilitate a closer examination of performance differ-

Table 1  
*Mean Relevance and Cognitive Effort Ratings for 20 Therapy-Related and Five Nontherapy-Related Activities for Participating Therapists*

Activity	n	Relevance		Cognitive effort	
		M	SD	M	SD
1. General clinical supervision as a supervisee (without review of audiovisual recordings of sessions)	13	6.92	2.63	7.46	2.63
2. Clinical supervision as a supervisee (with review of audiovisual recordings of sessions)	10	4.10	3.93	5.30	4.60
3. Clinical supervision as a supervisee (review of difficult/challenging cases and/or cases with nil improvement)	12	7.67	2.77	8.00 H	2.76
4. Live supervision provided during sessions (e.g., supervisor as cotherapist, one-way mirror/reflecting team)	10	3.40 L	4.01	5.80	5.01
5. Reading of journals pertaining to psychotherapy and counseling	14	7.21	1.67	6.71	1.68
6. Reading/rereading core counseling materials	11	6.55	3.24	6.00	2.79
7. Focused learning in specific model(s) of psychotherapy	11	7.27	2.87	7.45	2.98
8. Reviewing therapy recordings alone	10	4.00	3.71	4.40	3.98
9. Reviewing therapy recordings with peers	10	4.30	3.83	4.50	4.04
10. Reviewing difficult/challenging cases alone	14	8.00 H	2.77	7.43	2.59
11. Attending training workshops for specific models of therapy	12	8.00 H	2.17	8.25 H	1.96
12. Case discussion/conceptualization/formulation with a mentor/clinical supervisor	12	6.25	3.49	6.50	3.26
13. Mentally running through and reflecting on the past sessions in your mind	15	8.20 H	2.65	7.13	2.30
14. Mentally running through and reflecting on what to do in future sessions	15	8.40 H	2.38	7.07	2.05
15. Writing down your reflections of previous sessions	12	7.92	3.00	6.92	3.20
16. Writing down your plans for future sessions	13	7.00	2.89	6.15	2.97
17. Case discussion/conceptualization/formulation with peers	12	6.67	2.64	6.75	2.45
18. Viewing master therapist videos, with the aim of developing specific therapeutic skills as a therapist	11	4.36	3.23	4.45	3.70
19. Reading case examples (e.g., narratives, transcripts, case studies)	12	4.33	3.23	4.92	3.40
20. Discussion of psychotherapy-related subjects with contemporaries/peers/mentors	13	6.85	3.02	5.69	2.66
21. Tending to self-care activities and emotional needs	13	7.31	2.29	4.85	3.58
22. Socializing	15	6.80	2.83	5.13	2.83
23. Exercising	15	6.27	3.63	4.40	3.68
24. Rest (e.g., naps in the day, going for a walk, engaging in a nontherapeutic activity that is enjoyable)	15	7.40	2.59	4.13	3.54
25. Others <sup>a</sup> (Please specify)	7	5.00	4.00	5.57	4.43

Note. Grand means were 6.41 (relevance) and 6.04 (cognitive effort). Some therapists did not provide the rating of particular activities as they did not engage in the activity in question. H = significantly higher than the grand mean at  $p < .05$ ; L = significantly lower than the grand mean at  $p < .05$ .

<sup>a</sup> These included musical activities, meditation, spiritual practices, child rearing, reading nontherapy-related topics, and so on.

ences. Differences between the four groups of therapists on a variety of outcome variables are reported in Table 2. Subsequent analyses used the full dimensional value of these therapist variables—that is, without grouping therapists into quartiles.

### Subsample: DP

Four GLMM analyses were conducted for the effect of DP on outcomes: (a) time spent alone in DP, (b) time in solitary activities, (c) time in nonsolitary activities, and (d) time spent in nontherapeutic activities.<sup>2</sup>

The first variable, average number of hours per week spent alone in DP, was grand-mean centered and entered in the regression model. This was a significant predictor of the adjusted client outcomes ( $b = -0.016$ ,  $SE = 0.007$ ),  $t(1,549) = -2.09$ ,  $p = .037$ ,  $\eta^2 = .003$ . In other words, a reduction in client distress was predicted by therapists who spent more time alone outside of work in DP activities. None of the other three predictors was significantly related to client outcomes ( $ps > .05$ ; see the supplemental materials).

