



2008-07-16

The Group Questionnaire: A New Measure of the Group Relationship

JulieAnn Krogel

Brigham Young University - Provo

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



Part of the [Psychology Commons](#)

BYU ScholarsArchive Citation

Krogel, JulieAnn, "The Group Questionnaire: A New Measure of the Group Relationship" (2008). *All Theses and Dissertations*. 1732.
<https://scholarsarchive.byu.edu/etd/1732>

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

THE GROUP QUESTIONNAIRE:
A NEW MEASURE OF THE GROUP RELATIONSHIP

By

JulieAnn Krogel

A dissertation submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology

Brigham Young University

August 2008

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

JulieAnn Krogel

This dissertation has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date

Gary M. Burlingame, Chair

Date

Robert Gleave

Date

Mark Beecher

Date

Sally H. Barlow

Date

Bruce N. Carpenter

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of JulieAnn Krogel in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Gary M. Burlingame
Chair, Graduate Committee

Accepted for the Department

Date

Sally H. Barlow
Director of Clinical Training

Accepted for the College

Date

Susan Rugh
Associate Dean, College of Family,
Home and Social Sciences

ABSTRACT

THE GROUP QUESTIONNAIRE: A NEW MEASURE OF THE GROUP RELATIONSHIP

JulieAnn Krogel

Department of Psychology

Doctor of Philosophy

The Group Questionnaire (GQ) is a 30-item, self-report measure of the group relationship that was developed in the present study. It is based off of Johnson's new three factor model of the group relationship which includes Positive Bonding, Positive Working, and Negative Relationship. This study involved two parts, the creation of the GQ followed by the validation and refinement of the GQ using 486 participants from three populations - outpatient university counseling center, inpatient state hospital, and non-patient AGPA process groups. Confirmatory Factor Analysis (CFA) was used to test the goodness-of-fit of Johnson's model to each of the populations separately and as a whole. Following the refinement process, results showed the GQ to have good fit to the model for each population. Distinct differences in response pattern were found between the three populations. Reliability estimates were predominantly in the good range. Implications for future utility of the GQ include using it as a clinically relevant and efficient assessment tool to inform clinicians of the quality of the group psychotherapy relationship and to potentially predict group outcome.

TABLE OF CONTENTS

Title.....	i
Graduate Committee Approval.....	ii
Final Reading Approval and Acceptance	iii
Abstract.....	iv
Table of Contents.....	v
List of Tables	vi
List of Figures.....	vii
Chapter 1: Introduction.....	1
Chapter 2: Literature Review.....	6
Chapter 3: Methods.....	28
Chapter 4: Results.....	46
Chapter 5: Discussion	84
References.....	94

LIST OF TABLES

Table 1-1: Measures Used in Johnson et al. 2005	10
Table 3-1: Specification of Fit Indices.....	37
Table 4-1: Second Order Factor Definitions.....	47
Table 4-2: GQ Items	48
Table 4-3: Items Dropped or Reworded to Create the GQ	50
Table 4-4: GQ Standardized Regression Weights	55
Table 4-5: 3x3 MANOVA	62
Table 4-6: Population Factor Means and Standard Deviations	62
Table 4-7: Reliability of the GQ	66
Table 4-8: Adjusted Reliability of the GQ	67
Table 4-9: Items Dropped from the Original GQ	71
Table 4-10: Refined GQ Standardized Regression Weights	73
Table 4-11: Population by Factor Means and Standard Deviations	81
Table 4-12: 3x3 MANOVA	81
Table 4-13: Reliability for the Refined GQ	82
Table 4-14: Adjusted Reliability for the Refined GQ	83

LIST OF FIGURES

Figure 2-1: Johnson’s Model of Group Therapeutic Relationship	24
Figure 2-2: Second Order Factors of Johnson’s Model.....	25
Figure 3-1: GQ Structure Tested in Hypothesis 1	41
Figure 3-2: GQ Model without 1 st Order Factors	43
Figure 4-1: Hypothesis 1 – Confirmatory Factor Analysis.....	53
Figure 4-2: Hypothesis 3 – CFA without 1 st Order Factors.....	56
Figure 4-3: Hypothesis 4 – Original Model with UCC.....	58
Figure 4-4: Hypothesis 4 – Original Model with AGPA.....	59
Figure 4-5: Hypothesis 4 – Original Model with USH.....	60
Figure 4-6: Reliability Estimates	65
Figure 4-7: Refined GQ Structural Model.....	69
Figure 4-8: Hypothesis 1 – CFA of Refined GQ.....	72
Figure 4-9: Hypothesis 3 – Refined GQ without 1 st Order Factors	75
Figure 4-10: Hypothesis 4 – New Model with UCC	78
Figure 4-11: Hypothesis 4 – New Model with AGPA.....	79
Figure 4-12: Hypothesis 4 – New Model with USH.....	80

Chapter 1

Introduction

Recent evidence suggests that the demand for group treatment in clinical practice may be growing in North America (Burlingame, MacKenzie, & Strauss, 2004). The trend may be partially due to the beneficial cost-benefit ratio when group format is compared to individual therapy (Heinzel, Breyer, & Klein, 1998). Additionally, several reviews have now been conducted of the literature comparing group to individual treatment. These have shown group treatment to be at least as effective as individual psychotherapy (Burlingame et al., 2004; Fuhriman & Burlingame, 1994; McRoberts, Burlingame, & Hoag, 1998; Tschuschke, 1999). With the effectiveness and cost-benefit of group psychotherapy firmly established, researchers have turned their attention to examining why group treatments work. Some of the proposed mechanisms of change of group treatment relate to specific theories of change such as cognitive restructuring, exposure, and unconditional positive regard. Other mechanisms of change have been identified which transcend the specific theories and instead focus on properties of the group format itself. The therapeutic relationship that occurs within the group has been identified as one of the most important general mechanisms of change in group treatment (Johnson, Burlingame, Olsen, Davies, & Gleave, 2005). Group relationships that are both supportive and challenging have been shown to be a consistent contributing factor to patient improvement and low dropout rates (Burlingame, Fuhriman, & Johnson, 2002; Castonguay, Pincus, Agras, & Hines, 1998; Marziali, Munroe-Blum, & McCleary, 1997). It is presumed that these types of relationships provide the type of safe environment needed to conduct the challenging work that takes place in psychotherapy groups.

Many different aspects of the curative relationship have been examined in relation to client improvement; however, they have only recently begun to be studied in relationship to each other. The most notable attempt was a large study conducted by Johnson et al. in 2005. Johnson administered subscales from the Therapeutic Factors Inventory (TFI), the Group Climate Questionnaire (GCQ), the Working Alliance Inventory (WAI) and the Empathy Scale (ES) to 662 participants from 111 counseling center and personal growth groups. As hypothesized, she found a significant degree of overlap between the different aspects of the curative relationship (cohesion, alliance, empathy, and group climate). She tested three different structural equation models and found none had an adequate fit to the data. She found an exploratory model with Bonding, Working, and Negative factors provided the best fit to the data (Johnson et al., 2005). This study was a notable step toward creating a comprehensive understanding of the therapeutic relationship: a challenge that has been historically difficult in group literature.

One obstacle that has historically made it difficult to empirically study the therapeutic relationship is the existence of two principally isolated views. The first conceptualizes the therapeutic relationship by way of the group's structural properties. More specifically, relationships can be formed between a group member and the leader (member-leader), between group members (member-member), or between group members and the group as a whole (member-group). The literature has not given equal attention to the testing and description of these dimensions. A recent review of this literature found that two-thirds of the literature investigated the member-group relationship. In contrast, far fewer studies tested the member-member relationship, with

the member-leader relationship being nearly neglected (Burlingame, Fuhriman, & Johnson, 2002). The definite bias in the literature makes it difficult to make an empirically based judgment as to the relevance and relatedness of the three structural definitions of the therapeutic relationship in groups.

A second view conceptualizes the therapeutic relationship in terms of the specific constructs theorized to make up the therapeutic relationship in group. Many therapeutic relationship constructs have been hypothesized to exist. Two recent reviews identified a core set of therapeutic relationship constructs that have been shown to be related to group process and client improvement (Burlingame, Fuhriman, & Johnson, 2002). Among these were the constructs of group cohesion (a sense of belonging to the group), alliance (a fond working relationship between the therapist and the member), group climate (a sense of constructive interpersonal investigation) and empathy (a sense of being understood). Each of these core constructs has been shown to be linked to group process and client improvement, and Johnson has used each to create a comprehensive model of the therapeutic relationship (Johnson et al., 2005).

Prior to the Johnson et al.'s (2005) study, only measures of cohesion and alliance had been empirically compared in the same investigation. Thus, it was unknown how the former relationship constructs related to one another, which lead to some distinct limitations. The first and most prominent limitation was definitional. Authors of existing measures of these constructs seem to have varying definitions for the same concept. Furthermore, variations in instrumentation and source (for example, self-report vs. observation) further complicate links between constructs. More specifically, even if

constructs were clearly defined, differences in methods of instrumentation and source have well-known effects on correlations (Orlinsky, Ronnestad, & Willutzki, 2003).

A final challenge in interpreting the group psychotherapy literature is defining the relationships between constructs both conceptually and empirically. As mentioned earlier, until Johnson et al. (2005), only the relationship between alliance and cohesion had been studied. There are many other group relationship constructs that have been found to be linked to outcome, such as mutual understanding (Roback & Smith, 1987), therapists' like or dislike of clients (Roback & Smith, 1987; Sexton, 1993), therapists' friendliness (Hurley, 1986), understanding and trust in groups (Roarck & Sharah, 1989), feelings of being accepted by the group (MacKenzie, 1998; Rugel & Berry, 1990), and emotional relatedness to other group members (Tuchuschke & Dies, 1994). These could be aspects of cohesion, alliance, empathy, and group climate; however, with the inconsistency of the labeling, it is difficult to piece these together.

Johnson et al.'s research (2005) addressed the aforementioned limitations in three ways. First, it used a common source: self-report. Second, it employed one commonly used measure to assess each of the four content domains (cohesion, alliance, empathy, and group climate). Third, it simultaneously looked at the four domains using a common sample and tested different types of groups after various numbers of sessions. This resulted in the creation of a new model of the group therapeutic relationship. The model clearly shows the relationship between its three main constructs (Positive Bonding, Positive Working, and negative factors) as well as the relationship of the four main relationship constructs (cohesion, alliance, empathy, and group climate) and the perspectives (member-member, member-leader, and member-group). Evidence for this

model was very recently validated and strengthened when two studies using two different European populations replicated Johnson's study and found the same factor structure using confirmatory factor analysis (Bormann & Straus, 2007; Bakali et al. 2008).

Evidence for this model was very recently validated and strengthened when Bormann and Straus (2007) replicated Johnson's study using a German population and found the same factor structure using confirmatory factor analysis.

The purpose of the present study was to create the Group Questionnaire (GQ) based off of the Johnson et al. (2005) model. The creation of the GQ relied heavily on both empirical results and clinical judgment to create a measure that is empirically sound, clinically relevant, and practically efficient. The creation of the GQ took place in two steps: First, using empirical data and clinical judgment, a team of experienced group researchers and clinicians pared down and adapted the 60 items used by Johnson et al. to create the GQ as a measure of the group relationship including Positive Bonding, Positive Working, and Negative Relationship. Second, the GQ was tested and revised using confirmatory factor analysis (CFA) and data from three populations: outpatient university counseling center (UCC), non-patient American Group Psychotherapy Association (AGPA), and inpatient Utah State Hospital (USH). The penultimate goal of this line of research, which goes beyond the scope of this study, is to create a measure of the group relationship that can be used to track process change previously linked to outcome in group members.

Chapter 2

Literature Review

This chapter reviews the previous literature on the group relationship, specifically the definitions, and empirical findings, of the four main relationship constructs (group climate, cohesion, alliance, and empathy) examined by Johnson et al. (2005). It also describes how previous group relationship constructs have been theoretically and empirically connected to member-group, member-member, and member-leader relationships. It then reviews the Johnson, et al. (2005) study in detail. The purpose of this literature review is to demonstrate the clinical importance of group relationship and the need for a new measure that is more empirically sound, clinically relevant, and practically short. It also provides the historical and theoretical background contributing to its development. As this study is a continuation of a line of research starting with Johnson et al.'s study, the current literature review parallels, updates, and summarizes the literature contained in this previous study.

Group Climate

Definitions: The popularity of the Group Climate Questionnaire-Short Form (GCQ; MacKenzie, 1983) has done a great deal to unify the construct of group climate. Due to the dominance of the GCQ, it will be used as the definition of Group Climate for this review. The GCQ is believed to reflect the functioning of the group as a whole rather than an individual's experience in the group. For example, items are phrased "the group members did _____," rather than "I felt _____ in group today." This distinction is important as the perceptions of group climate as a whole may be unrelated to an individual member's experience being part of the group. For this reason, MacKenzie and

Tschuschke (1993) hypothesized that the GCQ did not predict outcome in their study because the unsuccessful patients may have thought, “the group is really working well, but I am not a part of it” (p.154). However, the GCQ does provide information on the perceptions of the group by its members, not the reality of the group, which may be more relevant to the individuals’ receptiveness to particular types of interaction (MacKenzie, 1984).

The current version of the GCQ, the GCQ-S for short form (GCQ-S; MacKenzie, 1983), contains three subscales of Engagement, Avoidance, and Conflict. This shorter, 12-item scale has almost entirely replaced the original longer version, which is why the GCQ-S is now almost universally referred to with the original acronym GCQ. The original, longer form included the subscales of engagement, support, practicality, disclosure, cognition, challenge, conflict, and control and was designed to study social environments and how they influence behavior (GCQ-L; MacKenzie, 1981). From this point forward all references to the GCQ are referring to the GCQ-S and its three subscales. The Engagement scale from the GCQ reflects a positive working atmosphere and a sense of constructive interpersonal investigation, including the importance of the group for members and a sense of closeness between them. The Avoidance scale reflects the perception of avoidance of responsibility by the members for their own change process. The Conflict scale reflects interpersonal conflict and distrust of the members (MacKenzie, 1983).

Outcome Findings: The Engagement scale of the GCQ has been most consistently related to outcome. Findings suggest that groups with higher engagement scores tend to have better outcomes (Johnson, 2004). In the largest study conducted

using the GCQ, MacKenzie et al. (1987) found that the 8 groups (out of 54) reporting the greatest learning at the American Group Psychotherapy Association Institutes were significantly more engaged at all times than the 8 groups reporting the lowest learning. Additionally, reported levels of engagement in these groups increased over the lifetime of these groups. A second study (Tschuschke & Greene, 2002) using the same population used a slightly altered Engagement subscale and predicted learning about self-peer relationships, authority, and leadership. Other studies have also found engagement to be positively related to both interpersonal outcomes (Braaten, 1989) and general outcomes in eating disorder groups (Castonguay, Pincus, Agras, & Hines, 1998) as well as group treatment for complicated grief (Ogrodniczuk & Piper, 2003). However, engagement has not been shown to predict outcome in long-term, psychoanalytically-oriented groups for inpatient neurotics (MacKenzie & Tschuschke, 1993) or in 12-week, process-oriented groups for females sexually abused as children (Cortez-Ison, 1997).

The Conflict scale of the GCQ typically correlates negatively with outcomes (Johnson, 2004). It helps to remember that conflict as measured by the GCQ is a measure of hostility rather than confrontation. MacKenzie et al. (1987) found that self-reported levels of conflict decreased over time in personal growth groups that experienced the most learning, whereas self-reported conflict in groups that reported less learning started at the same level, but remained unchanged over the lifetime of the group. Phipps and Zastowny (1988) found that their most poorly functioning groups reported high levels of conflict.

Some research does not find conflict to be related to outcome, and other research finds it to be positively related to outcome under certain conditions. Ogrodniczuk and

Piper (2003) failed to find an effect for conflict, reporting instead that members who were less engaged and more avoidant during times of group conflict had better outcomes.

These authors suggested that a high level of involvement in conflict as measured by the GCQ does not appear to have been helpful to members and that withdrawing from the group may serve as a buffer from the negative consequences. However, another study by Tschuschke and Greene (2002) showed members of more conflictual personal growth groups report greater learning. Along the same lines, Castonguay et al. (1998), with a group of eating disorder patients, found that conflict in the middle of the group's lifetime had good outcomes so long as early and later group climates were positive.

In comparison to engagement and conflict, avoidance seldom predicts improvements in outcome (Ogrodniczuk & Piper, 2003). The few findings that do exist seem contradictory, as seen with conflict. Braaten (1989) found that outpatient groups reporting less avoidance predicted improved interpersonal outcomes. However, Tschuschke and Green (2002) found that personal growth groups reporting more avoidance predicted greater learning about self-peer relationships.

Operationalization: Johnson et al. (2005) chose the GCQ to represent group climate for their study due to its overwhelming popularity and good psychometrics. A review of other measures of group climate, as well as the other three constructs (cohesion, alliance, and empathy) is contained in Johnson (2004). Table 1-1 provides a summary of the measures used by Johnson et al. (2005) as well as their uses and quality of psychometrics. According to MacKenzie (2003), its creator, the GCQ is believed to contain the three factors just reviewed: Engagement, Conflict, and Avoidance. They may, however, operate differently or be experienced differently by different groups.

Table 1-1: *Measures Used in Johnson et. al. 2005*

Measure	Number of Items and Subscales	Construct Coverage	Response Format	Psychometric Properties	Use in the Current Group Psychotherapy Literature
Group Climate Questionnaire (MacKenzie, 1983)	12 items 3 subscales	Engagement in group work, Avoidance of responsibility for group work, and group conflict and hostility	7-point Likert scale	Extensively Tested	Most commonly used group process measure, used by more than 40 studies over the last 20 years (MacKenzie, 1998)
Cohesion subscale of the Therapeutic Factors Inventory (Lese & MacNair-Semands, 2000)	12 items 1 of 11 subscales	Feeling accepted, sense of belonging, trust, caring, security and working together	7-point Likert, self report	Good reliability, validity	Relatively new scale, but used in several studies over the past few years
Working Alliance Inventory (WAI; Horvath & Greenburg, 1989)	36 items 3 subscales	Bond between therapist (T) and client (C), T-C agreement on tasks and goals of therapy bond between Therapist (T) and client (C), T- C	7-point Likert scale, self-report	Very well studied and very good	Probably the most commonly used measure of alliance in the current literature
Empathy Scale (ES; Burns, 1994)	10 items	Clients' perceptions of the therapist's warmth, genuineness and empathy	Various Likert scales, self- report	Adequate internal consistency, predicts outcome strongly	Probably the most commonly used measure of empathy in the individual literature

Other studies, including Johnson et al. (2005), have questioned this factor structure (Hurley & Brooks, 1987; Hurley & Brooks, 1988; Kay, 1996). For example, Hurley & Brooks (1987) suggest the GCQ to have one factor, which they named Affiliative vs. Disaffiliative. However, both this one factor and the original three factor model, point out that a few items load only moderately or even poorly onto the factor structure.

Cohesion

Definitions: Johnson, (2004) completed a thorough review of the Cohesion literature and concluded cohesion to be the least clearly defined group relationship construct. Kaul and Bednar (1986) have called cohesion “a spectacular embarrassment to group theory and research” because 30 years of effort have not produced an accepted definition of the term (p.707). In a recent review of therapeutic relationships in group psychotherapy, Burlingame, Fuhriman, and Johnson, (2002) identified 23 different measures of “cohesion” in only 31 studies. They point out that no one instrument was used more than 4 times. In addition to the plethora of cohesion measures, a given definition may be thought to represent group-as-a-whole, member-group, member-member, or member-leader relationships.

The concept of cohesion began in the 1940’s and 1950’s with the work of Festinger and colleagues. They defined cohesion as a “field of forces” acting on individuals to remain in the group (Dion, 2000). Following this, Lewin defined cohesion as the set of forces keeping members together and Frank (1957) called it the attraction of a group for its members (Dion, 2000). By the 1950’s, critics had already noted that different measures of cohesion often failed to correlate with one another. Around the mid-1960’s, cohesion was defined as interpersonal attraction: a group property inferred

from the number and strength of mutually positive attitudes among group members. Similar definitions reflected a group's desire to stick together, attraction, morale, or solidarity. It has also been described as a sense of group belongingness and group atmosphere, and correspondingly measured by the use of "we" rather than "I" (Dion, 2000).

More recent attempts at defining cohesion have focused on divisions in the understanding of the construct. Task vs. socioemotional cohesion and group-level vs. individual-level are two major divisions in the construct (Dion, 2000; see also Carron, Widmeyer, & Brawley, 1985). A factor analysis of therapy groups produced another division: vertical vs. horizontal cohesion (Griffith's, 1988). Vertical cohesion refers to a member's relationship with the group leader, whereas horizontal cohesion refers to members' relationships with other group members. Still other categorizations arise from the measurement perspective of cohesion. Some seek to measure perceived cohesion and individual members' perception of their sense of belonging to a group vs. a more "objective" rating of cohesion, as in Bollen and Hoyle (1990). An example would be items like "I feel a sense of belonging to the group" vs. "How the group sticks together." Additional dichotomies arise over appraisals at the affective vs. cognitive levels. For example, "I feel my group cares about me" vs. "My group shows appreciation for me." To clarify the complexity of these dichotomies, Dion (2000) suggests "[c]ohesiveness...means different things for different groups with different goals and tasks" (p. 22). Johnson (2004) suggests "[t]he cohesion definition most relevant to therapeutic relationships in psychotherapy groups involves the subjective experiences of individual group members with both the group leader and other group members, on both

task and socioemotional levels” (Johnson, p. 15). She uses this definition as well as popularity and psychometrics to guide her selection of the Cohesion subscale of the Therapeutic Factors Inventory to represent cohesion in her study.

