Friend, don’t be a perfectionist. Perfectionism is a curse and a strain. For you tremble lest you miss the bull’s eye. You are perfect if you let be.

Friend, don’t be afraid of mistakes. Mistakes are not sins. Mistakes are ways of doing something different, perhaps creatively new.

Friend, don’t be sorry for your mistakes. Be proud of them. You had the courage to give something of yourself. (Fritz Perls, 1969)

Perfection is for assholes. (Taylor Mac (2017). A 24-Decade History of Popular Music. 2017 Melbourne Festival)

The GTFS is a measure developed by Fogarty et al. (2015, 2016, 2019a, 2019b) that has been demonstrated to have scientifically robust content validity, face validity, criterion validity and reliability (Fogarty et al., 2019a). There are many ways to develop a fidelity scale for treatment adherence. The choices made in developing the GTFS were made under the supervision of a professor of psychology with extensive experience in scale development. The research used to develop the GTFS was then reviewed by, and accepted for publication in, the leading academic journal in the field, Psychotherapy Research (PR). This is the first time that an article about Gestalt therapy (GT) has ever been published in that journal; esteemed Gestalt colleagues have previously published in PR, but have not been able to refer to the modality as Gestalt (because of the lack of a scientific basis for Gestalt). So Hosemans’ claim that the GTFS is ‘a seemingly invalid, unreliable, and ineffective scale’ is both at odds with the views of the academic reviewers for PR; and, as the word ‘seemingly’ betrays, an assertion based on conjecture. It would take scientific research to determine whether and how the GTFS might be improved to reflect more holistically the phenomenological foundations of GT.

This does not mean that there is no substance to any of Hosemans’ criticisms. And whether the GTFS stands up to his criticisms is a matter of opinion; and, therefore, up to you, dear reader, to determine. Hosemans offers some productive lines of enquiry for the exploration and development of the GTFS. He, and the editorial team of the BGJ, who have invited me to comment, provide an opportunity for readers to consider some of the misconceptions about the GTFS that are prevailing in the international Gestalt community. As with any publication, the choices made by the developers of the GTFS about what to include or exclude in publications about the research were constrained by word limit, the need for thematic coherence and by the stipulations of the editor of the journal in question. Several of the issues raised by Hosemans were considered during the research, but have not yet made it to publication. So I welcome the opportunity that his critique provides.

There are a number of points on which Hosemans and I agree:

- Fidelity scales are necessary in research to determine ‘treatment adherence’: does the therapeutic approach offered in session correspond with the therapy that was intended to be delivered?
- Any research question regarding the effectiveness of a therapeutic method requires a fidelity scale in order to be published in a scientific journal.
- Fidelity scales are a measure of therapist behaviours.
- Fidelity scales are used by independent raters, they are not self-report instruments.
- Fidelity scales are limited by the constraints of the observability of data.
- Fidelity scales are a necessary part of any Randomised Control Trials (RCT) or efficacy research.
- Fidelity scales represent a conceptual form of a therapeutic modality.

Therefore (and we continue to agree) fidelity scales are not adequate:

- To describe the therapeutic dyad.
- To describe the aesthetic field of the dyad (Spagnuolo Lobb, 2013; Francesetti, 2015).
- To account for the phenomenological basis of the therapeutic encounter.
- To account for the wider field that may encompass complex familial, societal, economic, racial, gendered or political factors.
Differentiation

Beyond these shared points, Hosemans and I part ways. Hosemans is clearly well educated in statistics, but as a trainee therapist at Gestalt Therapy Australia, I am curious about the apparent authority with which he criticises the GTFS for failing to reflect fundamental GT principles. As O'Shea (2014) asks, ‘Why are we motivated to do research?’ And what are Hosemans’ motives for publishing such an aggressive and (as we shall see) unsubstantiated critique of the GTFS?