To further examine the impact of this first variable, time spent alone on DP, the therapists from the top quartile were compared with the rest of the cohort. Of the original sample of 69 therapists, the 17 were ranked 1, 3, 5, 7, 8, 13, 14, 15, 22, 27, 29, 30, 31, 36, 43, 44, and 54 in terms of client outcomes. As can be seen in Figure 1, on average, the top quartile group of therapists invested

about 2.81 times more time on DP alone in a typical work week than did the rest of the therapists.

Along with the amount of time spent for each of the 20 domain-specific and five nontherapy-related activities, each respondent was asked to rate the following on a Likert-type scale: (a) the relevance of the item to improving effectiveness and (b) the cognitive effort required to engage in the activity (with 0 representing *not at all relevant/no effort extended* and 10 representing *highly relevant/highest possible effort exerted*). The relevance and cognitive effort ratings were not included in the MLM. Instead, on the basis of recommendations made by Keppel and Wickens (2004), the items was rank ordered on the basis of their means. A series of one-sample  $t$  tests was conducted comparing the grand mean for the relevance rating to each item mean. This strategy was repeated for ratings of cognitive effort. The results are summarized in Table 1.

Of the 25 different activities, the following received significantly higher than average relevance ratings as regards improving

<sup>2</sup> Separate GLMMs were conducted on five clinician demographic variables (years of experience, gender, age [three levels], profession [five levels], and highest qualification [seven levels]) to determine the relationship between the characteristics and outcomes. None proved to be significantly related to client outcomes. Subsequent analysis also showed that caseload and integrative theoretical orientation were likewise unrelated to outcomes (see the appendix in the online supplemental materials).

Table 2  
Therapists Grouped into Quartiles (Qs) of Adjusted Client Outcomes ( $N = 69$ )

Q <sup>a</sup>	Initial score: <i>M (SD)</i>	Adjusted CORE score <sup>b</sup>	Raw ES <sup>c</sup>	RCI mean (%) <sup>d</sup>	CS mean (%) <sup>e</sup>	Deterioration (%)	No change (%)	Mean number of sessions	Mean planned ending (%)	Mean unplanned ending (%)
1	19.57 (7.17)	8.75	1.50	76.6	58.9	1.6	22.0	4.52	73.4	12.2
2	20.10 (7.37)	10.75	1.25	67.6	46.5	2.2	30.2	4.44	60.5	15.7
3	19.02 (7.45)	12.07	0.99	57.8	34.2	3.4	38.9	5.74	51.7	23.4
4	20.05 (7.57)	14.17	0.76	50.3	25.2	4.7	44.5	6.30	27.7	22.5

<sup>a</sup> Q 1 = best performing therapists, Q 4 = poorest performing therapists. Sample sizes were 17, 17, 17, and 18 for Q 1, Q 2, Q 3, and Q 4, respectively. <sup>b</sup> Adjusted Clinical Outcomes in Routine Evaluation (CORE) score = posttreatment CORE scores, adjusted for initial (pretreatment) CORE scores. Lower scores indicate better outcomes. <sup>c</sup> Raw ES = single-group pre-post effect size using the standard deviation of the pretreatment scores. <sup>d</sup> Percentage of clients with a score change  $\geq$ RCI (reliable change index; Jacobson & Truax, 1991). <sup>e</sup> Percentage of clients with a score change  $\geq$ CS (clinical significance; Jacobson & Truax, 1991).

effectiveness: “Reviewing difficult/challenging cases alone,” “Attending training workshops for specific models of therapy,” “Mentally running through and reflecting on the past sessions in your mind,” and “Mentally running through and reflecting on what to do in future sessions.” Live supervision provided during sessions was rated the least relevant (all  $ps < .05$ ).

