Italian adaptation of the Group Questionnaire: validity and factorial structure

Francesca Giannone,[1] Cinzia Guarnaccia,[2] Salvatore Gullo,[1] Maria Di Blasi,[1] Cecilia Giordano,[1] Gianluca Lo Coco,[1] Gary Burlingame[3]

[1]Department of Psychology, Educational Science and Human Movement, University of Palermo, Palermo, Italy; [2]Université Rennes 2, LP3C, Laboratoire de Psychologie: Cognition, Comportement, Communication, Rennes, France; [3]Department of Psychology, Brigham Young University, Provo, USA

ABSTRACT

The Group Questionnaire (GQ) is a measure recently developed by Krogel et al. (2013) for the evaluation of the therapeutic relationship in group. The GQ identifies a three-factor model of the relationship that allows to measure *quality* (Positive Bonding, Positive Working and Negative Relationship) and *structure* (member-member, member-leader and member-group), dimensions in group. This work shows the results of a first study on the Italian validation of the GQ. In this study the GQ was administered to 536 subjects from 32 non-clinical groups of undergraduate students. The cross-cultural validity of the GQ in the Italian population has been examined by comparing the psychometric properties and equivalence in factor structure and scores of the Italian GQ with the original American version. Multilevel confirmatory factor analysis was used to examine both the between- and within-group structures. Data concerning reliability and validity of GQ and the results for different SEM in Multilevel CFA confirm the three factors structure of the GQ. Data from the Italian population have a good fit with the original proposed model. Finally, we discuss the importance of an instrument like GQ, short but consistent, for the evaluation of the therapeutic relationship in clinical and training group.

Key words: Group Questionnaire; group process; group psychotherapy evaluation; therapeutic relationship; cross-cultural validity.

Introduction

Empirical literature agrees in identifying the therapeutic relationship as the main factor of change in group therapy (Burlingame, Strauss & Joyce, 2013; Johnson, Burlingame, Olsen, Davies, & Gleave, 2005; Johnson, Burlingame, Strauss, & Bormann, 2008, Marogna & Caccamo, 2014). Although there is substantial evidence that group treatment is effective with several disorders (Burlingame et al., 2018b, Burlingame, McClendon & Alonso, 2011; Burlingame et al., 2013), the underlying mechanisms by which patient change is achieved have several challenges. There is a structural complexity in the group relationship which may represent a challenge when investigating its development over time (Lo Coco, Gullo, Di Fratello, Giordano, & Kivlighan, 2016a; Lo Coco et al., 2019a). An individual group member develops a therapeutic relationship with the therapist, but also with the other group members and with the group-as-a-whole and most of previous research did not address this structural complexity by focusing mainly on the relationship between therapist and patient in the group (Lo Coco, Gullo & Kivlighan, 2012). Moreover, the disagreement about how to define the group therapeutic relationship has resulted in a variety of theoretical and empirical definitions and in a multiplicity of tools to measure the different components of the exchange in group. However, in past decades studies have emphasized the existence of high interrelations and overlapping group relational constructs (Johnson et al., 2005). These challenges highlight the necessity to develop empirically supported models to represent the constructs, moving the definition of a unified paradigm for understanding the therapeutic process, tak-

Correspondence: Francesca Giannone, University of Palermo, Department of Psychology, Educational Science and Human Movement, Viale delle Scienze, Edificio 15, 90128, Palermo, Italy.
E-mail: francesca.giannone@unipa.it
Tel.: +39 333 7422452.

Citation: Giannone, F., Guarnaccia, C., Gullo, S., Di Blasi, M., Giordano, C., Lo Coco, G., Burlingame, G. (2020). Italian adaptation of the Group Questionnaire: validity and factorial structure. *Research in Psychotherapy: Psychopathology, Process and Outcome, 23*(2), 133-144. doi: 10.4081/ripppo.2020.443

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

Received for publication: 17 December 2019.
Accepted for publication: 21 May 2020.

This work is licensed under a Creative Commons Attribution Non-Commercial 4.0 License (CC BY-NC 4.0).

©Copyright: the Author(s), 2020
Licensee PAGEPress, Italy
Research in Psychotherapy:
Psychopathology, Process and Outcome 2020; 23:133-144
doi:10.4081/ripppo.2020.443

ing into account the most relevant of the investigated variables (Bakali, Baldwin, & Lorentzen, 2009; Bakali, Wilberg, Hagtvet, & Lorentzen, 2010; Compare, Tasca, Lo Coco, Kivlighan, 2016; Giannone et al. 2019; Gullo et al., 2015; Lo Coco et al, 2016b).

