Theses and Dissertations

2023-06-01

Group Therapeutic Relationships Codevelopment in Short-Term Therapy: A Conceptual Replication and Extension of Lo Coco et al. (2019)

Tate M. Paxton
Brigham Young University

Group Therapeutic Relationships Codevelopment in Short-Term Therapy:

A Conceptual Replication and Extension

of Lo Coco et al. (2019)

Tate M. Paxton

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
Joseph A. Olsen
Scott A. Baldwin

Department of Psychology

Brigham Young University

ABSTRACT

Group Therapeutic Relationships Codevelopment in Short-Term Therapy:
A Conceptual Replication and Extension
of Lo Coco et al. (2019)

Tate M. Paxton
Department of Clinical Psychology, BYU
Master of Science

Objective: The empirical study of member-group mutual influence in group therapy is an emerging area of study. However, few replications have tried to generalize prior findings related to mutual influence. The aim of this study is to conceptually replicate the longitudinal group actor-partner interdependence models (LGAPIM) employed by Lo Coco et al. (2019) measuring how mutual influence and other factors affect the codevelopment of the group therapeutic relationships.

Method: 343 clients were included in this archival analysis. Group Questionnaire (GQ) main subscales, positive bond (PB), positive work (PW), and negative relationship (NR), were gathered from early, middle, and late timepoints. These timepoints were used to generate intraclass correlation coefficients (ICC) and LGAPIM models for the GQ subscales.

Results: Significant and increasing ICC over time for PB, PW, and NR support codevelopment; however, mutual influence was not significant for PB or PW. There was significant mutual influence between the middle and late stages on the NR subscale. All relationship constructs showed significant similarity, stability, and residual correlation of individual member and the other group member scores.

Conclusions: The current findings support codevelopment of group therapeutic relationships, replicating parts of Lo Coco and colleagues' (2019) findings. However, the role of mutual influence was mixed. For PB and PW, we did not find mutual influence, in contrast with Lo Coco et al.'s (2019) findings. With NR, the pattern of mutual influence was different than Lo Coco et al. (2019), partially replicating their results. Our results on similarity, stability, and residual correlations were all significant, in contrast to Lo Coco et al. (2019). This underscores the importance of replication and power when examining mutual influence.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF FIGURES

Group Therapeutic Relationships Codevelopment in Short-Term Therapy: A Conceptual

Replication and Extension of Lo Coco et al. (2019)

Evidence for equivalent outcomes in group psychotherapy compared to individual

therapy is now supported by many studies (Burlingame et al., 2016), but there is robust empirical

and theoretical debate to explain its effects. Prominent theoretical assumptions of group

psychotherapy posit that the group itself is a primary mechanism of change (Yalom & Leszcz,

2020). Practitioners and researchers point to many interwoven mechanisms which function at the

group level with Yalom and Leszcz (2020) listing eleven: instillation of hope, universality,

imparting information, altruism, the corrective recapitulation of the primary family group,

development of socializing techniques, imitative behavior, interpersonal learning, group

cohesiveness, catharsis, and existential factors. While some of these factors are based on

valuable theory and practice, others such as the group therapeutic relationship (e.g., group

cohesiveness and alliance) have robust empirical support (Burlingame, McClendon et al., 2018;

Alldredge et al., 2021; Lo Coco et al., 2022).

Group therapeutic relationships include relationships between member and leader,

between members, and between a member and the whole group. The relationship between leader

and member is generally defined as alliance—very similar to the individual therapy construct

(Alldredge et al., 2021). The relationship between members is usually defined as cohesion

(Burlingame, McClendon et al., 2018). Climate generally refers to group-as-a-whole

relationships (Bonsaksen, 2013). Research exploring the effect of group therapeutic relationships

on outcome has expanded in the last two decades. In three recent meta-analyses, alliance and

cohesion were both found to be a more important predictors of outcome than the theoretical

orientation (Burlingame, McClendon et al., 2018; Alldredge et al., 2021; Lo Coco et al., 2022).

Increased alliance and cohesion predicted better outcomes, and the predictive power of cohesion increased as the length of the group increased (Burlingame, McClendon et al., 2018). Given the prominence of group therapeutic relationships in theory, practice, and research, a better understanding of its formation and influence will be key for executing effective groups.

Nonindependence

Necessarily, interactions between members in a group create non-independence. Non-independence happens when "persons who are in the same group are more similar (or dissimilar) to one another than are persons who are members of different groups" (Kenny et al., 2002, p. 126). This member nonindependence—or member/group influence—is a central clinical tenet of group treatment. For instance, Yalom and Leszcz's (2020) mechanisms of change rely on the influences and relationships occurring in the group. More specifically, cohesion is possible because the group is building a non-independent conception of their shared group identity which includes relationships, which, in turn, is associated with improvement. This understanding of cohesion relies heavily on the group as the vehicle of these relationships, as indicated by Forsyth (2021) and others (e.g., Fuhriman & Burlingame, 1990). The therapeutic factor of universality happens because members are sharing their experience, therefore impacting the understanding of other members about their situation and the shared member-group experience. Clinically, many group-level mechanisms of change operate due to group nonindependence and would be functionally inoperable if the members in a group were truly independent of one another.

This group nonindependence introduces a statistical problem. The assumptions of common statistical procedures used to analyze group data such as ANOVA and linear regression include independence of observations (Kenny et al., 2002; Baldwin et al., 2005). Failure to account for nonindependence if using these techniques results in distorted estimates of error

variance, which, in turn, affects standard errors, p-values, confidence intervals, and effect sizes (Kenny et al., 2002). These distortions inflate the Type I error rate, leading researchers to reject the null hypotheses when they are actually true (Baldwin et al., 2005; Burlingame et al., 1994). In practical terms, researchers might find a group treatment to have a significant effect using regression or ANOVA; however, if they would have corrected for nonindependence of groups, the "significance" might vanish. In a large body of group therapy studies, Baldwin and colleagues (2005) found that many results ceased to be statistically significant after accounting for nonindependence.

While statistical nonindependence was largely ignored for many years (Burlingame et al., 1994; Baldwin et al., 2005), more researchers have accounted for this problem in recent decades using a range of techniques. The intraclass correlation (ICC) is an essential tool for measuring the nonindependence of observations in group-nested data (Kenny et al., 2002). The ICC "measures the magnitude of the dependency among observations taken on members of the same group" (Baldwin et al., 2005, p. 924). Using the ICC, an estimate of the variance inflation factor (VIF) can be obtained. This, in turn, allows for correction of parameters based on the number of group members, number of groups, and number of conditions in a given study (Baldwin et al., 2005). Even when intraclass correlations are not statistically significant, this may be due to small sample sizes and small numbers of groups used in many studies, so maintaining the assumption of nonindependence is proper (Baldwin et al., 2005; Kenny et al., 2002). Other techniques to control for non-independence include multilevel approaches that control for its effect by nesting of data by specific group membership. The use of actor-partner interdependence models, and the use of cluster-robust standard errors when estimating parameters in structural equation models also controls for non-independence. Clearly, nonindependence is an essential consideration for

both the clinical interpretation and empirical understanding of research on psychotherapy groups. Therefore, the sources of nonindependence are beneficial to empirically identify and understand.

Sources of Nonindependence

Mutual Influence

Mutual influence is one of the most prominent sources of nonindependence in group therapy. Kivlighan and Narvaez (2021) define mutual influence as, "the dynamic interplay between members of therapy groups, often referred to as group-to-individual or individual-to-group influence, hereafter referred to as mutual influence" (p. 192). Mutual influence, then, captures both individual member influence on other members and the other members' influence on an individual. Kenny and colleagues (2002) point out two parts of mutual influence: direct and indirect. Direct influence occurs when "there are reciprocal paths from the scores of individuals to each other" (p. 127). In other words, one aspect of mutual influence is when a group member influences that same construct in other group members and vice versa. For example, an individual's appropriate disclosure could increase other members' appropriate disclosures. Indirect influence, on the other hand, is when "one aspect of a group member influences another aspect of other group members" (Kenny et al., 2002, p. 127). For example, a member's repeated inappropriate disclosures could decrease other members' attendance. In this study, we focus on the direct influence of group therapy relationship perception.

Codevelopment

Another aspect of nonindependence is codevelopment. In couples research, codevelopment is defined as "development in similar directions" of a construct of interest (Orth et al., 2018, p. 151). Lo Coco and colleagues (2019) extend this principle to groups and indicate that "both common sense and theory suggest that codevelopment will occur in groups" (p. 641).

Convergence describes a pattern of codevelopment wherein group members not only develop in the same direction, but also have greater similarity in their perceptions over time. Congruence is a related term that indicates the level of group-agreement about a given construct at a single point in time (with the ICC being an indicator of congruence). For example, imagine that group members rate their trust in other group members over time. This pattern might show high congruence at each time point, with people trusting each other very little at first and high trust at the end of treatment. The pattern may also show codevelopment, where trust increases for each group member over time. Finally, convergence would be shown if trust not only increased over time, but member agreement about trust grew as well. In other words, convergence is both codevelopment and increasing congruence over time.

Codevelopment describes both a source of nonindependence and can be also conceptualized as a byproduct of other sources of nonindependence. For example, members in a therapy group might learn a mindfulness technique and practice during group sessions. Therefore, they may perform this mindfulness skill more similarly than those outside the group who are independently practicing the mindfulness technique (nonindependence). As they continue mindfulness practice, they could improve, or develop in the same direction (codevelopment). This codevelopment also shows the overlapping role of mutual influence, as practicing a skill in group could entail individual group members influencing each other. There is empirical support for codevelopment of group relationships, drawn from research that use repeated measures. These studies typically show that groups, on average, form stronger relationships over time (e.g., Lorentzen et al., 2018), but do not account for what drives the codevelopment.

The Group Leader and Shared Group Environment

Another major source of nonindependence in therapy groups is the group leader. Group leaders directly influence the group by selecting, preparing, and leading groups (Yalom & Leszcz, 2020). During the course of group, the group leader influences the members through the formation of norms, modeling, self-disclosure, psychoeducation, and teaching skills, among many other things (Yalom & Leszcz, 2020; Borgers & Koenig, 1983). Despite this, the group leader's influence is often neglected in traditional research designs which focus on the relationships, perceptions, and outcomes of group members. In individual therapy, "therapist effects," or the differences between therapists that are relevant to outcomes, are well known and robust – therapists vary in their abilities, client outcomes, and client satisfaction (Okiishi et al., 2003; Project MATCH Research Group, 1998). Because it is more difficult to have individual group leaders run a large number of groups in most studies, therapist effects in group therapy have not been as directly examined. However, researchers have demonstrated that group therapist behaviors and interventions are associated with the formation of group therapeutic relationships (Phipps & Zastowny, 1988), client outcome (Chapman et al., 2010), and client satisfaction (Ogrodniczuk, 2007). The shared group environment is the immediate setting that the group takes place in. This is yet another source of nonindependence, although one that is difficult to tease apart from other parts of group therapy, particularly the group leader, who is an essential component of any shared environment.

Analyzing Mutual Influence: The Group Actor-Partner Interdependence Model (GAPIM)

Sources of nonindependence create the opportunity to study how they impact psychotherapy groups. Mutual influence is of primary concern because it is difficult to know whether the group or individual members are exerting more influence. This, in turn, has

implications for codevelopment of group relationships and the role of the group leader. Largely in the last decade, group therapy researchers have examined mutual influence using the group actor-partner interdependence model (GAPIM; Kivlighan & Narvaez, 2021). The GAPIM is modeled after the actor-partner interdependence model (APIM) which is often used in research on dyads (Gonzalez & Griffin, 2001; Kenny et al., 2002; Cook & Kenny, 2005). Figure 1 shows a descriptive version of the group actor-partner interdependence model (Kenny & Garcia, 2012).

Figure 1

Descriptive GAPIM
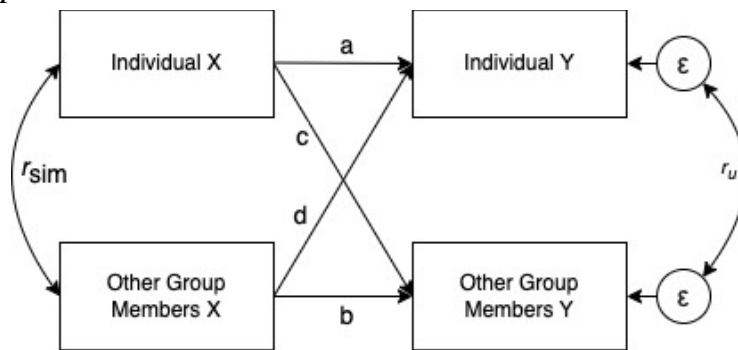
The first step in a typical GAPIM research is to pick two constructs: a composition variable and an outcome variable which are likely to be affected by mutual influence. A common composition and outcome construct in group therapy is the therapeutic relationship. For example, a GAPIM analysis could use the therapeutic relationship as the composition variable (X), and an outcome of interest, like symptomatic improvement (Y). The paths in this type of GAPIM represent (a) direct effect of X on Y at the individual level, (b) the direct effect of X on Y at the group level, (c) the effect of individual X on the other group members Y, and (d) the effect of the other group members X on individual Y. Paths (c) and (d) represent mutual influence, or the effect of individuals on the group and the group on individuals, as defined previously. The correlations are also important: r_{sim} is the similarity between individuals and their groups on X and r_u is the correlation of the unexplained variance of individual and other group member Y.