With regard to cognitive effort, therapists only rated “Clinical supervision as a supervisee (review of difficult/challenging cases and/or cases with nil improvement)” and “Attending training workshops for specific models of therapy” as requiring significantly higher than average effort (all  $ps < .05$ ). None of the relevance ratings were significantly correlated with therapist average outcomes. Among the cognitive effort ratings, only “Reviewing therapy recordings alone” was significantly correlated with therapist average outcomes ( $rs = -.665$ ,  $n = 10$ ,  $p = .036$ ). With the exception of “Live supervision provided during sessions,” “Reading/rereading core counseling materials,” “Writing

down your plans for future sessions,” and “Tending to self-care activities and emotional needs,” the perceived relevance of all items was significantly correlated with the cognitive effort ratings ( $rs$  ranging from .56 to .92).

## Discussion

After adjusting for initial severity and accounting for organizational-level data, the analysis determined that 5.1% of the variance in outcome was attributable to the therapist—a finding consistent with prior research in this area (Baldwin & Imel, 2013). Consistent with prior research, therapist demographic variables (age range, gender, highest qualification, professional discipline, years of experience) failed to predict client outcomes (Beutler et al., 2004). Other factors—including theoretical approach, degree of theoretical integration/eclecticism, and size of caseload—were tested but likewise did not predict client outcomes. With regard to caseload, researchers have noted the clinical relevance of feeling psychologically burdened with too many clients (Norcross & Guy, 2007). It is possible, however, that the caseload measure in this study was confounded, reflecting the total number of cases treated instead of a count of concurrent clients.

DP, or the amount of time therapists spent alone in DP, was significantly related to outcomes. This finding is consistent with results from numerous studies in different professional domains, including sports, chess, business, computer programming, teaching, and medicine and surgery (Duckworth, Kirby, Tsukayama, Berstein, & Ericsson, 2011; Ericsson et al., 1993; Gobet & Charness, 2006; Keith & Ericsson, 2007; Starkes, Deakin, Allard, Hodges, & Hayes, 1996). In contrast, no significant relationships were found between the amount of time spent in any of the specific activities surveyed and outcomes. In other words, no specific activity engaged in by therapists reliably led to better outcomes. Although contrary to expectations, this finding is consistent with that of Ericsson et al. (1993), whose investigation of violinists also yielded no profile differences in terms of the ratings of relevance, enjoyment, and effort of related activities. Given that the overall amount of DP was related to effectiveness, it may be that the utility of engaging in a specific activity depends on the needs, knowledge, skills, and competencies of the specific clinician. To date, no study has examined a taxonomy of DP activities for therapists to improve effectiveness.

To illustrate the impact of DP on the acquisition of expertise, trajectories plotting the amount of time performers of varying

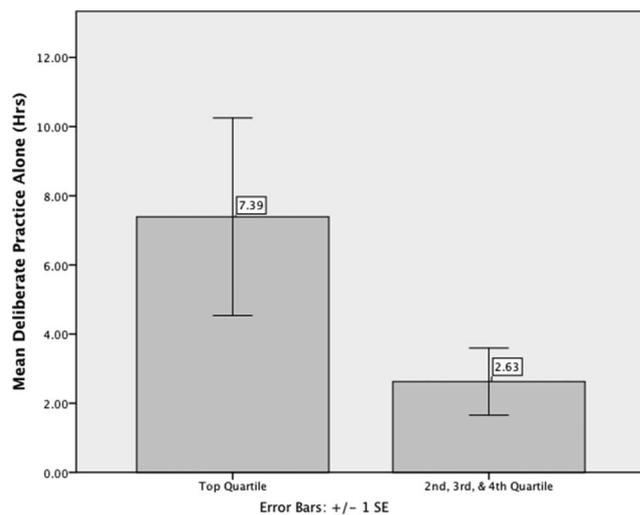


Figure 1. Comparing therapists from the top quartile with the others in the lower quartiles on the basis of their adjusted client outcomes as a function of estimated time spent on deliberate practice alone per typical work week. Groupings of therapists were based on the ranking of the complete cohort from in the first study. Two out of the 17 therapists in Study 2 did not complete this part of the questionnaire. Numbers of therapists in quartile grouping: first quartile = 7; second, third, and fourth quartiles = 8.

ability spend in DP over time were created for the therapists in this study. Using a methodology similar to those used in prior studies (Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005; Ericsson et al., 1993), the amount of time spent in DP was based on self-report in a typical work week. This figure, in turn, was first multiplied by 52 (weeks per year) and then by years of experience. As the average amount of experience for the subsample was approximately 8 years, only these years of professional experience were included (see Figure 2).

