



This is a repository copy of *An examination of therapists' professional characteristics as moderators of the effect of feedback on psychological treatment outcomes.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/212532/>

Version: Published Version

---

**Article:**

Delgadillo, J. [orcid.org/0000-0001-5349-230X](https://orcid.org/0000-0001-5349-230X), Nissen-Lie, H.A. [orcid.org/0000-0003-2197-5942](https://orcid.org/0000-0003-2197-5942), De Jong, K. [orcid.org/0000-0002-7621-9290](https://orcid.org/0000-0002-7621-9290) et al. (2 more authors) (2025) An examination of therapists' professional characteristics as moderators of the effect of feedback on psychological treatment outcomes. *Psychotherapy Research*, 35 (3). pp. 501-511. ISSN 1050-3307

<https://doi.org/10.1080/10503307.2024.2310635>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>



## An examination of therapists' professional characteristics as moderators of the effect of feedback on psychological treatment outcomes

Jaime Delgado, Helene A. Nissen-Lie, Kim De Jong, Thomas A. Schröder & Michael Barkham

**To cite this article:** Jaime Delgado, Helene A. Nissen-Lie, Kim De Jong, Thomas A. Schröder & Michael Barkham (13 Feb 2024): An examination of therapists' professional characteristics as moderators of the effect of feedback on psychological treatment outcomes, *Psychotherapy Research*, DOI: [10.1080/10503307.2024.2310635](https://doi.org/10.1080/10503307.2024.2310635)

**To link to this article:** <https://doi.org/10.1080/10503307.2024.2310635>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 13 Feb 2024.



[Submit your article to this journal](#)



Article views: 639



[View related articles](#)



[View Crossmark data](#)

RESEARCH ARTICLE

# An examination of therapists' professional characteristics as moderators of the effect of feedback on psychological treatment outcomes

JAIME DELGADILLO <sup>1</sup>, HELENE A. NISSEN-LIE <sup>2</sup>, KIM DE JONG <sup>3</sup>, THOMAS A. SCHRÖDER <sup>4</sup>, & MICHAEL BARKHAM <sup>1</sup>

<sup>1</sup>Clinical and Applied Psychology Unit, Department of Psychology, University of Sheffield, Sheffield, UK; <sup>2</sup>Department of Psychology, University of Oslo, Oslo, Norway; <sup>3</sup>Institute of Psychology, Leiden University, Leiden, Netherlands & <sup>4</sup>Division of Psychiatry and Applied Psychology, University of Nottingham, Nottingham, UK

(Received 6 December 2023; revised 16 January 2024; accepted 18 January 2024)

## Abstract

**Objective:** Feedback-informed treatment (FIT) has been shown to reduce the gap between more and less effective therapists. This study aimed to examine therapists' professional characteristics as potential moderators of the effect of feedback on treatment outcomes.

**Methods:** The IAPT-FIT Trial was a clinical trial where therapists were randomly assigned to a FIT group or a usual care control group. Treatment response was monitored using measures of depression (PHQ-9), anxiety (GAD-7) and functional impairment (WSAS). In a secondary analysis of this trial ( $n = 1,835$  patients;  $t = 67$  therapists), we used multilevel modelling to examine interactions between therapists' professional characteristics (e.g., attitude towards and self-efficacy regarding feedback utilization, decision-making style, job satisfaction, burnout, difficulties in practice, coping styles, caseload size) with random allocation (FIT vs. controls) to identify moderators of the effects of feedback.

**Results:** Between 9.6% and 10.8% of variability in treatment outcomes was attributable to therapist effects. Therapist-level caseload sizes and external feedback propensity (EFP) moderated the effect of feedback on depression outcomes. No statistically significant main effects were found for any of the included therapist characteristics.

**Conclusion:** FIT reduced variability in outcomes between therapists and was particularly effective for therapists with high EFP and larger caseloads.

**Keywords:** psychotherapy; depression; anxiety; feedback-informed treatment; therapist effects

**Clinical or methodological significance of this article:** Findings from this secondary analysis of data from a clinical trial of feedback-informed treatment conducted in the UK examined potential therapist-level moderators of the effects of feedback. Overall, feedback utilization was associated with better treatment outcomes and reduced the gap between more and less effective therapists. This observation is consistent with results from clinical trials in the USA, cumulatively indicating that the implementation of feedback reduces treatment outcome inequalities that are due to variability in therapist effectiveness. Furthermore, moderator analyses indicated that feedback was especially effective when therapists had large caseloads and had higher external feedback propensity.

Systematic differences in treatment outcomes between psychological therapists, referred to as *therapist effects* (TE), have been extensively documented in narrative and meta-analytic reviews (Baldwin

---

Correspondence concerning this article should be addressed to Jaime Delgado, Clinical and Applied Psychology Unit, Department of Psychology, University of Sheffield, Floor F, Cathedral Court, 1 Vicar Lane, S1 2LT, Sheffield, UK, Email: jaime.delgadillo@nhs.net

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

& Imel, 2013; Crits-Christoph et al., 1991; Johns et al., 2019; Wampold & Owen, 2021). In both randomized controlled trials and naturalistic studies, TE account for approximately 5% to 10% of variability in treatment outcomes after controlling for patients' characteristics (Baldwin & Imel, 2013; Johns et al., 2019). They are evident across several therapy processes and outcomes, including early response (e.g., Erekson et al., 2020), alliance (e.g., Del Re et al., 2021), dropout (e.g., Zimmermann et al., 2017), and symptomatic change (e.g., Baldwin & Imel, 2013). Moreover, TE have been observed across different psychotherapeutic interventions and treatment settings including inpatient, outpatient, and university counselling centres (Wampold & Owen, 2021). Motivated by this body of evidence, studies have sought to identify therapist characteristics that might explain TE. Contemporary reviews of this literature suggest that there are two broad domains that may be relevant: [1] interpersonal and [2] professional characteristics (Castonguay & Hill, 2017; Heinonen & Nissen-Lie, 2020; Wampold & Owen, 2021).