Outcome Findings: The variety of cohesion definitions and measures makes it difficult to interpret outcome findings. However, cohesion of several different definitions has been found to be related to better outcomes (Stokes, 1983). For example, one study found that the more members liked the groups they were in, the better the outcomes they had (Wright & Duncan, 1986). Similarly, more successful group therapy patients were found to experience a greater sense of emotional relatedness to other group members throughout therapy (Tschuschke & Dies, 1994). Two other studies found cohesion to be a significant predictor of positive goal attainment and symptom reduction (Braaten, 1989; Budman et al., 1989). Still another study (McCallum et al., 2002) found that in short-term groups for complicated grief, significantly lower rates of dropout were associated with high levels of therapist-rated cohesion, but not patient-rated cohesion.

Cohesion appears to be an even better predictor of interpersonally related outcomes. This was also observed with the construct of engagement. Cohesion, as well as other measures of positive relationship in group psychotherapy, seems to enhance self-concept (Johnson, 2004). Shechtman’s (1993) work was based on the theory that cohesion and other relationships in groups increase self-esteem because self-evaluation is the inner experience of esteem provided by peers. In support of this theory, both Falloon (1981) and Budman et al. (1989) found increases in self-esteem to be correlated with group cohesion. Shechtman (1993) found that the atmosphere of friendship provided by

children's psychotherapy groups predicted both greater self-esteem and capacity for intimate friendship.

While many studies suggest a positive relationship between cohesion and outcome, other studies suggest a more cautious relationship. Marziali et al. (1997) found that group alliance contributed more to outcome than cohesion did for psychotherapy groups for borderline personality disorder, although both contributed. Woody and Adessky (2002) failed to find any connection between either cohesion or alliance and outcome for socially phobic patients. Likewise, cohesion did not predict outcome in studies of interactional groups (modeled after Yalom, 1995) for substance abuse (Gillaspy et al., 2002) and psychodynamic groups for complicated grief (Kipnes et al., 2002). It is helpful to remember when interpreting these findings that these studies used very different measures and the different findings may be directly related to the differing measures and definitions. Unfortunately, the relationship between these measures is unknown and the studies that have measured the correlations between them show that they often fail to correlate (Dion, 2000), making it almost impossible to conclude much from the findings. Additionally, there may be an interaction with the different formal change theories or patient populations, although there is no empirically discernable pattern to date (Johnson, 2004).

Operationalization: Johnson et al. (2005) chose to measure cohesion using the Cohesion subscale of the Therapeutic Factors Inventory (TFI; Lese & MacNair-Semands, 2000; see Table 1-1). It reflects the individual and subjective perceptions that seem to underlie the relevance and influence of cohesion in group psychotherapy (see Bollen & Hoyle, 1990; Griffith, 1988; Lese & MacNair-Semands, 2000). Content domains of the

TFI include feeling accepted, a sense of belonging, trust, caring, security, and working together. This operationalization fits fairly well with the definition of cohesion given by Johnson (2004) that “[t]he cohesion definition most relevant to therapeutic relationships in psychotherapy groups involves the subjective experiences of individual group members with both the group leader and other group members, on both task and socioemotional levels” (p. 15). This measure also offers convenient psychometric properties and has been used several times in the past few years, unlike most of the other 31 measures. The items from this subscale are believed to load onto a single factor (Lese & MacNair-Semands, 2000) which was also supported by Johnson et al. (2005).

Alliance

Definitions: Alliance was first conceptualized by Freud with regards to individual psychotherapy. It later took on greater significance in the 1970’s when it was identified as one of the common factors associated with effective therapies (Horvath, 1994). Its identification as a common factor started with Rogers’ and Frank’s idea that factors such as empathy, congruence, and unconditional positive regard were enough on their own to cause change. Then in the 1970’s, when researchers noticed that different kinds of therapies achieved similar outcomes, they began to wonder what all these therapies had in common. Therapeutic alliance was one idea that arose (Horvath).

Much like cohesion, divisions have been created in alliance in an effort to give it more clarity. Some psychodynamic theorists (such as Sterba, 1934 and Zetzel, 1956) saw alliance as containing three elements: the *therapeutic alliance*, the *ego alliance*, and the *working alliance* (Horvath, 1994). Bordin (1975) conceptualized alliance in a slightly different way: the affective bond between the therapist and client, and their agreement on

tasks and goals. Since then group psychotherapists have worked to adapt the individual therapy construct of alliance to fit with group models. The simplest way to make this switch is to measure each member's alliance with the leader (see Pinsoff & Cahterall, 1986). However, others have conceptualized the group alliance concept as involving relationships between group members as well as the leader (Trad, 1993; see also Glatzer, 1990; Budman, 1989).

Some of the divisions made to alliance parallel those observed in cohesion. For example, Horvath and Luborsky, (1993) found two aspects of alliance: personal attachments or bonds (much like socioemotional from cohesion) and collaboration or willingness to invest in the therapy process (much like task cohesion). The Penn Helping Alliance Rating Scales (Luborsky, 1976) theorized Type I alliance, the degree to which the therapist is seen as supportive and helpful to the client, and Type II Alliance, a sense of working together in a joint struggle or "we-ness" (p.64). The first sounds very similar to the idea of socioemotional or attachments from cohesion, and "we-ness" is a word commonly used to describe cohesion. Glatzer (1978) divides alliance into member-member alliance (horizontal cohesion) and member-leader alliance (vertical cohesion). MacKenzie (1987) appears to have noticed this overlap between alliance and cohesion when he advocated both a group cohesion measure (Group Cohesiveness Scale; Budman, 1987) and a group climate measure (GCQ) as measures of group alliance. This apparent overlap and ambiguity between construct domains is what prompted the study by Johnson et al. (2005) in which the primary goal was empirically clarifying the relationships between these various group relationship constructs using selected measures.

Outcome findings: Outcome in group psychotherapy has been predicted by several kinds of alliances. Unfortunately, many of these studies use individual therapy instruments of alliance, which only measure the alliance with the therapist. For example, alliances with the therapist have been found to predict positive outcomes in Cognitive Behavior Therapy (CBT) groups for men violent to their partners (Brown & O’Leary, 2001; Taft et al., 2003), CBT groups to reduce risk for coronary heart disease (van Andel et al., 2003), psychoeducational groups for couples with marital distress (Bourgeois, Sabourin, & Wright, 1990), practical/supportive groups of non-psychotic outpatients (Sexton, 1993), and psychodynamic groups for long-term inpatients in Germany (Strauss & Burgmeier-Lohse, 1995).

In contrast, only two studies assessing alliance with the therapist failed to find a relationship between alliance and outcome. McCallum et al. (2002) found that alliance did not predict dropout in supportive or interpretive groups for complicated grief. Likewise, Woody & Adessky (2002) found that neither alliance nor change in alliance predicted outcome for socially phobic patients in CBT groups. The McCallum et al. study differed from those alliance studies predicting outcome in that it only used a four-item measure and predicted dropout instead of outcome. However, the Woody and Adessky study thoroughly assessed both alliance and outcome; thus the reasons for its negative findings are unclear. Nevertheless, the large majority of studies measuring alliances with group leaders using individual therapy instruments found this kind of alliance associated with positive outcomes (Johnson, 2004).

A smaller number of studies have measured alliance using the Pinsoff and Catherall (1986) conceptualization, which involves ideas of alliance and group dynamics

with all relationships still revolved around the leader. Under this model, group members rate their own, other group members', and the whole group's alliance with the leader.

The preliminary evidence for this more group-focused model of alliance is good; all three studies using this perspective found that alliance predicted positive outcomes. Bourgeois et al. (1990) studied psychoeducational groups for distressed married couples, Gillaspay et al. (2002) studied Yalom's interactional group therapy for substance abuse, and Marzialai et al. (1997) studied Interpersonal Group Therapy for borderline personality disorder.

Although there are, as of yet, too few studies on Pinsoff and Catherall's more systematic view of alliance to draw conclusions based on treatment models and populations, it is worth noting that the three existing studies used either an interpersonal/systemic treatment model (Marzialai et al., 1997; Gillaspay et al., 2002) or a population that is naturally more systemic (i.e. couples; Bourgeois et al., 1990).

Only one known study to date has attempted to look at alliance from a truly group-oriented perspective using member-member and member-group relationships. Using adult developmental model groups for nonpsychotic outpatients, Budman et al. (1989) found observer-rated alliances among group members to be strongly related to improved self-esteem and reduced symptom distress. This study used a very systemic definition of alliance, but deliberately excluded member-leader alliance. Perhaps this difference accounts for the very high correlation (above .9) found between alliance in this study and cohesion.

Operationalization: Johnson et al. (2005) used the Working Alliance Inventory (WAI; Harvath & Greenburg, 1989; see Table 1-1) to assess both vertical and horizontal alliances on both relational and task oriented dimensions. This allows measurement of

personal alliances with both the therapist and other group members to be reported from the individual group members' perspectives. The WAI is commonly used in the group psychotherapy literature and is thought to measure the three published factors of affective bond between therapist and client, agreement on the task of therapy, and agreement on the goals of therapy (Horvath & Greenburg). Its focus on alliance with both the therapist and other group members, as well as its good psychometrics and substantial literature to support it, made it a good choice to represent therapeutic alliance within a group.

Empathy

Definitions: Empathy seems to be more clearly defined in the literature than cohesion. As early as 1967, Rogers et al. defined empathy as “[t]he ability of the therapist (to) accurately and sensitively understand experiences and feelings and their meaning to the client...” or in other words, “[i]t is a sensing of the client’s inner world...” (p. 104-105). Psychodynamic, humanistic, and cognitive-behavioral schools of thought have each seen empathy as an important relationship variable in group psychotherapy and each has expressed this idea in its own language.

Psychodynamic theory and practice views empathy as enabling a good therapeutic relationship (see Stone & Gustafson, 1982; Kleinberg, 1999) or as curative on its own (Schain, 1989). It has been described as affective attunement (Schain), empathic attunement (Kleinberg), emotional availability (Shields, 1999) and containment (Dion, 2000).

Humanistic theorists often see empathy and positive regard as curative in and of themselves acting through the mechanism of increasing clients’ self-acceptance and helping clients become more congruent (Trad, 1993). Words such as warmth,

acceptance, positive regard, and democratic atmosphere have been used by humanistic group therapists to describe empathy (Johnson, 2004).

Cognitive behavioral perspective does not distinguish a difference between therapeutic relationships in groups and those in individual therapy relationships (Burlingame, MacKenzie & Strauss, 2004). Consequently, findings from the individual literature are often assumed to apply to CBT groups. Several authors in the CBT individual therapy literature (see Beck et al., 1993; Burns & Auerbach, 1996; Burns & Nolen-Hoeksema, 1992) have described the importance of empathy and warmth in the therapeutic relationship. They tend to define empathy more broadly to include respect, warmth, and genuineness. Empathy is seen to set the stage for client acceptance of more active therapist interventions and is not sufficient to enact change on its own (Burns & Auerbach, 1996).

Outcome findings: Only a few studies have addressed the beneficial effects of empathy in groups; however, there is a general consensus between group therapists and researchers that empathy is important (Braaten, 1990). More often, studies appear to find that a lack of empathy, including feeling discounted, misunderstood, unconnected to other members, or feeling attacked, lead to poorer outcomes (Doxee & Kivlighan, 1994; Karterud, 1988). On the positive side, a review by Raskin (1986) concluded that in client-centered group therapy, positive outcomes correlated with conditions like empathy, genuineness, and warmth.

Operationalization: Johnson et al. (2005) chose to use the Empathy Scale (ES; Burns, 1994; see Table 1-1) because of its brevity. This 10-item scale is thought to

contain one factor and defines empathy from the broader CBT perspective, including referrals to warmth and genuineness, as opposed to pure empathic understanding.

Other populations

The majority of group relationship research has been conducted using out-patient and some non-patient populations. However, a small body of literature is growing on the effects of the group relationship in severely mentally ill populations. Alliance literature contains the most consistent findings. It correlated positively with outcome for psychodynamic groups for long-term inpatients in Germany (Strauss & Burgmeier-Lohse, 1995), for groups of severely and persistently mentally ill inpatients in an inpatient institution (Jumes, 1996), and for long-term psychotherapy groups for inpatients in Germany (McKenzie & Tschuschke, 1993). Studies of cohesion in inpatient groups suggest that the construct of cohesion is important in theory, but only one study was found linking it with outcome. A study involving Turkish inpatients in interactional group therapy found cohesion to be one of the four most valued factors of Yalom's therapeutic factors (Kilic, Özbayrak, Ulusahin & Üstün, 1996). Geczy and Sultenfuss (1995) discuss how hospital procedures of admission and discharge disrupt the group cohesion and propose a way in which administration procedure could minimize this disruption. Tschuschke and Dies (1994) found cohesion to correlate positively with outcome and suggested that it was an important ingredient for effective group process to occur in long-term analytic inpatient groups. Dacey (1989) theorized cohesion to be important to a member's separation from the group in an inpatient setting. Very little research has been done relating group climate and empathy to inpatient populations. McLees et al. (1992) found the GCQ to be an effective measure for following brief 12-

week inpatient groups through developmental stages of the groups. Correale (1999) suggests that while empathy is highly valued by psychotic patients, a truly empathic understanding of the state of mind of these patients is extremely problematic.

Interrelationships

A careful look at the definitions and operationalizations of the four group relationship constructs just reviewed (group climate, cohesion, alliance, and empathy) shows the apparent overlap and lack of clear defining boundaries between these relationship constructs. As has been shown in this review of the literature, each has mixed to positive results predicting outcome in group treatment. Each is also theorized to provide a helpful therapeutic climate, which may lead to other positive therapeutic processes (Johnson, 2004). The few studies that have studied the relationship between these constructs have found them to be highly related to each other and have theorized that they serve the same function in group. Budman et al. (1989) studied the relationship between cohesion and alliance with nonpsychotic outpatients in time-limited psychotherapy groups and found a moderate to high correlation between cohesion and alliance. Cortez-Ison (1997) made a similar finding in psychotherapy groups for women sexually abused as children. Gillaspay et al. (2002) also found a positive correlation between cohesion and alliance with men in a residential group treatment program for substance abuse. Marziali, McCleary, and Munroe-Blum (1997) did a randomized controlled trial of borderline personality disorder clients in time-limited group treatment and found a significant correlation between cohesion and alliance.

Similar results have been found for relations between the other constructs, although there are fewer studies to support it. Roarck and Sarah (1989) studied

members of personal growth, psychotherapy, and DUI groups and found cohesion and empathy to be significantly correlated (.53). Likewise, alliance and empathy were reported to correlate positively (Horvath, 1994) and empathy and group climate were found to positively correlate in outpatient psychotherapy groups in a community mental health center (Phipps & Zastowny, 1988). This high level of interrelatedness and similar theoretical purpose suggests the possible presence of higher-order constructs.

Johnson et al. (2005) clarified the group relationship by proposing a new model of higher-order constructs of the group relationship. They used four measures: the Group Climate Questionnaire (GCQ), Cohesion scale of the Therapeutic Factors Inventory (TFI), Working Alliance Inventory (WAI), and Empathy Scale (ES) to operationalize these four constructs. The TFI and the GCQ represented member-group relationships, while the WAI and the ES represented member-member and member-leader relationships. These measures were administered to 662 participants from 11 different counseling centers and institute groups at the American Group Psychotherapy Association (AGPA). They then tested four different hypothesized models of the group relationship and found that none of them provided a good fit for the data. Then they used Exploratory Factor Analysis (EFA) to create a new model of the group relationship (see Figure 2-1). This model represents an exclusively empirical attempt to describe the group relationship. The three new second order factors (Positive Bonding, Positive Working, and Negative Relationship; see Figure 2-2) have not been previously named, defined or measured.

Figure 2-1

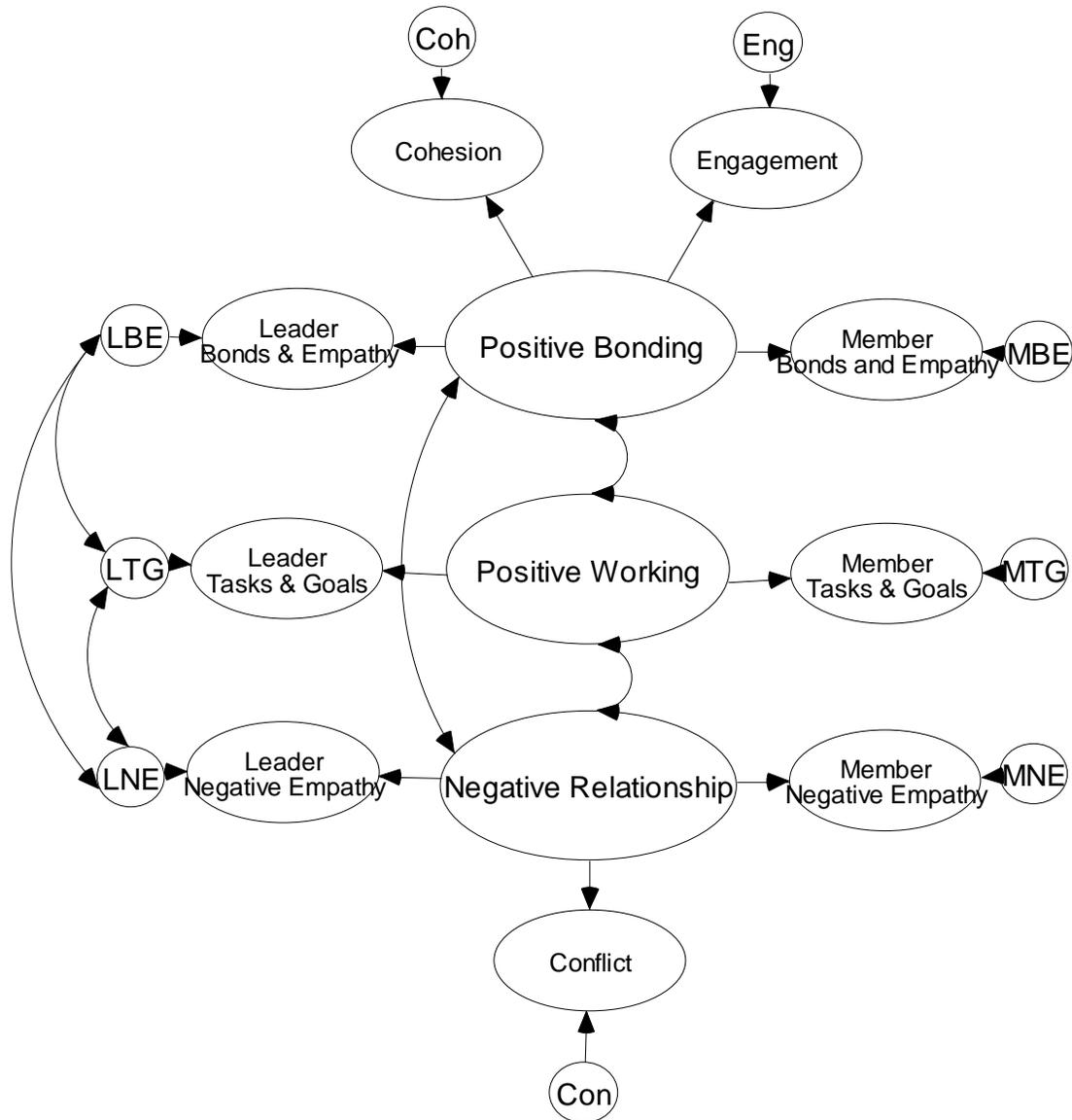
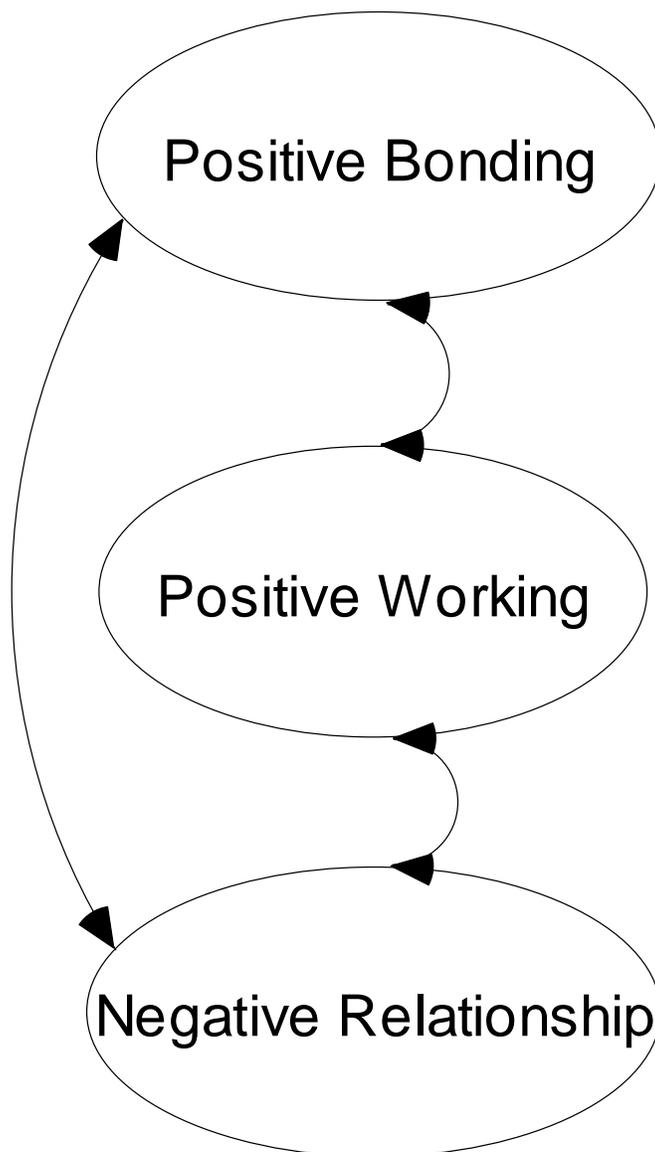
Johnson's Model of Group Therapeutic Relationship

Figure 2-2

Second Order Factors of Johnson's Model



The purpose of this study was to create the Group Questionnaire (GQ) based off of the Johnson et al. (2005) model. The creation of the GQ relied heavily on both empirical results and clinical judgment to create a measure that is empirically sound, clinically relevant, and practically efficient. The creation of the GQ took place in two steps: 1) Using empirical data and clinical judgment, a team of experienced group researchers and clinicians pared down and adapted the 60 items used by Johnson et al. to create the GQ as a measure of the group relationship including Positive Bonding, Positive Working, and Negative Relationship, 2) the GQ was then tested and revised using Confirmatory Factor Analysis (CFA) and data from three populations: university counseling centers (UCC), non-patient American Group Psychotherapy Association (AGPA), and inpatient Utah State Hospital (USH).