As Figure 2 illustrates, the estimated accumulative time spent by the top quartile (most effective therapists) was, on average, about 2.8 times more hours per week engaged in DP activities aimed at improving effectiveness than the rest of the other therapists. Although intriguing, there are some caveats. First and foremost, the sample was small. Second, the numbers of hours spent in DP were both self-reported and based on retrospective recall, not actual measurement. Third, therapists in the study were at different developmental stages in their professional careers. As a result, time estimates may have been related more to skills maintenance than skills acquisition, because the figure was indicative of each therapist's current ratings of the time spent in working to improve clinical skills.

It is possible to speculate on a differential effect of time spent engaging in DP during the early professional developmental phase of skills acquisition compared with the current maintenance of the relevant psychotherapeutic competencies. Given that most clinicians are likely to have spent more time in training during the initial phases of their careers, it would be advisable for future studies to include beginning therapists (e.g., Budge et al., 2013).

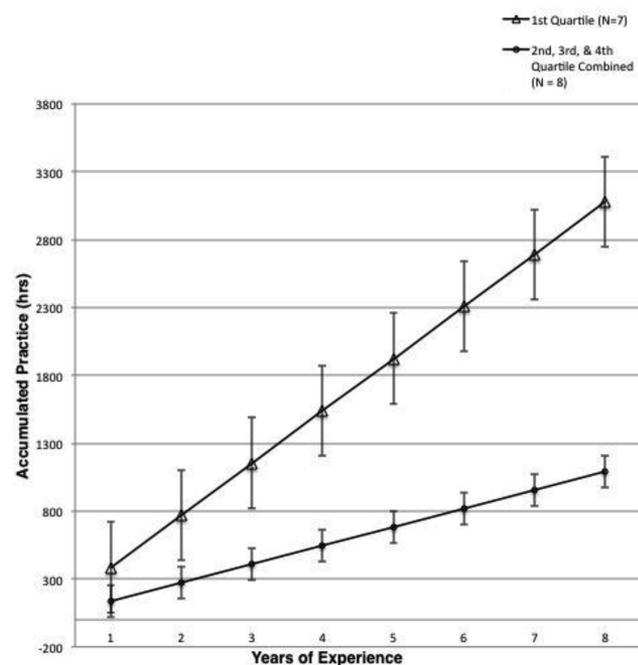


Figure 2. Comparing therapists from the top quartile with the others in the lower quartiles on the basis of their adjusted client outcomes as a function of their accumulative time spent on deliberate practice alone in the first 8 years of clinical practice. Error bars represent standard errors of the mean.

They would be less susceptible to bias recall of time spent in DP alone, and one could also test whether differences in early professional development are mediated by time spent in solitary practice.

In the subsample, four domain-specific activities received higher than average relevance ratings: (a) reviewing difficult/challenging cases alone, (b) attending training workshops for specific models of therapy, (c) mentally running through and reflecting on the past sessions in your mind, and (d) mentally running through and reflecting on what to do in future sessions. Given that these ratings were accumulated from clinicians of varying effectiveness (e.g., the most and least effective), their meaning and relationship to performance is unclear. Clinicians could, for example, rate a specific activity as highly relevant while simultaneously not engaging in it. Moreover, research on learning calls into question the reliability and validity of student and instructor self-ratings of effective methods for enhancing learning (Bjork & Bjork, 2011; Shea & Morgan, 1979).

