

Correspondence: Marjolein M. W. Koementas-de Vos, GGZ-Noord-Holland-Noord, Antwoordnummer 1506, 1850 VB Heiloo, The Netherlands.

E-mail: m.koementas@ggz-nhn.nl

Citation: Koementas-de Vos, M. M. W., Colleye, L. C., Tiemens, B., Engelsbel, F., de Jong, K., & Salcuni, Witteman, C. L. M., Nugter, M. A. (2022). The feasibility, acceptability and effectiveness of a feedback-informed group treatment (FIGT) tool for patients with anxiety or depressive disorders. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 25(2), 339-353. doi: 10.4081/ripppo.2022.647

Acknowledgements: Many thanks to the patients and therapists for participating in the study, and also special thanks to Ralph Feenstra, Jessica Apeldoorn, Evelien van der Ploeg, Shuna Vanner and Rinze Krol for all the work and support.

Contributions: MK, LC, AN, BT, KJ and CW developed the design, MK, LC and AN developed the FIGT tool in collaboration with therapists and technical specialists, MK was responsible for carrying out the research, MK and LC carried out the design, MK and FE performed the data analysis, all authors were involved in discussing the results and writing the manuscript.

Conflict of interest: the authors declare no potential conflict of interest.

Ethics approval: approval was granted by the internal scientific committee of the GGZ institution where the research took place from September 2017 until December 2021. The study protocol was investigated and registered by the Medical Ethics Committee of the University Medical Centre Groningen (MEtc 2016/471 and MEtc 2016/604) and the privacy protocol of the organization, based on the General Data Protection Regulation 2016/679 (The European Parliament and the Council, 2016), was followed.

Consent for participation and publication: participants agreed to participate in the study and to have their data published in an article anonymously.

Availability of data and material: the data files are available on request. Please contact the corresponding author, e-mail: m.koementas@ggz-nhn.nl

Received for publication: 19 June 2022.

Revision received: 1 September 2022.

Accepted for publication: 7 September 2022.

Publisher's note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

©Copyright: the Author(s), 2022

Licensee PAGEPress, Italy

Research in Psychotherapy:

Psychopathology, Process and Outcome 2022; 25:339-353

doi:10.4081/ripppo.2022.647

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The feasibility, acceptability and effectiveness of a feedback-informed group treatment (FIGT) tool for patients with anxiety or depressive disorders

Marjolein M. W. Koementas-de Vos,^{1,2}
Lucas C. Colleye,² Bea Tiemens,¹ Fabiana Engelsbel,²
Kim de Jong,³ Cilia L. M. Witteman,¹ M. Annet Nugter²

¹Radboud University, Nijmegen, The Netherlands; ²GGZ Noord-Holland-Noord, Heerhugowaard, The Netherlands; ³Leiden University, Institute of Psychology, Clinical Psychology Unit, Leiden, The Netherlands

ABSTRACT

Monitoring treatment progress by the use of standardized measures in individual therapy, also called feedback-informed treatment (FIT), has a small but significant effect on improving outcomes. Results of FIT in group therapy settings are mixed, possibly due to contextual factors. The goals of this study were to investigate the feasibility, acceptability and effectiveness of a feedback-informed group treatment (FIGT) tool, based on the principles of the Contextual Feedback Theory and earlier FIGT research. Patients with anxiety or depressive disorders following interpersonal or cognitive behavioural group psychotherapy (IPT-G or CBT-G) were randomized to either feedback ($n=104$) or Treatment As Usual (TAU; $n=93$). In the feedback condition, patients filled out the Outcome-Questionnaire 45 (OQ-45) weekly in a FIGT tool and therapists were instructed to discuss the results in each session. Dropout, attendance and outcomes were measured. Additionally, in the feedback condition, OQ-45 response, feedback discussions and acceptability by patients and therapists were assessed. Results showed no differences on dropout, but lower attendance rates in the feedback condition. Although therapists reported high rates of feedback use and helpfulness, patients experienced that results were discussed with them only half of the time and they were also less optimistic about its usefulness. The findings indicate that the FIGT instrument was partially feasible, more acceptable to therapists than patients, and was not effective as intended. Future research is needed to discover how feedback can be beneficial for both therapists and patients in group therapy.

Key words: FIGT; feedback-informed group treatment; group psychotherapy; feedback; feasibility.

Introduction

Group psychotherapy is an effective treatment for patients with a variety of psychiatric disorders, often with comparable results to individual psychotherapy (Burlingame *et al.*, 2016). Despite equal levels of efficacy,

group psychotherapy is a distinct treatment modality with unique therapeutic factors, such as vicarious learning (learning from each other), universality (observing others struggling with similar issues) and altruism (offering support to each other) (Fuhrman & Burlingame, 1990). In other words, patients in group psychotherapy have more people to learn from, identify with, share experiences and with whom to form therapeutic relationships.

As with individual therapy, not all patients benefit from group treatment. It is found that 5-15% of the patients in group psychotherapy have worsened by the end of treatment (Schuman *et al.*, 2014; Slone *et al.*, 2015). Therapists in both, individual and group settings, are often quite poor at predicting negative treatment outcomes and tend to overlook negative changes in their patients. (Chapman *et al.*, 2012; Hannan *et al.*, 2005; Hatfield *et al.*, 2010; Walfish *et al.*, 2012). Routinely monitoring patients' treatment progress and directly using this as feedback, so-called feedback-informed treatment (FIT), has been proposed as a possible solution to help therapists overcome their biases about treatment progress (*e.g.*, Bickman, 2008; Lutz *et al.*, 2015). FIT mostly entails the comparison of a patient's score on a questionnaire with earlier results or, in case of more advanced feedback systems, with an expected recovery trajectory. When the patient is not improving enough, both therapist and patient receive a signal, and some systems give treatment suggestions with the use of a clinical support tool (CST) (Lambert & Lo Coco, 2014; Lambert *et al.*, 2018).

The effects of FIT have mainly been studied in individual psychotherapy and it seems especially effective for patients who benefit less from treatment than expected, also known as not-on-track patients (NOT; Lambert *et al.*, 2018). The efficacy increases when both therapists and patients receive the feedback results instead of only informing therapists, and it increases even more when therapists receive additional treatment suggestions in case of deterioration (Shimokawa *et al.*, 2010). In a multilevel meta-analysis De Jong and colleagues (2021) found that FIT has a small but significant effect on symptom reduction and dropout rates not only for NOT-patients ($d=0.15$), but for on-track (OT) patients as well ($d=0.17$).

In group psychotherapy, therapists have to monitor several patients simultaneously, which makes identifying negative change processes more difficult than in individual therapy. Feedback-informed group treatment (FIGT; Gleave *et al.*, 2017) has therefore been developed and the Task Force of the American Group Psychotherapy Association recommends the use of a measurement battery to, periodically or continuously, monitor both outcome and process factors in group psychotherapy (Strauss *et al.*, 2008).

The effects FIGT have been assessed in seven controlled group psychotherapy studies and results are mixed. Four studies described a positive effect on symptom re-

duction or well-being for both NOT and OT patients (Hutson & Page, 2020; Koementas-de Vos *et al.*, 2018; Schuman *et al.*, 2014; Slone *et al.*, 2015), two studies only for NOT patients (Burlingame *et al.*, 2018; Newnham *et al.*, 2010) and one study found no effect at all (Davidsen *et al.*, 2017). There are also indications that FIGT may have an effect on attendance, but the results are inconclusive. In two studies, patients appeared to attend more therapy sessions and showed greater symptom reduction at the end of treatment when using feedback, suggesting improved treatment retention due to FIGT (Schuman *et al.*; Slone *et al.*). Patients in another study attended fewer treatment sessions but showed similar symptom reduction at the end of treatment when using feedback in comparison to group treatment as usual (TAU), indicating an efficiency of the group treatment (Koementas-de Vos *et al.*). Furthermore, only one study described long-term effects, whereby FIGT was associated with fewer readmissions six months after treatment for patients who had remained on track during therapy (Byrne *et al.*, 2012). In conclusion, despite the mixed results, FIGT seems to have the potential to improve outcomes, but more studies are needed to explore when feedback in group therapy is effective, in what form and for whom.