Hypotheses included:

1. *The factor structure including Positive Bonding, Positive Working, and Negative Relationship as identified by Johnson et al. will provide a good fit for the total data generated using the Group Questionnaire.*
2. *Subscales representing positive aspects of the therapeutic relationship will be significantly positively correlated with one another and will be significantly negatively correlated with subscales reflecting negative aspects of the therapeutic relationship as specified by the model. Subscales representing negative aspects of the group relationship will also be significantly correlated with each other as specified by the model.*

3. *Model divisions between member-member, member-leader, and member-group will provide a good fit to the data. Specifically, the model will not fit when these divisions are removed.*
4. *The model will provide good fit to the data from each of the three populations, (university counseling center, severely mentally ill, and non-clinical) when analyzed separately.*
5. *Differences in response styles will be seen between the three populations. Specifically, the factor means for Positive Bonding and Working will be lower and Negative Relationship will be higher for data from the severely mentally ill sample than the university counseling center and non-clinical samples.*
6. *Measurement refinement procedures, in line with accepted structural equation model (SEM) practice (Byrne, 2001), will improve model fit resulting in minor modifications to the Group Questionnaire and underlying model.*

Chapter 3

Methods

This study was conducted in two parts. In part 1, the Group Questionnaire (GQ) was created; in part 2, the GQ was tested and refined. The methods and results section are organized sequentially into these two parts. Thus, Chapter 3 presents two sets of participants and two sets of procedures.

Part 1: GQ Creation

Participants

The GQ was created by a team of three experienced group clinician/researchers and an advanced doctoral student. Cumulatively, their post doctoral clinical experience totaled 65 years. There were three males and one female. All four team members were Caucasian. Their ages at the time of the GQ's creation ranged from 21 to 54, with a mean of 40. All members were affiliated with Brigham Young University and represented the departments of Clinical Psychology, Counseling Psychology, and the Counseling and Career Center.

Procedures

Creation of the GQ relied heavily on both empirical data and the clinical judgment of the team which created it. The resulting questionnaire, entitled the Group Questionnaire (GQ), is a 40-item, 7-point, Likert-type, self-report measure of the therapeutic relationship in group (see Appendix A). The GQ was designed to measure the newly defined group relationship model identified by Johnson et al. (2005) (Figure 2-1) in a shorter and more clinically relevant manner. Johnson originally measured this

model using 60 empirically identified questions from the Working Alliance Inventory (WAI; Horvath & Greenberg, 1989), Group Climate Questionnaire-Short Form (GCQ; MacKenzie, 1983), the Cohesion subscale from the Therapeutic Factors Inventory (TFI; Lese & Mac Nair-Semands, 2000), and the Empathy Scale (ES; Burns, 1994). No clinical judgment was used in selecting these 60 items. This resulted in a set of 60 questions that were at times redundant, inconsistently worded, and lengthy.

In creating the GQ, empirical analyses were conducted first, followed by a thorough review of proposed items using clinical judgment. The described procedures follow in that order. The process began with a review of Johnson's statistical analyses of the original 60 questions used to create this model. Johnson provided access to the data set of 662 participants used in the Johnson et al. (2005) study to regenerate selected analyses and statistical tables needed to provide specific information on individual items. The goal of the empirical process was to identify from an empirical perspective which items from each first order factor were the strongest and which items could be dropped due to redundancy. Tables of factor loadings (aka. Standardized regression weights) and inter-item correlations were generated and evaluated to determine the strength and uniqueness of individual items. In general, items with smaller factor loadings or high correlations with other items were identified for consideration of being dropped.

Items associated with the second order factor Positive Bonding tended to have higher factor loading, than did items associated with the second order factors Positive Working or Negative Relationship. Due to the plethora of items with strong loading associated with the second order factor Positive Bonding, items with loadings less than .70 were considered for elimination. Factor loadings of items associated with second

order factors Positive Working Relationship and Negative Relationship were not as strong over all, and there were far fewer items associated with these factors. Items with the strongest factor loadings were identified, but no cut-off was used. It was not proposed that any items be dropped based on empirical analysis because of the low number of items. This was done in order to maintain factor stability of these first order factors by including a minimum of three or more items.

Once the item loadings had been evaluated, an inter-item correlation matrix was constructed of the Positive Bonding items. A cutoff of .65 was chosen to designate items which correlated highly. It is more standard to use a cut off .7 or .8 in defining high levels of correlation; however, in order to identify a greater number of significantly correlated items on the Positive Bonding factor, this cutoff was lowered. These high inter-item correlations were taken into consideration as evidence of redundancy as the new measure was being constructed.

The GQ creation team met and discussed the content domain of each subscale as reflected by the items it contained. A clinically relevant definition was provided to best describe each subscale after reviewing the items that comprised the scale. Items were considered for inclusion based on their clinical relevance to these refined construct definitions and their empirical support. Using this process, items were selected, eliminated, reworded, and combined to create the Group Questionnaire. A thorough description of the GQ is provided in part 1 of the results.

Part 2: GQ Testing and Refinement

Participants

Participants for this study included 486 individuals from three group populations: University Counseling Center (UCC), American Group Psychotherapy Association (AGPA), and Utah State Hospital (USH). Each population reflects a very different group setting: out-patient, inpatient, and non-patient respectively. The UCC patients and AGPA institute member populations were specifically chosen for this study because they were used previously by Johnson et al. (2005). They also represent both out-patient and non-patient samples, both of which have been used extensively for the past study of these group relationship constructs, as can be seen from the literature review. The third setting of inpatients was chosen to potentially broaden the scope of the GQ. It was encouraging that Johnson et al. (2005) found the same factor structure for both the UCC and AGPA groups, and it was hypothesized that the same factor structure would adequately fit the inpatient groups at the USH. The three populations are described briefly below. More demographic information is provided in the results section.

The University Counseling Center (UCC) data was contributed by 207 participants from two sites: the Brigham Young University Counseling and Career Center, and the University of North Carolina-Charlotte Counseling Center. These UCC groups were led by licensed, doctoral-level psychologists and often a student or intern co-leader. Groups were both psychoeducational and process focused. Group members were typically full-time university students referred to out-patient group therapy for a variety of problems common to a college population such as anxiety, depression, adjustment,

relationship, and sexual issues. They were typically young adults between the ages of 17 and 25.

The second population included 160 participants from the 2-day institute groups at the 2008 annual meeting of the American Group Psychotherapy Association (AGPA). Participants in these groups form a non-clinical group population, typically professionals (mostly clinicians) in the field of group psychotherapy and a few graduate students. The AGPA groups were process-oriented groups designed to furnish experiential training and growth in the field of group psychotherapy. Participants of these groups register ahead of time, pay a fee to the conference, and must be professionals or students in the area of group psychotherapy. These participants are generally more knowledgeable about group psychotherapy than the average outpatient client. Tuschuschke and Green (2002) found GCQ scores of AGPA (nonclinical) training groups to be higher on all three subscales (Engagement, Conflict, Avoidance) than those of outpatient clinical groups.

Participants from the Utah State Hospital (USH) comprised the third population and included 118 group participants. The USH is an inpatient hospital for severely mentally ill patients. Participants in this population were members of inpatient psychiatric groups. Patients at the USH are typically diagnosed with some combination of severe psychotic, bipolar, and/or affective disorders. Groups are structured, psychoeducational groups focused on self-care and social skills. Groups are typically led by nurses who have been trained using a manual on how to conduct these groups.

Procedures

Administrative approval and support for each population was obtained first and the study was presented on an institutional level. This greatly improved the willingness

of the group leaders to support the study. Group leaders from each of the three populations were given a brief verbal and written explanation of the study and asked if they would be willing to administer the GQ once to their group members. The GQ was administered during the last ten minutes of a regular group session for the UCC and USH populations and at the end of the first day's institute meeting for the AGPA population. The UCC and AGPA population group leaders administered the questionnaires without the help of researchers in order to increase confidentiality. The USH participants required extra assistance at times, and a researcher was available to assist participants when needed. A special version of the GQ was also created for the USH population using simplified language in the instructions, a non-parallel item format, and larger font. The items on the GQ were not changed. Cashews were used as an incentive to participate at the USH, but no incentive was needed for the other populations. Group leaders in the UCC and AGPA populations were also asked to complete a form describing their group, its population, size, theoretical orientation, number of sessions the group has met, and whether it followed a process or psychoeducational format.

Yalom (1995) suggests that group process progresses through developmental stages and the group cohesion and tension are present in the middle stages of group process. Data for this study was collected during the middle or working stage of the group. The UCC and USH population groups were required to have met at least three times and not be conducting their final group session in order to be eligible to participate. AGPA groups completed the GQ at the end of the first day of the 2-day institute group, after approximately 8 hours of group time. These parameters were designed to allow

time for a relationship to form between the group, its leader(s), and its members. Similar constraints were used by Johnson et al. (2005).

Statistical Analysis

Exploratory factor analysis (EFA) was used by Johnson et al. (2005) in explaining the previously unknown links between the observed and latent variables. A confirmatory factor analysis (CFA) is appropriately used when the researcher has some knowledge of the underlying latent variable structure (Byrne, 2001). Since Johnson has already specified a model of the latent variable structure, the data collected in this study was analyzed using a CFA to test its plausibility. The primary task in this model-testing procedure is to determine the goodness of fit between the hypothesized model and the sample data. The data was analyzed as a whole first and then broken down by population to test if the model demonstrates goodness of fit for each of the three populations used in this study.

Introduction to structural equation modeling. Structural equation modeling is a special form of regression in which relationships among variables are represented pictorially by a series of structural (i.e. regression) equations. The pictorial representation enables a clearer conceptualization of the theory under study (Byrne, 2001). Byrne explains:

The hypothesized model can then be tested statistically in a simultaneous analysis of the entire system of variables to determine the extent to which it is consistent with the data. If the goodness of fit is adequate, the model argues for the plausibility of postulated relationships among variables; if it is inadequate, the tenability of such relationships is rejected. (p.3)

Byrne further explains that unlike the older generation of multivariate procedures, structural equations modeling (SEM) provides researchers with more information that is particularly useful in this study. First, SEM allows for a confirmatory approach to data analysis which can be used for hypothesis testing. The majority of older multivariate procedures are essentially descriptive in nature, such as exploratory factor analysis, but it is difficult to employ them in hypothesis testing. Second, SEM provides explicit estimates of the error variance parameters, while traditional multivariate procedures are incapable of either assessing or correcting for measurement error. Third, SEM procedures allow both the unobserved and observed variables to be incorporated, and the graphical representation of the statistical relationships facilitates easy communication of model information.

In SEM, abstract theoretical constructs are referred to as latent variables or factors. In this study, Positive Bonding, Positive Working, and Negative Relationship as well as member-member, member-leader, and member group relationships are theoretical constructs that cannot be observed directly; thus, they cannot be measured directly. The latent variables must first be operationally defined in terms of the behavior they are believed to represent. In this manner, the unobserved variable can be linked to one that is observable and therefore measurable (Byrne, 2001). This allows for the indirect measurement of the unobserved variable or underlying construct by using the observed and measured variables as indicators.

The factor-analytic model, which includes EFA and CFA, focuses exclusively on how and to what extent the observed variables are linked to their underlying latent factors. Byrne specifies that the observed variables (i.e. factor loadings) are of primary

interest. They measure the strengths of the regression paths from the factors to the observed variables, thus reflecting the extent to which the observed variables are generated by the underlying latent constructs. Factor-analytic models explain how the observed and latent variables are related to one another (Byrne, 2001). Exploratory factor analysis was used appropriately by Johnson et al. (2005) because the relationships between the observed and latent variables were unknown or uncertain. In the present study, confirmatory factor analysis was used because there was some knowledge of the underlying latent variable structure based on Johnson's research. This follows the typical pattern of factor-analytic research where a researcher postulates a statistical model based on their knowledge of the related theory, on empirical research, or, as in this study, on a combination of both. After the model is specified, the researcher uses sample data comprised of all the observed variables in the model to test its plausibility.

The goal of this model-testing procedure is to determine the goodness of fit between the sample data and the hypothesized model. The residual refers to the difference between the model and the data. Fit statistics reflect the amounts of variance and covariance in the data that are accounted for by the model vs. the residual. Goodness of fit indices provide a way to evaluate how well the hypothetical model agrees with the empirical data. Over 30 goodness of fit statistics exist, but there is no "universally accepted" set. Multiple indicators of "success" are desirable (Breckler, 1990) and their use should be supported with a rationale. Table 3-1 provides a summary of the goodness of fit indices used in this study and standards for how to evaluate the adequacy of fit on each index.

Chi-square (χ^2) is one of the most enduring and widely used indicators of fit from the early days of SEM. A good fit requires the χ^2 test for the model to be approximately twice its degrees of freedom or smaller and have a significant p value. Although traditional in its use, Hu and Bentler (1995 Byrne) find the criterion of a non-significant χ^2 too stringent and report that it may easily result in high type-II error rates. Therefore, while the χ^2 statistic will be reported and used for evaluation of goodness of fit, it will be considered one of the least important of the fit indices for evaluating model fit.

Table 3-1**Specification of Fit Indices**

Fit Index	Poor Fit	Moderate Fit	Good Fit
χ^2	Significant p. value Or $\chi^2 >$ Twice the DF	-NA-	Non-significant p. value Or $\chi^2 <$ Twice the DF
CFI	Below .90	.90-.95	.95 or above
RMSEA	Greater than .08	.06-.08	.05 or below
PCLOSE	Below .05	-NA-	Above .05

Three other fit indices were used in this study to evaluate goodness of fit. Each is based on slightly different calculations, and together they provide a more robust measure of fit. They are the same indices used by Johnson et al. (2005), and will enable easy comparison between results. These indices are summarized in Table 3-1 and include the Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and the Test of Close Fit p-value. A long-standing rule of thumb for determining whether or not a SEM has a good fit based on the CFI fit indices has been a cutoff point of “.90 or above” (Byrne, 2001). However, some (i.e., Hu & Bentler, 1995) have suggested that

this guideline for good fit has been too forgiving and frequently resulted in poor-fitting models being classified as a good fit. On the basis of these cautions, a criterion of .95 or above was used to describe a *good fit* and the .90 to .95 range for a *moderate fit*.

The CFA model is often referred to as a *measurement model* since it defines and measures the relationships between the observed and unobserved variables. Unlike the *structural model*, which defines the relations among unobserved variables, the *measurement model* links the scores on measurement instruments (i.e. observed variables) to the underlying constructs they are designed to measure.

AMOS (Arbuckle & Wothke, 1999) portrays structural equation models graphically using particular configurations of four geometric symbols – an ellipse, a rectangle, a single-headed arrow, and a double-headed arrow. Ellipses (or circles, as they sometimes appear) represent unobserved latent factors. Rectangles, or squares, represent observed variables. Single headed arrows represent the effect of one variable on another, specifically the path coefficient for regression of an observed variable onto an unobserved variable or factor; numeric values are printed next to the single-headed arrows representing the factor loadings, also known as regression weights, of the items onto the factors. Double-headed arrows represent correlations or covariances between two variables; the numerical values printed next to the covariance symbols represent the correlations between two items. Error terms, which represent measurement error both random and unique, appear as small bubbles next to each observed variable. An error covariance is represented by a double-headed arrow connecting two error terms and represents shared error variance. It is used when two observed variables within a subscale are more correlated with one another than is explained by their common

association to the underlying factor or subscale and they are allowed to covary (Byrne, 2001).

Missing data is a common occurrence within social science research, and group research is no exception. It is not uncommon for group members to skip or miss certain questions on a questionnaire for a variety of reasons, presenting a challenge for any group study. AMOS allows for a full information maximum likelihood estimation (FIML) to be used to replace missing data, allowing the entire data set to be used. The FIML has several advantages over both the listwise and pairwise deletion approaches, including consistency, efficiency, and reduction of bias (Byrne, 2001). In cases where values are missing at random, FIML estimates are asymptotically unbiased. In cases where values are not missing at random, FIML estimates will exhibit the least bias of any current estimation technique, including providing reasonable estimates of standard errors (Byrnes, 2001; Duncan, Duncan, Strycker, Li, & Alpert, 1999; Enders & Bandalos, 2001). Stated simply, FIML uses patterns in the available data to estimate the most likely values for the missing data. This method was employed for all SEM statistics in this study.

Hypothesis 1

Hypothesis 1 proposed that the higher order factor structure of Positive Bonding, Positive Working, and Negative Relationship identified by Johnson et al. would provide a good fit for the total data generated using the Group Questionnaire. This hypothesis was tested using the *measurement model*, which, again, looks at the relationship between individual items to their scale and measures the extent to which all items included in the analysis are good representations of the underlying theoretical constructs they purport to

measure. This hypothesis was tested using a confirmatory factor analysis of the total data sample to test if Johnson's model of group relationship provided a good fit to the data from the GQ. The tested structure of the GQ items and subscales is represented in Figure 3-1.

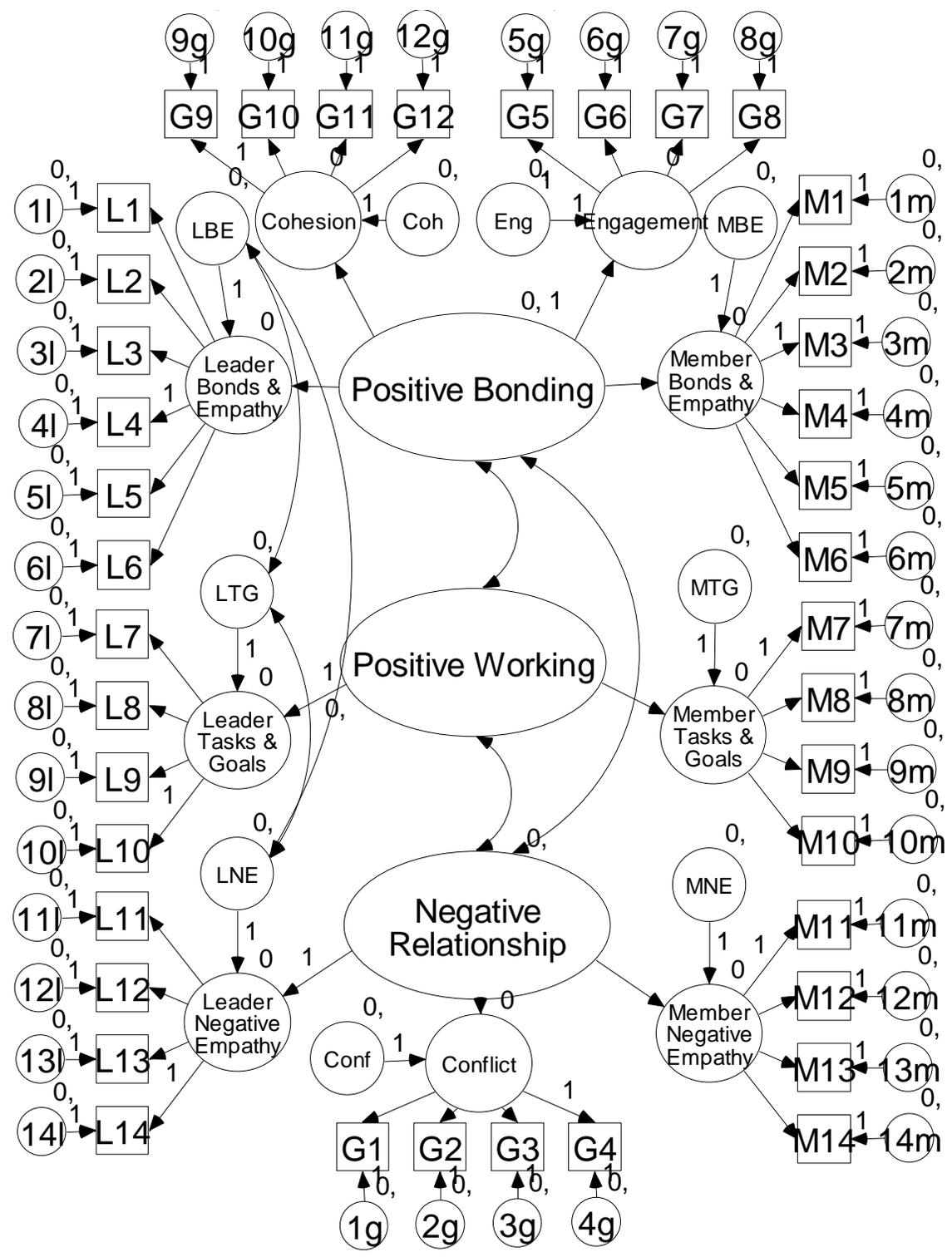
Hypothesis 2

Hypothesis 2 proposed that the first order factors representing positive aspects of the therapeutic relationship, those associated with Positive Bonding and Positive Working, would be significantly positively correlated with their associated second order factor as specified by the model. First order factors representing negative aspects of the group relationship would be significantly positively correlated with Negative Relationship as specified by the model. The second order factors were hypothesized to correlate with each other, with Positive Bonding and Positive Working being significantly positively correlated to each other and both negatively correlated with Negative Relationship. This hypothesis was tested by calculating correlations and regression weights among all the subscales and testing them for statistical significance using a 1-tailed test. Significance for this study was set at .05.