I was motivated to undertake the research after completing a Grad Dip in Psychology 2010–2012. This was a choice that followed the Australian government’s decision to grant health insurance for treatment by psychologists (but not psychotherapists) and to reduce funding for institutionalised care. With six children, I could not afford to lose my income. So I went back to university, after seventeen years in a private practice that was now threatened by a change in government policy. When I returned to university, I discovered why so many of my private practice clients had come to me, broken, shamed and confused by the previous psychological assistance they had received. CBT and its extensions had trained psychologists to pathologise the client and to move them towards ‘normality’ (whatever that is). This experience garnered my determination to bring GT back into psychology curriculums, in order to assist the clinical population that I have learnt from, connected with and served for the past seventeen years. Little did I know what a massive learning curve this undertaking would put me on. But what motivates Hosemans?

Hosemans articulates some misapprehensions about the GTFS that are shared by others in the GT community. He claims that the GTFS could be used in ‘determining the efficacy of GT in and of itself’. This is simply ‘fake news’. The authors of the GTFS have never made such a claim, and it is nonsensical. The GTFS is simply a fidelity scale. It is only one (necessary) part of the research process for determining the efficacy of GT. Unfortunately, this misapprehension seems to drive his critique.

Hosemans also agrees with Haynes’ claim that fidelity scales can potentially restrict the practices of the therapeutic modality to those that are stipulated by the scale itself, and turn the modality into a ‘“cookie-cutter” approach, which can be applied in the same way to all clients’. Again, this is highly unlikely in the case of the GTFS; but if it does happen, it is surely a problem arising from the (mis)use of the scale, than from the scale itself.

Nevertheless, it is a concern that has been raised by others. During the research, independent raters, trained in the use of the scale (see Fogarty et al., 2016 and 2019a), rated publicly available clinical work by GT therapists as diverse as Margherita Spagnuolo Lobb, Bob Resnick, Gordon Wheeler, Lynne Jacobs, Michael Clemmens and Peter Philippson. All of these highly regarded GT therapists were found to be practising GT according to the GTFS. However, there have also been practitioners who call themselves Gestalt therapists but whose work has little connection with the fundamental principles and practices of GT, whose work has not passed the GTFS. These studies were not carried out under rigorous research conditions, and so cannot be reported in a scientific journal. But I will lean into Hosemans’ style and suggest that there is false alarm about the potential of the GTFS to restrict practices, though it may exclude some practitioners and possibly training centres from being included within GT clinical practices. Is this a bad thing? Do we really want a situation where anyone can call themselves a Gestalt therapist, regardless of whether their work conforms to the fundamental principles and practices of GT?

It is also helpful to remember that fidelity scales have been developed for a number of other humanistic psychotherapies including Emotion-Focused Therapy (EFT) (Denton, Johnson and Burleson, 2009), Person-Centred and Experiential Psychotherapy (PCEP) (Freire, Elliott and Westwell, 2014) and Acceptance and Commitment Therapy (ACT) (McGrath, 2009; Smout et al., 2012). Furthermore, these modalities have drawn from GT and developed an evidence base, whilst GT is fading (Raffagnino, 2019).

Hosemans also challenges the parsimony of the GTFS as reductive. He reminds us that ‘interaction-based interventions are complex’ and ‘messy’ and therefore difficult to represent within a fidelity scale. As anyone involved with the GTFS development (several hundred people in total) would attest, the ease of application of the scale is one of its most appreciated features (correspondence from sixteen countries). The parsimony of the scale also opens out on to a dimension of apprehension that has previously been missing in GT. Any therapist knows that each session, each moment of a session is unique, and ‘messy’. But a scale that facilitates an apprehension of therapeutic interventions does not limit the therapist’s range, it merely offers a simple ‘checklist’ through which to deepen the comprehension of the complexity involved. As Atul Gawande, a professor of surgery at Harvard Medical School and a staff writer at The New Yorker, demonstrates in his best-selling book The Checklist Manifesto (2010):

Good checklists . . . are precise. They are efficient, to the point, and easy to use even in the most difficult situations. They do not try to spell out everything – a checklist cannot fly a plane. Instead, they provide reminders of only the most critical and important steps
many other GT authors, are now dead). It somehow feels beneath us to use a checklist, an embarrassment. It runs counter to deeply held beliefs about how the truly great among us – those we aspire to be – handle situations of high stakes and complexity. The truly great are daring. They improvise. They do not have protocols and checklists. Maybe our idea of heroism needs updating. (p. 173)