The following example, drawn from a published article, illustrates the use of this model using the labels from Figure 1. Kivlighan and colleagues (2012) tested whether attachment avoidance (X, characteristic variable) impacts perceptions of group climate, including engagement and conflict (Y, outcome variable). They found that higher other group member attachment avoidance predicted lower other group member engagement (b). Higher other group member attachment avoidance also predicted increased other group member perception of conflict (b) and individual perception of conflict (d). This indicated that the group was a source of mutual influence but did not indicate statistically significant mutual influence from individuals to the group (Kivlighan et al., 2012; see models on p. 524 for details). In other words, when groups were higher in average attachment avoidance, this impacted both the engagement and conflict components of group climate.

Kivlighan and Narvaez (2021) reviewed research on mutual influence in group psychotherapy and cited 15 studies which used GAPIM since 2011. At least two studies have been published since the publication of that review (Lo Coco et al., 2019; Carlucci et al., 2022). Of these, only a few analyzed group therapeutic relationship contribution to outcome. Gullo and colleagues (2014) used a time-lagged GAPIM approach to measure the effect of individual and other group members alliance over time. They found that other group member ratings of alliance predicted individual member alliance later on, although the effect was small due to the time-lagged design. They also found that perception of alliance and outcome had a significant interaction. Lo Coco, Gullo, Di Fratello and colleagues (2016) used the Group Questionnaire to measure group relationships and their impact on interpersonal functioning over the course of graduate student interpersonal process groups. They found that early and late measures of positive bond were associated with improved interpersonal problems for actors and partners.

Additionally, late measures of positive bonding and early measures of positive work were negatively associated with interpersonal improvement.

To date, analyses which have aimed to measure mutual influence in group treatment share a constellation of limitations. First, when using GAPIM to calculate an effect for the other group member effects, the number of groups is functionally the sample size. Therefore, statistical power to detect small effects (if present) has been low, with existing studies using a low number of groups (n = 6-20). Given the cost of conducting group research and the difficulty of gathering the requisite expertise and client base to run many groups, this limitation is hard to avoid. Additionally, many GAPIM researchers have used samples with large average group sizes (10+ members per group), which can decrease power to detect other group member effects. For some variables, there is not much variation between groups (particularly when using group means), which further limits the ability to detect group-level effects (see Baldwin et al., 2012 for a discussion of a similar conundrum). While the current study cannot solve all of these problems, a larger sample size (n = 53 groups), a relatively less specialized sample (undergraduates across 3 universities), and smaller average group size (n = 8 members) address some past limitations.

Lo Coco and Colleagues' (2019) Longitudinal GAPIM

The current study is a conceptual replication and extension of Lo Coco et al., 2019. This study used a longitudinal group actor partner interdependence model (LGAPIM) to investigate what contributes to the codevelopment of group relationship constructs as measured by the Group Questionnaire at 3 timepoints (session 3, session 6, and session 12). The constructs included in the GQ are positive bond, positive work, and negative relationship (Krogel et al., 2013; Burlingame et al., 2017). Lo Coco and colleagues (2019) sample included 210 people (20 groups) being treated with short-term cognitive behavioral group therapy in conjunction with

physical activity for obesity. A general version of the LGAPIM employed by Lo Coco and colleagues (2019) can be found in Figure 2.

Figure 2

Longitudinal Group Actor-Partner Interdependence Model
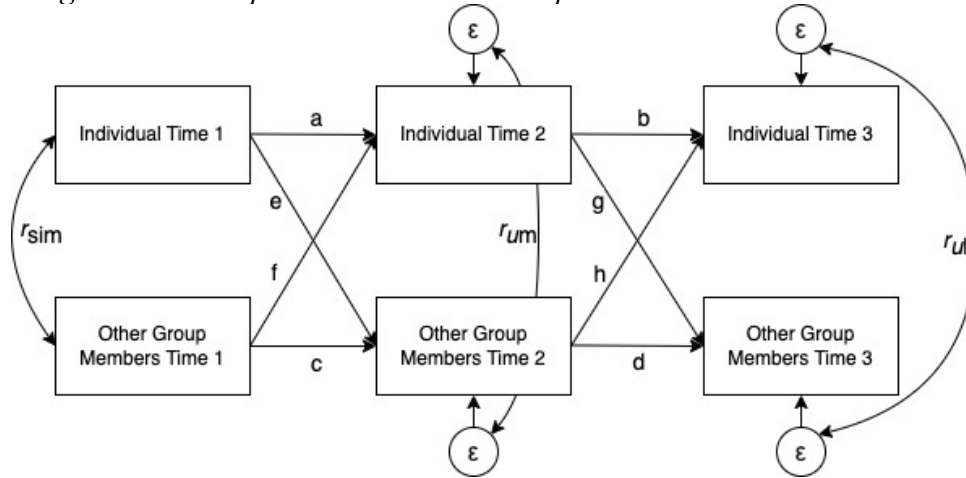
Paths a and b are autocorrelations between individual member scores across time. Paths c and d are autocorrelations between the other group members at the designated timepoints. These autocorrelations represent stability in perceptions of group relationships at the individual and group levels. Paths e, f, g & h are mutual influence effects. In the context of the LGAPIM, they represent individual member effects on the other group members (e & g), and the other group member effects on the individual member (f & h) between the designated timepoints. r_{sim} is the correlation of scores between the actors and partners at time 1 and represents actor similarity to the group. r_{um} and r_{ul} are the correlations between the individual and other group members unexplained errors at times 2 and 3.

Lo Coco and colleagues (2019) found evidence of mutual influence in the development of all three group therapeutic relationships at some point during therapy. For positive bond, path (h) was significant (.14), indicating that other group member positive bond at the midpoint of therapy contributes to individual positive bond at the end of therapy. For positive work, paths (e)

10

and (f) were significant (.20), indicating that higher early individual positive work contributed to higher other group member positive work at the midpoint. Higher other group member positive work was associated with an increase in midpoint positive work (.20). The authors inferred these findings on positive bond and positive work to be compatible with Bennis & Shepard's (1956) two-stage model of group development, wherein groups first deal with power dynamics and then focus on intimate relationships. For negative relationship, both early and middle individual negative relationship was associated with increased middle and late other group member negative relationship, respectively. They interpreted this finding as evidence of the "bad apple" hypothesis, whereby a single group member high in aversive traits can negatively impact the group (Wellen & Neale, 2006, p. 165).

Lo Coco and colleagues (2019) also found significant autocorrelation effects for individual and other group members for positive bond, positive work, and negative relationship. Positive work significant autocorrelations were confined to the effects between Times 1 and 2. These autocorrelations indicated significant stability of member relationship ratings across time. The correlation at Time 1 (r_{sim}) is a measure of similarity between an individual and their group in their perception of group therapeutic relationships (Orth et al., 2018). None of these correlations were statistically significant. The authors, building on couples therapy, interpreted the association of unexplained variance (r_{ul}) to measure the impact of a shared environment on group therapeutic relationships (Lo Coco et al., 2019; Orth et al., 2018). However, none of these correlations were significant. They concluded that while mutual influence plays a role in the codevelopment of group relationships over time, the shared group environment does not.

In sum, the degree to which individuals influence and are influenced by the other group members in their therapeutic relationships is an emerging area of empirical research. The

primary purpose of this paper is to conceptually replicate and extend the findings from Lo Coco et al., 2019 in a new sample on a different continent with a variety of group therapy theoretical orientations using the same measure—GQ. The LGAPIM allows us to examine the factors (mutual influence) that contribute to the codevelopment of group relationships. More broadly, we aim to add to research on nonindependence and probe mutual influence, codevelopment, and other factors that impact relationship development. Our hypotheses were informed by the hypotheses and findings of Lo Coco et al., 2019. Due to the low number of overall findings in this area, some of our hypotheses are in line with null results.

Hypotheses

H1: Codevelopment and convergence of group therapeutic relationships will be shown through significant, increasing intraclass correlations over time for all GQ subscales.

H2: Group therapeutic relationships (GQ subscales) will be highly correlated across time at the individual and other group member levels, indicating stability of group relationship perception over time.

H3: Mutual influence between individual group members and the rest of the group will be present at some point for positive bond, positive work, and negative relationship (at least one of Figure 2 paths c, d, g, & h will be significant for each construct). This hypothesis is in line with Lo Coco et al.'s 2019 findings and other GAPIM studies which have found mutual influence effects but have not found a consistent path where these effects occur.

H4: In line with Lo Coco et al.'s (2019) findings, we hypothesize that r_{sim}, r_{um} and r_{ul} will not be significant for positive bond, positive work, and negative relationship. We do, however, note that Lo Coco and colleagues (2019) hypothesized that these correlations would be significant and significant correlations based upon couples research that also used a longitudinal

actor-partner interdependence framework (Orth et al., 2018), so one could argue a hypothesis predicting significance based on past couples research.

Methods

Procedure

The current study is an archival analysis of Burlingame, Whitcomb and colleagues (2018) study, which is the original source of the data. Refer to that article for additional procedure details, particularly about the feedback conditions. Feedback refers to the change in scores and alerts that the instruments (OQ and GQ) automatically provide based on the client score change or trajectory. Our current analysis is attempting to measure group relationships theoretically common to all groups. Therefore, we will analyze the data from all groups together.

Upon joining a group, the group leader and research assistant described the study and obtained informed consent. Participants agreed to complete the routine OQ and the additional GQ measures weekly. They received $10 upon enrolling in the study, and $5 for completing the OQ and GQ weekly (compensation contingent on completing both). A $20 bonus was awarded those who completed the group. Measures were completed online through emailed URL links.

Sample

Participants

Table 1 provides the demographic information of the sample used in this secondary analysis, with the information from the original analysis included as a comparison. The sample is not racially/ethnically representative of the United States of America, nor of the communities where the universities reside (United States Census Bureau, 2023). Importantly, the sample was drawn from institutions of higher education, which are not representative of the broader community in education level. Furthermore, cultural features of the majority religion (Church of

Jesus Christ of Latter-day Saints), geographic region (intermountain west) and higher education level may play a role in group relationships. The college setting also likely differentiates these groups from groups formed in the community.

Table 1

Demographic Information of Current Sample And Original Sample

Variable	Categories	Current Sample	Original Sample
Sex (% female)		54.9	55.8
Age (mean)		23.9	23.5
Race (%)	White	88.1	88.9
	Hispanic	2.1	2.9
	Asian	2.4	1.8
	Multiracial	4.6	4.2
Class standing (%)	First year	10.8	11
	Second year	18.4	18
	Third year	38.2	39
	Fourth year	20.1	19
	Graduate Student	12	12.6
Religious affiliation (% LDS)		80.4	80.2
Presenting problems (%)	Depression	18.7	19
	Anxiety	14.2	17
	Relationship concerns	15.1	15
	Adjustment	12.1	8
	Self-esteem	4.2	7
	Eating Disorder	4.8	7
	Sexual concerns	10.9	6
	Social skills	5.4	5
	Impulse control	3	3
Disability status (%)		6.1	6.2
Previous counseling (%)		66.4	65.8
Medication status (%)	No medication	73.8	87.1
	Antidepressant	13.5	5.6
	Unspecified psychotropic	9.7	5.7
	Antipsychotic	2.9	1.5

Inclusion criteria

There were three main conditions of participation. Participants were excluded if they had no email address, as measures were administered through emailed URL links. One additional participant-level inclusion criterion in the current study was that participants needed to have attended at least 6 sessions. At the group level, we excluded groups that had 8 or fewer sessions, as this is too short to have comparable early, middle, and late stages of therapy to other groups.

Therapists

Sixteen doctoral-level psychologists volunteered for the study. Recruitment occurred through counseling center-wide meetings where a brief description of the feedback literature and the proposed study were provided. Each therapist had a graduate student co-leader. Theoretical orientation varied, with cognitive behavioral, humanistic, psychodynamic, systems and integrated approaches represented.

Groups

Of the 58 total groups included in the original study, we included groups which had a minimum of 9 possible sessions. This excluded 5 groups, including three interpersonal process groups and two groups for specific problems (personality and eating disorders). We also only included the first semester of a group, even if the group continued on to additional semesters to measure the impact of short-term group therapy. The treatment orientation of the 53 groups in the current sample was mostly interpersonal (68%). About a third of the groups focused on specific concerns, including sexual concerns (11%), eating disorder (6%), trauma (5%), anxiety (4%), personality disorders (4%), and autism (2%). Groups met weekly for two hours. and had between 7 and 12 members, with an average of 7.7 members per group.

Measures

The Group Questionnaire-30 (GQ) was used to measure group therapeutic relationships. The Outcome Questionnaire-45.2 (OQ) was used to measure general psychological distress throughout the study, although this measure is not used in the current analysis. The GQ was administered weekly after each group. See Burlingame, Whitcomb and colleagues (2018) for a list of other measures included in the original study.

The Group Questionnaire-30 (GQ-30)

The Group Questionnaire-30 is a 30 item self-report measure about relationships in a session of group treatment (Krogel et al., 2013; Burlingame et al., 2017). Respondents use a 7-point Likert scale ranging from "not true at all" to "very true" and each question is scored from 0-6 as outlined in the manual (Burlingame et al., 2017). No total score is calculated, as the GQ aims to measure group therapeutic relationships through three subscales: positive bonding relationship, positive working relationship, and negative relationship. Each subscale score is calculated by summing the items which make up the subscale. Positive bond refers to "emotional engagement with group, cohesion with members, and positive empathy toward leader" (Burlingame et al., 2017, p. 9). Positive work is defined as the "collaborative working relationship with members and leaders" (Burlingame et al., 2017, p. 9). Negative relationship refers to "group events that may adversely affect member attachment and impede outcome including negative empathy with leader and conflict with members" (Burlingame et al., 2017, p. 9). The structural subscales measure the types of group relationships within each quality subscale and include member-member, member-group, and member-leader (Burlingame et al., 2017 see Figure 3; Thayer & Burlingame, 2014 see Figures 1 & 2).