In individual settings, it is assumed that feedback alone is insufficient, but that certain contextual factors are needed, as described by Sapyta and colleagues (2005) in The Contextual Feedback Intervention Theory. Important factors for successful feedback are its specificity, reliability, quick delivery, and fit with the patient's and therapist's goals. Although there is a limited amount of research on FIGT, contextual aspects also appear to play a role in the use of feedback in a group setting. First, the use of technology, such as touchscreens or web applications, appears to be associated with increased satisfaction and engagement of both patients and therapists, as well as collaboration between them (Newnham *et al.*, 2012). Immediate and accurate feedback on treatment progress are also important for both patients and therapists, with therapists preferring an overview of the group's progress as a whole (Koementas-de Vos *et al.*, 2018). Furthermore, flexibility in the treatment protocol is necessary in order to make optimal use of feedback. Davidsen and colleagues (2017) suggested that the lack of effect of FIGT in their study was mainly due to the lack of opportunities for therapists to adjust the group therapy based on the feedback.

In line with The Contextual Feedback Intervention Theory (Sapyta *et al.*, 2005) and findings of contextual factors in group psychotherapy, a web-based application for the use of feedback in group psychotherapy was developed in this study, also called FIGT tool. This tool includes a digital dashboard in which the progress of the therapy is visible for both patient and therapist. For patients, it was possible to fill out a questionnaire before every session and to see the results in a graph with visual

and textual explanation immediately. Therapists could also view the results directly on the dashboard, giving them an overview of the group as a whole and the ability to look at the individual patients' level as well.

The main goals of this study were to investigate the feasibility and acceptability of the developed FIGT tool for patients with depressive or anxiety disorders following interpersonal or cognitive-behavioural psychotherapy in a group (IPT-G or CBT-G). The third goal was to get a first impression of the effectiveness of the FIGT tool in comparison to group psychotherapy without feedback. Patients were randomized in two conditions: feedback or treatment as usual (TAU). In the feedback condition, patients were asked to fill out the Outcome Questionnaire 45 (OQ-45; Lambert, 2013) before every group therapy session. Therapists were instructed to review the feedback results prior to each session and to discuss the feedback results in the group session. In TAU, patients also followed IPT-G or CBT-G, but did not use the OQ-45, and neither patients nor therapists received any feedback.

First, with regard to feasibility, we hypothesized that the rate of dropouts in the feedback condition would be similar to or lower compared to TAU. Given the mixed results on attendance from previous studies (Koementas-de Vos *et al.*, 2018; Schuman *et al.*, 2014; Slone *et al.*, 2015), *i.e.* attendance rates could be higher or lower when using feedback, we decided to only describe the rate of sessions attended in both conditions without a hypothesis. Furthermore, feasibility was considered satisfactory when at least 70% of the OQ-45 were completed by patients in the feedback condition and that those patients indicated that at least 70% of the filled out OQ-45s were discussed with them in the treatment sessions. The percentage of 70% is an estimated threshold based on two FIGT studies reporting percentages between 70-90% in which patients completed weekly questionnaires (Hutson *et al.*, 2020; Koementas *et al.*, 2018), other FIGT studies did not describe completion rates. Second, acceptability was explored through a questionnaire for patients and therapists on aspects such as self-reported use of the tool, helpfulness, relevance and specificity of the feedback, and we expected patients and therapist to be generally positive (at least 70% is positive). Third, with regard to effectiveness, the use of feedback was expected to have more positive effects on symptom reduction and on quality of life compared to TAU.

Three exploratory analyses were conducted to test: i) if there were differences in clinical significance (rates of deterioration, no change, improvement and recovery) at the end of treatment between the feedback and TAU condition; ii) if there were correlations between outcomes and the feasibility and acceptability indicators, and; iii) whether differences on feasibility, acceptability and outcomes were found between patients who are NOT or OT during treatment, in line with previous FIGT studies.

Materials and methods

Design

The study took place at a medium sized mental health institution for patients with mental health disorders in the Netherlands. The study had a randomized cluster controlled design in which randomization took place on the group level by the use of a computer algorithm. The main reason for randomization at group level, not at patient level, was to prevent drastic allocation changes. Patients were allocated to group therapy at a certain location based on their place of residence. There was no possibility to provide similar group therapies with and without FIGT at the same location, because the number of patients at one location was too low. A total of 42 therapy groups participated, 16 IPT-G and 26 CBT-G, and each group was randomized to either a feedback or a TAU condition.

Participants

Patients

Patients were eligible if they followed an IPT-G or CBT-G between September 2017 and December 2021, see Figure 1 for the flow chart. All patients were classified with a major depressive disorder or anxiety disorder based on the criteria of the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5; American Psychiatric Association, 2013) by a psychiatrist or psychologist. Exclusion criteria were age below 18 years or above 65 years, an IQ below 80, substance abuse or dependence, psychotic symptoms, acute (hypo)manic symptoms, lack of motivation, a severe suicide risk and participation in both feedback and TAU condition. The reason for exclusion of patients younger than 18 years and older than 65 years was that the department of the institution was only organized for patients 18 to 65 years of age.

Therapists

Most groups in the current study were led by an experienced psychologist accompanied by a psychotherapist, mental health psychologist, mental health psychologist in training or a psychologist with a master's degree in training. In total, 35 therapists participated in the study, five male and 30 female, with a mean age of 39.4 years (SD=10.1 years) and mean experience as a group therapist was on average 6.9 years (SD=5.5 years).

Treatment

In this study patients followed IPT-G or CBT-G. IPT-G is a time-limited, half-open psychotherapy group for patients with major depressive disorders. Similar to individual IPT, the group format of IPT has a structure of initial, middle and termination stages, with a focus on the relationship between feelings and interpersonal situations, working on common themes and interpersonal problems

(Weissman *et al.*, 2017). In this study, patients could enter and exit the IPT-G every 8 sessions and they could follow this group to a maximum of 24 sessions. This means that the minimum number of sessions was 8 and maximum was 24 session for the IPT-G. With regard to the assignment to the research conditions, feedback or TAU, it was decided to randomize the IPT-G every 24 sessions. Only patients who followed the IPT group in one research condition were included for the data-analysis. Patients who had been in both research conditions while participating in the IPT-G were excluded. CBT-G is also a time-limited but closed, semi-structured psychotherapy group for patients with major depressive and anxiety disorders. The duration of the CBT groups varied between 5 and 14 sessions and there were no blocks to enter or exit during the group. In CBT-G, patients are taught the cognitive behavioural model, learn to recognize their thoughts related to situations, feelings and behaviours, and are encouraged to change automatic processes within a group format (Whitfield, 2010). Each CBT-G was randomized into a feedback or TAU research condition and re-randomization was not needed, because of the closed group format. Both IPT-G and CBT-G had an average frequency of one session per week, the minimum number of patients in each group was four and maximum was nine.

Intervention

Feedback

A web based application was developed (FIGT tool) with a personalized dashboard for each patient and therapist. The Dutch version of the OQ-45 (Lambert *et al.*, 2013) was used to assess patients' progress during treatment (De Jong, 2008). The OQ-45 is a brief self-report questionnaire that consists of 45 items focussing on the patient's functioning during the last week. Questions are rated on a 5-point rating scale from never (0) to almost/always (4). Three domains of functioning are measured, namely Symptom Distress (SD), Interpersonal Relationships (IR), and Social Role (SR) performance. The Dutch version of the OQ-45 has an additional domain, called Anxiety and Somatic Distress (ASD), consisting of 12 items. The total score of the OQ-45 ranges between 0 and 180, with higher total scores being indicative of greater levels of psychological distress. There are five risk items in the questionnaire about substance abuse, suicide, and violence. An example item is 'I have thoughts about ending my life.' If a patient scores one or higher on a risk item, the therapist is alerted in the web based tool.

