



The
University
Of
Sheffield.

**Facilitative Interpersonal Skills (FIS) and Alliance: An examination of associated
variables**

Annie Faulds

A thesis submitted in partial fulfilment of the requirements for the award of
Doctor of Clinical Psychology

Clinical Psychology Unit
Department of Psychology
The University of Sheffield
May 2021

The results, discussions and conclusions presented herein are identical to those in the printed version. This electronic version of the thesis has been edited solely to ensure conformance with copyright legislation and all excisions are noted in the text. The final, awarded and examined version is available for consultation via the University Library.

Declaration

I declare that that the work has not been submitted for any other degree or to any other institution.

Word Counts

Section one: Literature Review

Word count excluding tables and references: 7,657

Word count including tables and references: 13,335

Section two: Empirical Study

Word count excluding tables and references: 7,934

Word count including tables and references: 11,996

Total word count

Excluding tables and references: 15,591

Including tables and references: 25,331

Lay Summary

Research has shown that there are several relationship factors between therapists and their clients which are important to psychotherapy working better for people. Research has examined these factors in more detail in the hope that this can inform therapist training and improve people's outcomes in psychotherapy.

Section one is a review of the existing research which has examined the influence of clients' interpersonal problems and interpersonal functioning on the therapeutic relationship, also termed the therapeutic alliance. Interpersonal problems are difficulties in interacting and communicating with others which can lead to distress for the person experiencing these difficulties. Nineteen studies were included, and the results suggested that interpersonal problems may influence the development of the therapeutic relationship. Some studies found this was more important for early relationship development. Other studies found it applied to specific interpersonal problems or specific parts of the therapeutic relationship. Quality of the studies was assessed, and most were deemed poor. The studies also used different approaches to examine this relationship. For these reasons, the findings are treated cautiously. Interpersonal problems may be important for therapists to consider when developing therapeutic relationships with their clients, but more research is needed to examine these factors.

Section two examined another relationship factor which has been found to be important for improving people's outcomes in psychotherapy, called Facilitative Interpersonal Skills (FIS). FIS are the interpersonal communications of therapists which help clients to address their problems in therapy. There are eight FIS which capture the therapists' ability to approach their client with warm, expressive, understanding, and collaborative communication. FIS also captures the therapists' ability to instil hope in the client and persuade them to consider things from different perspectives which might

help them with their problems. Therapists' FIS are rated by observers and traditionally this has been done using simulated video clips of clients presenting with difficulties, which therapists are asked to respond to. Little research has examined therapists' FIS in actual therapy sessions. The current study examined FIS in therapy sessions which were deemed to be "not on track" for achieving a positive outcome for the client. Therapy (Cognitive Behaviour Therapy; CBT or Person-Centred Experiential Therapy; PCET), therapist, and client variables were examined in relation to FIS. FIS was also examined in three sections of the session to consider if it varied over time. Results found no relationship between therapy, therapist and client variables and therapists' overall FIS score. However, there were relationships between specific FIS and the therapy model used by the therapist. Higher hope scores were found for CBT compared to higher emotional expression, warmth, and empathy scores for PCET. In CBT, the more a therapist aligned themselves with their therapy model the higher their FIS scores. FIS scores were also lower at the start of the therapy session. In conclusion, specific FIS may differ between therapy models and FIS may vary across a session. More research is needed to examine this further. Findings could have implications for therapist training and supervision.

Acknowledgements

I would like to thank my supervisors Professor Gillian Hardy and Professor Michael Barkham for your support throughout the project and the time you took to offer a space to think things through. Thank you to Dr Dave Saxon for your expertise, knowledge, and time throughout this process. I want to thank everyone who has assisted in the development of this project. Thank you to Assistant Professor Kim De Jong and Kane Steggles for your expertise and support in delivering FIS training. Merinda Hake for your time and expertise and Dr Mel Simmonds-Buckley for your assistance in identifying the participant sample.

A very special thank you to my thesis partner Charlotte Bentham. I am very grateful that we were able to share reflections and ideas together and I can honestly say I enjoyed the project, despite the hard work, and that is thanks to you.

Thank you to my wonderful cohort. I feel very lucky to have trained with such inspiring, supportive, and kind people. You have always made things feel more manageable even in the hard times. Finally, thank you to my Mum, Dad, and all my family and friends who have been there always, especially my partner Dawson who has ridden all the waves with me.

Table of Contents

Declaration	ii
Word Count	iii
Lay Summary	iv
Acknowledgements	vi
Section one: Literature Review	1
Abstract	2
Introduction	4
Methods	9
Results	14
Discussion	36
References	46
Appendix A: Search Terms	62
Appendix B: EPHPP Quality Assessment Tool	65
Appendix C: EPHPP Scoring Manual	68
Section two: Empirical Study	77
Abstract	78
Introduction	80
Methods	86
Results	97
Discussion	107
References	115
Appendix A: Outline of two FIS research projects and distinct aims	128
Appendix B-Ethical Approval Letter	129
Appendix C- PHQ-9	130

Appendix D-FIS Rating Scale (Anderson & Patterson, 2013)	131
Appendix E Problem Responsiveness FIS Domain Scale	140
Appendix F Beck Depression Inventory-II (BDI-II)	141
Appendix G Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM)	145
Appendix H- Connor-Davidson-Resilience Scale (CD-RISC; Connor & Davidson, 2003)	147
Appendix I- Client expectancy and credibility questionnaires	149
Appendix J- Session Adherence and Competence Scales for PCET and CBT	150
Appendix K- FIS Rating Calibration Meeting Minutes	152
Appendix L- FIS Distribution	159

Section one: Literature Review

The relationship between client interpersonal characteristics and the therapeutic alliance: A systematic review.

Abstract

Objectives The current systematic review aimed to summarise, and quality appraise the literature examining client interpersonal problems and therapeutic alliance in adult psychotherapy.

Methods A systematic search of Scopus, PsycInfo, PubMed, Web of Science was conducted in March 2021. Research papers examining the relationship between client interpersonal problems or functioning and therapeutic alliance from an adult individual psychotherapy sample were included. Results were narratively synthesised and quality appraised.

Results Nineteen studies were included in the final review. Most of the studies received a “weak” global quality assessment rating. There was a high level of heterogeneity between studies. Findings for the relationship between total interpersonal problems and therapeutic alliance were inconsistent. Several relationships between individual interpersonal problems and therapeutic alliance were identified. High hostility, coldness/detachment, and social inhibition were associated with poorer alliances in several studies. More affiliative or non-assertive problems were associated with higher alliances. Interpersonal problems were found to be more important for early alliance in several studies. Specific alliance components had stronger associations in some studies.

Conclusion The current review found some evidence for a relationship between interpersonal problems and therapeutic alliance. High heterogeneity and poor quality ratings limit the ability to adequately compare studies and infer a wider conclusion.

Practitioner Points

- Specific client interpersonal problems are associated with the therapeutic alliance, but further research is required to establish this finding.

- Client interpersonal problems may be more important for early alliance.

Limitations

- High level of heterogeneity between included studies limits comparability.
- Most studies did not account for nested data structures so findings should be treated cautiously.

Introduction

Therapeutic Alliance

The therapeutic alliance (Sterba, 1934; Zetzel, 1956) is an extensively researched concept within the field of psychotherapy (e.g., Norcross & Lambert, 2019). There are several definitions (e.g., Bordin, 1979; Gaston, 1990; Greenson, 1965; Frieswyk et al., 1984; Luborsky et al., 1983) but the therapeutic alliance is commonly viewed as a pan-theoretical construct (Bordin, 1979; Horvath, 2018) operationalised by collaboration, mutuality and engagement (Horvath & Symonds, 1991). Bordin's (1979) definition outlined three components: therapeutic bond, goal agreement and task agreement, which are understood as the foundation of forming supportive and trusting working relationships within the therapist-client dyad (e.g., Flückiger, Del Re et al., 2020; Norcross & Goldfried, 2019; Norcross & Lambert, 2019; Wampold & Imel, 2015). Gelso's (2014) tripartite model posits that there is a real relationship, the working alliance, and the transference. The working alliance is viewed as the part of the relationship where therapeutic technique and change takes place (Gelso, 2014).

Alliance and Outcome

The therapeutic relationship has long been considered a fundamental component of change (Rogers, 1957; Safran & Muran, 2000). An association between therapeutic alliance and clinical outcomes has been well established within the literature, with those who experience positive working alliances with their therapist demonstrating better treatment outcomes (e.g., Castonguay & Beutler, 2006; Castonguay et al., 2006; Crits-Cristoph et al., 2013; Del Re et al., 2012; Flückiger, Del Re et al., 2012; Horvath et al., 2011). Such an association has been demonstrated across therapeutic models and clinical problems (Castonguay et al., 2006). Flückiger et al.'s (2018) meta-analysis demonstrated an average effect size of $r = .28$ for the alliance-outcome association.

Psychotherapy research has investigated this in greater detail by examining alliance-outcome associations over time and across trajectories. Early alliance has been found to be a greater predictor of clinical outcome than alliance measured at other timepoints (e.g., Barber et al., 1999; Hilsenroth et al., 2004; Piper et al., 1991; Zilcha-Mano & Errázuriz, 2015). The early relationship is seen as key, particularly with the high proportion of clients who dropout after the first session (Connell et al., 2006). Horvath et al.'s (2011) meta-analysis indicated that a changing alliance trajectory was associated with better outcome. A U-shaped alliance trajectory has also been associated with improved outcomes as has linear growth in longer term therapies (De Roten et al., 2004; Kivlighan & Shaughnessy, 2000). A range of other alliance patterns have been identified and associated with improved outcomes (see Stiles et al., 2004; Constantino & Smith-Hansen, 2008). The alliance-outcome research is inconsistent, and the association includes much variability (Horvath et al., 2011). A debate has arisen within the field about whether symptomatic change or alliance comes first in the prediction of outcome (see Zilcha-Mano, 2017).

Factors Associated with the Therapeutic Alliance

With the therapeutic alliance seemingly a key ingredient for better clinical outcomes, research has investigated alliance as the dependent variable (Dinger et al., 2009), and examined the factors which influence its development (Baldwin et al., 2007; Connolly Gibbons et al., 2003; Castonguay et al., 2006), particularly as many studies adjust for intake characteristics (Flückiger, Del Re et al., 2020). Understanding the conditions and moderators which promote, or hinder alliance development is necessary for informing and improving clinical practice (Constantino, 2000; Constantino et al., 2005; Zilcha-Mano & Errázuriz, 2015) and developing adaptive interventions for specific clients (DeRubeis et al., 2014).

Client Characteristics

Client pre-treatment characteristics have provided the basis for much of this research, particularly as clients enter a relationship with their therapist with existing alliance building skills (DeRubeis et al., 2005). Horvath (1994) posited that client pre-treatment characteristics are key to understanding the alliance and subsequent therapeutic change. Zilcha-Mano (2017) posits that the alliance can be seen as having “trait-like” and “state-like” components. The former is described as the client’s ability to form relationships based on their internalised object relations and relationship expectations (Zilcha-Mano et al., 2014), whereas the latter is characterised as the alliance developing in action. Focusing on client characteristics has also been viewed as essential because of the evidence that patient-rated alliance and patient factors most strongly predict outcome (Ollila et al., 2016).

A range of client characteristics have been investigated with inconsistent results. Broadly speaking client demographic characteristics (Constantino et al., 2002) and client symptom severity (e.g., Flückiger, Del Re et al., 2020; Connolly Gibbons et al., 2003; Hersoug et al., 2002) have not, or are only marginally, associated with alliance development, however several studies have found evidence to the contrary. Flückiger, Del Re et al.’s (2020) meta-analytic comparison examining the alliance-outcome association found evidence to support the alliance as a distinct variable predicting treatment outcome, even when controlling for confounding client intake characteristics. Flückiger, Del Re et al. (2020) highlighted that within their selected studies interpersonal client characteristics were not systematically examined. Indeed, client interpersonal characteristics such as interpersonal problems (Horowitz et al., 1988) have demonstrated associations with therapeutic alliance within the extant literature, with more interpersonal problems leading to poorer commitment to the therapist and therapy (Andersen & Przybylinski, 2012). For example, Zilcha-Mano et al. (2014) found that patients’ pre-treatment representations of

others explained 32-54% of the alliance. Falkenström et al. (2013) found that personality difficulties moderated the alliance-outcome association, whereby alliance was more important in influencing outcome for those with personality difficulties.

Interpersonal Characteristics

A specific interpersonal characteristic is client interpersonal problems, understood within the framework of interpersonal theory (Puschner et al., 2005) whereby interpersonal communications or behaviours occur within a two-way relationship and these behaviours are experienced as either complementary or non-complementary (Horowitz, 2004). Non-complementary interactions are understood to increase distress and their continual pattern, or the deficit or excess of difficult interpersonal behaviours, are defined as interpersonal problems (Puschner et al., 2005; Gurtman & Lee, 2009). Interpersonal problems are conceptualised as deriving from internalized relationship schemas and have thus been implicated as a possible influence on the therapeutic dyad (Beretta et al., 2005), particularly as they are viewed as key reasons for people to seek psychotherapy (Horowitz et al., 1988; Horowitz et al., 2000). The interpersonal circumplex (IPC) orthogonal model (Leary, 1957; Horowitz et al., 2006) is one way of capturing interpersonal characteristics and examines interpersonal behaviours in terms of Communion/Affiliation (Connected/Close-Disconnected/Distant) and Dominance/Agency (Dominant-Submissive) and has been developed into several measures including the Inventory of Interpersonal Problems (IIP; Horowitz et al., 2000).

Interpersonal Characteristics and Outcome

Interpersonal characteristics feature within outcome research, with interpersonal functioning being seen as key in treating depression (Follette & Greenberg, 2006). High scores on the “communion” dimension of the IIP has been positively associated with symptom improvement (Filak et al., 1986; Schauenburg et al., 2000) whereas high

“agency” has been negatively associated with outcome (Borkovec et al., 2002), however these findings are not consistent within the literature (e.g., Ruiz et al., 2004).

Interpersonal Characteristics and Alliance

Researchers have theorised that those with greater interpersonal problems may develop poorer alliances, particularly those high in reactance (Beutler et al., 1991) or hostility (Safran & Muran, 1996). Gelso and Carter (1985) highlight several interpersonal capacities necessary for the formation of a successful working alliance, including the client’s ability to develop attachments, trust of others, and responsibility within the relationship. Client interpersonal difficulties are thus expressed within the therapeutic relationship (Henry et al., 1986; Kiesler, 1996; Safran & Muran, 2000) and those with interpersonal difficulties may demonstrate an attentional bias to therapists’ negative relational actions (Kiesler & Watkins, 1989). Indeed, Horvath and colleagues (Horvath & Symonds, 1991; Horvath & Greenberg, 1994) found moderate correlations between interpersonal characteristics and alliance.

Current Review

The current systematic review aimed to address Flückiger, Del Re et al.’s (2020) limitation concerning interpersonal characteristics and systematically examine interpersonal functioning or problems, and the therapeutic alliance. Flückiger, Del Re et al., (2020) focused on the alliance-outcome association, however for the current review interpersonal problems and the alliance were examined, to gain a detailed understanding of the interpersonal problems and alliance association, in the first instance.

Aims

The review had the following aims:

1. To examine the existing literature on interpersonal functioning or problems and their relationship to the therapeutic alliance.

2. To critically appraise the quality of included studies and highlight limitations to be addressed in future research.
3. To consider the clinical implications arising from the review and areas for future research

Method

Search Strategy

An initial scoping search to review the extent of existing literature and identify appropriate search terms took place in February 2021. In March 2021, a systematic search of Scopus, PsycInfo, PubMed and Web of Science was conducted, and the search strategy was registered on PROSPERO (CRD42021242445). Broad search terms were selected to ensure an optimal number of studies were selected. Search terms “psychotherapy”, “client characteristics”, “therapeutic alliance”, and their variations, were combined with Boolean operators. A comprehensive list of terms used across databases can be found in Appendix A. No constraints were placed on year of publication or subject type.

Pre-determined inclusion criteria based on the adapted PICOS framework (Methley et al. 2014; Table 1) were applied to titles and abstracts. The framework was adapted as there was no requirement for a comparator of interventions. A wide range of therapeutic interventions were considered in accordance with the alliance as a pan-theoretical construct (Bordin, 1979). Alliance was also considered at all timepoints during the therapy and as either a single-point measure or as a development trajectory to capture a broad understanding of the relationship between interpersonal problems and the alliance.

Search Procedure

The search was conducted adhering to the Preferred Reporting Items for Systematic and Meta-Analyses (PRISMA; Moher et al., 2009) framework, presented in Figure 1.

A search across all databases yielded 1,232 papers. Duplicates were removed electronically using EndNote software and by hand, following which, titles and abstracts

Table 1

Inclusion and exclusion criteria using PICOS Framework

PICOS Domain	Inclusion	Exclusion
Population	Adults (aged 18+) receiving individual psychotherapy	Children and adolescents. Adults not receiving a form of psychotherapy.
	Participant sample size of ≥ 10	Participant sample size of < 10
	Therapist sample size of ≥ 5	Therapist sample size of < 5
Intervention	Individual psychotherapy defined as that which adheres to psychological theory or principles and is delivered by trained or trainee therapists	Group psychotherapy, couples' psychotherapy, psychotherapy not adhering to psychological principles or theory, psychotherapy which is not defined
Comparator	All quantitative study designs (including cohort) regardless of having a comparator arm or not	No exclusions based on presence or absence of comparator arm
Outcome	Therapeutic alliance measured by a validated alliance scale and rated by client and/or therapist and/or observer prospectively during psychotherapy. Alliance measured at all	Therapeutic alliance not measured by a validated scale, therapeutic alliance ratings applied retrospectively to psychotherapy.

	timepoints and as a trajectory.	
	AND	Studies which did not examine client interpersonal variables.
	Client interpersonal functioning/style or interpersonal problems	Studies which did not examine the relationship between client interpersonal variables and therapeutic alliance.
Study design	Empirical research papers with quantitative data (e.g., Randomised controlled trials, RCTs; Controlled Clinical Trials, CCTs; Cohort studies)	Qualitative studies, editorials, reviews, book chapters, measure validations, position papers
	Papers written in the English language or where English translations were available	Papers not written in English
	Published and unpublished research	

were screened against the inclusion criteria. Seventy-eight full texts were reviewed, and 68 papers were excluded. The main reason for exclusion was that the paper did not examine client interpersonal functioning or interpersonal problems. Further handsearching was conducted to capture as many papers as possible. Handsearching included a backward citation search of selected papers and relevant reviews' reference lists, a forward citation search of selected papers, and searching on grey literature databases (OpenGrey, ProQuest, ETHOS). Handsearching identified a further nine papers for inclusion, resulting in a total of 19 included studies. Twenty percent of the full text sample were screened by a

second researcher (CB) achieving 92.31 % inter-rater agreement, and papers were discussed until consensus was reached.

Quality Assessment

The Effective Public Health Practice Project (EPHPP) Quality Assessment Tool for Quantitative Studies (Thomas et al., 2004) was selected to critically appraise included papers (Appendix B). The EPHPP tool was selected for its specific purpose of assessing the evidence base for systematic reviews and because it can be used on a range of quantitative study methodologies. Construct validity and inter-rater reliability have been established (Thomas et al. 2004). Eight assessment domains are included in the EPHPP tool (selection bias, study design, confounders, blinding, data collection methods, withdrawals and dropouts, intervention integrity, analysis). The first six domains are rated as “good”, “fair” or “poor” based on the scoring manual (Appendix C) and domains seven and eight are not assigned a score. A global rating (strong, moderate, weak) is then assigned to the study based on the domain scores. “Strong” is indicated when the study has no “poor” domain scores, “moderate” is indicated when there is no more than one “poor” domain score, and “weak” is assigned when there are two or more “poor” scores. All papers were included regardless of their quality rating to gain a thorough understanding of the state of the current literature.

To increase the robustness of the quality assessment process a second rater (CB) critically appraised 20% of the papers. Quality assessment was compared between raters and any disagreements were discussed and resolved. An intraclass correlation coefficient was calculated based on a two-way mixed effects, average rater, absolute agreement model (Koo & Li, 2016).

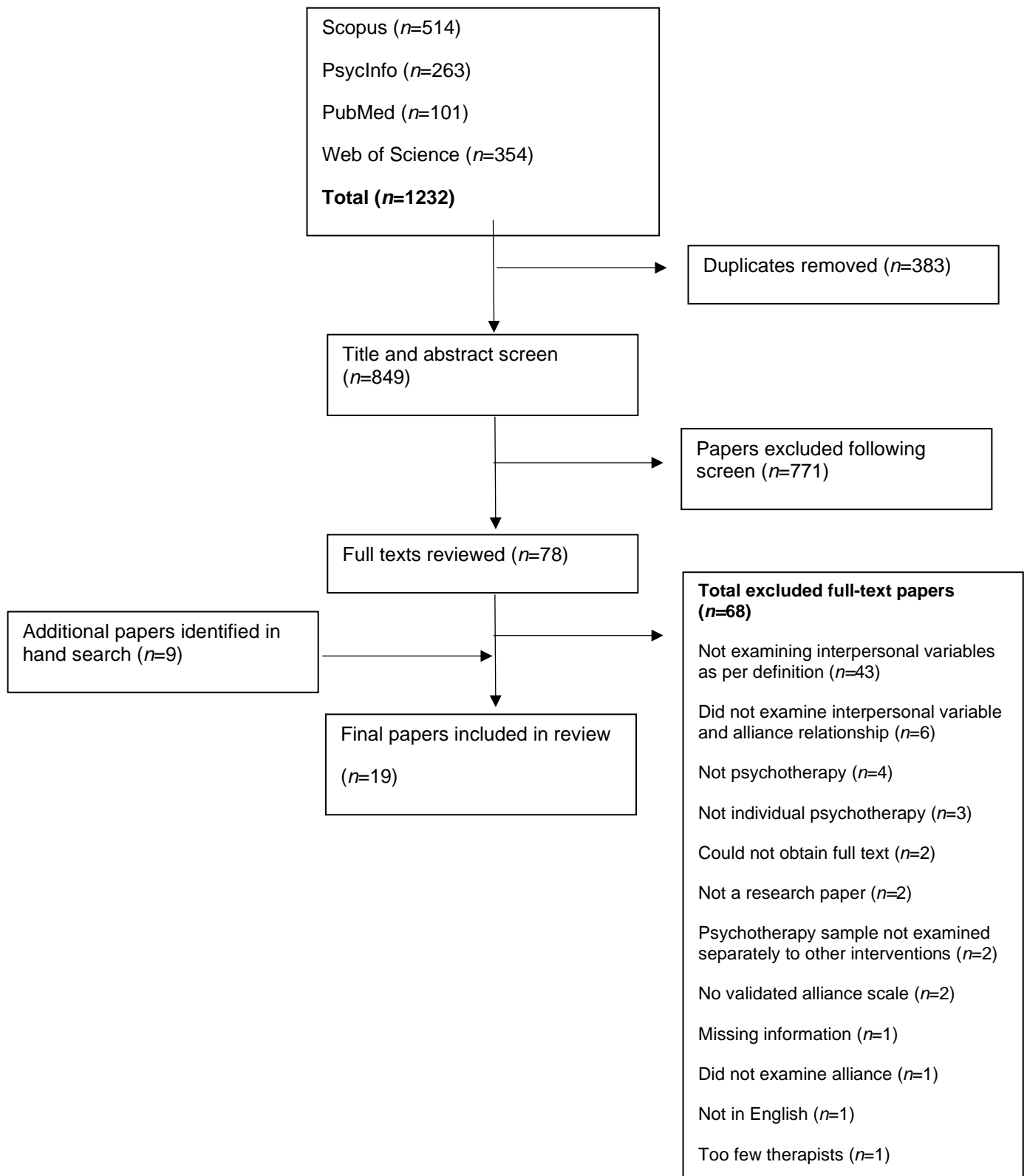
Data Synthesis

There was a high degree of heterogeneity between studies in terms of design, participant sample, psychotherapy models, length of therapy, the timepoint alliance was

measured, the rater of alliance, and the interpersonal variable assessed. A narrative synthesis was, therefore, undertaken as opposed to a meta-analytic approach in line with Higgins and Green's (2008) recommendations regarding appropriate conditions for proceeding with a meta-analysis.

Figure 1

PRISMA diagram of systematic search



Results

Data extraction

An overview of the study characteristics and key findings of the 19 included studies is presented in Table 2. Twenty percent of papers had data extracted by a second researcher (CB, the paired project), achieving 97.22% inter-rater agreement. Constantino et al. (2005) and Constantino and Smith-Hansen (2008) are included as separate studies but used data drawn from the same dataset. Likewise, the three papers by Hersoug and colleagues used the same dataset. Studies were examined separately as they examined the relationship between interpersonal problems and alliance in different ways, so their findings and study characteristics are reported separately.

Study Design

The majority ($n=11$) of included papers used a prospective cohort design whereby a single participant group was examined pre-and post-intervention. Four of the included studies were randomised-controlled trials (RCTs) where a formal process of participant randomisation was utilised. Four included studies were controlled-clinical trials where no formal randomisation process occurred, but two treatment arms were compared.

Sample

Studies demonstrated a range of participant sample sizes. Six studies had smaller samples ranging from 30-66 participants. Six studies had sample sizes ranging from 128-144 and seven studies had larger sample sizes ranging from 201-714 participants. Eight studies recruited participants diagnostically (Major Depressive Disorder ($n=3$); anxiety and mood disorders ($n=2$); schizophrenia/schizoaffective disorder ($n=1$); and bulimia nervosa ($n=2$)). The eleven remaining studies were characterised as a generic adult psychotherapy population.

Therapists and Therapy

Therapist sample sizes ranged from 8-471 therapists. Most studies utilised trained therapists ($n=16$), one study used a mix of trained and trainee therapists, and two studies used a solely trainee cohort. Six studies compared Cognitive-Behaviour Therapy (CBT) to another therapeutic model (Interpersonal Psychodynamic Therapy, Supportive Expressive Psychotherapy, Process-Experiential Therapy, Psychoeducation). Five studies included a mix of therapy modalities, four of which stated that psychodynamic was the dominant orientation. Other therapeutic modalities included psychodynamic therapy ($n=2$), Brief Psychodynamic Therapy ($n=1$), cognitive therapy ($n=2$), experiential therapy ($n=2$), and counselling ($n=2$).

Measures

A range of therapeutic alliance measures were utilised in included studies. Most studies ($n=9$) utilised the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989) or the WAI-Short Form (WAI-S; Tracey & Kokotovic, 1989). Four utilised the Helping Alliance Questionnaire (HAQ; Alexander & Luborsky, 1986). Two utilised the California Alliance Psychotherapy Scales (CALPAS; Gaston et al., 1991). One utilised the Bern Post-Session Report (Flückiger et al., 2010). One used the Therapeutic Bond Scales-Revised (Saunders, 1999). Most studies used client-rated versions of the measure ($n=12$) whereas some used both client-rated and therapist-rated versions ($n=7$). A range of alliance data collection points were used across studies ranging from single-point data collection to sessional collection.

Interpersonal functioning or interpersonal problems were assessed using a range of measures and several studies used multiple measures. The most common measure was the IIP (Horowitz et al., 2000) or a version of it ($n=15$). Other measures of interpersonal problems were not shared between studies (see Table 2).