Hosemans’ challenges continue. He argues that contact, rather than raising awareness (Yontef, 1993), is the underlying factor of GT. There was not one participant in the Delphi study who agreed with Hosemans on this point, though many provided qualitative (and highly differentiated) feedback about the definition and importance of contact. Substantively, contact was the only ‘construct’ where the expert panel could not reach a consensus about how it should be described, even though the panel did reach a consensus about the therapist behaviours that reflected it (Fogarty et al., 2016, 2019a). At one stage in the process of drafting the operationalisations for the Delphi, Leanne O’Shea jokingly shared with me ‘that if we held a conference on the theme of “contact in GT” half of the participants would kill each other!’ Jokes aside, I wonder whether Hosemans may be too new to the theory, practice and history of GT to apprehend the complexity of his assertion.

Scale development

Hosemans challenges almost every aspect of the scale development, from the lack of a systematic literature review to critiquing both the choice and procedure of the Delphi study, to the response format, the scale’s internal structure, and the statistical analysis.

The literature on GT is vast (see, for example, the full reference list to Fogarty et al., 2016, available at both the BGJ’s and the author’s websites). Might a different person conducting that review have come up with different results? Of course. But does that invalidate the study? Not according to the academic reviewers for PR. A ‘systematic’ literature review may have been far less broad: see, for example, Raffagnino (2019).

Incidentally, Hosemans’ suggestion that it would have been beneficial to have replaced the broad-based Delphi study with a study based on ‘fewer experts, such as those who wrote the literature from which the initial principles and concepts were derived’ suggests that he has little appreciation of the breadth of Gestalt literature or of the number of contributing authors (or of the fact that Perls, Hefferline and Goodman, and many other GT authors, are now dead).

Hosemans’ insight that the embodied experience of GT differs from conceptual understanding is correct. Over twenty years’ experience as a GT practitioner and supervisor has deepened my awareness of that fact. However, Hosemans’ assertion that consultation on the operationalisation of concepts was limited to Dr O’Shea is incorrect. The concepts were worked over five days at the Second International Conference for Gestalt Therapy Research in Cape Cod, May 2015. Esteemed colleagues such as Phil Brownell, Joe Melnick, Jan Roubal, Vincent Beja, Mark Reck and Leslie Greenberg were amongst those present. I then spent a month in New York regularly consulting with Dan Bloom and Iris Fodor. On returning to Australia, I reworked the concepts with my supervisors before loading them into Qualtrics for the Delphi survey (Fogarty et al., 2016). So I am grateful for the opportunity to underscore that the brainstorming phase was not ‘skipped’. If any researcher bias remained (as it often does in scale development) it was not commented on by any of the blind peer reviewers at any stages of the publications (Fogarty et al., 2016, 2019a).

The details of the Delphi questionnaire have been described elsewhere (ibid., 2016, 2019a). At every stage of the consultation, experts were asked to provide modifications, additions and eliminations. The assumption that experts need to be equal is both ludic in its conception and refuted by many. There were no participants with only a few years of clinical experience, so the question about weighting the regression to the mean is moot.

Time constraints were not reported as problematic for any of the sixty-three participants, no doubt because Qualtrics allows for the participants to go in and out of the survey at their convenience. It is axiomatic that a different group of experts may yield a different consensus. And if anyone is prepared to conduct one, then they can. One hundred and fifty-four experts were invited to contribute. Those that chose not to had their reasons, but that remains irrelevant to the results.

Hosemans also implicitly criticises the fact that it took only two iterations of the Delphi to achieve consensus. That this is another unfair slur only becomes apparent when one reads the source material cited by Hosemans. In the version of the Delphi method that Hsu and Sandford (2007) and Iqbal and Pipon-Young (2009) describe, the first iteration is usually a ‘brainstorming’ round used to generate the items for the remaining rounds. This round was omitted in this study (which used a different method for generating the items). However, once the items have been generated, it usually only takes two further iterations to collect the necessary data (Hsu and Sandford, 2007). The Delphi study for the GTFS was therefore typical in this regard. As Hsu and Sandford
(2007) and Iqbal and Pipon-Young (2009) also explain, further iterations can lead to falling participation rates (affecting the validity of the data) and only slight increases in the degree of consensus. Nor was the criterion of 80% agreement ‘arbitrarily defined’ (as Hosemans suggests), but has been recommended in the literature (as has a lower level of 70% agreement) (Hsu and Sandford, 2007).