The OQ-45 has good psychometric properties and is validated in the Dutch population. The OQ-45 is sensitive

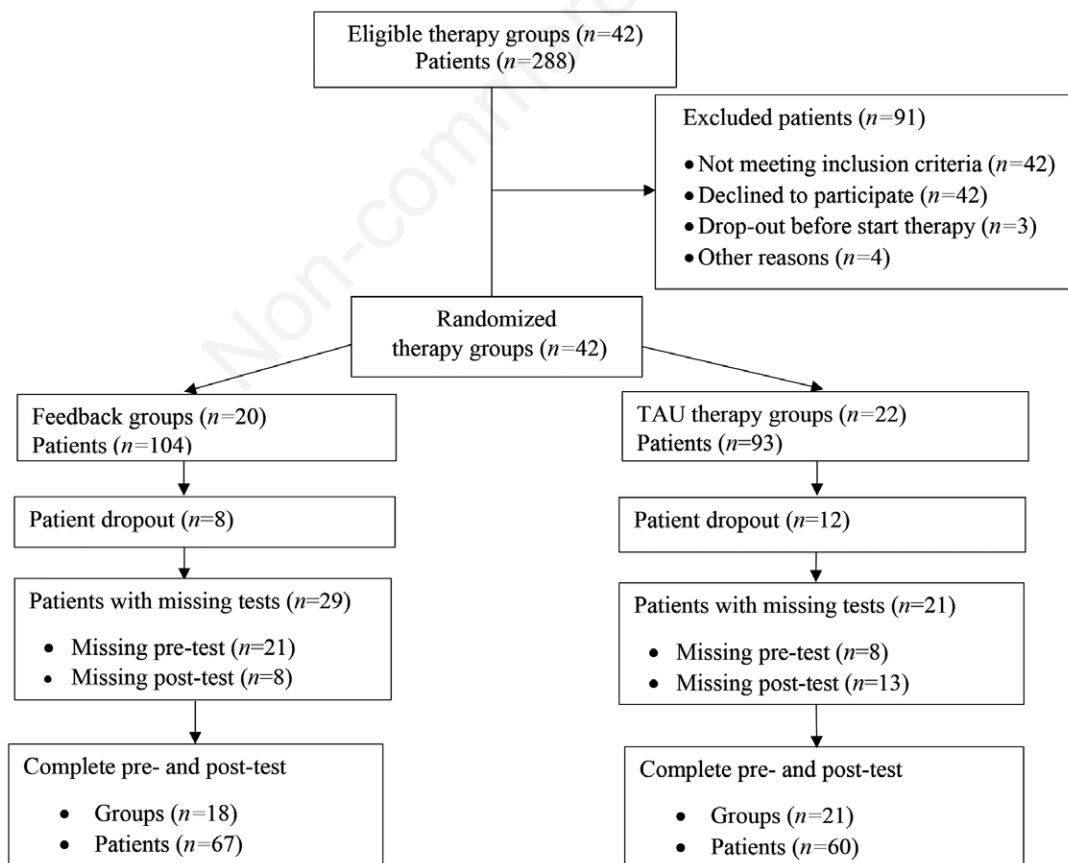


Figure 1. Flowchart of participants.

to change, whereas the scores remain stable in untreated patients, and it is therefore a good instrument for progress feedback (De Jong *et al.*, 2007). In the current study, the internal consistency was found to be $\alpha=0.95$ for the total score and $\alpha=0.95$ (SD), $\alpha=0.85$ (IR), $\alpha=0.74$ (SR), and $\alpha=0.90$ (ASD) for the subscales. The reliable change index for the OQ-45 (Timman *et al.*, 2015) is 18 points or more, and the cut-off score for normal functioning is 56 or less. This means that deterioration was defined as an increase of 18 points or more between the current session and the first or previous session only when the score was above 56. If patients showed no significant changes with scores above 56 during three consecutive sessions, their scores were signalled as 'no change in three sessions'. Improvement was defined as a decrease of 18 points or more, and recovery was defined as both a decrease of 18 points or more and a score below 56 (Timman *et al.*, 2015). Our method deviates from the method used by Lambert and other authors in the sense that no expected recovery curves were used to determine if a patient was NOT. Our method is in line with the approaches by Newnham *et al.* (2010) and De Jong *et al.* (2012; 2014), but with an added comparison with the previous session, in order to take sudden losses of functioning into account.

After completing the OQ-45, the results were immediately available on the application dashboard, for both patients and their therapists. On the patients' dashboards, individual progress on the OQ-45 was presented in a graph

and accompanied by coloured dots and texts to explain the type of change between current and previous session, and the current and first session (red=deterioration, orange=no change and green=improvement). Also a brief conclusion of the scores was given combined with a constructive feedback message, for example: 'In comparison with the previous and the first session you experience more symptoms. It may help to investigate with your group members and therapist possible explanations and solutions to achieve remediation of symptoms'. Information about the scores on the subscales of the OQ-45 was visible in a graph as well and answers on each item were presented. On the main OQ-45 progress graph, patients were able to click on earlier or further measurements and then forward to the visual display of the results of that moment.

As seen in Figure 2, a group graph of progress on the OQ-45 for the group as a whole was made for the therapists' dashboard. The graph was also accompanied by coloured dots which showed the type of change for each patient between the current and previous session, and the current and first session (red=deterioration, orange=no change and green=improvement). Therapists were also able to click on each individual measurement moment in the graph and were then forwarded to the results of that time, similar to what patients saw in their dashboard. The only difference was that the text of the risk items of the OQ-45 was coloured red for therapists so that they were extra alerted to these items.

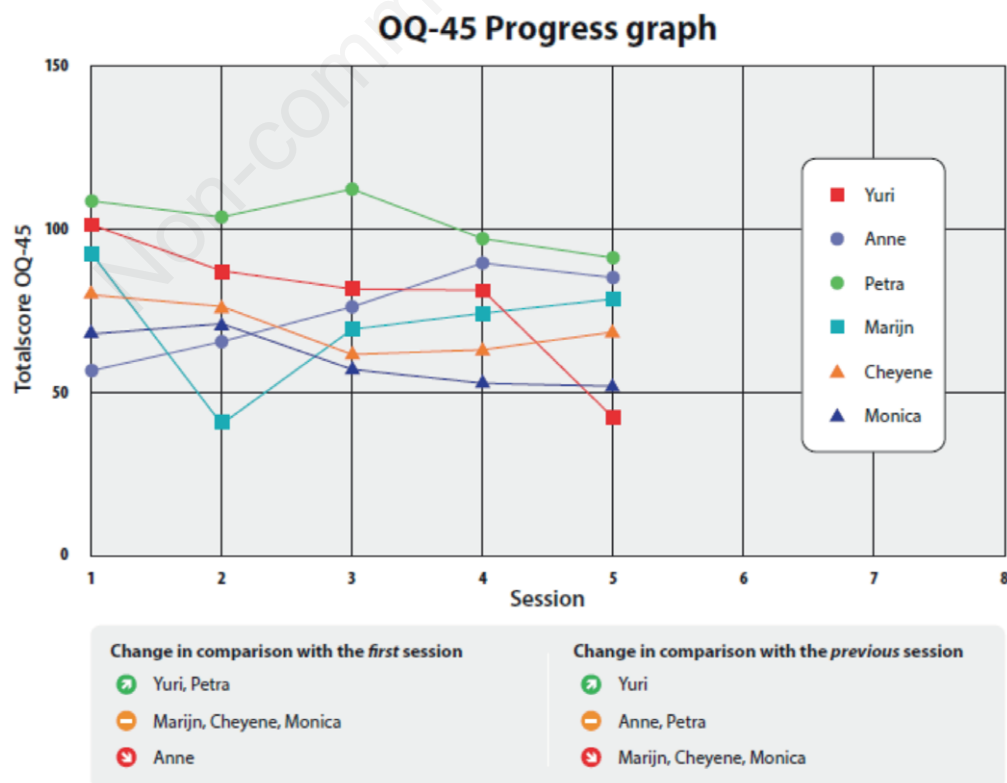


Figure 2. Example progress feedback to therapists.

Outcome measures

Feasibility: dropout and attendance

To test for feasibility of the FIGT-tool, dropout and attendance rates were compared between both research conditions. When a patient discontinued therapy prematurely before the end of a block of sessions, it was defined as dropout. For example, each block of the IPT-G had eight sessions. When a patient stopped treatment before the end of the eight sessions, it was noted as a drop-out. When dropout was higher in the feedback condition compared to TAU, feasibility was considered unsatisfactory. Attendance rate was calculated by the number of actual attended sessions divided by the maximum number of sessions offered $\times 100$. For example, if a patient attended 12 of the 14 CBT-G sessions, attendance rate was $12/14 \times 100 = 85.7\%$. And if a patients followed 3 of the 5 CBT-G sessions, the attendance rate was $3/5 \times 100 = 60\%$.

Feasibility: response rate of the OQ-45 and percentage of experienced feedback discussions

In the feedback condition, the response rate on the OQ-45 per patient was calculated by the number of completed OQ-45s divided by the number of OQ-45s offered $\times 100$. When the mean response rate OQ-45 of all patients in the feedback condition was less than 70%, feasibility was considered insufficient. In other words, when less than 70% of the offered OQ-45s have been completed by patients, the FIGT tool is inadequately used and not feasible for practice.