The study of Johnson et al. (2005) was the first that analyzed the associations between the most studied group processes, conducting an explorative factor analysis on the most used tools in the empirical group literature. On this basis was developed a two-dimensional model, which offered a promising, parsimonious, and empirically based definition of the latent structure of group relationship (Burlingame et al., 2011). The model identified a latent relationship *quality* dimension with three factors (positive bonding, positive working, and negative relationship) crossed with a latent relationship *structure* dimension (member-member, member-leader and member to group) that describes the main relationships occurring in group treatments.

Quality refers to constructs as cohesion, therapeutic alliance, group climate and empathy as the main variables responsible for the success of treatment (Burlingame, Fuhriman, & Johnson, 2001; Burlingame, MacKenzie, & Strauss, 2004; Marziali, Munroe-Blum, & McCleary, 1997). In this three-factor model, the *positive bonding relationship factor* merges the constructs cohesion, engagement, and emotional bond, the *positive working relationship factor* encompasses the agreement on therapeutic tasks and goals and the *negative relationship factor* includes elements of conflict and empathic failure (Johnson et al., 2005; Krogel et al., 2013). *Structure* refers particularly to the group's structural properties, as mentioned, member to leader, member to member, and member to group (group as a whole) relationships (Krogel et al., 2013).

The theoretical relevance of this model resulted in the construction of a measure: the Group Questionnaire (GQ; Burlingame et al., 2017; Krogel et al., 2013) that assesses a complete and complex set of relationships developed in the group setting and that is at the same time indicator of the quality and of the structure of the relationships in the group. The GQ was created by a team of experienced group clinicians/researchers from Brigham Young University. The creation of the GQ relied heavily on both empirical data analysis and clinical judgment. Starting from the 60 items version (composed by questions selected from existing measures) used by Johnson et al. (2005) by a series of subsequent refinements, Krogel et al. (2013) obtained the GQ, a Likert self-report measure of therapeutic relationship in the groups. In the final version GQ is a short questionnaire of 30 items, filling in 5-10 minutes that provides information on the quality of the relationships in the group compared to 3 main areas *Positive Bonding*, *Positive Working*, and *Negative Relationship*. It also provides specific information about the group structure: on how the relationships in the group *(member - member, member - leader and member - group)* are relatively strong or weak. Member-leader and member-member items are presented in a parallel format, with member-group items following.

There are several studies which tested the Johnson's et al. (2005) model across US and Europe countries. Most of these studies have supported the quality dimension of the therapeutic relationship replicating the three-factor model, also demonstrating that adding the relationship structure (member-member, member-leader and member to group) improved the model fit. In particular, using the 30-item GQ, Bakali et al. (2009) and Krogel et al. (2013) obtained good fit indexes, consistent with the Johnson's et al. (2005) finding and indicated that the refined 30-item GQ adequately captures the original model proposed by Johnson with one minor structural change.

A German version of the GQ-30 has also been validated by Bormann, Burlingame and Strauss (2011). They analyzed 67 groups (453 members) from 15 psychiatric hospitals in Germany & Switzerland and obtained a best fit on the 3-factor model, confirming the Johnson model. Recently, Thayer and Burlingame (2014) provided support that the GQ is an empirically valid measure of the quality of the group therapeutic relationship on US data from 65 treatment groups at 4 university counseling centers and 1 community mental health clinic.

Janis, Burlingame and Olsen (2018) using all data collected with GQ on 2479 group members of 202 groups found that quality and structure models hold across settings and countries, but separate norms are required for different settings in order to provide tailored alerts to group leaders, based upon clinical population. It is worth noting that the characteristics of an Italian version of the GQ and its specific norms that can be used by clinicians as a group monitoring feedback strategy are still lacking in the Italian context. Group therapy in Italy has been increasingly adopted in mental health services for inpatients and outpatients, due to its favorable cost-effectiveness (Giannone, Giordano & Di Blasi, 2015), and it is delivered in both public and private settings following different theoretical approach, such as psychodynamic, interpersonal and cognitive-behavioral (*e.g.*, Lo Coco et al., 2012; Popolo et al., 2018). Thus, the examination of the psychometric properties of the GQ could provide a suitable tool for analyzing the group therapy process to both clinicians and group researchers, by further improving the practice-based evidence. To date, it is not clear whether cross-cultural differences between the Italian and North American cultures may influence the group process and its measurement with group members. Cross-cultural studies traditionally suggested that there are several differences between collectivistic and individualistic cultures, however nowadays the individualism is a value shared by American, British, Dutch, French and Italian population (Hofstede, 2001; Hofstede & Minkov, 2010). Thus, it can be challenging to study if the interpersonal group relations in the Italian context have the same structure of those in USA.