Hypothesis 3

Hypothesis 3 proposed that the model divisions between levels of the group relationship (member-member, member-leader, and member-group) would provide a good fit to the data. Specifically, the GQ data was expected to fit poorly when the first order factors, which represented the level of relationship in the model, were removed and the data was related directly to the second order factors (Positive Bonding, Positive Working, and Negative Relationship) without accounting for level of relationship. The

Figure 3-1: GQ Structure Tested in Hypothesis 1



relationship levels have been theoretically important in the group literature, and confirmation of this hypothesis would provide some empirical support for this theoretical assumption. This hypothesis was tested by relating the items directly to their associated second order factors and excluding the first order factors as shown in Figure 3-2.

Goodness of fit was determined using the before-mentioned fit-indices and standards for interpretation.

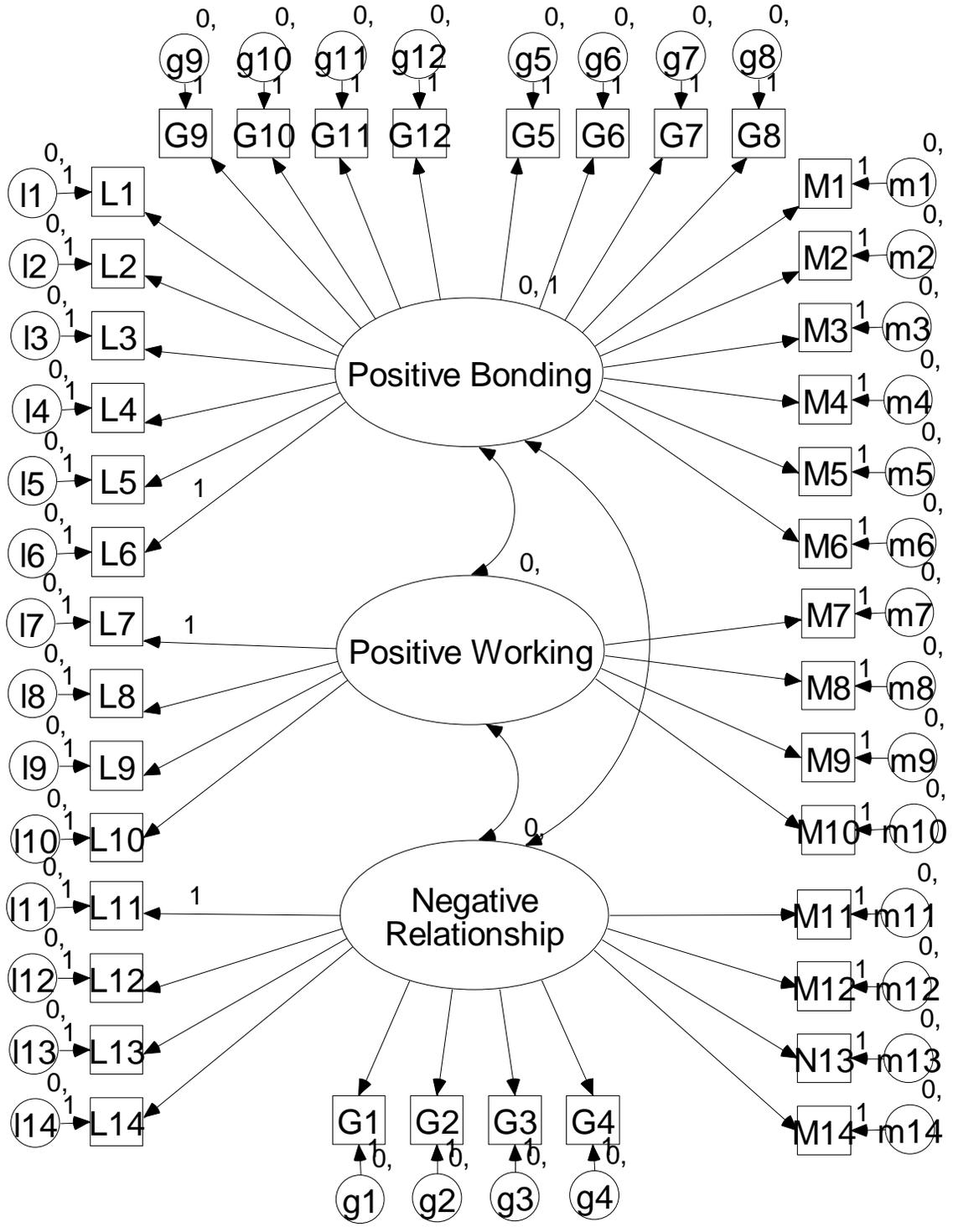
Hypothesis 4

Hypothesis 4 proposed that Johnson's model would provide good fit to the data from each of the three populations (UCC, AGPA, and USH) when analyzed separately. For Hypotheses 1–3, the entire data set was used in the analyses. Hypothesis 4 was tested by performing a separate CFA on each of the three sample populations. Goodness of fit statistics for each were reported separately and represent the appropriateness of the scale for each population. Again, Figure 3-1 represents the model tested.

Hypothesis 5

Hypothesis 5 stated that differences in response styles would be seen between the three populations. Specifically, the factor means for Positive Bonding and Working would be lower and Negative Relationship would be higher for data from the severely mentally ill sample than the university counseling center and non-clinical samples. A three by three multivariate analysis of variance (MANOVA) was conducted to test for statistically significant differences between the total scores for the 3 second order factor subscales for each of the three populations. Subscale means and standard deviations for Bonding, Working, and Negative Relationship were reported for each population.

Figure 3-2: GQ Model without 1st Order Factors



Test reliability is an important prerequisite to establishing a test's validity for any population or model. Thus, Cronbach's Alpha was used to calculate the test reliability for the measure as a whole using data from all three populations. Reliability estimates were also calculated for each of the populations separately and for each of the subscales.

Hypothesis 6

Hypothesis 6 proposed that model refinement procedures, in line with accepted SEM practice (Byrne, 2001), would improve model fit for the total data sample, resulting in minor modifications to the Group Questionnaire. Due to the modifications made to the GQ reducing it from the original 60 items to present 40, it was anticipated that the goodness of fit might be less strong than the fit found by Johnson and might have the potential of being improved by refining the measure using the new data set. The rationale behind using model respecification procedures in SEM was to locate the sources of the misfit in the model and to determine a model that better described the sample data. According to some researchers, this procedure can provide a more pure measure of the construct in question. It may be driven by theory or data, but the ultimate goal is to create a model that is theoretically meaningful and statistically sound (Byrne, 2001).

The following model respecification procedures were used, and they follow standard SEM practices as outlined by Byrne (2001). They were designed to minimize the likelihood that chance variations in the data overly influenced the model. The first step was to identify items with factor loadings on to the intended subscale which were not statistically significant. These items were then systematically excluded one item at a time from the analysis; fit statistics for the entire model were recalculated after each change, since a change in one sector of the model can effect other areas. Using this iterative

process minimized the likelihood of modeling random variation. The second step was to allow item errors to covary one at a time when there was theoretical support for doing so. AMOS provides a set of indicators termed modification indices that indicate how much the fit of the model would change by adding each potential error covariance. The covariance is considered when these modification indices are above 4.0. Higher error covariances are expected when a high-degree overlap exists in the item content. Items with a high degree of covariance were allowed to covary, but only if they 1) made theoretical sense, 2) made a worth while increase to the model fit to justify the increased model complexity, and 3) the covariance was consistent across parallel domains of the measure (i.e. member-member and member-leader domains). For example, two items would not be allowed to covary if they were associated in the member-leader bonds, but not the member-member bonds. These steps were done in an effort to reduce the chances of modeling random variation unique to the data set.

This set of procedures is designed to provide the best possible model for the data while still accounting for the constraints of theory and statistical consistency across similar models and is a common standard in SEM (see Byrne, 2001). It is conceded that deciding what makes “theoretical sense” and what level of complexity is “worthwhile” in a model ultimately are matters of judgment and are similar to some of those made in very complex multiple regression models.

Once refinements on the GQ were complete, Hypotheses 1–5 were retested using the refined GQ measurement model. Results for the refined GQ are presented following the results from the original GQ measurement model.

Chapter 4

Results

Results for the study are reported in two parts. Part 1 reports the results of the GQ Creation and describes the GQ. Part 2 reports the testing and refinement of the GQ.

Part 1: Creation of the GQ

Creation of the GQ was successful and resulted in a 40-item, 7-point, Likert-type, self-report measure of the group relationship. Using both clinical judgment and empirical support, the 60 items originally used by Johnson et al. (2005) were selected and refined down to 40 items. Definitions were also created for each of the 3 second order factors as shown in table 4-1. First order factors are not defined individually, but are thought to reflect the definition of their associated second order factor from the member-leader, member-member, or group as a whole relationship dimension. The following are the definitions for the three second order factors. Table 4-2 lists the items comprising the GQ. The item number corresponds to the item numbers on the models shown throughout the study (see Figure 3-1). Table 4-3 lists the items dropped or changed from the original 60 used by Johnson.

Part 2: Testing and Refinement

Demographics for the UCC population revealed a mean therapist age of 42 (range 22-65 with 65% (13) of the therapists being male and 35% (7) female. Self-reported ethnicity of the therapists was predominantly Caucasian (95% or 19 of 20) and 5% (1 of 20) being Asian. Most groups had a primary and secondary group leader. The primary group leaders were

Table 4-1: Second Order Factor Definitions

<i>Positive Bonding</i>	The sense of belonging or attraction that a member has to the group, its members, and its leader(s) that creates a positive atmosphere which allows members to feel genuinely understood and appreciated.
<i>Positive Working</i>	The ability of the group to agree upon and work toward treatment goals in an effective manner.
<i>Negative Relationship</i>	A lack of trust, genuineness, and understanding as well as friction and distance that might exist between the group, its members, or its leaders.

Table 4-2: GQ Items

<u>Item Number</u>	<u>Item</u>
<u>Positive Bonding - Leader Bonds and Empathy</u>	
L1	I felt that I could trust the group leaders during today's session.
L2	The group leaders understood what I said during today's session.
L3	I am confident in the group leaders' ability to help me.
L4	The group leaders and I respect each other.
L5	I feel the group leaders care about me even when I do things that they do not approve of.
L6	The group leaders were friendly and warm toward me.
<u>Positive Working – Leader Tasks and Goals</u>	
L7	The group leaders and I agree about the things I will need to do in therapy.
L8	The group leaders and I are working together toward mutually agreed upon goals.
L9	The group leaders and I agree on what is important to work on.
L10	The group leaders and I established a good understanding of the kind of changes that would be good for me.
<u>Negative Relationship – Leader Negative Empathy</u>	
L11	Sometimes the group leaders did not seem to be completely genuine.
L12	The group leaders did not always seem to care about me.
L13	The group leaders did not always understand the way I felt inside.
L14	The group leaders acted condescending and talked down to me.
<u>Positive Bonding – Member Bonds and Empathy</u>	
M1	I felt that I could trust the other group members during today's session.
M2	The other group members understood what I said during today's session.
M3	I am confident in the other group members' ability to help me.
M4	The other group members and I respect each other.
M5	I feel the other group members care about me even when I do things that they do not approve of.
M6	The other group members were friendly and warm toward me.

Table 4-2: GQ Items Continued

<u>Item Number</u>	<u>Item</u>
<u>Positive Working –Member Tasks and Goals Subscale</u>	
M7	The other group members and I agree about the things I will need to do in therapy.
M8	The other group members and I are working together toward mutually agreed upon goals.
M9	The other group members and I agree on what is important to work on.
M10	The other group members and I established a good understanding of the kind of changes that would be good for me.
<u>Negative Relationship – Member Negative Empathy</u>	
M11	Sometimes the other group members did not seem to be completely genuine.
M12	The other group members did not always seem to care about me.
M13	The other group members did not always understand the way I felt inside.
M14	The other group members acted condescending and talked down to me.
<u>Negative Relationship – Conflict Subscale</u>	
G1	There was friction and anger between members
G2	The members were distant and withdrawn from each other.
G3	The members rejected and distrusted each other.
G4	There was tension and anxiety between the members.
<u>Positive Bonding– Engagement</u>	
G5	The members liked and cared about each other.
G6	The members tried to understand why they do the things they do, tried to reason it out.
G7	The members felt what was happening was important and there was a sense of participation.
G8	The members revealed sensitive personal information or feelings.
<u>Positive Bonding – Cohesion</u>	
G9	The group members accept one another
G10	We cooperate and work together in group.
G11	There is a sense of loss when a member of our group is absent
G12	Even though we have differences, our group feels secure to me.

Table 4-3: Items Dropped or Reworded to Create the GQ

Leader Positive Bonds and Empathy

My relationships with the group leaders are very important to me.
 The group leaders were sympathetic and concerned about me.
 I believe the group leaders like me.
 I feel group leaders appreciate me.
 The group leaders and I trust one another.
 I believe the group leaders are genuinely concerned for my welfare.
 The group leaders felt I was worth while.
 The group leaders and I understand each other.

Member Positive Bonds and Empathy

My relationships with the other group members are very important to me.
 The other group members were sympathetic and concerned about me.
 I believe the other group members like me.
 I feel the other group members appreciate me.
 The other group members and I trust one another.
 I believe the other group members are genuinely concerned for my welfare.
 The other group members felt I was worthwhile.
 The group members and I understand each other.

Cohesion

I feel accepted by the group
 I feel a sense of belonging in this group
 I feel good about being a part of this group.
 Even though others may disagree with me sometimes, I feel accepted in group.

Engagement

The members liked and cared about each other.

Member Negative Empathy

The other group members pretended to like me more than they really do.

Leader Negative Empathy

The group leaders pretended to like me more than they really do.

Conflict

The members appeared tense and anxious.

all doctoral-level clinicians. The secondary group leaders were doctoral students or other doctoral-level clinicians. Of the group leaders, 70% (14) had doctoral degrees, 10% (2) had master's degrees, and 20% (4) were doctoral students. They had a mean number of years of clinical experience of 17 (range 3–38). Demographic information on the 207 UCC participants was not gathered for this study. However, previous data from this same population gathered by Johnson et al. (2005) found that 64% of the participants were female and 36% male with an average age of 24 (range 17–50, SD=5). Like the therapists, most participants reported ethnicity as Caucasian (78%) while minority representation was 5% Hispanic, 2% African American, 9% Asian/Pacific Islander, 5% were “Other”, and 1% did not respond to the question about ethnicity. Religious affiliation was not assessed; however, Brigham Young University is a religiously affiliated university and typically over 95% of the students identify as Latter-Day Saints.

Demographics for the AGPA group leaders revealed that 65% (9 of 14) were female and 35% (5 of 14) were male. Self-reported ethnicities of the leaders were predominantly Caucasian (79% or 11), with minority representations including Jewish (14% or 2) and Eskimo (7% or 1). The theoretical orientations of the group leaders can all be broadly grouped as psychodynamic theories. The group leaders had a mean number of years of group experience of 30 (range 10–56). They had mean age of 64 with a range of 50–85. In general they were significantly older and more experienced than their UCC counterparts. The leaders reported professional degrees including Ph. D. (50% or 7), MSW/LCSW (53% or 5), and M.D. (14% or 2). Demographic information on the AGPA participants was not gathered for this study. However, previous data from

this same population gathered by Johnson et al. (2005) found that group participants had an average age of 50 (range 21–83, SD = 12), 57% female, 31% male, and 12% not responding. Self-reported ethnicities were predominantly Caucasian (65%). Reported minorities included 3% “European,” and 3% “Jewish”. “Other” ethnicities made up 6% of the participants and 23% did not respond to the question.

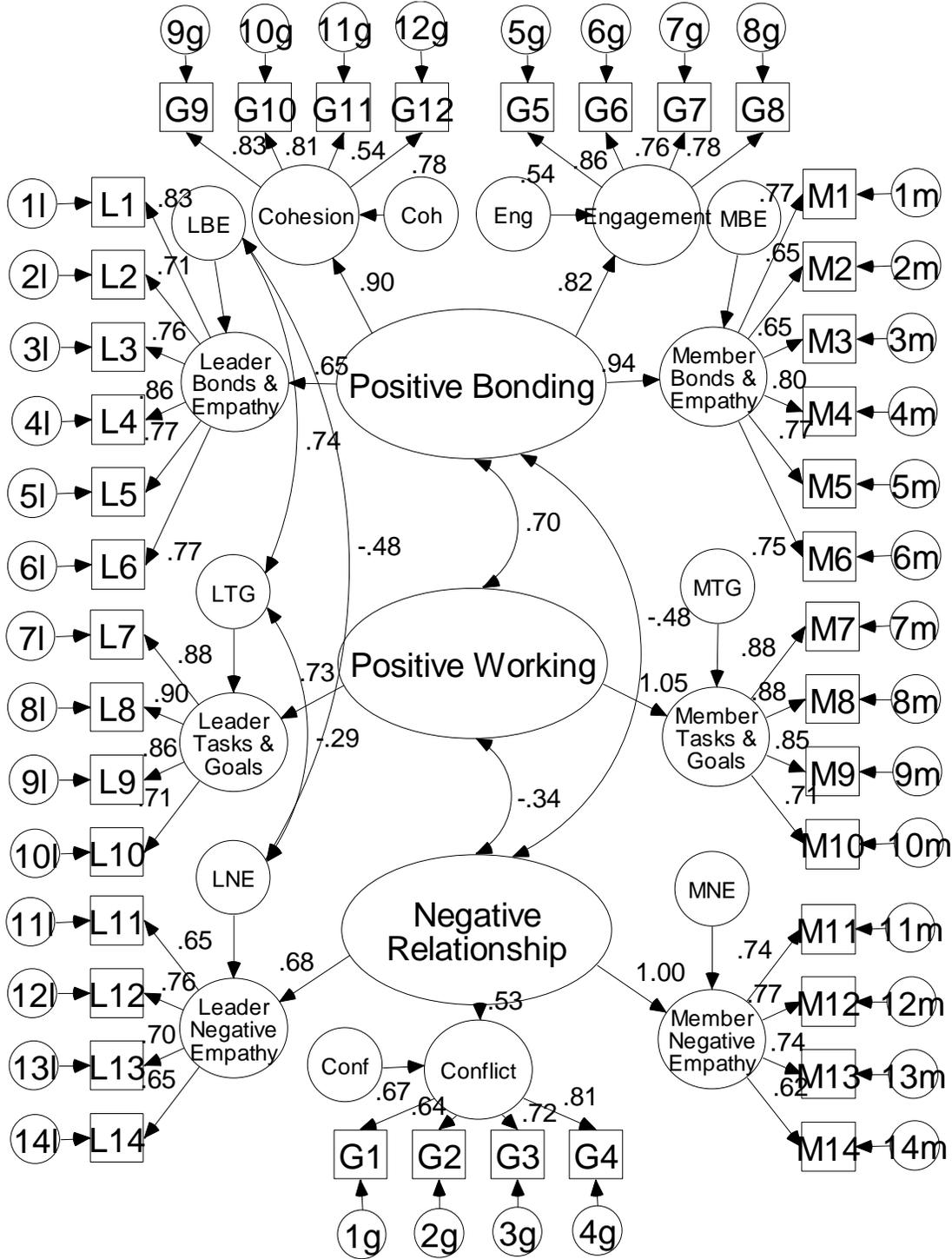
Demographic information for the inpatient population from the USH was not gathered for this study. However, Burlingame et al. (2006) also used a large sample from the USH and found that for the adult psychiatric sample there was a mean age of 39 (range 22–89, SD=15) with 53% male and 47% female. The ethnicities reported were 89% Caucasian, 1% African American, 5% Hispanic, 2% American Indian, 1% Asian /Pacific Islander, and 1% were unreported. Demographic information for the group leaders was not available; however, they were all bachelors level nurses, trained to use structure group protocols.

Tests of Hypothesis 1

Results of the confirmatory factor analysis for the Group Questionnaire (GQ) using the total data from all three populations are described below. As stated previously, this confirmatory factor analysis tests whether all items and subscales provide a good measure of the relevant theoretical constructs in these samples.

Results from the CFA of the GQ provided an inadequate fit to the data (Chi-Squared: 2555, df = 725, Comparative Fit index: .849, RMSEA: .072, p for Test of Close Fit: .000; see figure 4-1) and Hypothesis 1 was unsupported. The Chi-squared value was significant and much more than twice its degrees of freedom, CFI was well below the .95

Figure 4-1: Hypothesis 1 – GQ Confirmatory Factor Analysis



Fit Statistics
 Chi-Squared = 2555, df = 725
 Comparative Fit index = .849
 RMSEA = .072
 P for Test of Close Fit = .000

level, RMSEA was well above .05, and the p for Test of Close Fit was significant (see figure 4-1). The 40-item GQ measurement model does not adequately represent the relevant theoretical constructs in these samples.

Tests of Hypothesis 2

Hypothesis 2 was supported. All subscales representing positive and negative aspects of the group relationship model were positively and negatively correlated as predicted by the model. Specifically, the first order factors cohesion, engagement, Leader Positive Bonds, and Member Positive Bonds were each significantly positively correlated with their second order factor Positive Bonds. Likewise, the first order factors Leader Positive Work, and Member Positive Work were each significantly positively correlated with their second order factor Positive Work as specified by the model. The first order factors Leader Negative Relationship, Member Negative Relationship, and conflict were also significantly positively related to their second order factor Negative Relationship. Table 4-4 shows the standardized regression weights of each second order factor to its associated first order factor and the level of significance. This table also shows the correlations of the second order factors to each other and the associated significance levels. As predicted, both Positive Bonding and Positive Working were significantly positively correlated. They were each negatively correlated with negative. These correlations and regression weights can also be seen on Figure 4-4 along side the arrows representing the relationships between factors.