The percentage of feedback discussions was assessed by an added item to the weekly OQ-45: ‘Has your therapist discussed the results of the progress feedback with you?’ with the answer options yes and no. The number of feedback discussions was computed per patient by dividing the number of confirmed feedback discussions by the number of total completed OQ-45s $\times 100$. When the mean percentage of confirmed feedback discussions was below 70%, the use of feedback was considered insufficient.

Acceptability: experienced usability in the feedback condition

A 13-item user experience questionnaire was developed for patients to rate the FIGT tool on usability, relevance, reliability, specificity, completeness, effectiveness and future use. An example item is ‘Has the feedback given you insight into your treatment progress?’. Items 1-12 were measured on a five-point rating scale ranging from never/not at all (0) to a lot/often (4). Item 13 was a multiple choice question to determine the most desirable improvement for the feedback tool.

A 10-item therapist user questionnaire was designed to assess how therapists rated the FIGT tool on the aspects of helpfulness, relevance, clarity, specificity, usefulness and helpfulness to patients and future use. An example item is: ‘4. How relevant did you find the feedback given?’ A similar five-point scale was used, ranging from

never/not at all (0) to a lot/often (4). When more than 70% of the patients or therapists scored higher than a 2 (neutral) on an item, it was concluded that the acceptability of the tool was sufficient on that aspect. Item 13 was an open ended question ‘Do you have suggestions for improvement for the future?’

Effectiveness of feedback on symptoms and quality of life

A second aim of the study was to assess the effectiveness of the FIGT tool. For this purpose, the Depression Anxiety Stress Scale 21 Revised (DASS-21-R; De Beurs *et al.*, 2001) and the MANchester Short Assessment of quality of life (MANSA, Priebe *et al.*, 1999; Dutch translation; Van Nieuwenhuizen & Koeter, 2000) were used.

The DASS-21-R is a 21-item self-report scale with three subscales that assess levels of depression, stress and anxiety. Each subscale consists of seven items with four response options ranging from 0 (did not apply to me at all) to 3 (applied to me much, or most of the time). Scales are calculated independently with a minimum score of 0 and a maximum score of 42, *i.e.*, each scale is multiplied by two to make scores comparable to the DASS-42 (De Beurs *et al.*, 2001). Higher scores on each subscale indicates elevated depression, anxiety or stress. An example item of the depression scale is ‘I couldn’t seem to experience any positive feelings at all’. The DASS-21-R has been validated for the Dutch population and has a good internal consistency. In the current study, Cronbach’s alpha for the subscales were good: for the depression scale 0.92, the anxiety scale 0.83 and the stress scale 0.88.

The MANSA is a brief instrument for assessing quality of life. The MANSA consists of three sections and in this study we only used the twelve subjective items of the third section. These items ask about satisfaction with life as a whole, job (or sheltered employment, or training/education, or unemployment/retirement), financial situation, number and quality of friendships, leisure activities, accommodation, personal safety, people that the patient lives with (or living alone), sex life, relationship with family, physical health, and mental health. Items were scored on a seven-point rating scale (1=could not be worse, 7=could not be better) and the range of the total scores is between 12 and 84. An example item is ‘How satisfied are you with your mental health?’. A higher total score means more experienced quality of life. Internal consistency has been found to be sufficient ($\alpha=0.74$; Priebe *et al.*, 1999) and good ($\alpha=0.81$; Bjorkman *et al.*, 2005) in earlier studies. In this study the Cronbach’s alpha coefficient was good, namely 0.80.

Procedure

Patients were invited to participate in the study during the standard intake for the IPT-G or CBT-G. If patients agreed to participate, they were asked to sign an informed consent form after at least a week of reflection time. All patients received an email 24 hours prior to the first ses-

sion with the request to fill out the DASS-21-R and MANSA. Patients in the feedback condition were also asked to complete the OQ-45 prior to every group therapy session. A day before the start of the group therapy, therapists were informed about the randomization of their group: feedback or TAU.

In the feedback condition, therapists were asked to take time before every group session to look at the feedback results in the FIGT tool and to spend 10-15 minutes at the beginning of each therapy session to discuss the results in the group setting. Therapists were advised to show the results of the OQ-45 group progress graph groupwise and ask each patient if the results matched their own perception and if there were any explanations for the results. In case patients and therapists concluded that they needed more time deliberating the progress, for instance when a patient's progress was deteriorating, the instruction was to keep this important point in mind and discuss it after each patient's treatment progress was discussed.

Prior to the last session of group psychotherapy, all patients in both research conditions were requested to complete the DASS-21-R and the MANSA. Patients in the feedback condition were asked to complete the 13-item user experience questionnaire. When therapists ended their role as group therapist during the study or were still participating in the study at the end of 2021, they were asked to complete the 10-item user experience questionnaire.

Statistical analyses

IBM SPSS (version 27) was used for all data-analyses, a P-value <0.05 was considered significant for all tests. To test for pre-treatment differences chi square test, Mann Whitney U tests and t-tests were applied. For the feasibility hypotheses, the chi-square test and the t-test were also used. If the assumptions of these tests were violated, the non-parametric Mann-Whitney U or Wilcoxon Matched-Pair tests were conducted. Regarding the acceptability of the FIGT tool, a descriptive analysis was applied.

A multilevel analysis was performed to evaluate the effectiveness of the FIGT tool on symptoms (DASS-21-R) and quality of life (MANSA). Multilevel analysis provides the ability to take nested data and thus interdependency into account, as well as missing values (Peugh, 2010). The model for this study was a two-level analysis with repeated-measurement data on the DASS-21-R and MANSA in time (Level 1) nested within patients (Level 2). An unconditional model was postulated to decide whether the dependent variable (DASS-21-R or MANSA) was affected by the grouping variable 'patient', and thus if multilevel modeling was necessary. Next, a growth model was estimated in which time was added as a fixed parameter. Time was defined as the number of days between pre- and post- measurement of the DASS-21-R/MANSA. Days were used, not treatment sessions, as sessions were not always given weekly because of hol-

idays or other reasons. Next, a multi-level linear model was estimated that included covariates that differed at baseline between the feedback and the TAU condition: gender, treatment modality (IPT/CBT), and education. Next, for each covariate separately, the interaction with time was added, and only included in the model if it resulted in a better fitting model. Finally, condition and the interaction of condition with time was added to the model to test whether FIGT had an effect on outcomes over time. Determining the fit of the model was done by comparing the log-likelihood between two models.

To test for differences between the feedback and the TAU condition for clinical significance, a chi-square test or Mann Whitney U-tests were conducted. A correlation analysis was applied in the exploratory analysis with regard to relationships between feasibility, acceptability and outcomes. Furthermore, in the feedback condition, the number of OT and NOT patients was determined based on the following criterion: NOT means that patients showed an increase of 18 points or more between the current session and the first or previous session at some point in the treatment. Other patients were defined as OT. Then, similar analyses as for the main hypotheses were applied, but now focusing on the differences between OT and NOT patients in the feedback condition.

The effect sizes were calculated for all tests: *r* for chi-square tests with 2×2 contingency table and Mann-Whitney U tests, Cramer's *V* for a chi-squared test with a 2×4 contingency table and Cohen's *d* for t-tests and multilevel analyses. Effect size for the multilevel analyses were computed by Equation 1 or 2. For the feedback and TAU condition, Cohen's *d* was calculated by the difference between the estimate at time point *t* and the baseline estimate divided by the baseline standard deviation. Then, Cohen's *d* of the feedback condition was subtracted from the *d* of the TAU condition.

$$d_{DASS-21-R \text{ subscales}} = \frac{\text{Estimate}_{\text{baseline}} - \text{Estimate}_t}{sd_{\text{baseline}}} \quad (1)$$

$$d_{MANSA} = \frac{\text{Estimate}_t - \text{Estimate}_{\text{baseline}}}{sd_{\text{baseline}}} \quad (2)$$

Results

Preliminary analysis

As shown in Table 1, there were no significant differences between the feedback and TAU condition at baseline with regard to age, primary diagnosis, ethnicity, employment, marital status, presence of comorbid DSM-5 classification, and pre-treatment scores on the DASS-21-R and MANSA. There were, however, differences between the research conditions in terms of gender, treat-

ment modality (IPT/CBT) and education. This means that there were proportionally more men in the feedback condition than in the TAU condition, more patients allocated to IPT-G instead of CBT-G in the feedback condition, and more patients with a lower education level in the feedback condition than in TAU.