Studies on DP highlight the role of cognitive effort in the acquisition of domain-specific knowledge and skills (Ericsson et al., 1993; Ericsson & Lehmann, 1996). Of the 25 activities listed, one significant correlation was found between outcomes and the cognitive effort ratings of "Reviewing of therapy recordings alone." In other words, therapists with better outcomes rated the activity of reviewing of therapy recordings alone as requiring more cognitive effort than other activities. Researchers have noted the important role that reviewing performance recordings can play in the identification and remediation of errors (Abbass, 2004; Binder, 1999; Ericsson, 1996, 2006; Norcross & VandenBos, 2011). It is also possible to speculate that with more cognitive demands in a given practice activity, the learner is more likely to benefit in the process (Bjork & Bjork, 2011).

Across the entire cohort, the significant correlations between relevance and cognitive effort ratings for the majority of practice activities was consistent with the findings of previous studies (e.g., Ericsson et al., 1993; Starkes et al., 1996), suggesting that activities deemed highly relevant also tended to be perceived as requiring high cognitive effort. In addition, two items received higher than average cognitive effort ratings: (a) clinical supervision as a supervisee (review of difficult/challenging cases and/or nonprogressing cases) and (b) attending training/workshops for specific models of therapy. Once again, given that these ratings were accumulated from clinicians across clinicians of varying effectiveness (e.g., the most and least effective), their meaning and relationship to performance is unclear. Clinicians could have rated a particular item as cognitively demanding while never actually engaging in the activity. A replication of the present study with a significantly larger sample is ongoing and will hopefully provide the statistical power necessary for a more robust examination of the relationship between outcomes and both the relevance of and cognitive effort required for engaging in specific domain-related activities.

Although the present results are intriguing and point to a heretofore unexplored topic in the area of professional development, this study does suffer from a number of limitations. Given its exploratory nature, many variables were included in the analyses, thereby increasing the risk of Type I error. To address this concern, Bonferroni corrections were used in the second and smaller study in accordance with the number of GLMMs. Still, the risk remains. These initial findings await replication.

As already noted, the portion of the study related to DP had 17 participants. Although previous studies of therapist effects have had a range of sample sizes—from as large as 91 within a university counseling center (Okiishi, Lambert, Nielsen, & Ogles, 2003) to as small as nine therapists in a mental health clinic for male veterans (Luborsky, McLellan, Woody, O'Brien, & Auerbach, 1985)—the number of participants in the DP portion of this study limits generalizability. More important, perhaps, is the nature of the sample. To begin, top performers were overrepresented, thereby increasing the homogeneity of therapist effectiveness and preventing more robust comparisons with poorly performing clinicians. Not surprisingly, perhaps, therapists in the least effective group were less likely to participate in this study. The combination clearly leads to a risk of a self-selection bias limiting the generalizability of the results.

One additional limitation that deserves mention is the use of retrospective methods in assessing the amount of time spent in therapy-related and unrelated activities. Although consistent with other DP research, the validity of such methods is a matter of debate (e.g., Charness et al., 2005; Ericsson et al., 1993; Law, Côté, & Ericsson, 2007). Unlike other performance domains (e.g., sports, music) in which confirmation of DP time is more feasible (e.g., interviews with coaches, teachers, parents), the highly individualized and, to date, private nature of psychotherapy practice makes cross-validation with independent raters impractical (excepting, perhaps, time spent in supervision, mentoring, and staffing). Rating done by others of an adult's solitary behavior presents obvious challenges. Prospective journaling, in which individuals keep logs of activity over a set period of time may allow for more accurate estimates of engagement in DP (Shiffman, Stone, & Hufford, 2008). In real-world clinical settings, such as those examined in this study, any such efforts would need to be carefully balanced with time pressures faced by practitioners. Finally, compared with Clark et al.'s (2009) average number of sessions of 6.35, the limited length of treatment in this study ( $M = 4.72$ ) might constrain its implications when considering samples with longer treatments. Nonetheless, the majority of clients had a planned termination of the treatment process, which might suggest a briefer approach among this cohort.