Table 2*Summary of included studies*

Paper	Country	Design	Participants & Therapists	Therapy model(s)	Alliance Measure, Rater, and Alliance Timepoint	Interpersonal client characteristic (Measure)	Key findings
Altenstein-Yamanaka et al. (2017)	Switzerland	Randomised-controlled trial (RCT)	144 adults meeting criteria for Major Depressive Disorder (MDD) 25 trainee therapists	CBT & Exposure-based CBT (ECBT)	Bern Post-Session Report-Patient Form Client-rated Mean score of first 5 sessions	Interpersonal Problems (Inventory of Interpersonal Problems; IIP, Impact Message Inventory; IMI)	Positive correlation between IIP (communion) and early alliance retained in predictor model No significant association between IMI (communion) and early alliance
Beretta et al. (2005)	Switzerland	Prospective cohort	60 adults seeking treatment for anxiety and mood disorders 10 therapists (1 psychotherapist, 9 psychiatrist-psychotherapists)	Brief Psychodynamic Intervention (BPI)	Helping Alliance Questionnaire (HAQ) Client-rated	Interpersonal Problems (IIP, Conflictual Relationship Themes; CCRT)	Alliance associated with client's wish to be close and see others as trustworthy and helpful Low and stable alliances associated with the wish to feel close to others yet perceive others responses negatively

					Sessional		Low and stable alliances have more interpersonal problems related to low affiliation capacity.
Connolly Gibbons et al. (2003)	United States	CCT	141 adults 36 trained therapists (with masters, doctoral, or medical degree)	Supportive Expressive Psychotherapy (SE) or Cognitive therapy	CALPAS Client-rated	Interpersonal Problems (IIP)	Session 10 patient-rated alliance predicted by pre-treatment interpersonal distress Hostile-dominant interpersonal problems associated with poor alliance (in an exploratory analysis)
Constantino et al. (2005).	United States	RCT	220 adults meeting criteria for Bulimia Nervosa (BN) 8 therapists (7 doctoral-level psychologists, 1 psychiatrist)	CBT for BN; Interpersonal Psychotherapy for BN (IPT)	HAQ Client-rated	Interpersonal Problems (IIP)	No association between interpersonal problems and early alliance Degree of interpersonal problems were only associated with middle alliance development in IPT
Constantino & Smith Hansen (2008)	United States	RCT	207 adults meeting criteria for Bulimia Nervosa (BN)	CBT for BN; Interpersonal Psychotherapy for BN (IPT)	HAQ Client-rated	Interpersonal Problems (IIP)	High interpersonal distress associated with low early and middle alliance

8 therapists (7 doctoral-level psychologists, 1 psychiatrist)

Sessions 4, 12 and 18

More affiliative problems positively associated with higher early and middle alliance

Interaction between alliance within treatment groups (CBT vs IPT) and rigidity, control and affiliation

Three alliance trajectories identified.

Higher interpersonal distress and hostile submissiveness in Group 2 (low initial alliance significant growth over time) than Group 3 (high initial alliance and significant growth over time)

Couture et al. (2006)

United States

Controlled clinical trial (CCT)

30 adults with a diagnosis of schizophrenia/schizoaffective disorder

Cognitive-behavioural intervention or Psychoeducation

Working Alliance Inventory (WAI)

Social Functioning (Social Functioning Scale; SFS)

No association between client predictors and client-rated alliance

9 trained therapists with experience of CBT

Higher therapist-rated alliance associated with higher social functioning

					Client and therapist-rated		(accounting for 16% of variance)
					WAI collection timepoint not stated		
Gaston et al. (1988)	United States	CCT	60 older adults meeting criteria for MDD	Behavioural or Cognitive or Brief Dynamic	California Psychotherapy Alliance Scales (CALPAS)	Interpersonal functioning (Young Loneliness Inventory)	No association between pre-treatment interpersonal functioning and in-treatment patient contribution to alliance
			10 doctoral psychologists with at least 1 year post-doctoral training in specialist modality		Client and therapist-rated		
					Session 5, 10 and 15		
Hersoug et al. (2002)	Norway	Prospective cohort	270 adult psychotherapy clinic patients	Mixed (mostly psychodynamic orientation/training)	WAI	Interpersonal problems and relationship functioning (Dynamic Scales, DS Dynamic Relations; IIP; Quantity and Quality of Social Relations, QQSR)	IIP (cold) negatively associated with patient and therapist-rated alliance
			59 trained therapists (psychologists, psychiatrists, social workers, nurses)		Client and therapist-rated		6 relationship variables (current and past) associated with early alliance and 4 with later alliance
					Session 3 and 12		

							Current relationships were more predictive of alliance than past relationships
							A positive predictor relationship was found for QQSR and Dynamic relations and alliance
Hersoug et al. (2009)	Norway	Prospective cohort	201 adult psychotherapy clinic patients	Mixed (mostly psychodynamic orientation/training)	WAI	Interpersonal Problems (IIP-64)	IIP and PFS were not associated with WAI-T
			61 therapists (psychologists, psychiatrists, social workers, nurses)		Client and therapist-rated	Interpersonal Functioning (Psychodynamic Functioning Scale, Interpersonal-PFS)	Lower WAI-P scores associated with higher IIP (cold/detached)
					Session 3, 12, 20 and every successive 20 th session		Better WAI-P related to higher interpersonal functioning
							IIP (cold/detached) was related to alliance growth which improved over time and by session 20 IIP was no longer related
Hersoug et al. (2010)	Norway	Prospective cohort	201 adult psychotherapy clinic patients	Mixed (mostly psychodynamic orientation/training)	WAI	Interpersonal Problems (IIP-64)	No significant difference between the stable, improving or deteriorating alliance groups and interpersonal problems
			61 therapists (psychologists, psychiatrists, social workers, nurses)		Client-rated	Interpersonal Functioning (Psychodynamic	

					Session 3, 12, 20 and every successive 20 th session	Functioning Scale, Interpersonal-PFS)	
Kalogerakos (2009)	Canada	CCT (doctoral thesis)	66 adults diagnosed with MDD	Process-experiential therapy (PET) or CBT	WAI-Short Form (WAI-S)	Interpersonal Problems (IIP)	No significant difference between two alliance groups (Stable Linear vs Steep Linear Quadratic) and total IIP score
			15 trained therapists		Client-rated		
					Sessional		Higher "cold/distant" scores in Steep Linear Quadratic group (accounting for 11% of variance in alliance shape patterns)
							Overall baseline interpersonal problems predicted overall alliance strength (accounting for 8% of variance)
							Higher "social inhibition", "non-assertiveness" and "overly-accommodating" scores associated with low-mid-strength alliance shape and level compared to high-strength
							Total baseline interpersonal problems

							nor subscales were not associated with "task" or "agreement on goals" WAI subscale
							Total baseline interpersonal problems and "domineering/controlling" and "socially inhibited" subscales predicted average "bond" strength
Kokotovic & Tracey (1990)	United States	Prospective cohort	144 adults	Counselling	WAI	Interpersonal Relationships (Interpersonal Relationship Scale, IRS)	More hostile and poorer relationships were significantly associated with poorer alliance
			15 therapists (13 doctoral-level psychologists, 2 interns)		Client and therapist-rated		
					First session		
Krieg & Tracey (2016)	United States	Prospective cohort	144 adults	Counselling	WAI-S	Interpersonal Problems (IIP-C-Item Response Theory, IIP-C-IRT)	No association between IIP total and alliance
			44 trainee counsellors		Client-rated		"Vindictive/self centred" subscale associated with lower initial alliance
					Third session		Interaction between "dominance" and therapist gender on the WAI-S was found (higher dominance associated

with higher initial working alliance with male counsellors; lower dominance associated with higher initial working alliance with female counsellors)

Muran et al. (1994)	Canada	Prospective cohort	32 adults	Cognitive therapy	WAI	Interpersonal Problems (IIP)	Positive association between "hard to be assertive", "overly nurturant", "exploitable", and total alliance, "task agreement", and "goal agreement"
			11 trained therapists		Client-rated		Positive association between "non-assertive" and "task" and "goal agreement"
					Third session		Positive association between "socially avoidant" and "task agreement"
							No association with "hostile-dominant" subscales and WAI subscales

Ollila et al. (2016)	Finland	Prospective cohort	128 adults experiencing mood/anxiety disorders 41 trained therapists (82.5% psychologists, 5% psychiatrists, 12.5% other professions)	Psychodynamic	WAI Client and therapist-rated 7, 12 and 36 months from start of therapy	Interpersonal Problems (IIP)	Higher total IIP showed statistically significant improvement in patient-rated WAI by the end of therapy IIP subscale scores showed differing alliance slopes Therapist-rated WAI was more improved at end of therapy with moderate to high IIP clients and showed a faster improvement for high IIP compared to patient-rated WAI (subscales showed differential improvement trends)
Paivio & Bahr (1998)	Canada	Prospective cohort	33 adults 8 trained therapists	Experiential psychotherapy with empty-chair dialogue	WAI Client-rated Session 3 and last session (usually session 12)	Interpersonal Problems (IIP)	IIP significantly related to early "bond" development Positive association between "affiliation" and "bond" development early in therapy Negative association between "social avoidance" and total alliance, "goal", and

							"bond development throughout therapy
							Moderate negative association between "non-assertiveness" and "bond" development late in therapy
							Higher "hostility" associated with weaker "bond"
Puschner et al. (2005)	Germany	Prospective cohort	714 adults receiving psychotherapy in private healthcare company 471 trained therapists	Mixed (51.7% psychodynamic; 31.8% CBT; 16.4% psychoanalytic)	HAQ Client and therapist-rated HAQ collection timepoint not stated	Interpersonal Problems (IIP)	Main effect between the "hostile" and "friendly" dimensions on IIP and all four dimensions of the HAQ (higher hostility negatively associated with patient and therapist HAQ scores prior to/in early therapy sessions) HAQ dimensions not associated with "dominance" or "affiliation x dominance" Along the "affiliation-dominance" dimension there was an association with mean location of interpersonal distress and HAQ at early stages

							Positive correlation between patient and therapist-rated alliance and "affiliation"
							No association between early IIP scores and HAQ at 18 months follow-up (except, higher "hostility" or "dominance" associated with poorer alliance)
Renner et al. (2012)	United States	RCT	523 adult outpatients	Cognitive therapy	WAI	Interpersonal Problems (Interpersonal Problems-Circumplex; IIP-C)	Negative correlation between WAI and "distress" and "agency"
			15 trained therapists		Client-rated		Positive correlation between WAI and "communion"
					Middle and end of acute phase (12-14 weeks)		Pattern evident for all subscales except general distress (IIP-C) and task (WAI-C)
							Total IIP-C variance for WAI-C was small and was not changed by the addition of baseline severity

Saunders (2001)	United States	Prospective cohort	141 adults deemed appropriate for long-term psychodynamic psychotherapy	Psychodynamic	Therapeutic Bond Scales- Revised (TBS- R)	Interpersonal Problems (IIP)	Negative association between "too detached" and "global bond"
			20 therapists		Client-rated		Negative association between "too detached" and "too self-effacing" and "empathic resonance"
					Third session		Negative association between "too detached" and "mutual affirmation"
							No associations between IIP and "role investments"

Quality Assessment

A summary of the EPHPP (Thomas et al., 2004) quality assessment is presented in Table 3. Twenty percent ($n=4$) of papers were quality assessed by a second rater. A two-way mixed effects, average-rater ($n=2$), absolute agreement ICC of .96 (95% CI: .92-.98) was reached, suggesting "excellent" inter-rater reliability (Koo & Li, 2016). Many of the included studies used datasets derived from other trials, so the original trial paper was also considered to maximise the available information for assessment. Most of the studies ($n=11$) received a global rating of "weak" due to receiving two or more "poor" ratings at the domain level. Five studies received a global rating of "moderate" only receiving a "poor" rating in one domain. Only three studies received a global rating of "strong" receiving no "poor" ratings

Table 3*Summary of EPHPP Quality Assessment*

Paper	EPHPP Domain						Global Rating
	Selection Bias	Study Design	Confounders	Blinding	Data collection methods	Withdrawals and dropouts	
Altenstein-Yamanaka et al. (2017)	Poor	Good	Good	Poor	Good	Good	Weak
Beretta et al. (2005)	Poor	Fair	Poor	Poor	Good	Poor	Weak
Constantino et al. (2005).	Good	Good	Poor	Fair	Good	Fair	Strong
Constantino & Smith Hansen (2008)	Good	Good	Good	Fair	Good	Fair	Moderate
Couture et al. (2006)	Fair	Good	Good	Fair	Good	Poor	Strong
Gaston et al. (1988)	Fair	Fair	Good	Poor	Good	Poor	Weak
Connolly Gibbons et al. (2003)	Fair	Fair	Good	Poor	Good	Poor	Weak
Hersoug et al. (2002)	Fair	Fair	Poor	Poor	Good	Poor	Weak
Hersoug et al. (2009)	Fair	Fair	Poor	Poor	Good	Poor	Weak
Hersoug et al. (2010)	Fair	Fair	Poor	Poor	Good	Poor	Weak
Kalogerakos (2009)	Good	Good	Good	Poor	Good	Fair	Moderate
Kokotovic & Tracey (1990)	Fair	Fair	Good	Poor	Good	Poor	Moderate
Krieg & Tracey (2016)	Fair	Fair	Good	Poor	Poor	Poor	Weak
Muran et al. (1994)	Fair	Fair	Good	Fair	Good	Poor	Moderate
Ollila et al. (2016)	Fair	Good	Good	Poor	Good	Good	Moderate
Paivio & Bahr (1998)	Poor	Fair	Poor	Poor	Good	Poor	Weak
Puschner et al. (2005)	Poor	Fair	Good	Poor	Good	Poor	Weak
Renner et al. (2012)	Fair	Good	Fair	Fair	Good	Poor	Strong
Saunders (2001)	Fair	Fair	Good	Poor	Good	Poor	Weak

at the domain level.

Selection Bias

Most studies achieved “good” or “fair”. Studies scored “fair” largely due to selecting samples from clinic settings and not extending study advertisement to the general population. Four achieved “poor” ratings for selection bias. Altenstein-Yamanakada et al. (2017) selected from a broad sample yet less than 60% of those selected agreed to participate. Beretta et al (2005) and Paivio and Bahr (1998) did not provide information on their study selection. Puschner et al. (2005) selected from a private healthcare company meaning the generalisability of the sample to the wider population was limited.

Study Design

All studies achieved “good” or “fair”. “Good” was achieved for the seven studies which utilised either an RCT or CCT design and “fair” was applied to the remaining studies that utilised a prospective cohort design. One study was a doctoral dissertation.

Confounders

To account for the proportion of cohort studies this domain was applied not only to inspect differences between experimental groups but also between completers and non-completers. Most studies achieved “good” or “fair”. A “fair” rating was applied to Renner et al. (2012) as they did not outline how recruitment site differences were accounted for within the study. Five studies (Beretta et al., 2005; Hersoug et al., 2002; Hersoug et al., 2009; Hersoug et al., 2010; Paivio & Bhar, 1998) did not provide information to enable an accurate assessment of potential confounding. Although, in Constantino and Smith-Hansen’s (2008) study several confounders were addressed in the original trial, through using an independent quality controller to manage site differences and group assignment being stratified by anorexia nervosa, their study did not account for potential group or site differences within the analysis.

Blinding

Overall blinding of outcome assessors or participants was poorly reported across studies. No studies reported on whether participants were aware of the research question. Where “fair” was achieved, studies had used independent outcome assessors or ensured assessors were blind to treatment condition.

Data Collection Methods

All except one study achieved “good” for their use of data collection tools. Studies included tools which were both valid and reliable. Gaston et al. (1986) utilised the Young Loneliness Inventory. Validity and reliability were not reported, and it was not possible to establish this through use of references and searching.

Withdrawals and Dropouts

Only two studies achieved “good” in this domain. Three studies were scored as “fair” because numbers and reasons for withdrawal or dropout were reported but the percentage of completers was between 60-79%. Fourteen studies did not report on the numbers and reasons for withdrawal and dropout and one study provided no information at all (Saunders et al., 2001), receiving “poor” ratings.

Intervention Integrity

Five studies did not report adequately on the numbers receiving the allocated intervention (Connolly Gibbons et al., 2003; Paivio & Bahr, 1998; Puschner et al., 2005; Renner et al., 2012; Saunders, 2001). The remaining studies achieved between 60-100% of participants receiving the intervention. Only seven studies measured the consistency of the intervention delivered (Altenstein-Yamanaka et al., 2017; Constantino et al., 2005, Constantino & Smith-Hansen, 2008; Couture et al., 2006; Gaston et al., 1988; Kalogerakos, 2009; Renner et al., 2005), with the remaining studies either not measuring consistency or not reporting on this adequately. Most studies did not report on how the risk of intervention contamination was considered or monitored. Altenstein-Yamakada et al.

(2017), Constantino et al. (2005), and Constantino and Smith-Hansen (2008) all reported on methods of maintaining intervention integrity.

Analyses

Most studies utilised an appropriate statistical method. Six studies completed analysis based on intervention allocation (i.e. intention to treat analysis; Altenstein-Yamanaka et al., 2017; Constantino et al., 2005, Constantino & Smith-Hansen, 2008; Couture et al., 2006; Ollila et al., 2016; Renner et al., 2005). Kalogerakos (2009) and Puschner et al (2005) utilised ANOVAs when independence of observations may have been violated due to the risk that the same therapist saw multiple clients.

Findings

Overall Interpersonal Problems and Alliance

Studies presented mixed findings for the relationship between overall interpersonal problems and functioning and the alliance. Kalogerakos (2009) found that higher total IIP predicted lower total alliance and “bond” but no relationship was found for “task” and “goals”. Pavo and Bahr (1998) found that higher total IIP was associated with poorer early “bond” development. Contrastingly, Ollila et al. (2016) found that higher total pre-treatment IIP predicted greater improvement in client-rated alliance scores by the end of therapy and high to moderate IIP predicted greater improvement in therapist-rated alliance scores by the end of therapy. Several studies which utilised the IIP found no association between total IIP score and alliance (Krieg & Tracey, 2016), early alliance (Constantino et al., 2005), and “role investments” on the alliance measure (Saunders, 2001).

On other measures, Gaston et al. (1988) found no association between interpersonal functioning and alliance. Couture et al. (2006) and Hersoug et al. (2009) found no association between client interpersonal predictors and client-rated alliance and therapist-rated alliance, respectively.

Subscales or Specific Interpersonal Problems and Alliance

Several studies found relationships between specific interpersonal difficulties, or subscales of the IIP, and alliance. For a description, the circumplex of interpersonal problems (IIP-C) is presented diagrammatically in Figure 2 and descriptions of each octant are presented in Table 4 summarised from Horowitz et al. (2000).

Figure 2

Circumplex of interpersonal problems (IIP-C)

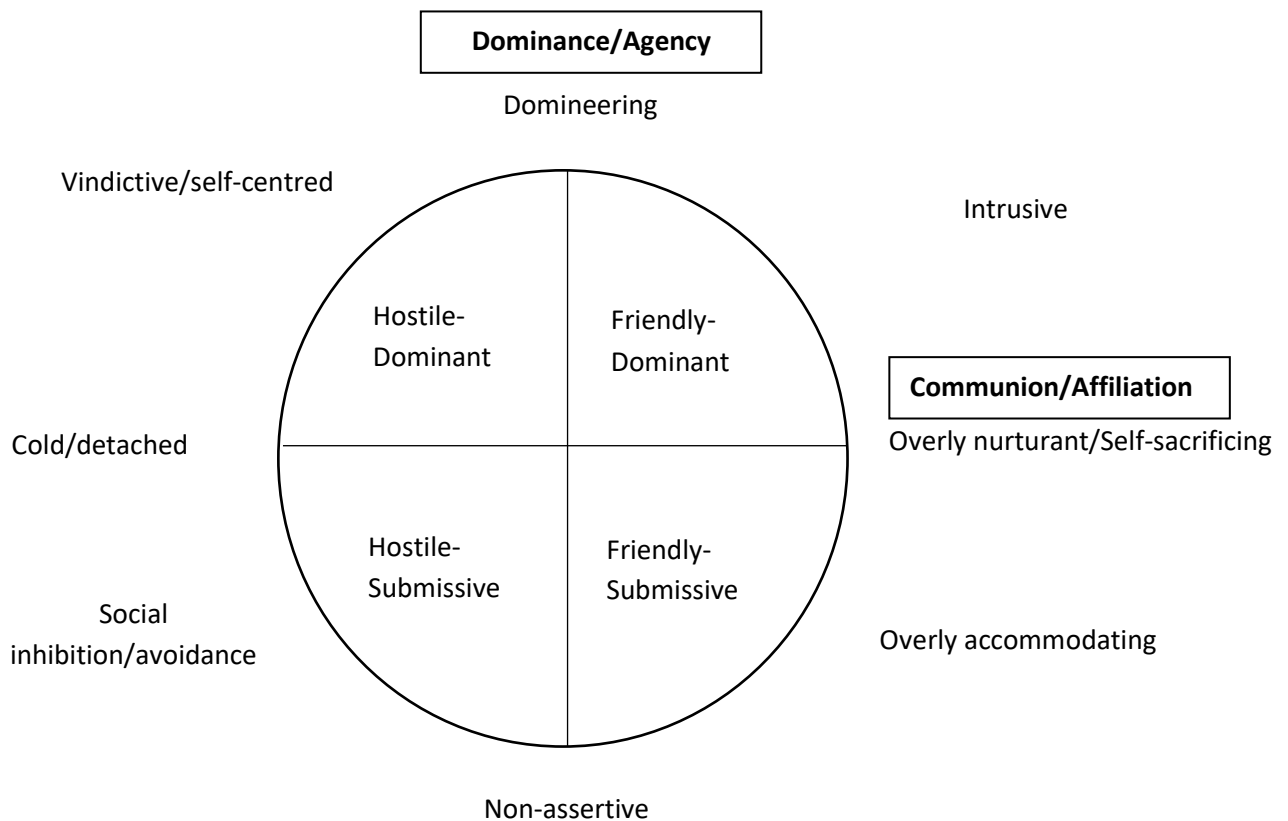


Table 4*Description of IIP-C octants*

IIP Octant	Description
Domineering	Being too aggressive
Intrusive	Seeking attention inappropriately
Overly nurturant/self-sacrificing	Being too eager to please others
Overly accommodating	Being too trusting and permissive
Non-assertive	Failing to be forceful
Social inhibition/avoidance	Being socially anxious or shy
Cold/detached	Having difficulty with affection or sympathy
Vindictive/self-centred	Being suspicious and distrustful

For the “communion/affiliation” dimension several relationships were identified. A positive association between overly affiliative interpersonal problems and alliance was found by Altenstein-Yamanaka et al. (2017), Renner et al. (2012), Puschner et al. (2005), and Constantino and Smith-Hansen (2008). This was identified for early and middle alliance by Constantino and Smith-Hansen (2008) and client and therapist-rated alliance by Puschner et al. (2005). Paivio and Bahr (1998) also found a positive association between “communion/affiliation” and early therapeutic “bond”. Therefore, those with more affiliative problems demonstrated higher alliances. However, Altenstein-Yamanaka et al. (2017) did not find this for the “communion/affiliation” dimension of the IMI which examined significant-other reports on interpersonal functioning.

Several studies identified significant relationships between alliance and the “non-assertive” pole or “friendly-submissive” quartile. Non-assertiveness was associated with higher total alliance, “task” and “goal” agreement (Muran et al., 1994) and negatively associated with “bond” (Paivio & Bahr, 1998). Higher scores on “overly nurturant” or

“exploitable” subscales were associated with higher total alliance, “task” and “goal” agreement (Muran et al., 1994). Saunders (2001) found those who were “too self-effacing” had higher scores on the “empathic resonance” alliance subscale. Renner et al. (2012) also found a negative association with alliance score and “dominance/agency”. Therefore, those characterised as “friendly-submissive” (i.e. eager to please others, being too trusting, and non-assertive) demonstrated higher alliances.

Several studies found relationships between the “hostile-dominant” quartile of the IIP and alliance. Many found that the more “hostile-dominant” interpersonal problems the poorer the alliance (Connolly Gibbons et al., 2003; Kalogerakos, 2009; Kokotovic & Tracey, 1990; Krieg & Tracey, 2016; Paivio & Bahr, 1998; Puschner et al., 2005). Kalogerakos (2009) found this relationship for overall alliance and for the “bond” subscale but not for “task” and “goals”. Conversely, Puschner et al. (2005) found this relationship for “hostility” specifically, but not for the overall “dominance” dimension and Muran et al. (1994) found no association with the “hostile-dominant” subscales and alliance. The relationship between hostile-dominance and alliance is therefore inconsistent.

Three studies identified relationships between the “cold/detached” dimension and alliance. Higher “cold/detached” scores were associated with poorer alliance (Hersoug et al., 2002; Hersoug et al., 2009; Saunders, 2001). Saunders (2001) found this relationship for “global bond” and the “mutual affirmation” and “empathic resonance” alliance subscales. Therefore, those who have difficulties with “coldness/detachment” (i.e. problems with affection and sympathy) demonstrated poorer alliances.

Three studies identified relationships between the “social inhibition/ avoidance” subscales and alliance. Paivio and Bahr (1998) found a negative association between “social inhibition/avoidance” and total alliance, “goal” and “bond”, and Kalogerakos (2009) found the same relationship with “bond”. Muran et al. (1994) found a positive association between “social avoidance” and “task agreement”. Using the SFS, Couture et al. (2006)

found higher social functioning was associated with higher therapist-rated alliance. “Social inhibition/avoidance” subscales were found to influence separate components of the alliance differentially.

IIP distress scores were also examined by some of the studies. Constantino and Smith-Hansen (2008) found that higher distress was associated with lower early and middle alliance scores. Renner et al. (2012) found a negative association between alliance scores and distress, however this pattern was not evident for the “task” alliance subscale. Higher distress was therefore associated with lower alliance, but this was not present for all components.

Hersoug et al. (2002) and Kokotovic and Tracey (1990) examined specific relationship variables and the development of the therapeutic alliance. Hersoug et al. (2002) found that current relationships were more predictive of alliance than past relationships with current relationships explaining 7% of early alliance. Dynamic relations were also associated with alliance. Kokotovic and Tracey (1990) found that poorer quality of current and past relationships was associated with poorer alliance. Therefore, quality of current and past relationships was associated with the alliance but how this influenced the alliance differed between studies.

Interpersonal Problems and Alliance Timepoints

Some studies examined the relationship between interpersonal problems or functioning and alliance at different timepoints in therapy. Constantino and Smith-Hansen (2008) found IIP subscale relationships (distress, affiliation) for both early and middle alliance. Early “bond” development was also related to interpersonal problems (IIP “affiliation”; Paivio & Bahr, 1998). Early alliance was also negatively associated with “hostility” on the IIP (Puschner et al., 2005). Puschner et al. (2005) found that pre-treatment IIP scores were not associated with alliance scores at 18-month follow up as they had been earlier in therapy, apart from those with high “hostility” and “dominance”

who had poorer alliance scores. Similarly, Hersoug et al. (2009) found that “cold/detached” IIP was associated with earlier alliance growth but this was not sustained by session twenty. Hersoug et al. (2002) also found six current and past relationship variables were associated with early alliance compared to four relationship variables for later alliance. The variance in alliance explained by current relationships was 7% for early alliance but this was not found for later alliance. Conversely, Constantino et al. (2005) did not find a relationship between early alliance and IIP but did find that middle alliance development was associated with IIP, however this was specific to only one treatment arm (IPT). Connolly Gibbons et al. (2003) also found a negative association for middle (session 10) patient-rated alliance and interpersonal distress. Most studies found a stronger association between interpersonal problems and early alliance, however this was not consistent across all studies.

Interpersonal Problems and Alliance Patterns

Five studies examined alliance development patterns and investigated relationships with interpersonal problems. Kalogerakos (2009) identified “stable linear” and “steep linear quadratic” groups and found no between-group differences on overall IIP scores, however the “steep linear quadratic” group (initial below average alliance and a steep growth curve over treatment) had a significantly higher score on the “cold/distant” subscale accounting for 11% of the variance in alliance patterns. Hersoug et al. (2010) also found no between-group differences (“stable”, “improving”, “deteriorating”) on overall IIP scores.

Beretta et al. (2005) found that the “low and stable” alliance group differed from the “high and stable” and “progressive” groups in that they were significantly more likely to have higher “cold/distant” and “social inhibition” subscale scores. Ollila et al. (2016) identified differences in those with stable alliance trajectories and steeper alliance slopes. Generally, those with the lowest IIP scores demonstrated the most stable alliance. However, those with the lowest “cold/distant” score demonstrated higher 36-months

alliance scores compared to 7-months. Furthermore, those with lowest "dominant", "vindictive", "intrusive", and "socially inhibited" scores had a higher alliance score at 12-months compared to 7-months. Those with high "vindictive", "cold", "socially inhibited" scores had a steeper slope of client-rated alliance at 12 and 36 months. Constantino and Smith-Hansen (2008) also identified subscale differences between groups. Significantly higher "distress" and "hostility" scores were evident in the "low initial" compared to the "high initial" alliance group.