Feasibility

Feasibility: dropout and attendance

Mean dropout rates were low in both research conditions and there were no significant differences in median dropout rates between the two research conditions. 8% of the patients in the feedback condition dropped out before the end of treatment and 13% of the patients in the TAU condition, but significant differences between feedback and TAU were not found. Although attendance rates were high in both study conditions, *i.e.* 95% in the feedback condition

and 96% in the TAU condition, patients in the TAU condition showed significantly higher attendance rates compared to patients in the feedback condition ($U(N_{\text{tau}}=93, N_{\text{feedback}}=104)=4070.50, z=-2.24, P=0.03, d=-0.14$).

Feasibility: OQ-45 response rate and percentage confirmed feedback discussions

In the feedback condition, it was found that 68% of OQ-45s were completed by patients which is not significantly different from the 70% cut-off. Of the 104 patients, 62.5% of the patients completed at least 70% of the OQ-45s offered, 37.5% completed all OQ-45s and 11.5% did not fill out any OQ-45. Patients completed at average 9.4 questionnaires ($SD=7.1$) during group treatment. With regard to the percentage of confirmed feedback discussions, patients experienced that the results of the OQ-45 were discussed with them in 52% of the sessions, which is significantly lower than the 70% cut-off, $t(103)=-5.975, P<0.001, d=0.19$ of the

Table 1. Patient characteristics and differences between experimental conditions.

	Feedback n (%) / Mean \pm SD	TAU n (%) / Mean \pm SD	Test statistic	P-value	Effect size
Gender					
Male	46 (44%)	26 (28%)	$\chi^2 [1]=5.607$	0.02	$r=0.17$
Female	58 (56%)	67 (72%)			
Age	41.2 \pm 13.5	40.6 \pm 13.3	$t [195]=-0.307$	0.65	$d=-0.04$
Primary diagnosis					
Major depressive disorder	85 (82%)	75 (81%)	$\chi^2 [1]=0.038$	0.84	$r=0.01$
Anxiety disorder	19 (18%)	18 (19%)			
Ethnicity					
Indigenous origin	92 (88%)	87 (94%)	$\chi^2 [1]=1.530$	0.22	$r=0.09$
Employment					
Employed	54 (52%)	46 (49%)	$\chi^2 [1]=0.183$	0.91	$r=0.03$
Education					
Low	13 (13%)	5 (5%)	$U=4032.50$ $z=-2.253$	<0.05	$d=0.29$
Intermediate	58 (55%)	46 (50%)			
High	33 (32%)	42 (45%)			
Marital status					
Married	24 (23%)	27 (29%)	$\chi^2 [3]=1.867$	0.60	$V=0.10$
Single	37 (36%)	36 (39%)			
Divorced	12 (12%)	10 (11%)			
Other	30 (29%)	20 (21%)			
Comorbid DSM-5 disorder					
Yes	50 (48%)	44 (47%)	$\chi^2 [1]=0.012$	0.92	$r=0.00$
No	54 (52%)	49 (53%)			
Type of followed group psychotherapy					
IPT	47 (45%)	25 (27%)	$\chi^2 [1]=7.098$	<0.01	$r=0.18$
CBT	57 (55%)	68 (73%)			
DASS-21-R pre-test score					
Depression	11.21 \pm 5.31	10.70 \pm 4.9	$t [124]=-0.558$	0.58	$d=-0.10$
Anxiety	7.73 \pm 4.41	67.18 \pm 3.81	$t [124]=-0.735$	0.46	$d=-0.13$
Stress	10.86 \pm 4.76	10.33 \pm 4.32	$t [124]=-0.653$	0.52	$d=-0.12$
MANSA pre-test score					
Total score	50.21 \pm 10.67	50.23 \pm 10.77	$t [124]=0.011$	0.99	$d=0.00$

patients reported that the OQ-45 results were never discussed with them, while 2% of the patients experienced that the OQ-45 results were discussed with them every session. Comparing CBT-G with IPT-G, the percentage of feedback discussions related to the number of sessions was significant higher in the CBT-G than in the IPT-G ($U(N_{\text{CBT-G}}=58, N_{\text{IPT-G}}=46)=990.50, z=-2.26, P=0.02, d=0.45$), with a median in the CBT-G of 58.4% and in the IPT-G a median of 45.0%.

Acceptability

Acceptability by patients

As can be seen in Figure 3, 72% of the patients rated the feedback as reliable (item 6) and specific (item 7), and 67% responded that they used the tool regularly or most of the time (item 2). Less than 70% of the patients scored positively on the other items: 56% of the patients responded that the feedback was complete (item 8), 47% of the patients rated the feedback as contributing to the effect of therapy (item 9) as well as being helpful for the therapeutic relationship (item 10), 41% experienced that the

feedback led to more agreement about the treatment goals (item 11), 37% of the patients found the feedback relevant (item 5) and none of the patients experienced that the feedback has given them insight in the treatment progress (item 4). Furthermore, 42% of the patients reported that the feedback results were discussed with them (item 3) and 30% of the patients would like to use the FIGT tool in the future (item 12). When asked what could be improved (item 13), 37% of the patients preferred an open text field to communicate with the therapist, 24% of patients wanted additional questionnaires to investigate the causes of their symptoms and 18% preferred a specific symptom questionnaire. There were no differences between patients who followed IPT-G or CBT-G.

Acceptability by therapists

As can be seen in Figure 4, more than 70% of the therapists scored a 3 or higher on actual use (item 1), use every session (item 2) and its helpfulness (item 3). Also, 68% rated the feedback as relevant (item 4). Less than 70% of the ther-

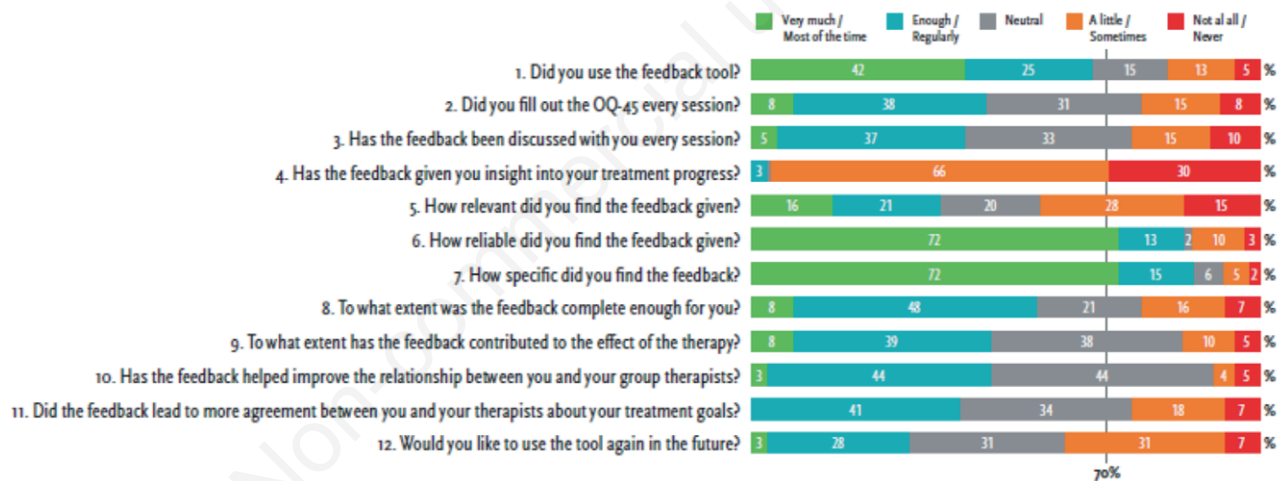


Figure 3. Acceptability of the feedback tool rated by patients (N=61).

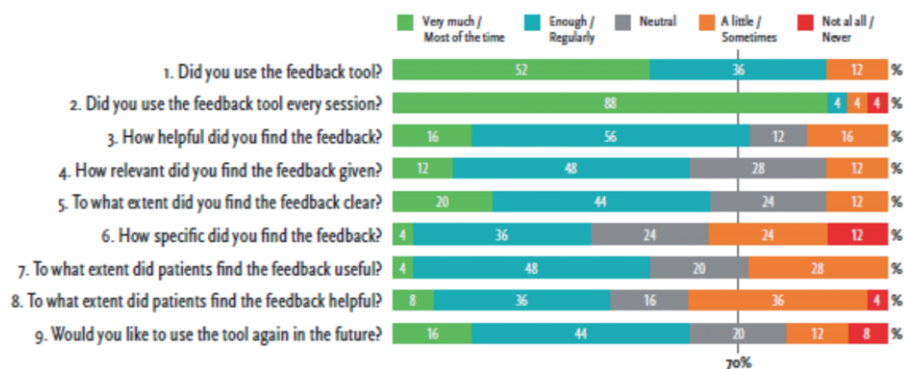


Figure 4. Acceptability of the feedback tool rated by therapists (N=25).

apists scored positively on the other items: 60% of the therapists experienced the feedback as clear (item 5) and would like to use the FIGT tool in the future again (item 9), 52% found the feedback useful for patients (item 7), 44% scored the feedback as helpful for patients (item 8) and 40% experienced the specificity of the feedback as sufficient (item 6). With regard to the question about improvements to the FIGT tool (item 10), six therapists did not respond, seven therapists suggested adding disorder-specific questionnaires and personalized goals to better tailor the feedback to the patient, five therapists needed more training and support in use of the tool, three had suggestions to improve the technical aspects of the tool and one therapist suggested adjusting the threshold of the OQ-45. With regard to the last point, the therapist explained that the recovery threshold of 56 points for patients with severe psychiatric disorders is too low and can be demotivating.

