



Development and validation of the Gestalt Therapy Fidelity Scale

Madeleine Fogarty, Sunil Bhar & Stephen Theiler

To cite this article: Madeleine Fogarty, Sunil Bhar & Stephen Theiler (2019): Development and validation of the Gestalt Therapy Fidelity Scale, Psychotherapy Research, DOI: [10.1080/10503307.2019.1571688](https://doi.org/10.1080/10503307.2019.1571688)

To link to this article: <https://doi.org/10.1080/10503307.2019.1571688>



Published online: 03 Feb 2019.



Submit your article to this journal [↗](#)



Article views: 181



View Crossmark data [↗](#)

Development and validation of the Gestalt Therapy Fidelity Scale

MADELEINE FOGARTY, SUNIL BHAR, & STEPHEN THEILER

Department of Psychological Sciences, Swinburne University of Technology, Hawthorn, VIC, Australia

(Received 12 April 2018; revised 24 December 2018; accepted 2 January 2019)

Abstract

Objective: Three studies were conducted to develop and validate the Gestalt Therapy Fidelity Scale (GTFS), a 21-item measure of treatment adherence for Gestalt Therapy (GT).

Method: Thirty five items for possible inclusion in the GTFS were generated on the basis of a literature review. In Study 1, a Delphi methodology consulting 63 international GT experts was used to select items for the GTFS. In Study 2, six experts used the scale to rate video-based sessions of GT, and provided feedback on the usability of the scale. In Study 3, 176 participants from 18 countries used the GTFS to rate GT and not-GT video recorded sessions.

Results: The Delphi study consensus method resulted in 25 items for consideration in the GTFS. The scoring system and items were subsequently revised following further feedback from experts (Study 2). The GTFS was found to significantly discriminate between GT and not-GT based sessions (Study 3): raters scored GT sessions significantly higher than not-GT sessions. High levels of internal and inter-rater reliability were found.

Conclusion: The GTFS is supported as a psychometrically sound measure of treatment adherence for GT, and hence can be used to assess the degree to which therapists are administering GT.

Keywords: psychoanalytic/psychodynamic therapy; test development; Gestalt therapy; treatment fidelity

The ability to assess treatment fidelity is an essential precondition to conducting research into the efficacy of a psychotherapeutic modality (Perepletchikova, 2011; Waltz, Addis, Koerner, & Jacobson, 1993). Without an adequate treatment fidelity measure, it is impossible to know whether the treatment delivered by a therapist was consistent with the therapeutic modality, and it cannot therefore be established that any clinical results can be attributed to that modality. The development of treatment fidelity measures has been recognised as a crucial step in the advancement of research and clinical practice (Fairburn & Patel, 2014; Stiles, Hill, & Elliot, 2015).

Gestalt Therapy

Gestalt Therapy (GT) is still a widely practiced form of psychotherapy. For example, among the 5732 holders of the European Certificate of Psychotherapy listed on the European Association for Psychotherapy (EAP)'s website in February 2018, 949 described

their modality as one of Gestalt Therapy, Integrative Gestalt Therapy or Gestalt Theoretical Psychotherapy, compared to only 262 who described their modality as Psychodynamic or Psychoanalytical Psychotherapy, and 869 who took a multi-modal approach (EAP, 2018).

GT is a dynamic, integrative, embodied approach that addresses human existence as a fundamentally relational modality in which the experience of self is situated in a dynamic organism-environment field (Bloom, 2016; Francesetti, 2015; Wheeler & Axelson, 2015). The cornerstone of GT is to increase awareness and thereby awaken new possibilities for growth (Yontef, 1993). GT focuses on direct experience, rather than secondary interpretation. It is a hermeneutic phenomenological approach focused on the present moment and anchored in a deeply dialogic relationship (Jacobs & Hycner, 2010). GT was first presented in writing through the text *Excitement and Growth in the Human Personality* (PHG: Perls, Hefferline, & Goodman, 1951). Since the publication of

Correspondence concerning this article should be addressed to Madeleine Fogarty, Department of Psychological Sciences, Swinburne University of Technology, Hawthorn, VIC 3122, Australia. Email: madeleine@madeleinefogarty.com

This article has been republished with minor changes. These changes do not impact the academic content of the article.

PHG, GT has been written about extensively and practiced diversely across numerous countries (O'Leary, 2013). GT has also influenced other therapeutic modalities, including Process Experiential Therapy (Greenberg, Rice, & Elliott, 1993), Dialectical Behaviour Therapy (Linehan, 2000), Interpersonal Psychotherapy (Weissman, Markowitz, & Klerman, 2000), Emotion Focused Therapy (EFT) (Greenberg, 2010), Schema Therapy (Young, Klosko, & Weishaar, 2003), and Pluralistic Psychotherapy (Cooper & McLeod, 2012).

GT does not have a generally accepted treatment manual, and can be tailored to different clinical populations (Francesetti, Gecele, & Roubal, 2013; Greenberg, 2016; Roubal & Rihacek, 2016; Roubal, Francesetti, & Gecele, 2017; Spagnuolo Lobb, 2013; Yontef, 1993), in a variety of clinical settings (Wheeler & Axelsson, 2015). Research in GT has increased over the last decade, particularly practice-based and phenomenological research (Barber, 2006; Brownell, 2008, 2014, 2016; Butollo, Koenig, Karl, Henkel, & Rosner, 2014; Elliott & Hendricks, 2017; Goss & Stevens, 2016; McConville, 2014; Meara & Fogarty, 2017; Roubal, 2016; Stevens, Stringfellow, Wakelin, & Waring, 2011). However, research into the efficacy of GT has been hampered by the lack of a treatment fidelity measure (Fogarty, Bhar, & Theiler, 2015). The establishment and adoption of treatment fidelity measures for newer psychotherapeutic modalities, such as Acceptance and Commitment Therapy, has been cited as one of the major reasons for the increased acceptance and adoption of those modalities (Smout et al., 2010; Smout, Hayes, Atkins, Klausen, & Duguid, 2012).

Treatment Fidelity Measures

Many treatment fidelity measures rely on the assessment of clinical sessions by an independent rater. Hence, items of such measures must be capable of verification by the independent rater; for example, instead of requiring a rater to assess whether a therapist possesses a quality such as empathy, a measure can ask the rater to look for observable behavioural correlates of empathy (Watson & Prosser, 2002). Treatment fidelity measures that are designed to be applied by independent raters typically therefore comprise items describing observable therapist behaviours. Waltz et al. (1993) recommended that such items be assessed through the analysis of audio or video recordings of clinical sessions. There is evidence that treatment fidelity can be reliably assessed on the basis of relatively short (15 min) extracts of sessions (rather than requiring the assessment of

entire or multiple sessions) (Luborsky, McLellan, Woody, Brian, & Auerbach, 1985; Weck, Bohn, Ginzburg, & Stangier, 2011). This is important, because there is also evidence that the longer it takes a rater to apply a treatment fidelity measure, the less likely the measure is to be adopted (Weck et al., 2011; Weck, Grikscheit, Höfling, & Stangier, 2014).

Treatment fidelity has three dimensions: treatment adherence, therapist competence and treatment specificity (Barber, Liese, & Abrams, 2003; McGrath, 2009; Perepletchikova, 2011). Treatment adherence refers to the extent to which the therapist delivered the treatment in accordance with its design, manual or theoretical model: were the behaviours exhibited by the therapist consistent with the treatment model? Competence refers to the therapist's skill, knowledge and judgment in delivering the treatment. A significantly higher level of rater expertise is required to make an assessment of treatment competence than to make an assessment of treatment adherence (Weck, Hilling, Schermelleh-Engel, Rudari, & Stangier, 2011). Treatment specificity is concerned with ensuring that treatments under investigation can be distinguished from each other (Grikscheit et al., 2015). By default, a treatment adherence measure will also function as a measure of treatment specificity (Perepletchikova, 2011). Some treatment fidelity measures address adherence and competence in a single scale (Barber et al., 2003; Denton, Johnson, & Burleson, 2009; Freire, Elliott & Westwell, 2014); others have separate scales for adherence and competence (Gutermann et al., 2015; Segal, Teasdale, Williams, & Gemar, 2002).