Multiple studies have provided evidence for the GQ's convergent and structural validity, and construct validity is promising. The development of the GQ relied on factor analysis of existing group therapeutic relationship measures (Johnson et al., 2005; Krogel et al., 2013). Researchers have confirmed the factor structure for the three quality scales (positive bond, positive work, and negative relationship) in multiple treatment settings in English (Thayer & Burlingame, 2014; Pederson et al., 2021), German (Burlingame et al., 2017 see Table 1), and Italian (Giannone et al., 2020). Convergent validity was established through tests of correlation with existing measures of group therapeutic relationships, including cohesion, alliance, and climate. (Thayer & Burlingame, 2014). Correlation of positive bond with other measures of cohesion yielded moderate to strong results, including with the Working Alliance Inventory (WAI), Bond ($r = .76$) and the Therapeutic Factors Inventory, Cohesion Scale ($r = .81$). Correlation of positive work with theoretically related measures also yielded moderate results, including WAI Task ($r = .79$) and WAI Goal ($r = .71$). Construct validity of the GQ is more difficult to assess, but some emerging studies may indicate its clinical utility and thus its success in measuring the targeted construct. Burlingame, Whitcomb and colleagues (2018) gave clinicians feedback based on weekly administration of the GQ, which involved alerting clinicians to deterioration or low levels of positive relationships for individuals in their groups. They found that clinicians given GQ feedback had fewer relationship failures compared to clinicians given no feedback. This could indicate developing construct validity, as clinicians were able to stabilize relationship deterioration when alerted.

Researchers have provided internal consistency evidence in multiple studies and languages, and test-retest reliability evidence is developing. Internal consistency for the GQ quality scales was demonstrated at its creation with positive bond Cronbach's $\alpha = .92$, positive

work α = .90, and negative relationship α = .80 (Krogel et al., 2013). Since then, these reliabilities have been replicated several times in English and Italian samples with similar results for the three quality scales (Thayer & Burlingame, 2014; Lo Coco, Gullo, Oieni et al., 2016). Recently, more studies have used repeated measures of the GQ to measure group relationships across time, as these relationships are theorized to be dynamic.

Analysis

To measure how group relationships codevelop over time, we used a longitudinal group actor-partner interdependence model (LGAPIM; Kenny & Garcia, 2012; Bonito, 2021; Orth et al., 2018; Lo Coco et al., 2019). This model aims to measure the effect of group composition on outcomes of interest, including the influence of individuals on other group members and vice versa. The model uses the calculation of individual scores and group means to construct a path analysis. Additionally, GAPIM employs a calculation of a group mean excluding the "actor" individual to avoid double counting when measuring individuals against the group mean. In the case of our analysis, the LGAPIM will be conducted with early, middle, and late measures of the major subscales of the GQ (positive bond, positive work, and negative relationship). See Figure 2 and the accompanying explanation in the introduction for a review of the current model. Structural notation is given in Figure 3 (Lleras, 2005).

Because we plan to include a measure of the correlation of errors at the middle timepoint, our models will be just identified ($df = 1$), and fit statistics will not be meaningful. We chose to do this because with low degrees of freedom, fit statistics, particularly RMSEA, but also SRMR, are often biased toward arbitrarily better fit. For example, Shi and colleagues (2022) found that with 2 degrees of freedom, SRMR was biased toward better fit, and CFI was also marginally biased but more resilient. However, they noted that sample size should be at least 200 before

using conventional cutoffs, even for CFI (see p. 201). We are primarily interested in conceptual replication at the structural coefficient level. Therefore, the current study will not replicate fit statistics nor employ the chi-square difference test of constrained vs. unconstrained model as Lo Coco et al. (2019) did.

Figure 3

Structural Notation of Longitudinal Group Actor-Partner Interdependence Model as Path Analysis
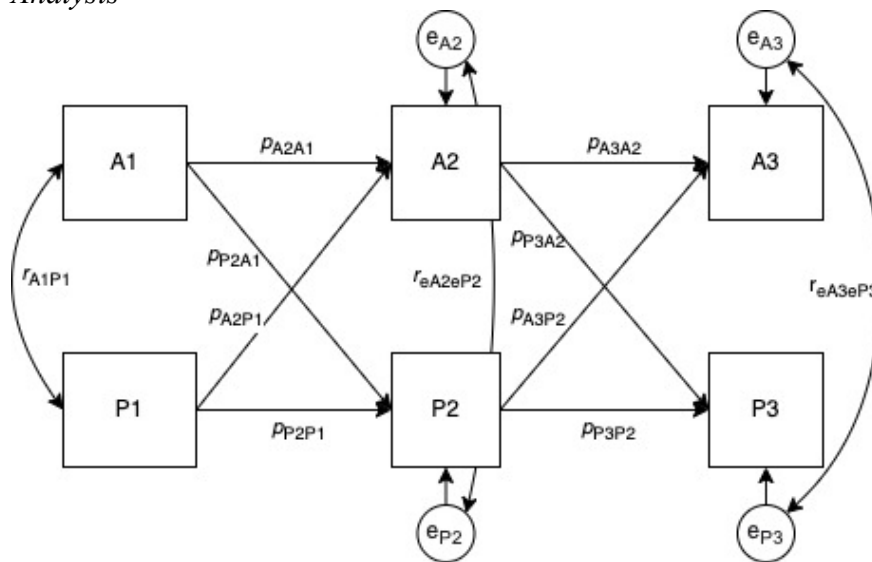
Note: A = Actor (individual member), P = Partner (other group members), p = path, e = error. Arabic numerals indicate timepoints (1, 2, 3).

Structural equations:

Actor Time 2 = $p_{A2A1}A1 + p_{A2P1}P1 + e_{A2}$ (1)

Partner Time 2 = $p_{P2A1}A1 + p_{P2P1}P1 + e_{P2}$ (2)

Actor Time 3 = $p_{A3A2}A2 + p_{A3P2}P2 + e_{A3}$ (3)

Partner Time 3 = $p_{P3A2}A2 + p_{P3P2}P2 + e_{P3}$ (4)

Power

Precise power calculations for GAPIM analyses are largely accomplished through simulation, as shown by Bonito (2021). However, power analyses using simple correlations based on the number of participants and number of groups can approximate power for the individual member and the other group member effects, respectively. We used G*Power (Faul et al., 2007) to calculate power based on two fixed sample sizes, n = 53 for the number of groups and n = 348 for the number of individuals. We used a range of small to medium effect sizes (*r* = .1 to *r* = .5). Lo Coco and colleagues' (2019) mutual influence terms (cross-lags) ranged from -.04 to .20, with the significant effects ranging from .09 to .20. We assume that we are unlikely to find effects larger than those for our mutual influence terms.

Figure 4

Power by Effect Size with 348 Participants
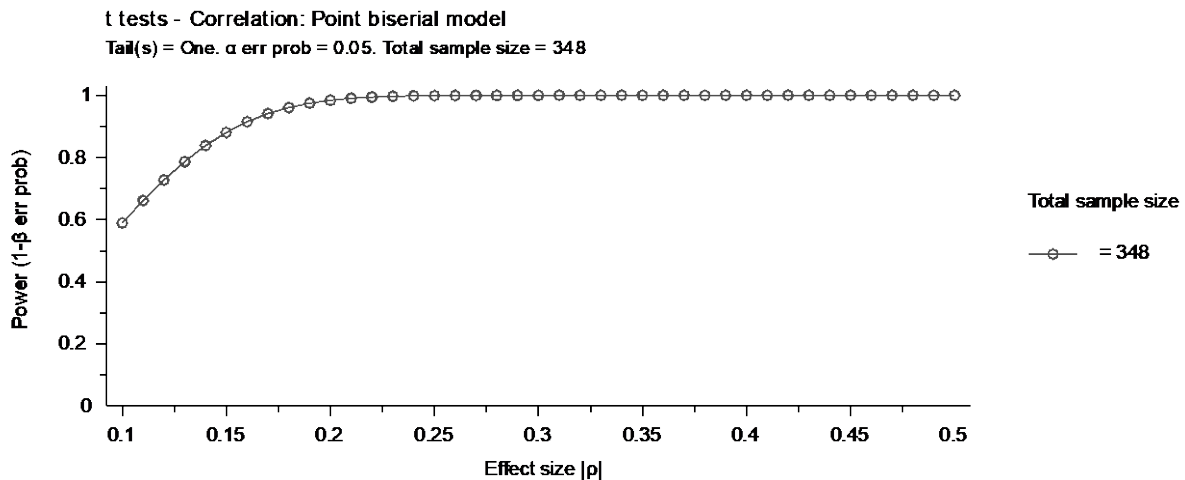
We used G*Power to create visualizations of power by effect size (*r*) at two fixed sample sizes, n = 348 for the total participants (Figure 4) and n = 53 for the number of groups (Figure 5). To provide more information we also created graphs of power by sample size at correlations ranging from *r* = .1 to *r* = .5 for sample sizes of n = 10 to n = 400 and n = 10 to n = 70 (see Appendix A). For the actor effect, a correlation of approximately *r* = .14 would achieve power of

.8. For the partner effect, an effect of approximately *r* = .35 would achieve power of .8. Given the low probability of achieving an effect as large as *r* = .35, this is likely a limitation of our study. However, the number of groups in our study (n = 53) is greater than any other available study using an GAPIM model for group psychotherapy. While these results may seem conservative, Bonito's (2021) results indicate that our approximation is overestimating power. In his simulation, as the number of people in groups increased, partner effect power decreased, and higher intraclass correlation also decreased power.

Figure 5

Power for Group Effects with 58 Groups

t tests - Correlation: Point biserial model
Tail(s) = Two. α err prob = 0.05. Total sample size = 53

Total sample size
—⊙— = 53
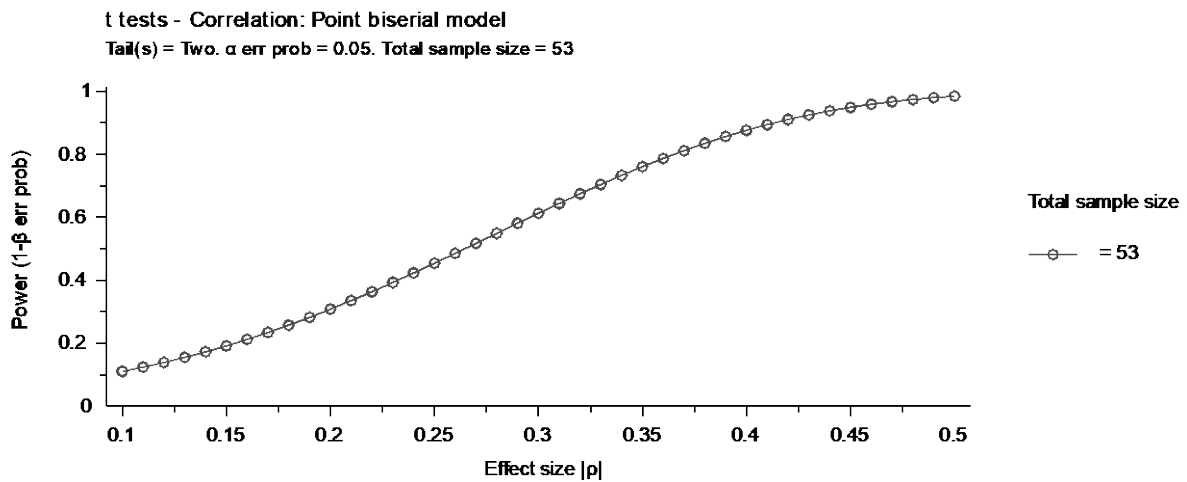
Data Management

For the current study, we used R Markdown (R Core Team, 2022) to structure the data and create .dta files compatible with Stata. Data management and structuring was done with tidyverse packages (Wickham et al., 2019), mainly dplyr (Wickham, François, et al., 2022). Other packages used included knitr for tables (Xie, 2022), naniar and haven for reading in SPSS and exporting Stata data (Tierney et al., 2021; Wickham, Miller, et al. 2022), psych for descriptive statistics (Revelle, 2022), rstatix for frequency tables (Kassambara, 2022), and ICC

for computing intraclass correlations (Wolak, 2022). We used Stata (StataCorp, 2021) to run the LGAPIM models, using the clustered standard error option by group to account for group nonindependence. The data used can possibly be shared by contacting the primary author on the original study (Burlingame, Whitcomb et al., 2018).

Missingness was a substantial problem, as is often the case in group treatments in naturalistic settings. The three types of missingness were session absence, questionnaire noncompletion for attended sessions, and participant dropout. The next paragraphs outline the approach to solving the first two types of missingness. To attempt to solve for the third type of missingness (participant dropout) I excluded participants who attended 5 or fewer sessions. This excluded largely participants who attended early sessions but were absent for the rest of the sessions. This type of missingness is appropriate to exclude because imputing scores for those who are true dropouts does not accurately reflect the membership of the group.