In recent years, the GQ has become the leading measure for providing therapists with feedback on client's progress in therapy. Since the three GQ quality scales show high ICC values demonstrating to be sensitive in capturing the effect-group on members, the GQ-30 has demonstrated good validity in monitoring client progress over time, by generating alerts when patients are *off-track* and leading clinicians to make adjustments to treatment (Burlingame, McClendon, & Yang, 2018a; Marmarosh, 2018). The feedback in therapy can be a useful strategy also in Italy in order to help clinicians to treat their patients more effectively and reduce the drop-out rates. Some good outcome measures are available in the Italian context (*i.e.*, the OQ-45 and CORE-OM) and a validated measure of group relationship is still needed.

Finally, in a more general perspective, in the last years there was a call for empirically-validated treatments in mental health (Kazdin, 2017) and both researchers and clinicians are highly demanding for sound and validated measures of therapy relationship in individual and group settings. The Italian validation of the GQ can also be a contribution in this direction.

Considering the attention that has recently been paid to the GQ by group psychotherapy research, the main purpose of this study is to test the Italian adaptation of the GQ 30 item version (Kroegel et al., 2013). More specifically, the present study aims to examine the internal consistency (see H1), the factor structure (see H2-H4) and the criterion validity of the GQ (see H5). The specific aims of the study are to test the following five hypotheses:

H1: The omega scores for the subscales of 30 item Italian GQ confirm a good internal consistency for the questionnaire;

H2: The hypothesized factor structure consisting of 3 second-order relationship quality subscales and 8 first-order relationship structure subscales (Figure 1) is confirmed by data from Italian sample;

H3: The three dimensions: member-member, member-leader, and member-group are necessary for the model to provide a good fit to the data and the model demonstrates poorer fit when these dimensions are not considered.

H4: The hypothesized factor model provides at least acceptable fit at both the between- and within group levels;

H5: The GQ shows significant correlations with other measures of group process (Alliance and Cohesion) supporting its criterion-related validity. Specifically, it was hypothesized that Bonding and Working scales of GQ would show positive associations with all the four California Psychotherapy Alliance Scale – Group Version (CALPAS) subscales as well as with the three subscales of Group/Member/Leader Cohesion Scale (GMLCS), vice versa it was hypothesized that the Negative scale of GQ would show negative associations with all the CALPAS and GLMCS subscales.

Materials and Method

Participants

The study involved 536 participants (88.1% females, 11.6 males, ranging in age from 20 to 66, with a mean of 23.16 and SD of 3.59) recruited among undergraduate students who were enrolled in 32 classes on group processes from bachelor's and master's degree in clinical psychology at the University of Palermo, who attended training groups focused on group dynamics during the course of their academic years. The initial number of participants in the groups was of 545, we removed from the dataset incomplete questionnaires and those with a very large number of missing data that did not allow the calculation of scores. Each participant involved in the study expressed written informed consensus. The average size of groups was of 17 subjects (ranged from 10 to 25 subjects), groups met weekly for 8 weeks for a total of 8 sessions. Sessions lasted 2 hours and 30 minutes. The theoretical approach taken in leading these groups was interpersonal and psychodynamic, with a focus on members' interpersonal themes in the here-and-now experience of the group (Yalom & Leczecs, 2005). The group leaders facilitated the communication between members by exploring the implicit references to the counselor or other group members to create a here-and-now experience. More specifically, this training practice allows students to learn and refine basic helping skills such as listening, feelings recognition and expression, and empathy.

Measures

The Group Questionnaire and two others self-report measures to evaluate different variables of the group process were administered by research assistants:

Group Questionnaire (GQ)

The 'forward-backward' procedure was applied to translate the 30-item GQ from English into Italian. Two general practitioners translated the questionnaire and these were backward translated into English by a professional translator (van de Vijver & Hambleton, 1996). The Italian translation has obtained final approval by the authors.

