Describing Therapeutic Relationship Change and Failure in Group Psychotherapy

Harold Thomas Svien

Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

BYU ScholarsArchive Citation
Svien, Harold Thomas, "Describing Therapeutic Relationship Change and Failure in Group Psychotherapy" (2019). Theses and Dissertations. 8570.
https://scholarsarchive.byu.edu/etd/8570

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Describing Therapeutic Relationship Change and

Failure in Group Psychotherapy

Harold Thomas Svien

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Gary M. Burlingame, Chair
Patrick Steffen
Joseph A. Olsen

Department of Psychology
Brigham Young University

Copyright © 2019 Harold Thomas Svien
All Rights Reserved
ABSTRACT

Describing Therapeutic Relationship Change and Failure in Group Psychotherapy

Harold Thomas Svien
Department of Psychology, BYU
Master of Science

Objectives. This study reanalyzed data from Burlingame and colleagues’ (2018) randomized controlled trial on the effect of adding Group Questionnaire (GQ) to Outcome Questionnaire (OQ-45) feedback. These data were assessed for the feedback effect using the amount of GQ alerts in one session reported by the group member to track change in GQ subscales as a measure of reversing therapeutic relationship failure.

Methods. 374 participants engaged in 58 psychotherapy groups. Every participant provided GQ measurements after every group session. These GQ measurements formed ‘person-session units’ representing whether or not each type of alert was present following each group meeting. Person-session units showing one, two, and three or more GQ alerts were selected for analysis. The GQ subscales of positive bond (PB), positive work (PW), and negative relationship (NR) were tracked over the following two sessions using hierarchical linear models (HLMs) to correct for group membership and analyze slopes of change between GQ feedback and no-feedback conditions.

Results. Insignificant results were shown in condition by session interactions for every GQ subscale following every specified amount of co-occurring GQ alerts. These results contrast with Burlingame and colleagues’ (2018) findings that half of all condition by session interactions shown were significant using GQ change and status alerts to trigger analyses.

Conclusions. The results of this study do not appear to better discriminate the effect of adding GQ to OQ feedback for group leaders. Thus, it does not appear that group leaders can better reverse the tide of relationship failure in psychotherapy groups when there are specific numbers of GQ alerts presented to them versus the alert types offered in GQ feedback reports.

Keywords: group questionnaire, mental health, psychotherapy, outcome questionnaire, relationship failure
TABLE OF CONTENTS

Abstract ........................................................................................................................................... ii

List of Figures .................................................................................................................................. v

List of Tables ................................................................................................................................... vi

Describing Therapeutic Relationship Change and Failure in Group Psychotherapy ............... 1

Routine Outcome Monitoring ......................................................................................................... 2

ROM in Individual Therapy ............................................................................................................ 3

ROM in Group Therapy .................................................................................................................... 8

Burlingame and Colleagues ......................................................................................................... 11

The Current Study ......................................................................................................................... 12

Aims and Hypotheses ..................................................................................................................... 13

Methods ...................................................................................................................................... 13

Participants ................................................................................................................................. 13

Psychotherapy Groups ................................................................................................................. 14

Measures .................................................................................................................................. 15

Procedures ................................................................................................................................. 15

Analysis ................................................................................................................................... 16

Results ..................................................................................................................................... 17

Configuration of Alerts ................................................................................................................ 17

Alert Patterns ............................................................................................................................... 18
Alert Number Effects ............................................................................................................ 19

Discussion ................................................................................................................................. 22

References ..................................................................................................................................... 25
LIST OF FIGURES

Figure 1. Formula for computing reliable change indices ................................................................. 4

Figure 2. Example of an OQ-45 individual client report ................................................................. 5

Figure 3. Sample GQ/OQ report. ................................................................................................. 11
LIST OF TABLES

Table 1  Frequency of Patterns of Signal Alerts for Scales of the ASC. n = 107 ................. 7

Table 2  Single and Double Change and Status Group Questionnaire (GQ) Alert Combinations. n = 508 ................................................................. 17

Table 3  Triple and Above Change and Status Alert Combinations. n = 212 ......................... 18

Table 4  Results of Multilevel Models Testing Group Questionnaire (GQ) Subscale Change Over Time for Therapeutic Encounters with One GQ Alert..................................................... 19

Table 5  Results of Multilevel Models Testing Group Questionnaire (GQ) Subscale Change Over Time for Therapeutic Encounters with Two GQ Alerts................................. 20

Table 6  Results of Multilevel Models Testing Group Questionnaire (GQ) Subscale Change Over Time for Therapeutic Encounters with Three or More GQ Alerts ..................... 21

Table 7  Comparison of Slope p-Values Following Specified Alert(s) by Study .................... 22
Describing Therapeutic Relationship Change and Failure in Group Psychotherapy

The process of providing continual feedback to therapists over the course of psychotherapy is now well-defined, accepted, and has ‘come of age’ related to its use in mental health delivery systems (Wampold, 2015). The essence of routine outcome monitoring (ROM) in these contexts focuses on continual evaluation of the client’s overall mental health functioning (Lambert, 2015). As Wampold noted, the use of ROM is producing considerable data on how psychotherapy works in naturalistic settings. The current study examines feedback from the OQ Analyst (OQ-A) system, which has the most robust existing evidence basis (Burlingame et al., 2018).