The first domain refers to personal or interpersonal characteristics, which reflect how therapists relate to people and deal with relational difficulties. There is replicated evidence that clinical outcomes are associated with interpersonal skills such as collaboration, empathy, positive regard, alliance building and alliance rupture-repair ability, among others (Norcross & Lambert, 2018). Consistent with this literature, TE studies examining interpersonal behaviours have found that therapists displaying greater *facilitative interpersonal skills* (e.g., emotional expressiveness, warmth, positive regard, hopefulness, empathy, verbal fluency, capacity to repair alliance ruptures) are more effective than others, and this applies to both professionally qualified therapists and lay helpers (Anderson et al., 2009, 2016). Therapists experiencing lower indices of *negative personal reaction* (e.g., a reaction of dislike or intolerance) towards their patients have been found to attain better treatment response (Nissen-Lie et al., 2013). Heinonen et al. (2014) found that the use of *constructive coping skills* (e.g., self-reflection, consultation, problem-solving) is advantageous, while the use of *avoidant coping* (e.g., avoiding stressful interactions) is disadvantageous to therapist effectiveness. Furthermore, therapists with a greater capacity for *reflective functioning* (i.e., ability to conceptualize one's own and others' mental states) were found to attain better treatment outcomes in one study (Cologon et al., 2017).

The second domain of professional characteristics broadly reflects how therapists relate to their work and practice development. Therapists that spend more time on their professional development through

*deliberate practice* of skills may be more effective than those who spend less time according to Chow et al. (2015). One interpretation is that deliberate practice is a proxy indicator of training and experience; however, years of experience and professional qualifications typically have been found to be unrelated to treatment outcomes (Beutler et al., 2004; Goldberg et al., 2016). Hence, engagement in deliberate practice may possibly reflect an attitude of commitment towards learning and improving, which may be favourable for clinical performance (see Ericsson, 2004). In another line of research on professional characteristics, Nissen-Lie et al. (2017) examined therapists' reports of their difficulties in practice (such as feeling frustration or self-doubt in one's clinical work) and different coping mechanisms. This study found that therapists with higher *professional self-doubt* (doubting one's efficacy in helping clients) attained better treatment outcomes. This finding is consistent with evidence from another study showing that therapists who underestimated their effectiveness attained better treatment outcomes compared to therapists who overestimated their effectiveness (Constantino et al., 2023). Contextual factors, such as job satisfaction and occupational burnout, have also been studied in relation to patients' outcomes. In one study, therapists with low job satisfaction and high occupational burnout (specifically an attitude of *disengagement* with work) were found to be less effective than those who reported lower levels of burnout (Delgado et al., 2018a).

Another aspect related to professional attitudes and practice concerns the use of outcome monitoring and feedback. The effectiveness of *progress feedback* as a method to improve clinical outcomes is supported by numerous meta-analyses of clinical trials (see review by Barkham et al., 2023). According to a meta-analysis of clinical trials conducted in the USA, feedback utilization also narrows the gap between more and less effective therapists, helping services to be more equitable and effective (Delgado et al., 2022). Furthermore, the combination of feedback and deliberate practice has been found to improve treatment outcomes at an organizational level (Goldberg et al., 2016). This is an area with much promise in terms of improving the overall effectiveness of mental health services as well as the performance of individual therapists (Wampold & Owen, 2021). However, some studies indicate that therapists who are more committed to using progress feedback obtain better treatment outcomes (e.g., de Jong et al., 2012; Lutz et al., 2021). Such evidence indicates that the effects of feedback vary between therapists, and it is plausible that effective feedback utilization is moderated by therapist-level variables, although there are few studies examining potential moderator variables (de Jong et al., 2021).

While the available findings provide some basis for characterizing effective therapists, the research base is fragmented, with some replicated findings regarding interpersonal skills, some findings derived from a single study, and other findings with mixed and inconclusive evidence. Importantly, very few studies to date have attempted to examine characteristics such as those described above in multivariable analyses. To address this research gap, the present study collected a broad set of therapist variables selected *a priori* based on the literature reviewed above, and which were feasible to obtain through self-reported measures (rather than observed tasks or practice) in a sample of psychological therapists that participated in a clinical trial of feedback-informed treatment (FIT). We aimed to identify potential therapist-level moderators of the effective use of FIT taking an exploratory, inductive and data-driven approach to analysis. This exploratory approach included literature-based professional characteristics described above, demographic characteristics, and additional variables plausibly related to feedback utilization, such as feedback propensity, self-efficacy, decision-making style, alliance in clinical supervision, and clinical caseload size.

## Methods

### Design and Setting

The IAPT-FIT Trial was a multi-site cluster randomized controlled trial that included  $N = 2,233$  patients treated by 77 psychological therapists across eight healthcare organizations in England. These services were part of the *Improving Access to Psychological Therapies* (IAPT) programme in England, a national treatment system offering evidence-based interventions for depression and anxiety disorders (Clark, 2018), and which is now known as *NHS Talking Therapies Services*. Participating therapists were randomly assigned to a FIT group or a usual care control group. All patients who received at least two sessions of individual therapy with participating therapists during a one-year study period were included in the trial sample, including completers and dropouts. Patients attending group therapies were excluded; and those who only attended a single therapy session were excluded since the FIT technology only starts to provide feedback signals after session 1. The trial was pre-registered in the international register of controlled trials (ISRCTN12459454) and approved by an independent NHS Research Ethics Committee (Ref: 15/LO/2200). Further details about the study design, inclusion and exclusion criteria, consort diagram,

sample size calculation and primary results are available elsewhere (Delgado et al., 2018b).

### Interventions

Participating services delivered evidence-based psychological interventions for common mental disorders organized in a stepped care model, following national guidelines (National Institute for Health and Care Excellence, 2011). These were protocol-driven interventions delivered under regular supervision (equivalent of 1h per week) to ensure adherence to treatment-specific competency standards (National IAPT Team, 2015; Roth & Pilling, 2008). In this stepped care model, most patients initially accessed low intensity treatments, which were brief ( $\leq 8$  sessions) psychoeducational interventions based on principles of cognitive behavioural therapy (CBT). High intensity treatments were offered to patients whose symptoms persisted after receiving a low intensity treatment, and to patients with more severe symptoms or with conditions for which only psychotherapy is indicated (e.g., post-traumatic stress disorder). High intensity treatments included empirically supported therapies lasting up to 20 sessions, such as CBT, interpersonal psychotherapy and person-centred experiential counselling for depression. All participating therapists had post-graduate qualifications to deliver one of the above interventions. Some therapists had specialist training in clinical areas such as the treatment of post-traumatic stress disorder, couples therapy for marital problems, or mental health problems complicated by comorbid chronic illnesses.