The 30 items were scored on a 7-point Likert scale ranging from 1 (not true at all) to 7 (very true) and provided scores on 3 quality factors: i) Positive Bonding (13 items), that evaluates the sense of belonging or attraction that a member has to the group, its members, and its leader(s) that creates a positive atmosphere and allows members to feel genuinely understood and appreciated; ii) Positive Working (8 items), that evaluates the ability of the group to agree upon and work toward treatment goals in an effective manner; iii) Negative Relationship (9 items), that evaluates lack of trust, genuineness, and understanding as well as distance that might exist between the group, its members, or its leaders.

The GQ is scored by summing items for each subscale, higher scores indicate a better bond and work relationship as well as higher negative relationship.

California Psychotherapy Alliance Scale – Group Version (CALPAS-G)

The CALPAS–G (Gaston & Marmar, 1994) is a parallel scale of the CALPAS patient version (CALPAS–P; Gaston, 1991), to evaluate the therapeutic alliance into the group. In this study, the term *therapist* was replaced by *group members* or *the group*. An example item is: *Did you feel that you were working together with the group members, that you were joined in a struggle to overcome your problems?* The CALPAS–G is a 12-item self-report measure scored on a 7-point Likert scale ranging from 1 (not at all) to 7 (very much so). The scale provides four sub-

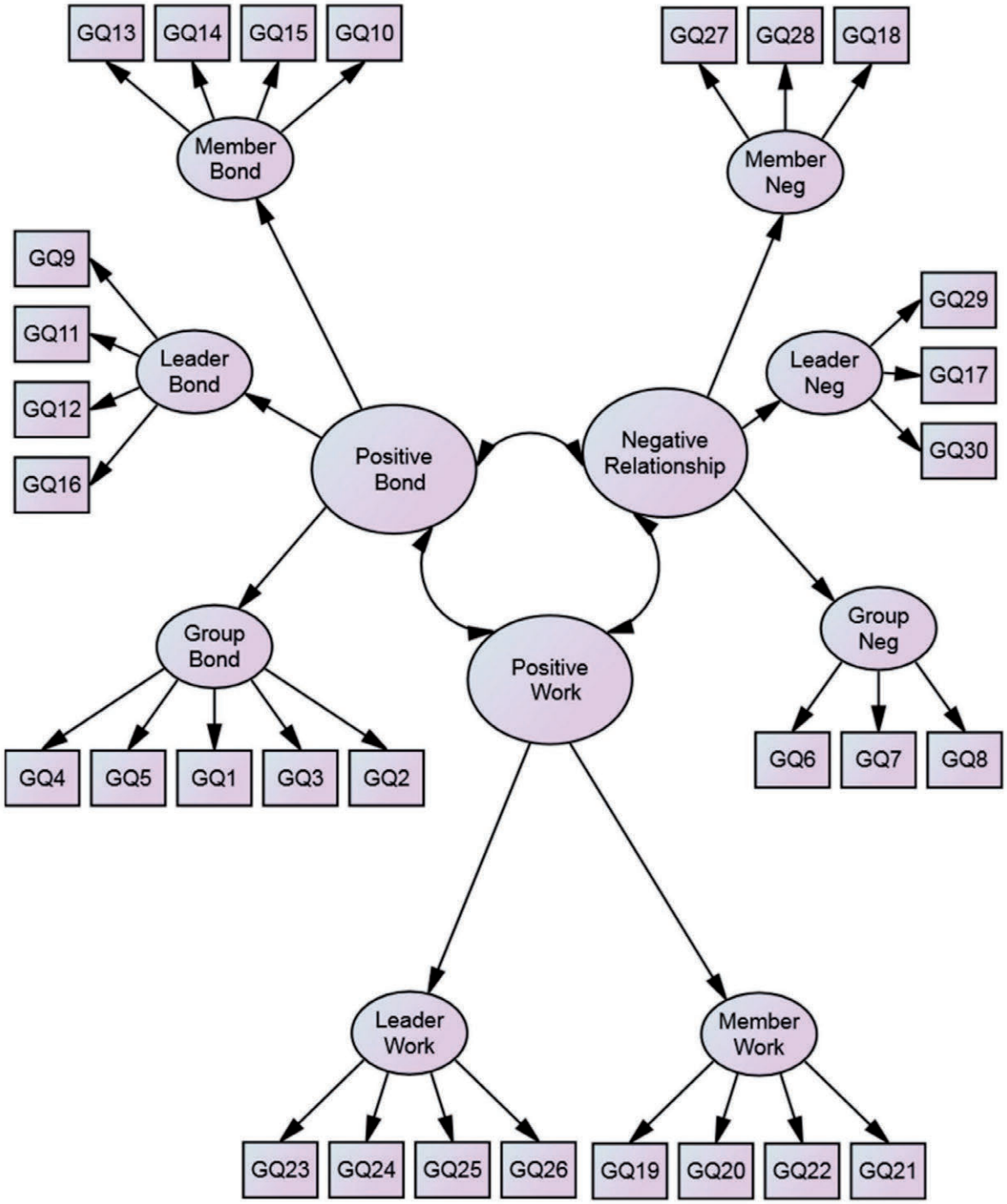


Figure 1. Factor Structure of GQ. Empty model.