The aim of this study is to assess the effect of providing ROM feedback to group psychotherapy leaders. Existing research shows considerable evidence that ROM feedback augments treatment outcomes (Shimokawa, Lambert, & Smart, 2010). However, most existing ROM measures focus only on individual psychotherapy. Previous research has only begun to assess providing feedback to group treatment leaders (Burlingame et al., 2018).

ROM feedback can be augmented by the use of clinical tools that assess reliable moderators of treatment success and give targeted problem-solving considerations to the therapist (Lambert, 2015). Prior research indicates that clients have reliable types of problems on these moderators, shown by such assessment tools (White et al., 2015). Other research has shown that the use of these tools results in better outcomes for patients at risk of treatment failure (Whipple et al., 2003).

Burlingame and colleagues (2018) were the first to evaluate a feedback tool designed specifically for group psychotherapy (i.e., the Group Questionnaire [GQ]; Burlingame et al.,
RELATIONSHIP FAILURE IN GROUP PSYCHOTHERAPY

2017). The GQ is similar to individual therapy clinical tools, and has been shown to have criterion validity with other measures of the therapeutic relationship. However, its use is different than individual clinical tools in that the group relationship is continually assessed on a session by session basis, while individual support tools are only administered when possible treatment failure is indicated by feedback. The GQ was designed to function as a warning system for relationship failure in group therapy (Thayer & Burlingame, 2014; Burlingame et al., 2017).

The current study focuses on secondary analysis of a randomized controlled trial (RCT) testing progress feedback from the 45-item Outcome Questionnaire (OQ-45) against OQ-45 and GQ feedback in treatment groups (Burlingame et al., 2018). Burlingame and colleagues compared clients randomly assigned to conditions to determine if GQ feedback led to reduced rates of therapeutic relationship deterioration and failure after therapists were alerted to relationship challenges. Their findings were mixed: GQ feedback resulted in improvements on two subscales of positive bonding and positive working in the group, but not on the subscale tapping negative feelings toward the group. Their conclusions were qualified by the fact that analyses took place after an alert on a subscale regardless of co-occurring alerts that could have also taken place. This, along with the lack of uniformity in the feedback effect, led Burlingame and colleagues to ask if their results could be broken down further by trends in alerts.

The question involved in this study is to determine if there is a more careful analysis of alerting types relates to the efficacy of outcome feedback; namely, does the number of co-occurring alerts reliably predict differences in the effect of feedback?

Routine Outcome Monitoring

The use of ROM in mental health treatment involves clinicians receiving progress feedback on their clients who complete an outcome assessment on a session by session basis.
RELATIONSHIP FAILURE IN GROUP PSYCHOTHERAPY

(Shimokawa et al., 2010). Exactly how therapists use that feedback on outcome measures is somewhat unclear, and varies by the ROM in question (Wampold, 2015); however, notable evidence exists indicating feedback is effective.

Research on ROM systems has shown the efficacy of using ROM feedback to enhance psychosocial treatment outcomes. A review of existing RCTs concerning ROM feedback vs. no feedback conditions found that a majority of studies reviewed showed positive effects when including any ROM measure during mental health treatment (Carlier et al., 2012). Results also show feedback decreasing the need for further sessions in clients who respond to treatment and feedback keeping clients who are at risk for treatment failure in treatment longer, resulting in better outcomes (Shimokawa et al., 2010; Whipple et al., 2003).

Two ROM systems have gained sufficient evidence bases to be included in the Substance Abuse and Mental Health Services Administration’s National Registry of Evidence-based Programs and Practices (NREPP; SAMHSA, 2015): the OQ-A system and the Partners for Change Outcome Management System (PCOMS; Burlingame et al., 2018). This review will focus on the OQ-A system, which has a strong evidence basis for preventing deterioration of therapy clients identified to be at-risk for treatment failure and improving outcomes (Wampold, 2015).

ROM in Individual Therapy

The majority of research has been completed on ROM systems in individual therapy. Thus, ROM feedback for individual therapy clients in the OQ-A system will be discussed prior to ROM feedback in group treatment.

**OQ-A routine monitoring.** The OQ-A system uses a global measure of psychiatric distress: the OQ-45’s total score (Lambert et al., 2013), which tracks distress using data collected
before the client’s weekly session of psychotherapy, as the basis for feedback. Lambert’s (2015) review summarized outcome trials of the OQ-45 and found an average effect size just below the medium threshold ($d = 0.48$) across trials.

The OQ-A system uses normative data from undergraduate, community, inpatient, and outpatient samples to provide three types of alerts to clinicians: reliable change from the client’s initial distress level, change from clinical to non-clinical levels of distress, and change compared to an expected course of recovery (Lambert et al., 2013).

The first type of alert concerns reliable change from the client’s presenting level of distress determined using Jacobson and Truax’s (1991) concept of a reliable change index (RCI). The formula for computing RCIs is shown below in Figure 1.

$$\text{RCI} = \frac{(\text{pre-treatment}) - (\text{post-treatment})}{\text{S}_{\text{diff}}} = 1.96$$

Where $\text{S}_{\text{diff}} = \sqrt{2\text{S}_{\text{E}}^2}$ and $\text{S}_{\text{E}} = \text{SD} \sqrt{1 - r_{xx}}$

*Figure 1.* Formula for computing reliable change indices.