Therapists randomized to the FIT group had access to a computerized outcome monitoring and feedback tool that compared patients' depression and anxiety symptoms to normative clinical data, classifying patients as "on track" or "not on track". These therapists were trained to monitor and discuss this feedback with patients at the start of every therapy session, and they were instructed to prioritize "not on track" cases for discussion with their clinical supervisors, in order to identify obstacles for improvement and to adjust the treatment plan accordingly. All other aspects of treatment (e.g., interventions, frequency of clinical supervision) were standardized across the FIT and the control group.

### Measures

**Patient-level data.** Trial participants completed two patient-reported outcome measures at the start of each therapy session. The PHQ-9 is a measure

of depression symptoms, where each of 9 questions is rated from 0 to 3, yielding an overall severity score between 0 and 27, and a cut-off of  $\geq 10$  is used to screen for clinically significant symptoms (Kroenke et al., 2001). Cronbach's alpha as an index of reliability (i.e., internal consistency) in the present sample was  $\alpha = 0.86$ . The GAD-7 is a 7-item measure of generalized anxiety disorder. Each item is rated between 0 and 3, with a total severity score between 0 and 21, and a cut-off score  $\geq 8$  is recommended to screen for clinically significant anxiety disorders (Kroenke et al., 2007). The sample reliability was  $\alpha = 0.86$ .

Patient-reported demographic data included age, gender, ethnicity, employment status, comorbid long-term conditions (chronic illnesses such as diabetes, heart disease, pulmonary disease, etc.). The index of multiple deprivation (IMD) is a neighbourhood-level index of socioeconomic deprivation derived from each patient's home postcode (Department for Communities & Local Government, 2011). The Work and Social Adjustment Scale (WSAS) is a composite measure of functional impairment (Mundt et al., 2002) across five domains (work, home management, social life, leisure activities, family and relationships); reliability in the present sample was  $\alpha = 0.79$ . Additional clinical information retrieved from healthcare records included the level of intensity of treatment (low vs. high intensity) and use of antidepressant medications.

**Therapist-level data.** A battery of self-reported measures was completed by participating therapists prior to randomization. The selection of measures was chosen to gather self-reported information about each therapist's role, training, approach to practice development, use of supervision, attitude towards and self-efficacy regarding feedback utilization, decision-making style, job satisfaction, burnout, difficulties in practice and coping styles.

Demographic variables included age, gender and ethnicity. Occupational variables included years of clinical training, years of clinical experience, the level of intensity of treatment they delivered (low or high intensity in the stepped care model), the number of monthly hours of clinical supervision they received in a typical month.

Informed by previous research on deliberate practice (Chow et al., 2015), we assessed the amount of time therapists spent in practice outside of work aimed at improving clinical knowledge and skills. Therapists self-reported how often they engaged in the following practice development activities during the last year (Likert scale from 0 = not at all, to 5 = very often): independent professional reading; journal club or

discussion group with colleagues; clinical training seminars/courses; experiential workshops; practice of therapy skills (items = 5; sample  $\alpha = 0.68$ ).

The Leeds Alliance in Supervision Scale (LASS) is a brief questionnaire based on 3 visual analogue scales that together assess respondents' impressions about the approach, relationship and helpfulness of clinical supervision sessions (Wainwright, 2010). The sample reliability was  $\alpha = 0.71$ .

The internal and external feedback propensity scale is a 12-item questionnaire by Herold and Fedor (2003). Internal feedback propensity (items = 6; sample  $\alpha = 0.76$ ) refers to a preference for evaluating task performance based on one's own perspective rather than other people's opinions or perspectives. External feedback propensity (items = 6; sample  $\alpha = 0.69$ ) refers to a preference for evaluating task performance based on other people's perspective and opinions.

An adapted version of the Contextualized Feedback Intervention Theory (CFIT) questionnaire (items = 7; sample  $\alpha = 0.86$ ) was used to measure therapists' self-efficacy in engaging patients in therapy, establishing a collaborative relationship, monitoring progress and identifying and resolving difficulties (Riemer & Bickman, 2011).

The decision-making style (DMS) questionnaire is based on the literature on rational and intuitive decision styles (Hamilton et al., 2016), clinical decision-making in stepped care services (Delgado et al., 2015), and was developed for the present study. The questionnaire measures two domains. Intuitive decision-making (items = 9; sample  $\alpha = 0.87$ ) reflects a proclivity for reliance on clinical intuition, "gut instinct", feelings and perceptions about the therapist-patient relationship. Empirical decision-making reflects a proclivity for reliance on outcome measures, clinical guidelines and research evidence (items = 9; sample  $\alpha = 0.83$ ).

The Oldenburg Burnout Inventory (OLBI; Demerouti et al., 2001) is a 16-item questionnaire that measures two facets of burnout, emotional exhaustion (OLBI-E; items = 8; sample  $\alpha = 0.70$ ) and disengagement (OLBI-D; items = 8; sample  $\alpha = 0.79$ ). The Job Discrepancy and Satisfaction Scale (JDSS; sample  $\alpha = 0.64$ ) is an 8-item measure that addresses the extent to which practitioners are satisfied with their current working conditions (Nagy, 2002).

The Development of Psychotherapists Common Core Questionnaire (DPCCQ) is a multi-domain self-assessment designed to collect information about therapists' professional characteristics, attitudes and practice (Orlinsky & Rønnestad, 2005). We included four DPCCQ domains: professional self-doubt (PSD; items = 9; sample  $\alpha = 0.86$ ), negative personal reaction (NPR; items = 6; sample  $\alpha = 0.79$ ), constructive coping (items = 10; sample  $\alpha =$

0.73) and unconstructive coping (items = 8; sample  $\alpha$  = 0.66).

**Sample Selection and Characteristics**

To meet the minimal sample size criteria to investigate therapist effects using multilevel modelling (Schiefele et al., 2017), we only included therapists that treated a minimum of four patients. Therefore, the present study analysed a subsample ( $n = 1,835$  patients;  $t = 67$  therapists) of the IAPT-FIT Trial ( $N = 2,233$ ;  $t = 77$ ). Table I summarizes participating therapists' characteristics in the full sample, and subsamples of therapists randomized to the feedback and control groups. Most therapists were white British (83.6%) females (83.6%), with a mean age of 41.01 (SD = 11.33) and with an average of 7.56 (SD = 5.68) years of clinical experience. The majority were qualified cognitive behavioural therapists (58.2%) or psychological wellbeing practitioners (31.3%). Caseload sizes ranged widely (4–101), with an average of 35.57 (SD = 24.08) for psychological wellbeing practitioners delivering low

intensity treatments, and 23.65 (SD = 17.31) for therapists delivering high intensity treatments.