scales: Patient Working Capacity (PWC), Patient Commitment (PC), Working Strategy Consensus (WSC), Member Understanding and Involvement (MUI). The CALPAS–G is scored by summing items for each subscale, higher scores indicate a better alliance with the group. An Italian version of the CALPAS-G was adopted (Kivlighan, Lo Coco, & Gullo, 2015) and, in the present study, average of inter-item correlation of the four CALPAS–G subscales were .40, .29, .34, .31, respectively. In order to test its factorial validity, we performed a CFA based on our data with the AMOS statistical package. We tested the four full correlated factors model proposed by authors and the hypothesized model appears to be a good fit to the data (χ^2=98.46, df =46, χ^2/df = 2.14, CFI = .94, RMSEA=.06).

Group/Member/Leader Cohesion Scale (GMLCS)

GMLCS (Piper, Marrache, Lacroix, Richardsen & Jones, 1983) is a self-report of 9 items, with a 6-point Likert scale scale ranging from 1 (very little) to 6 (very much). We used the version that evaluates the cohesion to the group. An example item is: *I think the group should meet more often*. This form provides scores on three subscales: Mutual Stimulation Effect (MSE); Commitment to the group (ComG); Compatibility of the group (CompG). The *GMLCS* is scored by summing items for each subscale, higher scores indicate a better cohesion in the group. In the present study average of inter-item correlation of the three GLMCS subscales were .20, .21 and .38, respectively. In order to test its factorial validity, we performed a CFA based on our data with the AMOS statistical package. We tested the three full correlated factors model proposed by authors and the hypothesized model appears to be a good fit to the data (χ^2=46.87, df =21, χ^2/df = 2.23, CFI = .95, RMSEA=.06).

Procedure

The measures were administrated at the end of the group session after a brief presentation and a research assistant was available to assist participants that need extra information. Participants, recruited on basis of voluntary participation, completed a form containing demographic information and an informed consent. According to the suggestions of Johnson et al. (2005) and by Krogel et al. (2013), data were collected at the end of the 3rd group session, when a relationship between the members and the group leader(s) had already formed. The study procedures were carried out in accordance with the Declaration of Helsinki and the project was approved by the university of the corresponding author.

Data analysis

After verifying the univariate normality of distributions using the Skewness and Kurtosis indices, we used the Kurtosis multivariate Mardia coefficient to demonstrate the multivariate normality of the variables (Barbaranelli, 2006). Then, we calculated descriptive statistics and correlations between variables.

The internal consistency of the CALPAS and GLMCS subscales was measured through the average of inter-item correlation given the small number of items (fewer than 5) in each of these scales. According to Clark and Watson (1995), average inter-item correlations should fall somewhere between .15 and .50 as anything below .15 would be too broad of a construct while anything above .50 would indicate redundancy of items on the scale.

Analysis strategy was driven by the aim of replicate the analyses conducted by Krogel et al. (2013) in order to have results that could be more directly comparable with those form the previous validation of the GQ.

A Confirmatory Factor Analysis (CFA) was performed using AMOS statistical software (Arbuckle & Wothke, 1999) to determine the dimensionality of GQ, CALPAS and GMLCS and to evaluate the goodness of fit between different models and the sample data. A CFA is a type of structural equation modeling, which is a special series of regression represented by a series of structural (*i.e.* regression) equations. A Full Information Maximum Likelihood estimation (FIML) was used to replace missing data and allowed the entire data set to be used. Multiple indicators of *goodness* are desirable when evaluating model fit. Three fit indices were used in this study: i) The Carmines–McIver Index: This index is the ratio of the x 2 to the degree of freedom (x 2/df). A result within the range of 2–3 indicates an acceptable fit (Carmines & McIver, 1981); ii) Comparative Fit Index (CFI) must be above .90 (Bentler, 1990); iii) Root Mean Square Error of Approximation (RMSEA) must be below .06 (Browne & Cudeck, 1993).