The RCI for the total OQ-45 score is 14 points of deviation from the client’s initial distress level in either direction (Lambert et al., 2013). This feedback appears in the top right of Figure 2, with no reliable change indicated from the client’s initial distress level.
Figure 2. Example of an OQ-45 individual client report.

The second type of alert uses a weighted midpoint to determine a score between clinical and nonclinical distress (at or above an OQ-45 value of 64 for total distress in the clinical range). Figure 2 shows distress consistently above the clinical cutoff, indicated by the dotted yellow line in the graph portion of the figure.
The third type of alert in the OQ-A system uses expected recovery curves (ERCs) based on comparable clients assessed at intake (Probst, Lambert, Loew, Dahlbender, & Tritt, 2015). Clients deviating from ERCs in a negative direction of at least one standard deviation are called not-on-track (NOT; Probst et al., 2015) cases. The OQ-A system uses color-coded progress alerts, with NOT cases being flagged by yellow and red alerts that indicate high and mixed chance of negative treatment outcome, on-track white alerts indicating distress below the clinical cutoff, and green indicating expected progress (Burlingame et al., 2017). Figure 2 illustrates a client on-track for positive outcomes with a green progress alert in the top right portion of the graph.

**OQ-A clinical support tools.** NOT clients, as Whipple et al. (2003) noted, often have poor outcomes even when their therapist is given feedback. This finding led to research on assessing moderators (e.g., the working alliance) of change that had a positive relationship to successful treatment outcomes (Whipple et al., 2003). These assessments are called clinical support tools (CSTs; Whipple et al., 2003), which assess why each client can be considered NOT for positive outcomes in psychotherapy (Burlingame et al., 2017).

The CST used in the OQ-A system for individual therapy, called the Assessment for Signal Clients (ASC; Witold, Lambert, Harris, Busath, & Vazquez, 2012), provides stepwise problem solving to assess for therapeutic and extratherapeutic factors impeding treatment. The ASC focuses on problems in the therapeutic alliance, level of motivation for change in psychotherapy, social support in the client’s life, and distressing life events to accomplish this (Witold et al., 2012). The ASC alerts therapists to clients significantly below normative ranges in those domains, with the intent to reverse treatment failure and increase positive outcomes (White et al., 2015). ASC alerts are also color-coded when clients’ reported levels for each scale are
below a cutoff value, similar to the OQ-45’s total score and subscales (see Lambert [2015] or an excellent example).

**Characterizing ASC alerts.** The investigation by White and colleagues (2015) explained patterns on the ASC for clients who signaled as a NOT case. Their findings included three general types of off-track clients: those with alliance/motivational difficulties, those with social support/life event problems, and those with no distinguishable pattern (White et al., 2015). White and colleagues’ results of their cluster analysis are recreated below in Table 1.

<table>
<thead>
<tr>
<th>ASC scales(s) with signal alert</th>
<th>Total number (percentage) of participants with signal alerts for scale(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No signal alerts for ASC scales</td>
<td>45 (41.4%)</td>
</tr>
<tr>
<td>SS (only)</td>
<td>20 (18.69%)</td>
</tr>
<tr>
<td>LE (only)</td>
<td>9 (8.4%)</td>
</tr>
<tr>
<td>M (only)</td>
<td>6 (5.6%)</td>
</tr>
<tr>
<td>TA (only)</td>
<td>5 (4.6%)</td>
</tr>
<tr>
<td>SS and M</td>
<td>4 (3.7%)</td>
</tr>
<tr>
<td>TA, SS, and M</td>
<td>4 (3.7%)</td>
</tr>
<tr>
<td>TA and M</td>
<td>4 (3.7%)</td>
</tr>
<tr>
<td>TA and LE</td>
<td>3 (2.8%)</td>
</tr>
<tr>
<td>TA and SS</td>
<td>2 (1.8%)</td>
</tr>
<tr>
<td>TA, SS, M, and LE</td>
<td>1 (0.9%)</td>
</tr>
</tbody>
</table>

**Note.** TA = therapeutic alliance; M = motivation; SS = social support; LE = life events

The findings by White and colleagues are a first step of characterizing what areas of treatment problems are most frequently endorsed by clients. However, their analysis focused on a subset of clients that are NOT. White and colleagues’ study leads to the question of what patterns of alerts one would find in group psychotherapy.
ROM in Group Therapy

As ROM feedback is shown to be effective, consideration of differences by treatment format are salient. That is, the effect of feedback does not vary when considering individual versus group psychotherapy (Burlingame et al., 2017). Burlingame and colleagues (2016) conducted an in-depth analysis of outcome differences by format and found equivalence (Hedges’s $g = -0.01$) in 46 studies comparing identical treatments, patients, and doses on primary outcome measures. Thus, treatment equivalence between individual and group formats has a solid empirical foundation.