Item Generation

In the absence of a generally accepted treatment manual, the current series of studies aimed to develop and evaluate a measure of treatment adherence for GT. Item generation was based on two steps.

First, based on an extensive literature review (Fogarty, Bhar, Theiler, & O'Shea, 2016), the researchers identified eight concepts that appeared to be collectively capable of accounting for GT's theoretical foundations and methods. Those concepts were: developing awareness, dialogic relating, working in the here and now, phenomenological practice, working with embodied awareness, field sensitive practice, working with contacting processes, and experimental attitude.

Second, each of these concepts, except for "developing awareness" was operationalised by behavioural descriptions. The concept of developing awareness appeared to reflect an aim common to the other

concepts. That is, developing awareness appeared to be an underlying aim of each of the other seven concepts and of all of the operationalisations. For that reason, developing awareness was not used as an organising category for the operationalisations. The seven remaining concepts were operationalised with four to seven descriptions each, resulting in a total of 35 descriptions, with each ranging in length from eight to 23 words.

Overview of Studies

The specific aim of the current research was to develop a measure of treatment adherence for GT and to assess its content validity, face validity, criterion validity and reliability. Three studies were conducted. Study 1 focussed on assessing the content validity of the measure. It employed a “Delphi” methodology to examine if the concepts and items describing therapist behaviours accurately and adequately reflected GT’s theoretical foundations and clinical practices. Study 2 assessed the face validity and clarity of the descriptions of the therapist behaviours in the measure, and whether those behaviours were capable of being observed by an independent rater. Study 3 involved an evaluation of the criterion validity and reliability of the measure.

Ethics

Protocols and procedures for the three stages of the project were approved by the Swinburne University human research ethics committee. Participants in each of the studies provided informed consent before participating.

Study 1

The aim of Study 1 was to assess the content validity of a measure of treatment adherence for GT: the Gestalt Therapy Fidelity Scale (GTFS). The research question examined in this study was “Do the items in the GTFS accurately and adequately reflect the fundamental clinical practices of GT?”

Method

Delphi method. Study 1 used the Delphi method, which is a protocol for establishing an expert consensus. A panel of experts engage in an on-going process in which feedback and results from one round are integrated into subsequent rounds until a consensus of 80% or more agreement is reached (de Villiers, de Villiers, & Kent, 2005).

Participants. Potential participants in the study were identified on the basis that they had edited one of the three major Gestalt journals, had published one or more books or articles about GT, and/or had been a trainer or director at a GT training centre and were able to complete study tasks in English. A further consideration in selecting panel members was that the panel would reflect the diversity and geographic spread of GT (Fogarty et al., 2016). More than one hundred and sixty people were invited to participate in the study. Sixty-three experts, from 17 countries, accepted the invitation (58 participated in both rounds of the study, two participated in only the first round, and three participated in only the second round). There were 24 participants from Europe, 22 from North America (USA and Canada), 11 from Oceania (Australia and New Zealand), four from Latin America, and one each from Israel and Japan. Thirty-five of the participants were male, 28 were female. No age data was collected. The 63 participants had practised as Gestalt therapists for an average of 26.8 years ($SD = 10.9$); 60 of the participants were, or had been, Gestalt trainers and had held these positions for an average of 21.2 years ($SD = 10.3$); 52 of the participants had published books or peer-reviewed articles about GT, and had an average of 9.7 such publications ($SD = 6.9$); 35 of the participants were or had been directors of Gestalt training centres, and had held these positions for an average of 11.6 years ($SD = 6.2$); and 21 of the participants were or had been editors of Gestalt journals, and had held these positions for an average of nine years ($SD = 5.4$).

Measures. The survey was administered in two rounds. The first round comprised questions about the eight concepts identified in the literature review. For each concept, participants were presented with:

- A label or descriptor for the concept (for example, “developing awareness”, or “experimental attitude”);
- A description of the concept’s theoretical foundations and clinical manifestations; and
- A list of associated items describing therapist behaviours (for all of the concepts apart from “developing awareness”; developing awareness was not used as an organising category for the operationalisations because the literature review had suggested that developing awareness was an underlying aim of each of the other seven concepts and of all of the therapist behaviours).

For each concept, participants were asked a combination of qualitative and quantitative questions. The response option for the quantitative questions was

based on a 5-point Likert scale: strongly agree, agree, neither agree nor disagree, disagree, strongly disagree. The questions were:

- (1) Do you agree that the concept is fundamental to Gestalt therapy? (Likert scale).
- (2) Can you think of a more accurate label/descriptor for this concept? (Qualitative).
- (3) Do you agree that the description of the concept is an accurate description? (Likert scale).
- (4) Can you suggest a more accurate way to describe the concept? (Qualitative).
- (5) Do you agree that the following therapist behaviours reflect the concept (as described in the survey)? (Likert scale for each behaviour).
- (6) Can you suggest any modifications or additions to these therapist behaviours? (Qualitative).

The first two questions were asked immediately after participants were presented with the label or descriptor for the concept. The next two questions were asked after participants were presented with the description for the concept. The fifth question was asked immediately after each of the therapist behaviours; and the sixth question was asked after participants had been presented with all of the behaviours for the concept.

Twenty-three of the descriptions of therapist behaviours achieved consensus in the first round (defined as 80% or more of participants agreeing or strongly agreeing with the question). Two of those descriptions were combined into a single description: The therapist “makes observations and enquires about the client’s embodiment” and “... makes observations about the client’s breathing” became “The therapist makes observations and enquires about the client’s embodiment (including breathing)”. Nine items were eliminated, on the basis that they did not achieve consensus and the qualitative feedback suggested that there were substantive disagreements within the expert panel that were likely to prevent them from doing so: the therapist “relates subjectively to the client (sharing themselves transparently in the relationship)”, “... offers self-disclosure that relates to the client’s experience”, “... supports the client, when the client refers to past events or to the future, to connect these to his or her own present experience”, “... shares his or her own somatic and sensory experiences with the client”, “... shares his or her own embodied experiences with the client”, “... explores how past events and future longings affect present thinking, feeling and behavior”, “... supports the client to embrace the process of change as constant”, “... identifies the client’s patterns of making contact

across the contact continuum including desensitization/hypersensitivity, egotism/spontaneity, introjection/questioning, retroreflection/expressing, projection/owning and confluence/differentiating”, “... identifies when a style of contact is useful or habitual”, and “... identifies the presence of an absence in the contact between client and therapist”. The remaining items, including several of those that had achieved consensus, were revised or refined in light of the participants’ responses.

The 25 resulting descriptions, grouped according to concept, were submitted to the expert panel in the second round of the survey. For each concept, participants were asked (using the same Likert scale as the first round) “Do you agree that the following therapist behaviours reflect the concept?” At the end of the second round, participants were given the opportunity to provide qualitative feedback in response to the question “Would you like to suggest any modifications or additions to these therapist behaviours?” The survey also asked participants to provide demographic information.

Procedure. The study was conducted through an online platform, Qualtrics. Participants received an email inviting them to participate, including a link to the survey that was specific to them. Participants completed the survey independently and in their own time. The first round was open for a period of approximately six weeks. The second round was open for approximately four months. Before each round was closed, reminders were sent to participants who had not started the survey.

Results

In the first round there was a consensus (that is, 80% of the participants either agreed or strongly agreed) that each of the eight concepts was fundamental to GT. There was also a consensus that six of the descriptions of the concepts were accurate (see [Table I](#)). Two descriptions did not achieve an 80% consensus: Working with Contacting Processes (67.7%) and Experimental Attitude (79.7%). All eight of the descriptions were published, together with a summary of the qualitative feedback, summarising the main points of disagreement and difference between participants in relation to the descriptions (Fogarty et al., 2016).

In the second round there was an 80% consensus that 24 of the 25 items reflected one or more of the concepts (see [Table I](#)). A descriptive analysis of the participants’ qualitative responses was conducted in order to identify whether there were any recurrent themes that might indicate that further items

Table I. Percentage agreement in relation to the concepts and associated therapist behaviours of GT.