Kalogerakos (2009) also compared alliance strength groups in terms of IIP and found that higher scores on "social inhibition", "non-assertiveness" and "overly-accommodating" subscales were associated with the "low-mid alliance strength shape" (initial low-mid alliance with no change over time) compared to the "high strength shape" (initial high alliance increasing linearly over time).

Additional Variables

Some studies identified additional variables which influenced the relationship between interpersonal problems and functioning and the alliance. Constantino et al. (2005) found a treatment group difference. IIP only influenced middle alliance in the IPT group and not in the CBT group. Constantino and Smith-Hansen (2008) found an interaction between "rigidity", "control" and "affiliation" and treatment group. Higher "rigidity" and "control" resulted in lower middle alliance scores in CBT and higher middle alliance scores in IPT. More affiliative problems were associated with higher middle alliance in CBT and unrelated to IPT.

Krieg and Tracey (2016) found an interaction between "dominance" and therapist gender and alliance. Higher dominance was associated with higher initial alliance with male counsellors whereas lower dominance was associated with higher initial alliance with female counsellors.

Several studies utilised both client and therapist-rated alliance. Of these, two (Gaston et al., 1988; Kokotovic & Tracey, 1990) did not separate these out and two (Hersoug et al., 2002; Puschner et al., 2005) found similar relationships between interpersonal problems and client and therapist-alliance. Three studies found differences in the relationship between interpersonal problems and alliance scale. Couture et al. (2006) found no relationship with client-rated alliance but a relationship between therapist-rated alliance and social functioning. Conversely, Hersoug et al. (2009) found no relationship with interpersonal variables and therapist-rated alliance but several relationships with client-rated alliance. Ollila et al. (2016) also found differences in the rate of improved alliance scores with therapist-rated alliance showing faster improvement than client-rated alliance for moderate to high IIP clients.

Discussion

The current review aimed to examine the relationship between interpersonal problems and functioning and therapeutic alliance to address the limitation outlined by Flückiger, Del Re et al.'s (2020) recent meta-analysis. The 19 included studies examined this relationship in diverse ways, looking at total IIP or interpersonal measure scores and alliance, specific interpersonal variables or subscales and alliance, alliance measured at different timepoints, or alliance development patterns. The main findings are discussed with attention to the extant literature, the quality of included studies, and the strengths and limitations of the current review.

The quality of the included studies is important to consider when interpreting and giving weight to the findings. Most of the studies received a global rating of weak due to various design and methodological shortcomings. Controlling for possible confounders, appropriately reporting and considering blinding of researchers and participants, and reporting on withdrawals and dropouts were the three main areas for concern across the studies. This could increase the risk of type 1 errors, researcher bias, and sampling bias.

Furthermore, few studies achieved a “good” rating for selection bias. There was a high level of heterogeneity between studies in terms of alliance measures, the way alliance was measured, the person rating the alliance, treatment modality and client presenting problems, which is also important to consider when interpreting the findings as this significantly limits the ability to compare studies.

There were mixed findings regarding the association between overall interpersonal problems or functioning and alliance. Three studies (Kalogerakos, 2009; Ollila et al., 2016; Paivio & Bahr, 1998) found relationships between total pre-treatment IIP score and alliance. This seemed to differ depending on the alliance subscale, where “bond” was more pertinent (Kalogerakos, 2009; Paivio & Bahr, 1998). Kalogerakos (2009) and Paivio & Bahr (1998) found the more severe interpersonal problems the poorer the alliance. Interestingly higher interpersonal problems predicted greater improvement in alliance over the course of therapy in Ollila et al.’s (2016) study. The authors provide a range of hypotheses regarding this finding, including higher motivation for change and working together with the therapist, and greater satisfaction as problems are alleviated in the relationship for those with higher interpersonal problems. Furthermore, Ollila et al. (2016) examined long-term psychotherapy which may explain the differential findings.

The relationship between overall interpersonal problems and alliance was either not replicated (Gaston et al., 1988; Krieg & Tracey, 1990) or appeared more nuanced in other studies. The timepoint alliance was measured (Constantino et al., 2005), the person rating the alliance (Couture et al., 2006; Hersoug et al., 2009), or the specific alliance subscale (Saunders, 2001) appeared to influence these findings. Higher interpersonal distress was also associated with poorer alliance (Constantino & Smith-Hansen, 2008; Renner et al., 2012).

More specific interpersonal problems or subscale and alliance relationships were identified in the studies. Specifically, items on the IIP showed differential influences on the

alliance. Overall higher “communion/affiliative” problems were associated with higher alliance scores (Altenstein-Yamanaka et al., 2017; Constantino & Smith-Hansen, 2008; Paivio & Bahr, 1998; Puschner et al., 2005; Renner et al., 2012), as well as those who scored highly on the “non-assertive” pole or “friendly-submissive” quartile (Paivio & Bahr, 1988; Renner et al., 2012; Saunders, 2001; Muran et al., 1994).. Several studies identified that those who scored highly on the “hostile-dominant” quartile had poorer alliance scores (Connolly Gibbons et al., 2003; Kalogerakos, 2009; Kokotovic & Tracey, 1990; Krieg & Tracey, 2016; Paivio & Bahr, 1998; Puschner et al., 2005) and, when other interpersonal problems influenced the alliance less as therapy went on, these factors sustained their influence (Puschner et al., 2005). Paivio and Bahr’s (1998) also controlled for hostility at therapy selection suggesting this association is present even with more moderate hostility.

Such findings support the literature that suggests that those in the “friendly-submissive” quartile compliment therapists’ “friendly-dominant” behaviour (Kiesler, 1983; Kiesler & Watkins, 1989) whereas “hostility” may prompt more hostile responses from therapists (Safran & Muran, 1996) and interfere with therapeutic action and goals (Tracey, 1993). Lower dominance may be associated with fewer power struggles and a willingness to give up control in the relationship (Beretta et al., 2005). This lends support to the clinical practice implications generated from the literature which posit that to maintain positive alliances therapists must learn not to respond defensively to hostility and approach clients’ negative responses (Horvath et al., 2011). Muran et al. (1994) suggest their finding may indicate that alliance is instead measuring compliance and “friendly-submissive” individuals are likely to comply but may not be truly benefitting from the relationship because they did not find the association with “bond”

However, Kalogerakos (2009) found hostility impacted total alliance and “bond” only, suggesting that specific parts of the therapeutic relationship may be affected whereas other elements (“task” and “goals”) are not. Interestingly, this contradicts the theory that

higher hostility will conflict with directive therapy action or modalities (Beutler et al., 1991) as “tasks” and “goals” could be viewed as the more directive elements of the relationship. This finding could also be suggestive of a separation between the components of the alliance measures. Gelso (2011) posits that the real relationship describes the more personal parts of the relationship whereas, the alliance describes the therapeutic work. The extant literature has found that the “bond” subscale attaches more onto the real relationship (Gelso, 2014; Kelley et al., 2010) thus a client’s interpersonal problems may be more influential within this domain because they reflect something more deeply personal within the therapeutic relationship. Conversely, Muran et al. (1994) did not find a relationship between the “hostile-dominant” subscale and alliance and Puschner et al.’s (2005) finding only applied to “hostility” not dominance. However, Muran et al. (1994) examined early alliance and, because “hostility” may have a longitudinal influence this may not have been captured. The generalisability of Puschner et al.’s (2005) sample can also be questioned, which may have influenced this differing finding.

The “cold/detached” dimension was associated with poorer alliance in several studies (Hersoug et al., 2002; Hersoug et al., 2009; Saunders, 2001) and “social inhibition/avoidance” (Kalogerakos, 2009; Muran et al., 1994; Paivio & Bahr, 1998) and low social functioning and therapist-rated alliance (Couture et al., 2006). This suggests that a level of approachability or willingness to enter a relationship influences the relationship’s success; avoidant individuals may be more distant and reluctant to engage in therapeutic tasks involving emotional processes (Mallinckrodt, 2010). This is in accordance with the extant literature that more detachment may lead to more rejection of therapists (e.g., von der Lippe et al., 2008). Indeed, Hersoug et al. (2002) and Kokotovic and Tracey (1990) found that the quality of current and past relationships influenced alliance quality, although current relationships were found to be stronger predictors (Hersoug et al., 2002). This supports the literature on client attachment style and social ability influencing client’s

alliance forming capabilities (Mallinckrodt, 2000; Mallinckrodt, 2010) and demonstrates the presence of “trait-like” components in alliance formation (Zilcha-Mano, 2017).

Interpersonal problems were identified as having more influence on early alliance in several studies (Constantino & Smith-Hansen, 2008; Hersoug et al., 2009; Paivio & Bahr, 1998; Puschner et al., 2005). Puschner et al. (2005) and Hersoug et al. (2009) found that early IIP subscales lost their influence towards the end of therapy, yet “hostile-dominance” stood the test of time (Puschner et al., 2005). This is interesting in that early alliance has often been positioned within the literature as more predictive of outcome (e.g., Barber et al., 1999; Hilsenroth et al., 2004; Piper et al., 1991) and it appears that interpersonal problems may have more of an impact on the early relationship. Extant literature suggests that interpersonal characteristics can change over the course of therapy (e.g., Driessen et al., 2015; Zilcha-Mano et al., 2016). It may, therefore, make sense that interpersonal difficulties which have not yet had corrective experiences within the therapeutic relationship (Hersoug et al., 2002; Zilcha-Mano, 2017) would be more influential in disrupting the alliance in early therapy.

Several alliance patterns were also examined in relation to interpersonal problems. Kalogerakos (2009) and Hersoug et al. (2010) found no difference between alliance pattern groups and overall IIP scores. Like findings on single-point or average alliance, specific interpersonal variables or subscales differentially related to alliance pattern groups (Beretta et al., 2005; Constantino & Smith-Hansen, 2008; Kalogerakos, 2009; Ollila et al., 2016). Differential methods of defining alliance trajectories limited comparison, however, overall, those patterns deemed to show a more negative alliance pattern were associated with more interpersonal problems, except for Ollila et al. (2016) who found steeper alliance growth in those with more interpersonal problems.

Treatment group differences (Constantino et al., 2005; Constantino & Smith-Hansen, 2008) suggest that although the alliance is a pan-theoretical construct (Bordin,

1979) it may be influenced differently by client interpersonal problems and therapeutic modality. Indeed, Paivio and Bahr (1998) attribute their differential findings from previous studies to the specific components of the experiential therapy model. Gelso (2014) highlights the interaction between therapeutic technique and alliance, suggesting that the dyadic experience between therapist and client influences the therapists' selection of therapeutic techniques. It may be that interpersonal problems arising within the relationship are responded to differently depending on a therapists' treatment modality, thus creating differential alliance patterns between therapy modalities. Indeed, Zilcha-Mano & Errázuriz (2015) found that integrative therapists were more successful in utilising the alliance to improve outcomes, suggesting that a level of flexibility within technique selection may be beneficial, this may be more necessary with the influence of high interpersonal problems.

Therapist gender differences were also identified by Krieg and Tracey (2016), which highlight the importance of examining the dyadic relations between client and therapist. Krieg and Tracey (2016) suggest that internalised societal stereotypes about gender roles may differentially interact with the poles of dominance dimension.

Whether the client or therapist completed the alliance measure also seemed to influence where the relationships between interpersonal problems and alliance were found. Although client-rated assessments have largely been favoured for their predictive relationship to alliance (Horvath et al., 2011), this supports the argument that there is still traction in examining the interdependence of client and therapist-ratings of alliance (e.g., Krasikova & LeBreton, 2012; Rozmarin et al., 2008). Several studies have found that both client and therapist-rated alliance predict client treatment outcome (Marcus et al., 2009; Zuroff et al., 2010).

Strengths and Limitations

The current review had several strengths. The use of broad search terms, multiple sources for handsearching, and including grey literature attempted to reduce the risk of publication bias and capture a wide range of literature. Employing a double rating procedure for data extraction and quality assessment improved rigour. However, there are several pertinent limitations. The exclusion of non-English language papers may have increased the risk of publication bias. The generalisability of the findings is limited both by the inclusion of studies originating wholly from western countries and the primary measures of interpersonal problems being culturally specific and derived from westernised perspectives on mental health and distress. Although the decision to include multiple therapeutic modalities and sample populations was derived from a theoretical understanding of the alliance as a pan-theoretical construct, this in turn limited the comparability between studies. Importantly, the way in which alliance was measured (single timepoints or trajectory) made comparison more difficult. Interpersonal problems were also reported dimensionally by some papers and as sub-groups by others, again challenging comparability. Theoretically, isolating client interpersonal problems and functioning from that of therapists ignores the dyadic and relational nature of the therapeutic alliance. These limitations lead to several important considerations for future research.

Future Research and Clinical Implications

Measuring alliance throughout treatment would enhance the understanding of the relationship between interpersonal problems and alliance (Zilcha-Mano et al., 2015; Zilcha-Mano, 2017). Indeed, several studies demonstrated changes in the relationship over time. Capturing alliance rated by both patient and therapist would also increase methodological rigour and examine possible interactions in interpersonal problems and alliance as measured by the relational dyad. It may be that clients and therapists consider different

aspects of the alliance (Eugster & Wampold, 1996; Marcus, et al., 2011), particularly as they have different roles within the relationship (Zilcha-Mano, 2017). Observer-rated alliance was not used by any of the included studies, but this could further enhance understanding.

Measuring interpersonal problems at multiple timepoints across therapy would also enhance understanding of change of time (Flückiger, Rubel et al., 2020). It may be possible to examine alliance and interpersonal problem trajectories to elucidate the relationship further. Indeed, Boswell et al. (2010) highlight the importance of looking at session outcome when examining process in therapy. Examining treatment model as a possible covariate may also be important in future research. Addressing the limitations of nested data within these studies would increase analytical rigour (Kahn, 2011) by examining variance from a multi-level model perspective (Marcus et al., 2009; Raudenbush & Bryk, 2002). Indeed, Krieg and Tracey (2016) utilised this approach and found fewer positive relationships between interpersonal problems and alliance, suggesting a risk of Type 1 errors in the current literature.

The therapeutic relationship is dyadic; thus, it is important to consider the role of the therapist and the possible interaction between complementary or non-complementary interpersonal styles. The characteristics of the therapist may influence clients' experience within the relationship (Gelso, 2014) and outcome (Anderson et al., 2009). Some studies have found that therapist characteristics which promote better alliances predict more successful outcomes (e.g., Baldwin & Imel, 2013; Heinonen & Niessen-Lie, 2010). There is also value in going beyond this and focusing on the dyadic relations themselves rather than isolating therapist and client characteristics (DuRubeis et al., 2005; Marcus et al., 2009). Returning to Zilcha-Mano's (2017) theoretical distinction between "trait-like" and "state-like" components of the alliance may be one way of examining these in tandem and the interaction between them (Zilcha-Mano, 2017).

Importantly, examining the role of interpersonal problems and functioning on the alliance-outcome association is an important next step to further understanding and add to Flückiger et al.'s (2020a) review. Zilcha-Mano (2017) highlights the importance of examining “trait-like” components of alliance at the outcome level.

Understanding client interpersonal problems and functioning and the impact on therapeutic alliance has several clinical implications. The evidence that clients with high interpersonal problems, such as hostility, dominance, social avoidance, and coldness/detachment, may have more difficulty forming positive therapeutic alliances has several implications for therapists' training, supervision, and responsiveness in therapy. Horvath et al. (2011) views alliance development as a skill in which therapists should be trained in. Identifying specific interpersonal problems or patterns at therapy outset would alert therapists to the need for responsive alliance building skills. More relational supervision models may also be appropriate in these instances. Ackerman and Hilsenroth (2003) identified specific skills expressed by the therapist which were associated with better alliances. Alliance-based training may also improve therapists' capacity to form positive alliances (e.g., Crits-Christoph et al., 2006; Eubanks-Carter et al., 2015). Therapists could also anticipate early challenges in the relationship when clients present with specific interpersonal problems and adapt practice accordingly (Constantino & Smith-Hansen, 2008). The weight of the early alliance also provides optimism about the capacity for alliances to develop over time even in the face of interpersonal problems.

The suggestion that interpersonal problems may differentially influence alliance in different treatment modalities has further implications. It may be that specific techniques help or hinder alliance when working with clients with interpersonal difficulties, thus therapy training programmes could also address this.

There may also be an argument for matching clients and therapists. Wampold (2015) suggests that clients with interpersonal difficulties and challenging relationship

histories may benefit from being paired with therapists who are better at developing alliances where these challenges exist. DeRubeis et al. (2014) highlights the possibility of adapting interventions for individual clients.

Conclusion

Findings from the current review identified several associations with client interpersonal problems and functioning and therapeutic alliance. Included studies had several design and methodological limitations and addressing these is important for future research. Future research which examines interpersonal problems and alliance on a sessional basis over time would enhance understanding as well as considering both client and therapist-alliance. Examining the interdependence between client and therapist interpersonal characteristics would develop theory further. Finally, there appears to be enough evidence of an association between client interpersonal problems and therapeutic alliance to warrant further investigation at the alliance-outcome level to further enhance Flückiger et al.'s (2020a) recent meta-analytic comparison review.

References

- Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review, 23*(1), 1–33. [https://doi.org/10.1016/S0272-7358\(02\)00146-0](https://doi.org/10.1016/S0272-7358(02)00146-0)
- Alexander, L. B., & Luborsky, L. (1986). *The Penn Helping Alliance Scales*. In L. S. Greenberg & W. M. Pinsof (Eds.), *Guilford clinical psychology and psychotherapy series. The psychotherapeutic process: A research handbook* (p. 325–366). Guilford Press.
- Altenstein-Yamanaka, D., Zimmermann, J., Krieger, T., Dörig, N., & Grosse Holtforth, M. (2017). Self-reported interpersonal problems and impact messages as perceived by significant others are differentially associated with the process and outcome of depression therapy. *Journal of Counseling Psychology, 64*(4), 410–423. <https://doi.org/10.1037/cou0000221>
- Andersen, S. M., & Przybylinski, E. (2012). Experiments on transference in interpersonal relations: Implications for treatment. *Psychotherapy, 49*(3), 370–383. <https://doi.org/10.1037/a0029116>
- 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., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology, 75*(6), 842–852. <https://doi.org/10.1037/0022-006X.75.6.842>
- Baldwin, S., & Imel, Z. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 258 –297). Wiley.

- Barber, J., Luborsky, L., Crits-Christoph, P., Thase, M., Weiss, R., Frank, A., Onken, L., & Gallop, R. (1999). Therapeutic Alliance as a Predictor of Outcome in Treatment of Cocaine Dependence. *Psychotherapy Research*, 9(1), 54–73.
<https://doi.org/10.1080/10503309912331332591>
- Beretta, V., de Roten, Y., Stigler, M., Drapeau, M., Fischer, M., & Despland, J.-N. (2005). The Influence of Patient's Interpersonal Schemas on Early Alliance Building. *Swiss Journal of Psychology / Schweizerische Zeitschrift für Psychologie / Revue Suisse de Psychologie*, 64(1), 13–20. <https://doi.org/10.1024/1421-0185.64.1.13>
- Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., & Merry, W. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *Journal of Consulting and Clinical Psychology*, 59(2), 333–340. <https://doi.org/10.1037//0022-006x.59.2.333>
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, 16(3), 252–260. <https://doi.org/10.1037/h0085885>
- Borkovec, T. D., Newman, M. G., Pincus, A. L., & Lytle, R. (2002). A component analysis of cognitive-behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *Journal of Consulting and Clinical Psychology*, 70(2), 288–298.
- Boswell, J. F., Castonguay, L. G., & Wasserman, R. H. (2010). Effects of psychotherapy training and intervention use on session outcome. *Journal of Consulting and Clinical Psychology*, 78(5), 717–723. <https://doi.org/10.1037/a0020088>
- Castonguay, L. G., & Beutler, L. E. (2006). Principles of therapeutic change: a task force on participants, relationships, and techniques factors. *Journal of Clinical Psychology*, 62(6), 631–638. <https://doi.org/10.1002/jclp.20256>

- Castonguay, L. G., Constantino, M. J., & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy: Theory, Research, Practice, Training*, 43(3), 271–279. <https://doi.org/10.1037/0033-3204.43.3.271>
- Connell, J., Grant, S., & Mullin, T. (2006). Client initiated termination of therapy at NHS primary care counselling services. *Counselling and Psychotherapy Research*, 6(1), 60–67. <https://doi.org/10.1080/14733140600581507>
- Connolly Gibbons, M. B., Crits-Christoph, P., de la Cruz, C., Barber, J. P., Siqueland, L., & Gladis, M. (2003). Pretreatment expectations, interpersonal functioning, and symptoms in the prediction of the therapeutic alliance across supportive-expressive psychotherapy and cognitive therapy. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 13(1), 59–76. <https://doi.org/10.1093/ptr/kpg007>
- Constantino, M. J. (2000). Interpersonal process in psychotherapy through the lens of the Structural Analysis of Social Behavior. *Applied & Preventive Psychology*, 9(3), 153–172. [https://doi.org/10.1016/S0962-1849\(05\)80002-2](https://doi.org/10.1016/S0962-1849(05)80002-2)
- Constantino, M. J., Castonguay, L. G., & Schut, A. J. (2002). The working alliance: A flagship for the scientist-practitioner model in psychotherapy. In G. S. Tryon (Ed.), *Counseling Based on Process Research: Applying what we know* (pp. 81–131). Allyn & Bacon
- Constantino, M. J., Arnow, B. A., Blasey, C., & Agras, W. S. (2005). The association between patient characteristics and the therapeutic alliance in cognitive-behavioral and interpersonal therapy for bulimia nervosa. *Journal of Consulting and Clinical Psychology*, 73(2), 203–211. <https://doi.org/10.1037/0022-006X.73.2.203>
- Constantino, M., & Smith-Hansen, L. (2008). Patient interpersonal factors and the therapeutic alliance in two treatments for bulimia nervosa. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 18(6), 683–698. <https://doi.org/10.1080/10503300802183702>

- Couture, S. M., Roberts, D. L., Penn, D. L., Cather, C., Otto, M. W., & Goff, D. (2006). Do baseline client characteristics predict the therapeutic alliance in the treatment of schizophrenia? *The Journal of Nervous and Mental Disease*, *194*(1), 10–14.
<https://doi.org/10.1097/01.nmd.0000195315.39196.52>
- Crits-Christoph, P., Connolly Gibbons, M. B., Crits-Christoph, K., Narducci, J., Schamberger, M., & Gallop, R. (2006). Can therapists be trained to improve their alliances? A preliminary study of alliance-fostering psychotherapy. *Psychotherapy Research*, *16*(3), 268–281. <https://doi.org/10.1080/10503300500268557>
- Crits-Christoph, P., Gibbons, M. B. C., & Mukherjee, D. (2013). Psychotherapy Process-Outcome Research. In *Bergin and Garfield's handbook of psychotherapy and behavior change* (6 ed., pp. 298-340). John Wiley and Sons. <https://www.wiley.com/en-us/Bergin+and+Garfield%27s+Handbook+of+Psychotherapy+and+Behavior+Change%2C+6th+Edition-p-9781118415924>
- Del Re, A. C., Flückiger, C., Horvath, A. O., Symonds, D., & Wampold, B. E. (2012). Therapist effects in the therapeutic alliance-outcome relationship: a restricted-maximum likelihood meta-analysis. *Clinical Psychology Review*, *32*(7), 642–649.
<https://doi.org/10.1016/j.cpr.2012.07.002>
- DeRubeis, R. J., Gelfand, L. A., German, R. E., Fournier, J. C., & Forand, N. R. (2014). Understanding processes of change: How some patients reveal more than others—and some groups of therapists less—about what matters in psychotherapy. *Psychotherapy Research*, *24*(3), 419–428. <https://doi.org/10.1080/10503307.2013.838654>
- Derubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A Conceptual and Methodological Analysis of the Nonspecifics Argument. *Clinical Psychology: Science and Practice*, *12*(2), 174–183. <https://doi.org/10.1093/clipsy.bpi022>
- de Roten, Y., Fischer, M., Drapeau, M., Beretta, V., Kramer, U., Favre, N. and Despland, J.N. (2004), Is one assessment enough? Patterns of helping alliance development and

outcome. *Clinical Psychology & Psychotherapy*, 11(5),324-

331. <https://doi.org/10.1002/cpp.420>

Dinger, U., Strack, M., Sachsse, T., & Schauenburg, H. (2009). Therapists' attachment, patients' interpersonal problems and alliance development over time in inpatient psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, 46(3), 277–290. <https://doi.org/10.1037/a0016913>

Driessen, E., Hegelmaier, L. M., Abbass, A. A., Barber, J. P., Dekker, J. J., Van, H. L., Jansma, E. P., & Cuijpers, P. (2015). The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis update. *Clinical Psychology Review*, 42, 1–15. <https://doi.org/10.1016/j.cpr.2015.07.004>

Eubanks-Carter, C., Muran, J. C., & Safran, J. D. (2015). Alliance-focused training. *Psychotherapy*, 52(2), 169–173. <https://doi.org/10.1037/a0037596>

Eugster, S. L., & Wampold, B. E. (1996). Systematic effects of participant role on evaluation of the psychotherapy session. *Journal of Consulting and Clinical Psychology*, 64(5), 1020–1028. <https://doi.org/10.1037/0022-006X.64.5.1020>

Falkenström, F., Granström, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology*, 60(3), 317–328. <https://doi.org/10.1037/a0032258>

Filak, J., Abeles, N., & Norquist, S. (1986). Clients' pretherapy interpersonal attitudes and psychotherapy outcome. *Professional Psychology: Research and Practice*, 17(3), 217–222. <https://doi.org/10.1037/0735-7028.17.3.217>

Flückiger, C., Del Re, A. C., Wlodusch, D., Horvath, A. O., Solomonov, N., & Wampold, B. E. (2020). Assessing the alliance-outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons. *Journal of Counseling Psychology*, 67(6), 706–711. <https://doi.org/10.1037/cou0000424>

- Flückiger, C., Rubel, J., Del Re, A. C., Horvath, A. O., Wampold, B. E., Crits-Christoph, P., Atzil-Slonim, D., Compare, A., Falkenström, F., Ekeblad, A., Errázuriz, P., Fisher, H., Hoffart, A., Huppert, J. D., Kivity, Y., Kumar, M., Lutz, W., Muran, J. C., Strunk, D. R., Tasca, G. A., ... Barber, J. P. (2020). The reciprocal relationship between alliance and early treatment symptoms: A two-stage individual participant data meta-analysis. *Journal of Consulting and Clinical Psychology, 88*(9), 829–843. <https://doi.org/10.1037/ccp0000594>
- Flückiger, C., Del Re, A. C., Wampold, B. E., Symonds, D., & Horvath, A. O. (2012). How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *Journal of Counseling Psychology, 59*(1), 10–17. <https://doi.org/10.1037/a0025749>
- Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy (Chicago, Ill.), 55*(4), 316–340. <https://doi.org/10.1037/pst0000172>
- Flückiger, C., Regli, D., Zwahlen, D., Hostettler, S., & Caspar, F. (2010). Der berner patienten- und therapeutenstundenbogen 2000: Ein instrument zur erfassung von therapieprozessen [The Bern Post Session Report 2000, patient and therapist versions: Measuring psychotherapeutic processes]. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis, 39*(2), 71–79. <https://doi.org/10.1026/1616-3443/a000015>
- Follette, W. C., & Greenberg, L. S. (2006). Technique factors in treating dysphoric disorders. In L. G. Castonguay & L. E. Beutler (Eds.), *Principles of therapeutic change that work* (pp. 83–109). Oxford University Press.
- Frieswyk, S. H., Colson, D. B., & Allen, J. G. (1984). Conceptualizing the therapeutic alliance from a psychoanalytic perspective. *Psychotherapy: Theory, Research, Practice, Training, 21*(4), 460–464. <https://doi.org/10.1037/h0085989>
- Gaston, L., Marmar, C. R., Thompson, L. W., & Gallagher, D. (1988). Relation of patient pretreatment characteristics to the therapeutic alliance in diverse psychotherapies. *Journal*

of Consulting and Clinical Psychology, 56(4), 483–489. <https://doi.org/10.1037/0022-006X.56.4.483>

- Gaston, L. (1990). The concept of the alliance and its role in psychotherapy: Theoretical and empirical considerations. *Psychotherapy: Theory, Research, Practice, Training*, 27(2), 143–153. <https://doi.org/10.1037/0033-3204.27.2.143>
- Gaston, L. (1991). Reliability and criterion-related validity of the California Psychotherapy Alliance Scales—patient version. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(1), 68–74. <https://doi.org/10.1037/1040-3590.3.1.68>
- Gelso, C. J. (2011). *The real relationship in psychotherapy: The hidden foundation of change*. American Psychological Association.
- Gelso C. (2014). A tripartite model of the therapeutic relationship: theory, research, and practice. *Psychotherapy research: Journal of the Society for Psychotherapy Research*, 24(2), 117–131. <https://doi.org/10.1080/10503307.2013.845920>
- Gelso, C. J., & Carter, J. A. (1985). The relationship in counseling and psychotherapy: Components, consequences, and theoretical antecedents. *The Counseling Psychologist*, 13(2), 155-243. <https://doi.org/10.1177/0011000085132001>
- Greenson, R. R. (1965) The Working Alliance and the Transference Neurosis. *The Psychoanalytic Quarterly*, 34(2), 155-181, <https://doi.org/10.1080/21674086.1965.11926343>
- Gurtman, M. B., & Lee, D. L. (2009). Sex differences in interpersonal problems: A circumplex analysis. *Psychological Assessment*, 21(4), 515–527. <https://doi.org/10.1037/a0017085>
- Heinonen, E., & Nissen-Lie, H. A. (2020). The professional and personal characteristics of effective psychotherapists: a systematic review. *Psychotherapy research: Journal of the Society for Psychotherapy Research*, 30(4), 417–432. <https://doi.org/10.1080/10503307.2019.1620366>

- Henry, W. P., Schacht, T. E., & Strupp, H. H. (1986). Structural analysis of social behavior: Application to a study of interpersonal process in differential psychotherapeutic outcome. *Journal of Consulting and Clinical Psychology*, 54(1), 27–31. <https://doi.org/10.1037/0022-006X.54.1.27>
- Hersoug, A. G., Monsen, J. T., Havik, O. E., & Høglend, P. (2002). Quality of early working alliance in psychotherapy: diagnoses, relationship and intrapsychic variables as predictors. *Psychotherapy and Psychosomatics*, 71(1), 18–27. <https://doi.org/10.1159/000049340>
- Hersoug, A. G., Høglend, P., Havik, O. E., von der Lippe, A., & Monsen, J. T. (2009). Pretreatment patient characteristics related to the level and development of working alliance in long-term psychotherapy. *Psychotherapy research: Journal of the Society for Psychotherapy Research*, 19(2), 172–180. <https://doi.org/10.1080/10503300802657374>
- Hersoug, A. G., Høglend, P., Havik, O. E., & Monsen, J. T. (2010). Development of working alliance over the course of psychotherapy. *Psychology and psychotherapy*, 83(2), 145–159. <https://doi.org/10.1348/147608309X471497>
- Higgins, J., & Green, S. (Eds). (2011). *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration.
- Hilsenroth, M. J., Peters, E. J., & Ackerman, S. J. (2004). The Development of Therapeutic Alliance During Psychological Assessment: Patient and Therapist Perspectives Across Treatment. *Journal of Personality Assessment*, 83(3), 332–344. https://doi.org/10.1207/s15327752jpa8303_14
- Horowitz, L.M. (2004). *Interpersonal foundations of psychopathology*. American Psychological Association.
- Horowitz, L. M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *IIP, Inventory of Interpersonal Problems Manual*. The Psychological Corporation.

- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*(6), 885–892. <https://doi.org/10.1037/0022-006X.56.6.885>
- Horowitz, L. M., Wilson, K. R., Turan, B., Zolotsev, P., Constantino, M. J., & Henderson, L. (2006). How interpersonal motives clarify the meaning of interpersonal behavior: a revised circumplex model. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc, 10*(1), 67–86. https://doi.org/10.1207/s15327957pspr1001_4
- Horvath, A. O. (1994). Empirical validation of Bordin's pantheoretical model of the alliance: The Working Alliance Inventory perspective. In A. Horvath, & L. S. Greenberg (Eds.), *The working alliance: Theory, research and practice* (pp. 109-128). Wiley.
- Horvath A. O. (2018). Research on the alliance: Knowledge in search of a theory. *Psychotherapy research: Journal of the Society for Psychotherapy Research, 28*(4), 499–516. <https://doi.org/10.1080/10503307.2017.1373204>
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*(2), 223-233. <https://doi.org/10.1037/0022-0167.36.2.223>
- Horvath, A. O., & Greenberg, L. S. (Eds.). (1994). *Wiley series on personality processes. The working alliance: Theory, research, and practice*. John Wiley & Sons.
- Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*(2), 139–149. <https://doi.org/10.1037/0022-0167.38.2.139>
- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*(1), 9–16. <https://doi.org/10.1037/a0022186>

- Kahn, J. H. (2011). Multilevel modeling: Overview and applications to research in counseling psychology. *Journal of Counseling Psychology, 58*(2), 257–271. <https://doi.org/10.1037/a0022680>
- Kalogerakos, A. F. (2009). *An examination of therapeutic alliance patterns, client attachment, client interpersonal problems, and therapy outcome in process-experiential and cognitive-behavioural treatment for depression*. [Unpublished doctoral thesis]. University of Toronto.
- Kelley, F. A., Gelso, C. J., Fuentes, J. N., Marmarosh, C., & Lanier, S. H. (2010). The Real Relationship Inventory: Development and psychometric investigation of the client form. *Psychotherapy: Theory, Research, Practice, Training, 47*(4), 540–553. <https://doi.org/10.1037/a0022082>
- Kiesler, D. J. (1983). The 1982 Interpersonal Circle: A taxonomy for complementarity in human transactions. *Psychological Review, 90*(3), 185–214. <https://doi.org/10.1037/0033-295X.90.3.185>
- Kiesler, D. J. (1996). *Contemporary interpersonal theory and research: Personality, psychopathology and psychotherapy*. John Wiley & Sons.
- Kiesler, D. J., & Watkins, L. M. (1989). Interpersonal complementarity and the therapeutic alliance: A study of relationship in psychotherapy. *Psychotherapy: Theory, Research, Practice, Training, 26*(2), 183–194. <https://doi.org/10.1037/h0085418>
- Kivlighan, D., & Shaughnessy, P. (2000). Patterns of working alliance development: A typology of client's working alliance ratings. *Journal of Counseling Psychology, 47*(3), 362-371.
- Kokotovic, A. M., & Tracey, T. J. (1990). Working alliance in the early phase of counseling. *Journal of Counseling Psychology, 37*(1), 16–21. <https://doi.org/10.1037/0022-0167.37.1.16>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>

- Krasikova, D. V., & LeBreton, J. M. (2012). Just the two of us: Misalignment of theory and methods in examining dyadic phenomena. *Journal of Applied Psychology, 97*(4), 739–757. <https://doi.org/10.1037/a0027962>
- Krieg, C. & Tracey, T. (2016). Client Interpersonal Problems and the Initial Working Alliance. *The European Journal of Counselling Psychology, 4*(2). 191-204.
<https://doi.org/10.5964/ejcop.v4i2.64>
- Leary, T. (1957). *Interpersonal diagnosis of personality: A functional theory and methodology for personality evaluation*. Ronald Press.
- Luborsky, L., Crits-Christoph, P., Alexander, L., Margolis, M., & Cohen, M. (1983). Two helping alliance methods for predicting outcomes of psychotherapy. A counting signs vs. a global rating method. *The Journal of Nervous and Mental Disease, 171*(8), 480–491.
<https://doi.org/10.1097/00005053-198308000-00005>
- Mallinckrodt, B. (2000). Attachment, social competencies, social support, and interpersonal process in psychotherapy. *Psychotherapy Research, 10*(3), 239–266. <https://doi.org/10.1093/ptr/10.3.239>
- Mallinckrodt, B. (2010). The psychotherapy relationship as attachment: Evidence and implications. *Journal of Social and Personal Relationships, 27*(2), 262–270. <https://doi.org/10.1177/0265407509360905>
- Marcus, D. K., Kashy, D. A., & Baldwin, S. A. (2009). Studying psychotherapy using the one-with-many design: The therapeutic alliance as an exemplar. *Journal of Counseling Psychology, 56*(4), 537–548. <https://doi.org/10.1037/a0017291>
- Marcus, D. K., Kashy, D. A., Wintersteen, M. B., & Diamond, G. S. (2011). The therapeutic alliance in adolescent substance abuse treatment: A one-with-many analysis. *Journal of Counseling Psychology, 58*(3), 449–455. <https://doi.org/10.1037/a0023196>
- Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools

for qualitative systematic reviews. *BMC Health Services Research*, 14(1), 579.

<https://doi.org/10.1186/s12913-014-0579-0>

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>

Muran, J. C., Segal, Z. V., Samstag, L. W., & Crawford, C. E. (1994). Patient pretreatment interpersonal problems and therapeutic alliance in short-term cognitive therapy. *Journal of Consulting and Clinical Psychology*, 62(1), 185–190. <https://doi.org/10.1037//0022-006x.62.1.185>

Norcross, J. C., & Goldfried, M. R. (Eds.). (2019). *Handbook of psychotherapy integration* (3rd ed.). Oxford University Press. <http://dx.doi.org/10.1093/med-psych/9780190690465.001.0001>

Norcross, J. C., & Lambert, M. J. (2018). Psychotherapy relationships that work III. *Psychotherapy*, 55(4), 303-315. <http://dx.doi.org/10.1037/pst0000193>

Ollila, P., Knekt, P., Heinonen, E., & Lindfors, O. (2016). Patients' pre-treatment interpersonal problems as predictors of therapeutic alliance in long-term psychodynamic psychotherapy. *Psychiatry Research*, 241, 110–117. <https://doi.org/10.1016/j.psychres.2016.04.093>

Paivio, S. C., & Bahr, L. M. (1998). Interpersonal problems, working alliance, and outcome in short-term experiential therapy. *Psychotherapy Research*, 8(4), 392–407. <https://doi.org/10.1093/ptr/8.4.392>

Piper, W. E., Azim, H. F., Joyce, A. S., McCallum, M., Nixon, G. W., & Segal, P. S. (1991). Quality of object relations versus interpersonal functioning as predictors of therapeutic alliance and psychotherapy outcome. *The Journal of Nervous and Mental Disease*, 179(7), 432–438. <https://doi.org/10.1097/00005053-199107000-00008>

- Puschner, B., Bauer, S., Horowitz, L. M., & Kordy, H. (2005). The relationship between interpersonal problems and the helping alliance. *Journal of Clinical Psychology, 61*(4), 415–429. <https://doi.org/10.1002/jclp.20050>
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed). Sage Publications.
- Renner, F., Jarrett, R. B., Vittengl, J. R., Barrett, M. S., Clark, L. A., & Thase, M. E. (2012). Interpersonal problems as predictors of therapeutic alliance and symptom improvement in cognitive therapy for depression. *Journal of Affective Disorders, 138*(3), 458–467. <https://doi.org/10.1016/j.jad.2011.12.044>
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*(2), 95–103. <https://doi.org/10.1037/h0045357>
- Rozmarin, E., Muran, J. C., Safran, J., Gorman, B., Nagy, J., & Winston, A. (2008). Subjective and intersubjective analyses of the therapeutic alliance in a brief relational therapy. *American Journal of Psychotherapy, 62*(3), 313–328. <https://doi.org/10.1176/appi.psychotherapy.2008.62.3.313>
- Ruiz, M. A., Pincus, A. L., Borkovec, T. D., Echemendia, R. J., Castonguay, L. G., & Ragusea, S. A. (2004). Validity of the Inventory of Interpersonal Problems for Predicting Treatment Outcome: An Investigation With The Pennsylvania Practice Research Network. *Journal of Personality Assessment, 83*(3), 213–222. https://doi.org/10.1207/s15327752jpa8303_05
- Safran, J. D., & Muran, J. C. (1996). The resolution of ruptures in the therapeutic alliance. *Journal of Consulting and Clinical Psychology, 64*(3), 447–458. <https://doi.org/10.1037/0022-006X.64.3.447>
- Safran, J.D., & Muran, J.C. (2000). *Negotiating the therapeutic alliance: A relational treatment guide*. Guilford Press.

- Saunders S. M. (2001). Pretreatment correlates of the therapeutic bond. *Journal of Clinical Psychology, 57*(12), 1339–1352. <https://doi.org/10.1002/jclp.1101>
- Schauenburg, H., Kuda, M., Sammet, I., & Strack, M.. (2000). The influence of interpersonal problems and symptom severity on the duration and outcome of short-term psychodynamic psychotherapy. *Psychotherapy Research, 10*(2), 133–146. <https://doi.org/10.1080/713663670>
- Sterba. R. F. (1934) The fate of the ego in analytic therapy. *International Journal of Psychoanalysis, 15*, 117-126. <https://doi.org/10.1177/000306519404200310>
- Stiles, W. B., Glick, M. J., Osatuke, K., Hardy, G. E., Shapiro, D. A., Agnew-Davies, R., Rees, A., & Barkham, M. (2004). Patterns of Alliance Development and the Rupture-Repair Hypothesis: Are Productive Relationships U-Shaped or V-Shaped? *Journal of Counseling Psychology, 51*(1), 81–92. <https://doi.org/10.1037/0022-0167.51.1.81>
- Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing, 1*(3), 176–184. <https://doi.org/10.1111/j.1524-475X.2004.04006.x>
- Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(3), 207–210. <https://doi.org/10.1037/1040-3590.1.3.207>
- Tracey, T. J. (1993). An interpersonal stage model of the therapeutic process. *Journal of Counseling Psychology, 40*(4), 396–409. <https://doi.org/10.1037/0022-0167.40.4.396>
- Von Der Lippe, A. L., Monsen, J. T., Rønnestad, M. H., & Eilertsen, D. E. (2008). Treatment failure in psychotherapy: The pull of hostility. *Psychotherapy Research, 18*(4), 420–432. <https://doi.org/10.1080/10503300701810793>
- Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*. Routledge.

- Wampold B. E. (2015). How important are the common factors in psychotherapy? An update. *World psychiatry: Official journal of the World Psychiatric Association (WPA)*, 14(3), 270–277. <https://doi.org/10.1002/wps.20238>
- Zetzel, E. R. (1956). Current concepts of transference. *The International Journal of Psychoanalysis*, 37, 369–375.
- Zilcha-Mano, S. (2017). Is the alliance really therapeutic? Revisiting this question in light of recent methodological advances. *American Psychologist*, 72(4), 311–325. <https://doi.org/10.1037/a0040435>
- Zilcha-Mano, S., & Errázuriz, P. (2015). One size does not fit all: Examining heterogeneity and identifying moderators of the alliance-outcome association. *Journal of Counseling Psychology*, 62(4), 579–591. <https://doi.org/10.1037/cou0000103>
- Zilcha-Mano, S., McCarthy, K. S., Dinger, U., & Barber, J. P. (2014). To what extent is alliance affected by transference? An empirical exploration. *Psychotherapy*, 51(3), 424–433. <https://doi.org/10.1037/a0036566>
- Zilcha-Mano, S., Muran, J. C., Hungr, C., Eubanks, C. F., Safran, J. D., & Winston, A. (2016). The relationship between alliance and outcome: Analysis of a two-person perspective on alliance and session outcome. *Journal of Consulting and Clinical Psychology*, 84(6), 484–496. <https://doi.org/10.1037/ccp0000058>
- Zuroff, D. C., Kelly, A. C., Leybman, M. J., Blatt, S. J., & Wampold, B. E. (2010). Between-therapist and within-therapist differences in the quality of the therapeutic relationship: effects on maladjustment and self-critical perfectionism. *Journal of Clinical Psychology*, 66(7), 681–697. <https://doi.org/10.1002/jclp.20683>

Appendices

Appendix A

Search Terms

Database	Term 1	Term 2	Term 3
	Psychotherapy	Client characteristics	Therapeutic Alliance
Scopus	Psychotherapy	Client characteristics	Alliance
	Counseling	Patient characteristics	Relationship*
	Counselling	Client qualit*	Therap* alliance
		Patient qualit*	Therap* relationship
		Client variable*	
		Patient variable*	
		Client factor*	
		Patient factor*	
		Client feature*	
		Patient feature*	
		Client trait* Patient trait*	
		Client propert* Patient propert**	
PsycInfo	Psychotherapy (MeSH)	Client characteristics (MeSH)	Therapeutic Alliance (MeSH)
	Counselling (MeSH)	Client qualit*	
		Patient qualit*	
		Client variable*	
		Patient variable*	
		Client factor*	
		Patient factor*	

		Client feature*	
		Patient feature*	
		Client trait* Patient trait*	
		Client propert* Patient propert*	
Pubmed	Psychotherapy	Client characteristics	Alliance
	Counseling	Patient characteristics	Relationship*
	Counselling	Client qualit*	Therap* alliance
		Patient qualit*	Therap* relationship
		Client variable*	
		Patient variable*	
		Client factor*	
		Patient factor*	
		Client feature*	
		Patient feature*	
		Client trait* Patient trait*	
		Client propert* Patient propert*	
Web of Science	Psychotherapy	Client characteristics	Alliance
	Counseling	Patient characteristics	Relationship*
	Counselling	Client qualit*	Therap* alliance
		Patient qualit*	Therap* relationship
		Client variable*	
		Patient variable*	
		Client factor*	
		Patient factor*	
		Client feature*	
		Patient feature*	

Client trait* Patient
trait*

Client propert*
Patient propert**

Appendix B

EPHPP Quality Assessment Tool

Domain	Questions	Answers
Selection Bias	Q1) Are the individuals selected to participate in the study likely to be representative of the target population?	1 Very likely 2 Somewhat likely 3 Not likely 4 Can't tell
	Q2) What percentage of selected individuals agreed to participate?	1 80 - 100% agreement 2 60 – 79% agreement 3 less than 60% agreement 4 Not applicable 5 Can't tell
Study Design	Q1) Indicate the study design	1 Randomized controlled trial 2 Controlled clinical trial 3 Cohort analytic (two group pre + post) 4 Case-control 5 Cohort (one group pre + post (before and after)) 6 Interrupted time series 7 Other specify 8 Can't tell
	Q2) Was the study described as randomized? If NO, go to Component C	No Yes
	Q3) If Yes, was the method of randomization described? (See dictionary)	No Yes
	Q4) If Yes, was the method appropriate? (See dictionary)	No Yes
Confounders	Q1) Were there important differences between groups prior to the intervention?	1 Yes 2 No 3 Can't tell
	The following are examples of confounders: 1 Race 2	

	Sex 3 Marital status/family 4 Age 5 SES (income or class) 6 Education 7 Health status 8 Pre-intervention score on outcome measure	
	Q2) If yes, indicate the percentage of relevant confounders that were controlled (either in the design (e.g., stratification, matching) or analysis)?	1 80 – 100% (most) 2 60 – 79% (some) 3 Less than 60% (few or none) 4 Can't Tell
Blinding	Q1) Was (were) the outcome assessor(s) aware of the intervention or exposure status of participants?	1 Yes 2 No 3 Can't tell
	Q2) Were the study participants aware of the research question?	1 Yes 2 No 3 Can't tell
Data Collection Methods	Q1) Were data collection tools shown to be valid?	1 Yes 2 No 3 Can't tell
	Q2) Were data collection tools shown to be reliable?	1 Yes 2 No 3 Can't tell
Withdrawals and dropouts	Q1) Were withdrawals and drop-outs reported in terms of numbers and/or reasons per group?	1 Yes 2 No 3 Can't tell 4 Not Applicable (i.e. one time surveys or interviews)
	Q2) Indicate the percentage of participants completing the study. (If the percentage differs by groups, record the lowest).	1 80 -100% 2 60 - 79% 3 less than 60% 4 Can't tell 5 Not Applicable (i.e. Retrospective case-control)
Intervention Integrity	Q1) What percentage of participants received the	1 80 -100% 2 60 - 79% 3 less than 60% 4 Can't te

allocated intervention or exposure of interest?

Q2) Was the consistency of the intervention measured? 1 Yes 2 No 3 Can't tell

Q3) Is it likely that subjects received an unintended intervention (contamination or co-intervention) that may influence the results? 1 Yes 2 No 3 Can't tell

Analyses

Q1) Indicate the unit of allocation (circle one) community organization/institution practice/office

individual

Q2) Indicate the unit of analysis (circle one) community organization/institution practice/office

individual

Q3) Are the statistical methods appropriate for the study design? 1 Yes 2 No 3 Can't tell

Q4) Is the analysis performed by intervention allocation status (i.e. intention to treat) rather than the actual intervention received? 1 Yes 2 No 3 Can't tell

Appendix C

EPHPP Scoring Manual

The purpose of this dictionary is to describe items in the tool thereby assisting raters to score study quality. Due to under-reporting or lack of clarity in the primary study, raters will need to make judgements about the extent that bias may be present. When making judgements about each component, raters should form their opinion based upon information contained in the study rather than making inferences about what the authors intended. Mixed methods studies can be quality assessed using this tool with the quantitative component of the study.

Domain	Scoring guidance	Ratings
A) Selection Bias	<p>(Q1) Participants are more likely to be representative of the target population if they are randomly selected from a comprehensive list of individuals in the target population (score very likely). They may not be representative if they are referred from a source (e.g., clinic) in a systematic manner (score somewhat likely) or self-referred (score not likely).</p> <p>(Q2) Refers to the % of subjects in the control and intervention groups that agreed to participate in the study before they were assigned to intervention or control groups.</p>	<p>Good: The selected individuals are very likely to be representative of the target population (Q1 is 1) and there is greater than 80% participation (Q2 is 1).</p> <p>Fair: The selected individuals are at least somewhat likely to be representative of the target population (Q1 is 1 or 2); and there is 60 - 79% participation (Q2 is 2). 'Moderate' may also be assigned if Q1 is 1 or 2 and Q2 is 5 (can't tell).</p> <p>Poor: The selected individuals are not likely to be representative of the target population (Q1 is 3); or there is less than 60% participation (Q2 is 3)</p>

B) Study Design

In this section, raters assess the likelihood of bias due to the allocation process in an experimental study. For observational studies, raters assess the extent that assessments of exposure and outcome are likely to be independent. Generally, the type of design is a good indicator of the extent of bias. In stronger designs, an equivalent control group is present and the allocation process is such that the investigators are unable to predict the sequence.

Randomized Controlled Trial (RCT) An experimental design where investigators randomly allocate eligible people to an intervention or control group. A rater should describe a study as an RCT if the randomization sequence allows each study participant to have the same chance of receiving each intervention and the investigators could not predict which intervention was next. If the investigators do not describe the allocation process and only use the words 'random' or 'randomly', the study is described as a controlled clinical trial. See below for more details.

Was the study described as randomized? Score YES, if the

or selection is not described (Q1 is 4); and the level of participation is not described (Q2 is 5).

Good: will be assigned to those articles that described RCTs and CCTs.

Fair: will be assigned to those that described a cohort analytic study, a case control study, a cohort design, or an interrupted time series.

Weak: will be assigned to those that used any other method or did not state the method used.

authors used words such as random allocation, randomly assigned, and random assignment. Score NO, if no mention of randomization is made.

Was the method of randomization described? Score YES, if the authors describe any method used to generate a random allocation sequence. Score NO, if the authors do not describe the allocation method or describe methods of allocation such as alternation, case record numbers, dates of birth, day of the week, and any allocation procedure that is entirely transparent before assignment, such as an open list of random numbers of assignments. If NO is scored, then the study is a controlled clinical trial.

Was the method appropriate? Score YES, if the randomization sequence allowed each study participant to have the same chance of receiving each intervention and the investigators could not predict which intervention was next. Examples of appropriate approaches include assignment of subjects by a central office unaware of subject characteristics, or sequentially numbered, sealed, opaque envelopes. Score NO, if the randomization sequence is open to the individuals responsible for recruiting and allocating participants or providing the intervention, since

those individuals can influence the allocation process, either knowingly or unknowingly. If NO is scored, then the study is a controlled clinical trial.

Controlled Clinical Trial (CCT)

An experimental study design where the method of allocating study subjects to intervention or control groups is open to individuals responsible for recruiting subjects or providing the intervention. The method of allocation is transparent before assignment, e.g., an open list of random numbers or allocation by date of birth, etc.

Cohort analytic (two group pre and post)

An observational study design where groups are assembled according to whether or not exposure to the intervention has occurred. Exposure to the intervention is not under the control of the investigators. Study groups might be non-equivalent or not comparable on some feature that affects outcome.

Case control study

A retrospective study design where the investigators gather 'cases' of people who already have the outcome of interest and 'controls' who do not. Both groups are then questioned or their records examined about

whether they received the intervention exposure of interest.

Cohort (one group pre + post (before and after))

The same group is pretested, given an intervention, and tested immediately after the intervention. The intervention group, by means of the pretest, act as their own control group.

Interrupted time series

A study that uses observations at multiple time points before and after an intervention (the 'interruption'). The design attempts to detect whether the intervention has had an effect significantly greater than any underlying trend over time.
Exclusion: Studies that do not have a clearly defined point in time when the intervention occurred and at least three data points before and three after the intervention

Other: One time surveys or interviews

C) Confounders

By definition, a confounder is a variable that is associated with the intervention or exposure and causally related to the outcome of interest. Even in a robust study design, groups may not be balanced with respect to important variables prior to the intervention. The authors should indicate if

Good: will be assigned to those articles that controlled for at least 80% of relevant confounders (Q1 is 2); or (Q2 is 1).

confounders were controlled in the design (by stratification or matching) or in the analysis. If the allocation to intervention and control groups is randomized, the authors must report that the groups were balanced at baseline with respect to confounders (either in the text or a table).

Fair: will be given to those studies that controlled for 60 – 79% of relevant confounders (Q1 is 1) and (Q2 is 2).

Poor: will be assigned when less than 60% of relevant confounders were controlled (Q1 is 1) and (Q2 is 3) or control of confounders was not described (Q1 is 3) and (Q2 is 4).

D) Blinding

(Q1) Assessors should be described as blinded to which participants were in the control and intervention groups. The purpose of blinding the outcome assessors (who might also be the care providers) is to protect against detection bias.

Good: The outcome assessor is not aware of the intervention status of participants (Q1 is 2); and the study participants are not aware of the research question (Q2 is 2).

(Q2) Study participants should not be aware of (i.e. blinded to) the research question. The purpose of blinding the participants is to protect against reporting bias.

Fair: The outcome assessor is not aware of the intervention status of participants (Q1 is 2); or the study participants are not aware of the research question (Q2 is 2).

Poor: The outcome assessor is aware of the intervention status of participants (Q1 is 1); and the study participants are aware of the

		research question (Q2 is 1); or blinding is not described (Q1 is 3 and Q2 is 3).
E) Data Collection Methods	<p>Tools for primary outcome measures must be described as reliable and valid. If 'face' validity or 'content' validity has been demonstrated, this is acceptable. Some sources from which data may be collected are described below: Self reported data includes data that is collected from participants in the study (e.g., completing a questionnaire, survey, answering questions during an interview, etc.).</p> <p>Assessment/Screening includes objective data that is retrieved by the researchers. (e.g., observations by investigators).</p> <p>Medical Records/Vital Statistics refers to the types of formal records used for the extraction of the data. Reliability and validity can be reported in the study or in a separate study. For example, some standard assessment tools have known reliability and validity.</p>	<p>Good: The data collection tools have been shown to be valid (Q1 is 1); and the data collection tools have been shown to be reliable (Q2 is 1).</p> <p>Fair: The data collection tools have been shown to be valid (Q1 is 1); and the data collection tools have not been shown to be reliable (Q2 is 2) or reliability is not described (Q2 is 3).</p> <p>Poor: The data collection tools have not been shown to be valid (Q1 is 2) or both reliability and validity are not described (Q1 is 3 and Q2 is 3).</p>
F) Withdrawals and dropouts	<p>Score YES if the authors describe BOTH the numbers and reasons for withdrawals and drop-outs. Score NO if either the numbers or reasons for withdrawals and drop-outs are not reported. Score NOT APPLICABLE if the study was a one-time interview or survey where there was not follow-up data reported. The percentage of participants completing the study refers to the % of subjects remaining in the study at the</p>	<p>Good: will be assigned when the follow-up rate is 80% or greater (Q1 is 1 and Q2 is 1).</p> <p>Fair: will be assigned when the follow-up rate is 60 – 79% (Q2 is 2) OR Q1 is 4 or Q2 is 5.</p>

final data collection period in all groups (i.e. control and intervention groups).