In accounting for the reasons professionals do not spend more time in DP (aside from the lack of financial compensation for practice time), Ericsson (2009) noted the following:

Most professionals—such as doctors, nurses, stockbrokers, and accountants—do not receive the constant pressure from performing in front of an audience of paying ticket holders, like actors, musicians, and athletes. The lack of scrutiny and perhaps feedback may be an important difference that explains why many doctors do not spontaneously adopt the best practice methods for treating their patients, and spend a rather modest amount of time engaged in deliberate practice and effortful training to improve and maintain their skills. . . . The greatest obstacle for deliberate practice during work is the lack of immediate objective feedback. (p. 422)

Ericsson (2009) pointed to one significant practice-related barrier to engaging in DP: lack of immediate feedback. Research documents that therapists routinely overestimate their effectiveness—on average, by about 65% (Miller, Hubble, & Duncan, 2007; Walfish, McAlister, O'Donnell, & Lambert, 2012). Miller et al. (2007) argued persuasively that such findings may in part

contribute to less time being devoted to improving performance. As both prior research and the present study confirm, experience is not a reliable predictor of outcomes (cf. Beutler et al., 2004). Indeed, despite the early gains of skill-based competencies in their careers, professionals tend to plateau in their development (Ericsson, 2009).

Recently, a number of valid, reliable, and feasible measures for systematically monitoring progress in clinical practice have become available to practitioners, the routine use of which have been shown in multiple randomized clinical trials to reduce dropout and deterioration rates (Hannan et al., 2005; Simon, Lambert, Harris, Busath, & Vazquez, 2012) while simultaneously improving treatment outcomes (Lambert & Shimokawa, 2011; Miller, Duncan, Brown, Sorrell, & Chalk, 2006). In this regard, Ericsson (2009) indicated 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). Possibly, DP in the psychotherapy profession can be specifically targeted the following areas: (a) improving outcomes of at-risk cases; (b) creating social experiments in naturalistic settings to test, recalibrate, and improve empathic accuracy (Sripada et al., 2011); (c) enhancing environments for targeted learning of fundamental therapeutic skills, such as rehearsing difficult conversations (Bjork & Bjork, 2011; Burns, 2009); (d) using standardized patients' simulated case vignettes to improve interaction with clients (Issenberg, McGaghie, Petrusa, Gordon, & Scalese, 2005; Ravitz et al., 2013), and (e) setting aside time to reflect and plan ahead (Lemov, Woolway, & Yezzi, 2012; Miller & Hubble, 2011).

The present study provides preliminary evidence for the role that DP plays in the development of highly effective therapists. Parallel to the development of expertise in sports (Starkes et al., 1996), highly effective therapists spent more time engaged in activities outside of practice specifically aimed at improving performance while practicing. DP might provide the necessary scaffolding for the development of therapeutic skills beyond a given therapist's current ability. Beyond devoting time to the process, it is not known what moderator variables may influence, enhance, or suppress the engagement in DP or the consolidation of the effects of DP on client outcomes. It is entirely possible, for example, that DP may need to occur in combination with other activities (e.g., traditional classroom training, work-life balance, time off for personal activities) to enhance learning, skills acquisition, and maintenance of expert performance.

No longer treated as a nuisance variable (Garfield, 1997), therapist effects have become a serious focus of clinical trials and naturalistic research (Baldwin & Imel, 2013). The findings reported here are consistent with those of prior research. Clearly, therapists vary in their ability to engage and help clients. The results from the present study point toward the important role of time spent in solitary practice in accounting for such differences.

## References

- Abbass, A. (2004). Small-group videotape training for psychotherapy skills development. *Academic Psychiatry*, 28, 151–155. <http://dx.doi.org/10.1176/appi.ap.28.2.151>