Potential effectiveness

Depression, anxiety and stress symptoms

For all the subscales of the DASS-21-R, separate multilevel analyses were performed, as can be seen in

Table 2. It was found that patients in both the Feedback and TAU conditions showed improvement on symptoms of depression, anxiety and stress at the end of treatment, see Table 3 for the means and standard deviations. Only for the subscale depression, the interaction Time * Condition significantly predicted depressive outcomes. Patients in the feedback condition showed less improvement in depressive symptoms at the end of treatment compared to patients in the TAU condition, $F(1,436.787)=1.123$, $P=0.022$, $d=-0.28$. There were no significant differences in anxiety and stress outcomes between the research conditions.

Quality of life

For the MANSA outcomes, both patients in the feedback and TAU condition improved on quality of life at the end of treatment. Also the interaction Time * Condition was significant: patients in the Feedback condition improved significantly less on quality of life compared to patients in the TAU condition at the end of treatment, $F(1,162.359)=4.567$, $P=0.034$, $d=-0.19$.

Table 2. Fixed and random effect estimates and standard errors for the multilevel linear models for the DASS-21-R and MANSA*.

	Parameter	Estimates (standard error)	P-value
Depression	Fixed effects		
	Intercept (Depression total score)	13.617 (1.051)	<0.01
	Time in days	-0.027 (0.008)	0.02
	Condition * Time in days	0.019 (0.008)	0.03
	Random effects		
	Residual	11.636 (1.433)	<0.01
Patient intercept variance	12.894 (2.287)	<0.01	
Anxiety	Fixed effects		
	Intercept (Anxiety total score)	9.369 (0.850)	<0.01
	Time in days	-0.017 (0.004)	<0.01
	Condition * Time in days	0.008 (0.006)	0.20
	Random effects		
	Residual	0.361 (0.805)	<0.01
Patient intercept variance	11.037 (1.705)	<0.01	
Stress	Fixed effects		
	Intercept (Stress total score)	11.955 (0.796)	<0.01
	Time in days	-0.021 (0.008)	<0.01
	Condition * Time in days	0.015 (0.008)	0.06
	Random effects		
	Residual	10.053 (1.239)	<0.01
Patient intercept variance	4.937 (1.361)	<0.01	
Quality of life	Fixed effects		
	Intercept (MANSA total score)	40.893 (2.341)	<0.01
	Time in days	0.058 (0.010)	<0.01
	Condition * Time in days	-0.029 (0.014)*	0.03
	Random effects		
	Residual	28.479 (3.676)	<0.01
Patient intercept variance	51.410 (7.457)	<0.01	

*Covariates Gender, treatment modality and education are not reported in this table.

Explorative analysis

Clinical significance

In the feedback and TAU conditions, the clinical significance rates on the DASS-21-R and MANSA were calculated, distinguishing four categories: significantly deteriorated, no change, significantly improved, and recovered. No significant differences were found in clinically significant change between the feedback and control conditions.

Correlations

There were no significant correlations between the feasibility indicators (dropout, attendance, OQ-45s response and confirmed feedback discussions) and outcome (DASS-21-R subscales and MANSA), nor significant correlations between the acceptability indicators and outcome.

NOT versus OT patients

In the feedback condition, 39 patients were not-on-track and 54 patients were on-track during treatment. There were no differences between NOT and OT patients with regard to pre-treatment variables (*i.e.*, gender, age, primary diagnosis, ethnicity, employment, education, marital status, comorbid disorders, distribution of followed IPT/CBT, DASS -21-R pre-test scores as well as MANSA pre-test scores), feasibility (*i.e.* attendance, OQ-45 response rate and number of confirmed feedback interviews) and acceptability indicators. For the DASS-21-R subscales, it is found that OT patients im-

proved more on the anxiety and stress subscales in comparison to NOT-patients at the end of treatment, resp. $F(1,95.296)=108.741$, $P<0.001$, $d= -0.12$, and $F(1,120.594)=201.196$, $P<0.001$, $d= -0.34$, see Table 4 for the means and standard deviations of both groups. For the MANSA, it was also found that OT patients improved more on quality of life in comparison to NOT patients at the end of treatment, $F(1,100.215)=743.490$, $P<0.001$, $d= -0.29$, see also Table 4 for the means and standard deviations. With regard to clinical significance, no significant differences were found between the NOT and OT patients.

Discussion and conclusions

This study examined whether the use of a FIGT tool, developed using The Contextual Feedback Intervention Theory (Sapyta *et al.*, 2005) and recommendations from previous FIGT studies, is feasible and acceptable for patients with depressive or anxiety symptoms and their therapists in IPT-G and CBT-G. In terms of feasibility, as expected, the use of feedback did not lead to differences in dropout compared to TAU. The attendance percentage in group psychotherapy with the FIGT tool was lower than TAU without feedback, but attendance percentages were high in both conditions. In the feedback condition, patients completed nearly 70% of all offered OQ-45s, but only, experienced that the feedback results were actually discussed with them in half of the cases. Regarding the acceptability indicators, it appears that therapists were generally more positive about the use of feedback than patients. Therapists indicated that they used the feedback almost every session and the majority of therapists rated

Table 3. Mean pre- and post-test scores and standard deviations of the feedback and TAU condition on the DASS-21-R and MANSA.

	Feedback Mean±SD				TAU Mean±SD			
	<i>n</i>	Pre-test	<i>n</i>	Post-test	<i>n</i>	Pre-test	<i>n</i>	Post-test
DASS-21-R score								
Depression	66	11.06±5.36	72	9.28±5.75	60	10.95±4.96	64	7.61±5.30
Anxiety	66	7.92±4.42	72	6.71±4.55	60	7.77±4.23	64	5.84±3.92
Stress	66	11.06±4.61	72	9.36±4.47	60	10.38±4.32	64	8.50±3.87
MANSA score	66	47.17±9.10	73	51.15±10.11	60	47.46±7.29	66	53.28±9.55

Table 4. Pre- and post-test scores and standard deviations (SD) of the OT and NOT patients on the DASS-21-R and MANSA.

	OT (<i>n</i> =54) Mean±SD		NOT (<i>n</i> =39) Mean±SD	
	Pre-test	Post-test	Pre-test	Post-test
DASS-21-R score				
Depression	11.06±5.36	8.51±5.78	10.95±4.96	9.57±5.67
Anxiety	7.92±4.42	6.46±4.29	7.77±4.23	6.86±4.81
Stress	11.06±4.61	8.89±4.13	10.38±4.32	9.80±4.79
MANSA score	47.17±9.10	52.83±10.26	47.46±7.29	49.86±9.30

the FIGT tool as helpful, relevant and clear. A lower percentage of therapists believed that the feedback was useful and helpful to patients and found the feedback specific enough. Most patients, on the other hand, found the feedback to be both specific and reliable and also used the FIGT tool regularly. Relatively lower percentages of patients found the feedback complete and a contribution to the effect of therapy, the therapeutic relationship and agreement on therapeutic goals. None of the patients experienced the feedback as helpful in gaining insight into the therapeutic process. Furthermore, it appears that the majority of therapists (60%) and minority of patients (30%) would prefer to use the tool again in the future, implying that the developed FIGT tool mainly contributed to the therapist's process and less to the process of the patient. Finally, using the FIGT tool in this study appears to have no effects on improvement in anxiety and stress at the end of treatment, but may have negative effects on the degree of improvement in depressive symptoms and quality of life compared to TAU. Overall, the FIGT tool developed appears to be partially feasible and acceptable for patients and therapists in group therapy and does not improve therapy outcomes in terms of symptoms and quality of life and may potentially have a negative effect on outcomes.