Poor: will be assigned when a follow-up rate is less than 60% (Q2 is 3) or if the withdrawals and drop-outs were not described (Q1 is No or Q2 is 4). Not Applicable: if Q1 is 4 or Q2 is 5.

G) Intervention Integrity

The number of participants receiving the intended intervention should be noted (consider both frequency and intensity). For example, the authors may have reported that at least 80 percent of the participants received the complete intervention. The authors should describe a method of measuring if the intervention was provided to all participants the same way. As well, the authors should indicate if subjects received an unintended intervention that may have influenced the outcomes. For example, co-intervention occurs when the study group receives an additional intervention (other than that intended). In this case, it is possible that the effect of the intervention may be overestimated. Contamination refers to situations where the control group accidentally receives the study intervention. This could result in an under-estimation of the impact of the intervention.

Not applicable

H) Analyses

Was the quantitative analysis appropriate to the research question being asked? An intention-to-treat analysis is one in which all the participants in a

Not applicable

trial are analyzed according to the intervention to which they were allocated, whether they received it or not. Intention-to-treat analyses are favoured in assessments of effectiveness as they mirror the noncompliance and treatment changes that are likely to occur when the intervention is used in practice, and because of the risk of attrition bias when participants are excluded from the analysis.

Section two: Empirical Study

An examination of the relationship between therapy, therapist and client variables and therapists' Facilitative Interpersonal Skills (FIS) during “not on track” therapy sessions.

Abstract

Objectives The current study examined therapists' Facilitative Interpersonal Skills (FIS) in "not on track" (NOT) therapy sessions in relation to therapy, therapist, and client variables. FIS variability within a session and within a therapist was examined.

Design A proof-of-concept quantitative methodology was applied to audio-recorded therapy sessions from the PRaCTICED trial examining Cognitive Behaviour Therapy (CBT) and Person-Centred Experiential Therapy (PCET).

Methods Therapists' FIS was assessed in NOT therapy sessions using an adapted in-session version of the FIS rating scale. Treatment modality, treatment adherence, therapist demographic characteristics, client symptom severity, client resilience, client demographic characteristics, and client expectancy and credibility scores were examined in relation to the FIS. An exploratory descriptive analysis examined FIS variability across a session and within a therapist.

Results Therapy, therapist and client variables were not associated with FIS total scores. Significant differences were found between treatment modality and FIS domain scores. Higher hope scores were found for CBT compared to higher emotional expression, warmth, and empathy scores for PCET. Higher therapist adherence scores were significantly associated with higher FIS in CBT only. Emerging evidence was found for variability in FIS across a session.

Conclusion The study provided proof-of-concept for the utility of in-session FIS ratings in NOT sessions. Treatment modality differences were identified on a FIS domain level and therapist adherence and FIS were related in CBT only. No relationship was found between client variables and FIS.

Practitioner Points

- Use of FIS may differ between treatment modalities.

- Therapist training and clinical supervision could consider the relation between treatment technique and model specific relational skills.

Limitations

- The small sample size limited the statistical methods and inferences made.
- No interpersonal client and therapist variables were examined which may be more pertinent when researching FIS.
- Use of audio-recordings may omit important interpersonal communication between client and therapist.

Introduction

Psychotherapy outcome research has increasingly identified variability amongst therapists (e.g., Johns et al., 2019; Saxon & Barkham, 2012; Wampold, 2001; Wampold & Bolt, 2006). Such findings have placed the focus on “The Person of the Therapist” (APA Division Taskforce 29, Norcross & Wampold 2011) and interpersonal variables within the therapeutic relationship (Norcross, 2002), such as therapeutic alliance (e.g., Wampold & Imel, 2015) and expressions of empathy (Elliott et al., 2011). Outcome research frequently identifies therapist effects accounting for significant proportions of the outcome variance (e.g., Baldwin & Imel, 2013; Johns et al., 2019; Wampold, 2001; Wampold & Bolt, 2006), however methods locating the source of this variability are a recent phenomenon (Anderson et al., 2020). A range of empirically supported specific interpersonal skills were identified by Norcross and Wampold (2011) and have been defined as skills which create facilitative conditions within therapy (Anderson, et al., 2016a).

Facilitative Interpersonal Skills (FIS)

The Facilitative Interpersonal Skills (FIS) measure (Anderson et al., 2006) was developed from existing research focusing on the role of alliance, empathy, (e.g., Bordin, 1979; Safran & Muran, 2000; Norcross, 2002; Wampold, 2001) warmth, and positive regard (Rogers, 1959; Farber & Doolin, 2011) in therapy outcomes. FIS are defined as a set of pan-theoretical (Anderson et al., 2020) interpersonal communications used when helping another address their problems (Anderson & Patterson, 2013).

The FIS performance task and corresponding measure (Anderson et al., 2006) operationalises eight specific interpersonal processes demonstrated by those performing a helping role (see Table 1, Anderson & Patterson, 2013). The FIS performance task was designed to examine therapists’ responses to therapeutic problems, or ‘difficult moments’, depicted in standardised simulated video vignettes (Anderson et al., 2020). Brief video vignettes are shown to participants, after which participants are asked to respond as if they

were the therapist. 'Difficult moments' are selected because they are viewed as key events or 'critical incidents' (Fitzpatrick, Janzen, Chamodraka, 2006; Janzen, 2007) within the formation of the therapeutic relationship, thus providing a good opportunity for demonstrating therapists' FIS (Anderson et al., 2020). Timely and non-defensive responses to challenges demonstrate high therapist responsiveness (Anderson et al. 2020).

Table 1

Description of Facilitative Interpersonal Skill (FIS) domains

Facilitative Interpersonal Skill/Domain	Domain Description
Verbal Fluency (VF)	Extent to which the participant is verbally at ease in communicating
Hope and Positive Expectation (HPE)	Expressions of hope, optimism and change
Persuasiveness (PERS)	Capacity to induce a view from the other that may be different to his/her own
Emotional Expression (EE)	Level of energy and emotion in response
Warmth, Acceptance and Understanding (WAU)	Ability to care for and accept the other
Empathy (EMP)	Capacity to respond with an expressed understanding of the subjective experience of the other
Alliance Bond Capacity (ABC)	Capacity to provide a collaborative environment
Alliance Rupture-Repair Responsiveness (ARR)	Responsiveness to an interpersonal issue/rupture.
OR	
Problem Responsiveness (PR)	Problem responsiveness is the alternative domain used for benign cases (when a rupture cannot be guaranteed). Problem responsiveness assesses the extent to which the participant offers solutions to the specific problem and whether solutions are likely to be helpful

High therapist FIS was found to be a significant predictor of better outcomes (e.g., Anderson et al., 2009), client-rated alliance (Anderson et al., 2016a), and trainee therapists' performance in subsequent years of training (Anderson et al., 2016b). FIS has been examined "in-session", using actual therapy sessions rather than performance tasks (FIS-IS; Uhlin et al., 2010). Uhlin (2011) did not find a relationship between FIS-IS and outcome but a predictive relationship between FIS-IS and trained therapist-rated alliance, when clients demonstrated high social skills. Armstrong (2013) investigated narrative processes and FIS-IS and found relationships between FIS-IS and narrative process variables, such as client-rated smoothness and depth and client and therapist emotional arousal. This suggests other variables may be associated with FIS, as dyadic processes in-session demonstrate an interplay with FIS.

Remaining Questions

FIS Measure

Existing research has generated further questions about the FIS measure itself. Extant literature has not used a consistent therapy model (Anderson et al., 2015), thus utilising the same therapeutic modality could increase validity by examining the nature of FIS with less variability in therapeutic approach (Anderson et al., 2016a). Furthermore, other than Uhlin (2011) and Armstrong (2013) who used FIS-IS, FIS during the performance task is assumed to be representative of FIS within the therapy room (Anderson et al., 2016a).

Responsiveness in psychotherapy is defined as the way in which therapists respond to an evolving interpersonal context and attend to clients' communications (Stiles, et al., 1998). FIS has only been examined prior to the delivery of therapy and viewed as a "trait-like" construct, thus its responsive or "state-like" quality within-therapists is yet to be established (Anderson et al., 2009). Examining variability in therapist FIS across a session and within a therapist (across clients) would address these questions.

Associated variables

Examination of the relationship between FIS and other therapy, therapist or client variables is yet to be explored (Anderson et al., 2016a). Treatment modality is one source of therapist variability and often different modalities approach FIS-related constructs, such as the alliance, differently (Zilcha-Mano & Errázuriz, 2015). Examining FIS within different modalities, particularly relationally focused therapies, is a new direction for research (Anderson et al., 2020) and could explain the interaction between FIS and treatment techniques (Anderson et al., 2016b). Therapist adherence may also be influential. Lambert (2013) suggests therapists may adhere to a model more readily with less complex clients or inversely increase model delivery when a client is not responding. Indeed, research has suggested greater adherence may interfere with interpersonal and collaborative factors within therapy (e.g., Ball et al., 2007).

Extant literature on related constructs (e.g., alliance) provides support for investigating therapist characteristics and FIS. Alliance has been related to a range of therapists' personal characteristics (Roth & Fonagy, 1996) and therapist attachment styles (Schaenberg et al., 2010). Research on therapist experience is inconsistent, with some finding more experience correlates with better alliance (e.g., Auerbach & Johnson, 1977; Meier et al., 2005) and others not observing this relationship (Anderson et al., 2009; Dunkle & Friedlander, 1996; Hersoug et al., 2001; Marcus et al., 2009; Wampold & Brown, 2005). Zilcha-Mano and Errázuriz (2015) proposed two opposing hypotheses; either experienced therapists use the alliance more effectively within therapy, or less experienced therapists rely more on alliance-based skills when their technical skills are underdeveloped. The relationship between experience and FIS warrants further investigation (Anderson et al., 2009). Generally, therapist demographic characteristics have not been implicated within research into therapist effects (e.g., Beutler et al., 2004;

Okiishi et al., 2006), yet they have still been examined within several FIS studies and gender differences were found in post-training therapist FIS (Perlman et al., 2020).

Anderson et al. (2016b) highlight the need to examine relationships between FIS, client characteristics and process variables. Examining FIS-related constructs enables consideration of possible variables. Generally, demographic characteristics have not been associated with alliance (e.g., Constantino et al., 2002; Constantino et al., 2005), however this is not consistent across all studies (e.g., Connolly Gibbons et al., 2003). Lambert et al. (2004) posit that with increasing client severity, common factors within the therapeutic relationship may be constrained. Anderson et al. (2016a) use this to argue relevance of further exploration between client symptom severity and FIS. Indeed, several constructs have been examined with client severity. High severity has been considered to hinder alliance development (Flückiger et al., 2018; Hersoug et al., 2002) and the relationship between positive regard and outcome (Farber et al., 2018). Kiesler et al. (1967) found lower observer-rated empathy for higher severity clients. However, there is evidence to the contrary (e.g., Connolly Gibbons et al., 2003; Gaston et al., 1988; Klein et al., 2003). Clients' expectations of their own role within therapy (Patterson et al., 2008) and improvement (Connolly Gibbons et al., 2003; Meyer et al., 2002) have been found to predict early alliance.

Current study

The current proof-of-concept study had three primary aims to expand on existing in-session FIS research. Firstly, the relationship between therapy and therapist variables, and secondly client variables, and FIS in actual therapy sessions was examined. With associated variables yet to be established with FIS, a range of available characteristics were examined. Thirdly, it was possible to examine if therapists adapted their interpersonal processes in-session by measuring FIS at three timepoints and examining within-therapist differences across clients.

A proof-of-concept design was utilised to gather evidence for the feasibility of examining FIS in actual therapy sessions using novel methodology. Firstly, the performance task (Anderson et al., 2006) was developed to examine therapists' FIS during problematic therapy processes, the current study aimed to replicate this by examining therapists' interpersonal behaviour during difficult therapy sessions. In this study, such sessions were operationalised as those where the client is not progressing as expected in symptom reduction towards recovery, termed "not on track" (NOT). NOT sessions were assumed likely to be more challenging for therapists as the client was "off track", as well as having a higher likelihood of "difficult moments" in the session. NOT sessions were identified using Delgadillo et al.'s (2018) statistical algorithm developed from outcome feedback technology, which creates a signal for those sessions which fall outside of expected treatment response norms, denoting poorer progress. Secondly, whole therapy sessions were rated by two trained and clinically experienced raters to establish the feasibility of using the FIS measure in this way.

Closer examination of FIS in this way could enhance current understanding of the interpersonal processes within therapy. Lambert (2013) highlighted the importance of attending to the multiplicity of variance at the therapist level within outcome research and viewing it as a dyadic process between patient and therapist, as well as considering within-therapist differences. Understanding the factors which may mediate FIS in the first place and examining the use of FIS within a session would assist in this task. Findings derived from the research could inform therapist training programmes and supervisory processes.

Aims

The current study had the following aims:

1. To determine the relationship between therapy/therapist variables (therapy model, therapist adherence and competence, therapist demographics: age, gender, years of experience) and FIS during NOT therapy sessions.

2. To determine the relationship between client variables (client symptom severity, client resilience, client expectancy/credibility, client demographics: age, gender, employment status, multiple deprivation score) and FIS during NOT therapy sessions.
3. To consider the stability of therapists' interpersonal communication style in NOT therapy sessions, by examining patterns of FIS variability within a session and across clients.

Method

Design

A proof-of-concept design utilising a quantitative methodology was applied to existing audio-recorded data from the pragmatic randomised controlled trial assessing the effectiveness of counselling for depression (PRaCTICED; Barkham et al., 2021; Saxon et al., 2017). The independent variables were therapy (therapy model), therapist (therapist adherence, therapist demographics: age, gender, years of experience), and client characteristics (client symptom severity, client resilience score, client expectancy/credibility score, client demographics: age, gender, employment status, multiple deprivation score). The dependent variable was therapist FIS, rated during NOT client sessions. The project is shared with another researcher (CB). Shared elements of the projects and their individual components are outlined in Appendix A.

Ethical implications

Existing ethical approval was granted for the PRaCTICED trial by the Health Research Authority (REC 14/YH/0001), which permitted utilising the data for secondary process research. Specific approval for the current study was granted by the University of Sheffield Ethics Board (Appendix B)

PRaCTICED trial

The PRaCTICED trial compared Person-Centred Experiential Therapy (PCET; Pearce et al., 2012) and Cognitive Behaviour Therapy (CBT; J.S. Beck, 2011; A. T. Beck et al., 1979) in the Sheffield Improving Access to Psychological Therapies (IAPT) service. Clients were included in the PRaCTICED trial if they were aged > 18 years, were receiving step 2 treatment, had a Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001, Appendix C) score of 12 or more and where depression was the focus of treatment. Clients were excluded if there was evidence of an organic condition, psychosis, drug or alcohol dependence, high clinical risk, or a diagnosis of a personality disorder.

Participants ($N = 510$) were randomised to receive up to 20 sessions of either PCET or CBT delivered by 50 therapists. Counselling for Depression (CfD), more widely known as PCET involves both person-centred counselling (Mearns & Thorne, 2007) and emotion-focused therapy approaches (Elliott et al., 2004). The IAPT model CBT is informed by Beckian approaches to cognitive therapy (J.S. Beck, 2011; A.T. Beck et al., 1979) and Martell's behavioural activation approach (Martell et al., 2001), however within the PRaCTICED trial Beckian CBT alone was the comparator arm. Model consistency was ensured through provision of specific training in PCET or workshops on CBT delivery adhering to Beckian principles (J.S. Beck, 2011). Therapist adherence and competence was assessed using a newly developed measure, clinical supervision, and audio recordings.

Participants

A participant sub-sample from the PRaCTICED trial was selected for the current study, consisting of client-therapist dyads where therapy was identified as NOT using Delgadillo et al.'s (2018) algorithm. Sample inclusion criteria is presented in Table 2.

Key Measures

Facilitative Interpersonal Skills Rating Scale (FIS; Anderson & Patterson, 2013)

Therapists FIS score was coded by raters (AF and CB) using Anderson and Patterson's (2013) rating scale. Like Uhlin et al.'s (2010) in-session adaptation of the FIS performance task (Anderson et al., 2006), Anderson and Patterson's (2013) updated rating scale was adapted for in-session ratings (Appendix D). The FIS rating scale includes eight domains measuring therapists' interpersonal skills: verbal fluency (VF); hope and positive expectations (HPE); persuasiveness (PERS); emotional expression (EE); warmth,

Table 2

Participant subsample inclusion criteria

Inclusion criterion	
1	Where therapy is identified as NOT during treatment, the session prior to the NOT signal was selected.
2	Participants who met the four-session minimum threshold for having received treatment (PRaCTICED trial protocol, Saxon et al., 2017).
3	Participants who had remained in the same treatment condition (PCET/CBT)
4	Participants who had completed the primary outcome measure PHQ-9 and scored ≥ 13 .
5	Participants who had at least 3 PHQ-9 scores necessary for determining NOT.
6	A maximum of 5 NOT sessions per therapist. If a single therapist had more than 5 clients identified as NOT a random 5 were selected.

acceptance and understanding (WAU); empathy (EMP); alliance-bond capacity (ABC); alliance rupture-repair responsiveness (ARR). In this study the recently developed “problem responsiveness (PR)” domain (De Jong et al., unpublished; Appendix E) was substituted for “alliance rupture-repair responsiveness”. PR measures therapists’ responses to general problems as ruptures could not be guaranteed within NOT sessions, although would still be rated under the PR domain. Domains are assessed on a five-point Likert scale with one denoting “not characteristic” and five denoting “extremely characteristic”. Specific descriptions are provided for scoring each domain. Raters begin at three and adapt the score based on therapists’ performance.

Inter-rater reliability for the FIS has been established at a good level with each item reaching $r > .70$ (Anderson et al., 2016a; Anderson et al., 2009). Good inter-rater reliability was achieved with doctoral level raters with psychotherapy experience, whereas more raters were required to achieve good reliability for undergraduate and graduate raters (Anderson & Patterson, 2013). Likewise, the FIS-IS achieved good inter-rater reliability when raters with clinical experience were utilised (Armstrong, 2013; Uhlin, 2011). A factor analysis found FIS to be a single construct (McClintock et al., 2012). Concurrent validity was established with significant positive correlations between FIS, social skills, empathy, and sociability (Anderson et al., 2016a). In the current study, the two raters both had psychotherapy experience and were trained to a high standard in rating FIS.

“Not on track” (NOT) Measure (Patient Health Questionnaire (PHQ-9))

NOT cases were selected by applying Delgadillo et al.’s (2018) statistical algorithm. Those whose PHQ-9 scores (Kroenke et al., 2001) fell outside of the expected treatment response norms (e.g., Finch et al., 2001) were identified.

The PHQ-9 is a measure of depression. Scores range from 0-27 with 5-9 indicating mild depression, 10-14 moderate depression, 15-19 moderate-severe depression, and 20-27 severe depression. A meta-analysis conducted by Manea et al. (2012) found cut-off

scores of 8-11 had good diagnostic properties for major depression. A cut-off score of 10 indicated diagnoses of major depression (Levis Brooke & Thombs, 2019). The PHQ-9 was validated in a large cross-sectional study by Kroenke et al. (2001).

Independent Variables

A range of therapy, therapist and client variables were selected from the PRaCTICED trial and examined in relation to FIS.

Therapy/Therapist Variables.

Therapy Model. Assignment to either PCET or CBT was examined.

Therapist Demographics. Therapist demographics included age, gender and overall years of experience.

Therapist Adherence and Competence. The PRaCTICED trial developed the Session Adherence and Competence Scale (SACS; Saxon et al., 2017) as a measure of therapist adherence and competence. The measure was derived from the Person Centred and Experiential Psychotherapy Rating Scale (PCEPS; Freire et al., 2014) for PCET and the Cognitive Therapy Scale-Revised (CTS-R; Blackburn et al., 2001) for CBT, termed SACS-CfD and SACS-CBT (Appendix, J). SACS scores were obtained for each client at session two, six and twelve and were rated by therapists' supervisors qualified within the relevant model. Psychometric properties of the SACS are yet to be established.

Client variables.

Client Demographics. Client demographics included age, gender, employment status and multiple deprivation score. The Index of Multiple Deprivation (IMD; Consumer Data Research Centre, CDRC, 2021) was used to generate a client multiple deprivation based on client postcode. The IMD provides relative deprivation scores for small geographical areas based on national population datasets. Areas are ranked into deciles with '1' denoting the 'most deprived' and 10 denoting the 'least deprived'.

Client Symptom Severity. A range of outcome measures were selected which captured client symptom severity. The BDI-II (BDI-II; A.T. Beck et al., 1996; see Appendix F) was selected to ensure independence from the NOT measure (PHQ-9) whilst also being depression-specific, as therapy was primarily targeting depression. The BDI-II is a 21-item measure designed to assess the severity of symptoms of depression in line with those outlined in the Diagnostic and Statistics Manual for Mental Disorders-IV (DSM-IV, American Psychiatric Association, 2000). Each item has four statements representing varying levels of depression severity. Good internal reliability has been found in several populations (e.g., Segal et al., 2008; Wang & Gorenstein, 2013).

Baseline Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Evans et al., 2002; see Appendix G) score was utilised to examine wider psychological constructs. The CORE-OM is a trans-diagnostic 34-item measure of psychological distress. It consists of a series of statements rated on a four-point Likert scale assessing frequency of experiences in the past week. Statements are aligned with four domains: well-being, functioning, symptoms, and risk. Responses are averaged to produce a total score indicating levels of psychological distress. A clinical cut-off score of 10 is utilised, with increasing scores representing higher severity (mild, moderate, moderate-severe, severe). Reliable change is indicated when clients' scores move by 5 points and clinically significant change is indicated when clients' scores move into the non-clinical population. Internal and test-retest reliability range from .75-.95 (Evans et al., 2002). Good convergent validity with a range of other outcome measures has been established (Barkham et al., 2001; Connell et al., 2007).

The GAD-7 (GAD-7; Spitzer et al., 2006) is a measure assessing generalised anxiety disorder and uses a four-point Likert scale to capture the frequency of experiences in the last fortnight. Item responses are summed to create a total score. Mild, moderate, and

severe GAD are denoted by cut-off scores of 5, 10, and 15, respectively. The GAD-7 was found to have good reliability and validity (Spitzer et al., 2006).

PRaCTICED trial participants were also placed within two subgroups based on depression severity denoting “moderate” and “severe” using the Clinical Interview Schedule-Revised (CIS-R; Lewis et al., 1992).

Client Resilience. The Connor-Davidson Resilience Scale (CD-RISC; Connor & Davidson, 2003; Appendix, H) was collected at baseline. The CD-RISC is a 25-item measure of resilience assessed on a five-point Likert Scale. There is a total score of 100 and higher scores denote more resilience. Good internal consistency has been found with a Cronbach’s α of .89 and good test-retest-reliability (Connor & Davidson, 2003). Convergent validity was established by Connor and Davidson (2003) with the CD-RISC positively correlating with a measure of hardiness and negatively correlating with perceived stress and stress vulnerability.

Client Expectancy and Credibility Client treatment preferences and their expectancy and credibility ratings for each therapy modality were gathered using adapted standard scales at baseline (Appendix I), regardless of the therapy modality to which participants were assigned.

Procedure

Identifying the NOT Sample

The sub-sample of therapist-client dyads was identified by an independent researcher (MBS). Application of the expected treatment response norms (ETR; e.g., Finch et al., 2001) statistical algorithm identified cases where the PHQ-9 scores denoted progress falling outside of these norms, thus indicating clients are NOT towards recovery (Delgadillo et al., 2018). Individualised algorithms were created for separate baseline groups, informed by PHQ-9 score range, and growth curve modelling informed the specific ETR for each group. Confidence intervals allowed for the identification of an upper boundary,

where sessional PHQ-9 scores which fell into the upper boundary generated a NOT signal. NOT sessions were defined as those occurring prior to the NOT signal because PHQ-9 scores were completed before the session. Cases were stratified for PCET and CBT so half the NOT sessions preceded another NOT session and half an OT session. To generate a NOT signal PHQ-9 scores were required before and after the NOT session, thus first and last sessions were not included. The algorithm identified 68 NOT sessions (n = 40 CBT; n = 28 PCET) of 27 therapists (n = 17 CBT; n = 10 PCET). Nine were excluded as recordings were unavailable resulting in a final sub-sample of 59 sessions.

FIS Rating Training

Researchers were trained in the FIS methodology by KDJ and FIS-trained rater KS. Initially researchers were familiarised with the FIS rating scale by practicing on FIS performance task tapes. Once competence was established in training sessions delivered by KS, researchers rated practice in-session recordings. In-session recordings had been rated to consensus by KDJ and KS and researchers own ratings were compared and discussed with KDJ. Based on a two-way mixed effects, average rater, absolute agreement model (Koo & Li, 2016) an intraclass correlation coefficient (ICC) of .71 (95% CI .57-.81) was reached between AF and CB before progressing onto study sample ratings, indicating “moderate-good” reliability.

Researchers were familiarised with both treatment modalities (PCET, CBT) to ensure greater recognition of therapeutic processes as opposed to specific therapeutic technique. Researchers had access to the PRaCTICED trial PCET and CBT treatment manuals and additional support from an expert therapist (MH).

Rating Using the FIS

FIS ratings were applied to audio-recordings of NOT sessions. To support the third aim examining variability of therapist FIS within-session, the sessions were divided into three segments (beginning, middle, end). Research defines therapist responsiveness as a

pattern instead of a single occurrence within therapy (Stiles et al., 1998), which can occur on a whole session and on a momentary level during therapy (Honos-Webb & Stiles, 2002). Therefore, each segment was rated individually to establish therapist FIS across multiple timepoints. Sessions were generally 50 minutes long, so they were divided into two 17-minute segments, with the final segment accounting for the remaining portion of time left. The FIS rating scale (Anderson & Patterson, 2013) was adapted for in-session ratings. Each segment had a total FIS composite score comprising the eight domains, as well as individual domain scores.

Where researchers decided therapists were not fully matching the criteria for either whole points on the scale, then half marks were permitted (e.g., 3.5). The novel approach to rating in-session recordings led to new considerations around rating PR. Unlike the performance task, where a single problem or rupture is present, the in-session recordings often had multiple problems arising. To account for this, a decision was made to separate out problems, when multiple problems were present, and apply individual scores, which were then averaged to create a total PR score.

Audio-recordings were rated in blocks of 10 (50% PCET; 50% CBT) NOT sessions. Researchers were blinded to therapist, client, and outcome data. Within each researchers' block were two audio-recordings (1 PCET; 1 CBT) rated by both researchers ($n=8$). Researchers were blinded to which audio-recordings were to be double-rated until completion. A calibration meeting was held at the end of each rating block where double-rated sessions were compared and discussed to reduce the risk of rater drift and ensure consistency. Discrepancies between scores were discussed until agreement was reached. Raters discussed any further queries generated during rating to ensure a consistent approach. Discussions were recorded for ongoing reference during future rating (Appendix K).

Double-rated sessions achieved a good level of agreement between raters. The average measure ICC was .79 (95% CI: .72-.84). Differences between researchers' overall FIS domain and total scores for the full sample were not significant, see Table 3.

Analytic method

Normality and Non-parametric Approach

FIS distribution was examined utilising histograms and the Kolmogorov-Smirnoff test for normality. Data appeared to have a bimodal formation (Appendix, L) and the Kolmogorov-Smirnoff test indicated the FIS did not follow a normal distribution, $D = .132$, $p < .05$, thus a non-parametric statistical approach was taken.