The extant research on ROM feedback largely involves research on individual therapy. Krägeloh, Czuba, Billington, Kersten, and Siegert (2015) reviewed 27 ROM feedback trials, finding the literature did not contain any ROM studies involving group treatment. Burlingame and colleagues (2018) found that all research on ROM feedback for group treatment has been published since the Krägeloh review and involved the PCOMS system. Burlingame and colleagues summarized each study conducted on PCOMS feedback, showing mixed findings related to dose-response effects of session frequency and methodological shortcomings in trial designs by way of uncontrolled intragroup dependency and other limitations. Clearly, the literature on using ROM feedback in group psychotherapy is relatively immature. As of now, there are two ROM systems that have been evaluated in group treatment (the PCOMS and the OQ-A system), however only the OQ-A system has a randomized controlled trial (RCT) testing the effect of group feedback. Thus, this review will focus on group feedback in the OQ-A system.

OQ-A ROM in group psychotherapy. The routine feedback tool for group treatment in the OQ-A system is similar to the OQ-45’s session-by-session measurement, but combines this
aspect with moderators of group treatment success to predict treatment failures from specific
depths, similar to the ASC. The Group Questionnaire (GQ; Krogel et al., 2013) is a 30-item
measure created through empirical analysis of the best predictors of group treatment success (i.e.,
cohesion, group climate, therapist empathy, and the working alliance), just like the measurement
of treatment moderators in individual treatment (Burlingame et al., 2017). The session by session
use of the GQ came from two findings: (1) Chapman and colleagues’ (2012) results showing
group leaders could not accurately predict a single case of treatment deterioration, and (2)
research found that clinicians regard continual feedback during treatment preferable (Burlingame
et al., 2018; Strauss, Burlingame, & Bormann, 2008).

The GQ has a three-factor structure including positive bond (PB), positive work (PW),
and negative relationship (NR), providing each subscale score for each group member to the
group leader after every session (Burlingame et al., 2017). The PB scale measures “the positive
emotional connection and attachment that a group member feels toward other members, the
leader, and group-as-a-whole”; the PW scale measures “collaborative engagement of the leader,
member, and group in therapeutic work toward consensus treatment [on] treatment tasks and
goals”; the NR scale measures “unpleasant aspects of the group relationship…[including]
alliance ruptures with the leader, an absence of member empathy, and unproductive and/or
unresolved conflict in the group” (Burlingame et al., 2017, p. 12).

Each GQ subscale uses change and status (termed “absolute” in the example) alerts,
similar to the OQ-45’s alerts. However, the GQ bases its change alerts on reliable changes from
the previous session’s subscale score (versus the OQ-45’s reliable change from the start of
treatment) to show reliable change on subdomains; its status alerts are based on being at or above
the 95th percentile (for positive status indications) and at or below the bottom 10th percentile of
the relevant normative sample (for negative status indications; Burlingame et al., 2017. Figure 3 shows a sample GQ report for a group, depicting various change and status alerts. For instance, “Heidi Ranger” shows a reliable positive change on PB and is in the top 5% on that subscale; “Neal Langley” is in the bottom 10% on PW and has shown reliable deterioration on that scale since his last GQ measurement.
Study of the GQ’s sensitivity to change is promising, showing linear change over time on GQ subscales at different stages of treatment and the GQ’s ability to predict improvement on measures of interpersonal problems (Lo Coco, Gullo, Di Fratello, Giordano, & Kivlighan, 2016; Lo Coco, Gullo, Oieni, et al., 2016). Research on feedback measures in group treatment is “an embryonic field,” and is very much still being tested (Burlingame et al., 2017, p. 14). To wit, more studies were needed after Lo Coco and colleagues’ research to elucidate if adding the GQ to OQ-45 feedback increased the group leader’s ability to reverse relationship failure.

**Burlingame and Colleagues**

The need for ROM feedback literature in group psychotherapy, with appropriate methodology, spurred Burlingame and colleagues’ (2018) investigation of such feedback when they set out to replicate Shimokawa and colleagues’ (2010) separation of ROM feedback effects in group psychotherapy.

Burlingame and colleagues conducted an RCT of psychotherapy groups in university counseling centers to compare the effects of giving group leaders OQ-45 feedback versus OQ-45 feedback coupled with GQ feedback. Each group leader ran pairs of groups, with one group receiving only OQ-45 feedback and the second group receiving OQ-45 and GQ-feedback. Groups were randomly assigned to one of these arms with the groups free to vary in accordance with the naturalistic settings of the study and purposes of the group. Each group member gave responses on the OQ-45 and GQ on a session by session basis. The study created “person-session units” (p. 122), OQ-45 and GQ subscales for each group member at each session, as the units of analysis. The person-session units also represented if the group member indicated a PB, PW, or NR change/status alert for that group session.
Analyses of these person-session units allowed Burlingame and colleagues to observe trends in GQ alerts, using the presence of specific GQ alerts to analyze subsequent change on that subscale. Burlingame and colleagues used hierarchical linear modeling (HLM [also called multilevel (or mixed) models]; Heck & Thomas, 2009) techniques to compare the effect of feedback after change and status alerts. Negative change (deterioration) and status (failure) alerts on all subscales were examined in the trial. The study used HLMs to analyze the slopes of change on each GQ subscale two sessions immediately after the session in which the GQ alert occurred. They predicted that the GQ feedback condition would have a change in slope (negative to positive) that exceeded the slope of the no-feedback condition. Their results showed significant slopes of change in the predicted direction between conditions on two GQ subscales (PB and PW), but not on the NR subscale.