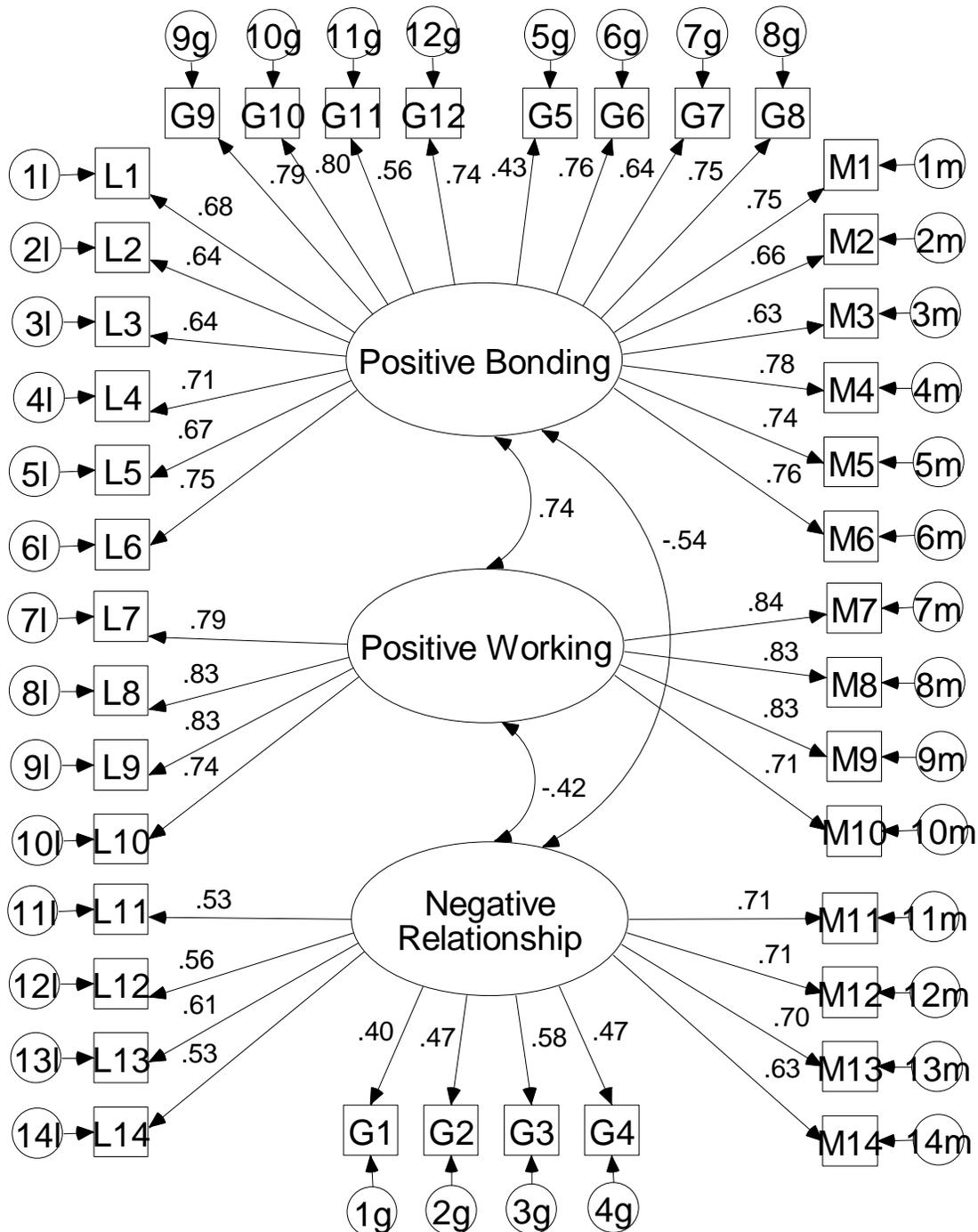
Table 4-4

GQ Standardized Regression Weights

First Order Factors	Second Order Factors	Stand. Regression Weight	Significance
Cohesion	<--- Positive Bonding	.904	.001
Engagement	<--- Positive Bonding	.817	.001
Member_Bonds &_Empathy	<--- Positive Bonding	.942	.001
Member_Tasks &_Goals	<--- Positive Working	1.052	.001
Leader_Tasks &_Goals	<--- Positive Working	.733	.001
Leader_Bonds &_Empathy	<--- Positive Bonding	.654	.001
Leader_Negative_Empathy	<--- Negative_Relationship	.680	.001
Conflict	<--- Negative_Relationship	.529	.001
Member_Negative_Empathy	<--- Negative_Relationship	.997	.001

GQ Correlations

Subscale1	Subscale 2	Correlation	Significance
Positive Bonding	<--> Positive Working	.695	.001
Negative_Relationship	<--> Positive Working	-.341	.001
Negative_Relationship	<--> Positive Bonding	-.482	.001

Figure 4-2: Hypothesis 3 – CFA without 1st Order FactorsFit Statistics

Chi-Squared = 4917, df = 738

Comparative Fit index = .655

RMSEA = .108

P for Test of Close Fit = .000

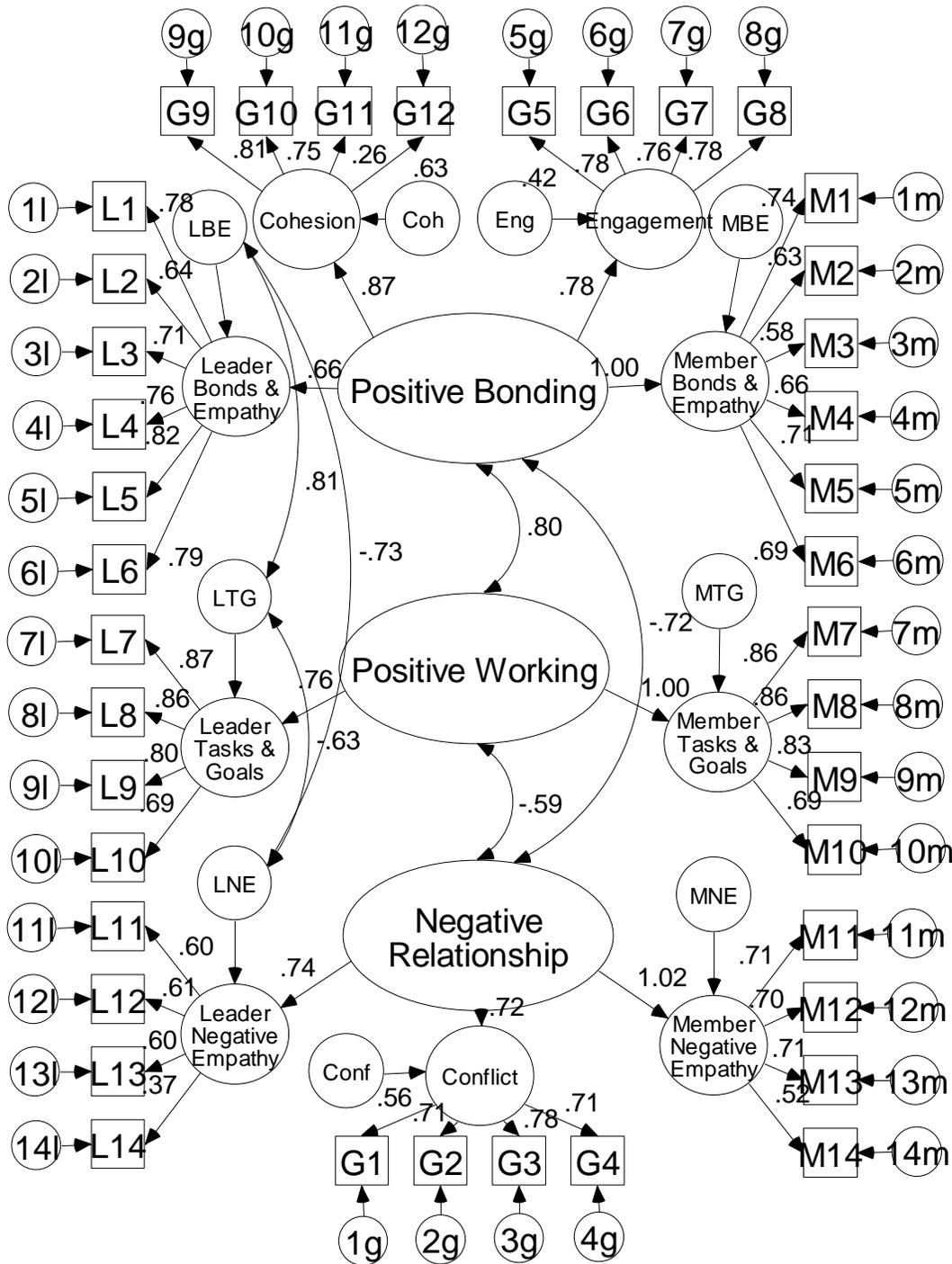
Tests of Hypothesis 3

Hypothesis 3 was tested by removing the first order factors, representing the member-member, member-leader, and member group aspects of the group relationship, and relating the items directly to the 3 second order factors. Fit indices were calculated for this adjusted model and fit indices as reported in Figure 4-2 reflect very poor fit. It should be noted that as reported in Hypothesis 1, the proposed GQ model also did not fit as predicted (Chi-Squared: 2555, $df = 725$, Comparative Fit index: .849, RMSEA: .072, P for Test of Close Fit: .000). However, the fit indices reported in Figure 4-2 are substantially poorer (Chi-Squared: 4917, $df = 738$, Comparative Fit index: .655, RMSEA: .108, P for Test of Close Fit: .000). The fit indices for the originally proposed model are only moderately poor and open the possibility that minor post hoc modifications might be made to improve the model fit. However, fit indices for second order factors only model are poor beyond the hope of refinements. Therefore Hypothesis 3 was accepted and no refinements were attempted.

Tests of Hypothesis 4

All three populations provided a poor fit to the original model. Figure 4-3 shows the results of the UCC population; figure 4-4 shows the results of the AGPA population and figure 4-5 shows the results of the USH population. All indices of fit indicate that the model fits poorly regardless of the population. Furthermore, the indices are not approximating a good or even moderate fit, and it is very unlikely that increased sample size will improve the fit substantially. Hypothesis 4 was unsupported: the model did not provide an adequate fit for any of the three populations.

Figure 4-3: Hypothesis 4 – Original Model with UCC



Fit Statistics

Chi-Squared = 1568, df = 725

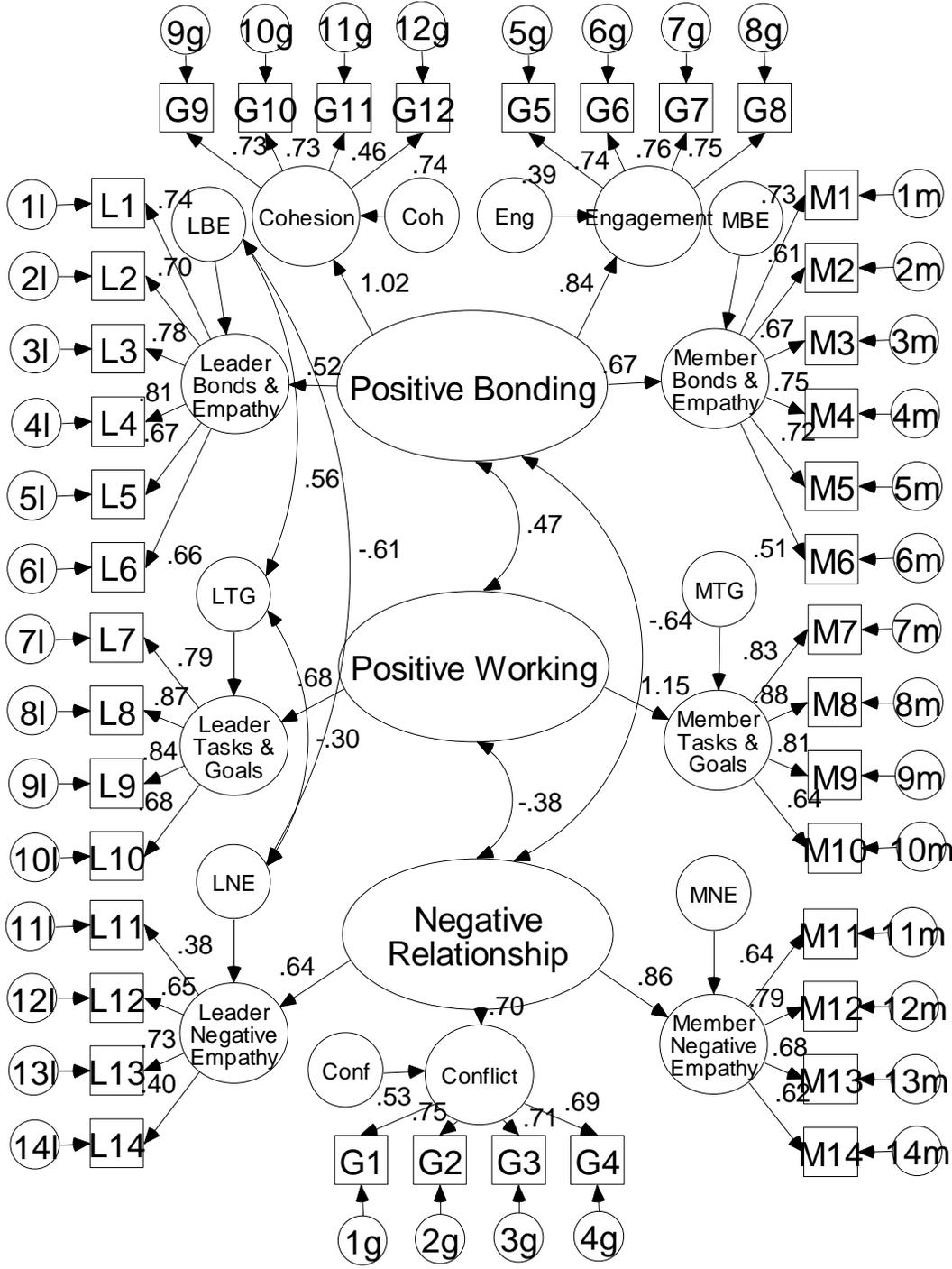
Comparative Fit index = .813

RMSEA = .075

P for Test of Close Fit = .000

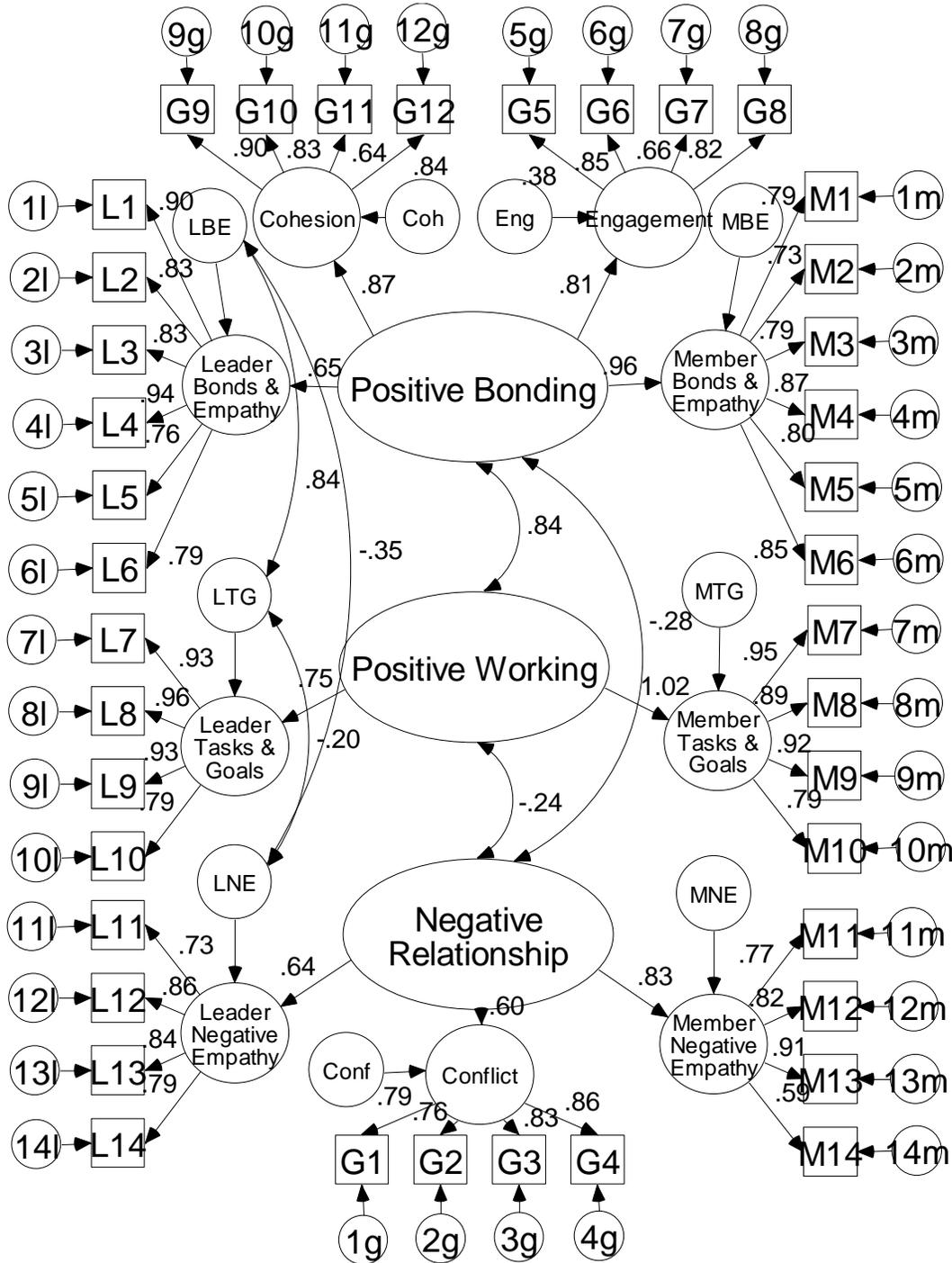
N = 207

Figure 4-4: Hypothesis 4 – Original Model with AGAP



Fit Statistics
 Chi-Squared = 1494, df = 725
 Comparative Fit index = .751
 RMSEA = .082
 P for Test of Close Fit = .000
 N = 160

Figure 4-5: Hypothesis 4 – Original Model with USH



Fit Statistics

Chi-Squared = 1467, df = 725

Comparative Fit index = .822

RMSEA = .094

P for Test of Close Fit = .000

N = 118

Test of Hypothesis 5

A three by three multivariate analysis of variance (MANOVA) was conducted to test for statistically significant differences between the total scores for the 3 second order factor subscales (Positive Bonding, Positive Working, and Negative Relationship) and each of the three populations (UCC, AGPA, and USH). All three populations showed statistically significant differences from one another, and each had a distinct response pattern. Tukey's HSD and Bonferroni were both used to test for statistical significance. Table 4-5 reports the results from Tukey's HSD test of statistical significance between factor means by population. Bonferroni test results were in complete agreement with Tukey's HSD and results are not printed due to redundancy. Table 4-6 reports subscale means and standard deviations for Bonding, Working, and Negative relationship for each population. Means reflect scale scores and not standard scores, making it necessary to make comparisons only within a subscale and not between subscales. Only cases containing complete data were used for the MANOVA and tests of reliability.

The UCC population (Mean = 118) scored significantly higher than the USH (Mean = 100) on Positive Bonding and statistically similar to the AGPA (Mean = 116) population. It scored statistically higher (UCC Mean = 42) than both the AGPA (Mean = 37) and the USH (Mean = 38) populations for Positive Working. It was also statistically lower than both other groups for Negative Relationship (UCC Mean = 23; AGPA Mean = 28; USH Mean = 28). Overall, the response pattern of the UCC population showed that scores were comparatively high on Positive Bonding and Working and low on Negative Relationship (see tables 4-5 and 4-6).

Table 4-5: 3x3 MANOVA

	Pop. 1	Pop. 2	Mean Difference	Significance
Bonds	UCC	AGPA	2.81	---
	UCC	USH	18.13	.000
	AGPA	USH	15.32	.000
Work	UCC	AGPA	5.76	.000
	UCC	USH	4.00	.015
	AGPA	USH	1.77	---
Negative	UCC	AGPA	5.33	.000
	UCC	USH	5.48	.000
	AGPA	USH	1.59	---

Table 4-6: Population Factor Means and Standard Deviations

	Population	Mean	Std. Deviation	N
Bonds	UCC	118	13.24	175
	AGPA	116	12.01	95
	USH	100	25.83	71
	Total	114	17.80	341
Work	UCC	42	8.76	175
	AGPA	37	9.78	95
	USH	38	13.47	71
	Total	40	10.47	341
Negative	UCC	23	8.12	175
	AGPA	28	8.89	95
	USH	28	14.94	71
	Total	25	10.43	341

The AGPA population (Mean = 116) was statistically similar to the UCC population (Mean = 118) for Positive Bonding and was statistically higher than the USH (Mean = 100). It was statistically lower (Mean = 37) for Positive Working than the UCC population (Mean = 42) and statistically similar to the USH population (Mean = 38). For Negative Relationship, the AGPA population (Mean = 28) was again similar to the USH population (Mean = 28) and statistically higher than the UCC population (Mean = 23). Overall, the AGPA population shows a response pattern of being comparatively high in Positive Bonding and Negative Relationship and low in Positive Working.

The USH population (Mean = 100) scored statistically lower than both the UCC (Mean = 118) and AGPA (Mean = 116) populations on Positive Bonding. On the Positive Working subscale, the USH population (Mean = 38) scored statistically similar to the AGPA population (Mean = 37) and statistically lower than the UCC population (Mean = 42). The USH population (Mean = 28) also scored statistically similar to the AGPA population (Mean = 28) and statistically higher than the UCC population (Mean = 23) on Negative Relationship. Overall, the USH population showed a response pattern of being comparatively low in Positive Bonding and Positive Working and high in Negative Relationship.