Concept	Fun (%)	Acc (%)	TBRC (%)
Developing awareness	98.3	87.9	
Dialogic relating	98.3	88.7	
(1) The therapist follows the client attentively, tracking the awareness process and the client's experience, not following a predetermined agenda			91.8
(2) The therapist responds non-judgementally to the client, creating the conditions that allow for the most effective client expression			90.2
(3) The therapist demonstrates a willingness to be uncertain and to work with creative indifference			93.4
(4) The therapist draws on their relationship with the client as the ground for challenge and growth			95.1
(5) The therapist seeks to identify and repair any ruptures in the relationship			95.1
Working in the here and now	96.7	93.3	
(6) The therapist enquires about the client's immediate presentation			96.7
(7) The therapist supports the client to stay with what is happening in the encounter between them, by enquiring and seeking to extend awareness about immediate sensation, affect, cognitions and somatic presentations			98.4
(8) The therapist supports the client to accept and deepen their awareness of their presenting issue rather than trying to change it			96.7
Phenomenological practice	98.3	84.7	
(9) The therapist supports the client to describe and deepen and become more present to their experience			100
(10) The therapist articulates the different perspectives/experience of the therapist and client			86.9
(11) The therapist encourages the client to widen their choices rather than establishing a programme for change			90.2
(12) The therapist shares (where appropriate) his or her own experiences that relate to the client's experience			82.0
Working with embodied awareness	94.9	88.1	
(13) The therapist makes observations and enquires about the client's embodiment (including breathing)			98.4
(14) The therapist invites the client to identify sensations, feelings, emotions, thoughts or images that emerge as a consequence of attending to somatic experiences			98.4
(15) The therapist invites the client to engage with their body through experiment			93.4
Field sensitive practice	94.9	89.8	
(16) The therapist investigates the ground (or context) from which the client's presenting figure emerges			95.1
(17) The therapist supports the client to identify how their perception of their environment and prior relationships and needs organise current experience			96.7
(18) The therapist supports the client to identify the uniqueness of their experience			82.0
Working with contacting processes	100	67.7	
(19) The therapist works with the client's interactional patterns as they emerge between client and therapist			93.4
(20) The therapist and the client identify the figure together			85.2
(21) The therapist co-creates a space in which the client and therapist explore how they are impacting each other			88.5
(22) The therapist identifies experiential processes that have not yet been named or overt and explores the impact of this on their awareness			75.4
Experimental attitude	90	79.7	
(23) The therapist uses material that emerges in the therapeutic encounter as the basis for introducing experiments to develop the client's awareness			98.4
(24) The therapist grades the experiment by eliciting feedback from the client regarding the degree of challenge and support that the client perceives			95.1
(25) The therapist supports the client to integrate learning and awareness that emerges from an experiment			96.7

Note. N = 60 for concepts (first round); N = 61 for therapist behaviours (second round); Fun = Concept is fundamental to GT; Acc = Description of concept is accurate; TBRC = Therapists behaviour reflects concept.

needed to be added to the GTFS in order to ensure that it adequately reflected the clinical practices of GT. No such items were identified.

Discussion

Study 1 provided evidence that the concepts were fundamental to GT, and that 24 of the 25 items

describing therapist behaviours reflect those concepts. One item – “The therapist identifies experiential processes that have not yet been named or overt and explores the impact of this on their awareness” – received less than 80% agreement, and was eliminated (item 22 in Table I). These results provide evidence for content validity for the GTFS comprising these 24 items.

Study 2

Study 2 was a pilot study in which experienced Gestalt clinicians used the GTFS to rate video recordings of clinical work by four therapists. The aims of the study were to assess the face validity and clarity of the items, and to assess the amount of training needed by such clinicians to use the GTFS.

Method

Participants. The sample comprised six participants (100% female), with an average age of 54 ($SD = 2.76$). Participants were therapists who had trained in GT and practised as Gestalt therapists for at least 5 years (mean = 18 years, $SD = 3.10$). Three of the participants were Gestalt trainers, and had been trainers for an average of 10 years ($SD = 5.29$). All of the participants lived in Melbourne, Australia, and were known to author MF and each other.

Measures. The GTFS comprised 24 items describing therapist behaviours (see [Table I](#)). For each item there were three response options: “Yes”, “No” or “Not applicable”. Participants were instructed to mark “Yes” if they observed a behaviour, regardless of how well the therapist performed that behaviour; “No” if they did not observe the behaviour; and “Not applicable” if they thought that there may have been legitimate reasons why the therapist did not engage in the behaviour in that session.

Stimuli. Stimuli were 25 min extracts from videos demonstrating work by three well-known Gestalt therapists (Michael Clemmens, Bob Resnick and Gordon Wheeler), and a well-known therapist from a different modality (Jeff Young, Schema Therapy). Given the evidence suggesting that the longer a rater has to spend reviewing videos, the less likely a measure is to be adopted (Weck et al., 2011; Weck et al., 2014), the aim in developing the GTFS was to develop a measure that could be used to assess relatively short (approximately 25 min) sessions or extracts from sessions that would be likely to showcase the therapeutic modality. The pilot study therefore needed to use recordings of similar length. The extracts included the beginning and end of each session, as well those that reflected the therapist’s approach to the presenting issues in the session.

The Wheeler (2004) and Young (2007) videos were part of the Systems of Psychotherapy Video Series (Washington, DC: American Psychological Association); the Clemmens and Resnick videos were provided to author MF by Clemmens and

Resnick respectively. In order to minimise the amount of time required from each participant in the Pilot Study, the researchers elected to compare the GT videos with only one other modality. Schema Therapy was chosen as the single point of comparison because it is a widely accepted integrative form of psychotherapy that incorporates concepts and approaches from a number of modalities, including GT (Lee, 2018). Although Schema Therapy is usually classified as a cognitive modality (whereas GT is classified as humanistic/experiential), it has some similarities to GT in that it uses relational and experiential interventions including experiment (for example, empty chair work) (Fromme, 2011).

Procedure. Participants were asked to familiarise themselves, in advance, with the GTFS and an article containing descriptions of the fundamental concepts (Fogarty et al., 2016). Participants then met as a group with author MF in a room at Gestalt Therapy Australia, Melbourne and discussed the following questions:

- Are there any words or concepts that you are not familiar with in the GTFS?
- Do you think you understand the therapist behaviours described in the GTFS?
- Do you think you would be able to identify those behaviours by watching a video of a session of a Gestalt therapist working with a client?
- Do you understand how to use the GTFS as a rating tool?

Participants asked questions about the GTFS (see [Appendix A](#)). Participants then watched and rated (in the same room) the selected extracts from the four videos. There was no discussion between participants during this process. After participants rated the videos, author MF facilitated a group discussion, using the following questions:

- Which of the therapist behaviours in the GTFS did you observe?
- Which of the therapist behaviours in the GTFS did you not observe?
- In relation to therapist behaviours that were observed by some but not all participants:
 - Can you [who observed the behaviour] describe what it was that you saw?
 - [Having heard this] Do you [who did not observe the behaviour] now agree that the behaviour was present?
- Can you think of any training or information that might have helped you to use the GTFS?

Results

All participants reported that the training they had received – the article containing descriptions of the fundamental concepts (Fogarty et al., 2016), the instructions provided by author MF, and the answers she had given to their questions – had been sufficient to enable them to use the GTFS. With the exception of the issues discussed below, all participants reported that they had been able to understand the items and to observe behaviours described by those items when watching the videos. However, at least half of the participants said they did not understand one item (item 18 in Table I): “The therapist supports the client to identify the uniqueness of their experience”. They also suggested that two of the items could be improved through minor revision (items 6 and 10 in Table I): the therapist “enquires about the client’s immediate presentation” and “... articulates the different perspectives/experience of the therapist and client”. At least half of the participants also reported that they had found four items to be repetitive or overlapping. These were items 7 and 9: the therapist “supports the client to stay with what is happening in the encounter between them, by enquiring and seeking to extend awareness about immediate sensation, affect, cognitions and somatic presentations” and “supports the client to describe and deepen and become more present to their experience”; and items 10 and 12: the therapist “articulates the different perspectives/experience of the therapist and client” and “shares (where appropriate) his or her own experiences that relate to the client’s experience”. They also reported that item 5 depended on factors that would not necessarily be present in a typical clinical session: “The therapist seeks to identify and repair any ruptures in the relationship”. All participants reported that they found the “Not applicable” option confusing.