Missing GQ data at any given session due to absence and questionnaire noncompletion also presented a problem, which we attempted to solve through three approaches. First, we gathered data based solely from sessions 3, 6, and the last session of each group. This yielded high levels of missingness and significantly decreased n due to a combination of session absence and noncompletion of measures. Next, we tried gathering data using a stage framework. Under this approach, we targeted specific sessions and gathered data from surrounding sessions if data could not be obtained from the target. For example, we gathered early session data primarily from session 3. If session 3 data was not available, we then gathered from session 4; if this was unavailable, session 2. If none of the "early" sessions provided data, the participant's Time 1 GQ was marked as missing. This approach decreased missingness significantly but introduced a thorny problem. Gathering scores from different sessions impacts scores and makes the other

group member effect not connected to the same session. In particular, more volatile scores like negative relationship can spike due to events in a single group session. While negative relationship is perhaps the most volatile, other measures also have session-to-session fluctuation.

Table 2

Descriptive Statistics of Data Gathered by Session Number

Time	Session	Groups	n	Percent (%)	Attendance (%)	GQ (%)	GQ mean		
							PB	PW	NR
Time 1	2	6	40	11.4	98.4	88.5	74.50	35.83	18.46
	3	36	233	66.2	97.0	90.4	74.49	40.46	19.07
	4	11	79	22.4	98.3	90.1	76.02	41.16	18.65
Time 2	5	12	74	20.5	94.6	86.5	77.29	43.08	16.92
	6	27	186	51.5	94.1	86.6	78.32	42.37	16.88
	7	14	101	28.0	96.0	90.1	76.36	39.77	19.71
Time 3	9	20	132	36.6	91.3	83.5	76.47	42.02	19.27
	10	11	81	22.4	87.7	82.7	79.85	44.64	17.54
	11	12	81	22.4	95.1	92.6	80.59	43.08	15.96
	12	7	48	13.3	87.5	83.3	78.13	41.30	17.88
	13	3	19	5.3	100	83.3	84.40	47.95	13.25

Note: GQ = Group Questionnaire. PB = Positive Bond. PW = Positive Work. NR = Negative Relationship. Attendance and GQ percentages include those who fulfilled inclusion criteria (attended 6 or more group sessions). Some groups had other members present not represented here.

To solve this problem, we established our final data gathering technique by focusing on the group as a whole at each stage (early, middle, late). We gathered data for each *group* from the *session* that had the highest amount of attendance and data provided in a given timeframe (early, middle, late). This also allowed us to account for many groups that had greater or fewer than 12 sessions. Table 2 presents the results of this final data gathering technique. Missingness for Time 1 GQ was 12.4%, Time 2 12.7%, and Time 3 14.9%. This data gathering technique

could have caused the LGAPIM models to diverge from Lo Coco et al.'s (2019), which will be addressed in limitations. Dealing with missing data in GAPIM models is an area without concrete guidelines, so we leaned on theory, trial and error, and transparency in making our decisions. We believe the approach was worth the possible downsides given the high level of data incompletion and scattered attendance pattern in some groups.

For 24 groups, Time 3 was the group's last session. Another ten groups' data were gathered from the second-to-last session. An additional nine groups' data were gathered two sessions before the last session. There were nine groups with Time 3 sessions more than 2 sessions before the group's last. In these cases, missing data made it impossible to obtain substantial data closer to the final session. Missingness at the client level can be found in Appendix A.

<div align="center">

Results

</div>

We first review descriptive statistics. Then, we outline the results relevant to Hypothesis 1 – intraclass correlations for GQ subscales at Times 1, 2, and 3. The LGAPIM results are organized by GQ subscale—positive bond, positive work, and negative relationship—with Hypotheses 2 through 4 addressed within each subscale section. Tables 3, 4, 5, and 6 include Lo Coco and colleagues' (2019) results for reader convenience. The results section ends with a very brief review of hypotheses, findings, and comparison with Lo Coco et al. (2019). A substantive comparison of the current study and Lo Coco and colleagues (2019) is addressed in the discussion section.

Descriptive Statistics

Means and standard deviations of the GQ subscales at each timepoint for the current

sample and Lo Coco and colleagues (2019) are found in Table 3. Positive bond and positive

work both increased over time, and negative relationship decreased. These findings are in line

with previous findings using the GQ (Lo Coco, Gullo, Di Fratello et al., 2016). The current

sample had higher standard deviations in both positive bond and positive work than Lo Coco et

al. (2019). In the current sample, average positive work was also lower at each timepoint. The

general consistency of the means across the two samples is encouraging for conceptual

replication. Distributions of subscale scores are found in Figures A.3, A.4, and A.5 in the

appendix.

Table 3

Means and Standard Deviations by Subscale and Time

GQ Subscale	Study	Time 1 M (SD)	Time 2 M (SD)	Time 3 M (SD)
Positive Bond	Current	74.75 (12.03)	77.55 (12.53)	78.93 (12.52)
	LC (2019)	73.18 (10.20)	75.67 (10.10)	77.40 (11.29)
Positive Work	Current	39.79 (10.10)	41.77 (10.98)	43.14 (10.46)
	LC (2019)	44.14 (6.91)	46.34 (6.95)	48.09 (9.19)
Negative Relationship	Current	18.89 (7.90)	17.70 (8.61)	17.51 (9.43)
	LC (2019)	19.24 (8.08)	18.90 (8.76)	18.82 (9.16)

Hypothesis 1: Codevelopment and Convergence

Mean ICCs for each GQ subscale at each timepoint are found in Table 4, with Lo Coco and

colleagues (2019) results included for comparison. As Lo Coco et al. (2019) noted, ICC values

serve as one measure of group member congruence, or agreement at a given point in time, in

their perception of group therapy relationships. Our results indicated significant intraclass

correlations for all subscales at each timepoint, indicating congruence. In the current sample,

positive bond and negative relationship ICC decreased slightly from Time 1 to Time 2 and

increased from Time 2 to Time 3. Positive bond and positive work ICCs were both higher at

Time 3 than at Time 1. Positive work ICC increased from Time 1 to Time 2 and Time 2 to Time

3. These results support our first hypothesis: greater congruence at the end of treatment

compared to the beginning indicates codevelopment and convergence, as agreement about group

therapeutic relationships ended higher than it started.

Table 4

Mean Intraclass Correlation and 95% Confidence Interval or p-value by Subscale and Time

GQ Subscale	Study	T1	95% CI or *p*-value	T2	95% CI or *p*-value	T3	95% CI or *p*-value
Positive Bond	Current	.22*	[.12, .35]	.16*	[.07, .29]	.43*	[.31, .56]
	LC (2019)	.03	*p* = .141	.09*	*p* = .036	.22*	*p* < .001
Positive Work	Current	.09*	[.01, .20]	.14*	[.05, .26]	.26*	[.15, .39]
	LC (2019)	.02	*p* = .447	.11*	*p* = .041	.25*	*p* < .001
Negative Relationship	Current	.26*	[.16, .40]	.14*	[.05, .26]	.50*	[.38, .63]
	LC (2019)	.06	*p* = .089	.08*	*p* = .049	.19*	*p* < .001

Note: * = *p* < .05. GQ = Group Questionnaire; T1 = Time 1; T2 = Time 2;

T3 = Time 3; LC (2019) = Lo Coco et al., 2019.

LGAPIM Models: Hypotheses 2, 3, & 4 by GQ Subscale

This section contains three LGAPIM models, organized by GQ subscales positive bond,

positive work, and negative relationship. Table 5 shows the results for the GQ subscales together

with Lo Coco et al.'s (2019) for convenience.

Positive Bond

Hypothesis 2: Stability. Figure 6 displays the model for positive bond. We found that

both individual (.58 & .58) and other group members' (.40 & .57) perceptions of positive bond

were fairly stable over the timepoints. This supports our hypothesis, that perception of positive bond would be correlated over time.

Hypothesis 3: Mutual Influence. Our findings do not provide evidence for mutual influence in the development of positive bond over time; thus, our third hypothesis is unsupported.

Hypothesis 4: Similarity and Unexplained Variance Correlations. We found significant similarity between individuals and their group at Time 1 ($r = .25$). In addition, the error for middle and late member and other member positive bond significantly correlated ($r = .22$ and $r = .35$, respectively; see Table 5 for 95% CI). This indicates that group members had similar perceptions of group relationships early and something is causing the error variance to be correlated at times 2 and 3. This does not support our fourth hypothesis for positive bond, which was in line with the null. All positive bond correlations were significant.

Figure 6

LGAPIM: Positive Bond
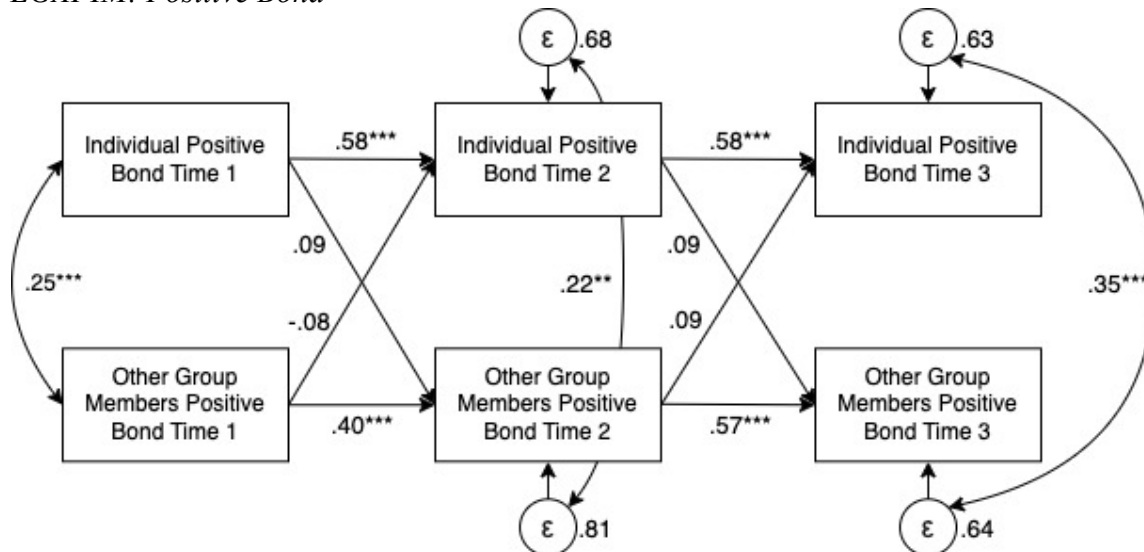
Note: * p < .05. ** p < .01. *** p <.001

Positive Work

Hypothesis 2: Stability. Figure 7 displays the LGAPIM model for positive work. There is significant stability for the individual autocorrelation between Time 1 and 2 (.64) and Time 2 and 3 (.60). There was also significant stability for the other group member autocorrelations between Times 1 and 2 (.42) and Times 2 and 3 (.54). These results support our second hypothesis, positive work was correlated over time for both individuals and groups.

Figure 7

LGAPIM Model: Positive Work
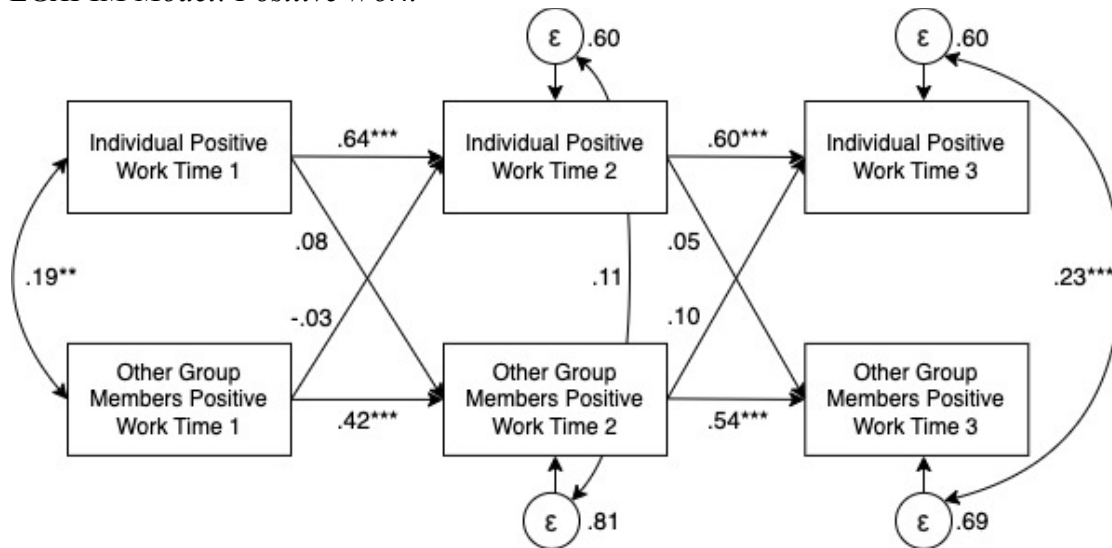
Note: * p < .05. ** p < .01. *** p <.001

Hypothesis 3: Mutual Influence. Our findings do not provide evidence for mutual influence in the development of positive work over time, with no significant effects. Our second hypothesis is unsupported.

Hypothesis 4: Similarity and Unexplained Variance Correlations. As with positive bond, there was significant similarity, meaning individual and other group member ratings of positive work at Time 1 were correlated ($r = .19$). The individual and other group member residuals correlation was not significant at Time 2 ($r = .11$, 95% CI [-.03, .24], $p = .11$). The residual correlation was significant at Time 3 ($r = .23$). Taken together, there was some similarity

in perceptions early-on in the group, and residual correlation in the late stage. This does not

support our fourth hypothesis given that two of the three correlations were significant.

Negative Relationship

Hypothesis 2: Stability. Figure 8 displays the LGAPIM model for negative relationship.

There is significant stability for individual member and other group member ratings of negative

relationship across the timepoints, supporting our second hypothesis.