They noted several caveats, with one of the most significant being the use of any combination of alerts to compare slopes of change on the GQ subscales. Stated differently, the lack of feedback’s effect on the NR subscale could be due to specific combinations of alerts. For instance, a single NR alert might be more responsive to leader interventions than when it co-occurs with alerts on both PB and PW. This idea relates to White and colleagues’ (2015) investigation of alerting typologies on the ASC. Their findings suggest that testing different alerting patterns on the GQ might be a fruitful inquiry explaining the absence of feedback’s effect on the NR subscale.

The Current Study

The current study is a secondary analysis of Burlingame and colleagues’ (2018) data, using a more fine-grained analysis of GQ alerts. The original person-session units were used to
test the effect of feedback on single and multiple GQ subscale alerts. The current study also used the same HLMs.

**Aims and Hypotheses**

**Aim 1**: To assess the effect of feedback which might vary in the presence of multiple GQ alerts in a single session. This will be assessed by testing the slopes of change on each GQ subscale two sessions after a specified amount of GQ alerts occurred in a person-session unit.

**Hypothesis 1**: As the number of simultaneous GQ change and status alerts increase, there will be a parallel weakening effect for feedback on each GQ subscale.

**Methods**

**Participants**

A total of 432 undergraduates consented to participate in the study. Of those participants, 374 satisfied inclusion and exclusion criteria over the course of therapy. More detailed characteristics are presented below.

**Demographics.** Of the total sample \((N = 432)\), 55.8% of participants were female and 44.2% male. Their mean age was 23.5. Caucasian was reported by 88.9% as their race, 4.2% reported being multiracial, 2.9% reported Hispanic, 1.8% Asian, and 9% not reporting their race. Membership in the Church of Jesus Christ of Latter-day Saints was the most frequently reported religion (80.2%). The most common class standing was junior (38.6%) with a nearly equal proportion of seniors and sophomores (19% and 18%, respectively), followed by freshman and graduate student (11.1% and 12.6%, respectively). Most participants reported having previous counseling (65.8%), while few reported being on academic probation (3%) or having disability status (6.2%).
Presenting complaints and group attendance. Participants’ presenting complaints were most frequently of depression (19%), anxiety (16.9%), and relationship concerns (15.3%), with adjustment challenges (8.4%), self-esteem (7.1%), disordered eating (6.9%), sexual concerns (6.4%), social skills (5.4%), and impulse control (3.3%) less common. Most group members (87.1%) did not report actively using any medication, with small proportions reporting use of antidepressant medications (5.6%), unspecified psychoactive medications (5.7%), and antipsychotic medications (1.5%). Most participants (91.4%) attended 13 or fewer sessions, with the mean number of sessions in the feedback and no-feedback conditions being very similar (9.9 and 9.4 sessions, respectively).

Inclusion criteria. Clients had to agree to two criteria for inclusion in the current study: (1) willingness to complete the GQ weekly, and (2) willingness to have group psychotherapy as their primary mode of psychosocial treatment, while receiving no more than one session of individual psychotherapy per every three sessions of group psychotherapy.

Exclusion criterion. Clients were excluded from the current study if they did not possess an email address. Missing information on outcome measures was obtained through emailing links to outcome measures.

Psychotherapy Groups

Participants engaged in a total of 58 psychotherapy groups. Groups had an average of 7.7 members per group with a range of between five and 12 members. Nearly two thirds (67%) of groups were interpersonal process groups, with the remaining third oriented to specific clinical concerns. These problem-oriented groups focused on sexual concerns (10%), eating disorders (7%), trauma (6%), and an equal number (3% each) focusing on anxiety, personality disorders, and autism spectrum disorder. Psychotherapy groups at each site ranged in the amount of time
each group ran, with the majority (85%) running for one semester. Eight groups ran for two semesters, and one group ran for four semesters. Groups were each co-led by one licensed psychologist and one trainee or intern.

**Measures**

**The Group Questionnaire.** The Group Questionnaire (GQ; Krogel et al., 2013) is a 30-item self-report measure assessing group psychotherapy clients’ perceptions of the therapeutic relationship. The GQ uses a 7-point Likert scale from 1 (“not true at all”) to 7 (“very true”) and uses three subscales to assess quality of the therapeutic relationship: 13 items on positive bonding (PB), eight items on positive work (PW), and nine items on negative relationship (NR). There is no total score on the GQ; each subscale yields a score. Scoring ranges on the GQ vary by subscale; PB scores range from 13 to 91, PW ranges from 8 to 56, and NR ranges from 9 to 63. The GQ has been shown to have good reliability, with reliability coefficients of .79 to .92 for PB, .85 to .91 for PW, and .86 to .87 for NR (Chapman et al., 2012; Krogel et al., 2013; Thayer & Burlingame, 2014).

**Procedures**

Archival data from Burlingame and colleagues (2018) were used. After clients requested services, they completed an OQ-45 and demographic information at their counseling center’s online intake program. Following this, clients were referred to groups according to normal scheduling procedures. During each group’s first session, the group leader and a research assistant described the study per a standardized script, determined client eligibility for the study, and obtained informed consent. Participants were incentivized with cash payments of $10 upon consenting to the study and $5 for every set of OQ-45 and GQ they completed. New members were allowed to join the group and study for up to four weeks following the study’s start.
Data collection occurred throughout the course of group treatment following the first meeting of each group. Members were instructed to complete the GQ immediately following the group using a tablet at their counseling center or by completing it online within a week of the group session. Reminders were sent by email to members who did not complete a GQ within three days of a group session.