In line with the FIGT study by Hutson and colleagues (2020), the FIGT tool appears to have been insufficiently used for its intended purpose: actively discussing treatment progress and making timely adjustments for NOT-patients so that patients get (back) on-track. Hutson and colleagues found that patients who were actively engaged in the discussion of the feedback results showed more consistent recovery rates. But group therapists in their study used the feedback mostly for patients who already benefited from treatment, rather than for NOT patients, indicating that the feedback was not used for the patients who needed it the most. Therapists in our study reported that they used the feedback every session, but patients only experienced in half of the cases that the feedback was actively discussed with them. It is possible that therapists have indeed not discussed the results sufficiently, but it is also possible that patients have different expectations than therapists when feedback is actually discussed. We also tested for differences between NOT and OT patients and found no differences in number of discussions perceived by patients, meaning that therapists discussed the results evenly with NOT and OT patients, in contrast to the study by Hutson and colleagues. However, NOT patients experienced significantly more symptoms and a lower quality of life at the end of treatment compared to OT patients, showing that feedback did not lead to better outcomes for the group that needed it most.

None of the seven FIGT studies so far report negative effects of feedback and only one study (Davidsen *et al.*, 2017) found no effect at all. It is surprising that the FIGT tool in this study did not have the intended effect, as both

the Contextual Feedback Theory (Sapyta *et al.*, 2005) and previous FIGT research on contextual factors were included in the development of the FIGT tool. For example, in contrast to the earlier study by Davidsen and colleagues, therapists in the current study had more flexibility to adjust their protocol based on the feedback results, but this did not lead to the desired effect. As described by Lewis and colleagues (2019), it is possible that other contextual factors may play a role as well, namely patient and therapist characteristics, which were probably not considered in this study. In terms of patient characteristics, it is clear that the FIGT tool did not sufficiently meet the needs of the patients. In this study, patients with long-lasting and recurrent mood and anxiety disorders participated and 50% of them had one or more comorbid disorders. Although patients with comorbid disorders can benefit just as much from treatment as patients without, it is found that the severity of symptoms at the start and end of treatment are higher (Banyard *et al.*, 2021). For these patients, the OQ-45 alone can be demotivating if they are regularly confronted with high scores. As Sapyta and colleagues (2005) suggest in their Contextual Feedback Intervention Theory, feedback needs to be quick, reliable, and specific, but also needs to fit one's goals. Patients with more severe symptoms may not have the goal of achieving full symptom recovery and have goals in other areas of life. Furthermore, patients in general have different needs than therapists for a useful feedback system, which also confirms that the OQ-45 alone is probably not sufficient for the patients in this study. Patients prefer a feedback system that is flexible and adapts to their situation, gives a holistic view of their situation in addition to symptoms, includes personal treatment goals, empowers and promotes collaboration with their therapist (Koementas-de Vos *et al.*, 2022; Moltu *et al.*, 2016; Solstad *et al.*, 2017).

Therapists' characteristics and attitudes toward feedback can also influence the actual use of feedback and its effectiveness. De Jong *et al.* (2013) found that feedback is more effective when therapists have an open attitude towards feedback, high self-efficacy and are committed to use feedback. Maladaptive attitudes are associated with less effective use of feedback, *e.g.* 'It takes too much time and effort', 'standardized measures are not as accurate as my own clinical judgement' and 'data could be used against me' (Lewis *et al.*, 2019). This study did not specifically reveal such attitudes, although attitudes were not asked directly. The therapists in this study were even more positive about the FIGT tool than patients, and most therapists would prefer to use the tool again in the future. It is likely that maladaptive attitudes do not explain the results, but rather a lack of knowledge about how to use the feedback effectively, which was also found in a previous study by Koementas-de Vos and colleagues (2022). In addition, it is possible that the feedback results provided insufficient information to therapists to help them how to improve the effects of therapy. For example, in individual

therapeutic settings, the effect of feedback has been found to increase when a clinical support tool is offered (Shimokawa *et al.*, 2010).

Not mentioned by Lewis and colleagues (2019) specifically, but group characteristics may play a role as well. In this study, the research conditions were not equally distributed in CBT and IPT groups despite randomization, so that more patients in the IPT group appeared to be in the feedback condition. There is some evidence that feedback works better with symptom-specific therapies, such as CBT, where monitoring of thoughts, emotions and behaviour is standard (Janse *et al.*, 2020). IPT is also a symptom-specific therapy but focuses mainly on changing interpersonal functioning and less on systematic monitoring of intrapersonal aspects. It is possible that using feedback in IPT is less effective than in CBT. In the feedback condition, it was also found that patients in the IPT-G experienced relatively less feedback discussions than patients in the CBT-G. This could mean that therapists discussed the results more often in CBT-G, but it could also mean that patients in CBT-G, where the use of registration exercises is standard, notice the feedback discussions more. Specific expectations about group processes could play a negative role as well. It appears that therapists may feel that discussing feedback in groups creates feelings of competition and insecurity in patients, which prevents them from discussing results in the group, which may explain the low number of confirmed feedback discussions in this study (Koementas-de Vos *et al.*, 2022). In contrast, patients prefer that feedback results are discussed in the group so they can learn from each other (Koementas-de Vos *et al.*, 2022). It is possible that therapists are insufficiently aware of this need and discuss the feedback results less actively in the group. It is therefore important to inform and train therapists so that patients may benefit from feedback in the group.

A major criticism of the FIGT studies in general is that the results should be interpreted with caution due to various kinds of bias. Due to the interdependence between observations in groups, for example, large samples are needed to obtain sufficient power. The randomized FIGT studies published to date have too small sample sizes and/or overestimate the a priori effect sizes. Non-randomized FIGT studies have been published with large sample sizes, but these studies risk other biases, such as selection bias. Furthermore, the lack of publications to date on negative effects of FIGT may indicate publication bias or allegiance effects in FIGT studies (Lambert *et al.*, 2018; Østergård *et al.*, 2020). It is clear that the effects of researchers' allegiance on obtaining positive results are not the case in this study. Replication of the present study is needed, but for ethical reasons the FIGT tool should be adjusted so that intentional harm can be prevented. The following adjustments are advisable: actively promoting feedback discussions between therapists and patients, adding an open text field for patients in the FIGT tool,

using specific questionnaires and personal treatment goals, adding a clinical support tool and providing training and supervision to therapists.

There are several limitations of this study. Despite randomization, there were differences in gender, modality (IPT/CBT) and educational level between the research conditions. In the multilevel analysis, we corrected for these differences by adding these variables as covariates, but it is unclear to what extent these differences played a role in the feasibility and acceptability results. In addition, the power of the multilevel analysis was too low. To have sufficient power, a minimum of 300 subjects per study condition was required, which was not possible within our study setting. The results should therefore be interpreted with caution. In the future, another design may be appropriate, such as a multi-baseline design, or a multicentre study to test the effectiveness of the use of feedback in group therapy in a more reliable way. Finally, this study does not include specific moderators, such as the degree of self-efficacy, fidelity and the ability of the therapist to use feedback effectively, or attitudes to feedback from both patients and therapists.

In conclusion, this study indicates that feedback-informed group treatment with our FIGT tool is only partially feasible and not effective in improving treatment outcomes. There is room for improvement in the design of the FIGT tool and in the way it is utilized. Attunement to the needs of both patient and therapist is required to support them as well as possible in the therapeutic process. More research is needed to discover how, for whom and in which setting feedback is feasible, effective, ineffective or even harmful.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Banyard, H., Behn, A.J. & Delgadillo, J. 2021. Personality disorders and their relation to treatment outcomes in cognitive behavioural therapy for depression: a systematic review and meta-analysis. *Cognitive Therapy Research*, 45, 561-576. doi:10.1007/s10608-021-10203-x
- Bickman L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47(10), 1114-1119. doi:10.1097/CHI.0b013e3181825af8.
- Burlingame, G. M., Seebeck, J. D., Janis, R. A., Whitcomb, K. E., Barkowski, S., Rosendahl, J., & Strauss, B. (2016). Outcome differences between individual and group formats when identical and nonidentical treatments, patients, and doses are compared: A 25-year meta-analytic perspective. *Psychotherapy (Chicago, Ill.)*, 53(4), 446-461. doi:10.1037/pst0000090.
- Burlingame, G. M., Whitcomb, K. E., Woodland, S. C., Olsen, J. A., Beecher, M., & Gleave, R. (2018). The effects of relationship and progress feedback in group psychotherapy using the Group Questionnaire and Outcome Questionnaire-