Figure 8

LGAPIM Model: Negative Relationship
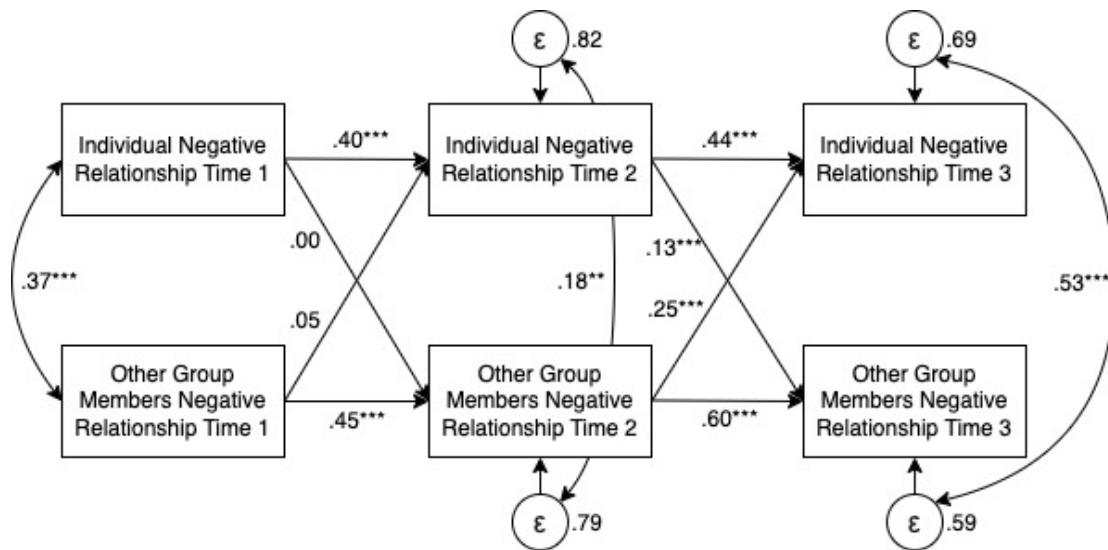
Note: * p < .05. ** p < .01. *** p <.001

Hypothesis 3: Mutual Influence. Individual member and other group member effects

were not significant between the early and middle stages of the group. However, between Times

2 and 3, both individual member (.13) and other group member (.25) effects were significant.

This indicates that higher individual member perception of negative relationship during the

middle of the group increased other group member perception of negative relationship at the end

of group. Similarly, when other group member perception of negative relationship was higher at

middle stage, individual member perception of negative relationship was higher at the late stage.

Taken together, these results indicate that higher perception of conflict in the middle sessions of group was associated with higher perceptions of conflict at the end of group. Therefore, in line with hypothesis three, we found evidence of mutual influence in the development of negative relationships at the mid to late stages of group. This influence occurred at the individual and the other group members levels.

Hypothesis 4: Similarity and Unexplained Variance Correlation. Individual and other group member Time 1 negative relationship perception was significantly correlated ($r = .37$), indicating similarity between individuals and their groups. Residual correlations of individual and other group member negative relationship ratings were significant at Time 2 ($r = .18$) and Time 3 ($r = .53$). These findings do not support our fourth hypothesis, which predicted no significant correlations.

Hypotheses Review and Brief Comparison with Lo Coco et al. (2019)

H1: Codevelopment and convergence of group therapeutic relationships will be shown through significant and increasing intraclass correlations over time for all GQ subscales. This hypothesis was supported for all three GQ subscales, and in the case of negative relationship, surpassed the magnitude of intraclass correlations expected. The similarity (Time 1) correlation was significant, in contrast with Lo Coco et al. (2019) findings for this timepoint. At Times 2 and 3, both the current study and Lo Coco et al. (2019) found significant ICC, and Time 3 ICC were the largest for the three GQ subscales (see Table 4).

H2: Group therapeutic relationships (GQ subscales) will be highly correlated across time at the individual and other group member levels, indicating stability of group relationship perception over time. This hypothesis was supported for all GQ subscales, and the direction and size of correlations was similar to Lo Coco et al. (2019).

H3: Mutual influence between individual group members and the rest of the group will be present at some point for positive bond, positive work, and negative relationship (at least one of Figure 2 paths c, d, g, & h will be significant for each construct). This hypothesis is in line with Lo Coco et al.'s 2019 findings and other GAPIM studies which have found mutual influence effects but have not found a consistent path where these effects occur. In the current study, this hypothesis was unsupported for both positive bond and positive work, which differs from Lo Coco et al. (2019) findings. The hypothesis was supported for negative relationship, which included significant individual and other group member effects from Time 2 to Time 3, which matches Lo Coco et al. (2019) for the Time 2 to Time 3 actor effect, but the significant other group member effects diverged from Lo Coco and colleagues' findings (see Table 6).

H4: In line with Lo Coco et al.'s (2019) findings, we hypothesized that r_{sim}, r_{um} and r_{ul} would not be significant for positive bond, positive work, and negative relationship. This hypothesis was unsupported in the current analysis. Every value of r_{sim}, r_{um} and r_{ul} was significant except positive bond Time 2 r_{um} (see Table 5).

Table 5

Correlation Coefficients (r) of Individual and Other Group Member by Subscale and Time

GQ Subscale	Study	Time 1	95% CI	Time 2 UV	95% CI	Time 3 UV	95% CI
Positive Bond	Current	.25***	[.11, .39]	.22**	[.07, .37]	.35***	[.15, .55]
	LC (2019)	.06				-.04	
Positive Work	Current	.19**	[.05, .33]	.11	[-.03, .37]	.23***	[.10, .37]
	LC (2019)	.02				.00	
Negative Relationship	Current	.37***	[.25, .48]	.18**	[.04, .31]	.53***	[.34, .73]
	LC (2019)	.13				.01	

Note: ** p < .01. *** p < .001. UV = unexplained variance, referring to the correlation between error terms. LC (2019) = Lo Coco et al. (2019). Current study correlations were significant except Positive Work Time 2 UV. No correlations from Lo Coco et al. (2019) were significant.

Table 6

LGAPIM Coefficients with 95% Confidence Intervals by GQ Subscale and Time

Subscale	Path	Study	Time 1 to Time 2		Time 2 to Time 3	
			Coeff.	95% CI	Coeff.	95% CI
Positive Bond	IM AC	Current	.58***	[.45, .72]	.58***	[.47, .68]
		LC (2019)	.50***		.56***	
	OGM AC	Current	.40***	[.19, .62]	.57***	[.39, .76]
		LC (2019)	.63***		.63***	
	IM effect	Current	.09	[-.03, .21]	.09	[-.01, .19]
		LC (2019)	-.05		.09	
	OGM effect	Current	-.08	[-.19, .02]	.09	[-.04, .23]
		LC (2019)	.03		.14*	
Positive Work	IM AC	Current	.64***	[.53, .74]	.60***	[.51, .69]
		LC (2019)	.34*		-.17	
	OGM AC	Current	.42***	[.21, .63]	.54***	[.37, .72]
		LC (2019)	.63***		.14	
	IM effect	Current	.08	[-.01, .16]	.05	[-.05, .14]
		LC (2019)	.20**		-.03	
	OGM effect	Current	-.03	[-.13, .08]	.10	[-.002, .21]
		LC (2019)	.20**		.03	
Negative Relationship	IM AC	Current	.40***	[.29, .51]	.44***	[.33, .56]
		LC (2019)	.64***		.36**	
	OGM AC	Current	.45***	[.29, .62]	.60***	[.46, .74]
		LC (2019)	.77***		.67***	
	IM effect	Current	.00	[-.10, .09]	.13***	[.05, .21]
		LC (2019)	.09*		.14*	
	OGM effect	Current	.05	[-.09, .18]	.25***	[.08, .42]
		LC (2019)	.09		.09	

Note: * p < .05. ** p < .01. *** p < .001. Coeff. = Coefficient; GQ = Group Questionnaire; LC (2019) = Lo Coco et al. (2019); IM = Individual Member; AC = autocorrelation; OGM = Other Group Members

Discussion

The importance of group therapeutic relationships has been outlined in many studies and is a mechanism of change in group therapy (Burlingame, McClendon et al., 2018; Alldredge et al., 2021; Lo Coco et al., 2022). The GQ is a factor analytically derived instrument to measure group therapeutic relationships with multiple replications of its structure and constructs (Krogel et al., 2013; Burlingame et al., 2017). Recent use of statistical techniques which model for group mutual influence (like GAPIM approaches) are promising for further illuminating the role of individual members and the rest of the group in forming relationships and instigating change. This study is a replication of Lo Coco et al. (2019) and used a longitudinal GAPIM to analyze individual and the other group member mutual influence in the formation of positive bond, positive work, and negative relationships. Therefore, this is the second study to use a LGAPIM model to investigate the codevelopment of GQ group therapeutic relationships across the early, middle, and late timepoints of short-term group therapy.

Our findings provided further evidence of codevelopment of group relationships through convergence, mutual influence, and some unclear mechanisms (e.g., the unexplained variance correlation). This discussion first addresses findings that were replicated in the current study. Then, the differences between the two studies are discussed, with clinical implications, followed by a discussion of replication and power in mutual influence research. The discussion then proceeds to strengths and limitations, future directions, and ends with a brief conclusion.

Replicated Findings

Means and Standard Deviations of GQ Subscales for Times 1, 2, and 3

The means and standard deviations at the three timepoints were largely comparable in the current study and Lo Coco et al.'s, as shown in Table 3. The only slight deviations were found in

positive work, where the current study had slightly lower means and higher standard deviations than Lo Coco et al. (2019). This may reflect the larger number of different group types included in our sample, which might have lower and more variable positive work than the single group type – cognitive behavioral therapy – which makes up the sample in Lo Coco et al., 2019.

Codevelopment and Convergence

We found evidence of codevelopment through convergence, indicated by ICCs over time. This broadly replicates the pattern found in Lo Coco et al. (2019), with ICCs increasing over time and Time 3 containing the highest ICCs. As Lo Coco et al. (2019, p. 645) noted, group members appear to have developed "a shared sense of positive bonding, positive working, and negative relationships" by the late stage of group treatment.

Stability

The pattern of stability, indicated by significant correlation of individual member scores over time and significant correlation other group member scores over time, was largely replicated. Both the current study and Lo Coco et al. (2019) found significant correlations for all three GQ subscales across most timepoints. This stability indicates that relationships formed early significantly predict how these relationships will be over the course of the group. This replication may also have psychometric implications, as it signals consistency of the GQ measurement in two very different samples. As will be outlined later, this stability could also signal "trait-like" parts of positive bond, positive work, and negative relationship.

Mutual Influence

Our results indicated codevelopment of group relationships through negative relationship mutual influence in the late group stage. Individual members influenced the rest of the group— high individual member perception of negative relationship at the middle stage was associated

with an increase in the other group members negative relationship at the late stage. This is consistent with Lo Coco et al. (2019) and provides evidence for the "bad apple" effect (Wellen & Neale, 2006, p. 165), where a single member can negatively impact the rest of the group.

We also largely replicated Lo Coco et al. (2019) findings that most mutual influence effects were not significant. Non-significant effects in the current study included all mutual influence effects for positive bond and positive work and early to middle stage mutual influence effects for negative relationship. This non-significance for most of the mutual influence effects is interesting, particularly given the theories that hold the group to be a major mechanism of change. But nonsignificant mutual influence effects are not simply confined to these two GAPIM studies—many others follow a similar pattern (e.g., Carlucci et al., 2022). Perhaps the typical group size used in mutual influence studies is too large. As group size increases, the effect of each member on the group mean is diminished, making mutual influence effects more difficult to detect. Given with the typically small effects found in mutual influence research, this poses a problem. One potential solution is maintaining smaller group size during data collection (e.g., 5-6 members per group instead of 8-10+). However, the tradeoffs of including fewer people in groups may not make clinical and even ethical sense in many settings. It's also important to note that two-thirds of the groups in our sample were process groups that loosely follow the Yalom interpersonal model (Yalom & Leszcz, 2020). These groups emphasize interpersonal relationships through member-to-member processing, so it is surprising that we did not find mutual influence in groups where this is supposed to be occurring.

Differences in Findings

Stability

While stability was largely replicated in both the current study and Lo Coco et al.'s, there were two prominent differences. First, Lo Coco et al. (2019) did not find significant autocorrelations from the middle to late stages for positive work. This is surprising given the similar trajectory of positive work mean scores and ICCs across the three timepoints in both samples (see Tables 3 and 4). Second, results differed in magnitude of negative relationship autocorrelations. The current study's autocorrelation effects appear weaker than Lo Coco and colleagues' findings between Times 1 and 2 for both individual (.40 and .64, respectively) and other group member (.45 and .77, respectively) ratings (see Lo Coco et al., 2019 Figure 4). More specifically, the 95% confidence intervals of the current study's early autocorrelation effects do not include the effects found by Lo Coco and colleagues (see Table 6). This indicates that perceptions of negative relationship were more stable from the early to middle parts of the group in the Lo Coco et al.'s sample than in the current sample. In interpersonal process groups that dominated the current sample, fluctuations and less stability in negative relationship may make sense since interpersonal conflicts arise and are intentionally brought out into the open by the group leader, introducing more opportunity for conflict and confrontation.