In scoring the videos, each item on the GTFS observed was awarded 1 point, with points added to constitute the total score. Participants rated the GT videos higher than the not-GT video. Scores were as follows: Michael Clemmens (mean = 21.3, range = 6, *SD* = 2.42), Gordon Wheeler (mean = 17.8, range = 6, *SD* = 2.17), Bob Resnick (mean = 17.8, range = 6, *SD* = 2.68), and Jeff Young (mean = 5.5, range = 4, *SD* = 1.64).

Discussion

The feedback from participants was used to refine the GTFS: four items were eliminated (items 5, 7, 12 and 18 in Table I): The therapist “seeks to identify and repair any ruptures in the relationship”, “supports

the client to stay with what is happening in the encounter between them, by enquiring and seeking to extend awareness about immediate sensation, affect, cognitions and somatic presentations”, “shares (where appropriate) his or her own experiences that relate to the client’s experience”, and “supports the client to identify the uniqueness of their experience”). Three items were revised (compare items 6, 9 and 10 in Table I and items 5, 7 and 8 in the GTFS respectively). The item “The therapist enquires about the client’s immediate presentation” was revised to “The therapist enquires about the client’s immediate experience”. The item “The therapist supports the client to describe and deepen and become more present to their experience” was revised to “The therapist supports the client to describe, deepen and become more present to their immediate sensation, affect, cognitions and/or somatic presentations”. The item “The therapist articulates the different perspectives/experience of the therapist and client” was revised to “The therapist describes and validates the different experiences of the therapist and client”.

The “Not applicable” option was replaced by an item that allowed raters to indicate and describe the presence of “significant unusual factors” that “justified the therapist not engaging in the behaviours described in this scale”. If raters indicate “Yes”, they were asked “what were those factors, and what departures did they justify?” (see Appendix B). For example, in rating a session where the client appeared to be highly shame bound or traumatised, a rater might have noticed that the therapist did not engage in experiment, but concluded that this was justified on the basis that an overly experimental attitude on the part of the therapist could have caused a rupture in the therapeutic relationship. Determining that significant unusual factors were present involves an exercise of judgment on the part of a rater. The presence of such factors disqualifies a session from being evaluated for treatment adherence using the GTFS.

Study 3

The aims of Study 3 were to assess the criterion validity and reliability of the GTFS as a measure to determine whether or not a therapist is delivering GT.

Method

Participants. Invitations to participate in the study were circulated at GT conferences and through GT discussion lists. One hundred and

seventy-six people from 18 countries participated: Australia, Austria, Belgium, Belorussia, China, Croatia, the Czech Republic, France, Germany, Greece, Hungary, Italy, Japan, the Netherlands, Slovakia, Spain, the United Kingdom, and the United States. The highest numbers of participants were from Italy ($n = 45$), Hungary ($n = 20$), and the United Kingdom ($n = 19$). There were 122 females (69.3%, mean age = 47.98, $SD = 10.95$), 46 males (26.7%, mean age = 50.37, $SD = 12.44$) and seven raters who did not provide gender or age information (4%). One hundred and fifty-two (86.4%) participants did not speak English as their first language although all participants indicated that they were able “to complete the study requirements in English”. One hundred and fifty-two (86.4%) of the raters were current Gestalt practitioners (mean years as Gestalt practitioners = 13.77, $SD = 9.81$). Additionally, 96 of the raters were current Gestalt therapy trainers (mean years as Gestalt therapy trainers = 10.53, $SD = 8.89$).

Measure. The final version of the GTFS was a 21-item scale describing therapist behaviours, grouped under seven concepts. For each of the 20 therapist behaviour items there were two response choices: “Yes” and “No”. Item 21 is described above.

Stimuli. The researchers arranged and recorded simulated clinical sessions with therapists trained in GT and therapists not trained in GT. Therapists with similar levels of experience were recruited from two sources in Melbourne, Australia. Therapists trained in GT were recruited from Gestalt Therapy Australia, where they were in their fourth year of clinical training, and were doing placements in the Connect Ground clinic. The not-GT therapists were Masters students doing placements in the Swinburne Psychology Clinic. People recruited to play the role of client had all previously undertaken psychotherapy. Therapist and client had not previously met.

The videos were recorded in the same room, with the same lighting and production values. Each session lasted approximately 25 minutes. In order to provide a degree of verisimilitude and authenticity, the sessions were based on a form of role-play in which the clients, in consultation with author MF, were asked to choose an issue that was real for them in the here and now, but also manageable in the circumstances of a simulated therapeutic encounter. Therapist and client were then introduced to each other and had a brief discussion about the issue before the session; as a result, the recorded session did not have to deal with preliminary, client intake-related

matters. Eleven sessions were recorded (six GT, and five not-GT), with eight different therapists (six GT and two not-GT, with one of the not-GT therapists featuring in four recordings) and eight clients (with three of the clients featuring in two recordings).

Two GT videos and two not-GT videos were selected. The therapeutic orientation of the not-GT videos was Cognitive Behavioural Therapy (CBT). Videos were selected for use in the study on the basis of gender balance, variety of modality and theme, and not having the same therapist or client appear twice. All therapists featured in the sessions were assessed as competent by a panel of three experienced clinicians, who all had training in both GT and CBT and had been practicing therapists for at least 10 years; two members of the panel were also clinical psychologists.

Procedure. People who responded to the invitations to participate as raters were asked to invite other GT therapists in their location to undertake the study with them. Author MF sent copies of the GTFS, the article containing the concept descriptions (Fogarty et al., 2016), a demographics questionnaire and the ethics documents to these local organisers, the extracts from the Clemmens and Young videos used in Study 2 (the test videos), and the four videos of simulated clinical sessions selected for Study 3 (the research videos). Author MF also provided them access to an instructional video in which she recorded instructions (based on those given to the participants in Study 2) about how to use the GTFS (Fogarty, 2018).

The local organisers arranged rooms with the technology to watch the videos and facilitated their group’s participation. Participants were asked to familiarise themselves in advance with the GTFS and the article. At the start of the day participants completed the demographics questionnaire. They then watched the instructional video. After this, participants rated the two test videos (without discussion). This was used as a test of their competence in using the GTFS: the results of Study 2 suggested that a competent rater should give a higher score to the Clemmens video than to the Young video; a participant who did not do this was therefore considered to be incompetent and their data would be excluded from the study. Finally, participants rated the four research videos. Raters were blinded to the modality of these videos, and were instructed not to engage in discussion with each other while rating them. Four participants undertook the procedures individually, without being in a group.

Results

Test videos. Every participant gave a higher score to the Clemmens video than to the Young video.

Repeated measures ANOVA with simple contrasts. Statistical tests were carried out in SPSS-23. A one-way repeated measures Analysis of Variance (ANOVA) was conducted to test for differences between GTFS ratings of the GT and not-GT sessions (Tabachnick & Fidell, 2013). Deviations from normal distribution were not found to influence the result's robustness. Mauchly's test of sphericity was extremely significant, $\chi^2 = 30.24(5)$, $p < .001$, $e = .90$, however as group sizes were equal and epsilon was greater than $e = .75$, interpretation could continue using Huynh-Feldt correction index (Mauchly, 1940; Tabachnick & Fidell, 2013).

The ANOVA revealed a significant difference in ratings across sessions, $F(2.71, 474.16) = 1177.578$, $p < .001$, $\eta^2 = .87$. Session type accounted for 87% of the variance in GTFS ratings. Mean rating and standard deviation for each session are presented in Figure 1. Scores were as follows: Not GT 1 (mean = 3.65, $SD = 0.21$), GT 1 (mean = 15.36, $SD = 0.23$), GT 2 (mean = 17.82, $SD = 0.18$), and Not GT 2 (mean = 5.86, $SD = 0.28$).

Simple contrasts produced significant results. The GT sessions were rated significantly higher than the not-GT sessions, with large effect sizes (between 81% and 94%) accounted for in contrasts between those sessions (Table II).