Group leaders were encouraged to incorporate GQ feedback into practice during the group. The leaders asked for group members’ preferences on how to use GQ feedback, with group decisions ranging from no formal integration to the group being informed on the GQ and OQ feedback the leader received during every session.

**Analysis**

**Person-session units.** The current study examined the person-session units produced by Burlingame and colleagues (2018). There were 1,249 person-session units in the feedback and 1,165 in the no-feedback condition. 720 person-session units showed some kind of alert over the course of treatment.

**Data structure.** This study examined longitudinal data. GQ data were continually obtained, leading to a long-form dataset with multiple person-session units for each participant. In the feedback condition, group leaders were provided with change and status alerts whenever possible (no alerts are possible if only one session of data are available). To analyze differences between conditions, GQ alerts were also computed for the no-feedback condition.

**Group and member selection.** Some groups in the study met for more than one semester during the course of the study. In those instances, only the first semester’s worth of data were used for analysis. Additionally, some students attended more than one group in the study. Only
their first semester’s worth of data were included in analyses. Participants from one ‘couples’
group were excluded from analyses.

Analytic models. HLMs were used to account for intragroup dependency and repeated
measurement by including random effects for groups and/or participants in statistical models,
when appropriate. Condition interaction effects are represented with $F$-tests of significance using
the Satterthwaite (1946) method to approximate degrees of freedom. Slope and intercept
estimates are also given for each condition on every subscale representing change from the
previous two sessions of GQ data.

Results

Configuration of Alerts

Alerts patterns on the GQ are represented by single, double, and triple-or-more alerts
occurring simultaneously at any given group psychotherapy session. Overall, 322 person-session
units showed a single alert, 186 a double alert, and 212 triple or more alerts. The combinations of
all 508 single and double alerts are depicted in Table 2.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Single and Double Change and Status Group Questionnaire (GQ) Alert Combinations. n = 508</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PB</td>
</tr>
<tr>
<td>PB</td>
<td>Change</td>
</tr>
<tr>
<td>PW</td>
<td>Change</td>
</tr>
<tr>
<td>NR</td>
<td>Change</td>
</tr>
<tr>
<td></td>
<td>Status</td>
</tr>
</tbody>
</table>

Note. Cells show number of person-session units (percent of units) with a single or double alert. PB = Positive Bond GQ subscale; PW = Positive Work GQ subscale; NR = Negative Relationship GQ subscale.
Alert Patterns

Nearly half \((n = 322, 45\%)\) of all alerts occurred on a single GQ subscale. Most \((n = 169, 52\%)\) were on the PW subscale, followed by the PB subscale \((n = 103)\), with the smallest number occurring on the NR subscale \((n = 50)\).

About a fourth \((n = 186; 26\%)\) of all alerts were double alerts, with PB subscale accounting for most \((n = 112; 60\%)\) combinations, followed by PW subscale \((n = 108; 58\%)\) combinations, and NR \((n = 64; 34\%)\) combinations. Some alert combinations are double-counted by being in both categories.

Approximately 29\% \((n = 212)\) of all alerts combinations were three or more, with almost one third \((n = 64)\) of triple and greater combinations involving all change alerts. All alerts occurring simultaneously account for over 13\% \((n = 28)\) of triple and greater GQ alert combinations. Table 3 shows the number and proportion of triple and greater alerts in a given session.

Table 3

<table>
<thead>
<tr>
<th>Alert Type</th>
<th>Number of person-session units (% units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB+NR Change Alert; PB+NR Status Alert</td>
<td>10 (4.7)</td>
</tr>
<tr>
<td>PB+PW Change Alert; PW Status Alert</td>
<td>10 (4.7)</td>
</tr>
<tr>
<td>PB+PW Change Alert; PB+PW Status Alert</td>
<td>14 (6.6)</td>
</tr>
<tr>
<td>All Change Alerts</td>
<td>26 (12.3)</td>
</tr>
<tr>
<td>All Change Alerts; PB Status Alert</td>
<td>10 (4.7)</td>
</tr>
<tr>
<td>All Alerts (Change+Status)</td>
<td>28 (13.2)</td>
</tr>
</tbody>
</table>

*Note.* Cells shown number of person-session units (percent of units) with three or more alerts. Configurations of alerts with ≥10 person-session units are represented. PB = Positive Bond GQ subscale; PW = Positive Work GQ subscale; NR = Negative Relationship GQ subscale.
Alert Number Effects

Condition by session interactions terms for single alerts on PB ($F[1, 508] = 0.15, p = .696$), PW ($F[1, 491] = 0.71, p = .401$), and NR ($F[1, 510] = 1.68, p = .196$) indicated that the slopes of change on GQ subscales did not vary significantly by condition. Estimated slopes of change on individual subscales by condition, shown in Table 4, illustrate that the feedback condition showed statistically significant improvement over time (relative to a slope of 0) on all GQ subscales after a single alert while the no-feedback condition showed an improvement over time on the PB subscale alone.