In order to verify the research hypotheses the following analytic strategy was used: for the hypotheses H2 and H3 two different models were created, the first *a priori* model contained both first and second order factors, - as depicted in Figure 1 the 30 items loaded into the 8 first order factors -, and these in turn loaded into the 3 second order factors (quality scales). The second model contained only the three second-order factors, with the 30 items loading directly on the three quality scales.

Finally, in order to verify the fourth hypothesis (H4) we made a multilevel factor analysis model with two sources of variation: between and within groups. The between-groups variation came from each group's mean score, while the within-group variation considered each group member's deviation (random component) from their own group's mean score.

In order to improve when necessary the fit to the data of the proposed models, modification indices greater than .10 were considered to modify the factor model.

The concurrent validity was assessed by calculating the Pearson correlation coefficient *r* between the GQ subscales and the other group process scales.

Results

Descriptive statistics and Internal Consistency of GQ scales

We checked the normality through univariate indices of Skewness and Kurtosis with acceptance threshold of ±1. No variables displayed violations of normality. The Kurtosis multivariate index of Mardia calculated on the 30 GQ item was equal to 188.7, below the critical cut-off of 960.

Table 1 reports descriptive statistics quality and structure GQ dimensions and internal consistency for the GQ quality subscales. In order to better evaluate internal consistency, McDonald's Omega was performed, the results shown Omega scores .90 for both Positive Bonding and Working scales and .79 for Negative Relationship scale. The results confirm H1 suggesting that the three subscales of 30-item Italian GQ have a good internal consistency. The cut scores for the Italian sample are also shown in Table 1.

As expected, (see Table 2), Positive Bonding and Positive Working were positively associated with each other and both were negatively associated with the Negative Relationship scale.

Dimensionality and Factorial Validity

Three out of the five hypotheses formulated (H2, H3 and H4) regard the factor structure of the GQ-30. Figure 1 shows the empty model that was tested. The hypothesis H2 tests whether the factor structure was confirmed with data from Italian sample. The first unmodified model retained only the presence of first and second order factors as specified in the theoretical model by Johnson and colleagues (2005). Since the model showed initial inadequate fit (χ^2=1774.88, df =396, χ^2/df = 4.48, CFI = .841, RMSEA=.081) refinements were made as indicated by modification indices (for all items that showed MI greater than .10) to improve the model. In the modified model, seven couples of item were allowed to covary, in particular covariance was added between those items (items 11/14, 12/15, 20/23, 21/26, 22/25, 27/29, 28/30) that were defined parallel, that is the same items that loaded on *Member* or *Leader* factor because their high degree of similarity in content (Krogel et al., 2013). This change determined a significant reduction of M.I. (no longer above .10 for all the items) and an improvement in all the fit indexes (χ^2=1099.12, df =387, χ^2/df = 2.84, CFI=.918, RMSEA=.059). The correlations and factor loadings between factors are depicted in Figure 1. In brief, CFA results confirmed the validity model proposed by Krogel et al. (2013) and replicates finding of previous GQ studies (Bormann et al., 2011; Thayer & Burlingame, 2014). The reduction of χ^2 value from the initial unmodified to the modified model resulted significant ($\Delta\chi^2$ = 675,76, Δdf= 9, p< .01), showing that the refined model is the preferable solution.

In order to verify the H3 hypothesis related to the GQ factor structure, a second model was tested. In this model, that comprised only the three second-order factors, the 30 items loaded directly on the three quality scales. The second unmodified model showed a poor fit with the data (χ^2 = 1794.07, df = 394, χ^2 /df= 4.55, CFI= .839, RMSEA= .081). On the basis of the modification indices (> .10), seven couples of item were allowed to covary, and particularly covariance was added between those items (items 11/14, 12/15, 20/23, 21/26, 22/25, 27/29, 28/30) which still showed an inadequate fit in the modified model (χ^2= 20018.48, df = 386, χ^2/df= 5.22, CFI= .812, RMSE= .089). Therefore, considering that both the first unmodified and the first modified models showed clear better fits in comparison with the second unmodified and the second modified models, respectively, it may be concluded that results provide support to the third hypothesis.

The hypothesis H4 deals with issues related to the

Table 1. Descriptive Statistics, Cut scores, Mc Donald's Omega and Intraclass coefficients for GQ subscales.