Internal reliability and inter-rater reliability. Internal reliability was evaluated using Cronbach's alpha. In general a Cronbach's alpha of > 0.70 is an

Table II. Within-subjects contrasts of not-GT therapy videos contrasted with GT videos.

	$F(1,161)$	η^2	M difference
Not GT 1 vs GT 2	1814.08***	.91	11.71
Not GT 1 vs GT 3	2830.67***	.94	14.17
Not GT 1 vs Not GT 4	66.54***	.28	2.21
Not GT 4 vs GT 3	1553.15***	.89	-11.96
Not GT 4 vs GT 2	766.60***	.81	-9.50

Note. *** = $p < .001$

indicator of good reliability for a scale (Nunnally, 1978). Internal reliability of the GTFS was excellent, with a Cronbach's alpha of .98.

Inter-rater reliability (IRR) was evaluated using a Two-Way Random model Intraclass Correlation Coefficient (ICC) for ratings agreement between raters across GT sessions and not-GT sessions (Landers, 2015). ICC is a measure of reliability and is typically a ratio of the variance of interest to the sum of the variance of interest plus error. It is used to measure the inter-rater reliability of continuous constructs. As a rule of thumb, ICC above 0.8 is considered excellent, those in the range of 0.7–0.8 are considered good and those in the 0.5–0.7 range are considered fair (Koo & Li, 2016). Excellent IRR was observed between the 176 raters, with an ICC of .88 ($p < .001$) and a 95% confidence interval of .70–.99.

Roc analysis of GTFS ratings across therapy sessions. A receiving operating characteristic (ROC) analysis was conducted to develop a cut-off score that accurately differentiated GT from not-GT sessions (Hanley & McNeil, 1982; Metz,

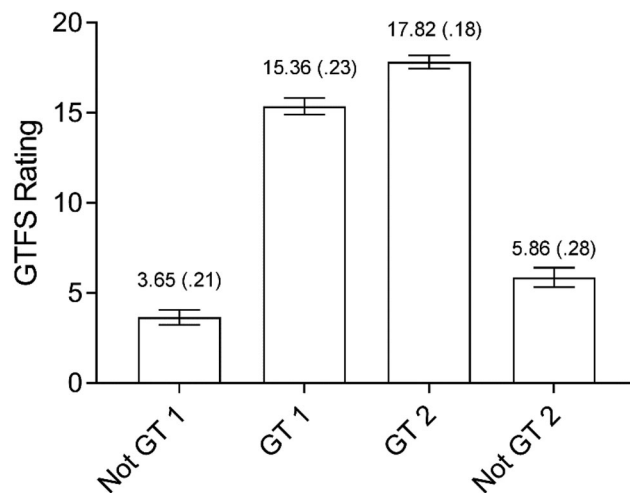


Figure 1. Mean GTFS ratings across session type ($N = 176$).

1978). The ROC curve, is a plotting of the true positive rate compared to the false positive rate, such that the diagnostic ability of a binary test can be shown as the scores on a test vary (Mason & Graham, 2002). This allows for an optimal cut off score to be chosen where there is the lowest amount of false negatives and lowest amount of false positives simultaneously. The results of the ROC analysis are presented in Table III. Results from the ROC analyses revealed that the GTFS achieved very good sensitivity and specificity in predicting the difference between the GT and not-GT sessions. A cut-off score of 11 achieved a very high efficiency with only a 2% false-negative rate and a 7.5% false-positive rate ($N=704$). Therefore, an optimal cut-off score of 11 on the GTFS may be used to predict whether a therapy session is a Gestalt therapy session, at least when compared to a session with a therapist trained in CBT.

Discussion

The results of Study 3 provided strong evidence for both the criterion validity and reliability of the GTFS as a treatment fidelity measure. The ROC analysis suggests that a therapist is adhering to GT if they achieve a score of 11 from an independent rater assessing in a single video recorded session. This might be subject to the caveat that the rater may have identified (in response to item 21) the presence of significant unusual factors that justified the therapist not engaging in certain behaviours. For example, in rating a session with an obese client, the rater might have observed that the therapist did not work with embodied awareness; however, the rater might conclude that this was justified on the basis that the therapist could have been concerned that such an intervention would have been shaming.

Table III. ROC analysis statistics for GTFS Agreement between raters of GT vs not GT videos.

	Statistics
True Positive	320
False Positive	26
False Negative	7
True Negative	339
AUC	.99
Sensitivity	.95
Specificity	.96
PPV	95.9%
NPV	94.9%
Efficiency	0.95

Note: Statistics based on a cut-off score of 11; AUC: Area Under the Curve of the Receiver Operator Characteristic; PPV: Positive Predictive Value; NPV: Negative Predictive Value.

General discussion

Aims and Findings

The specific aim of this research was to develop a measure of treatment adherence for GT and to evaluate its content validity, face validity, criterion validity and reliability. An extensive literature review was conducted to identify the fundamental concepts of GT and therapist behaviours that operationalised those concepts.

The aims of Study 1, a Delphi study, were to evaluate the content validity of the items generated in the literature review. Study 1 provided evidence that the eight concepts identified in the literature review were all fundamental to GT and that 24 descriptions of therapist behaviours reflected those concepts. This led to the development of a draft 24-item instrument with content validity.

The aims of Study 2 were to evaluate the face validity and clarity of those items, and to identify what training a rater needed to use the GTFS. The participants reported that the training they had received had been sufficient to enable them to use the GTFS. The participants also reported that they had been able to understand most of the items in the draft instrument and to observe the behaviours described by those items when watching videos of clinical sessions. Participant feedback about the other items led to the elimination of four items, the revision of three items, and the addition of a new item that asked raters to indicate and describe the presence of “significant unusual factors” that “justified the therapist not engaging in the behaviours described in this scale”.

The aims of Study 3 were to evaluate the criterion validity and reliability of the revised 21-item instrument as a measure for evaluating whether a therapist is administering GT. The results of Study 3 provided strong evidence that the GTFS has excellent criterion validity and reliability when used for that purpose. The ROC analysis suggests that a therapist is adhering to GT if they achieve a score of 11 from an independent rater assessing a single video recorded session. The breadth of participation in Studies 1 and 3 provided a degree of assurance that the GTFS is consistent with principles of GT endorsed by an international community.

Clinical or Methodological Significance of this article

Treatment fidelity scales measures are usually constructed to evaluate the adherence of treatment to specific treatment manuals. There is no generally-accepted GT treatment manual. This meant the development and validation of a treatment fidelity measure for GT required an innovative methodology

that could be used by other researchers seeking to construct an adherence scale for psychotherapy research. The current research therefore establishes a methodological framework for how to construct an adherence scale in psychotherapy research.

The operationalised items provide a foundation for identifying processes of change in psychotherapy and may also support the wider psychotherapy community by providing operationalised interventions relevant to common and specific factors research. From a clinical perspective, the GTFS may also provide a checklist for therapists and their trainers and supervisors, enabling them to determine, relatively easily, the extent to which the therapist is applying the observable interventions of GT. In institutional clinical settings, the GTFS also enables GT therapists to describe their therapeutic interventions in terms that are consistent with GT.

Limitations and Further Research

The GTFS is a parsimonious scale that can be assessed on a 25-minute video segment of clinical work. Prior research demonstrating the efficacy of GT may be substantiated by post-hoc assessment of video recordings of sessions of GT (Brownell, 2016; Roubal, 2016). Future research using single case time series analysis or Randomized Control Trials (RCTs) may be supported by the GTFS (Herrera, Mstibovskiy, Roubal, & Brownell, 2018), as ratings of adherence are critical for replication (Barber & Sharpless, 2015).

In developing the GTFS, we were guided by considerations of utility towards a treatment fidelity measure that could be used by raters who lacked the expertise necessary for assessments of competence (Weck et al., 2011). This suggested a measure that required raters to assess treatment adherence only, and to do so on the basis of binary response options concerned with whether or not the therapist behaviours in the scale were observed. A limitation in such a treatment fidelity measure is that it allows for the possibility that poor clinical results may be due to the treatment having been delivered incompetently. Further studies may develop the GTFS into a competence scale in order to eliminate this possible critique of studies relying on the GTFS. This is likely to require the development of descriptions of competence standards for each of the therapist behaviours, anchored to the points of a Likert scale. This has been the approach taken in other fidelity scales that do assess competence, including the Emotion-Focused Therapy–Therapist Fidelity Scale (Denton et al., 2009), the Person-Centred and Experiential Psychotherapy Scale (Freire, Elliott & Westwell, 2014) and the ACT/tCBT Adherence and

Competence Rating Scale (McGrath, 2009). Training or accreditation bodies could also develop such standards in order to facilitate their use of the GTFS as an assessment, training or supervision tool.