Cronbach's Alpha was used to calculate the test reliability for each subtest of the GQ (see Table 4-4). The overall reliability of the GQ was not reported because the GQ is designed to be evaluated based on its three subscales and there is no meaning at the level of a total score. Positive Bonding had a reliability estimate of .93, and Positive Working had a reliability estimate of .90. Both are in the good range for reliability as defined as .90 and above. The reliability estimate of Negative Relationship was substantially lower at .84.

Johnson et al. (2005) also cited problems with the reliability of Negative Relationship. They specifically identified the conflict scale, which originated from the Group Climate Questionnaire (GCQ), as being problematic. Some items from the GCQ Conflict scale have been retained in the GQ and are included in Negative Relationship (see figure 3-1).

In an effort to isolate the source of lowered reliability on the Negative Relationship factor and whole measure, further tests of reliability were run, separating out populations and looking for evidence of restricted range. Table 4-5 reports reliability estimates for the GQ. These subsequent analyses reveal large discrepancies in the reliability of the measure by population. The USH population demonstrated far superior reliability estimates for all subscales, especially the Negative Relationship (see Table 4-5). Even with a small sample size, these reliability estimates are all in the good range of reliability.

Reliability estimates for the UCC and AGPA populations dropped on all subscales and for the total measure (see Table 4-5). Reliability estimates are highly subject to the amount of variance present in the item responses. Table 4-2 reports the standard deviation of each subscale by population. In each case, the USH population has a 50-100% greater standard deviation than either the UCC or AGPA populations, which greatly resemble each other on standard deviation. Ghiselli (1981) provides a formula which estimates what the reliability of a measure would be given a different sample size. This formula was used to estimate the reliability of the UCC and AGPA populations if they had the same variance or range as the USH population. Results are shown in table 4-6. They indicate that if the sample variance were increased, then reliability for both the UCC and AGPA groups would be high for all three factors. This further suggests that the lower reliabilities for these two populations were due to a restriction of range and not a lack of reliability for the GQ.

Figure 4-6: Reliability Estimates

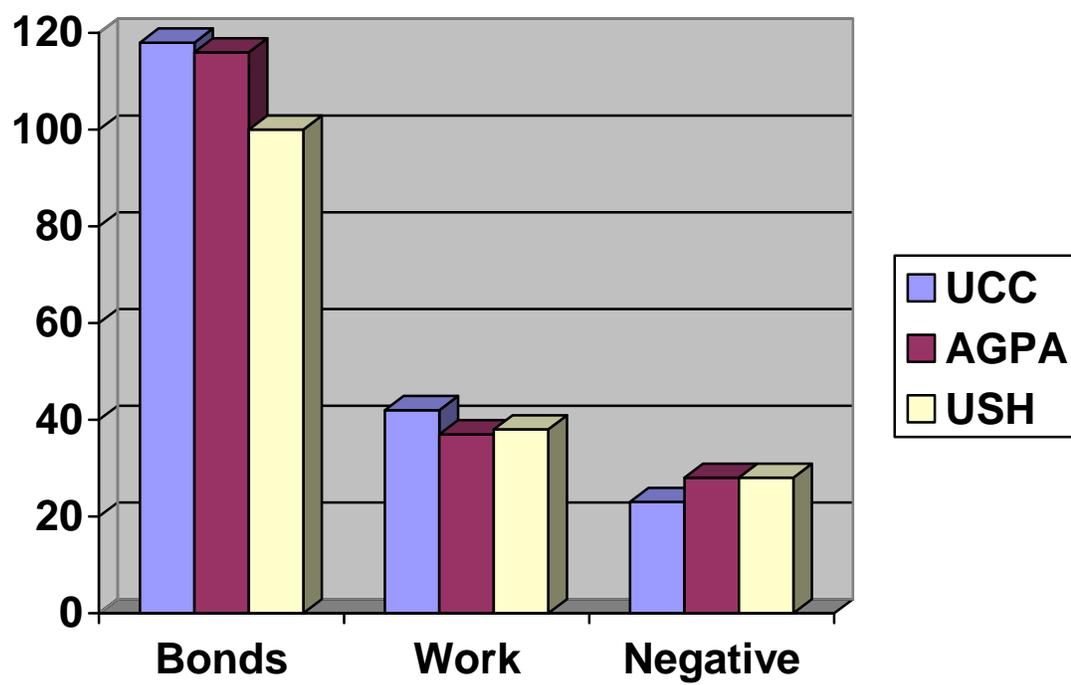


Table 4-7: Reliability of the GQ

	Cronbach's- Alpha	N of Items	N of Cases
<u>All populations</u>			
Bonding	.93	20	402
Working	.90	9	415
Negative	.84	13	431
<u>UCC</u>			
Bonding	.90	20	195
Working	.91	9	194
Negative	.79	13	194
<u>AGPA</u>			
Bonding	.89	20	115
Working	.90	9	123
Negative	.74	13	146
<u>USH</u>			
Bonding	.94	20	92
Working	.95	9	98
Negative	.89	13	90

Table 4-8: Adjusted Reliability of the GQ

	Cronbach's- Alpha	N of Items	N of Cases
<u>UCC</u>			
Bonding	.97	20	195
Working	.96	9	194
Negative	.94	13	194
<u>AGPA</u>			
Bonding	.98	20	115
Working	.95	9	123
Negative	.91	13	146

Tests of Hypothesis 6

Refinements were made to the GQ in AMOS following the procedures delineated in Chapter 3 to improve the model fit from Hypothesis 1. A series of minor modifications were made which resulted in a model with good fit. Figure 4-7 shows the new model. Due to the need to rearrange the appearance of the model after the modifications were complete, the changes visually appear to be greater than the actual statistical and structural changes to the model. Only one change was made to the structural model. The first order factors of Cohesion and engagement were found to correlate at .90. They both represent the whole group dimension of Positive Bonding; model fit did not vary based on their separation or combination. In an effort to create greater parsimony in the model and eliminate needless complication, a decision was made to combine the two into one new factor labeled Whole Group Bonding. Figure 4-7 shows the structural model for the new GQ.

A total of 10 items were dropped from the GQ due to small regression weights or factor loadings with their associated first order factor. Items with a parallel member-leader or member-member question were dropped only in pairs and only if neither item had a strong factor loading. During the refinement process, a high degree of correlation was found between the member-leader and member-member counterpart questions. Thus, items with poor factor loading on either member-leader or member-member side almost always performed similarly on the parallel side. Three sets of parallel questions were dropped, comprising six of the ten dropped items. Items L2, L3, L14, M2, M3, M14, G3, G5, G7 and G11 were dropped from the model and questionnaire. Table 4-9 provides a list of the items that were dropped.

Due to the high degree of correlation between the items that run parallel across member-leader and member-member dimensions, 7 of the 14 pairs were allowed to covary on their error terms. This decision greatly improved the model fit, and it makes theoretical sense given the high degree of similarity between the items. The first order factors were renamed to give greater clarity to the model. The original names were carry-overs from some of the scales used to help create the GQ. The model was also visually reorganized to neatly accommodate the new changes made.

The refined GQ was tested using Hypotheses 1–5, and the resulting analyses are presented below.

Hypothesis 1: Refined GQ

Results from the CFA of the refined GQ provided a good fit to the data (Chi-Squared: 775.4, $df = 381$; Comparative Fit index: .957; RMSEA: .046; P for Test of Close Fit: .930 $N = 486$; see figure 4-7) and Hypothesis 1 was supported. This suggests that the refined 30-item GQ measurement model does adequately represent the relevant theoretical constructs in these samples. Figure 4-8 shows the refined GQ model and fit indices.

Hypothesis 2: Refined GQ

Hypothesis 2 was supported for the refined GQ model in the same way it was with the original. All subscales representing positive and negative aspects of the group relationship model were positively and negatively correlated as predicted by the model. Table 4-10 shows the standardized regression weights of each second order factor to its associated first order factor and the level of significance. This table also shows the correlations of the second order factors to each other and the associated significance levels.

Table 4-9: Items Dropped from the Original GQ

- L 2 The group leaders understood what I said during today's session.
- M2 The other group members understood what I said during today's session.
- L3 I am confident in the group leaders' ability to help me.
- M3 I am confident in the other group members' ability to help me.
- L14 The group leaders acted condescending and talked down to me.
- M14 The other group members acted condescending and talked down to me.
- G3 The members rejected and distrusted each other.
- G5 The members revealed sensitive personal information or feelings
- G7 The members tried to understand why they do the things they do, tried to reason it out.
- G11 There is a sense of loss when a member of our group is absent

Table 4-10

Refined GQ Standardized Regression Weights

First Order Factors	Second Order Factors	Stand. Regression Weight	Significance
Whole_Group_Bonding	<--- Positive Bonding	.826	.001
Member_Bonding	<--- Positive Bonding	.975	.001
Member_Working	<--- Positive Working	1.058	.001
Leader_Working	<--- Positive Working	.717	.001
Leader_Bonding	<--- Positive Bonding	.639	.001
Leader_Negative	<--- Negative_Relationship	.595	.001
Whole_Group_Negative	<--- Negative_Relationship	.489	.001
Member_Negative	<--- Negative_Relationship	.936	.001

Refined GQ Correlations

Subscale1	Subscale 2	Correlation	Significance
Negative_Relationship	<--> Positive Working	-.387	.001
Positive Bonding	<--> Positive Working	.688	.001
Negative_Relationship	<--> Positive Bonding	-.513	.001

Hypothesis 3: Refined GQ

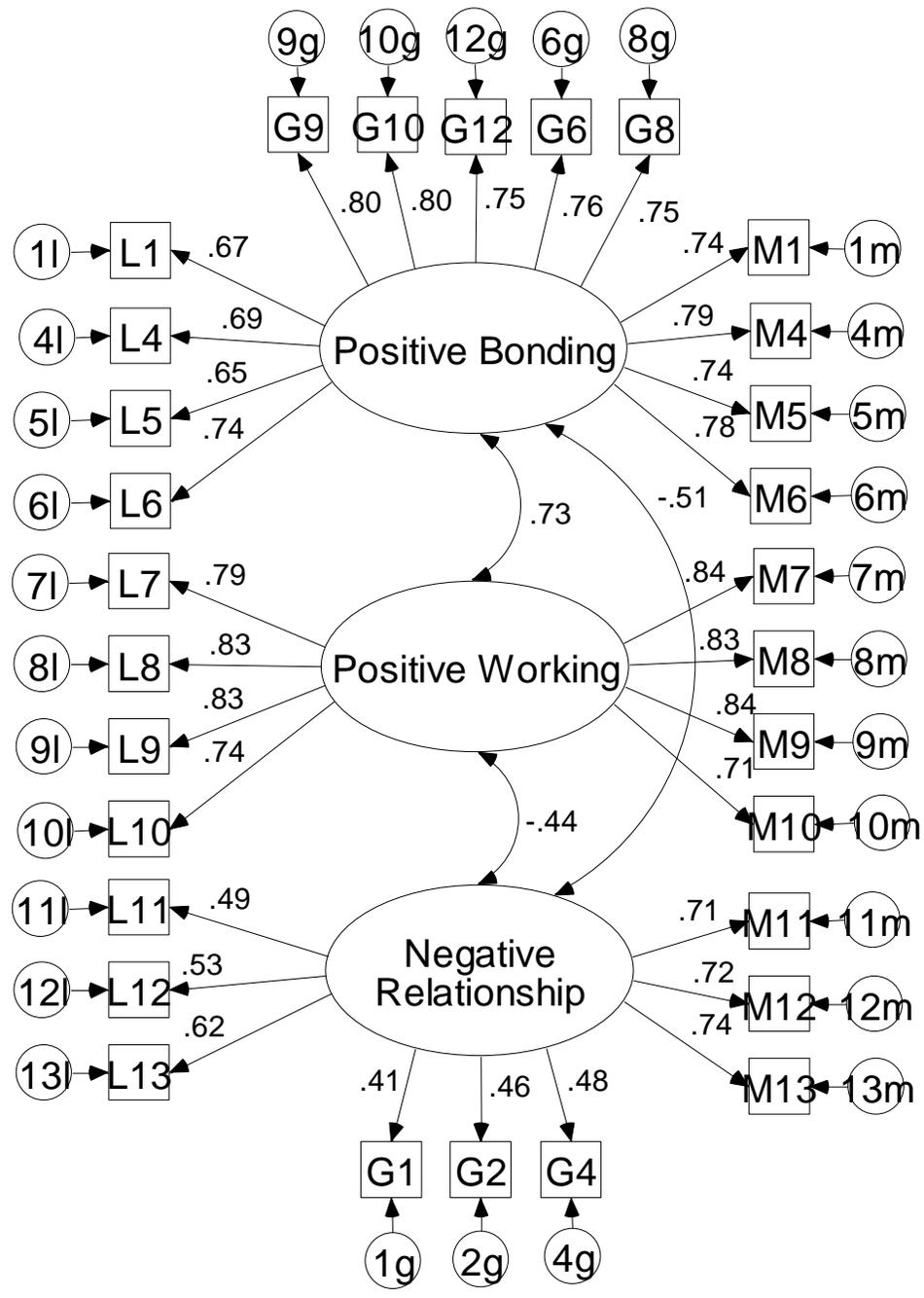
Hypothesis 3 was again tested by removing the first order factors; representing the member-member, member-leader, and member-group aspects of the group relationship; and relating the items directly to the 3 second order factors. Fit indices were calculated for this revised model. As with the original tests of Hypothesis 3, fit indices for the refined model without the first order factors, as reported in Figure 4-9, reflected poor fit.

Hypothesis 4: Refined GQ

Sample size presented a significant challenge in testing Hypothesis 4. The total data set contains data from 486 participants and provides statistically robust analyses for testing Hypotheses 1–3. However, when divided by population the UCC population contained an N of 207, the AGPA an N of 160 and the USH an N of 118. Each of these falls short of the recommended 200–300 participants needed for robust CFA tests. Low sample size makes achieving adequate goodness of fit difficult. As can be seen in the results reported below, the goodness of fit scales with the sample size of each population. This creates a challenge in interpretation and renders the results less conclusive.

The revised GQ model was tested separately on each of the three populations. It demonstrated moderate fit for each of the three populations. There was a surprisingly small degree of variation in the goodness of fit of the model for the three populations. The UCC population provided the best fit to the refined model (Chi-Squared: 654, $df = 413$; Comparative Fit index: .933; RMSEA: .054; P for Test of Close Fit: .209; see figure 4-10) and can be described as moderate to good. The Chi-Squared and P for Test of Close Fit both indicate a good fit; however the CFI and RMSEA fall just short of the cut-off for good fit and are in the moderate category. The sample size for the UCC population was 207. It is likely

Figure 4-9: Refined GQ without 1st Order Factors



Fit Statistics
 Chi-Squared = 3049.5, df = 403
 Comparative Fit index = .706
 RMSEA = .116
 p for Test of Close Fit = .000
 N = 486

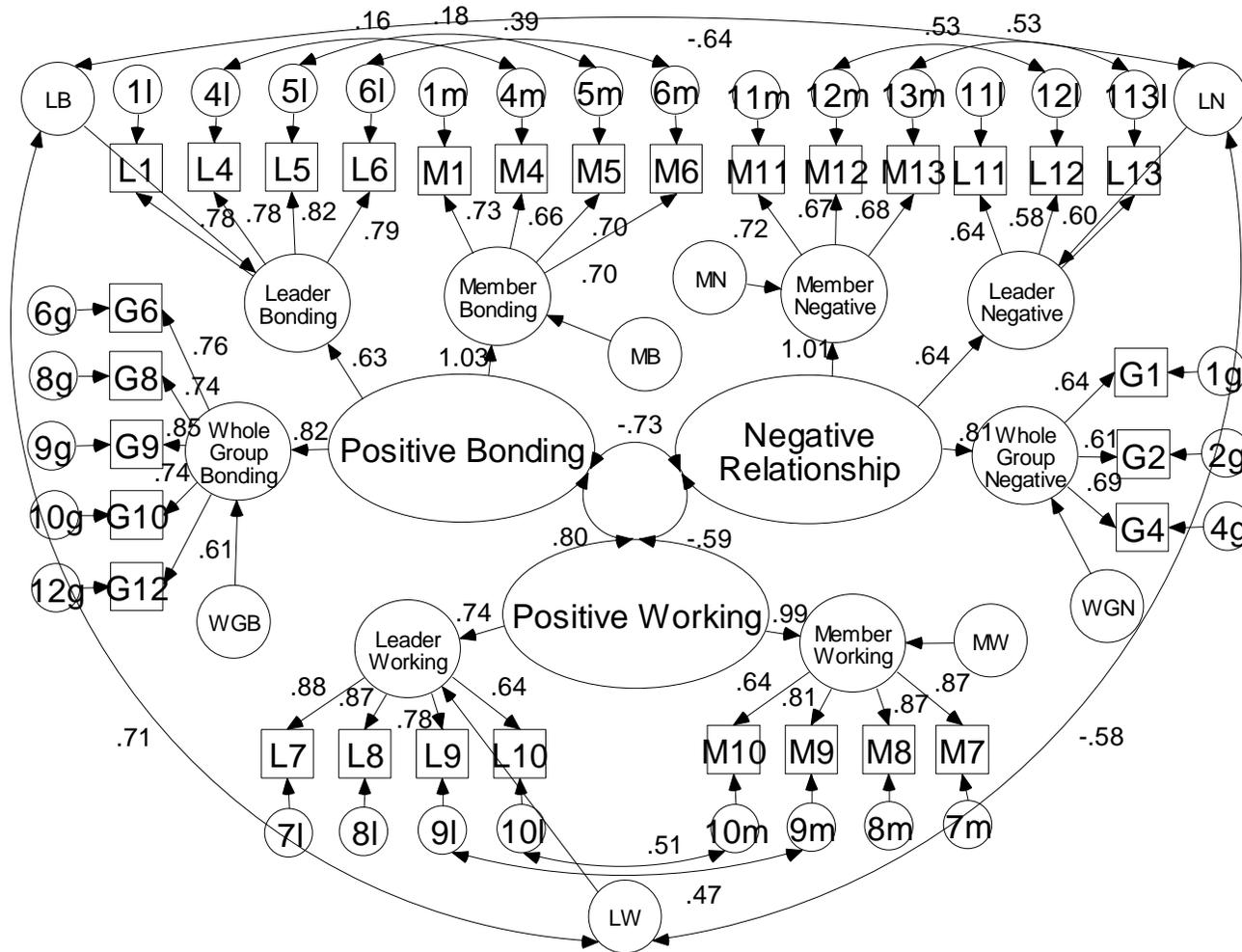
that with a larger sample size, the CFI and RMSEA would also reflect good fit. The AGPA population fit indices show it to be consistently in the range of moderate fit, with the exception of the P for Test of Close Fit, which fall just short of the .05 significance level (Chi-Squared: 648, df = 413; Comparative Fit index: .907; RMSEA: .059; P for Test of Close Fit: .061; see figure 4-11). This population contained an N of 160, which is insufficient to reliably support the CFA test conducted. The USH population had the poorest fit (Chi-Squared: 618, df = 384; Comparative Fit index: .922; RMSEA: .072; P for Test of Close Fit: .001; see figure 4-12), although not far behind the AGPA population. The USH population also had the smallest sample size, with a total of 118 participants. It is important to note that the populations' goodness of fit with the model is in order of population size. Johnson et al. found the model to fit both the UCC and AGPA population with a larger population size for each group. The results of Hypothesis 4 are therefore inconclusive. It is very possible that give a larger sample size, the goodness of fit would move from moderate to good. Hypothesis 4 is thus tentatively supported with a need for further research to verify these moderate results.

Hypothesis 5: Refined GQ

Results for Hypothesis 5 using the revised GQ were very similar to those already reported using the original model. Table 4-11 reports the population by factor means and standard deviation. Table 4-12 reports the results of Tukey's HSD. The only difference from the previous analysis that reaches statistical significance is that all three populations were significantly different from each other on Positive Bonding. Specifically, the AGPA population scored significantly higher on Positive Bonding than did the USH population. All other results were not significantly changed.