Codevelopment and Convergence

Notably, the current results yielded higher Time 3 ICCs than Lo Coco et al. (2019) for positive bond and negative relationship. The early ICCs in the current study were also significant, indicated in Table 5 with 95% confidence intervals. This early significance may be the result of a greater number of groups in our study. However, an increased number of groups is not the only thing that could explain the discrepancy with Lo Coco et al. There may be other

factors (like shared university environment or dominant religious culture) that are driving participants in the current sample toward more congruence in the early stage of group therapy, which, in turn, produce greater congruence in the late stage. Another possibility is that the high and significant ICCs are partially products of our data gathering technique, which gathered group data from different sessions at each stage, in contrast to the same session in Lo Coco. However, ICCs for positive bond and positive work were similar when employing other gathering techniques that we tried (including gathering from the same session for all groups), indicating that our results are unlikely to be solely due to data gathering.

Mutual Influence

For positive work, the current study found no evidence of mutual influence, whereas Lo Coco et al. (2019) found significant other group members influence on the individual from Times 2 to 3. This difference indicates that in Lo Coco's sample, higher middle stage positive bond in the group is associated with an increase in individual positive bond at the end of group. The general lack of mutual influence in the current sample, combined with high stability, could mean that positive bond has trait-like elements and is resilient over time. Indeed, the means and standard deviations seem to indicate that this construct starts high and marginally increases over time. Another possibility is that fluctuations in positive bond are of a short duration that would not be captured by the three-time-point design currently being used. Finally, in the original study (Burlingame, Whitcomb et al., 2018), slightly stronger effects for feedback on the GQ were found for structured groups that included CBT and ACT protocols. The preponderance of process groups in the current sample may have diluted effects from the structured groups that more closely resemble the Lo Coco study.

For positive bond, our results diverge with Lo Coco et al. (2019), who found significant individual member (.20) and other group member (.20) effects between Time 1 and Time 2. This may be due differences in group type between the two studies. Theoretically, manualized group CBT is more goal-oriented and task focused than general process groups, which make up most of the current sample. Therefore, CBT groups might develop a stronger sense of positive work that entails more mutual influence of members early on as they are socialized into the group. Indeed, in the original study (Burlingame, Whitcomb et al., 2018), the structured groups had higher positive bond and lower negative relationship scores compared to the process groups, indicating that structured groups may follow different patterns than unstructured groups.

For negative relationship, Lo Coco et al. (2019) found that individual member perception impacted other group member's perception from Time 1 and Time 2. This essentially extends the "bad apple" effect findings to early in the group as well as in the late stage, in contrast with our findings, which only found such an effect in the middle to late stage. In the current sample, the other group members effect was significant—higher group perception of negative relationship in the middle stage was associated with an increase in individual member perception of negative relationship at the late stage. This may indicate that broad member perceptions of negative relationships (like conflict) in the middle sessions of the group fosters negative relationships that are difficult to resolve, even in the late stage of short-term group therapy. The original study (Burlingame, Whitcomb et al., 2018) also may give insight into these negative relationship findings. Specifically, more OQ-45 alerts indicating symptomatic worsening occurred in later stages of group therapy. The significant correlation between negative relationship scores and symptomatic worsening suggests that these late stage OQ-45 alerts may co-occur with increased negative relationship impact on the group.

The negative relationship results of the original GQ feedback study (Burlingame, Whitcomb et al., 2018), Lo Coco and colleagues (2019) study, and the current study provide an opportunity for clinical reflection. Burlingame, Whitcomb and colleagues (2018) measured the impact of providing group leaders feedback on the GQ subscales. This feedback would indicate when relationship deterioration (increases in negative relationship) occurred and alert the group leader. They indicated that their results were partially "bad news" because "feedback failed to change relationship deterioration on the negative relationship scale" (Burlingame, Whitcomb et al., 2018, p. 128). Both the current study and Lo Coco et al. (2019) indicated that higher individual member negative relationship was associated with higher other group member negative relationship later on in the group. The current study added to this that higher other group member negative relationship at the middle stage was associated with higher individual member negative relationship at the late stage.

Clinically, these results could indicate that a single member rise in negative relationship is an early signal that may impact the group more broadly. Perhaps this presents an opportunity for the group leader to intervene quickly with individuals whose scores indicate a rise in negative relationships. This intervention would not entail prevention of conflict, but perhaps guidance in productive ways to bring elements of negative relationship into the group. The reasoning behind an early intervention is because alerts have not been sufficient to correct deterioration (Burlingame, Whitcomb et al., 2018), and by the time the negative relationship is present in the rest of the group members, this impacted individuals at the late stage. In other words, conflict that was high in the middle stage of group appeared to increase conflict in the late stage, even with some groups receiving GQ feedback. Introducing ways to deal with this negative relationship earlier in group may increase the chance of a satisfactory pattern of resolution for

individuals and the group as a whole. On the other hand, given the lack feedback effect, perhaps negative relationships in group are difficult for the group leader to resolve quickly through their own effort. Under this assumption, once negative relationships are present, they will simply take time to resolve among members of the group.

Similarity and Correlation of Unexplained Variance

The findings that most sharply diverged from Lo Coco et al. (2019) were the significant similarity and unexplained variances correlations for the three GQ subscales. None of these correlations were significant in Lo Coco et al. (2019), whereas all of them (except one) were significant in the current analysis (see Table 5). Interestingly, our results follow Lo Coco et al.'s initial hypotheses. Regarding interpretation, Lo Coco and colleagues (2019) followed couples researchers, who indicated this correlation reflected the influence of the shared environment (Orth et al., 2018). In groups, the interpretation of this term as the influence of a shared environment is more tenuous than in couples due to the limited nature of the immediate shared environment. Furthermore, an essential component of group therapy is the leader. In a GAPIM framework, the leader's impact is largely overlooked, but may be represented in these residual correlations. This points to a broader problem with GAPIM models in group therapy in that they do not account for the actions, perceptions, and characteristics of the leader. There were 16 leaders represented in the current sample, as compared to the single leader and coleader who led the groups in Lo Coco et al. (2019). This difference alone might explain the divergent findings.

While we are unsure about the exact factors impacting the Time 2 and Time 3 residual correlations, the current results were significant. According to Cohen (1992) the effect size of the significant correlations ranged from small ($r = .18$, negative relationship Time 2) to large ($r = .53$, negative relationship Time 3) with the average significant correlation being medium ($r =$

.30). The residual correlation was highest at the late stage of group for all relationship constructs. Our findings may be due to the fact that the clients in therapy groups all shared multiple layers of environment: university, location, and often religion. These similarities are greater than a sample that is recruited from a broader community in Lo Coco et al. (2019) who likely vary in age and demographic characteristics. Also likely is that the residual correlation is tapping influence of the group leader. Discovering how to best interpret this residual correlation in group psychotherapy will require more clarification and replication.

In all, many of our findings replicate the results of Lo Coco et al. (2019) analysis. The differences have implications for different group types and settings and invite future research for clarification. As more GAPIM studies are replicated, we will learn more about how individuals and groups affect each other and *when* these effects might hold across different clinical populations, group treatments and leadership styles.

Strengths and Limitations

The current study has key strengths compared to the broader GAPIM literature. The number of groups used is the highest to date of any published study, increasing statistical power to detect significant group effects. The sample was made up of clients with many different presenting problems, different sites, and the groups varied in theoretical orientations. Therefore, conclusions drawn from the sample may be pertinent to university counseling centers in the USA that included process groups and structured evidence-based protocols. We note, however, that the most groups were interpersonal process oriented. We also measured the residual correlation at both the middle and late stages of group to further examine codevelopment.

The current study also has limitations. While the number of groups was larger than previous studies, power to detect group effects was still quite low (see Figure 4). Missing data

was also a problem which necessitated a particular data gathering strategy for early, middle, and late stages of group. This strategy could have introduced substantial differences compared to studies which measure group relationships at fixed sessions. Furthermore, missing data was still substantial after using this strategy, and protocols for dealing with missingness in a GAPIM framework are still developing. For example, how might researchers deal with participants who attended group but did not complete measures? How might we deal with one-off absences? And true attrition? Or an extended incompletion of questionnaires? Maximum-likelihood approaches, as used in the current analysis, may reasonably be used for some types of missingness, but not all. Another issue is the number of tests we conducted – 39 across the three models. Caution is necessary in interpreting findings with p-values higher than $p = .01$ due to finding "significant" results by chance alone. The fact that the sample had clients with a range of presenting concerns attending groups with a range of theoretical orientations reflects a naturalistic setting, but this can also be seen as weakness compared to Lo Coco et al. (2019) who focused on a more homogenous sample and treatment strategy. Specific patterns of relationships may be different in different types of groups with varying levels of structure. The current results should not be overgeneralized.

Replication

Our findings have broader implications for GAPIM research replication. We posit that replication of mutual influence studies (largely GAPIM) is essential (Open Science Collaboration, 2015). Replication will allow the field to observe under what circumstances and for which groups and samples certain effects hold. Replication will also eventually allow for meta-analytic estimates. Research designs geared toward answering questions about mutual

influence using GAPIM are also necessary. For example, a protocol could examine differences among theoretical orientations in mutual influence, or differences among group leaders.

Power

One key problem in the current mutual influence literature is low power for group effects. As reviewed in the analysis section, partner effects are powered based on the number of groups, the number of group members, the intraclass correlation, and the effect size (Bonito, 2021). Research designs which have many groups and purposefully have fewer group members have a better chance at achieving sufficient power. Any such study would likely need to be multi-site, perhaps multi-year trial. These are massive endeavors with great cost in both clinician time, and research labor and money. Another option is the development of a shared protocol for a mutual influence study, to be carried out at multiple sites. This data would then be combined into a mega-analysis after the sites complete their trials. Accounting for the power problem will take careful planning and collaboration on the part of researchers.

Future Directions

Our findings regarding negative relationship are very clinically relevant and merit future research (along with the data source findings, Burlingame, Whitcomb et al., 2018, and Lo Coco et al., 2019). Future research could examine the impact of more passive feedback (like is provided with the GQ) versus a more direct and personal leader intervention for individual members with increases in negative relationship. Such a comparison could gauge how these different methods of "treating" patterns of negative relationship could be made more productive for groups in short-term therapy. Another direction for this research is to look more closely at the factors driving negative relationship. The GQ measures group relationships at member-member, member-leader, and member-group levels. As noted by Burlingame, Whitcomb and colleagues

(2018), the negative relationship may be more connected to certain relationships than others. For example, perhaps the group member is angry at the leader, but not the other group members. This research may bridge current efforts to investigate alliance ruptures in group therapy. As Tasca & Marmarosh (2023) indicate, positive bond and positive work are two core components of many definitions of alliance. Therefore, ruptures might reasonably be broken into positive bond, positive work, and negative relationship ruptures, as Burlingame and colleagues (2021) show with the GQ. With the many possible rupture types in group therapy, rupture and repair is a promising direction for a more nuanced understanding of the formation and fluctuation of group therapeutic relationships (Tasca & Maramarosh, 2023).

Future research might also account for the different components of group therapeutic relationship constructs. Partitioning state- and trait-like alliance using various statistical methods is a developing practice in individual therapy (Zilcha-Mano & Fisher, 2022) and group therapy research (Kivlighan et al., 2022). As Zilcha-Mano and Fisher explain, "trait-like alliance is a product of patients' and therapists' intrapersonal and interpersonal characteristics, whereas state-like alliance reflects the therapeutic process evolving between them" (2022, p. 194). In the current study, positive bond, positive work, and negative relationship all had significant stability over time. This could point to possible trait-like elements of these constructs. There were also fluctuations and changes in scores over time, pointing toward stat-like elements of the constructs, particularly negative relationship. Identifying trait- and state-like elements of these constructs could help researchers and practitioners understand how traits and states impact group therapeutic relationships and the group experience at large.

Another way to measure trait-like elements of group therapeutic relationships could be the use of well-established personality measures to observe connections between these and group

therapeutic relationships (Aafjes-van Doorn et al., 2022). As Bainbridge, Ludeke, and Smillie (2022) show, Big-5 personality traits are often correlated with popular measures of psychological phenomena, including early measures of the therapeutic alliance (Aafjes-van Doorn et al., 2022). Group therapeutic relationships measured by the GQ could also be connected to broad personality traits. Regardless of the method to disentangle state- and trait-like group therapeutic relationships, the endeavor would admittedly be complicated. Group therapeutic relationships are often a function of the entire group, which Forsyth (2021) pointed out in regard to cohesion. Perhaps groups, with their unique combinations of members, exhibit trait-like elements of group therapeutic relationships. For example, perhaps certain groups quickly form positive bond, and others are more liable to spikes in negative relationships. Despite the complexity, disentangling trait- and state-like components of the GQ subscales at both the individual and group levels could be fruitful for empirical and clinical GQ use.

Conclusion

Our results indicated that codevelopment of positive bond and positive work is not due to mutual influence, as measured in a GAPIM framework. Mutual influence of individuals and their groups does play a role in the codevelopment of negative relationships between the midpoint and end of group treatment. These results have important practical implications, as quick attention to negative relationships and working to resolve conflict in the middle of group may help improve relationships at the end of group.