Another limitation of the study is that all of the participants in Studies 2 and 3 were trained in GT and had some clinical experience. It is, therefore, uncertain whether raters who are not trained in GT would be able to apply the scale if they received the training in the use of the GTFS that was offered in Study 3. Further research may seek to establish whether the GTFS can also be effectively used by raters who are not trained in GT.

A current feature of the GTFS is that the instrument is written in English, and all of the videos of clinical sessions that participants rated in Studies 2 and 3 were also conducted in English. However, GT has been written about, taught and practised in many other languages for many years (O’Leary, 2013), and it is possible that the GTFS fails to account for any differences in the ways that GT may be understood and practised in those other languages. Future studies might seek to replicate the findings of the current research using translations of the GTFS, and videos of clinical sessions conducted in other languages.

A cut-off score was produced by the results of the ROC analysis that distinguishes between therapists trained in GT and therapists with a CBT orientation. However, it remains to be investigated if this score would accurately distinguish between therapists trained in GT and therapists trained in other humanistic/experiential modalities that have more similarities with GT than does CBT, such as EFT and Person-Centred and Experiential Psychotherapy. Further studies will be necessary to establish whether the GTFS is capable of distinguishing between therapists trained in those modalities and therapists trained in GT, and what the appropriate cut-off score would be.

Related to this issue is the question of whether the GTFS, with its focus on observable therapist behaviours, fails to capture the “aesthetic relational knowledge” (Francesetti, 2015; Spagnuolo Lobb, 2013): that is, an authentic, “felt”, relational and experiential essence of GT (Fogarty, 2017; Philippson, 2017). Phenomenological criteria that are relevant to clinical practice may not be accounted for within the paradigm of observability (Giorgi, 2009; Slife & Wendt, 2007; Slife, Wiggins, & Graham, 2005). Some GT writers have cautioned that the adoption of a scientific paradigm, with its measurement bias, would reduce GT to a positivist framework that excludes the aesthetic relational foundations of GT. In evidence-based practice (EBP), GT would be presented as a two-person psychology (rather than a system of environment/organism development (Bloom, 2016)) and its

therapeutic interventions limited to those that can be independently observed and measured (McConville, 2014; Philippon, 2017; Wheeler & Axelsson, 2015; Wollants, 2008). Further research could explore whether there is an expert consensus that these atmospheric conditions of clinical practice reflect fundamental GT concepts (along the lines of the Delphi study in Study 1); and, if so, whether they are capable of being observed by an independent rater (as the items in the GTFS were in Studies 2 and 3).

The consensus-based approach also led to the exclusion of almost all of the items that described therapist self-disclosure. This may have been due to a poor design in the original Delphi questionnaire, where there were four separate items dealing with therapist self-disclosure. This is a limitation in the GTFS, especially as self-disclosure has been shown to be efficacious in clinical practice (Pinto-Coelho, Hill, & Kivlighan Jr, 2016).

The results of this research nevertheless represent an important step towards establishing an evidence base for the efficacy of GT. Treatment fidelity measures have been fundamental in advancing the evidence base and adoption of psychotherapeutic modalities (Perepletchikova, 2011; Smout et al., 2012). There is concern in the wider psychodynamic community that RCTs comparing the efficacy of different treatment modalities are inherently flawed for several reasons (Wampold, 2001) including that the results tend to show that modality accounts for relatively little of the difference in outcome. It is possible that future EBP research may only establish that GT is as efficacious as other modalities. However, unless such research is conducted GT may cease to be taught or practised. In this context, the current research, in producing a valid and reliable fidelity scale, advances the opportunities for psychotherapy and the clients of psychotherapy to benefit from the 70 years of experience in developing and practising GT.

Conclusion

The demand for psychotherapies to be accountable in EBP is well-established. GT has been ill-equipped to meet this demand (Brownell, 2008; Burley, 2014; Fogarty et al., 2015; Frew, 2013; Gaudiano & Miller, 2013; Gold & Zahm, 2008; O'Leary, 2013; Strümpfel, 2006), despite the increase of GT research in the past decade (Roubal, 2016). In countries where GT research is flourishing (Europe and South America) GT training and practice is also increasing (EAP, 2018). A treatment fidelity measure is an essential pre-condition for research into the efficacy of any psychotherapeutic modality.

The three studies in this research provide strong evidence that the GTFS has face validity, content

validity, criterion validity and reliability as a measure for evaluating whether a therapist is delivering GT. The GTFS could also be used in professional training, supervision and continuing professional development. The collaborative methods used to develop the GTFS represent an extended effort to create dialogue between GT experts from many different nationalities and styles of clinical practice. The validation of the GTFS through studies conducted in numerous countries suggests that there is a consensus amongst GT practitioners that the GTFS represents the fundamental clinical practices of GT, without limiting those practices to a normative framework. That is, the GTFS identifies how GT is practiced, rather than the specificities of what happens in each unique clinical encounter. The utility of the scale is that it represents a common understanding of GT that has been long sought after and promises to assist psychotherapy practitioners and researchers in furthering our appreciation of clinical practice and efficacy.

Acknowledgment

The authors are grateful to all of the many people who participated in the research, including the members of the expert panel for Study 1, the pilot study for Study 2, and the trainee therapists, clients, raters and facilitators in Study 3. The authors would also like to acknowledge the support of Dan Bloom, Phil Brownell, Matthew Farrugia, Gianni Francesetti, Pablo Herrera Salinas, Illia Mstibovskiy, Leanne O'Shea, Jan Roubal, Mark Reck, Michele Settanni, Margherita Spagnuolo Lobb and Christine Stevens.

References

- Barber, P. (2006). *Becoming a practitioner researcher: A Gestalt approach to holistic inquiry*. London, UK: Middlesex University Press.
- Barber, J. P., Liese, B., & Abrams, M. J. (2003). Development of the cognitive therapy adherence and competence scale. *Psychotherapy Research, 13*, 205–221.
- Barber, J. P., & Sharpless, B. A. (2015). On the future of psychodynamic therapy research. *Psychotherapy Research, 25*(3), 309–320.
- Bloom, D. (2016). The relational function of self: Self function on the most human Plane. In J.-M. Robine (Ed.), *Self—A polyphony of contemporary Gestalt therapists* (pp. 665–84). St. Romaine-la-Virvée, France: L'Expressimie.
- Brownell, P. (ed.). (2008). *Handbook for theory, research and practice in Gestalt Therapy*. Newcastle, UK: Cambridge Scholars Publishing.
- Brownell, P. (2014). C'mon now: Let's get serious about research. *Gestalt Review, 18*(1), 6–23.
- Brownell, P. (2016). Contemporary Gestalt Therapy. In D. Cain, K. Keenan, & S. Rubin (Eds.), *Humanistic psychotherapies: Handbook of research and practice* (2nd ed.). (pp. 219–250). Washington, DC: American Psychological Association.
- Burley, T. (2014). Can we get back to being serious about the processes of experience, awareness, and the action of gestalt formation resolution? *Gestalt Review, 18*(1), 17–21.