Table 4

Results of Multilevel Models Testing Group Questionnaire (GQ) Subscale Change Over Time for Therapeutic Encounters with One GQ Alert

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept Feedback</td>
<td>68.92</td>
<td>2.09</td>
<td>211</td>
<td>33.00</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Control</td>
<td>69.48</td>
<td>2.19</td>
<td>229</td>
<td>31.73</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope</td>
<td>Feedback</td>
<td>0.74</td>
<td>0.26</td>
<td>507</td>
<td>2.80</td>
</tr>
<tr>
<td>Control</td>
<td>0.59</td>
<td>0.26</td>
<td>509</td>
<td>2.30</td>
<td>.022</td>
</tr>
<tr>
<td>PW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept Feedback</td>
<td>35.98</td>
<td>1.72</td>
<td>225</td>
<td>20.95</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Control</td>
<td>36.91</td>
<td>1.81</td>
<td>239</td>
<td>20.41</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope</td>
<td>Feedback</td>
<td>0.44</td>
<td>0.22</td>
<td>490</td>
<td>2.02</td>
</tr>
<tr>
<td>Control</td>
<td>0.18</td>
<td>0.21</td>
<td>491</td>
<td>0.86</td>
<td>.388</td>
</tr>
<tr>
<td>NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept Feedback</td>
<td>24.12</td>
<td>1.58</td>
<td>191</td>
<td>15.31</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Control</td>
<td>20.62</td>
<td>1.64</td>
<td>208</td>
<td>12.54</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope</td>
<td>Feedback</td>
<td>-0.57</td>
<td>0.19</td>
<td>508</td>
<td>-2.93</td>
</tr>
<tr>
<td>Control</td>
<td>-0.22</td>
<td>0.19</td>
<td>511</td>
<td>-1.14</td>
<td>.256</td>
</tr>
</tbody>
</table>

Note. Coefficients, standard errors, and t-values are rounded to two decimals. Degrees of freedom are rounded to the nearest integer. PB = Positive Bond GQ subscale; PW = Positive Work GQ subscale; NR = Negative Relationship GQ subscale.
Condition by session interactions terms for double alerts on PB ($F[1, 285] = 1.14, p = .286$), PW ($F[1, 279] = 0.58, p = .448$), and NR ($F[1, 289] = 2.40, p = .122$) indicated that the slopes of change on GQ subscales did not vary significantly by condition. Estimated slopes of change on individual subscales by condition, shown in Table 5, illustrate that the feedback and control conditions’ slopes of change did not vary from a slope of 0 in the two sessions after the double alert occurred.

Table 5  
Results of Multilevel Models Testing Group Questionnaire (GQ) Subscale Change Over Time for Therapeutic Encounters with Two GQ Alerts

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>Feedback</td>
<td>67.47</td>
<td>2.73</td>
<td>203</td>
<td>24.71</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>69.89</td>
<td>2.76</td>
<td>205</td>
<td>25.35</td>
</tr>
<tr>
<td>Slope</td>
<td>Feedback</td>
<td>0.51</td>
<td>0.36</td>
<td>281</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-0.03</td>
<td>0.35</td>
<td>283</td>
<td>-0.10</td>
</tr>
<tr>
<td>PW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>Feedback</td>
<td>35.65</td>
<td>2.23</td>
<td>286</td>
<td>15.96</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>32.73</td>
<td>2.28</td>
<td>272</td>
<td>14.35</td>
</tr>
<tr>
<td>Slope</td>
<td>Feedback</td>
<td>0.32</td>
<td>0.29</td>
<td>267</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.01</td>
<td>0.29</td>
<td>287</td>
<td>0.05</td>
</tr>
<tr>
<td>NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>Feedback</td>
<td>23.69</td>
<td>1.92</td>
<td>220</td>
<td>12.32</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>20.50</td>
<td>1.94</td>
<td>206</td>
<td>10.58</td>
</tr>
<tr>
<td>Slope</td>
<td>Feedback</td>
<td>-0.37</td>
<td>0.25</td>
<td>282</td>
<td>-1.50</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.16</td>
<td>0.24</td>
<td>291</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Note. Coefficients, standard errors, and t-values are rounded to two decimals. Degrees of freedom are rounded to the nearest integer. PB = Positive Bond GQ subscale; PW = Positive Work GQ subscale; NR = Negative Relationship GQ subscale.
Condition by session interactions terms for triple or more alerts on PB ($F[1, 238] = 1.63, p = .203$), PW ($F[1, 253] = 0.73, p = .394$), and NR ($F[1, 245] = 0.07, p = .791$) indicated that the slopes of change on GQ subscales did not vary significantly by condition. Estimated slopes of change on individual subscales by condition, shown in Table 6, illustrate that the feedback condition showed statistically significant improvement over time on the PB and PW scales after three or more alerts while the no-feedback condition did not improve on any subscale.