Most patients were white British (88.4%) females (66.4%), with a mean age of 39.82 (SD = 15.10), of whom 20.7% were unemployed and 35.5% had a comorbid long-term chronic health problem. Primary problems recorded in clinical records included affective disorders (34.7%), generalized anxiety disorder (13.6%), mixed anxiety and depressive disorder (15.6%), and other common mental health problems. In total, 41.2% accessed a low intensity treatment, 58.8% accessed a high intensity treatment, and no participants in this study sample accessed both types of treatment. Additional sample characteristics of the full sample and subgroups (feedback vs. controls) are available in the Supplemental Materials (S1).

**Statistical Analysis**

The statistical analysis aimed to identify therapist-level moderators of feedback effects, including all available variables listed in the measures section.

Table I. Therapist sample characteristics.

Variables	Full sample	Feedback group	Control group
Age: mean (SD)	41.01 (11.33)	41.67 (11.47)	40.38 (11.34)
Females (%)	56 (83.6%)	25 (75.8%)	31 (91.2%)
White British background (%)	56 (83.6%)	27 (81.8%)	29 (85.3%)
Qualifications*			
PWP (%)	21 (31.3%)	10 (30.3%)	11 (32.4%)
CBT (%)	39 (58.2%)	18 (54.5%)	21 (61.8%)
Counsellor (%)	5 (7.5%)	4 (12.1%)	1 (2.9%)
IPT (%)	6 (9.0%)	3 (9.1%)	3 (8.8%)
EMDR (%)	7 (10.4%)	5 (15.2%)	2 (5.9%)
DClinPsy (%)	2 (3.0%)	0	2 (5.9%)
Training: mean years (SD)	4.41 (2.79)	4.29 (2.81)	4.52 (2.82)
Experience: mean years (SD)	7.56 (5.68)	7.87 (5.91)	7.25 (5.51)
Caseload size: mean (SD)	27.39 (20.27)	26.73 (20.18)	28.03 (20.65)
Supervision: mean hours / month (SD)	3.19 (1.42)	3.32 (1.33)	3.07 (1.51)
LASS mean (SD)	4.05 (0.71)	4.03 (0.74)	4.08 (0.70)
Deliberate practice mean (SD)	2.69 (0.98)	2.61 (1.10)	2.76 (0.87)
OLBI-E mean (SD)	2.39 (0.45)	2.39 (0.46)	2.39 (0.45)
OLBI-D mean (SD)	2.17 (0.52)	2.19 (0.57)	2.14 (0.48)
JDSS mean (SD)	2.77 (0.42)	2.92 (0.47)	2.63 (0.32)
PSD mean (SD)	1.32 (0.70)	1.43 (0.72)	1.20 (0.66)
NPR mean (SD)	0.81 (0.60)	0.87 (0.64)	0.74 (0.56)
Constructive coping mean (SD)	3.44 (0.64)	3.25 (0.68)	3.62 (0.55)
Unconstructive coping mean (SD)	1.08 (0.59)	1.19 (0.49)	0.97 (0.67)
Self-efficacy mean (SD)	21.16 (3.99)	20.21 (4.26)	22.09 (3.52)
Internal FP mean (SD)	19.04 (3.66)	18.58 (3.91)	19.50 (3.40)
External FP mean (SD)	13.10 (3.38)	13.21 (3.42)	13.00 (3.39)
Intuitive DMS mean (SD)	1.56 (0.67)	1.71 (0.65)	1.42 (0.66)
Empirical DMS mean (SD)	2.80 (0.57)	2.64 (0.64)	2.96 (0.43)

\*Percentages for qualifications do not add to exactly 100% because some participants had more than one qualification; PWP = psychological wellbeing practitioner; CBT = cognitive behavioural therapy; Counsellor = person-centred experiential counselling for depression; IPT = interpersonal psychotherapy; EMDR = eye-movement desensitization and reprocessing; DClinPsy = doctorate in clinical psychology; LASS = Leeds alliance in supervision

This enabled us to examine variables that may be plausible moderators of feedback effects (e.g., internal/external feedback propensity, self-efficacy, decision-making style, etc.), while controlling for variables which we expected to be related to therapist effects (e.g., deliberate practice, therapeutic self-doubt, burnout, etc.). This analysis followed an exploratory (i.e., data-driven) approach, and investigated interactions between therapist-level variables and feedback utilization (feedback vs. control group). The analysis followed an iterative model-building process described below.

**Model-building approach and estimation of therapist effects.** We applied multilevel modeling (MLM) nesting patients (level 1) within therapists (level 2) within services (level 3). Consistent with conventional model-building guidelines (Raudenbush, 1993), continuous predictors were grand mean-centred and MLM was performed in sequential steps, starting with single-level models and eventually developing multi-level and covariate-adjusted models that optimized goodness-of-fit. Model fit was examined after each modelling step by inspecting the standard error of regression coefficients and the  $-2$  log-likelihood ratio test. The intraclass correlation coefficient (ICC) was also calculated at each step, as an index of variability in treatment outcomes attributable to higher-level nesting variables (therapists, services). We retained and interpreted the best-fitting and most parsimonious model achieved through this stepwise process.

Initial model-building followed five steps. Step 1 was a single-level null model (i.e., with no predictors), including post-treatment PHQ-9 as the dependent variable. Step 2 was a two-level model including a random intercept for the therapist level. Step 3 was a three-level model additionally including a random intercept for the service level. These initial steps confirmed that a two-level model had the best goodness-of-fit, and this structure was retained in the remaining steps. Step 4 entered all available patient-level variables (age, gender, ethnicity, IMD, employment status, comorbid chronic illness, use of antidepressant medication, intensity of treatment [low = 0, high = 1], baseline PHQ-9, GAD-7, WSAS). In step 5, non-significant patient-level predictors were removed to attain a parsimonious case-mix adjusted model. The same stepwise analysis strategy was applied to examine therapist effects in post-treatment anxiety (GAD-7) and functional impairment (WSAS). These models were performed in the full sample, without modelling the effects of feedback, and this enabled us to obtain parsimonious case-mix adjusted models.