References

Aafjes-van Doorn, K., Kealy, D., Ehrenthal, J. C., Ogrodniczuk, J. S., Joyce, A. S., & Weber, R.
(2022). The relationship between patients' personality traits, the alliance, and change in
interpersonal distress in intensive group treatment for personality dysfunction. *Journal of
Personality Disorders*, *36*(6), 731–748. https://doi.org/10.1521/pedi.2022.36.6.731

Alldredge, C. T., Burlingame, G. M., Yang, C., & Rosendahl, J. (2021). Alliance in group
therapy: A meta-analysis. *Group Dynamics: Theory, Research, and Practice*, *25*(1), 13–
28. https://doi.org/10.1037/gdn0000135

Bainbridge, T. F., Ludeke, S. G., & Smillie, L. D. (2022). Evaluating the Big Five as an
organizing framework for commonly used psychological trait scales. *Journal of
Personality and Social Psychology*, *122*(4), 749–777.
https://doi.org/10.1037/pspp0000395

Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or
type I errors? Problems with the analysis of data from group-administered treatments.
Journal of Consulting and Clinical Psychology, *73*(5), 924–935.
https://doi.org/10.1037/0022-006X.73.5.924

Baldwin, S., Imel, Z., & Atkins, D. (2012). The influence of therapist variance on the
dependability of therapists' alliance scores: A brief comment on "the dependability of
alliance assessments: The alliance-outcome correlation is larger than you think" (Crits-
Christoph et al. *J Consult Clin Psychol*, *80*, 947–951. https://doi.org/10.1037/a0027935

Bennis, W. G., & Shepard, H. A. (1956). A theory of group development. *Human Relations*,
9(4), 415–437. https://doi.org/10.1177/001872675600900403

Bonito, J. A. (2021). Extending the group actor-partner interdependence model: MSEM

    applications. *Small Group Research*, 10464964211012456.

    https://doi.org/10.1177/10464964211012457

Bonsaksen, T., Borge, F.-M., & Hoffart, A. (2013). Group climate as predictor of short- and

    long-term outcome in group therapy for social phobia. *International Journal of Group*

    Psychotherapy, *63*(3), 394–417. https://doi.org/10.1521/ijgp.2013.63.3.394

Borgers, S. B., & Koenig, R. W. (1983). Uses and effects of modeling by the therapist in group

    therapy. *The Journal for Specialists in Group Work*, *8*(3), 133–138.

    https://doi.org/10.1080/01933928308411743

Burlingame, G. M., Alldredge, C. T., & Arnold, R. A. (2021). Alliance rupture detection and

    repair in group therapy: Using the Group Questionnaire--GQ. *International Journal of*

    Group Psychotherapy, *71*(2), 338–370. https://doi.org/10.1080/00207284.2020.1844010

Burlingame, G. M., Gleave, R., Beecher, M., Griner, D., Hansen, K., Jensen, J., Worthen, V., &

    Svien, H. (2017). *Group Questionnaire OQ®-GQ*. 76.

Burlingame, G. M., Kircher, J. C., & Honts, C. R. (1994). Analysis of variance versus bootstrap

    procedures for analyzing dependent observations in small group research. *Small Group*

    Research, *25*(4), 486–501. https://doi.org/10.1177/1046496494254004

Burlingame, G. M., McClendon, D. T., & Yang, C. (2018). Cohesion in group therapy: A meta-

    analysis. *Psychotherapy*, *55*(4), 384–398. https://doi.org/10.1037/pst0000173

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*, *53*(4), 446–461. https://doi.org/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*, *55*(2), 116–131. https://doi.org/10.1037/pst0000133

Carlucci, S., Chyurlia, L., Presniak, M., Mcquaid, N., Wiley, J. C., Wiebe, S., Hill, R., Garceau, C., Baldwin, D., Slowikowski, C., Ivanova, I., Grenon, R., Balfour, L., & Tasca, G. A. (2022). A group's level of defensive functioning affects individual outcomes in group psychodynamic-interpersonal psychotherapy. *Psychotherapy*, *59*, 57–62. https://doi.org/10.1037/pst0000423

Chapman, C. L., Baker, E. L., Porter, G., Thayer, S. D., & Burlingame, G. M. (2010). Rating group therapist interventions: The validation of the Group Psychotherapy Intervention Rating Scale. *Group Dynamics: Theory, Research, and Practice*, *14*(1), 15–31. https://doi.org/10.1037/a0016628

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cook, W. L., & Kenny, D. A. (2005). The Actor–Partner Interdependence Model: A model of bidirectional effects in developmental studies. *International Journal of Behavioral Development*, *29*(2), 101–109. https://doi.org/10.1080/01650250444000405

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Forsyth, D. R. (2021). Recent advances in the study of group cohesion. *Group Dynamics: Theory, Research, and Practice*, *25*, 213–228. https://doi.org/10.1037/gdn0000163

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. https://doi.org/10.1177/0011000090181002

Giannone, F., Guarnaccia, C., Gullo, S., Di Blasi, M., Giordano, C., Lo Coco, G., & Burlingame, G. (2020). Italian adaptation of the Group Questionnaire: Validity and factorial structure. *Research in Psychotherapy : Psychopathology, Process, and Outcome*, *23*(2), 443. https://doi.org/10.4081/ripppo.2020.443

Gonzalez, R., & Griffin, D. (2001). A statistical framework for modeling homogeneity and interdependence in groups. *Handbook of Social Psychology*, *2*, 505–534.

Gullo, S., Lo Coco, G., Pazzagli, C., Piana, N., De Feo, P., Mazzeschi, C., & Kivlighan, D. (2014). a time-lagged, actor-partner interdependence analysis of alliance to the group as a whole and group member outcome in overweight and obesity treatment groups. *Journal of Counseling Psychology*, *61*. https://doi.org/10.1037/a0036084

Johnson, J. E., Burlingame, G. M., Olsen, J. A., Davies, D. R., & Gleave, R. L. (2005). Group climate, cohesion, alliance, and empathy in group psychotherapy: Multilevel structural equation models. *Journal of Counseling Psychology*, *52*(3), 310–321. https://doi.org/10.1037/0022-0167.52.3.310

Kassambara, A. (2022). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (0.7.1). https://CRAN.R-project.org/package=rstatix

Kenny, D. A., & Garcia, R. L. (2012). Using the actor–partner interdependence model to study the effects of group composition. *Small Group Research*, *43*(4), 468–496. https://doi.org/10.1177/1046496412441626

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, *83*(1), 126–137. https://doi.org/10.1037/0022-3514.83.1.126

Kivlighan, D. M., Cohen, K., Zilcha-Mano, S., Palgi, S., & Zilberstein, A. (2022). Examining state and trait alliance in group therapy: A within-person and between-person actor–partner interdependence model. *Group Dynamics: Theory, Research, and Practice*, *26*(2), 137–150. https://doi.org/10.1037/gdn0000152

Kivlighan, D. M. I., & Narvaez, R. C. (2021). Mutual influence in group psychotherapy: A review and application to group psychology. In C. D. Parks & G. A. Tasca (Eds.), *The psychology of groups: The intersection of social psychology and psychotherapy research.* (2020-56134-011; pp. 191–206). American Psychological Association. https://doi.org/10.1037/0000201-011

Kivlighan, D. M., Lo Coco, G., & Gullo, S. (2012). Attachment anxiety and avoidance and perceptions of group climate: An actor–partner interdependence analysis. *Journal of Counseling Psychology*, *59*(4), 518–527. https://doi.org/10.1037/a0030173

Krogel, J., Burlingame, G., Chapman, C., Renshaw, T., Gleave, R., Beecher, M., & MacNair-Semands, R. (2013). The Group Questionnaire: A clinical and empirically derived measure of group relationship. *Psychotherapy Research*, *23*(3), 344–354. https://doi.org/10.1080/10503307.2012.729868

Lleras, C. (2005). Path analysis. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 25-30). Elsevier. https://doi.org/10.1016/B0-12-369398-5/00483-7

Lo Coco, G., Gullo, S., Albano, G., Brugnera, A., Flückiger, C., & Tasca, G. A. (2022). The alliance-outcome association in group interventions: A multilevel meta-analysis. *Journal*

of Consulting and Clinical Psychology, *90*(6), 513-527.

https://doi.org/10.1037/ccp0000735

Lo Coco, G., Gullo, S., Di Fratello, C., Giordano, C., & Kivlighan, D. M. Jr. (2016). Group

relationships in early and late sessions and improvement in interpersonal problems.

Journal of Counseling Psychology, *63*(4), 419–428. https://doi.org/10.1037/cou0000153

Lo Coco, G., Gullo, S., Oieni, V., Giannone, F., Di Blasi, M., & Kivlighan, D. (2016). The

relationship between attachment dimensions and perceptions of group relationships over

time: An actor–partner interdependence analysis. *Group Dynamics: Theory, Research,*

and Practice, *20*. https://doi.org/10.1037/gdn0000056

Lo Coco, G., Gullo, S., Profita, G., Pazzagli, C., Mazzeschi, C., & Kivlighan, D. M. (2019). The

codevelopment of group relationships: The role of individual group member's and other

group members' mutual influence and shared group environment. *Journal of Counseling*

Psychology, *66*(5), 640–649. https://doi.org/10.1037/cou0000349

Lorentzen, S., Strauss, B., & Altmann, U. (2018). Process-outcome relationships in short- and

long-term psychodynamic group psychotherapy: Results from a randomized clinical trial.

Group Dynamics: Theory, Research, and Practice, *22*(2), 93–107.

https://doi.org/10.1037/gdn0000084

Ogrodniczuk, J. S., Joyce, A. S., & Piper, W. E. (2007). Effect of patient dissatisfaction with the

therapist on group therapy outcome. *Clinical Psychology & Psychotherapy*, *14*(2), 126–

134. https://doi.org/10.1002/cpp.526

Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An

empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy*, *10*(6), 361–

373. https://doi.org/10.1002/cpp.383

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Orth, U., Erol, R. Y., Ledermann, T., & Grob, A. (2018). Codevelopment of well-being and self-esteem in romantic partners: Disentangling the effects of mutual influence and shared environment. *Developmental Psychology*, *54*(1), 151–166. https://doi.org/10.1037/dev0000400

Pedersen, G., Kvarstein, E. H., Wilberg, T., Folmo, E. J., Burlingame, G. M., & Lorentzen, S. (2021). The Group Questionnaire (GQ)—Psychometric properties among outpatients with personality disorders. *Group Dynamics: Theory, Research, and Practice*, Advance online publication. https://doi.org/10.1037/gdn0000176

Phipps, L. B., & Zastowny, T. R. (1988). Leadership behavior, group climate and outcome in group psychotherapy: A study of outpatient psychotherapy groups. *Group*, *12*(3), 157–171. https://doi.org/10.1007/BF01456565

Project MATCH Research Group. (1998). Therapist effects in three treatments for alcohol problems. *Psychotherapy Research*, *8*(4), 455–474. https://doi.org/10.1093/ptr/8.4.455

R Core Team. (2022). *R: A language and environment for statistical computing* (4.2.2). R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research* (2.2.9). https://CRAN.R-project.org/package=psych

Shi, D., DiStefano, C., Maydeu-Olivares, A., & Lee, T. (2022). Evaluating SEM model fit with small degrees of freedom. *Multivariate Behavioral Research*, *57*(2–3), 179–207. https://doi.org/10.1080/00273171.2020.1868965

StataCorp. (2021). *Stata Statistical Software: Release 17* (Version 17). StataCorp.

    https://www.stata.com/

Tasca, G. A., & Marmarosh, C. (2023). Alliance rupture and repair in group psychotherapy. In C.

    F. Eubanks, L. W. Samstag, & J. C. Muran (Eds.), *Rupture and repair in psychotherapy:*

    A critical process for change. (pp. 53–71). American Psychological Association.

    https://doi.org/10.1037/0000306-003

Thayer, S. D., & Burlingame, G. M. (2014). The validity of the Group Questionnaire: Construct

    clarity or construct drift? *Group Dynamics: Theory, Research, and Practice*, *18*(4), 318–

    332. https://doi.org/10.1037/gdn0000015

Tierney, N., Cook, D., McBain, M., Fay, C., O'Hara-Wild, M., Hester, J., Smith, L., & Heiss, A.

    (2021). *naniar: Data Structures, Summaries, and Visualisations for Missing Data* (0.6.1).

    https://CRAN.R-project.org/package=naniar

United States Census Bureau. (n.d.). *U.S. Census Bureau QuickFacts: Utah*. Retrieved January

    18, 2023, from https://www.census.gov/quickfacts/UT

Wellen, J. M., & Neale, M. (2006). Deviance, self-typicality, and group cohesion: The corrosive

    effects of the bad apples on the barrel. *Small Group Research*, *37*(2), 165–186.

    https://doi.org/10.1177/1046496406286420

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund,

    G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M.,

    Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019).

    Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

    https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & RStudio. (2022). *dplyr: A Grammar of Data Manipulation* (1.0.10). https://CRAN.R-project.org/package=dplyr

Wickham, H., Miller, E., & Smith, D. (2022). *haven: Import and Export "SPSS", "Stata" and "SAS" Files* (2.5.1). https://CRAN.R-project.org/package=haven

Wolak, M. (2022). *ICC: Facilitating Estimation of the Intraclass Correlation Coefficient* (2.4.0). https://CRAN.R-project.org/package=ICC

Xie, Y. (2022). *knitr: A general-purpose package for dynamic report generation in R* (1.41). https://yihui.org/knitr/

Yalom, I. D. & Leszcz, M. (2020). *The theory and practice of group psychotherapy* (6th ed.). Basic Books.