- 45: A randomized clinical trial. *Psychotherapy (Chicago, Ill.)*, 55(2), 116-131. doi:10.1037/pst0000133.
- Byrne, S. L., Hooke, G. R., Newnham, E. A., & Page, A. C. (2012). The effects of progress monitoring on subsequent readmission to psychiatric care: a six-month follow-up. *Journal of Affective Disorders*, 137(1-3), 113-116. doi:10.1016/j.jad.2011.12.005.
- Chapman, C. L., Burlingame, G. M., Gleave, R., Rees, F., Beecher, M., & Porter, G. S. (2012). Clinical prediction in group psychotherapy. *Psychotherapy research: journal of the Society for Psychotherapy Research*, 22(6), 673-681. Doi:10.1080/10503307.2012.702512.
- Davidson, A. H., Poulsen, S., Lindschou, J., Winkel, P., Trøndarson, M. F., Waadegaard, M., & Lau, M. (2017). Feedback in group psychotherapy for eating disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 85(5), 484-494. doi:10.1037/ccp0000173.
- De Beurs, E., Van Dyck, R., Marquenie, L. A., Lange, A., & Blonk, R. W. (2001). De DASS: een vragenlijst voor het meten van depressie, angst en stress. *Gedragstherapie*, 34(1), 35-54.
- De Jong, K., Conijn, J. M., Gallagher, R., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, 85, 102002. doi:10.1016/j.cpr.2021.102002.
- De Jong, K., Nugter, M. A., Lambert, M. J., & Burlingame, G. M. (2008). *Handleiding voor afname van de Outcome Questionnaire OQ-45.2*. Salt Lake City, UT: OQ Measures LLC.
- De Jong, K., Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heiser, W. J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology & Psychotherapy*, 14(4), 288-301. doi:10.1002/cpp.529.
- De Jong, K., van Sluis, P., Nugter, M. A., Heiser, W. J., & Spinhoven, P. (2012). Understanding the differential impact of outcome monitoring: therapist variables that moderate feedback effects in a randomized clinical trial. *Psychotherapy research: journal of the Society for Psychotherapy Research*, 22(4), 464-474. doi:10.1080/10503307.2012.673023.
- Fuhriman, A., & Burlingame, G. M. (1990). Consistency of matter: a comparative analysis of individual and group process variables. *The Counseling Psychologist*, 18(1), 6-63. doi:10.1177/001100009018100.
- Gleave, R. L., Burlingame, G. M., Beecher, M. E., Griner, D., Hansen, K., & Jenkins, S. A. (2017). Feedback-informed group treatment: Application of the OQ-45 and Group Questionnaire. In *Feedback-informed treatment in clinical Practice: Reaching for excellence* (pp. 141-166). Washington, DC: American Psychological Association.
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, 61(2), 155-163. doi:10.1002/jclp.20108.
- Hatfield, D., McCullough, L., Frantz, S. H., & Krieger, K. (2010). Do we know when our clients get worse? an investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy*, 17(1), 25-32. doi:10.1002/cpp.656.
- Hutson, J., Hooke, G. R., & Page, A. C. (2020). Progress monitoring and feedback delivered in routine psychiatric care: Beneficial but not reaching those thought to need it most. *Psychotherapy research: journal of the Society for Psychotherapy Research*, 30(7), 843-856. doi:10.1080/10503307.2019.1710305.
- Koentmas-de Vos, M. W., van Dijk, M., Tiemens, B., de Jong, K., Witteman, C. L. M. & Nugter, M. A. (2022). Feedback-informed group treatment: a qualitative study of the experiences and needs of patients and therapists. *International Journal of Group Psychotherapy*, 72 (3), 193-227. doi:10.1080/00207284.2022.2086557.
- Koentmas-de Vos, M., Nugter, M. A., Engelsbel, F., & De Jong, K. (2018). Does progress feedback enhance the outcome of group psychotherapy?. *Psychotherapy (Chicago, Ill.)*, 55(2), 151-163. doi:10.1037/pst0000164.
- Lambert, M. J., Kahler, M., Harmon, C., Burlingame, G. M., Shimokawa, K., White, M. M. (2013). *Administration and Scoring Manual: Outcome Questionnaire OQ®-45.2*. Salt Lake City, UT: OQMeasures.
- Lambert, M. J., & Lo Coco, G. (2013). Simple methods for enhancing patient outcome in routine care: Measuring, monitoring, and feedback. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 16(2), 93-101. doi:10.4081/ripppo.2013.147.
- Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy (Chicago, Ill.)*, 55(4), 520-537. doi:10.1037/pst0000167.
- Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., Douglas, S., Simon, G., & Kroenke, K. (2019). Implementing measurement-based care in behavioral health: a review. *JAMA Psychiatry*, 76(3), 324-335. doi:10.1001/jamapsychiatry.2018.3329.
- Lutz, W., De Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in psychotherapy: Where are we and where do we want to go?. *Psychotherapy research: journal of the Society for Psychotherapy Research*, 25(6), 625-632. doi:10.1080/10503307.2015.1079661.
- Newnham, E. A., Hooke, G. R., & Page, A. C. (2010). Progress monitoring and feedback in psychiatric care reduces depressive symptoms. *Journal of Affective Disorders*, 127(1-3), 139-146. doi:10.1016/j.jad.2010.05.003.
- Østergård, O. K., Randa, H., & Hougaard, E. (2020). The effect of using the partners for change outcome management system as feedback tool in psychotherapy-A systematic review and meta-analysis. *Psychotherapy research: journal of the Society for Psychotherapy Research*, 30(2), 195-212. doi:10.1080/10503307.2018.1517949.
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). (2016). *Official Journal*, L 119, 1-88.
- Peugh J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48(1), 85-112. doi:10.1016/j.jsp.2009.09.002.
- Priebe, S., Huxley, P., Knight, S., & Evans, S. (1999). Application and results of the Manchester Short Assessment of Quality of Life (MANSA). *The International Journal of Social Psychiatry*, 45(1), 7-12. doi:10.1177/002076409904500102.
- Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to cli-

- nicians: theory, research, and practice. *Journal of Clinical Psychology*, 61(2), 145-153. doi:10.1002/jclp.20107.
- Schuman, D. L., Slone, N. C., Reese, R. J., & Duncan, B. (2015). Efficacy of client feedback in group psychotherapy with soldiers referred for substance abuse treatment. *Psychotherapy research: journal of the Society for Psychotherapy Research*, 25(4), 396-407. doi:10.1080/10503307.2014.900875.
- Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, 78(3), 298-311. doi:10.1037/a0019247.
- Slone, N. C., Reese, R. J., Mathews-Duvall, S., & Kodet, J. (2015). Evaluating the efficacy of client feedback in group psychotherapy. *Group Dynamics: Theory, Research, and Practice*, 19(2), 122-136. doi:10.1037/gdn0000026.
- Strauss, B., Burlingame, G. M., & Bormann, B. (2008). Using the CORE-R battery in group psychotherapy. *Journal of Clinical Psychology*, 64(11), 1225-1237. doi:10.1002/jclp.20535.
- Timman, R., de Jong, K., & de Neve-Enthoven, N. (2017). Cut-off scores and clinical change indices for the Dutch Outcome Questionnaire (OQ-45) in a large sample of normal and several psychotherapeutic populations. *Clinical Psychology & Psychotherapy*, 24(1), 72-81. doi:10.1002/cpp.1979.
- Nieuwenhuizen, C. van, Janssen-de Ruijter, E.A.W., & Nugter, M. (2017). *Handleiding Manchester Short Assessment of Quality of Life (MANSA)*. Eindhoven: Stichting QoLM.
- Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological reports*, 110(2), 639-644. doi:10.2466/02.07.17.PR0.110.2.639-644.
- Weissman, M. M., Markowitz, J. C., & Klerman, G. L. (2017). *The guide to interpersonal psychotherapy: updated and expanded edition*. Oxford University Press.
- Whitfield, G. (2010). Group cognitive-behavioural therapy for anxiety and depression. *Advances in Psychiatric Treatment*, 16(3), 219-227. doi:10.1192/apt.bp.108.005744.
- Wiebe, D. E., Remers, S., Nippak, P., & Meyer, J. (2021). Evaluation of an online system for routine outcome monitoring: cross-sectional survey study. *JMIR Mental Health*, 8(12), e29243. doi:10.2196/29243.