**Examination of moderators of the effects of feedback.** The exploration of potential moderators followed four steps. First, a two-level case-mix adjusted model (including only significant patient-level variables) was fitted in the control group sample, entering all available treatment (treatment intensity and type) and therapist-level variables. Second, the same model was fitted in the feedback group sample. We compared the outputs from both samples to select potential prognostic indicators (significant predictors in both samples) and potential moderators of feedback effects (significant predictors in only one of the models). This variable selection approach offers a parsimonious way to identify potential moderators, avoiding the need to introduce a large number of interaction terms into an exploratory model (e.g., 21 interaction terms in addition to 31 main effects) or to resort to bias-prone forward selection approaches (Heinze et al., 2018). In the third step, we introduced the selected main (i.e., fixed) effects and interaction terms (moderator\*feedback) into a two-level model that was fitted in the full sample, controlling for the main effect feedback (vs. control group). Finally, in the fourth step we refitted a parsimonious and best-fitting model only retaining statistically significant predictors. This strategy enabled us to avoid fitting noise variables into the final models, so as to model therapist effects and main effects with higher precision. This modelling strategy was repeated with each of the three outcome measures (PHQ-9, GAD-7, WSAS) as the dependent variable.

## Results

Statistical outputs from all steps in the modelling process are available in the Supplemental Materials (Tables S1-28). Parsimonious models controlling for statistically significant patient-level predictors indicated that therapist effects accounted for 9.6% of variability in depression outcomes (supplement S6), 9.9% of variability in anxiety outcomes (S15), and 10.8% of variability in functional impairment (S24). The most consistent predictors of treatment response across all outcome measures were patients' age, employment status, baseline depression (PHQ-9) and anxiety (GAD-7). Comorbidity of chronic illnesses predicted poor outcomes in depression (PHQ-9) and functional impairment (WSAS) models. Minority ethnic group also predicted higher post-treatment functional impairment (WSAS).

The effect of feedback on depression treatment outcomes was significantly moderated by only two therapist-level variables: [1] caseload size and [2] external feedback propensity (see Table II).

Table II. Moderators of feedback effects on depression (PHQ-9) treatment outcomes.

<i>Fixed effects</i>							
Variables	B	SE	t	p	CI-low	CI-high	
Intercept	8.669	0.372	23.333	<0.001	7.941	9.398	
Age (mc)	-0.027	0.009	-2.922	0.004	-0.045	-0.009	
Unemployed	2.726	0.330	8.260	<0.001	2.078	3.373	
LTC	0.849	0.286	2.967	0.003	0.288	1.411	
PHQ-9 (mc)	0.395	0.033	11.818	<0.001	0.330	0.461	
GAD-7 (mc)	0.088	0.035	2.512	0.012	0.019	0.156	
WSAS (mc)	0.063	0.020	3.161	0.002	0.024	0.103	
Caseload size (mc)	0.028	0.015	1.843	0.066	-0.002	0.058	
External FP (mc)	0.343	0.095	3.601	<0.001	0.156	0.530	
Group=FB	-0.944	0.509	-1.855	0.064	-1.942	0.054	
Caseload size (mc) * FB	-0.069	0.022	-3.223	0.001	-0.112	-0.027	
External FP (mc) * FB	-0.326	0.135	-2.427	0.015	-0.590	-0.063	
<i>Covariance parameters</i>							
Effects	Variance	SE	Z	p	CI-low	CI-high	ICC
Residual effect	29.476	0.992	29.721	<0.001	27.595	31.486	
Random effect (therapists)	2.098	0.627	3.345	0.001	1.168	3.769	0.066

B = regression coefficient; SE = standard error; CI = 95% confidence intervals; ICC = intraclass correlation coefficient; mc = mean centred; LTC = long-term condition; PHQ-9 = patient health questionnaire (depression); GAD-7 = generalized anxiety disorder questionnaire (anxiety); WSAS = work and social adjustment scale; external FP = external feedback propensity; group = feedback users (vs. controls); note that main (i.e., fixed) effects are retained for variables that have significant interaction terms; -2 log likelihood = 11515.435

Treatment outcomes were similar for therapists with smaller caseloads (<40 patients per year). However, as caseload sizes increased, therapists in the feedback group had considerably better treatment outcomes, as shown in Figure 1 - panel A. Furthermore, average post-treatment depression scores were in the clinical range (PHQ-9 ≥ 10) for control-group therapists with a yearly caseload size above 70, whereas average depression scores were in the sub-clinical range

for feedback-group therapists with comparably large caseloads.

As shown in Figure 1 - panel B, post-treatment depression scores were generally higher for therapists with a high external feedback propensity. However, this general trend for poorer outcomes at high-levels of external feedback propensity was mitigated by feedback utilization, since feedback-group therapists had significantly better treatment outcomes relative to those in the control group. Furthermore,