	Mean (SD)	Nof items	McDonald's Omega	Intraclass coefficients	Cut scores 10%	90%
GQ_PB	60.87 (11.18)	13	.908	.09	48	75
PB_Group	22.66 (4.90)			.21		
PB_Leader	20.23 (3.85)			-.06		
PB_Member	18.14 (3.88)			.03		
GQ_PW	34.23 (8.18)	8	.907	-.01	24	44
PW_Leader	17.65 (4.43)			-.01		
.PW_Member	16.60 (4.34)			.00		
GQ_NR	27.02 (7.72)	9	.798	-.05	16	37
NR_Group	10.02 (3.44)			-.06		
NR_Leader	7.06 (2.85)			-.03		
NR_Member	9.92 (3.40)			-.06		

N = 310; GQ PB, Positive Bonding; PW, Positive Working; NR, Negative Relationship.

group dependence effect. Recent research on GQ has demonstrated significant effect of group dependence among observations in rating the GQ scales as well as great variability. To measure magnitude of this group effect in our sample ICCs were calculated through a one-way ANOVA. The size of ICCs were moderate too low for the three quality scales ranging from .09 to -.05, in a similar way the ICCs were moderate too low for all the structure scales, ranging from .03 to -.06 with the exception of the GQ Bonding, which shows a high level of group dependence (ICC = .21) (see Table 1). Given the high variability of the ICC values, according to Janis et al. (2016) we established to treat our data using multilevel modeling that allows to test the factor structure of a measure while accounting for group dependence (see H4). Thus, the next step was to analyze the goodness of fit of the model as a whole that consider both the within and the between levels.

Results of the multilevel factor analysis model showed acceptable fit considering the traditional threshold (Hu & Bentler, 1999): $\chi^2 = 1098.22$, df =798, p< .001, χ^2/df=1.38, CFI=.926, TLI=.920 RMSEA=.037; $SRMR_{between}$=.245, $SRMR_{within}$=.081. Factor loadings showed significant regression paths of all items onto the latent variables as well as significant correlations between latent variables (Figures 2 and 3). However, the Root Mean Square Residual values showed that greater problems in fit raised from between group level, leading to suppose the small number of groups involved in the present study could have negatively affected these results.

Convergent validity

Finally, in order to test our last hypothesis (H5), convergent validity between GQ, CALPAS-G and GMLCS scales was also analyzed considering a subsample composed by 310 subjects who received all the measures. Descriptive for these measures are reported in Table 3.

In Table 2 the correlation coefficients show that most of the associations between the GQ scales and alliance and cohesion measures were significant and exceeds the threshold of .30. In particular, we highlight that the GQ Positive Bonding and Working scales were significantly and positively associated with all the four CALPAS subscales and the three GLMCS subscales. Both the GQ scales showed the highest (> .50) associations with the CALPAS Working Strategy Consensus (WSC) and Member Understanding and Involvement (MUI) scales, as well as with the GMLCS Compatibility of the group scale. The GQ Negative Relationship scale was significantly and negatively associated with three out of the four scales of the CALPAS and with two out of the three subscales of the GLMCS. The highest (< -.40) negative association was with the CALPAS Member Understanding and Involvement (MUI) scale. Overall, these results confirm the H5 hypothesis showing significant

Table 2. Pearson correlations between GQ, CALPAS and GMLCS subscales.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1.GMLCS_MSE																	
2.GMLCS_ComG	.434**																
3.GMLCS_CompG	.475**	.566**															
4.Calpas PWC	.292**	.117*	.161**														
5.Calpas PC	.289**	.376**	.433**	.096													
6.Calpas WSC	.506**	.396**	.523**	.143*	.470**												
7.Calpas MUI	.478**	.384**	.570**	.228**	.481**	.706**											
8.GQ_PB	.384**	.367**	.605**	.131*	.438**	.564**	.581**										
9.GQ_PW	.438**	.434**	.615**	.140*	.406**	.639**	.550**	.624**									
10.GQ_NR	-.079	-.181**	-.368**	-.095	-.349**	-.378**	-.451**	-.397**	-.309**								
11.PB_Group	.364**	.342**	.600**	.115*	.387**	.514**	.553**	.806**	.596**	-.348**							
12.PB_Leader	.341**	.347**	.439**	.105	.374**	.443**	.426**	.735**	.540**	-.297**	.479**						
13.PB_Member	.431**	.363**	.652**	.163**	.419**	.545**	.665**	.823**	.587**	-.410**	.717**	.639**					
14.PW_Leader	.407**	.419**	.542**	.156**	.405**	.560**	.452**	.558**	.921**	-.241**	.471**	.582**	.500**				
15.PW_Member	.382**	.372**	.592**	.104	.334**	.610**	.570**	.590**	.909**	-.332**	.629**	.407**	.589**	.686**			
16.NR_Group	.083	-.050	-.180**	.012	-.185**	-.168**	-.217**	-.153**	-.172**	.738**	-.210**	-.042	-.164**	-.097	-.230**		
17.NR_Leader	-.159**	-.230**	-.314**	-.100	-.317**	-.344**	-.307**	-.380**	-.304**	.740**	-.231**	-.463**	-.319**	-.332**	-.215**	.260**	
18.NR_Member	-.120*	-.149**	-.362**	-.134*	-.313**	-.370**	-.515**	-.396**	-.247**	.836**	-.360**	-.211**	-.468**	-.147**	-.319**	.414**	.507**