Table 6
Results of Multilevel Models Testing Group Questionnaire (GQ) Subscale Change Over Time for Therapeutic Encounters with Three or More GQ Alerts

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>Feedback</td>
<td>57.40</td>
<td>3.76</td>
<td>175</td>
<td>15.25</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>63.29</td>
<td>3.37</td>
<td>216</td>
<td>18.80</td>
</tr>
<tr>
<td>Slope</td>
<td>Feedback</td>
<td>1.38</td>
<td>0.47</td>
<td>218</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.61</td>
<td>0.39</td>
<td>243</td>
<td>1.56</td>
</tr>
<tr>
<td>PW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>Feedback</td>
<td>29.23</td>
<td>2.83</td>
<td>179</td>
<td>10.31</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>32.32</td>
<td>2.53</td>
<td>210</td>
<td>12.80</td>
</tr>
<tr>
<td>Slope</td>
<td>Feedback</td>
<td>0.72</td>
<td>0.35</td>
<td>234</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>0.33</td>
<td>0.29</td>
<td>252</td>
<td>1.16</td>
</tr>
<tr>
<td>NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>Feedback</td>
<td>26.71</td>
<td>2.70</td>
<td>161</td>
<td>9.88</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>25.99</td>
<td>2.50</td>
<td>219</td>
<td>10.41</td>
</tr>
<tr>
<td>Slope</td>
<td>Feedback</td>
<td>-0.37</td>
<td>0.34</td>
<td>212</td>
<td>-1.08</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>-0.25</td>
<td>0.30</td>
<td>270</td>
<td>-0.84</td>
</tr>
</tbody>
</table>

Note. Coefficients, standard errors, and $t$-values are rounded to two decimals. Degrees of freedom are rounded to the nearest integer. PB = Positive Bond GQ subscale; PW = Positive Work GQ subscale; NR = Negative Relationship GQ subscale.
The results of this study compared to Burlingame and colleagues’ (2018) results on slopes of change are shown in Table 7. Interaction $F$ tests of slopes and resulting $p$ values allow for comparisons between studies on each subscale. Results show no trend of increasing of $p$ values as more GQ alerts co-occur in the present study, while Burlingame and colleagues’ (2018) original analyses show more consistent significant condition by session interactions following specified alerts on half of all analyses.

Table 7

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Analysis</th>
<th>Burlingame et al.</th>
<th>Current Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB</td>
<td>Change alert</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Status alert</td>
<td>&lt;.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Single alert</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Double alert</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Three or more alerts</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>PW</td>
<td>Change alert</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Status alert</td>
<td>.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Single alert</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Double alert</td>
<td>.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Three or more alerts</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>NR</td>
<td>Change alert</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Status alert</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Single alert</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Double alert</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Three or more alerts</td>
<td>.79</td>
<td></td>
</tr>
</tbody>
</table>

*Note. p* values are rounded to two decimals. PB = Positive Bond GQ subscale; PW = Positive Work GQ subscale; NR = Negative Relationship GQ subscale.

**Discussion**

The primary objective of this study was a more fine-grained analysis of the effect of single versus multiple GQ alert permutations on feedback. The primary hypothesis was that the effect of feedback on GQ subscales would weaken as more GQ alerts occurred simultaneously. No specified number of alerts in any given session interacted with the treatment condition to
distinguish between slopes of change following those alert(s). This study demonstrated that there is not empirical support for this hypothesis, with \( p \) values not showing the expected linear increase on each subscale as more GQ alerts occur. It could be that there is an effect of specific alerts being combined simultaneously (e.g., NR change and status alerts in the same session) that are potent and differentiate between each condition’s trajectory following those alerts, as Burlingame and colleagues suggested. Such a combination might prove more difficult for the group leader to reverse as the group member indicates deterioration and failure of the group relationship at the same time, leaving the group leader with little time to effect a change.

There are limitations of the current study, including the reliability of GQ use. The possibility of GQ non-use by group leaders affecting results shown is possible; however, Burlingame and colleagues modelled GQ use as a moderator of outcome in their original analyses and found it to have a nonsignificant effect on their results. Nevertheless, it could be that GQ use interacts with number of alerts to produce differential results following those alerts.

Another limitation of the current study is the lack of discrete alerts. The current study tested GQ slopes in the sessions following a specified numbers of alerts; however, specifying alerts in a given session does not also mean that a different combination of alerts did not happen in the following sessions where GQ change was tracked (i.e., two alerts may follow a single alert and be included in the single alert analysis). This potential for surplus alerts in subsequent sessions may introduce uncertainty in the results. To remedy this, such variables need to be explained.

Including only the two sessions after any number of GQ alerts might also be too conservative to observe any meaningful changes in the group relationship. Future research could expand on the current design to include longer ‘lags’ in measurement between GQ alert
occurrences and when GQ subscales are subsequently measured. However, this may exacerbate the limitations noted above regarding additional alerts occurring in multiple sessions.

Despite these limitations, the current study expands on the effects found by Burlingame and colleagues and contributes important points into the body of knowledge concerning ROM systems. This study determined that no difference is shown between group psychotherapists who do and do not receive GQ feedback when their patients evidence different amounts of problems with the group treatment relationship. In contrast, Burlingame and colleagues’ original results showed more reliable trends in the group leader’s ability to use feedback in reversing deterioration on two GQ subscales, and relationship failure on one. The practical significance of these findings is that the effect of feedback to group leaders is not more easily discerned by the amount of alerts their clients evidence. It remains unclear if there are certain combinations of GQ alerts happening simultaneously or consequently that result in successful relationship deterioration or failure reversal. It is also unclear if the results shown in the current study generalize to other settings and populations.
References