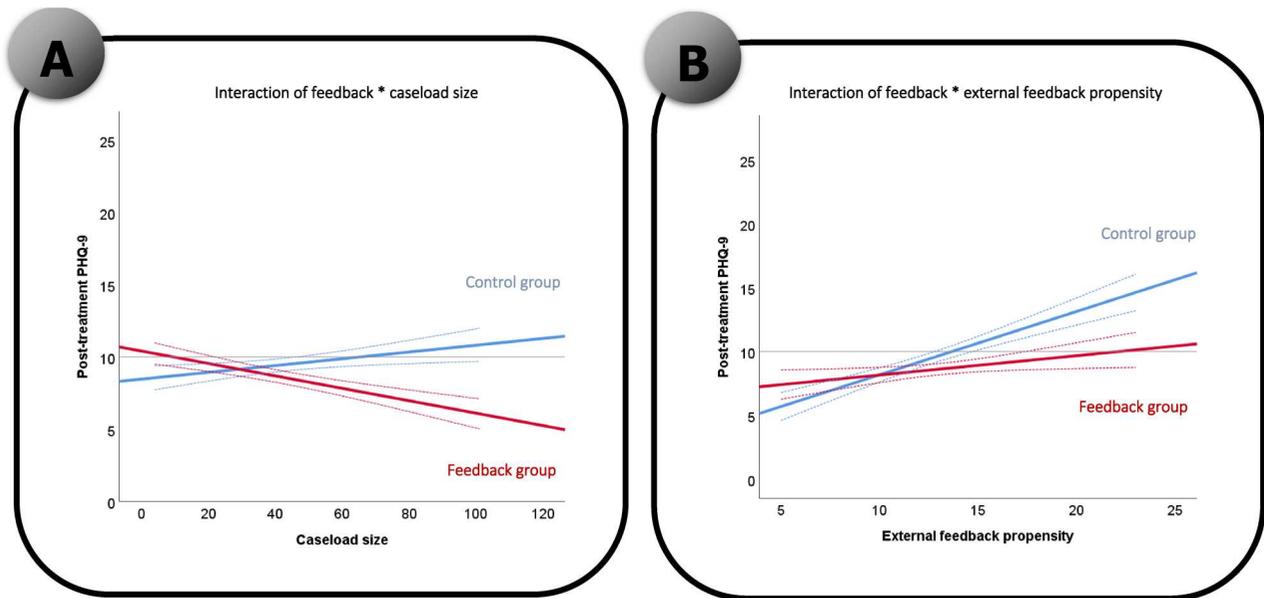


Figure 1. Moderators of the effect of feedback on depression (PHQ-9) treatment outcome.

only caseload size was a significant moderator of the effect of feedback in models examining anxiety (supplement S19) and functional impairment (supplement S28) as treatment outcomes. These models which included feedback and moderators of feedback utilization had smaller TE indices (ICC = 0.066 for depression, 0.088 for anxiety, 0.099 for functional impairment) in comparison to the initial models that only adjusted for patient-level variables (S6, S15, S24). No other therapist-level characteristics had replicated support as statistically significant main effects when they were examined in the feedback and control samples, nor in the full sample.

### Discussion

The literature on TE has identified several interpersonal and professional characteristics that may characterize effective therapists (Wampold & Owen, 2021). However, few of these findings have been replicated and there is a lack of multivariable analyses testing the robustness of multiple therapist characteristics in the same sample. Furthermore, there are few investigations of therapist characteristics as potential moderators of treatment effects in psychotherapy studies (Heinonen & Nissen-Lie, 2020) and in studies of feedback-informed treatment (de Jong et al., 2021). This study aimed to contribute to this literature by examining a selection of therapist characteristics that have been associated with treatment outcomes in prior studies, and by exploring potential moderators of the effects of progress feedback.

The results do not provide any support for the included variables that have been associated with treatment outcomes in prior studies. It is important to note that the majority of variables collected in this study are related to therapists' professional characteristics, and only few variables provided information about interpersonal aspects (negative personal reaction, constructive and unconstructive coping). Overall, the variables assessed in this study do not seem to enable us to reliably differentiate between more and less effective therapists in this treatment setting.

However, consistent with prior evidence (Baldwin & Imel, 2013; Johns et al., 2019), TEs were significant in this study sample, explaining around 10% of variability in treatment outcomes. The magnitude of the TE reduced once feedback utilization and moderators were included in the analysis. For example, this reduced from 9.6% to 6.6% in the model of depression treatment outcomes and from 9.9% to 8.8% in the model for anxiety outcomes. These findings indicate that feedback utilization reduces variability in performance between therapists, and that the beneficial effect of feedback is

moderated by two variables which we examined. Specifically, the effects of feedback were more positive when therapists had larger patient caseloads. This is plausibly explained by the experimental procedure in the FIT group, since these therapists were specifically instructed to prioritize discussing "not on track" cases in their clinical supervision consultations. Given that therapists with large caseloads cannot thoroughly discuss all their cases in the limited time available for supervision, it is likely that the FIT method of case selection optimally prioritizes those cases where careful reflection about the case formulation and modification of the treatment plan may be necessary. Prior evidence in the field of clinical supervision suggests that supervision time can often be devoted to a heterogeneous set of tasks and discussions (Milne & Gracie, 2001; Simpson-Southward et al., 2017), and methods of case selection and discussion are not always effective (e.g., Milne et al., 2009). It seems that the FIT method may correct for this potential problem, ensuring that clinical supervision is systematic and outcome-oriented. An alternative explanation could be that therapists with higher caseloads also get more practice at using the FIT method and become more effective for this reason. Of course, these explanations are not mutually exclusive, and it may be that a combination of practice and strategic selection of cases for clinical supervision may explain this finding.

The second moderator was external feedback propensity (EFP), referring to a preference for evaluating one's performance based on other people's opinions or perspectives (see Herold & Fedor, 2003). Therapists with low EFP generally attained better treatment outcomes than those with high EFP, as shown in Figure 1. However, the seemingly hindering effect of high EFP was attenuated by feedback-utilization. As EFP reflects a tendency to rely strongly on validation and reassurance from others (i.e., patients, peers, supervisors) in order to feel competent, it could be that therapists who strongly require validation may come across as less confident and credible to their patients, thus undermining their expectations about improvement. This interpretation fits with wider evidence that the perceived credibility of therapy is associated with treatment outcomes (Constantino et al., 2018). Furthermore, therapists in the FIT group were trained to have structured discussions with patients who were classed as "not on track", providing them a specific framework to explore potential obstacles to improvement in a collaborative way (Delgado et al., 2018b). Hence, feedback technology provided unambiguous information about whether patients are responding to treatment as expected, which is a concrete form of external feedback. The FIT procedure and training provided direction on how to approach

and solve difficulties in practice, potentially helping therapists with high EFP to behave in a more decisive, structured and credible manner.

### Strengths and Limitations

To our knowledge, this is one of the few studies that systematically gathered a wide array of therapist characteristics in a large enough sample (including enough patients per therapist) to model therapist effects reliably. The experimental design simultaneously enabled the examination of therapist variables as outcome predictors as well as moderators of the effect of feedback. In addition, the multi-centre and pragmatic design of this trial enabled us to study therapist effects in a diverse sample of patients and therapists recruited from eight health services from different regions of England. As reported above, all psychometric measures had adequate reliability and were selected *a priori* based on prior literature and evidence.

The findings should also be interpreted in light of some methodological shortcomings. Although we were able to gather over 20 therapist-level variables, these mostly reflect professional characteristics. Hence, we cannot draw firm conclusions about interpersonal aspects, since only few such variables were collected. All measures collected in this study were self-reported by patients and therapists, which may be influenced by social desirability bias and demand characteristics. The use of observer-rated assessments such as the facilitative interpersonal skills (FIS) task seems to be a promising area for research into therapist effects (Anderson et al., 2009, 2016). However, interpersonal variables such as the FIS are drawn from observation and ratings of clinical behaviours and are more difficult to collect in large-scale studies. Furthermore, a lack of therapy session recordings precluded us from investigating therapy processes in a more direct way.