- Butollo, W., Koenig, J., Karl, R., Henkel, C., & Rosner, R. (2014). Feasibility and outcome of dialogical exposure therapy for posttraumatic stress disorder: A pilot study with 25 outpatients. *Psychotherapy Research, 24*(4), 514–521.
- Cooper, M., & McLeod, J. (2012). *Pluralistic counselling and psychotherapy*. Los Angeles, CA: SAGE.
- Denton, W. H., Johnson, S. M., & Burlinson, B. R. (2009). Emotion-focused therapy – therapist fidelity scale (EFT-TFS): Conceptual development and content validity. *Journal of Couple & Relationship Therapy, 8*, 226–246.
- de Villiers, M. R., de Villiers, P. J., & Kent, A. P. (2005). The Delphi technique in health sciences education research. *Medical Teacher, 27*, 639–643.
- Elliott, R., & Hendricks, M. (2017). Abstracts of research on Gestalt Therapy. Retrieved from <http://www.experiential-researchers.org/studies.html> in June 2017
- European Association of Psychotherapy. (2018). Retrieved from <http://www.europsyche.org/ecp> on 27 February 2018
- Fairburn, C. G., & Patel, V. (2014). The global dissemination of psychological treatments: A road map for research and practice. *American Journal of Psychiatry, 171*(5), 495–498.
- Fogarty, M. (2017). The Gestalt Therapy fidelity scale: A reply to Philippson. *British Gestalt Journal, 26*(2), 59–60.
- Fogarty, M. (2018). The GTFS: How to use the Gestalt Therapy fidelity scale. [YouTube video]. Retrieved from <https://youtu.be/chDT-6KqJ9M>
- Fogarty, M., Bhar, S., & Theiler, S. (2015). Creating a fidelity scale for Gestalt therapy. *Gestalt Journal of Australia and New Zealand, 11*, 39–54.
- Fogarty, M., Bhar, S., Theiler, S., & O’Shea, L. (2016). What do Gestalt Therapists do in the clinic? The expert consensus. *British Gestalt Journal, 25*(1), 32–41.
- Francesetti, G. (2015). From individual symptoms to psychopathological fields. Towards a field perspective on clinical human suffering. *British Gestalt Journal, 24*(1), 5–19.
- Francesetti, G., Gecele, M., & Roubal, J. (Eds.). (2013). *Gestalt Therapy in clinical practice: From psychopathology to the aesthetics of contact*. Milan: FrancoAngeli.
- 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 and Psychotherapy Research, 14*(3), 220–226.
- Frew, J. (2013). Gestalt Therapy training and research: Holding our place at academic institutions in the USA. *Gestalt Review, 17*(3), 263–271.
- Fromme, D. K. (2011). *Systems of psychotherapy*. New York, NY: Springer.
- Gaudiano, B. A., & Miller, I. W. (2013). The evidence based practice of psychotherapy: Facing the challenges that lie ahead. *Clinical Psychology Review, 33*, 813–824.
- Giorgi, A. (2009). *The descriptive phenomenological method in psychology: A modified Husserlian approach*. Pittsburgh, PA: Duquesne University Press.
- Gold, E., & Zahm, S. (2008). The need for Gestalt Therapy research. In P. Brownell (Ed.), *Handbook for Theory, research and practice in Gestalt Therapy* (pp. 27–36). Newcastle, UK: Cambridge Scholars Publishing.
- Goss, S., & Stevens, C. (Eds.). (2016). *Making research Matter: Researching for change in the theory and practice of counselling and psychotherapy*. New York, NY: Routledge.
- Greenberg, E. (2016). *Borderline, narcissistic and schizoid adaptations: The pursuit of love, admiration, and safety*. New York, NY: Greenbrooke Press.
- Greenberg, L. (2010). *Emotion-focussed therapy*. Washington, DC: American Psychological Association.
- Greenberg, L., Rice, L. N., & Elliott, R. (1993). *Facilitating emotional change: The moment-by-moment process*. New York, NY: Guilford Press.
- Grikscheit, F., Weck, F., Hautzinger, M., Heidenreich, T., Weigel, M., Rudari, V., ... Stangier, U. (2015). Evaluation of treatment integrity: The treatment specificity index. *Journal of Clinical Psychology, 71*(7), 653–665.
- Gutermann, J., Schreiber, F., Matulis, S., Stangier, U., Rosner, R., & Steil, R. (2015). Therapeutic adherence and competence scales for Developmentally Adapted Cognitive Processing Therapy for adolescents with PTSD. *European Journal of Psychotraumatology, 6*. <https://doi.org/10.3402/ejpt.v6.26632>
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under the receiver operating characteristic (ROC) curve. *Radiology, 143*, 29–36.
- Herrera, P., Mstibovskiy, I., Roubal, J., & Brownell, P. (2018). Researching Gestalt Therapy for anxiety in practice-based settings: A single-case experimental design. *Revisit Argentina de Clinical Psychological, 27*(2), 321–352. doi:10.24205/03276716.2018.1066
- Jacobs, L., & Hycner, R. (2010). *Relational approaches in Gestalt Therapy*. New York, NY: Routledge, Taylor & Francis Group.
- 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), 155–63. doi:10.1016/j.jcm.2016.02.012
- Landers, R. N. (2015). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower, 2*, 1–4. doi:10.15200/winn.143518.81744
- Lee, C. (2018). Treating borderline personality disorder. *InPsych, 40*(2), 16–21.
- Linehan, M. M. (2000). Commentary on innovations in dialectical behavior therapy. *Cognitive and Behavioral Therapy, 7*, 478–481.
- Luborsky, L., McLellan, A. T., Woody, G. E., Brian, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry, 42*, 602–611.
- Mason, S. J., & Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society, 128*, 2145–2166. doi:10.1256/003590002320603584
- Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics, 11* (2), 204–209. doi:10.1214/aoms/1177731915
- McConville, M. (2014). Commentary 1: Gestalt Therapy, research and phenomenology. *Gestalt Review, 18*(1), 23–31.
- McGrath, J. B. (2009). Validation of the Drexel University ACT/CBT Adherence and Competence Rating Scale: Revised for use in a clinical population (Doctoral dissertation). Drexel University, Philadelphia.
- Meara, A., & Fogarty, M. (2017). Editorial: Excitement and growth in Gestalt therapy research. *Gestalt Journal of Australia and New Zealand, 13*(2), 2–7.
- Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8*(4), 283–298. doi:10.1016/S0001-2998(78)80014-2
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- O’Leary, E. (Ed.). (2013). *Gestalt Therapy around the world*. West Sussex, UK: Wiley-Blackwell.
- Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and Practice, 18*, 148–153.
- Perls, F., Hefferline, R., & Goodman, P. (1951). *Gestalt therapy: Excitement and growth in the human personality*. Highland, NY: The Gestalt Journal Press.
- Philippson, P. (2017). ‘What does a therapist do when s/he does Gestalt therapy?’ From the conference ‘exploring practice-