The large number of experts consulted was the foundational motivation for the choice of a Delphi method. As outlined, GT has been practised for over seventy years, in many countries, so broad consultation was considered the most representative method for the task. Hosemans criticises the size of the Delphi panel as excessively large; however, Delphi panels of more than 50 people (and as many as 1,000) have been used before (Hsu and Sandford, 2007; Iqbal and Pipon-Young, 2009); and one can readily imagine that if the panel had failed to represent the breadth of countries and styles in which GT is practised, Hosemans would have criticised it for that instead. That the breadth of the panel may have led to generalised items was precisely the point. As the excellent inter-rater reliability between the 176 raters (ICC = .88 (p < .001)) (Fogarty et al., 2019a) in the validation study showed, the GTFS is both essential enough to represent a wide range of views about contemporary GT across the world, and discriminatory enough to distinguish it from other modalities.

Hosemans questions the specificity of the items (Theiler also shared this concern). As a GT practitioner, I had confidence that those trained in GT would comprehend the meaning of each item. To test this assumption, a pilot study was conducted (with six participants) and items were refined according to feedback (Fogarty et al., 2019a). One of the avenues for further research identified in the PR article was whether ‘the GTFS can also be effectively used by raters who are not trained in GT’ (ibid., p. 11). Perhaps Hosemans’ difficulties in understanding the items suggests that the GTFS cannot be effectively used by raters such as Hosemans who have not completed their training in GT, and have never practised as a Gestalt therapist. Some items still remain the subject of debate and engagement between GT practitioners. For example, ‘creative indifference’ (Item 3 in the GTFS) is the main theme for the EAGT conference in Budapest in September 2019 (an example of how the GTFS provokes engagement with the practice of GT rather than restricting it).

Hosemans is a statistician. I am a GT practitioner/researcher. My curiosity is revived in every clinical encounter: what is happening now? What is going on for the client? How are we co-creating this therapeutic field? How am I, as therapist, contributing to ‘excitement and growth’ in human personality?

As you, dear reader, may have observed, Hosemans (a counselling psychologist, statistician and trainee in GT) takes us through a number of seeming errors in methodology and statistics to which the development of the GTFS is accused of having fallen foul. Hosemans makes a valuable point here. When undergoing research it is essential to have a good statistician by your side. Despite the expertise offered by my supervisor Professor Bhar, the statistical analysis of binary data was technically demanding. We engaged a number of other statisticians to assist. However, it was not until the final stage of the project that Matthew Farrugia (who could understand the project, work with binary data, and talk through the conceptual stats space) became involved.

I am not going to engage with Hosemans on the detail of his statistical critiques for several reasons. First, I have no doubt that he is more expert at statistics than I am (though I am not convinced that he is more expert than those who helped me). Second, I imagine there are few readers who will be able to work out who is right and who is wrong. Third, it is always possible to do more and better statistics; but (as Voltaire said) the perfect should never be allowed to be the enemy of the good. Fourth, anyone who wants to do further statistical analysis on the GTFS is welcome to do so! Fifth, whatever the shortcomings in the statistical

Response format

Binary options were chosen for parsimony and ease of application. We considered a continuum (Likert scale), but decided against it, as the issue of competency was more likely to vary according to country and practitioner style. (Feedback from the raters has confirmed this.) However, competency scales are being developed in a few training centres, and this development is promising. Binary data did initially prove to be difficult to manage for the statistical analysis (see below). However, when meeting with Professor Robert Elliot in Glasgow, before the BACP Conference in Chester in 2017, he gracefully informed me that I ‘had taught him something about stats and scale development!’ The continuous response items in the PCEP fidelity scale resulted in there being insufficient distinction between PCEP and EFT (Freire, Elliot and Westwell, 2014). My intuitive and holistic experiment seems to have paid off.

Internal structure

Hosemans is a statistician. I am a GT practitioner/researcher. My curiosity is revived in every clinical encounter: what is happening now? What is going on for the client? How are we co-creating this therapeutic field? How am I, as therapist, contributing to ‘excitement and growth’ in human personality?