Table 3

Independent samples t-tests comparing raters overall FIS scores

FIS Domain	Rater 1 (Mean)	Rater 2 (Mean)	<i>t</i>	<i>P</i>
	89.4	91.3	-0.42	0.68
Verbal Fluency	11.5	11.7	-0.23	0.82
Hope	10.5	10.9	-0.73	0.47
Persuasiveness	11.0	11.0	-0.00	1.00
Emotional Expression	11.7	12.3	-0.83	0.41
Warmth	11.0	12.1	-0.14	0.89
Empathy	10.9	11.4	-0.71	0.48
Alliance Bond Capacity	11.7	11.3	0.47	0.64
Problem Responsiveness	10.1	10.6	-0.731	0.47

FIS Scores

A FIS total score was calculated by summing all domains for each segment (1-3) and summing all segments, enabling examination of relationships between characteristics and therapists' overall FIS. Total scores for each segment were summed from all domain scores, enabling examination of FIS at the three session timepoints. Total domain scores for all eight domains were summed from the domain score for each segment, enabling examination of relationships between characteristics and specific FIS domains.

Aim 1: Examining Therapy/Therapist Characteristics and FIS

Relationships between therapy/therapist characteristics and FIS total score were examined. Bivariate correlations (Spearman's rho) were conducted for continuous variables (therapist years of experience, therapist adherence scores). Independent samples Mann-Whitney U Tests were conducted for categorical variables (treatment modality, therapist gender). Kruskal-Wallis *H* tests were utilised for variables containing multiple independent groups (therapist age), however are treated cautiously within the findings due to nested data meaning the assumption of independence of observations is violated (Marcus et al., 2009; Reise & Duan, 1999). Only therapist adherence scores from the SACS were utilized within the analysis as there was no variability between therapists' scores on the competence component of the measure. Some clients' therapists had incomplete adherence scores across the three data collection points, so an average was calculated from the available data. Relationships between therapy/therapist characteristics and the eight individual FIS domains were examined using the same statistical approach but using the total domain scores as the dependent variables.

Aim 2: Examining Client Characteristics and FIS

Relationships between client characteristics and FIS total scores were examined. Bivariate correlations (Spearman's rho) were conducted for continuous variables (client age, BDI-II, CORE-OM, GAD-7, CD-RISC). Independent samples Mann-Whitney U Tests were conducted for categorical variables (client gender, depression severity). Kruskal-Wallis tests were utilised for variables containing multiple independent groups (employment status, IMD, expectancy, credibility). IMD scores were transformed from deciles to quintiles for the analysis, to account for the small sample. Difference scores were calculated for each clients' expectancy and credibility score and transformed into three groups denoting: no difference, favouring CBT over PCET, or favouring PCET over CBT. The three expectancy and credibility scores were compared using Kruskal-Wallis

tests. Relationships between client characteristics and the eight individual FIS domains were examined using the same statistical approach but using the total domain scores as the dependent variable.

Aim 3: Examining FIS Variability within a Session and within a Therapist

FIS variability was first considered descriptively by examining the means for each segment and computing difference variables between segments. A Friedman's ANOVA was then utilised to compare within-group differences between the three session segments. A post-hoc Wilcoxon's signed rank was conducted to examine group differences more closely and the Bonferroni correction was applied, to reduce the risk of Type 1 error arising from multiple comparisons. Results are treated cautiously due to violating independence of observations. Examining within-therapist FIS variability by comparing FIS scores for multiple clients seen by the same therapist was limited by the small sample size. For the 14 therapists who saw multiple clients FIS total scores for each client were examined graphically and the scores were split by the whole sample mean to denote "high" and "low" therapist FIS.

Results

Sample characteristics

Client and therapist demographics are presented in Table 4. Fifty-nine clients were included in the sample. Selected NOT sessions had a mean session number of 7.49 ($SD=4.10$). Most clients were female ($n = 39$). Client age had a mean of 39.88 ($SD 12.13$). The sample identified as primarily White British ($n = 57$) and were employed ($n = 33$). Twenty-eight clients received PCET and 31 received CBT. The mean number of sessions attended was 15.14 ($SD = 4.84$).

Twenty-four therapists were included in the sample, most were female ($n = 19$) and over 50-years old (58.30%). PCET therapists were older than CBT therapists with all therapists falling in the 50-60+ range compared to CBT therapists falling in the 30-59

range only. The mean years of experience for the therapists was 13.13 ($SD=6.48$). Years of experience also differed between therapy modalities. On average CBT therapists had 8.67 years of experience compared to 18.39 in the PCET sample. Therapist ethnicity was not reported as part of the trial.

Table 4

Sample characteristics

Client Characteristics	N	%
Gender		
Female	38	64.4
Male	21	35.6
Ethnicity		
White British	56	94.92
Other	3	5.08
Employment Status		
Employed	33	55.93
Unemployed	14	23.72
In education	2	3.39
Missing Data	10	16.95
Treatment		
CBT	31	52.5
PCET	28	47.5
	M	SD
Age	39.88	12.13
Session total (<i>N=56, Missing Data N=3</i>)	15.14	4.84
Therapist Characteristics	N	%
Gender		
Female	19	91.55
Male	5	8.47
Female delivered sessions	42	71.19
Male delivered sessions	17	28.81
Age		
30-39	6	25
40-49	4	16
50-59	9	37.5
60+	5	20.83

Years of experience

13.13

6.48

FIS Descriptive Statistics

Descriptive statistics for the FIS total and domain scores are presented in Table 5.

FIS total scores had a possible range of 24-120 and FIS domain scores had a possible range of 3-15. The mean score for the FIS total ($M = 90.11$, $SD = 16.56$) fell below the median (93.5) and in the third quartile of the possible range of FIS total scores (72-96).

HPE, PERS, and PR had the lowest mean scores and WAU had the highest mean score.

Table 5

FIS Descriptive Statistics

FIS Domain	Total Mean (SD)	Range	CBT Mean (SD)	PCET Mean (SD)
FIS Total	90.11 (16.56)	53.00-116.00	88.17 (18.42)	92.25 (14.25)
VF	11.59 (2.71)	6.00-15.00	11.47 (3.11)	11.73 (2.23)
HPE	10.68 (1.85)	7.00-15.00	11.53 (1.79)	9.73 (1.41)
PERS	10.89 (1.97)	7.00-15.00	11.35 (2.18)	10.38 (1.59)
EE	11.97 (2.88)	6.00-15.00	11.19 (3.07)	12.82 (2.43)
WAU	12.03 (2.43)	7.00-15.00	10.98 (2.40)	13.18 (1.91)
EMP	11.19 (2.19)	6.00-15.00	10.42 (2.04)	12.05 (2.05)
ABC	11.37 (2.87)	6.00-15.00	10.97 (3.06)	11.82 (2.61)
PR	10.39 (2.79)	6.00-15.00	10.25 (2.98)	10.54 (2.60)

Note: SD= Standard Deviation, VF=Verbal Fluency, HPE=Hope, PERS=Persuasiveness, EE= Emotional Expression, WAU=Warmth, acceptance and understanding, EMP=Empathy, ABC=Alliance Bond Capacity, PR=Problem Responsiveness

Aim 1: Therapy and therapist characteristics and FIS

Treatment modality

Mann-Whitney U test statistics examining treatment modality differences are presented in Table 6. A Mann-Whitney U test found no significant difference in FIS total score between treatment modalities $U(N_{CBT} = 31, N_{PCET} = 28,) = 382.00, p = .43$.

When examining treatment modality on a FIS domain level, several significant differences were found. A significant difference between treatment modalities was found for HPE, ($U(N_{CBT}=31, N_{PCET}=28) = 673, p < .001$), with higher HPE scores evident in CBT compared to PCET. Significant differences between treatment modalities were also found for EE ($U(N_{CBT}=31, N_{PCET}=28) = 292.50, p < .05$), WAU ($U(N_{CBT}=31, N_{PCET}=28) = 208, p < .01$) and EMP ($U(N_{CBT}=31, N_{PCET}=28) = 255, p < .01$) with higher domain scores in PCET compared to CBT.

Therapist Adherence

Spearman's Rho correlation coefficients for therapist adherence across each FIS domain and between treatments are presented in Table 7. No significant correlation was found for therapist adherence scores and FIS total score when examining the relationship across therapies ($r_s = .22, p = .11, N = 59$). When examining therapist adherence scores and FIS total score at a therapy modality level there was a significant positive correlation between therapist adherence and FIS total score in the CBT group ($r_s = .54, p < .01, N = 59$).

Table 6

Mann-Whitney U tests for treatment modality

FIS Domain	Mann-Whitney (U)
FIS Total	382.00
VF	425.00
HPE	673.00***
PERS	552.00
EE	292.50*
WAU	208.00**
EMP	255.00**
ABC	369.00
PR	414.00

Note: VF=Verbal Fluency, HPE=Hope, PERS=Persuasiveness, EE= Emotional Expression, WAU=Warmth, acceptance and understanding, EMP=Empathy, ABC=Alliance Bond Capacity, PR=Problem Responsiveness

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

Several significant relationships were found between therapist adherence and FIS domain scores, all of which were found in the CBT group only when examining therapy modalities separately. Significant positive correlations were found in the CBT group only between therapist adherence and VF ($r_s=.54$, $p<.05$, $N=59$; CBT: $r_s=.60$, $p<.01$, $N=59$), HPE ($r_s=.44$, $p<.01$, $N=59$; CBT: $r_s=.49$, $p<.05$, $N=59$), and PERS ($r_s=.33$, $p<.05$, $N=59$; CBT: $r_s=.41$, $p<.05$, $N=59$). Significant positive correlations were not found on a whole sample level but identified in CBT between therapist adherence and EE ($r_s=.48$, $p<.05$, $N=59$), WAU ($r_s=.41$, $p<.05$, $N=59$), ABC ($r_s=.48$, $p<.01$, $N=59$), and PR ($r_s=.46$, $p<.05$, $N=59$).

Therapist Demographics

Statistical analysis examining therapist demographics (age, gender, years of experience) using a Kruskal-Wallis test, Mann-Whitney U test and Spearman's Rho correlation respectively, are presented in Table 8. No significant differences were found between therapist gender and FIS scores on a total or domain level.

Table 7

Spearman's Rho correlations between therapist adherence and therapist FIS

FIS Domain	Adherence	CBT Adherence	PCET Adherence
FIS Total	.22	.54**	-.05
VF	.28*	.60**	-.15
HPE	.44**	.49*	.13
PERS	.33*	.41*	.00
EE	.05	.48*	-.21
WAU	-.02	.41*	-.13
EMP	-.07	.34	-.21
ABC	.24	.48**	.04
PR	.23	.46*	.02

Note: CBT= Cognitive Behaviour Therapy, PCET= Person Centred Experiential Therapy, VF=Verbal Fluency, HPE=Hope, PERS=Persuasiveness, EE= Emotional Expression, WAU=Warmth, acceptance and understanding, EMP=Empathy, ABC=Alliance Bond Capacity, PR=Problem Responsiveness

* $p<.05$, ** $p<.01$

A comparison of the four age groups using a Kruskal-Wallis test found no significant differences between FIS total scores. For HPE, WAU, and EMP some significant age

group differences were identified. These are treated cautiously as they may instead reflect a treatment modality difference because of the differential age groups between therapies. For HPE a significant difference was found $H(3) = 10.16, p < .05$. A post-hoc Mann Whitney U found a significant difference between the 60+ group and the 40-49 group which withstood a Bonferroni correction with an adjusted alpha level of .0083 ($0.05/4$), $U(N_{40-49} = 10, N_{60+} = 14,) = 19.96, p = .004$, indicating higher HPE in the younger group. For WAU a significant difference was found, $H(3) = 8.07, p < .05$. A post-hoc Mann Whitney U identified significant differences between the 30-39 and 60+ and the 30-39 and 50-59 age groups but these did not withstand Bonferroni correction. For EMP a significant difference was found $H(3) = 8.75, p < .05$. A post-hoc Mann Whitney U found a significant difference between the 30-39 group and the 50-59 group which withstood a Bonferroni correction, $U(N_{30-39} = 15, N_{50-59} = 20,) = -16.86, p = .004$, indicating higher EMP in the older group.

A Spearman's Rho found no significant relationship between therapist years of experience and FIS total score ($r_s = .07, p = .603, N = 59$). Several significant relationships were identified between years of experience and FIS domain scores. Again, these are treated cautiously due to the differences in years of experience between therapy modalities. A significant negative correlation was found between years of experience and HPE ($r_s = -.42, p < .01, N = 59$) with higher HPE scores for less experienced therapists. A significant positive correlation was found between years of experience and WAU ($r_s = -.42, p < .01, N = 59$) and EMP ($r_s = -.42, p < .01, N = 59$) with higher scores for more experienced therapists. Indeed, this reflects the treatment modality differences with higher HPE scores found in CBT and higher WAU, and EMP scores found in PCET so this may be reflective of treatment differences rather than years of experience.

Table 8

Inferential statistical analysis for therapist demographics and therapist FIS

FIS Domain	Therapist variable (<i>Test statistic</i>)
------------	--

	Age (<i>H</i>)	Gender (<i>U</i>)	Years of experience (ρ)
FIS Total	1.53	349.50	.07
VF	2.61	426.50	.03
HPE	10.16*	338.00	-.42**
PERS	2.97	325.50	-.12
EE	2.79	345.50	.16
WAU	8.07*	361.00	.28*
EMP	8.75*	359.00	.27*
ABC	1.85	290.50	.14
PR	.87	345.50	.08

Note: *H*= Kruskal-Wallis test statistic, *U*= Mann-Whitney U test statistic, ρ =Spearman's Rho correlation coefficient, VF=Verbal Fluency, HPE=Hope, PERS=Persuasiveness, EE= Emotional Expression, WAU=Warmth, acceptance and understanding, EMP=Empathy, ABC=Alliance Bond Capacity, PR=Problem Responsiveness

* $p < .05$, ** $p < .01$

Aim 2: Client Characteristics and FIS

Overall, no significant differences were found between client characteristics and therapists' FIS total or FIS domain scores, presented in Table 9. One exception was between client gender and HPE where a Mann-Whitney U test found a significant difference ($U(N_{CBT} = 31, N_{PCET} = 28) = 557.50, p < .05$), with higher therapist HPE scores for female clients. However, this is treated with caution due to the number of statistical tests increasing the risk of a Type 1 error.

Aim 3: Examining FIS Variability within a Session and within a Therapist

FIS Variability within a Session

FIS variability was examined by comparing FIS scores across the three session segments. Descriptive statistics are presented in Table 10. The FIS score mean ($M = 29.16, SD = 4.89$) for segment one was lower compared to segment two and three and with smaller standard deviation suggesting less variability. Initially, the FIS scores for the three session segments were examined descriptively by examining the mean differences (see Table 11), which indicated further analysis.

The FIS scores were compared across segments using Friedman's ANOVA. A significant difference between the three segments was found, $\chi^2(2) = 7.29, p < .05$. Post-hoc tests using a Wilcoxon signed-rank test with a Bonferroni adjusted alpha level of .017

(0.05/3) found FIS scores at segment three were significantly higher than FIS scores at segment one ($Z = -2.41$, $p = .015$). Differences between segment one and segment two did not withstand a Bonferroni correction.

Table 9*Inferential statistics for client variables and therapist FIS*

FIS Domain	Client Variable (<i>Test statistic</i>)										
	Age (ρ)	Gender (U)	Employment Status (H)	Deprivation (H)	BDI-II (ρ)	CORE-OM (ρ)	GAD-7 (ρ)	Depression severity (U)	CD-RISC (ρ)	Expectancy (H)	Credibility (H)
FIS Total	-.05	478.50	.10	.18	-.21	-.18	-.01	319.00	.05	3.97	1.67
VF	-.02	488.00	.95	1.48	-.21	-.14	.02	320.50	.03	2.62	1.77
HPE	.05	557.50*	.28	1.58	-.21	-.16	-.05	322.50	.22	.19	.14
PERS	-.02	479.50	2.08	1.89	-.24	-.23	.02	329.00	.05	.99	.69
EE	-.03	493.50	.17	.57	-.13	-.09	.00	324.50	-.03	4.92	2.28
WAU	-.08	477.50	.85	1.51	-.04	-.08	.06	339.00	.00	4.96	.01
EMP	-.06	426.50	2.74	.79	-.14	-.21	-.07	387.00	-.08	3.47	.23
ABC	-.05	445.50	.03	.16	-.24	-.15	.06	294.00	.04	3.51	4.45
PR	-.07	448.50	.70	.32	-.25	-.18	-.08	345.50	-.01	2.35	.19

Note: ρ =Spearman's Rho correlation coefficient, U =Mann-Whitney U test statistic, H =Kruskal-Wallis test statistic, VF=Verbal Fluency, HPE=Hope, PERS=Persuasiveness, EE=Emotional Expression, WAU=Warmth, acceptance and understanding, EMP=Empathy, ABC=Alliance Bond Capacity, PR=Problem Responsiveness, BDI-II= Beck's Depression Inventory-II, CORE-OM= Clinical Outcomes in Routine Evaluation-Outcome Measure, GAD-7= Generalised Anxiety Disorder-7 measure, CD-RISC= Connor-Davidson Resilience Scale

* $p < .05$

FIS Variability within a Therapist

FIS was compared within a therapist for the 14 therapists who saw multiple clients. Due to small numbers this was examined graphically by splitting the FIS total score for each client into either a "high FIS" or "low FIS" group. This was split by the whole sample mean ($M = 90.11$) with scores above denoting "high FIS" and scores below denoting "low FIS".

Table 10*Descriptive statistics for FIS scores across time*

FIS Timepoint	Mean	SD	Range
FIS Segment 1	29.16	4.89	20.00-38.00
FIS Segment 2	30.32	6.29	17.00-40.00
FIS Segment 3	30.63	6.66	16.00-40.00

Note: SD= Standard Deviation**Table 11***Mean differences between FIS scores at three timepoints*

FIS Timepoint	Mean Difference	SD	Range
FIS 1 & 2	-1.17	4.11	-11.00-8.50
FIS 2 & 3	-.31	3.49	-12.00-7.50
FIS 1 & 3	-1.47	4.50	-12.00-9.50

Note: SD= Standard Deviation

Therapists' FIS scores for each client were then presented graphically as either "high" or "low" FIS. Seven therapists demonstrated a stable trend with FIS scores consistently being "high" or "low" across clients (Figure 1). Seven therapists showed more variability, with FIS scores differing across clients (Figure 2).

Figure 1

Graph showing FIS stability within a therapist (between their clients)

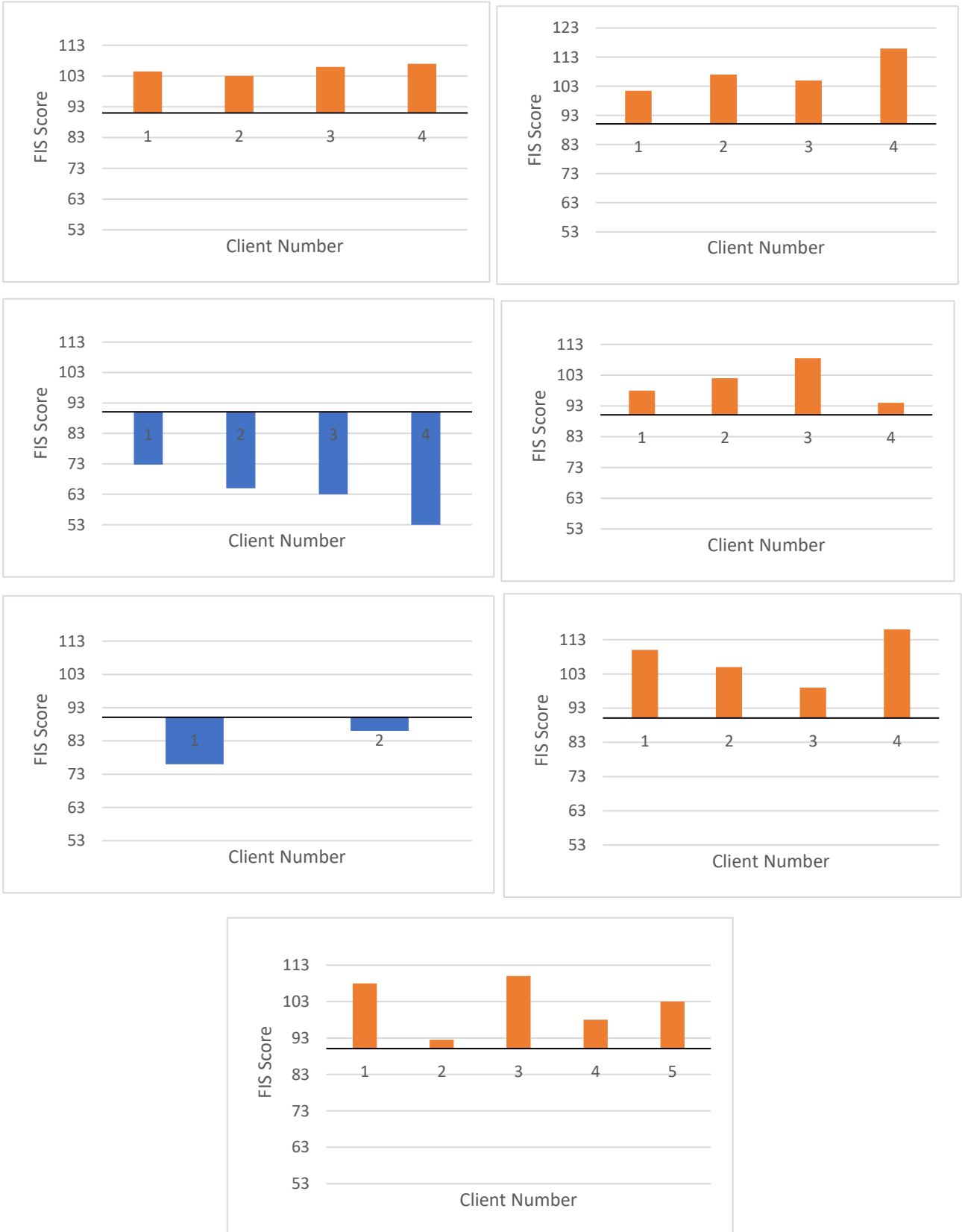
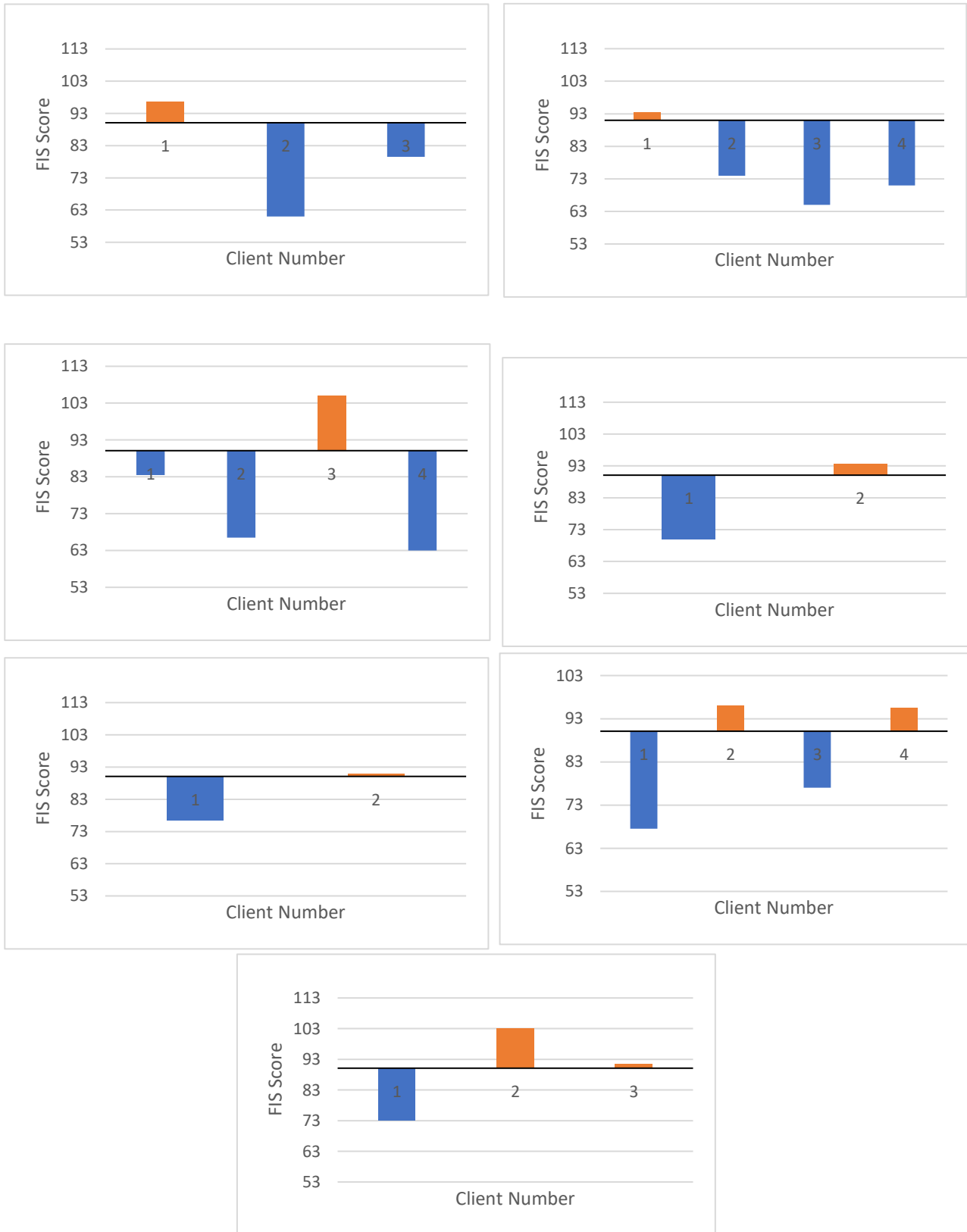


Figure 2

Graphs showing FIS variability within a therapist (between their clients)



Discussion

The aim of the current proof-of-concept study was to examine therapist FIS within actual therapy sessions, deemed to be NOT. Primarily, the relationship between therapy, therapist, and client variables and FIS was examined. A further exploratory aim examined whether FIS varied within a therapy session and within a therapist.

Therapy and Therapist Characteristics and FIS

The first aim of the current study examined the relationships between therapy model and therapist characteristics and FIS. Although, no significant differences were found between therapy models and overall therapist FIS, several differences were identified on a FIS domain level. A significant difference between therapy models was found for HPE, with higher scores within CBT compared to PCET. Significant differences between therapy models were also found for EE, WAU, and EMP, with higher scores within PCET compared to CBT. This finding could be understood within the theoretical and change based components of the therapy models. CBT has a focus on goals (J.S. Beck, 2011; A.T. Beck et al., 1979), thus explicitly addresses change processes from the outset with the client. It could be argued this goal focus means instances of therapists expressing hope and positive expectations, which are defined as expressions of hope, optimism, and change, are more inherent to the CBT model. PCET, on the other hand, draws on an emotion-focused approach (Elliott et al., 2004) with an emphasis on working with the emotional experience the client brings (Murphy, 2019). EE, WAU, and EMP could be viewed as key components of the PCET model more so than CBT. Although, FIS is defined as pan-theoretical (Anderson et al, 2020), the interplay between FIS and treatment-specific techniques has been an area of interest (Anderson et al., 2016b). Relationship and treatment factors have demonstrated an interdependence (e.g., Bedi et al., 2005; Hatcher & Barends, 2006; Horvath et al., 2011) and Anderson, et al. (2016b) hypothesised specific FIS domains may influence the delivery of specific treatment

components. It may be specific treatment components are thus enhanced by a therapist who can deliver parts of the treatment model on an interpersonal level additionally to the treatment manual approach (Anderson, et al., 2016b; Wampold & Imel, 2015).