- Andrews, W. P., Wislocki, A. P., Short, F., Chow, D., & Minami, T. (2013). A five-year evaluation of the human givens therapy using a practice research network. *Mental Health Review Journal, 18*, 165–176. <http://dx.doi.org/10.1108/MHRJ-04-2013-0011>
- Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 259–297). 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* (5th ed., pp. 227–306). New York: Wiley.
- Binder, J. L. (1999). Issues in teaching and learning time-limited psychodynamic psychotherapy. *Clinical Psychology Review, 19*, 705–719. [http://dx.doi.org/10.1016/S0272-7358\(98\)00078-6](http://dx.doi.org/10.1016/S0272-7358(98)00078-6)
- 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: Worth.
- 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: Theory, Research & Practice, 50*, 150–157. <http://dx.doi.org/10.1037/a0029565>
- Burns, D. D. (2009, December). *Advanced empathy skills*. Lecture presented at the Evolution of Psychotherapy Conference, Anaheim, CA.
- Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19*, 151–165. <http://dx.doi.org/10.1002/acp.1106>
- Chow, D. (2014). *The study of supershrinks: Development and deliberate practices of highly effective psychotherapists*. Unpublished doctoral dissertation, Curtin University.
- Chow, D., & Miller, S. D. (2012). *Retrospective analysis of psychotherapists' involvement in deliberate practice (RAPIDPractice)*. Unpublished manuscript.
- Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy, 47*, 910–920. <http://dx.doi.org/10.1016/j.brat.2009.07.010>
- Connell, J., & Barkham, M. (2007). *CORE-10 user manual, Version 1.1*. Rugby, UK: CORE System Trust & CORE Information Management Systems Ltd.
- Côté, J., Ericsson, K., & Law, M. P. (2005). Tracing the development of athletes using retrospective interview methods: A proposed interview and validation procedure for reported information. *Journal of Applied Sport Psychology, 17*, 1–19. <http://dx.doi.org/10.1080/10413200590907531>
- Crits-Christoph, P., Gibbons, M. B., Hamilton, J., Ring-Kurtz, S., & Gallop, R. (2011). The dependability of alliance assessments: The alliance–outcome correlation is larger than you might think. *Journal of Consulting and Clinical Psychology, 79*, 267–278. <http://dx.doi.org/10.1037/a0023668>
- Duckworth, A. L., Kirby, T. A., Tsukayama, E., Berstein, H., & Ericsson, K. A. (2011). Deliberate practice spells success: Why grittier competitors triumph at the National Spelling Bee. *Social Psychological and Personality Science, 2*, 174–181. <http://dx.doi.org/10.1177/1948550610385872>
- Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 1–50). Mahwah, NJ: Erlbaum.
- Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(Suppl. 10), S70–S81. <http://dx.doi.org/10.1097/00001888-200410001-00022>
- Ericsson, K. A. (2006). The Influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–704). Cambridge, England: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511816796.038>
- Ericsson, K. A. (2009). Enhancing the development of professional performance: Implications from the study of deliberate practice. In K. A. Ericsson (Ed.), *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 405–431). New York: Cambridge University Press.
- Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406. <http://dx.doi.org/10.1037/0033-295X.100.3.363>
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273–305. <http://dx.doi.org/10.1146/annurev.psych.47.1.273>
- Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2009). Toward a science of exceptional achievement: Attaining superior performance through deliberate practice. In W. C. Bushell, E. L. Olivio, & N. D. Theise (Eds.), *Longevity, regeneration, and optimal health: Integrating Eastern and Western perspectives: Annals of the New York Academy of Sciences* (Vol. 1172, pp. 199–217). Hoboken, NJ: Wiley-Blackwell.
- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry, 180*, 51–60. <http://dx.doi.org/10.1192/bjp.180.1.51>
- Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & McGrath, G. (2000). CORE: Clinical outcomes in routine evaluation. *Journal of Mental Health, 9*, 247–255. <http://dx.doi.org/10.1080/713680250>
- Garfield, S. L. (1997). The therapist as a neglected variable in psychotherapy research. *Clinical Psychology: Science and Practice, 4*, 40–43. <http://dx.doi.org/10.1111/j.1468-2850.1997.tb00097.x>
- Gobet, F., & Charness, N. (2006). Expertise in chess. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 523–538). New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511816796.030>
- 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>
- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*, 9–16. <http://dx.doi.org/10.1037/a0022186>
- Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Gordon, D. L., & 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>
- 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>
- Keith, N., & Ericsson, K. A. (2007). A deliberate practice account of typing proficiency in everyday typists. *Journal of Experimental Psychology: Applied, 13*, 135–145. <http://dx.doi.org/10.1037/1076-898X.13.3.135>

- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General*, *125*, 331–359. <http://dx.doi.org/10.1037/0096-3445.125.4.331>
- Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, *48*, 72–79. <http://dx.doi.org/10.1037/a0022238>
- Law, M. P., Côté, J., & Ericsson, K. (2008). Characteristics of expert development in rhythmic gymnastics: A retrospective study. *International Journal of Sport and Exercise Psychology*, *5*, 82–103. <http://dx.doi.org/10.1080/1612197X.2008.9671814>
- Lemov, D., Woolway, E., & Yezzi, K. (2012). *Practice perfect: 42 rules for getting better at getting better*. San Francisco: Jossey-Bass.
- Levitt, H. M., & Williams, D. C. (2010). Facilitating client change: Principles based upon the experience of eminent psychotherapists. *Psychotherapy Research*, *20*, 337–352. <http://dx.doi.org/10.1080/10503300903476708>
- Luborsky, L., McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, *42*, 602–611. <http://dx.doi.org/10.1001/archpsyc.1985.01790290084010>
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*, 173–181. <http://dx.doi.org/10.1023/A:1026595011371>
- McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, linear, and mixed models* (2nd ed.). New York: Wiley.
- Miller, S. D., Duncan, B. L., Brown, J., Sorrell, R., & Chalk, M. B. (2006). Using formal client feedback to improve retention and outcome: Making ongoing real-time assessment feasible. *Journal of Brief Therapy*, *5*, 5–22.
- Miller, S. D., & Hubble, M. (2011). The road to mastery. *Psychotherapy Networker*, *35*, 22–31.
- Miller, S. D., Hubble, M. A., Chow, D. L., & Seidel, J. A. (2013). The outcome of psychotherapy: Yesterday, today, and tomorrow. *Psychotherapy*, *50*, 88–97. <http://dx.doi.org/10.1037/a0031097>
- Miller, S. D., Hubble, M. A., & Duncan, B. L. (2007). Supershrinks. *Psychotherapy Networker*, *31*(6).
- Norcross, J. C., & Guy, J. D. (2007). *Leaving It at the office: A guide to psychotherapist self-care*. New York: Guilford Press.
- Norcross, J. C., & VandenBos, G. R. (2011). Training audiotapes and videotapes. In J. C. Norcross, G. R. VandenBos & D. K. Freedheim (Eds.), *History of psychotherapy: Continuity and change* (2nd ed., pp. 693–702). Washington, DC: American Psychological Association.
- Norman, G., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in medicine and surgery. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 339–354). New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511816796.019>
- Okiishi, J. C., 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). *How psychotherapists develop: A study of therapeutic work and professional growth*. Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/11157-000>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- 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>
- Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 179–187.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32. <http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Simon, W., Lambert, M. J., Harris, M. W., Busath, G., & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: Effects on patients at risk of treatment failure. *Psychotherapy Research*, *22*, 638–647. <http://dx.doi.org/10.1080/10503307.2012.698918>
- Skovholt, T., & Jennings, L. (2005). Mastery and expertise in counseling. *Journal of Mental Health Counseling*, *27*, 13–18.
- Sonnentag, S., & Kleibe, B. M. (2000). Deliberate practice at work: A study with insurance agents. *Journal of Occupational and Organizational Psychology*, *73*, 87–102. <http://dx.doi.org/10.1348/096317900166895>
- 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>
- Starkes, J. L., Deakin, J. M., Allard, F., Hodges, N., & Hayes, A. (1996). Deliberate practice in sports: What is it anyway? In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 81–106). Mahwah, NJ: Erlbaum.
- 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. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.
- Wampold, B. E. (2005). Establishing specificity in psychotherapy scientifically: Design and evidence issues. *Clinical Psychology: Science & Practice*, *12*, 194–197.
- 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>

Received January 22, 2015

Accepted January 23, 2015 ■