- based research in Gestalt Therapy', Paris, May 2017. *British Gestalt Journal*, 26(2), 57–58.
- Pinto-Coelho, K. G., Hill, C. E., & Kivlighan Jr, D. M. (2016). Therapist self-disclosure in psychodynamic psychotherapy: A mixed methods investigation. *Counselling Psychology Quarterly*, 29(1), 29–52.
- Roubal, J. (Ed.). (2016). *Towards a research Tradition in Gestalt Therapy*. Newcastle, UK: Cambridge Scholars Publishing.
- Roubal, J., Francesetti, G., & Gecele, M. (2017). Aesthetic diagnosis in Gestalt Therapy. *Behavioral Sciences*, 7(70), doi:10.3390/bs7040070
- Roubal, J., & Rihacek, T. (2016). Therapists' in-session experiences with depressive clients: A grounded theory. *Psychotherapy Research*, 26(2), 206–219.
- Segal, Z. V., Teasdale, J. D., Williams, J. M., & Gemar, M. C. (2002). The mindfulness-based Cognitive Therapy adherence scale: Inter-rater reliability, adherence to protocol and treatment distinctiveness. *Clinical Psychology and Psychotherapy*, 9, 131–138.
- Slife, B. D., & Wendt, D. C. (2007). The next step in the evidence-based practice movement. In B. D. Slife (Ed.), *Taking sides: Clashing views on controversial psychological issues* (15th ed.) (pp. 81–89). New York, NY: McGraw-Hill.
- Slife, B. D., Wiggins, B. J., & Graham, J. T. (2005). Avoiding an EST monopoly: Toward a pluralism of philosophies and methods. *Journal of Contemporary Psychotherapy*, 35, 83–97.
- Smout, M. F., Hayes, L., Atkins, P. W. B., Klausen, J., & Duguid, J. E. (2012). The empirically supported status of acceptance and commitment therapy: An update. *Clinical Psychologist*, 16, 97–109.
- Smout, M. F., Longo, M., Harrison, S., Minniti, R., Wickes, W., & White, J. M. (2010). Psychosocial treatment for methamphetamine use disorders: A preliminary randomized controlled trial of Cognitive Behavior therapy and Acceptance and Commitment Therapy. *Substance Abuse*, 31(2), 98–107.
- Spagnuolo Lobb, M. (2013). *The now-for-next in psychotherapy: Gestalt Therapy recounted in post-modern society*. Milan: FrancoAngeli.
- Stevens, C., Stringfellow, J., Wakelin, K., & Waring, J. (2011). The UK Gestalt psychotherapy CORE research project: The findings. *British Gestalt Journal*, 20, 22–27.
- Stiles, W. B., Hill, C. E., & Elliot, R. (2015). Looking both ways. *Psychotherapy Research*, 25(3), 282–293.
- Strümpfel, U. (2006). *Therapie der Gefühle. Forschungsbefunde zur Gestalttherapie*. Friedhelm, Germany: EHP.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61, 620–630.
- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Hillsdale, NJ: Erlbaum.
- Watson, J. C., & Prosser, M. (2002). Development of an observer rated measure of therapist empathy. In J. C. Watson, R. Goldman, & M. Warner (Eds.), *Client-centered and experiential psychotherapy in the 21st century: Advances in theory, research and practice* (pp. 303–314). Ross on Wye, UK: PCCS Books.
- Weck, F., Bohn, C., Ginzburg, D. M., & Stangier, U. (2011). Assessment of adherence and competence in cognitive therapy: Comparing session segments with entire session. *Psychotherapy Research*, 21, 658–669.
- Weck, F., Grikscheit, F., Höfling, V., & Stangier, U. (2014). Assessing treatment integrity in Cognitive-Behavioral therapy: Comparing session segments with entire sessions. *Behavior Therapy*, 45, 541–552.
- Weck, F., Hilling, C., Schermelleh-Engel, K., Rudari, V., & Stangier, U. (2011). Reliability of adherence and competence assessment in Cognitive Behavioral therapy: Influence of clinical experience. *The Journal of Nervous and Mental Disease*, 199, 276–279.
- Weissman, M., Markowitz, J., & Klerman, G. (2000). *Comprehensive guide to interpersonal psychotherapy*. New York, NY: Basic Books.
- Wheeler, G. (2004). *Gestalt Therapy (Systems of Psychotherapy Video Series)*. Washington, DC: American Psychological Association.
- Wheeler, G., & Axelsson, L. (2015). *Gestalt Therapy*. Washington, DC: American Psychological Association.
- Wollants, G. (2008). *Gestalt Therapy: Therapy of the situation*. Turnhout, Belgium: Faculteit voor Mens en Samenleving.
- Yontef, G. M. (1993). *Awareness, dialogue and process: Essays on Gestalt Therapy*. Highland, NY: Gestalt Journal Press.
- Young, J. (2007). *Schema therapy (Systems of Psychotherapy Video Series)*. Washington, DC: American Psychological Association.
- Young, J., Klosko, J., & Weishaar, M. (2003). *Schema therapy: A practitioner's guide*. New York, NY: Guilford Press.

Appendices

Appendix A. Questions addressed in the instructional video

Is there a particular order in which we should expect to observe the therapist behaviours? [No.]

Might a therapist who has not been trained in Gestalt still exhibit some of the therapist behaviours described in the GTFS? [Yes.]

Am I trying to judge how well a person uses a particular Gestalt method, or just whether I saw them doing so? [No, yes.]

What do I do if I think I saw a particular therapist behaviour, but am not completely sure? [You probably observed it and should indicate “Yes”.]

Should I answer “yes” if the behaviour happened just once? [Yes.] Or does it need to be more frequent and pervasive than that? [No.]

Could be a person be a good therapist and not score well very well on the GTFS; and conversely, could someone score well on the GTFS without being a very good therapist? [Yes to both.]

(Item 3) What is “creative indifference”? How does it differ from not responding judgmentally (Item 2)?

(Item 4) In these videos the therapist and client have not worked together before (although they sometimes pretend this is the second session): given that, how

could the therapist draw on “their relationship with the client as the ground for challenge and growth”?

(Item 5) What is the difference between enquiring about the client’s experience and enquiring about their “immediate” experience?

(Item 8) What do you mean when you say that the therapist “describes and validates the different experiences of the therapist and client”?

(Item 9) What does “widening the client’s choices” mean? How is that different from establishing a programme for change?

(Item 11) Does it count if the therapist and client do this in the opposite order: first identifying emotion then coming to somatic experiences?

(Item 15) How do you know if the therapist is working with the client’s “interactional patterns as they emerge between client and therapist”?

(Item 16) What do you mean when you say that the therapist and client “identify the figure together”?

(Item 18) What does an “experiment” look like?

(Item 19) How would the therapist elicit feedback to grade the experiment?

Appendix B. The GTFS

		YES	NO
Dialogic relating			
1	The therapist follows the client attentively, tracking the awareness process and the client’s experience, not following a predetermined agenda	<input type="checkbox"/>	<input type="checkbox"/>
2	The therapist responds non-judgmentally to the client, creating the conditions that allow for the most effective client expression	<input type="checkbox"/>	<input type="checkbox"/>
3	The therapist demonstrates a willingness to be uncertain and to work with creative indifference	<input type="checkbox"/>	<input type="checkbox"/>
4	The therapist draws on their relationship with the client as the ground for challenge and growth	<input type="checkbox"/>	<input type="checkbox"/>
Working in the here and now			
5	The therapist enquires about the client’s immediate experience	<input type="checkbox"/>	<input type="checkbox"/>
6	The therapist supports the client to accept and deepen their awareness of their presenting issue rather than trying to change it	<input type="checkbox"/>	<input type="checkbox"/>
Phenomenological practice			

(Continued)

Continued.

		YES	NO
7	The therapist supports the client to describe, deepen and become more present to their immediate sensation, affect, cognitions and/or somatic presentations	<input type="checkbox"/>	<input type="checkbox"/>
8	The therapist describes and validates the different experiences of the therapist and client	<input type="checkbox"/>	<input type="checkbox"/>
9	The therapist encourages the client to widen their choices rather than establishing a programme for change	<input type="checkbox"/>	<input type="checkbox"/>
Working with embodied awareness			
10	The therapist makes observations and enquires about the client’s embodiment (including breathing)	<input type="checkbox"/>	<input type="checkbox"/>
11	The therapist invites the client to identify sensations, feelings, emotions, thoughts or images that emerge as a consequence of attending to somatic experiences	<input type="checkbox"/>	<input type="checkbox"/>
12	The therapist invites the client to engage with their body through experiment	<input type="checkbox"/>	<input type="checkbox"/>
Field sensitive practice			
13	The therapist investigates the ground (or context) from which the client’s presenting figure emerges	<input type="checkbox"/>	<input type="checkbox"/>
14	The therapist supports the client to identify how their perception of their environment and prior relationships and needs organise current experience	<input type="checkbox"/>	<input type="checkbox"/>
Working with contacting processes			
15	The therapist works with the client’s interactional patterns as they emerge between client and therapist	<input type="checkbox"/>	<input type="checkbox"/>
16	The therapist and the client identify the figure together	<input type="checkbox"/>	<input type="checkbox"/>
17	The therapist co-creates a space in which the client and therapist explore how they are impacting each other	<input type="checkbox"/>	<input type="checkbox"/>
Experimental attitude			
18	The therapist uses material that emerges in the therapeutic encounter as the basis for introducing experiments to develop the client’s awareness	<input type="checkbox"/>	<input type="checkbox"/>
19	The therapist grades the experiment by eliciting feedback from the client regarding the degree of challenge and support that the client perceives	<input type="checkbox"/>	<input type="checkbox"/>
20	The therapist supports the client to integrate learning and awareness that emerges from an experiment	<input type="checkbox"/>	<input type="checkbox"/>
Presence of unusual factors?			
21	In your view, were there any significant unusual factors about this clinical session that justified the therapist not engaging in the behaviours described in this scale? If “Yes” what were those factors, and what departures did they justify?	<input type="checkbox"/>	<input type="checkbox"/>