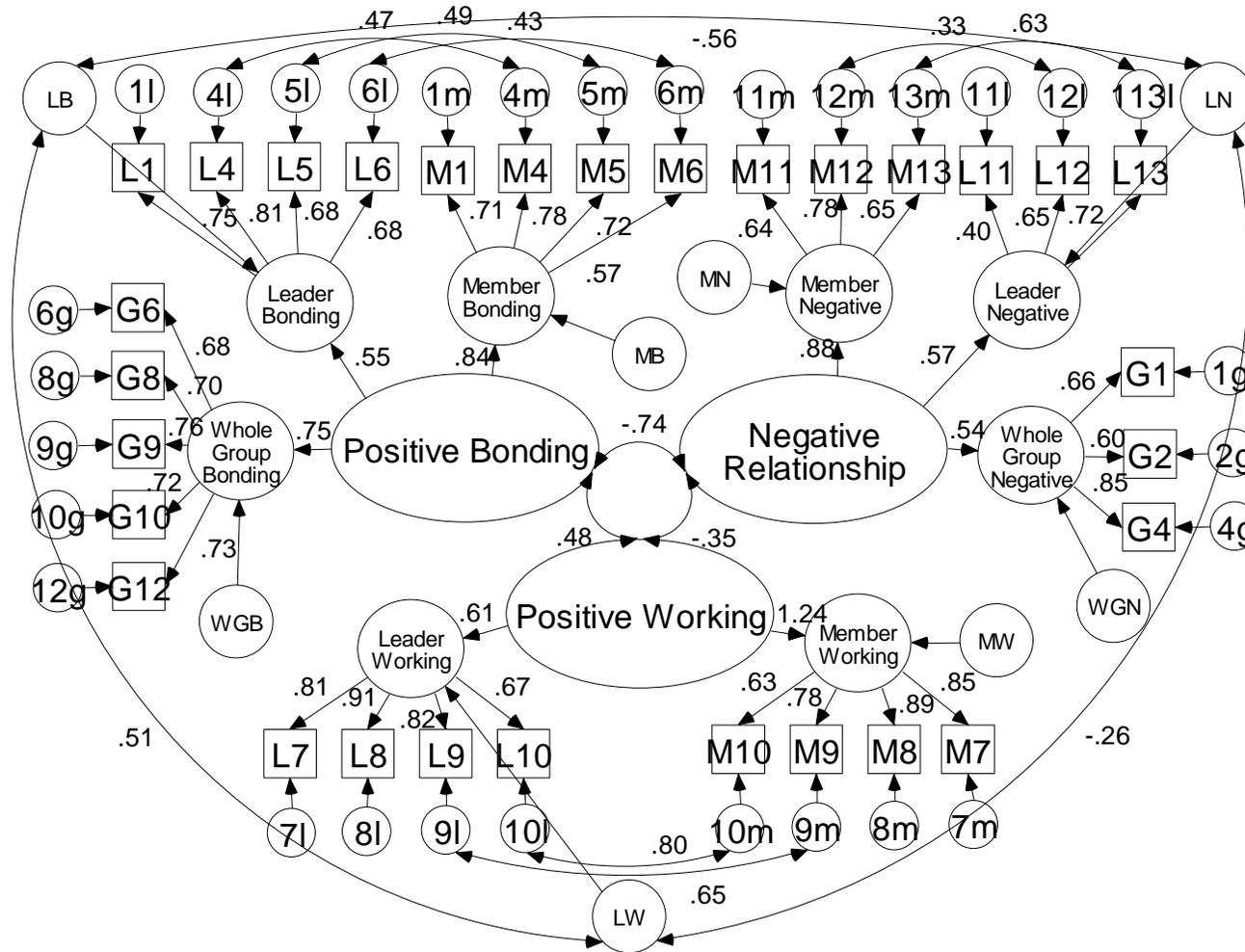
Chronbach's Alpha coefficients dropped slightly on the revised model. This is most likely a result of the shorter test length since reliability depends heavily on the number of items included on a test. Even with the slight drop, overall results for reliability remained consistent with previous findings. Reliability is highest for Positive Bonding, followed by Positive Working. Reliability for Negative Relationship is poor for the UCC and AGPA populations, but moderate for the USH population. As before, the restriction of range on the UCC and AGPA populations significantly decreased the reliability estimates. Table 4-13 reports the reliability for the revised GQ and table 4-14 reports the reliability with adjustments made for restriction of range.

Figure 4-10: New Model with UCC



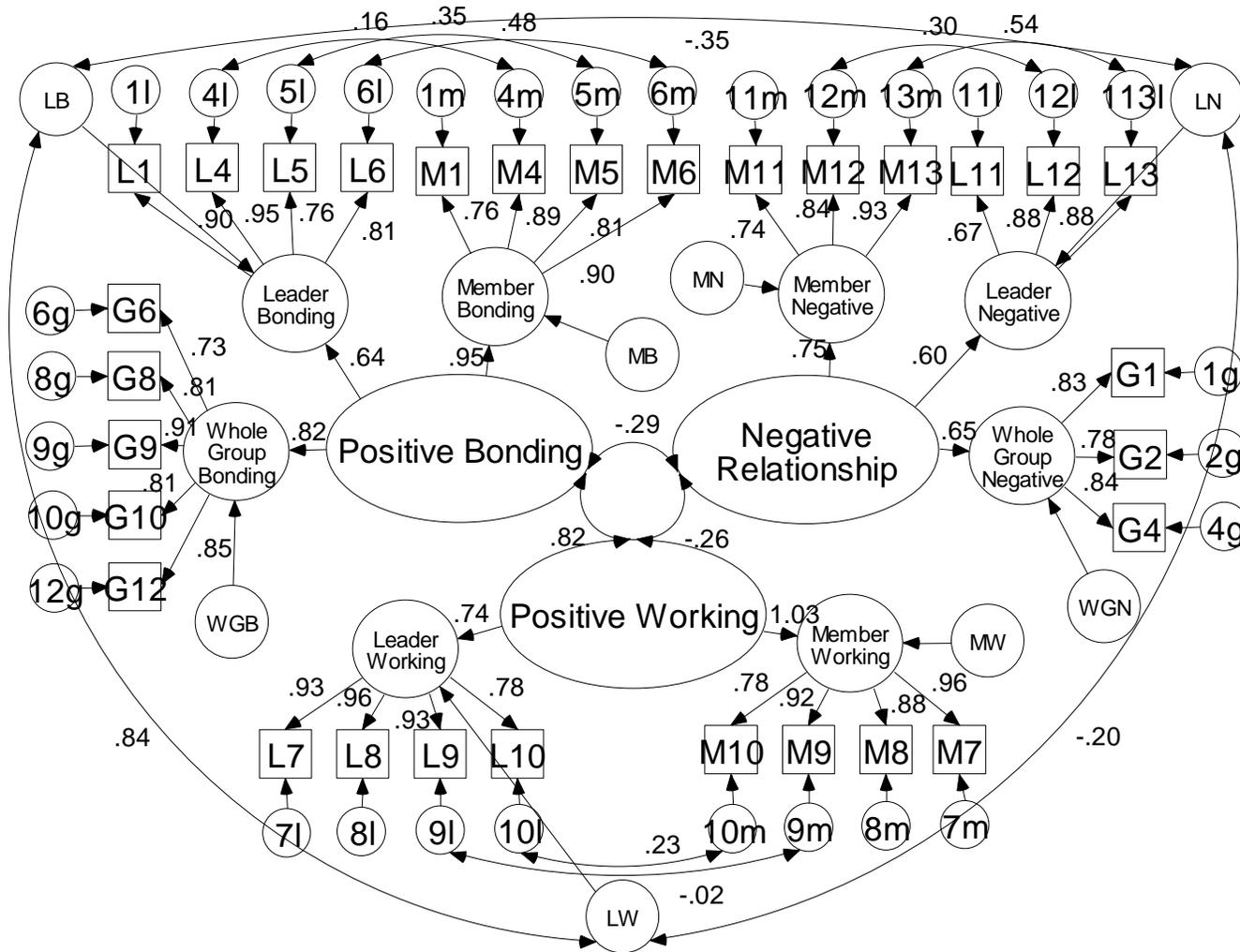
Fit Statistics: Chi-Squared = 654, df = 413; Comparative Fit index = .933; RMSEA = .054; *p* for Test of Close Fit = .209; N = 207

Figure 4- 11: New Model with AGPA



Fit Statistics: Chi-Squared = 648, df = 413; Comparative Fit index = .907; RMSEA = .059; *p* for Test of Close Fit = .061; N = 160

Figure 4-12: New Model with USH



Fit Statistics: Chi-Squared = 618, df = 384; Comparative Fit index = .922; RMSEA = .072; P for Test of Close Fit = .001; N = 118

Table 4-11: Population by Factor Means and Standard Deviations for the Revised GQ

	Population	Mean	Std. Deviation	<i>n</i>
Bonds	UCC	79	9.01	179
	AGPA	75	8.78	112
	USH	66	17.30	79
	Total	75	12.16	370
Work	UCC	42	8.93	179
	AGPA	36	9.59	112
	USH	38	13.00	79
	Total	40	10.45	370
Negative	UCC	18	6.77	179
	AGPA	22	7.57	112
	USH	22	11.56	79
	Total	20	8.46	370

Table 4-12: 3x3 MANOVA

	Pop. 1	Pop. 2	Mean Difference	Significance
Bonds	UCC	AGPA	3.46	.030
	UCC	USH	12.23	.000
	AGPA	USH	8.77	.000
Work	UCC	AGPA	6.02	.000
	UCC	USH	4.13	.008
	AGPA	USH	1.89	---
Negative	UCC	AGPA	3.96	.000
	UCC	USH	3.89	.002
	AGPA	USH	0.07	--

Table 4-13: Reliability for the refined GQ

	Cronbach's- Alpha	<i>n</i> of Items	<i>n</i> of Cases
<u>All populations</u>			
Bonding	.92	20	402
Working	.90	9	415
Negative	.80	13	431
<u>UCC</u>			
Bonding	.91	20	195
Working	.91	9	194
Negative	.71	13	194
<u>AGPA</u>			
Bonding	.87	20	115
Working	.90	9	123
Negative	.71	13	146
<u>USH</u>			
Bonding	.93	20	92
Working	.95	9	98
Negative	.86	13	90

Table 4-14: Adjusted Reliability for the refined GQ

	Cronbach's- Alpha	<i>n</i> of Items	<i>n</i> of Cases
<u>UCC</u>			
Bonding	.98	20	195
Negative	.90	13	194
<u>AGPA</u>			
Bonding	.97	20	115
Negative	.88	13	146

Chapter 5

Discussion

This study successfully created the Group Questionnaire (GQ) as a measure of the group therapeutic relationship. Creation of the GQ represents a significant contribution to both researchers of group psychotherapy and to clinicians conducting group psychotherapy. It differs from other previously created measures of group psychotherapy because it is based on a unified theory of the group relationship that was empirically supported by Johnson et al. (2005) and validated by both Bormann and Straus (2007) and Bakali et al. (2008). Johnson's new theoretical model simplified and unified the overlap between previously defined group relationship constructs of Group Climate, Cohesion, Empathy, and Working Alliance. In this study, the GQ successfully measured this new unified theory of group relationship in a brief and clinically relevant manner.

One of the major advantages of the GQ is that it was created using both clinical judgment and empirical evidence. A large effort was made in creating the GQ to allow practice and research to inform each other. The result is that the GQ is a brief, 30-item questionnaire that takes only 5–10 minutes to complete and provides the leader valuable clinical information. It can inform the leader about the quality of the group relationship in three core areas: Positive Bonding, Positive Working, and Negative Relationship. It can also offer specific information about which group relationships (member-member, member-leader, or member-group) are relatively strong or weak. Previous research has linked group relationship factors to client outcome (Johnson et al., 2004); the GQ can provide a group leader with insight that can be used to inform therapy interventions and strengthen the group relationship.

A second major advantage of the GQ is that it is not only clinically relevant, but it is also empirically sound. In the present study, Hypotheses 1–5 were tested on the GQ twice, first with the initial version and then with the revisions.

Validation of the GQ

The initial GQ did not provide adequate fit to the model based on SEM goodness of fit indices when Hypothesis 1 was tested. However, the fit was only moderately poor and after minor refinements were made to the GQ it demonstrated good fit on all fit indices. This suggests that Johnson's group relationship model is a good representation of the theoretical constructs underlying the GQ. Given the number of items dropped or changed to create the GQ, it was not surprising that refinements were needed to improve the model fit, and in the process increase the efficiency of the measure. Refinement processes are a standard part of measurement development. It would be highly improbable that a measure would be created in the optimal way on the first try. This study is part of a series of research studies—past, present, and future—that will be used to create and validate the GQ and its uses.

Hypothesis 2 tested if the positively and negatively oriented factors in the model were significantly correlated with each other in the predicted direction. The predicted correlations between factors were universally supported for both the original and revised versions of the GQ tested. These findings suggest that while each factor is separate and unique, they are significantly positively or negatively correlated with each other in predictable ways. For example, Positive Bonding has a positive moderate correlation with positive work, and a negative moderate correlation with Negative Relationship. Thus, a group that is bonding well together is likely to also be working well together and

have fewer aspects of Negative Relationship.

It is also important to note that the correlation between Positive Bonding and Negative Relationship is only moderate and not approaching strong. It might be tempting to conceptualize Negative Relationship as the opposite of Positive Bonding; however, the correlation coefficients suggest that while they are negatively correlated, they are not inverses of each other. Negative Relationship is not the opposite of Positive Bonding, but rather a separate and negatively correlated construct. At this point in the development of the GQ it is conceptually unclear why and how Negative Relationship is not the opposite of Positive Relationship. A better qualitative description of the Negative Relationship is needed to enhance the utility of the scale. It is also possible that Negative Relationship may be different for different populations. For example, it could be theorized that Negative Relationship in a UCC population may represent a lack of connection or genuineness from group, where as a population like the USH may experience more overt negative comments and behaviors from group. Future research may find a different, more specific name than Negative Relationship to be more descriptive of what is actually being measured and thus improve the clinical utility.

One question that arose during the creation of the GQ was the necessity of the first order factors reflecting the member-leader, member-member, and member-group aspects of the relationship. Chapter 1 presented the research supporting the importance of these three dimensions of the group relationship. The distinctions make good theoretical sense and are clinically useful in conceptualizing the group process; however, the empirical need for these constructs was questioned. The model would be much simpler without the first order constructs, leaving only the three main second order

factors to interpret. Hypothesis 3 tested both the original and revised GQ and found both to have very poor fit to the model when the first order factors were removed. This gives strong empirical support for the retention of the first order factor on the GQ model. On a broader scale, it may offer some empirical support for theoretically and clinically important differences between the relationships in general.

An interesting outcome of this study is the performance of the GQ across three very different populations. It might be expected that GQ model would not be an appropriate fit for one or more of the populations tested. There were several significant differences in the construction of the groups as well as the populations involved. For example, the USH groups were all psychoeducational groups using manualized treatments that were conducted by trained nurses and ran for a period of months. In contrast, the AGPA groups were all process oriented, had no set agendas, were conducted by very experienced group leaders, and ran for a period of two days. The UCC groups displayed greater variety in their format, leaders, and length. Despite these dramatic differences in structure and population, the testing of Hypothesis 4 showed all three populations tested to demonstrate moderate fit to the model. Although the sample sizes are insufficient, rendering the analysis inconclusive, these preliminary findings parallel Johnson's findings that the group relationship model fit both the UCC and AGPA populations used in that study.

The populations did not differ appreciably in their degree of fit to the model; however, each showed a unique response pattern. These response patterns provide insight into how the GQ performs with different populations and also reveal some group relationship patterns for the populations. The findings make clinical sense given a

knowledge of the populations. The UCC population had the best overall scores for the GQ. Results suggest that UCC group members in this study experienced a strong Positive Bonding relationship, a strong Positive Working relationship, and minimal amounts of Negative Relationship experiences. These groups could also be described as having high cohesion and empathy, working productively together, and experiencing little friction and negative affect. This fits with the tendency of UCC clients to have generally positive therapy experiences and to engage in a positive response bias.

The response pattern for the AGPA population reflected moderate amounts of Positive Bonding, low levels of Positive Working and small amounts of Negative Relationship. This pattern also makes sense in light of the AGPA population. These groups had met for only one day, although they had spent about 8 hours together during that day. The recency of these groups makes the relationships new and it is likely that less time had transpired in which to form a strong cohesive, trusting, bonding relationship. Positive Working was substantially lower for the AGPA than for the UCC population. This is may be due to the fact that the groups were not therapy groups and often lacked set objectives. Group members were not coming to group to work on a specific personal problem, and thus some of the Positive Working questions had little relevance to this population. In fact, it was apparent from looking at the questionnaires that many participants did not respond to these questions or wrote NA to the side of them. This systematic problem of missing data was accommodated for statistically; however, it yields good clinical information about the relevance of the Positive Working scale for the AGPA population. An alternative interpretation to the low Positive Working other than the group working poorly together is that people endorsed the items weakly because they

lacked relevance. The AGPA population was also low in Negative Relationship, which may be due to aspects of the population of professionals and also to the brevity of their time in group. They may have had limited time to develop feelings of friction and conflict.

The response pattern of the USH population had the overall poorest clinical profile. The group members scored relatively low on Positive Bonding and Positive Working and high on Negative Relationship. Clinicians who have run groups for severely and persistently mentally ill populations may have had similar experiences with their groups being generally less positive and more conflictual. This is likely enhanced because many group members may not be attending voluntarily, and there is often a distinct division between the patients and the staff members. All of these factors make the group relationship more problematic, as reflected by the GQ results for the USH population.

Reliability for the GQ was evaluated on the level of the second order factors, which is the level offering the most clinical significance. Reliability coefficients were generally good; however, there were distinct discrepancies between different subscales and populations. All reliability estimates for both the original and revised GQ models were in the good range for both Positive Bonding and Positive Working. The Negative Relationship factor was more problematic due to a restriction of range in the UCC and AGPA populations. In general, the USH population demonstrated the highest reliability on the GQ across factors, especially for Negative Relationship. Further analysis of the reliability showed that this was due to the substantially larger range of responses given by the USH population. When restriction of range was accounted for on the UCC and

AGPA populations, they had an estimated reliability that was higher than the USH population. Other measures have documented similar problems with measuring constructs like Negative Relationship. Specifically, the conflict scale on the Group Climate Questionnaire (GCQ) has been documented to have problems with reliability (Johnson et al., 2004). The difficulty seems to lie in the nature of the construct. It is generally more difficult for people to endorse negatively oriented questions than positively oriented questions. The difficulty of reliability on the Negative Relationship scale, or others like it, does not diminish its clinical relevance. Clinicians are aware of the importance of Negative Relationship aspects to the group even if group members are reluctant to identify it. However, further study of the GQ is may reveal varying relevance or irrelevance of individual scales for specific populations. For example, with the range restriction present in Negative Relationship for the UCC and AGPA populations, it could be argued that it affords less or different clinical utility for these populations.

Limitations

The present study is limited by the insufficient sample size of each of the populations individually. The total sample collected is sufficient to run robust CFA analyses; however, the individual populations are not large enough to thoroughly test the fit of the GQ model for each of the populations. In this case, the type of population is confounded with the size of the sample, making it impossible to determine from this data the reason for the slight drop in model fit.

Shrinkage is a common and expected occurrence when replicating factor analyses. Including the present study, Johnson's model has been successfully replicated three times (Bormann & Straus, 2007; Bakali et al., 2008). Although each replication found a good fit

for the model, some shrinkage occurred with each replication. Further replications using the newly created GQ may experience further shrinkage of the goodness of fit. Depending on the degree of shrinkage in following replications, the fit may drop into the moderate range.

Another weakness of the samples used in this study is the restricted range of responses for both the UCC and AGPA populations. This specifically affects the reliability coefficients on the Negative Relationship scale by weakening the reported reliability. Adjustments for the restriction of range suggest that the GQ itself has good reliability; however, it is dependent on the range of the population. It seems safe to assume that the GQ is reliable when used with severely mentally ill populations like the USH or others with a wider range of responses. However, the reliability of the GQ Negative Relationship factor for populations such as the UCC and AGPA that predictably yield a small range of responses is more difficult to interpret. The GQ itself does appear to be reliable and the argument could be made that the sample tested does not change the test and the test may be accurately reflecting a lack of Negative Relationship. However, it could also be questioned if the GQ Negative Relationship scale is appropriate for these populations when group members are inclined to under report their negative experiences. Perhaps it is advisable to use caution in the clinical interpretation of the Negative Relationship scale when applying it to populations that may struggle with addressing negative aspects of their group experience.

The use of three very different populations in the creation of the GQ greatly strengthens its generalizability and provides important information about its use with the groups tested. However, despite the range in the clinical sample, it must not be assumed that these groups are a good representation of all groups or populations. The sample used

primarily reflects a wide range of clinical severity (from severely mentally ill to non-patients). Due to the geographical location of the sites, minority populations were not well represented and this study was not designed to test for differences between these smaller minority groups. There is also no information on the effect that age, diagnosis, or gender may have on the GQ. It seems likely that there are significant differences in how these populations might behave in group and respond to the GQ.

The process of creating the GQ in this study resulted in the elimination or change of half of the original items used to create the measure model. Given the amount of change, there is a chance that in the process construct drift might have occurred. That is, over the course of changing the GQ, the current three factors may not represent exactly the same three factors from Johnson's study. Further testing could be done comparing the original 60 items with the current 30 to check for construct drift. One of the protective advantages of the way the GQ was constructed was the use of clinical judgment as well as empirical data in creating the GQ. Before any items were eliminated or changed, the team of experienced group researcher/clinicians articulated the essence of each factor and worked to reflect this essence by creating a more pure measure of each of these factors. While this cannot assure the absence of construct drift, it does protect against arbitrary and unguided drift due to statistical nuances of the data. A weakness of team which created the GQ is that the dominant clinical experience of its members was with a UCC population. This may unintentionally bias the GQ to be more relevant for that population.

Future Directions

This study was one in a line of research focused on creating and validating a new

measure of the group relationship with high clinical utility. Future studies are already being designed to continue this work. Among the first is an effort to translate the GQ into German and replicate these findings. If the results of this study can be replicated, they will add strength to the validity of the GQ and an important next first step to making the GQ clinically relevant. Other upcoming studies that are being proposed include the creation of an on-line version of the GQ that could be taken in many convenient locations and scored and interpreted with great ease. Again, these efforts speak to the intent to make the GQ a very clinically relevant and accessible tool.

Other areas for future research on the GQ include creating a set of normative data and proposing methods for scoring and clinical interpretation. The GQ has the potential of offering information on both the individual and group level. Perhaps qualitative data could be collected on the best ways to administer the GQ to clients and on group leaders' perceptions of its utility and ease of administration. It has also been suggested that GQ might be used to track changes in a group over time. This would require multiple administrations to a group to establish the GQ's sensitivity to change. If the GQ were successful at tracking changes in the group relationship, a study might be done on its ability to predict client change. Ultimately, the aim of this research should be to improve the clinical utility and empirical soundness of the GQ so that it may be use to inform group leaders, increase knowledge of group psychotherapy, and promote client change.