Therapist adherence scores were associated with higher FIS scores in the CBT model only. Interestingly, this contradicts some of the research findings which have suggested higher adherence to the model may hinder a therapists relational and interpersonal skills (e.g., Ball et al., 2007), as well as suggestions within the literature about the inverse employment of treatment or relational skills when one of these areas is not going well (e.g., Anderson, et al., 2016b; Castonguay et al., 1996). Lambert (2013) posited two ideas regarding adherence; either therapists increase their delivery of the model when clients are not responding to treatment or find they can more readily adhere to the model when there is less complexity. It can be assumed NOT sessions may involve more complexity for therapists to respond to yet disentangling this finding would require further investigation. It may be therapists who adhere to the therapy model are more able to utilise their interpersonal skills in session as they have a better grasp of the treatment components. Alternatively, therapists may be simultaneously increasing adherence and their FIS in a responsive nature to the complexity.

Therapist demographics were also examined. Although, some age group differences were identified these cannot be separated from the issues regarding differential age brackets for CBT compared to PCET therapists. Therefore, it is not possible to draw a conclusion regarding age as a possible related characteristic to therapist FIS. The same issue was present for years of experience, due to much more experienced therapists within the PCET group. Therefore, although some differences were identified for HPE and WAU, this cannot provide further clarity to the already inconsistent research findings without further investigation.

Client Characteristics and FIS

The client characteristics investigated were not associated with therapist FIS, except for HPE and client gender. However, due to the numerous client variables and FIS domains investigated, this finding should be treated cautiously due to the high risk of a Type I error. The lack of evidence for a relationship between client characteristics and FIS is not surprising due to the already mixed findings for client symptom severity (Connolly Gibbons et al., 2003; Farber et al., 2018; Fluckiger et al., 2018; Gaston et al., 1988; Hersoug et al. 2002; Kiesler et al., 1967; Klein et al., 2003) and alliance. The current study supports existing research which has not found associations between client demographic characteristics and common factors (e.g., Constantino et al., 2002; Constantino et al., 2005). The lack of evidence for client expectancy or credibility and therapist FIS differs from some of the research which has found a relationship between clients' treatment expectation and alliance (e.g., Connolly Gibbons et al., 2003; Patterson et al., 2008; Meyer et al., 2002), however this may be due the way in which the scale was used in the analysis to detect expectation and credibility differences between therapies.

FIS Variability

The exploratory aim examining FIS variability within a session and within a therapist (seeing multiple clients) provided evidence that this under-researched area of the FIS warrants further investigation. On a whole sample level, FIS scores appeared to increase as the session progressed, suggesting therapists took time to “warm up” to their optimal level of FIS skills. FIS had previously been treated as a “trait-like” construct within the extant literature (e.g., Anderson et al., 2009; Anderson et al., 2016a, Anderson et al., 2016b). Although, these preliminary findings are exploratory and inferences cannot be drawn, some therapists' FIS scores fell into both “high” and “low FIS” categories when divided by the whole sample means. It may be therapist FIS is dependent on several other factors occurring within the therapeutic relationship, which warrants further investigation.

Using the FIS Methodology In-session

Selecting NOT therapy sessions for using the FIS in-session was in line with Anderson et al.'s (2020) rationalisation that “difficult moments” provide a good basis for evaluating a therapists’ FIS. Indeed, raters found NOT sessions provided a feasible basis for rating FIS and there was opportunity to rate “difficult moments” within these sessions. Rating two therapy models was also a novel approach within the extant FIS research. The way in which the FIS measure was used to rate CBT and PCET tapes led to several considerations about how therapists demonstrated FIS within their model (see Appendix K) and may be an interesting area for further investigation. The use of the PR domain in place of the alliance rupture-repair responsiveness domain also demonstrated utility in rating in-session recordings where a specific rupture may not have been present. The PR domain was also used to capture multiple problems as a way of managing the array of problems which arose in full therapy sessions. Dividing the therapy session into three segments proved a manageable means of rating FIS in-session and, also provided the opportunity to examine FIS over time. The ICC calculation between the two FIS raters demonstrated that good inter-rater reliability can be achieved when rating FIS in-session.

Limitations

The findings presented are important to consider within the context of several limitations. The small sample size of 59 means findings should be treated cautiously due to the possibility of variance being due to sampling error, limitations to the reliability of therapist means, and low power (Baldwin & Imel, 2013; Crits-Cristoph et al., 2006). The small sample also led to several other statistical limitations. The nested data structure of multiple clients to single therapists limited the scope of completing regression analysis, yet the small sample size did not permit examining this within a multi-level modelling (MLM) approach. This means the Kruskal-Wallis and Friedman’s ANOVA tests should be treated very cautiously, as not accounting for nested data leads to high risk of statistical error (Baldwin et al., 2005; Kahn, 2011; Marcus et al., 2009). An MLM approach would have

provided a way to manage the violation of independence of observations assumption within the data (Marcus et al., 2009; Reise & Duan, 1999). Separating out the different FIS domains within the analysis could be questioned, as Anderson et al. (2016a) highlights there is much intercorrelation between individual skills and it remains up for debate as to whether the FIS captures a general relational component or eight individual ones.

The client and therapist variables included within the current study were limited by those which were available within the PRaCTICED trial dataset. Examining more interpersonal characteristics may be of greater interest. Indeed, client attachment style (Diener & Monroe, 2011; Goldman & Anderson, 2007; Mallinckrodt & Jeong, 2015), object relations (e.g., Piper et al., 2004), and interpersonal problems (see Faulds, 2021) may have been more appropriate to consider. Heinonen and Nissen-Lie's (2020) recent review found evidence for therapists' personal and professional characteristics relating to patient outcome, thus, therapist variables with more of an interpersonal focus may have had greater utility than the variables examined here.

Several methodological issues are important to discuss. The availability of audio-recorded sessions was a potential limitation. The FIS performance task was developed as a video task, thus the absence of therapist and client body language and facial expressions within the social exchange may have limited the scope of the FIS ratings and increased the risk of omitting pertinent interpersonal information. The use of two clinically experienced raters was in line with Anderson and Patterson's (2013) evidence for high inter-rater agreement. However, it is always important to consider the risk of potential bias and the subjective understanding raters reached together. The participant sample was primarily White British and therapist ethnicity was not recorded as part of the trial. Lack of diversity within the sample limits the generalisability of the findings to wider populations. Increasing the diversity and size of the sample is important for future research. Involvement of experts by experience could have deepened understanding around the

utility of FIS within the study design. It is important to acknowledge that the FIS measure has been developed within Western research institutes, thus, several cultural assumptions are made about the expressions of FIS within the therapy session limiting its wider applicability.

Finally, examining FIS in-session although contributing to the FIS literature and increasing its ecological validity has its own theoretical issues. In standardising the client problems and examples in the FIS performance task the variance coming from the client is reduced (Anderson et al., 2020). Indeed, Anderson et al. (2016a) highlight an issue in the extant process literature is the interdependence between the therapist and client within the dyad, limiting the ability to isolate therapists' behaviour, and examining FIS in-session encounters this issue. Utilising the NOT algorithm provided a standardised way of replicating problematic sessions, however, client outcomes may be NOT for reasons external to the therapy, such as difficult life events.

Future Research and Clinical Implications

Future research could address several of these limitations. Increasing the sample size could overcome the statistical limitations and increase understanding by being able to isolate variability through the employment of an MLM approach. This would greatly assist in the examination of FIS variability by being able to examine this on an individual therapist and sessional level. Investigating the relationship between FIS and more interpersonal client and therapist variables would also broaden understanding of the FIS concept and possible clinical implications. Examining the interdependence between these factors could enhance understanding further, for example examining clients' interpersonal skills and how they interact with the therapists' FIS. Therapy specific differences in FIS is an interesting finding and future research could investigate this further with a bigger sample. To add to the research question regarding FIS as a "trait-like" or "state-like" quality, therapist FIS

could be measured on a performance task and again in-session, as suggested by Anderson et al. (2016a).

The proof-of-concept nature of the current study means clinical implications are limited at this stage. The suggestion that different therapeutic modalities may use FIS differently could have implications for therapeutic training and supervision by being able to focus on the model aligned FIS domains. Considering the current findings in tandem with Bentham's (2021) doctoral thesis on FIS and outcome could further explicate the findings.

Conclusion

Rating FIS in NOT therapy sessions is a novel approach and rating in-session had only been examined previously by Uhlin (2011) and Armstrong (2013). The current study has provided proof-of-concept for the utility of in-session FIS ratings and the replication of "problematic sessions" through employing the NOT algorithm. Further research questions regarding therapists' FIS in different treatment modalities, the relationship between adherence and FIS, and FIS as a responsive and "state-like" construct have been generated. Future research could address these through employing larger samples, MLM approaches, and including a wider range of therapist and client variables.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- Anderson, T., Finkelstein, J. D., & Horvath, S. A. (2020). The facilitative interpersonal skills method: Difficult psychotherapy moments and appropriate therapist responsiveness. *Counselling and Psychotherapy Research, 20*(3), 463–469. <https://doi.org/10.1002/capr.12302>
- Anderson, T., Crowley, M. E., Himawan, L., Holmberg, J. K., & Uhlin, B. D. (2016a). Therapist facilitative interpersonal skills and training status: A randomized clinical trial on alliance and outcome. *Psychotherapy Research: Journal of the Society for Psychotherapy Research, 26*(5), 511–529. <https://doi.org/10.1080/10503307.2015.1049671>
- Anderson, T., McClintock, A. S., Himawan, L., Song, X., & Patterson, C. L. (2016b). 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>
- Anderson, T., & Patterson, C. (2013). Facilitative interpersonal skill task and rating method. Retrieved from: http://clinica.ispa.pt/ficheiros/areas_utilizador/user11/facilitative_interpersonal_skill_manual_and_rating_scale.pdf
- Anderson, T., Patterson, C., & Weiss, A. C. (2006). *Facilitative interpersonal skills performance analysis rating method*. [Unpublished manuscript]. Ohio University, Ohio.

- Armstrong, C. (2013). *Towards an understanding of the therapeutic relationship: Narrative processes and facilitative interpersonal skills in psychotherapy provided by trainee therapists* [Unpublished doctoral dissertation]. McGill University, Canada. Retrieved from:
http://digitool.library.mcgill.ca/webclient/StreamGate?folder_id=0&dvs=1567361012436~971
- Auerbach, H. & Johnson, M. (1977). Research on the therapist's level of experience. In A.S. Gurman, & A.M. Razin (Eds). *Effective Psychotherapy: A handbook of research* (pp. 84–102). Pergamon.
- Baldwin, S., & Imel, Z. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 258 –297). Wiley.
- Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology, 73*(5), 924–935.
<https://doi.org/10.1037/0022-006X.73.5.924>
- Ball, S. A., Martino, S., Nich, C., Frankforter, T. L., Van Horn, D., Crits-Christoph, P., Woody, G. E., Obert, J. L., Farentinos, C., Carroll, K. M., & National Institute on Drug Abuse Clinical Trials Network (2007). Site matters: multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. *Journal of Consulting and Clinical Psychology, 75*(4), 556–567. <https://doi.org/10.1037/0022-006X.75.4.556>
- Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K., & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. *Clinical Outcomes in Routine Evaluation-Outcome*

Measures. *Journal of Consulting and Clinical Psychology*, 69(2), 184–196.

<https://doi.org/10.1037//0022-006x.69.2.184>

Barkham, M., Saxon, D., Hardy, G. E., Bradburn, M., Galloway, D., Wickramasekera, N., Keetharuth, A. D., Bower, P., King, M., Elliott, R., Gabriel, L., Kellett, S., Shaw, S., Wilkinson, T., Connell, J., Harrison, P., Arden, K., Bishop-Edwards, L., Ashley, K., & Ohlsen, S. (2021). Person-centred experiential therapy versus cognitive behavioural therapy delivered in the English Improving Access to Psychological Therapies service for the treatment of moderate or severe depression (PRaCTICED): a pragmatic, randomised, non-inferiority trial. *The Lancet*, 8(6), 487–499.

[https://doi.org/10.1016/S2215-0366\(21\)00083-3](https://doi.org/10.1016/S2215-0366(21)00083-3)

Beck, J. S. (2011). *Cognitive therapy: Basics and beyond*. (2nd. ed.) Guilford.

Beck, A. T., Rush, J. A., Shaw, B. F., Emery, G. (1979). *Cognitive therapy of depression*. The Guilford Press.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck depression inventory-II*. Psychological Corporation.

Bedi, R. P., Davis, M. D., & Williams, M. (2005). Critical incidents in the formation of the therapeutic alliance from the client's perspective. *Psychotherapy: Theory, Research, Practice, Training*, 42(3), 311–323. <https://doi.org/10.1037/0033-3204.42.3.311>

Bentham, C. (2012) *The role of therapists' interpersonal skills during "not on track" therapy sessions in predicting response to counselling person-centred experiential and cognitive behavioural therapy interventions for depression*. [Unpublished Doctoral Thesis] University of Sheffield.

Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S. & Wong, E. (2003). Therapist variables. In M. J. Lambert, J. Bergin & S. Garfield (Eds.) *Handbook of psychotherapy and behavior change*. (5th edition., pp. 227-306). John Wiley & Sons.

- Blackburn, I. M., James, I.A., Milne, D.L., Baker, C., Standard, S., Garland, A., & Reichelt, F.K. (2001). The Revised Cognitive Therapy Scale (CTS-R): psychometric properties. *Behavioural and Cognitive Psychotherapy*, 29(4) 431–46.
<http://dx.doi.org/10.1017/S1352465801004040>
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, 16(3), 252–260. <https://doi.org/10.1037/h0085885>
- Castonguay, L. G., Goldfried, M. R., Wisner, S., Raue, P. J., & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology*, 64(3), 497–504. <https://doi.org/10.1037/0022-006X.64.3.497>
- Connell, J., Barkham, M., & Mellor-Clark, J. (2007). CORE-OM mental health norms of students attending university counselling services benchmarked against an age-matched primary care sample. *British Journal of Guidance & Counselling*, 35(1), 41–57. <https://doi.org/10.1080/03069880601106781>
- Connolly Gibbons, M. B., Crits-Christoph, P., de la Cruz, C., Barber, J. P., Siqueland, L., & Gladis, M. (2003). Pretreatment expectations, interpersonal functioning, and symptoms in the prediction of the therapeutic alliance across supportive-expressive psychotherapy and cognitive therapy. *Psychotherapy research: Journal of the Society for Psychotherapy Research*, 13(1), 59–76. <https://doi.org/10.1093/ptr/kpg007>
- Connor, K. M., & Davidson, J. R. (2003). Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety*, 18(2), 76–82. <https://doi.org/10.1002/da.10113>
- Constantino, M. J., Castonguay, L. G., & Schut, A. J. (2002). The working alliance: A flagship for the scientist-practitioner model in psychotherapy. In G. S. Tryon (Ed.),

Counseling based on process research: Applying what we know (pp. 81–131). Boston: Allyn & Bacon.

Constantino, M. J., Arnow, B. A., Blasey, C., & Agras, W. S. (2005). The association between patient characteristics and the therapeutic alliance in cognitive-behavioral and interpersonal therapy for bulimia nervosa. *Journal of Consulting and Clinical Psychology, 73*(2), 203–211. <https://doi.org/10.1037/0022-006X.73.2.203>

Consumer Data Research Centre (2021) *Index of Multiple Deprivation*.

<https://data.cdrc.ac.uk/dataset/index-multiple-deprivation-imd>

Crits-Christoph, P., Gibbons, M. B. C., Crits-Christoph, K., Narducci, J., Schamberger, M., & Gallop, R. (2006). Can therapists be trained to improve their alliances? A preliminary study of alliance-fostering psychotherapy. *Psychotherapy Research, 16*(3), 268 –281. <http://dx.doi.org/10.1080/10503300500268557>

Delgadillo, 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., Sainty, A., Spencer, P., McMillan, D. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: A multisite, open-label, cluster randomised controlled trial. *The Lancet Psychiatry, 5* (7), 564-572. doi: 10.1016/S2215-0366(18)30162-7

De Jong, K., Steggle, K., & Anderson, T. (2019). *Problem Responsiveness Item*.

[Unpublished work] Department of Clinical Psychology, Universiteit Leiden.

Diener, M. J., & Monroe, J. M. (2011). The relationship between adult attachment style and therapeutic alliance in individual psychotherapy: A meta-analytic review. *Psychotherapy, 48*(3), 237–248. <https://doi.org/10.1037/a0022425>.

Dunkle, J. H., & Friedlander, M. L. (1996). Contribution of therapist experience and personal characteristics to the working alliance. *Journal of Counseling Psychology, 43*(4), 456–460. <https://doi.org/10.1037/0022-0167.43.4.456>

- Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. *Psychotherapy, 48*(1), 43–49. <https://doi.org/10.1037/a0022187>
- Elliott, R., Watson, J. C., Goldman, R. N., & Greenberg, L. S. (2004). *Learning emotion-focused therapy: The process-experiential approach to change*. American Psychological Association. <https://doi.org/10.1037/10725-000>
- 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. *The British Journal of Psychiatry: The Journal of Mental Science, 180*(1), 51–60. <https://doi.org/10.1192/bjp.180.1.51>
- Farber, B. A., & Doolin, E.M. (2011). Positive regard and affirmation. In J. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 168–186). Oxford University Press.
- Farber, B. A., Suzuki, J. Y., & Lynch, D. A. (2018). Positive regard and psychotherapy outcome: A meta-analytic review. *Psychotherapy, 55*(4), 411–423. <https://doi.org/10.1037/pst0000171>
- Faulds, A. (2021). *The relationship between client interpersonal characteristics and the therapeutic alliance: A systematic review*. [Unpublished doctoral dissertation]. University of Sheffield.
- Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology & Psychotherapy, 8*(4), 231–242. <https://doi.org/10.1002/cpp.286>
- Fitzpatrick, M. R., Janzen, J., Chamodraka, M., & Park, J. (2006). Client critical incidents in the process of early alliance development: A positive emotion-exploration spiral. *Psychotherapy Research, 16*(4), 486–498. <https://doi.org/10.1080/10503300500485391>

- Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy (Chicago, Ill.)*, 55(4), 316–340. <https://doi.org/10.1037/pst0000172>
- Freire, E., Elliott, R., & Westwell, G. (2014). Person-Centred and Experiential Psychotherapy Scale: Development and reliability of an adherence/competence measure for person-centred and experiential psychotherapies. *Counselling & Psychotherapy Research*, 14(3), 220–226. <https://doi.org/10.1080/14733145.2013.808682>
- Gaston, L., Marmar, C. R., Thompson, L. W., & Gallagher, D. (1988). Relation of patient pretreatment characteristics to the therapeutic alliance in diverse psychotherapies. *Journal of Consulting and Clinical Psychology*, 56(4), 483–489. <https://doi.org/10.1037/0022-006X.56.4.483>
- Goldman, G. A., & Anderson, T. (2007). Quality of object relations and security of attachment as predictors of early therapeutic alliance. *Journal of Counseling Psychology*, 54(2), 111–117. <https://doi.org/10.1037/0022-0167.54.2.111>
- Hatcher, R. L., & Barends, A. W. (2006). How a return to theory could help alliance research. *Psychotherapy: Theory, Research, Practice, Training*, 43(3), 292–299. <https://doi.org/10.1037/0033-3204.43.3.292>
- Heinonen, E., & Nissen-Lie, H. A. (2020). The professional and personal characteristics of effective psychotherapists: a systematic review. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 30(4), 417–432. <https://doi.org/10.1080/10503307.2019.1620366>
- Hersoug, A. G., Hoglend P., Monsen, J. T., & Havik, O. E. (2001). Quality of working alliance in psychotherapy: therapist variables and patient/therapist similarity as predictors. *Journal of Psychotherapy Practice Research*, 10(4), 205-216.

- Hersoug, A. G., Monsen, J. T., Havik, O. E., & Høglend, P. (2002). Quality of early working alliance in psychotherapy: Diagnoses, relationship and intrapsychic variables as predictors. *Psychotherapy and Psychosomatics*, *71*(1), 18–27. <https://doi.org/10.1159/000049340>
- Honos-Webb, L., & Stiles, W. B. (2002). Assimilative integration and responsive use of the assimilation model. *Journal of Psychotherapy Integration*, *12*(4), 406-420. <http://dx.doi.org/10.1037/1053-0479.12.4.406>
- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, *48*(1), 9–16. <https://doi.org/10.1037/a0022186>
- Janzen, J. (2007). *Attachment and facilitative interpersonal skill in critical therapy incidents*. [Unpublished doctoral dissertation]. McGill University.
- Johns, R. G., Barkham, M., Kellett, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to review. *Clinical Psychology Review*, *67*, 78–93. <https://doi.org/10.1016/j.cpr.2018.08.004>
- Kahn, J. H. (2011). Multilevel modeling: Overview and applications to research in counseling psychology. *Journal of Counseling Psychology*, *58*(2), 257–271. <https://doi.org/10.1037/a0022680>
- Kiesler, D. J., Klein, M. H., Mathieu, P. L., & Schoeninger, D. (1967). Constructive personality change for therapy and control patients. In C. R. Rogers et al. (Eds.), *The therapeutic relationship and its impact* (pp. 251–294), University of Wisconsin Press.
- Klein, D. N., Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C., Castonguay, L. G., Arnow, B., Blalock, J. A., Manber, R., Markowitz, J. C., Riso, L. P., Rothbaum, B., McCullough, J. P., Thase, M. E., Borian, F. E., Miller, I. W., & Keller, M. B. (2003). Therapeutic alliance in depression treatment: controlling for prior change and patient characteristics. *Journal of Consulting and Clinical Psychology*, *71*(6), 997–1006. <https://doi.org/10.1037/0022-006X.71.6.997>

- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (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>
- Lambert, M. J. (2013). *Bergin and Garfield's handbook of psychotherapy and behavior change*. John Wiley & Sons.
- Lambert, M. J., Garfield, S. L., & Bergin, A. E. (2004). Overview, trends, and future issues. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 805– 821). Wiley
- Levis Brooke, B.A., Thombs B. D. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for Screening to Detect Major Depression: Individual Participant Data Meta-analysis, *British Medical Journal*, 365 (l1476). <https://doi.org/10.1136/bmj.l1476>
- Lewis, G., Pelosi, A. J., Araya, R., & Dunn, G. (1992). Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychological Medicine*, 22(2), 465–486. <https://doi.org/10.1017/s0033291700030415>
- Mallinckrodt, B., & Jeong, J. (2015). Meta-analysis of client attachment to therapist: associations with working alliance and client pretherapy attachment. *Psychotherapy (Chicago, Ill.)*, 52(1), 134–139. <https://doi.org/10.1037/a0036890>
- Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Canadian Medical Association Journal*, 184(3), 191–196. <https://doi.org/10.1503/cmaj.110829>
- Marcus, D. K., Kashy, D. A., & Baldwin, S. A. (2009). Studying psychotherapy using the one-with-many design: The therapeutic alliance as an exemplar. *Journal of Counseling Psychology*, 56(4), 537–548. <https://doi.org/10.1037/a0017291>

- Martell, C. R., Addis, M. E., & Jacobson, N. S. (2001). *Depression in context: Strategies for guided action*. WW Norton & Co.
- McClintock, A. S., Song, X., Jameson, G., Anderson, T., Patterson, C., Goldman, G. A., Uhlin, B., MacFarlane, P., Goldman, E. D. (2012). *Reliability of therapist facilitative interpersonal skills for brief responses to "thin slice" client video simulations*. Paper presented at the annual meeting of the International Society for Psychotherapy Research, Virginia Beach, VA.
- Mearns, D., & Thorne, B. (2007). *Person-centred counselling in action*. 3rd ed. London: Sage Pubs
- Meier, P. S., Donmall, M. C., Barrowclough, C., McElduff, P., & Heller, R. F. (2005). Predicting the early therapeutic alliance in the treatment of drug misuse. *Addiction*, 100(4), 500–511. <http://dx.doi.org/10.1111/j.1360-0443.2005.01031.x>
- Meyer, B., Pilkonis, P. A., Krupnick, J. L., Egan, M. K., Simmens, S. J., & Sotsky, S. M. (2002). Treatment expectancies, patient alliance and outcome: Further analyses from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, 70(4), 1051–1055. <https://doi.org/10.1037/0022-006X.70.4.1051>
- Murphy D. (2019). *Person-centred experiential counselling for depression: a manual for training and practice*. SAGE Publishing.
- Norcross, J. C. (Ed.). (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patient needs*. Oxford University Press.
- Norcross, J. C., & Wampold, B. E. (2011). Evidence-based therapy relationships: research conclusions and clinical practices. *Psychotherapy (Chicago, Ill.)*, 48(1), 98–102. <https://doi.org/10.1037/a0022161>
- Okiishi, J. C., Lambert, M. J., Eggett, D., Nielsen, L., Dayton, D. D., & Vermeersch, D. A. (2006). An analysis of therapist treatment effects: toward providing feedback to

individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology*, 62(9), 1157–1172. <https://doi.org/10.1002/jclp.20272>

Patterson, C. L., Uhlin, B., & Anderson, T. (2008). Clients' pretreatment counseling expectations as predictors of the working alliance. *Journal of Counseling Psychology*, 55(4), 528–534. <https://doi.org/10.1037/a0013289>

Pearce, P., Sewell, R., Hill, A., & Coles, H. (2012). Counselling for depression. *Therapy Today*, 23, 20-23.

Perlman, M. R., Anderson, T., Foley, V. K., Mimnaugh, S., & Safran, J. D. (2020). The impact of alliance-focused and facilitative interpersonal relationship training on therapist skills: An RCT of brief training. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, 30(7), 871–884.

<https://doi.org/10.1080/10503307.2020.1722862>

Piper, W. E., Azim, H. F., Joyce, A. S., McCallum, M., Nixon, G. W., & Segal, P. S. (1991). Quality of object relations versus interpersonal functioning as predictors of therapeutic alliance and psychotherapy outcome. *The Journal of Nervous and Mental Disease*, 179(7), 432–438. <https://doi.org/10.1097/00005053-199107000-00008>

Reise, S. P., & Duan, N. (1999). Multilevel modeling and its application in counseling psychology research. *The Counseling Psychologist*, 27(4), 528–551. <https://doi.org/10.1177/0011000099274003>

Rogers, C. R. (1959). A theory of therapy, personality, and interpersonal relationships as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3., pp. 184-256). McGraw-Hill.

Roth A., & Fonagy, P. (1996). *The relationship between outcome and therapist training, experience and technique*. In A. Roth., & P. Fonagy (Eds.) *What works for whom? A critical review of psychotherapy research*. Guilford

- Safran, J.D., & Muran, J.C. (2000). *Negotiating the therapeutic alliance: A relational treatment guide*. Guilford Press.
- Saxon, D., Ashley, K., Bishop-Edwards, L., Connell, J., Harrison, P., Ohlsen, S., Hardy, G. E., Kellett, S., Mukuria, C., Mank, T., Bower, P., Bradburn, M., Brazier, J., Elliott, R., Gabriel, L., King, M., Pilling, S., Shaw, S., Waller, G., & Barkham, M. (2017). A pragmatic randomised controlled trial assessing the non-inferiority of counselling for depression versus cognitive-behaviour therapy for patients in primary care meeting a diagnosis of moderate or severe depression (PRaCTICED): Study protocol for a randomised controlled trial. *Trials*, *18*(1), 93. <https://doi.org/10.1186/s13063-017-1834-6>
- Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology*, *80*(4), 535–546. <https://doi.org/10.1037/a0028898>
- Schauenburg, H., Buchheim, A., Beckh, K., Nolte, T., Brenk-Franz, K., Leichsenring, F., Strack, M., & Dinger, U. (2010). The influence of psychodynamically oriented therapists' attachment representations on outcome and alliance in inpatient psychotherapy. *Psychotherapy Research*, *20*(2), 193–202. <https://doi.org/10.1080/10503300903204043>
- Segal, D. L., Coolidge, F. L., Cahill, B. S., & O'Riley, A. A. (2008). Psychometric Properties of the Beck Depression Inventory—II (BDI-II) Among Community-Dwelling Older Adults. *Behavior Modification*, *32*(1), 3-20. <https://doi.org/10.1177/0145445507303833>
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Stiles, W. B., Honos-Webb, L., & Surko, M. (1998). Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice*, *5*(4), 439-458. <http://dx.doi.org/10.1111/j.1468-2850.1998.tb00166.x>

- Uhlen, B. (2011). *Therapist in-session rated facilitative interpersonal skills (FIS-IS) in the psychotherapy process* [Unpublished doctoral dissertation]. Ohio University. Retrieved from: https://etd.ohiolink.edu/pg_10?::NO:10:P10_ETD_SUBID:61851
- Uhlen, B., Anderson, T., & MacFarlane, P. (2010). *The Facilitative Interpersonal Skills In-Session Rating Manual*. [Unpublished Manuscript], Ohio University.
- Wampold, B. E. (2001). *The great psychotherapy debate*. Lawrence Erlbaum Associates.
- Wampold, B. E., & Bolt, D. M. (2006). Therapist effects: Clever ways to make them (and everything else) disappear. *Psychotherapy Research*, 16(2), 184–187. <https://doi.org/10.1080/10503300500265181>
- Wampold, B. E., & Brown, G. S. (J.). (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, 73(5), 914–923. <https://doi.org/10.1037/0022-006X.73.5.914>
- Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*. Routledge.
- Wang, Y.P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. *Revista Brasileira de Psiquiatria*, 35(4), 416-431. <http://dx.doi.org/10.1590/1516-4446-2012-1048>
- Zilcha-Mano, S., & Errázuriz, P. (2015). One size does not fit all: Examining heterogeneity and identifying moderators of the alliance-outcome association. *Journal of Counseling Psychology*, 62(4), 579–591. <https://doi.org/10.1037/cou0000103>

Appendix A

Outline of two FIS research projects and distinct aims

AF and CB completed a shared project focusing on therapists' use of Facilitative Interpersonal Skills (FIS) in "not on track" sessions using the FIS rating scale. The shared part of the project was the identification of the "not on track" sub-sample and the subsequent rating of the sessions using the FIS. The individual projects had different aims and proposed analyses (see Table 1a) and consequently differing clinical implications.