Zilcha-Mano, S., & Fisher, H. (2022). Distinct roles of state-like and trait-like patient–therapist alliance in psychotherapy. *Nature Reviews Psychology*, *1*(4), 194–210. https://doi.org/10.1038/s44159-022-00029-z

Appendix A

Figure A.1

Power by Sample Size (n=10-400) with Effect Size r = .1 to r = .5

t tests - Correlation: Point biserial model
Tail(s) = Two. α err prob = 0.05
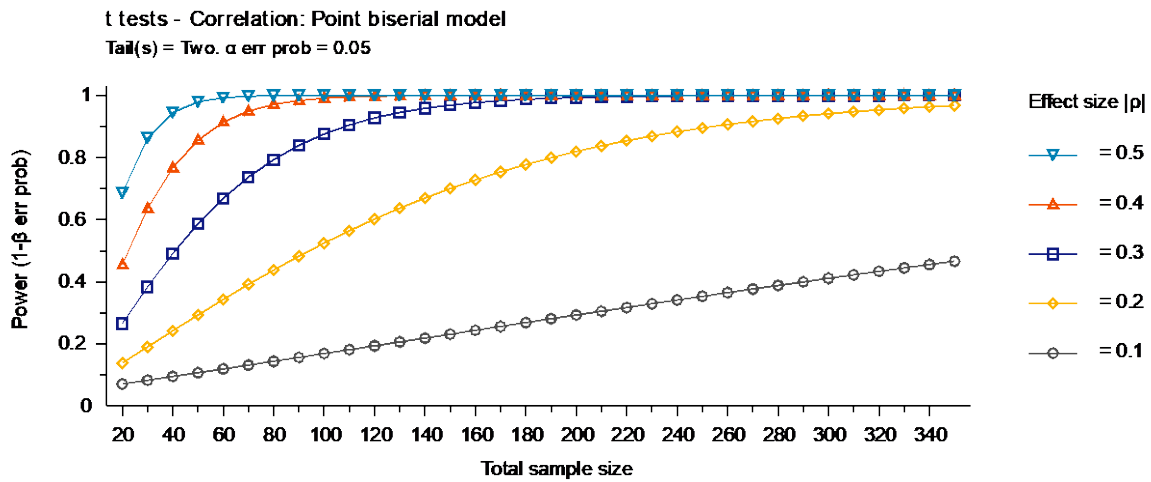
Figure A.2

Power by Sample Size (n=10-70) with Effect Size r = .1 to r = .5

t tests - Correlation: Point biserial model
Tail(s) = Two. α err prob = 0.05
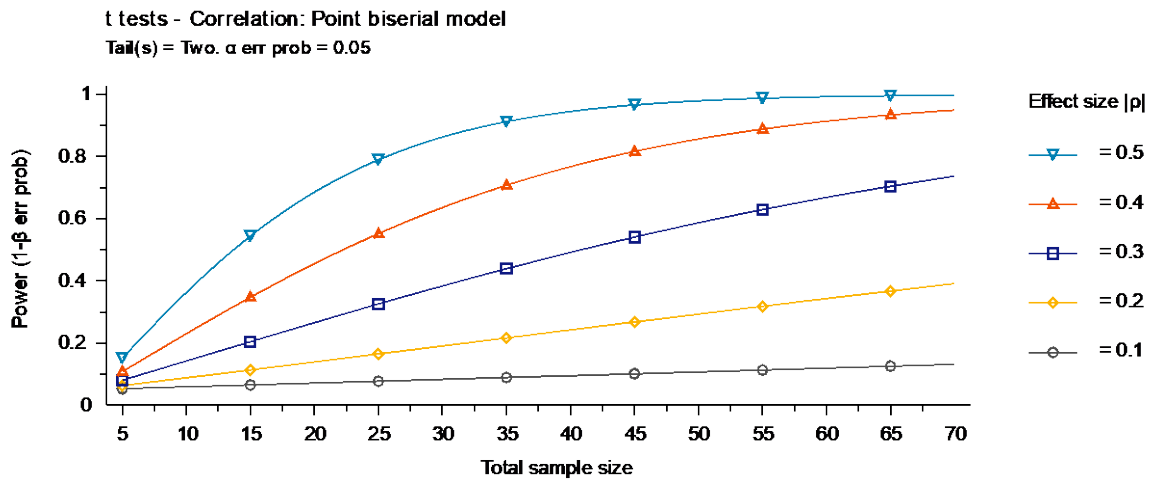
Table A.1

Participant Data Missingness by Stage and Overall

Stage	Missing	
	n	%

Early	33	9.5%
Middle	32	9.2%
Late	40	11.5%

Observations Obtained	n	%
3 of 3	259	74.4%
2 of 3	73	21%
1 of 3	16	4.6%

Figure A.3

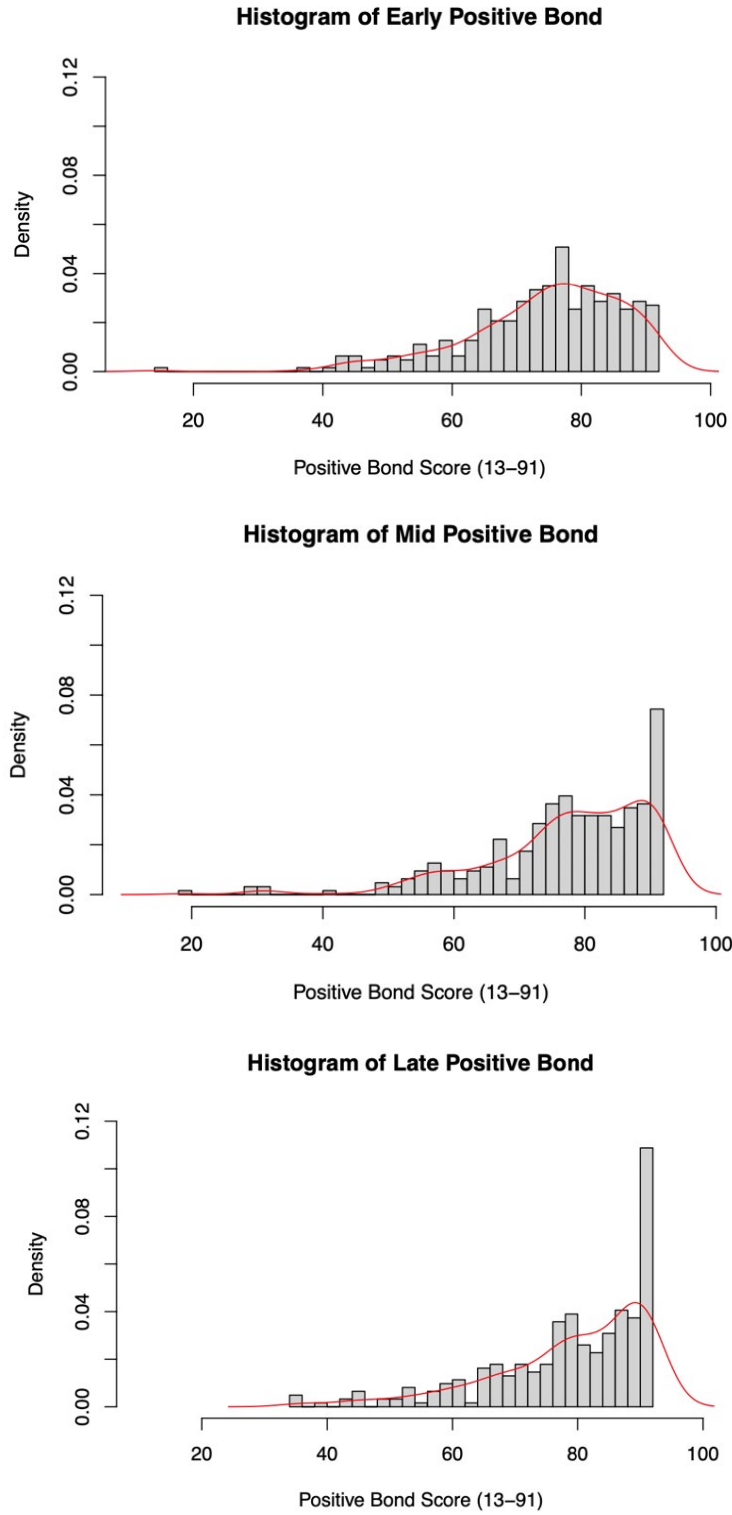*Histograms of Positive Bond at Three Stages (Early, Middle, Late)*

Figure A.4

Histograms of Positive Work at Three Stages (Early, Middle, Late)
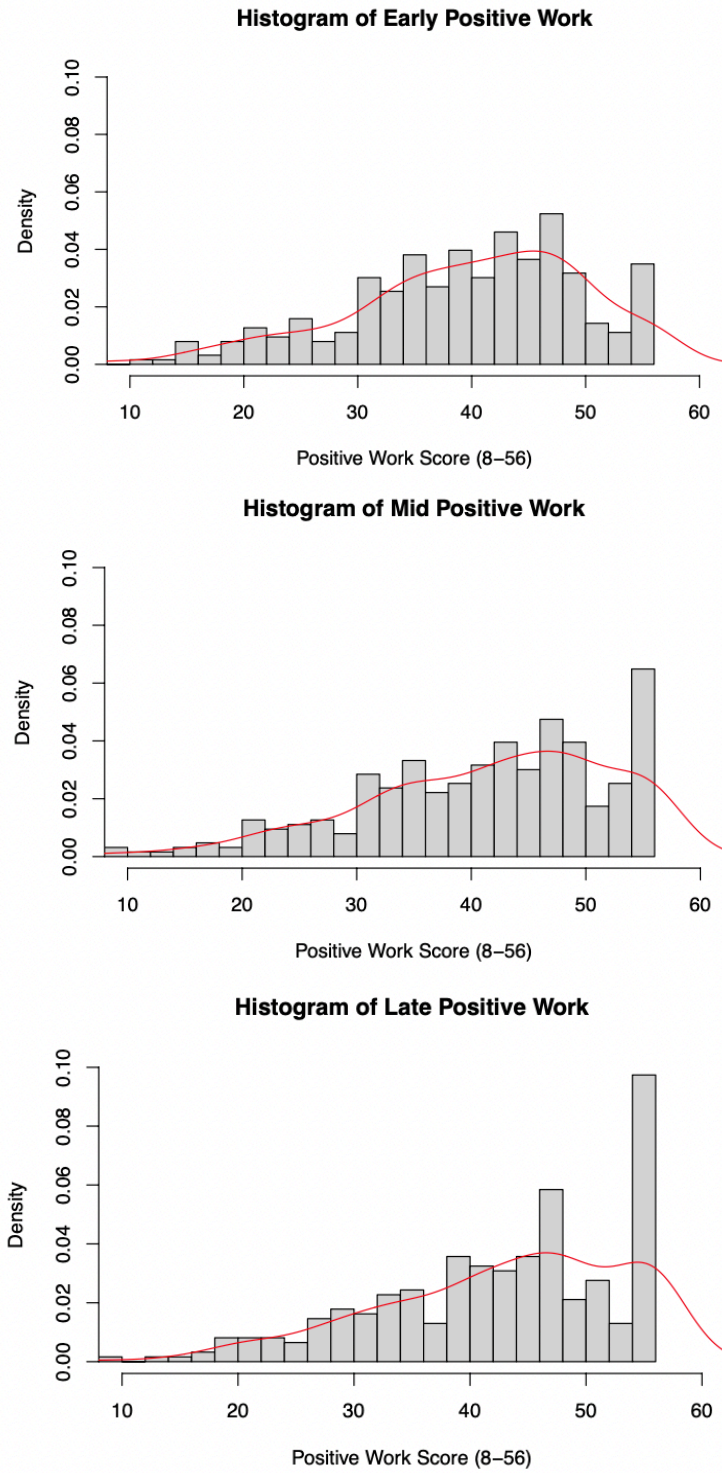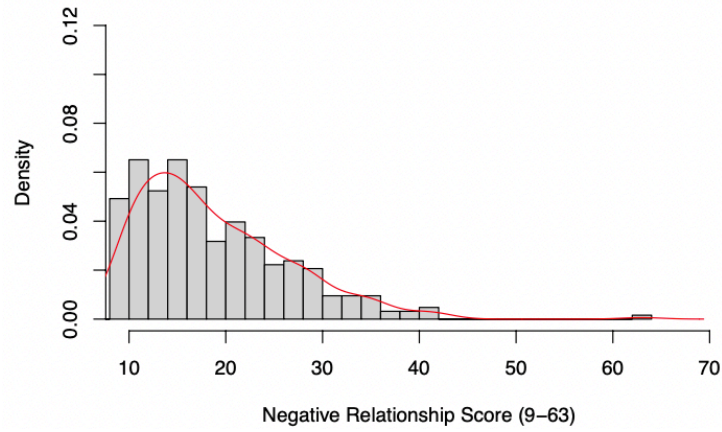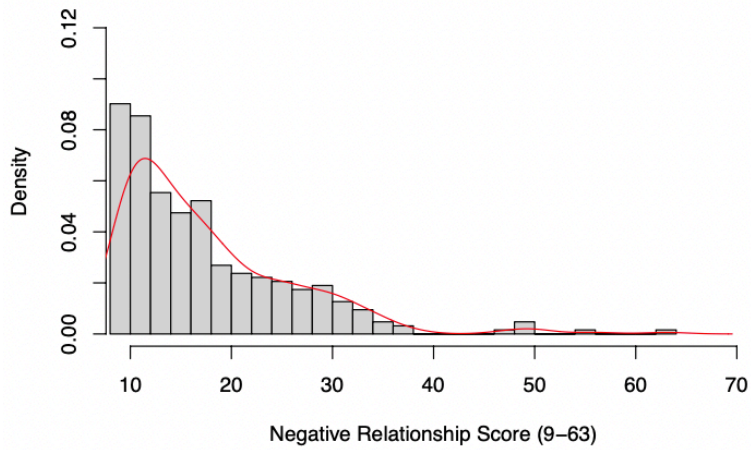
Histogram of Early Positive Work

Histogram of Mid Positive Work

Histogram of Late Positive Work

Figure A.5

Histograms of Negative Relationship at Three Stages (Early, Middle, Late)

Histogram of Early Negative Relationship

Histogram of Mid Negative Relationship

Histogram of Late Negative Relationship

59