References

- Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. California: Wadsworth.
- Antonuccio, D.O., Davis, C., Lewinson, P. M., & Breckenridge, J. S. (1987). Therapist variables related to cohesiveness in a group treatment for depression. *Small Group Behavior, 18*(4), 557–564.
- Arbuckle, J. L., & Wothke, W. (1999). *AMOS 4.0 user's guide*. Chicago: Smallwaters.
- Bakali, J.V., Baldwin, S., Lorentzen, S. (2008). Measuring group processes: A factor analysis of the Group Climate Questionnaire, the Therapeutic Factors Inventory, (subscale cohesiveness). *A paper presented at the annual meeting of the Society for Psychotherapy Research – Spain*.
- Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs, 76*(7), 562.
- Beck, A. T., Wright, F. D., Newman, C. F., & Liese, B. S. (1993). *Cognitive therapy of substance abuse*. New York: Guilford Press.
- Bergin, A. E., & Lambert, M. J. (1978). The evaluation of therapeutic outcomes. In Garfield, S. L., & Bergin, A. E. (Eds.), *Handbook of psychotherapy and behavior change: An empirical analysis* (2nd ed.) (pp. 139–190). New York: Wiley.
- Bloch, S., & Crouch, E. (1985). *Therapeutic factors in group psychotherapy*. New York: Oxford University Press.
- Bollen, K A., & Hoyle, R. H. (1990). Perceived cohesion: A conceptual and empirical examination. *Social Forces, 69*, 479–504.

- Bormann, B., & Straus, B. (2007). Gruppenklima, Kohäsion, Allianz und Empathie als Komponenten der therapeutischen Beziehung in Gruppenpsychotherapien—Überprüfung eines Mehrebenen-Modells. *Gruppenpsychotherapie und Gruppendynamik*, 43(1), 1–20.
- Braaten, L. J. (1989). Predicting positive goal attainment and symptom reduction from early group climate dimensions. *International Journal of Group Psychotherapy*, 39(3), 377–387.
- Braaten, L.J. (1990). The different patterns of group climate critical incidents in high and low cohesion sessions of group psychotherapy. *International Journal of Group Psychotherapy*, 40(4), 477–493.
- Brown, P. D., & O'Leary, K. D. (2001). Therapeutic alliance: Predicting continuance and success in group treatment for spouse abuse. *Journal of Consulting and Clinical Psychology*, 68, 340–345.
- Budman, S. H., Demby, A., Feldstein, M., Redondo, J., Scherz, B., Bennett, M. J., et al. (1987). Preliminary findings on a new instrument to measure cohesion in group psychotherapy. *International Journal of Group Psychotherapy*, 37(1), 75–94.
- Budman, S. H., Soldz, S., Demby, A., Feldstein, M., Springer, T., & Davis, M. S. (1989). Cohesion, alliance, and outcome in group psychotherapy. *Psychiatry*, 52, 339–350.
- Burlingame, G. M., Furhiman, A., & Johnson, J. (2002). Cohesion in group psychotherapy. In J. Norcross (Ed.), *A guide to psychotherapy relationships that work* (pp. 71–88). Oxford: Oxford University Press.
- Burlingame, G. M., Johnson, J. E., & MacKenzie, R. (2002, February). *Process measurement in group psychotherapy*. Paper presented at the annual meeting of the American

- Group Psychotherapy Association, New Orleans, LA.
- Burlingame, G. M., Kircher, J. C., & Ronts, C. R. (1994). Analysis of variance versus bootstrap procedures for analyzing dependent observations in small group research. *Small Group Research*, 25, 486–501.
- Burlingame, G. M., Kircher, J. C., & Taylor, S. (1994). Methodological considerations in group psychotherapy research: Past, present, and future practices. *Handbook of group psychotherapy: An empirical and clinical synthesis* (pp. 41–82). New York: Wiley.
- Burlingame, G. M., MacKenzie, K. R., & Strauss, B. (2004). Small group treatment: Evidence for effectiveness and mechanisms of change. In M. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change*. Wiley: NY.
- Burns, D. D., & Auerbach, A. (1996). Therapeutic empathy in cognitive-behavioral therapy: Does it really make a difference? In P. M. Salkovskis (Ed.), *Frontiers of cognitive therapy* (pp. 135–164). New York: Guilford.
- Burns, D. D., & Nolen-Hoeksema, S. (1992). Therapeutic empathy and recovery from depression in cognitive-behavioral therapy: A structural equation model. *Journal of Consulting and Clinical Psychology*, 59(2), 305–311.
- Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. New Jersey: Lawrence Erlbaum Associates.
- Carron, A. V., Widmeyer, W. N., & Brawley, C. R. (1985). The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire. *Journal of Sport Psychology*, 7, 244–266.

- Castongauy, L. G., Pincus, A. L., Agras, W. S., & Hines, C. E. (1998). The role of emotion in group cognitive-behavioral therapy for binge-eating disorder: When things have to feel worse before they get better. *Psychotherapy Research*, 8(2), 225–238.
- Cortez-Ison, E. A. (1997). *Perceptions of early parental bonding as related to group cohesion, working alliance and treatment outcome among female sexual abuse survivors in group psychotherapy*. Unpublished doctoral dissertation, University of California, Santa-Barbara.
- Cota, A. A., Dion, K. L., & Evans (1993). A reexamination of the structure of the Gross Cohesiveness Scale. *Educational and Psychological Measurement*, 53, 499–506.
- Cota, A. A., Longman, R. S., Evans, C. R., Dion, K. L., & Kilik, L. (1995). Using and misusing factor analysis to explore group cohesion. *Journal of Clinical Psychology*, 51(2), 308–306.
- Dion, K. L. (2000). Group cohesion: From "field of forces" to multidimensional construct. *Group Dynamics*, 4(1), 7–26.
- Doxsee, D. J. & Kivlighan, D. M. (1994). Hindering events in interpersonal relations groups for counselor trainees. *Journal of Counseling & Development*, 72, 621–626.
- Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. New Jersey: Lawrence Erlbaum Associates.
- Edwards, Q. T. (1999). *Member perceptions and the relationship between leader behavior, gender, and group climate*. Unpublished doctoral dissertation, University of Missouri, Columbia.

- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430–457.
- Evans, N. J., & Jarvis, P. A. (1986). The Group Attitude Scale: A measure of attraction to group. *Small Group Behavior, 17*, 203–216.
- Falloon, R. H. (1981). Interpersonal variables in behavioral group therapy. *British Journal of Medical Psychology, 54*(2), 133–141.
- Fenton, L. R., Cecero, J. J., Nich, C., Frankforter, T. L., & Carroll, K. M. (2001). Perspective is everything: The predictive validity of six working alliance instruments. *Journal of Psychotherapy Practice and Research, 10*(4), 262–268.
- Fischer, J., & Corcoran, K. (1994). *Measures for clinical practice: A sourcebook* (2nd Ed.). New York: The Free Press.
- Friedman, W. H. (1989). *Practical group therapy: A guide for clinicians*. San Francisco: Jossey-Bass.
- Fuehrer, A., & Keys, C. (1988). Group development in self-help groups for college students. *Small Group Behavior, 19*, 325–341.
- Fuhriman, A. & Burlingame, G. M. (1994). Group psychotherapy: Research and practice. In A. Fuhriman & G. M. Burlingame (Eds.), *Handbook of group psychotherapy: An empirical and clinical synthesis* (pp. 3–40). New York: Wiley.
- Fuhriman, A. & Burlingame, G. M. (2000). The Hill Interaction Matrix: Therapy through dialogue. In A. P. Beck & C. M. Lewis (Eds.), *The process of group psychotherapy: Systems for analyzing change* (pp. 135–174). Washington, DC: American Psychological Association.

- Fuhriman, A., Drescher, S., & Burlingame, G. (1984). Conceptualizing small group process. *Small Group Behavior, 15*, 427–440.
- Gaston, L. (1991). Reliability and criterion-related validity of the California Psychotherapy Alliance Scales—Patient Version. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*, 68–74.
- Gaston, L., Thompson, L., Gallagher, D., Cournoyer, L. G., & Gagnon, R. (1998). Alliance, technique, and their interactions in predicting outcome of behavioral, cognitive, and brief dynamic therapy. *Psychotherapy Research, 8*, 190–209.
- Gillaspy, J. A., Wright, A. R., Campbell, C., Stokes, S., & Adinoff, B. (2002). Group alliance and cohesion as predictors of drug and alcohol abuse treatment outcomes. *Psychotherapy Research, 12*, 213–229.
- Glatzer, H. T. (1978). The working alliance in analytic group psychotherapy. *International Journal of Group Psychotherapy, 28*, 147–161.
- Glatzer, H. T. (1990). Psychoanalytic group psychotherapy. In L. Kutash, & A. Wolf (Eds.), *The group psychotherapist's handbook: Contemporary theory and technique. Personality, psychopathology, and psychotherapy: Theoretical and clinical perspectives* (pp. 46–60). New York: Columbia University Press.
- Griffith, J. (1988). Measurement of group cohesion in U.S. Army units. *Basic and Applied Social Psychology, 9*, 149–171.
- Hatcher, R. L., Barends, A., Hansell, J., & Gutfreund, J. (1995). Patients' and therapists' shared and unique views of the therapeutic alliance: An investigation using confirmatory factor analysis in a nested design. *Journal of Consulting and Clinical Psychology, 63*, 636–643.

- Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. New Jersey: Lawrence Erlbaum Associates.
- Heinzel, R., Breyer, F., & Klein, T. (1998). Ambulante analytische Einzel und Gruppenpsychotherapie [Outpatient psychoanalytic therapy in Germany]. *Gruppenpsychotherapie und Gruppendynamik*, 34, 135–152.
- Henry, W. P., & Strupp, H. H. (1994). The therapeutic alliance as interpersonal process. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 51–84). New York: John Wiley & Sons.
- Horvath, A. O. (1994). Empirical validation of Bordin's pantheoretical model of the alliance: The Working Alliance Inventory perspective. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 109–130). New York: John Wiley & Sons.
- Horvath, A. O. & Greenberg, L. S. (1986). The development of the Working Alliance Inventory. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook. Guilford Clinical Psychology and Psychotherapy Series* (pp. 529–556). New York: The Guilford Press.
- Horvath, A. O. & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, 36, 223–233.
- Horvath, A. O. & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology*, 61, 561–573.
- Horvath, A. O. & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy. *Journal of Counseling Psychology*, 38, 139–149.
- Hox, J. (2002). *Multilevel Analysis: Techniques and Applications*. New Jersey: Lawrence Erlbaum Associates.

- Hurley, J. R. (1986). Leader's behavior and group members' interpersonal gains. *Group, 10*(3), 161–176.
- Hurley, J. R., & Brooks, L. J. (1987). Brief reports of group climate's principal dimension: Affiliation. *International Journal of Group Psychotherapy, 37*(3), 441–448.
- Hurley, J. R. & Brooks, L. A. (1988). Primacy of affiliativeness in ratings of group climate. *Psychological Reports, 62*, 123–133.
- Johnson, J. Burlingame, G. M., Olsen, D., Davies, 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.
- Kahn, E. M., Sturke T., & Schaeffer, J. (1992). Inpatient group processes parallel unit dynamics. *International Journal of Group Psychotherapy, 42*(3), 407–418.
- Kanas, N., & Barr, M. A., (1986). Process and content in a short-term inpatient schizophrenic group. *Small Group Behavior, 17*(3), 355–363.
- Kanas, N., & Ziegler, J. (1984). Group climate in a stress discussion group for medical interns. *Group, 8*(1), 35–38.
- Karterud, S. (1988). The influence of task definition, leadership, and therapeutic style on inpatient group cultures. *International Journal of Therapeutic Communities, 9*, 231–247.
- Kaul, T., & Bednar, R. (1986). Research on group and related therapies. In S. Garfield and A. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed.), (pp.671–714). New York: Wiley.
- Kay, S. (1996). *Effects of gender on small group cohesion: MacKenzie's Group Climate*

- Questionnaire analyzed.* Unpublished doctoral dissertation, Brigham Young University.
- Kilic, C., Özbayrak, K., Ulusahin, A., & Üstün, B. (1996). Therapeutic factors in interactional inpatient groups: Results from a Turkish sample. *Group, 20*(3), 241-249.
- Kipnes, D. R., Piper, W. E., & Joyce, A. S. (2002). Cohesion and outcome in short-term psychodynamic groups for complicated grief. *International Journal of Group Psychotherapy, 52*, 483–509.
- Kivlighan, D. M., Jr., & Angelone, E. O. (1992). Interpersonal problems: Variables influencing participants' perception of group climate. *Journal of Counseling Psychology, 39*(4), 468–472.
- Kivlighan, D. M., & Goldfine, D. C. (1991). Endorsement of therapeutic factors as a function of stage of group development and participant interpersonal attitudes. *Journal of Counseling Psychology, 38*(2), 150–158.
- Kivlighan, D. M. & Jauquet, C. A. (1990). Quality of group member agendas and group session climate. *Small Group Research, 21*, 205–219.
- Kleinberg, J. L. (1999). The supervisory alliance and the training of psychodynamic group psychotherapists. *International Journal of Group Psychotherapy, 49*(2), 159–179.
- Krupnick, J. L., Sotsky, S. M., Simmens, S., Moyer, J., Elkin, I., Watkins, J., et al. (1996). The role of the therapeutic alliance in psychotherapy and psychopharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of*

Consulting and Clinical Psychology, 64, 532–539.

- Lese, K. P. & MacNair-Semands, R. (2000). The Therapeutic Factors Inventory: Development of a scale. *Group*, 24(4), 303–317.
- Livingston, M. S. (1999). Vulnerability, tenderness, and the experience of self object relationship: A self psychological view of deepening curative process in group psychotherapy. *International Journal of Group Psychotherapy*, 49(1), 19–40.
- Luborsky, L. (1976). Helping alliances in psychotherapy. In J. L. Cleghorn (Ed.), *Successful psychotherapy* (pp. 92–116). New York: Brunner/Maze.
- MacKenzie, K. R. (1981). Measurement of group climate. *International Journal of Group Psychotherapy*, 31, 287–296.
- MacKenzie, K. R. (1983). The clinical application of group measure. In R. R. Dies & K. R. MacKenzie (Eds.), *Advances in group psychotherapy: Integrating research and practice* (pp. 159–170). New York: International Universities Press.
- MacKenzie, K. R. (1984). Commentary on "Process and content in an outpatient schizophrenic group," by Kanas, DiLella and Jones. *Group*, 8(3), 31–34.
- MacKenzie, K. R. (1998). The alliance in time-limited group psychotherapy. In J. D. Safran & J. C. Muran (Eds.), *The therapeutic alliance in brief psychotherapy* (pp. 193–215). Washington, DC: American Psychological Association.
- MacKenzie, K. R., Dies, R. R., Coche, E., Rutan, J. S., & Stone, W. N. (1987). An analysis of AGPA Institute groups. *International Journal of Group Psychotherapy*, 37(1), 55–74.
- MacKenzie, K. R., & Tschuschke, V. (1993). Relatedness, group work, and outcome in long-term inpatient psychotherapy groups. *The Journal of Psychotherapy*

- Practice and Research*, 2, 147–156.
- Marziali, E., Mumoe-Blum, H., & McCleary, L. (1997). The contribution of group cohesion and group alliance to the outcome of group psychotherapy. *International Journal of Group Psychotherapy*, 47(4), 475–497.
- Marziali, E., Mumoe-Blum, H., & McCleary, L. (1999). The effects of the therapeutic alliance on the outcomes of individual and group psychotherapy with borderline personality disorder. *Psychotherapy Research*, 9(4), 424–436.
- McBride, L. (1995). *Toward group process understanding: Leadership and group climate*. Unpublished doctoral dissertation, University of Missouri-Columbia.
- McCallum, M., Piper, W. E., Ogrodniczuk, J. S., & Joyce, A. S. (2002). Early Process and dropping out from short-term therapy for complicated grief. *Group Dynamics*, 6, 243–254.
- McLees, E., Margo, G. M., Waterman, S., & Beeber, A. (1992). Group climate and group development in a community meeting on a short-term inpatient psychiatric unit. *Group*, 16(1), 18–30.
- McRoberts, C., Burlingame, G. M., & Hoag, M. J. (1998). Comparative efficacy of individual and group psychotherapy: A meta-analytic perspective. *Group Dynamics*, 2(2), 101–117.
- Moos, R. F. (1974). *Evaluating treatment environments*. New York: Wiley.
- Muthen, L. K., & Muthen, B. O. (1998). *Mplus: The comprehensive modeling program for applied researchers: User's guide*. California: Muthen & Muthen.
- Muthen, L. K., & Muthen, B. O. (2002). *Mplus VERSION 2.12: Addendum to the Mplus user's guide*. California: Muthen & Muthen.

- Ogrodniczuk, J. S., & Piper, W. E. (2003). The effect of group climate on outcome in two forms of short-term group therapy. *Group Dynamics*, 7, 64–76.
- Orlinsky, D. E., Grawe, K., & Parks, B. K. (1994). Process and outcome in psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed.) (pp. 270–376). New York: Wiley.
- Orlinsky, D. E., Ronnestad, M. H., & Willutzki, U. (2003). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. L. Lambert (Ed.), *Bergin and Garfield handbook of psychotherapy and behavior change* (5th ed) (pp. 307–389). New York: Wiley.
- Persons, J., & Burns, D. D. (1985). Mechanism of action of cognitive therapy: Relative contribution of technical and interpersonal intervention. *Cognitive Therapy and Research*, 9(5), 539–551.
- 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.
- Pinsoff & Catherall (1986). The integrative psychotherapy alliance: Family, couple and individual therapy scales. *Journal of Marital and Family Therapy*, 12, 137–151.
- Piper, W. E., Marrache, M., Lacroix, R., Richardsen, A. M., & Jones, B. D. (1983). Cohesion as a basic bond in groups. *Human Relations*, 36, 93–108.
- Raskin, N. J. (1986). Client-centered group psychotherapy, part II: Research on client centered groups. *Person-Centered Review*, 1(4), 389–408.
- Roarck, A. E. & Sarah, H. S. (1989). Factors related to group cohesiveness. *Small Group Behavior*, 20(1), 62–69.

- Roback, H. B. & Smith, M. (1987). Patient attrition in dynamically oriented treatment groups. *American Journal of Psychiatry*, *144*(4), 426–431.
- Rogers, C. R., Gendlin, G. T., Kiesler, D. V., & Truax, L. B. (1967). *The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics*. Madison, Wisconsin: University of Wisconsin Press.
- Rugel, R. P. & Barry, D. (1990). Overcoming denial through the group: A test of acceptance therapy. *Small Group Research*, *21*, 45–58.
- Safran, J. D., Wallner, L. K. (1991). The relative predictive validity of the therapeutic alliance measures in cognitive therapy. *Psychological Assessment*, *3*, 188–195.
- Schain, J. (1989). The new infant research: Same implications far group therapy. *Group*, *13*, 112–121.
- Schreiber, M. (1987). Combined therapy. *Issues in Ego Psychology*, *10*(1), 74–77.
- Sexton, H. (1993). Exploring a psychotherapeutic change sequence: Relating process to intersessional and posttreatment outcome. *Journal of Consulting and Clinical Psychology*, *61*(1), 128–136.
- Shechtman, Z. (1993). Group psychotherapy for the enhancement of intimate friendship and self-esteem among troubled elementary-school children. *Journal of Social and Personal Relationships*, *10*, 483–494.
- Shields, W. (1999). Hope and the inclination to be troublesome: Winnicott and the treatment of character disorder in group therapy. *International Journal of Group Psychotherapy*, *50*(1), 87–103.
- Silbergeld, S., Koenig, G. R., Manderscheid, R. W., Meeker, B. F., & Hornung, C. (1975). Assessment of environment-therapy systems: The Group Atmosphere

- Scale. *Journal of Consulting and Clinical Psychology*, 43, 460–469.
- Stokes, J. P. (1983). Toward an understanding of cohesion in personal change groups. *International Journal of Group Psychotherapy*, 33(4), 449–467.
- Stone, W. N. & Gustafson, L. P. (1982). Technique in group psychotherapy of narcissistic and borderline patients. *International Journal of Group Psychotherapy*, 32(1), 29–47.
- Strauss, B., & Burgmeier-Lohse, M. (1995). The "fit" between therapist and patient as a determinant of treatment outcome following inpatient group psychotherapy. *Zeitschrift fuer Psychosomatische Medizin und Psychoanalyse*, 41, 127–140.
- Tracey, T. J. & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment*, 1, 207–210.
- Trad, P. V. (1993). Using the prospective approach as an adjunct to established models of group psychotherapy. *Group*, 17(1), 43–60.
- Tschuschke, V. (1999). Gruppentherapie versus Einzeltherapie—gleich wirksam? [Group versus individual psychotherapy—equally effective?]. *Gruppenpsychotherapie und Gruppendynamik*, 35, 257–274.
- Tschuschke, V. & Dies, R. R. (1994). Intensive analysis of therapeutic factors and outcome in long-term inpatient groups. *International Journal of Group Psychotherapy*, 44(2), 185–208.
- Tschuschke, V., & Greene, L. R. (2002). Group therapists' training: What predicts learning? *International Journal of Group Psychotherapy*, 52, 463–482.
- Walsh, J. (1994). Social support resource outcomes for the clients of two assertive community treatment teams. *Research in Social Work Practice*, 4(4), 448–463.

- Wessler, R. L. & Hankin-Wessler, S. (1989). Cognitive group therapy. In A. Freeman & K. M. Simon (Eds.), *Comprehensive handbook of cognitive therapy*. (pp. 559–581). New York: Plenum Press.
- Woody, S. R., & Adessky, R. S. (2002). Therapeutic alliance, group cohesion, and homework compliance during cognitive-behavioral group treatment of social phobia. *Behavior Therapy*, *33*, 5–27.
- Wright, T. L. & Duncan, D. (1986). Attraction to group, group cohesiveness, and individual outcome: A study of training groups. *Small Group Behavior*, *17*, 487.
- Yalom, I. (1995). *The theory and practice of group psychotherapy* (4th Ed.). New York: Basic Books