Table 1a

Outline of project differences

	AF	CB
Aims	1. To investigate the relationship between therapy/therapist variables and FIS. 2. To investigate the relationship between client variables and FIS. 3. To consider the stability of therapists' interpersonal communication style across a session and between clients.	1) To assess if therapists' facilitative interpersonal skills during NOT client sessions are predictive of client outcome. 2) To consider interactions between treatment modalities and therapist skills by assessing if facilitative interpersonal skills are more important for client outcome in PCET or CBT.
Analyses	Exploratory non-parametric analyses between therapist and client variables and FIS Exploratory non-parametric and descriptive analyses examining FIS variability	Multiple regression

Appendix B

Ethical Approval Letter



Downloaded: 07/02/2020
Approved: 31/01/2020

Annie Faulds
Registration number: 180157066
Psychology
Programme: Doctorate in Clinical Psychology

Dear Annie

PROJECT TITLE: An examination of the relationship between client and therapist variables and therapists Facilitative Interpersonal Skills (FIS) during not on track therapy sessions.

APPLICATION: Reference Number 032726

This letter confirms that you have signed a University Research Ethics Committee-approved self-declaration to confirm that your research will involve only existing research, clinical or other data that has been robustly anonymised. You have judged it to be unlikely that this project would cause offence to those who originally provided the data, should they become aware of it.

As such, on behalf of the University Research Ethics Committee, I can confirm that your project can go ahead on the basis of this self-declaration.

If during the course of the project you need to [deviate significantly from the above-approved documentation](#) please inform me since full ethical review may be required.

Yours sincerely

Thomas Webb
Departmental Ethics Administrator

Appendix C

(Removed for copyright reasons)

Appendix D

FIS Rating Scale (Anderson & Patterson, 2013)

Ratings for each item are made on a 5 point Likert-type scale. The rating scale ranges from Not Characteristic ("1") to Extremely Characteristic ("5"). More elaborate descriptions of each rating level are provided to help identify the correct rating level.

Response Set: Clearly, people differ in how they evaluate the skillfulness of helper interventions. The descriptions of these items are lengthy in order to provide the context for what is intended by each particular FIS domain / item. Items are written in order to assist the rater in using specific evidence from the tape to inform rating decisions. In rating a response, instances of an "average" level of a helping behavior would merit a rating of 3. Ratings of 3 are thought of as the default rating for all items and are considered ordinary helping or facilitative interpersonal skills. Thus, a 3 should be the starting point for ratings and the participant's response may influence you to increase or decrease your rating from a 3.

Item One: Verbal Fluency

This item is a rating of the extent to which the participant is verbally capable and at-ease in communicating. The response is delivered in a relaxed manner and without significant signs of anxiety (e.g., broken speech, extended and awkward pauses, and clarity in communication). However, the content of what is said is not rated, but rather how it is spoken.

- 5 The participant is at great ease and communicates ideas with no anxiety, reflecting a desire to "approach" the other. The verbal quality of the response may have a "melodic," rhythmical quality and is easy to follow; the response is fluent.

- 4 The response is fluent, and there is little that is difficult to follow.
- 3 A moderate level of verbal fluency indicates that the participant's response is conversational and mostly easy to follow.
- 2 Fluency is disrupted by the participant's anxiety and avoidance of verbal expression. The respondent may be obviously anxious and struggling to formulate a response. At times, the communication may be choppy, even halting.
[Note: In some rare instances a response could represent an avoidance of the interpersonal situation through anxious rambling. It would need to be clear that the participants' ramblings are the result of anxiety over communicating with another].
- 1 The participant has great difficulty verbalizing a response (e.g., obviously anxious, sounds shaky or timid), reflecting a clear avoidance or anxiety. The participant may lack confidence and is clearly uncertain or even difficult to follow.

Item Two: Hope & Positive Expectations

This item rates expressions of hope, optimism, and positive expectations for change.

Staats (1989, 2001) defines hope as the interaction between wishes and expectations.

The interpersonal skills needed for hope involve facilitating a) personal agency and b) building the pathways needed for attaining desired goals and expectations (Steed, 2002).

Hope is related to persuasiveness and collaboration in the sense that hope and positive expectations are often built through offering a rationale, friendliness, and enthusiasm. As defined here, hope focuses more on building client agency for actions that will facilitate meeting the client's goals whereas persuasion is based more on a plausible explanation (which may or may not include hope).

- 5 The participant's response expresses clear hope about the client's future and/or positive expectations about therapeutic work. In addition, for a response to be

coded as a “5” there needs to be an allusion to building the client’s agency as well as how the client might participate or do something that will help move toward his/her desired goals (i.e., pathways).

- 4 A general sense of optimism about the client’s situation is detected. Specifically, the participant's response is directed toward building the client’s agency OR facilitating the building of pathways to meet the client’s goals.
- 3 The response is ordinary OR the optimism of the response is not discernible. There may be some hopefulness expressed, but with little confidence or reason for being hopeful.
- 2 The participant responds with some hopelessness, including subtle expressions of feeling unable to help the client.
- 1 The participant's response is hopeless or is even pessimistic. For example, the participant may address only issues or concerns beyond the control of the other or subtly suggests that the other cannot change or improve his/her problems.

Item Three: Persuasiveness

Persuasiveness is the capacity to induce the other to accept a view that may be different from his or her own view. It involves that ability to convey a clear, organized understanding about the meaning of the other’s source of distress. Persuasiveness implies an ability to communicate what Jerome Frank called a “believable myth.” This capacity implies that the persuasive therapist must be convincing in communicating this belief-system.

Rating Notes: For rating purposes, the response does not necessarily need to convey an entire world view, but a point of view that is implied to be at least slightly different from the client in the video clip. High ratings require that the participant provide a clear belief in a point of view or rationale. It is necessary that the rationale be relevant to the other’s

problems and at least somewhat novel to the other's experience. For this item, the rater should disregard personal beliefs about the validity of the participant's rationale, but instead rate the extent to which the participant might persuade another (i.e., ability to "sell" their rationale).

- 5 The participant is highly persuasive. Persuasive persons may speak with great confidence, certainty, and authority. Advice may or may not be given, but the participant must offer some rationale or re-framing of the other's experience.
- 4 The participant speaks persuasively. The rationale may be more implicit and it is even possible that the rationale, though present, may be unclear, superficial, or marginally relevant to the other's problems.
- 3 The participant's response conveys little sense of persuasiveness.
- 2 The participant is unpersuasive. Unpersuasive responses may be characterized by either
 - a) a rationale that lacks credibility and there is little reason to believe that it could be convincing. It is important here to try to be aware of your personal biases in judging credibility. As a rule, you can accept most explanations offered as being credible unless there is a clear logical flaw in the process of explaining their particular belief.
 - b) a response that is expressed with a lack of confidence, lethargy, or uncertainty by the respondent will be low in persuasiveness. Even responses that don't contain a rationale may be coded as low in persuasiveness
- 1 The participant's response is unorganized, incoherent, and difficult to follow. The participant may also not know what to say.

Item Four: Emotional Expression

This item rates the energy and emotion in the participant's response. This item rates the extent to which the participant's response is delivered with effective expressions of emotion.

- 5 There is affect and prosody in the participant's voice. The response is delivered in a highly emotional and engaging manner. The primary criterion is that the vocal expression conveys emotion. There may be a more focused delivery of emotional intonations to emphasize meanings that influence other processes (e.g. persuasion). The participant may even be somewhat provocative or challenging in delivering an emotion-based response toward the client in the video clip. However, a "5" should not be rated if the affect is primarily demeaning or hostile toward the other (in which case a "3" would be the maximum rating possible).
- 4 The participant is emotionally expressive at a moderate level. There is more emotion than found in ordinary speech, but it is not as focused in its delivery as the maximum rating of 5.
- 3 The participant has prosody, but it is the amount of emotion that one might find in ordinary conversation.
- 2 The participant may display some sense of interest or curiosity, but the response is not emotionally engaging. Prosody is somewhat less than typical to casual conversation.
- 1 The participant speaks with little or no affect and may be dull or boring (e.g., speaking in a near monotone voice and without intensity).

Item Five: Warmth, Acceptance & Understanding

This item is a rating of the ability of the participant to care for and accept the other.

Therapist behaviors/attitudes that might indicate an absence of acceptance and understanding include: a judgmental attitude, condescension, rudeness, disapproval, guilt-

induction, exasperation, or annoyance. Often it will be necessary to avoid rating what the participant is doing (e.g., giving advice), but rate how it is being done. Note that accepting does not necessarily mean approval, but rather a caring attitude and determination to help the other.

- 5 The participant expresses clear and obvious warmth, concern and acceptance. The participant may, for example, make a compassionate attempt to relate to the other's experience.
- 4 The participant's response is genuinely nonjudgmental and gently explores the other's thoughts, feelings, alternatives for dealing with future situations, etc. The participant appears concerned for and respectful of the client.
- 3 There is an "ordinary" level of courtesy and warmth in the response OR the participant's opinion of the other may not be clearly discernable from the response.
- 2 The participant conveys a subtle lack of respect, acceptance, or concern of the other (e.g., sarcasm, exasperation, annoyance).
- 1 The participant has an obvious lack of respect, acceptance, or warmth for the other (e.g., clearly pejorative comments, judgmental attitude, condescension, disapproval, guilt induction, blaming the other).

Item Six: Empathy

The capacity to respond with an expressed understanding of the subjective experience of the client. The response must also convey an accurate understanding of the thoughts and emotions expressed in the video clip. Therefore, it is especially important that the rater have an accurate understanding of the client's experiences in the video clips.

- 5 Participant alludes to the client's experience so that it is clear that he/she has not only listened, but obtained an exceptional comprehension of what the other is

experiencing. In order to receive a "5" the participant must infer something about the other's experience that is not explicitly stated by the other but is clearly identifiable in the client's nonverbal expression.

- 4 Participant comments accurately on the other's experience but not to the extent required to receive a "5" rating. The distinction between the 4 and 5 ratings are matters of intensity.
- 3 Participant is generally accurate about the other's experience but only perceives the more obvious aspects of the other's experience or concerns.
- 2 Participant does not communicate an awareness or understanding of the other's experience, and/or there are minor distortions of the other's experience. Some aspects of the participant's response may be irrelevant to the other's concerns (when clearly not an attempt to change the other's focus).
- 1 Participant clearly distorts the other's experience. That is, the participant misidentifies a significant component of the other's complaints, beliefs, emotions, etc. Give a rating of 1 if the response indicates a clear disregard of the other's experience.

Item Seven: Alliance Bond Capacity

This item rates the participant's capacity to provide a collaborative environment, one in which there is recognition of the need to work with the client jointly on problems.

- 5 Specific actions on the part of the participant help create a collaborative atmosphere. There should be a sense that the participant is attempting to work with the other to create a "we-ness" that is implied in the participant's behavior (e.g., participant checks with the other by asking questions about the "fit" of interpretations, conclusions, goals, etc.).

- 4 Some effort to collaborate is made but not as strong as a “5” (e.g., subtle invitations to engage in working with the client).
- 3 The participant neither undermines nor attempts to enhance a collaborative effort.
- 2 The participant may slightly undermine the building of a collaborative atmosphere, although it may be unintentional or superficial.
- 1 The participant actively undermines a mutual collaboration. The participant may respond in a way that is over-involved or reactive (e.g., moralistic lecturing, "preaching" to the other, assuming all responsibility). The rupture may also involve withdrawal or under-involvement in the participant's response (e.g., putting all the responsibility for change on the other).

Item Eight: Alliance Rupture-Repair Responsiveness

Background: Each client in video clips is expressing an interpersonal issue that involves the patient-therapist relationship. Each video clip places the participant in the middle of alliance rupture episodes. Further, these rupture episodes take place at different locations within the interpersonal circle, which requires interpersonal flexibility for the therapist. The interpersonal problem with the client-therapist relationship threatens the development of the alliance.

This item rates the extent to which the therapist appears responsive to the interpersonal issue. In some cases, the problem is clearly stated as when Suzie angrily berates the therapist for being ineffective. In other cases, the problem is more implicit such as when Lauren idealizes the therapist to the extent of leaving herself overly vulnerable to disappointment.

- 5 Participant makes attempts repair the interpersonal issue by engaging the client in a direct discussion of the immediate moment-to-moment interaction. This may

include how specific relational messages are being expressed by the client in the video clip. Optimal responses will include attempts to not only allude to the interpersonal tension, but make some attempt to repair that interpersonal issue.

- 4 Participant recognizes the other's interpersonal issue, and may discuss this further in more general terms (or discuss some secondary element of the other's issue or the relationship).
- 3 There may be more casual recognition of the interpersonal situation, but the response does not draw for further discussion of the issue or the relationship.
- 2 Participant addresses an issue to discuss that it is tangentially related to the interpersonal issue presented, but directs the discussion away from the present relationship situation.
- 1 Participant reacts to the interpersonal tension in a way that is nonproductive or in a way that likely exacerbates the rupture (e.g., responding negatively to a hostile client; responding to a controlling client with counter-control). Low scores also may be given when the participant fails to respond to the interpersonal issue involved in a way that indicates that the participant is avoiding the interpersonal issue or the relationship altogether.

Appendix E

Problem Responsiveness FIS Domain Scale (De Jong et al., unpublished)

Problem Responsiveness Item

Item: This item measures the extent to which the therapist appears to offer solutions to the specific problem AND whether said solutions are likely to be helpful. One must note that the suggestions offered do not need to be direct solutions to the problem (advice) but can also be prompts toward the client to help themselves in this situation e.g. encouraging self-reflection or reappraisal of the situation. The rater must have a good understanding of the specific issue each client is presenting in order to accurately rate this item. A perfect response, with a wrongful understanding of the issue at hand cannot lead to a score above two on this item.

5 – The therapist makes direct suggestions in tackling the issue at hand, these can include the direct modification of the issue or encouraging the client to resolve or reappraise the issue themselves. The suggestions are considered feasible for the client to carry out and aren't asking too much of the client, risking further issues. Achieving a 5 on this item can only occur if the therapist displays an understanding of the issue, if not any suggestions cannot be accurate.

4 – The therapist recognizes the issue and discusses solutions in a more general sense or alludes to the client reappraising or resolving the issue themselves, without providing direct guidance on it. The suggestions are considered feasible but may not be able to be easily implemented due to the lack of specific advice or support.

3 – There is recognition of the issue but there is no indication of any tangible solutions to the issue, nor is the client stimulated to search for solutions themselves.

2 – The therapist either ignores the issue or brings up a similar yet unrelated topic. Advice/suggestions unrelated to the issue may be provided.

1 – The therapist responds in a way that is obviously unhelpful to the situation; this could include minimizing the importance of the issue and/or offering a completely infeasible solution to the problem. Similarly, if the therapist actively ignores the problem or shows a willingness to disregard the issue.

Appendix F

(Removed for copyright reasons)

(Removed for copyright reasons)

(Removed for copyright reasons)

(Removed for copyright reasons)

Appendix G

(Removed for copyright reasons)

(Removed for copyright reasons)

Appendix H-

(Removed for copyright reasons)

(Removed for copyright reasons)

Appendix I

(Removed for copyright reasons)

Appendix J

(Removed for copyright reasons)

(Removed for copyright reasons)

Appendix K
FIS Rating Calibration Meeting Minutes
Calibration Meeting 02/11/2020

Identified an issue with timings – agreed that we would start rating the recording from the first moment it is identified the client and therapist are interacting.

ID: XXX

Section 1

Persuasiveness agreed a 4 based on the therapist rationale provided for thoughts and completing homework. For both emotional expression and warmth agreed on the 4 because it feels a bit superficial. We have gone for obvious empathy meaning a score of 3.

Section 2

Drawn out two problems in the section. We have rated the therapist a 5 for the reappraisal responsiveness however we have downrated for the potential rupture. Agreed to capture this in a separate problem responsiveness section. Have agreed on a 4 for warmth because of this incident. Agreed on a 2 for ABC because of the lack of collaboration/forcefulness and a 3 because of obvious empathy.

Section 3

Rated hope as a 5 due to identified pathways and agency. Agreed persuasiveness is a 5. Warmth 4 due to superficial nature. Empathy 4. Decided to separate our PR into problems – the first been reappraisal of the clients problems through thought challenging (scored 5) and the second been the therapists lack of responsiveness/subtle dismissiveness/too much reassurance of the clients concerns/anxieties – scored a 2 for this problem.

ID: XXX

Section 1

Difference in warm – agreed on a 5 based on initiative sense. ABC agreed 5. PR agreed reappraisal is a 5 (agreed that PCET can be scored a 5 if there are direct attempts to reappraise the issue).

Section 2

Agreed VF is a 5. W&A is a 5. ABC agreed a 5. Problem responsiveness is a 5.

Section 3

VF agreed 5. Hope agreed that pathways and agency both tangibly present scoring a 5.

Other queries

AF queried what to do in longer sections when there feels like stark differences between different points in the segment. Agreed that we would aggregate on VF etc., but mark two different scores for PR based on the rupture and repair phases.

CB queried the scoring for empathy when the therapist is building meaning over time – this might mean that questions and summaries in the first sections seem superficial but are more complex towards the end. Agreed this was fine to rate this way and was an interesting point of therapy.

AF queried how to score a therapist when they use buzz words indicating collaboration e.g. we etc. but seem forceful and silencing of the client. Agreed that this justifies a low score.

CB queried how to score hope in PCET when there may not be explicit examples of hope or positive expectations. Agreed to include more subtle indicators of hope e.g. “that is interesting”. If not explicit example it is ok to rate a 3.

AF queried if a therapist is asking exploratory questions that lead a client to a different viewpoint can this be considered persuasiveness? Agreed that if it feels like intentionally directing the client then this could be considered persuasive.

CB queried if can rate a 2 if a therapist does not feel like an active part of the session – just makes simple reflections/questions. Agreed it was ok to rate a 2 if therapist felt under involved.

Calibration Meeting 16/11/2020

ID XXX

Section 1

Hope- Rated as a 3. Agreed that the rating of a 4 was not warranted as therapist does not show enough.

Warmth- Downrated to a 3 as it does not feel enough to warrant a 4

PR- Decided to separate out into Problem 1 (interpersonal) and Problem 2 (Rational approach vs emotional thoughts) so scored 2 and a 3.

Section 2

Empathy- Agreed to rate as a 3 as it just felt mostly the obvious aspects.

ABC- Downrated as 2 because any suggestions made were not done collaboratively and the dominance of client left the therapist under involved

Section 3

Persuasiveness- Agreed to score a 2 because therapist attempts persuasion but this is not received well by client thus ineffective.

Empathy- agreed to go for a 3 in the middle (score of 2 & 4) as therapist attempts empathy but client is the one to take it to the next level

PR- separated out into two PRs. Relational component was a 2 and rated the identification of behavioural patterns as a 3

ID XXX

Section 1

VF- Decided to rate as a 4 because therapist was fluent but maybe not super approaching

Persuasiveness- Example of the therapist using questioning to lead the client to an alternative place or way of thinking so rated a 4

Warmth- Agreed to rate as a 4 because therapist is showing acceptance and understanding

ABC- Agreed on a 5 because the therapist sets a shared agenda, sense of we-ness

Section 2

VF- Agreed on a 2 due to early example of hesitation and possible anxiety.

Hope- Agreed on a 3 because there's not enough evidence for pathways/agency

Persuasiveness- Rated as a 2 because therapist's response lacks a lot energy and has lethargy

EE- Rated as a 2 because there is a distinct lack of energy in the therapist's responses.

Empathy- Agreed on a 4 as there are some examples of the therapist showing an understanding beyond the obvious points the client shares.

PR- Agreed on a 4 that it does superficially work to solve the problem and the passive/lethargic nature of doing this is captured in EE and Persuasiveness

VF- Agreed on a 3 as neither have strong evidence for a 2/5. Agreed that there is a sense of choppiness/therapist anxiety- enough to warrant nothing higher than 3.

Hope- Agreed on 3.5 because some evidence of facilitating pathways to goals but maybe not quite enough for demonstrating optimism about client's experience.

Persuasiveness- Agreed to score a 2 because of how therapist tries to share rationale (with lethargy)

EE- Agreed on 3 due to lack of examples of higher EE

ABC- Agreed on 3 due to lack of true collaboration and slightly prescriptive therapist approach.

Other issues

Discussion about hope. Recognised that for a 5 rating hopefulness and how someone achieves their goals is required whereas only one is required for a 4.

Discussed how to rate when therapist is doing a good job at PR but is missing a wider issue. Ok to score lower when there seems to be an omission.

Discussed a tape where the main problems are around anxiety rather than depression. Discussed that because the trial selects clients based on severe depression scores then this is probably ok. It may be that the focus on anxiety is the reason the client goes off track

Discussed whether being unresponsive to client humour can be rated in empathy. Agreed would be captured under EE and PR.

Discussed situation involving a possible rupture in the eyes of the rater, but the client does not necessarily sound like they find it inappropriate. AF and CB listened to it together and agreed this warranted a separate problem and would be scored a 1.

Calibration Meeting 04/12/2020

ID1XXX

Section 1

Verbal fluency calibrated at a 5 – CB assumptions may have brought this down in the initial section. Hope moved to a 5 – reminder that pathways and agency would constitute a rating of 5. Emotional expression rated 5. Scored Warmth as a 5 – CB not to be held back by “compassionate attempt to relate” in the manual. Agreed PR is a 5 – collaboratively drawing up list of solutions with the client.

Section 2

Agreed persuasiveness would constitute a score of 5. Agreed to score empathy as a 5 – CB discussed two examples demonstrating good levels of empathy. Agreed to separate problems out into two separate PR items based on CB’s distinction. The second problem agreed to rate a 5.

Section 3

Agreed PR is a 5. Warmth rated as a 4 – no evidence of compassionate attempt to relate. Empathy rated a 5 feels that the therapist makes reference to how clients behavior leaves them stuck which goes beyond narrative. ABC rated as a 4 – AF felt there were some examples in the section of a mismatch between the client and therapist. Agreed to separate out PR into 2 separate problems – Problem 1: continuing to explore automatic thoughts and depressed thinking scored 5. Problem 2: homework barriers which we agreed to score a 5.

IDXXX

Section 1

VF scored as a 3.5 – discussed how we are likely to give a higher score because we have a sense that this is a stylistic characteristic of the therapist (familiarity from previous sessions). Rated EE as a 5. Empathy rated as a 4. ABC rated as a 3. Rated PR as a 2 based on evidence from CB.

Section 2

Scored Hope as 4. Rated PR as a 4 based on evidence of pull to provide solutions in CB’s notes.

Section 3

Scored VF 3.5. Listened again to the end of section 3 to rate hope and persuasiveness -agreed with AF's observation about the spirit in which the comments from the therapist were received by the client rather than a focus on the content – rated 2.5. Rated EE as a 4, rated warmth as a 4. Agreed empathy would be a 5. Agreed with AF's observations that that PR would be based on the partial rupture at the end of session – decided that although the therapist addressed some of the issues it felt as though the client will be left with some hopelessness at the end of the section – rated 3.5.

Other issues

AF raised a moment where client becomes tearful and therapist does not acknowledge directly but is generally warm. Feels missed opportunity, so rated PR as a 4, otherwise would be a 5. Therapist also switches off the tape before the end of the session when planning endings –query regarding potential for rupture. Agreed to mark what is present.

AF asked if working towards goals count as hope – agreed it did and reflected in score given. Therapist highlights a vicious cycle in the client's narrative, agreed this could be both an example of empathy, but also part of PR.

CB raised an issue regarding PR - in PCET agreed that being alongside the client can warrant a higher score for PR, when it feels like reappraisal or solution giving would not be responsive to the client need.

Discussed our own biases against a diagnostic model as CPs.

CB discussed incident when it felt as though the therapist disregards clients wishes not to complete measures – agreed can be raised in PR as well as ABC.

AF highlighted a good example of using an analogy for counselling, agreed can be considered persuasiveness.

AF raised an incident in which the client is tearful and therapist quickly moves on. Happens in two sections. Agreed to mark it as an omission in the PR section of this tape.

AF raised an incident in which the therapist is facilitating exploration of emotions but not reappraisal yet, agreed can be higher if this feels as though it is what the client needs at the time (PCET).

AF raised an incidence when the therapist is actively making suggestions but it feels as though it is more surface level, feel as though misses something deeper. Hope being created – because of pathways.

Calibration Meeting 18/12/2020

IDXXX

Section 1

Verbal Fluency- both agreed on a 2 due to some initial halting

Warmth- Agreed on a 4 due to some evidence of gentle exploration of client's thoughts

ABC- Agreed that due to the therapist being under involved this can be a 2

PR- Separate out into two problems (fear of rejection/abandonment- scored as a 3 and then interpersonal issue of client interrupting and the therapist not addressing this- scored as a 2)

Section 2

Empathy- Agreed on a 4 due to the therapist reflecting back on the client's fear of abandonment and how this impacts on behaviours

ABC- Agreed that due to under involvement this can be a 2

PR- Separated out into two problems (fixed and self-critical beliefs which are left unchecked- 3 and client continuing to interrupt- 3)

Section 3

Warmth- Agreed on a 2 due to therapist showing signs of feelings exasperated

PR- Separate out into two problems (therapist trying to offer something about thought challenging which is unhelpful-2; interpersonal issue of interrupting client-2)

ID XXXSection 1

Warmth- Agreed on a 4 because some examples of warmth

ABC- Agreed on a 3 because not much evidence of it

Section 2

Persuasiveness- not enough to warrant more than a 3 so agreed on this. Although a link was made it is not enough for persuasiveness

EE- Agreed on a 2 because therapist presented flat in expression

Warmth- Agreed on a 3 because there isn't enough feeling/ warmth here

Empathy- Agreed on a 4 as both teetering on this and because persuasiveness has not been rated up for this instance

ABC- Agreed on a 3 as the therapist does offer something so maybe not as harsh as a 2.

PR- Agreed on a 2 as both feel that they are really under involved and ignoring the issues

Section 3

Hope- Agreed that we can rate the omission of any action from the therapist here as tipping into hopelessness so agreed on a 2

EE_ Both agreed on a 2 because of the reasons outlined in other sections

Warmth- Agreed that the therapist's absence from the session is communicating a lack of concerns and this can be rated a 2

ABC_ Agreed that the under involvement warrants a 2 here.

PR- Agreed that we can separate out into two problems (problem 1 is the issues around work and not offering any solutions- scored 2, problem 2 is the therapist not engaging at all with the client raising stressful life events- scored 2)

Appendix L

FIS Distribution