### Conclusions

We found little evidence that therapists' professional characteristics are associated with differential treatment outcomes. Consistent with prior evidence from the USA (Delgado et al., 2022), feedback-informed treatment reduced the magnitude of therapist effects at a service-wide level in this UK treatment setting, enhancing the effectiveness of psychological therapy and minimizing the gap between more and less effective therapists. Therapists with a strong external feedback propensity benefitted most from using progress feedback, since this provided them with a clear and structured way to support patients who were not on

track. These findings provide additional reasons to implement feedback in routine care, and particularly in services where therapists may have large caseloads.

### Disclosure Statement

No potential conflict of interest was reported by the authors.

### Data Sharing Policy

In line with the requirements of the ethics review board for this study, requests for access to data are to be made in writing to the corresponding author.

### Supplemental data

Supplemental data for this article can be accessed <http://dx.doi.org/10.1080/10503307.2024.2310635>.

### ORCID

*Jaime Delgado*  <http://orcid.org/0000-0001-5349-230X>

*Helene A. Nissen-lie*  <http://orcid.org/0000-0003-2197-5942>

*Kim de Jong*  <http://orcid.org/0000-0002-7621-9290>

*Thomas A. Schröder*  <http://orcid.org/0000-0002-6466-4960>

*Michael Barkham*  <http://orcid.org/0000-0003-1687-6376>

### References

- Anderson, T., McClintock, A. S., Himawan, L., Song, X., & Patterson, C. L. (2016). A prospective study of therapist facilitative interpersonal skills as a predictor of treatment outcome. *Journal of Consulting and Clinical Psychology, 84*(1), 57–66. <https://doi.org/10.1037/ccp0000060>
- Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology, 65*(7), 755–768. <https://doi.org/10.1002/jclp.20583>
- Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert, & M. J. Lambert (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change, 6th ed* (pp. 258–297). Wiley.
- Barkham, M., De Jong, K., Delgado, J., & Lutz, W. (2023). Routine Outcome Monitoring (ROM) and feedback: Research review and recommendations. *Psychotherapy Research, 33*(7), 841–855. <https://doi.org/10.1080/10503307.2023.2181114>
- 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). Wiley.
- Castonguay, L. G., & Hill, C. E. (Eds.). (2017). *How and why are some therapists better than others?: Understanding therapist effects*. American Psychological Association. <https://doi.org/10.1037/000034-000>
- Chow, D. L., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J. A., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy*, 52(3), 337–345. <https://doi.org/10.1037/pst0000015>
- Clark, D. M. (2018). Realizing the mass public benefit of evidence-based psychological therapies: The IAPT program. *Annual Review of Clinical Psychology*, 14(1), 159–183. <https://doi.org/10.1146/annurev-clinpsy-050817-084833>
- Cologon, J., Schweitzer, R. D., King, R., & Nolte, T. (2017). Therapist reflective functioning, therapist attachment style and therapist effectiveness. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(5), 614–625. <https://doi.org/10.1007/s10488-017-0790-5>
- Constantino, M. J., Boswell, J. F., Coyne, A. E., Muir, H. J., Gaines, A. N., & Kraus, D. R. (2023). Therapist perceptions of their own measurement-based, problem-specific effectiveness. *Journal of Consulting and Clinical Psychology*, 91(8), 474–484. <https://doi.org/10.1037/ccp0000813>
- Constantino, M. J., Coyne, A. E., Boswell, J. F., Iles, B. R., & Vislä, A. (2018). A meta-analysis of the association between patients' early perception of treatment credibility and their posttreatment outcomes. *Psychotherapy*, 55(4), 486–495. <https://doi.org/10.1037/pst0000168>
- Crits-Christoph, P., Baranackie, K., Kurcias, J., Beck, A., Carroll, K., Perry, K., Luborsky, L., McLellan, A., Woody, G., Thompson, L., Gallagher, D., & Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research*, 1(2), 81–91. <https://doi.org/10.1080/10503309112331335511>
- de Jong, K., Conijn, J. M., Gallagher, R. A., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, 102002. <https://doi.org/10.1016/j.cpr.2021.102002>
- de Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: Therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy Research*, 22(4), 464–474. <https://doi.org/10.1080/10503307.2012.673023>
- Delgado, J., Deisenhofer, A.-K., Probst, T., Shimokawa, K., Lambert, M. J., & Kleinstäuber, M. (2022). Progress feedback narrows the gap between more and less effective therapists: A therapist effects meta-analysis of clinical trials. *Journal of Consulting and Clinical Psychology*, 90(7), 559–567. <https://doi.org/10.1037/ccp0000747>
- Delgado, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., Ali, S., Aguirre, E., Appleton, M., Nevin, J., O'Hayon, H., Patel, U., Sainy, A., Spencer, P., & McMillan, D. (2018b). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multi-site, open-label, cluster randomised controlled trial. *The Lancet Psychiatry*, 5(7), 564–572. [https://doi.org/10.1016/S2215-0366\(18\)30162-7](https://doi.org/10.1016/S2215-0366(18)30162-7)
- Delgado, J., Gellatly, J., & Stephenson-Bellwood, S. (2015). Decision making in stepped care: How do therapists decide whether to prolong treatment or not? *Behavioural and Cognitive Psychotherapy*, 43(3), 328–341. <https://doi.org/10.1017/S135246581300091X>
- Delgado, J., Saxon, D., & Barkham, M. (2018a). Associations between therapists' occupational burnout and their patients' depression and anxiety treatment outcomes. *Depression and Anxiety*, 35(9), 844–850. <https://doi.org/10.1002/da.22766>
- Del Re, A. C., Flückiger, C., Horvath, A. O., & Wampold, B. E. (2021). Examining therapist effects in the alliance–outcome relationship: A multilevel meta-analysis. *Journal of Consulting and Clinical Psychology*, 89(5), 371–378. <https://doi.org/10.1037/ccp0000637>
- Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands–resources model of burnout. *Journal of Applied Psychology*, 86(3), 499–512. <https://doi.org/10.1037/0021-9010.86.3.499>
- Department for Communities and Local Government. (2011). The English indices of deprivation 2010. <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010>.
- Erekson, D. M., Clayson, R., Park, S. Y., & Tass, S. (2020). Therapist effects on early change in psychotherapy in a naturalistic setting. *Psychotherapy Research*, 30(1), 68–78. <https://doi.org/10.1080/10503307.2018.1556824>
- Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, 79(10), S70–S81. <https://doi.org/10.1097/00001888-200410001-00022>
- Goldberg, S. B., Babins-Wagner, R., Rousmaniere, T., Berzins, S., Hoyt, W. T., Whipple, J. L., Miller, S. D., & Wampold, B. E. (2016). Creating a climate for therapist improvement: A case study of an agency focused on outcomes and deliberate practice. *Psychotherapy*, 53(3), 367–375. <https://psycnet.apa.org/doi/10.1037/pst0000060>.
- Hamilton, K., Shih, S. I., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of Personality Assessment*, 98(5), 523–535. <https://doi.org/10.1080/00223891.2015.1132426>
- Heinonen, E., Knekt, P., Jääskeläinen, T., & Lindfors, O. (2014). Therapists' professional and personal characteristics as predictors of outcome in long-term psychodynamic psychotherapy and psychoanalysis. *European Psychiatry*, 29(5), 265–274. <https://doi.org/10.1016/j.eurpsy.2013.07.002>
- Heinonen, E., & Nissen-Lie, H. (2020). The professional and personal characteristics of effective psychotherapists: A systematic review. *Psychotherapy Research*, 30(4), 417–432. <https://doi.org/10.1080/10503307.2019.1620366>
- Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection—a review and recommendations for the practicing statistician. *Biometrical Journal*, 60(3), 431–449. <https://doi.org/10.1002/bimj.201700067>
- Herold, D. M., & Fedor, D. B. (2003). Individual differences in feedback propensities and training performance. *Human Resource Management Review*, 13(4), 675–689. <https://doi.org/10.1016/j.hrmr.2003.11.008>
- Johns, R. G., Barkham, M., Kelleth, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to baldwin and imel's (2013) review. *Clinical Psychology Review*, 67, 78–93. <https://doi.org/10.1016/j.cpr.2018.08.004>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, 146(5), 317–325. <https://doi.org/10.7326/0003-4819-146-5-200703060-00004>
- Lutz, W., Deisenhofer, A.-K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2021).

- Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology*, 90(1), 90–106. <https://doi.org/10.1037/ccp0000642>
- Milne, D. L., & Gracie, J. (2001). The role of the supervisee: 20 ways to facilitate clinical supervision. *Clinical Psychology Forum*, 5, 13–15.
- Milne, D. L., Leck, C., & Choudhri, N. (2009). Collusion in clinical supervision: Review and case study in self-reflection. *The Cognitive Behaviour Therapist*, 2(2), 106–114. <https://doi.org/10.1017/S1754470X0900018X>
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The work and social adjustment scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*, 180(05), 461–464. <https://doi.org/10.1192/bjp.180.5.461>
- Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. *Journal of Occupational and Organizational Psychology*, 75(1), 77–86. <https://doi.org/10.1348/096317902167658>
- National IAPT Team. (2015). *National curriculum for the education of psychological wellbeing practitioners* (3rd ed.). NHS England/Department of Health.
- National Institute for Health and Care Excellence. (2011). *Common mental health problems: Identification and pathways to care*. National Collaborating Centre for Mental Health.
- Nissen-Lie, H. A., Monsen, J. T., Ulleberg, P., & Rønnestad, M. H. (2013). Psychotherapists' self-reports of their interpersonal functioning and difficulties in practice as predictors of patient outcome. *Psychotherapy Research*, 23(1), 86–104. <https://doi.org/10.1080/10503307.2012.735775>
- Nissen-Lie, H. A., Rønnestad, M. H., Høglend, P. A., Havik, O. E., Solbakken, O. A., Stiles, T. C., & Monsen, J. T. (2017). Love yourself as a person, doubt yourself as a therapist? *Clinical Psychology & Psychotherapy*, 24(1), 48–60. <https://doi.org/10.1002/cpp.1977>
- Norcross, J. C., & Lambert, M. J. (2018). Psychotherapy relationships that work III. *Psychotherapy*, 55(4), 303–315. <https://doi.org/10.1037/pst0000193>
- Orlinsky, D. E., & Rønnestad, M. H. (2005). *How psychotherapists develop: A study of therapeutic work and professional growth*. American Psychological Association.
- Raudenbush, S. W. (1993). Hierarchical linear models and experimental design. In L. Edwards (Ed.), *Applied analysis of variance in behavioral science* (pp. 459–496). Marcel Dekker.
- Riemer, M., & Bickman, L. (2011). Using program theory to link social psychology and program evaluation. In M. M. Mark, S. I. Donaldson, & B. Campbell (Eds.), *Social psychology and evaluation* (pp. 104–139). Guilford Press.
- Roth, A. D., & Pilling, S. (2008). Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy*, 36(02), 129–147. <https://doi.org/10.1017/S1352465808004141>
- Schiefele, A. K., Lutz, W., Barkham, M., Rubel, J., Böhnke, J., Delgadillo, J., Kopta, M., Schulte, D., Saxon, D., Nielsen, S. L., & Lambert, M. J. (2017). Reliability of therapist effects in practice-based psychotherapy research: A guide for the planning of future studies. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(5), 598–613. <https://doi.org/10.1007/s10488-016-0736-3>
- Simpson-Southward, C., Waller, G., & Hardy, G. E. (2017). How do we know what makes for “best practice” in clinical supervision for psychological therapists? A content analysis of supervisory models and approaches. *Clinical Psychology & Psychotherapy*, 24(6), 1228–1245. <https://doi.org/10.1002/cpp.2084>
- Wainwright, N. A. (2010). *The development of the Leeds Alliance in Supervision Scale (LASS): a brief sessional measure of the supervisory alliance*. University of Leeds.
- Wampold, B. E., & Owen, J. (2021). Therapist effects: History, methods, magnitude, and characteristics of effective therapists. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change, 7th ed* (pp. 301–330). Wiley.
- Zimmermann, D., Rubel, J., Page, A. C., & Lutz, W. (2017). Therapist effects on and predictors of non-consensual dropout in psychotherapy. *Clinical Psychology & Psychotherapy*, 24(2), 312–321. <https://doi.org/10.1002/cpp.2022>