As you, dear reader, may have observed, Hosemans (a counselling psychologist, statistician and trainee in GT) takes us through a number of seeming errors in methodology and statistics to which the development of the GTFS is accused of having fallen foul. Hosemans makes a valuable point here. When undergoing research it is essential to have a good statistician by your side. Despite the expertise offered by my supervisor Professor Bhar, the statistical analysis of binary data was technically demanding. We engaged a number of other statisticians to assist. However, it was not until the final stage of the project that Matthew Farrugia (who could understand the project, work with binary data, and talk through the conceptual stats space) became involved.

I am not going to engage with Hosemans on the detail of his statistical critiques for several reasons. First, I have no doubt that he is more expert at statistics than I am (though I am not convinced that he is more expert than those who helped me). Second, I imagine there are few readers who will be able to work out who is right and who is wrong. Third, it is always possible to do more and better statistics; but (as Voltaire said) the perfect should never be allowed to be the enemy of the good. Fourth, anyone who wants to do further statistical analysis on the GTFS is welcome to do so! Fifth, whatever the shortcomings in the statistical
analysis that we actually did, these were not enough to stop the expert reviewers from PR finding the paper to be worthy of publication.

Conclusions

Hosemans’ fundamental critiques of the GTFS seem to be that it is subject to the same limitations as any fidelity scale; that better statistical analysis could have been applied to it; and that more, and better, research might have resulted in a different instrument. Let it be. There are really two main ways of responding to the GTFS. First, we might see it as an instrument, flawed, imperfect and human, that opens the door to evidence-based practice (EBP). A tool we can use to try to preserve GT in a world that demands evidence of efficacy. Alternatively, we might see it as the enemy. Something that confines and constrains us. The lackey to a system of EBP we reject. However, the experience of developing and validating the GTFS has suggested a third way, in which the GTFS both opens the door to EBP, co-cretes an international community of support for research within GT, and stimulates and generates excitement and growth within the Gestalt therapy community. If, instead of being threatened by the GTFS as Hosemans appears to be, we both embrace it and are willing to voice our difference, then we have nothing to fear. What concerns me is that, despite Hosemans’ claim that fidelity scales are outmoded (because therapeutic modality accounts for little of the variance in treatment outcome), without a fidelity scale, GT will be lost to common factors. This issue, and the question of folding into Wampold’s contextual model (2017), were the centre of lively debate at the 3rd International Conference for Gestalt Therapy Research in Paris, May 2017. As we speak, 2,500 members of the Spanish Association of Gestalt Therapy are working overtime to find sufficient citations in scientific journals to shore up GT’s currently liminal status in Spain. David Picó and Renata Mizerska are developing a database for these and other GT references (https://gestaltresearch.org/). The only countries where GT is growing are in South America and Russia, where there is little or no governmental regulation. If we are to preserve GT in the face of governmental demands for an evidence base, then we need, at the very least, to be able to establish that GT is no less efficacious than other modalities (the ‘dodo-bird effect’: Rosenzweig, 1936; Wampold, 2001). And to do that we need a fidelity scale. The GTFS is, no doubt, an imperfect instrument. But unless and until a better instrument has been developed, criticising the GTFS may be cutting off our nose to spite our face.

Acknowledgement

I would like to thank Andrew Palmer, Pablo Herrera-Salinas and Guiomar Miranda Gomez for support in drafting this response.

References


Smout, M.F., Hayes, L., Atkins, P.W.B., Klausen, J. and Duguid, J.E. (2012). The empirically supported status of acceptance and


Madeleine Fogarty has been writing for GT publications since 2007. Her work has been translated into six languages. She has been a teacher, an academic, and a theatre director. For the past twenty years Madeleine has worked as a Gestalt therapist and now supervisor in mental health and forensic settings. More recently her focus has been in running private practice, supervision groups and concentrating on research and the development of an evidence base for GT. Madeleine regularly offers workshops and presents at psychotherapy conferences in Australia and overseas. Madeleine is the Research Director for GANZ, on the Leadership group for Psychotherapy in PACFA, a member of the EAGT scientific board and a friend and colleague to many in the GT community (to whom she is grateful).

Address for correspondence: madeleine@madeleinfogarty.com