N = 310; **. Correlation is significant at the 0.01 level *. Correlation is significant at the 0.05 level. GMLCS_MSE, Mutual Stimulation and Effect; ComG, Commitment to the group; CompG, Compatibility with the group; CALPAS PWC, Patient Working Capacity; PC, Patient Commitment; WSC, Working Strategy Consensus; MUI, Member Understanding and Involvement; GQ PB, Positive Bonding; PW, Positive Working; NR, Negative Relationship.

correlations between the GQ scales and the other measures of group process (CALPAS-G and GMLCS) and support its criterion-related validity.

Discussion

In the current study we reported data on the psychometric properties of the Italian version of the GQ (Krogel et al, 2013). The results showed its good psychometric properties with Italian samples, further supporting this group measure as a valid and reliable instrument for the assessment of the group process.

The main goals of the current study were to verify the validity and the factor structure of the GQ questionnaire, in order to propose it as a useful tool for the evaluation of the group process.

We hypothesized that the data collected on the Italian population would confirm the factor structure proposed by the authors of the instrument. In particular, our results confirmed the hypothesis of a structure with related quality factors, given that the model with first and second order factors showed better fit of the data compared to the model that didn't consider the first-order (structure level). These values are consistent with previous findings (Bormann et al., 2011; Johnson et al., 2005; Krogel et al., 2013; Thayer & Burlingame, 2014) and confirm that multilevel confirmatory factor analysis is a valuable technique for evaluating factor structures of measures used in grouped data (Janis, Burlingame, & Olsen, 2016). However, some weakness in the results should be highlighted, first of all the addition of covariance between measurement errors of several items indicates that there is a shared variance between those items that is not captured by the latent variables. Moreover, our results showed a poorer fit of the model at the between group level, and this finding seems in line with those by Janis et al., (2016). Thus, our findings provide further support to the importance of looking at the factor structure of the GQ by separating the within and between levels. It seems that the GQ is a more powerful tool for analyzing how group members differ from each other in their perceptions of group relationship, whereas it shows some weaknesses in providing information about how groups differ from each other. Further research involving a higher number of groups is needed to further test the robustness of the model fit of the GQ at the between level.

Table 3. Descriptive Statistics for CALPAS and GMLCS subscales.

	Mean	SD
GMLCS_MSE	10.07	2.579
GMLCS_ComG	14.05	2.253
GMLCS_CompG	12.39	2.417
Calpas PWC	9.76	3.215
Calpas PC	15.00	3.107
Calpas WSC	11.45	3.150
Calpas MUI	11.84	3.105

N= 310; GMLCS_MSE, Mutual Stimulation and Effect; ComG, Commitment to the group; CompG, Compatibility with the group; CALPAS PWC, Patient Working Capacity; PC, Patient Commitment; WSC, Working Strategy Consensus; MUI, Member Understanding and Involvement.

Concerning the second aim of this study, we tested the validity of GQ comparing the questionnaire scores with those obtained by the same subjects on other measures typically used for the validation of the group process, in particular therapeutic alliance and group cohesion. We found high correlations between the GQ and the other measures in all subscales which confirm that the GQ can be used effectively to detect these constructs. Although there is evidence of some overlap in measures of group relationship, therapeutic alliance and cohesion (Burlingame et al., 2018a; Johnson et al., 2005; Lo Coco et al., 2019b), the findings of the current study further supported the robustness of the construct of group relationship and the validity of its measurement, accordingly with previous studies conducted in USA (Janis et al., 2016; Kroegel et al., 2013).

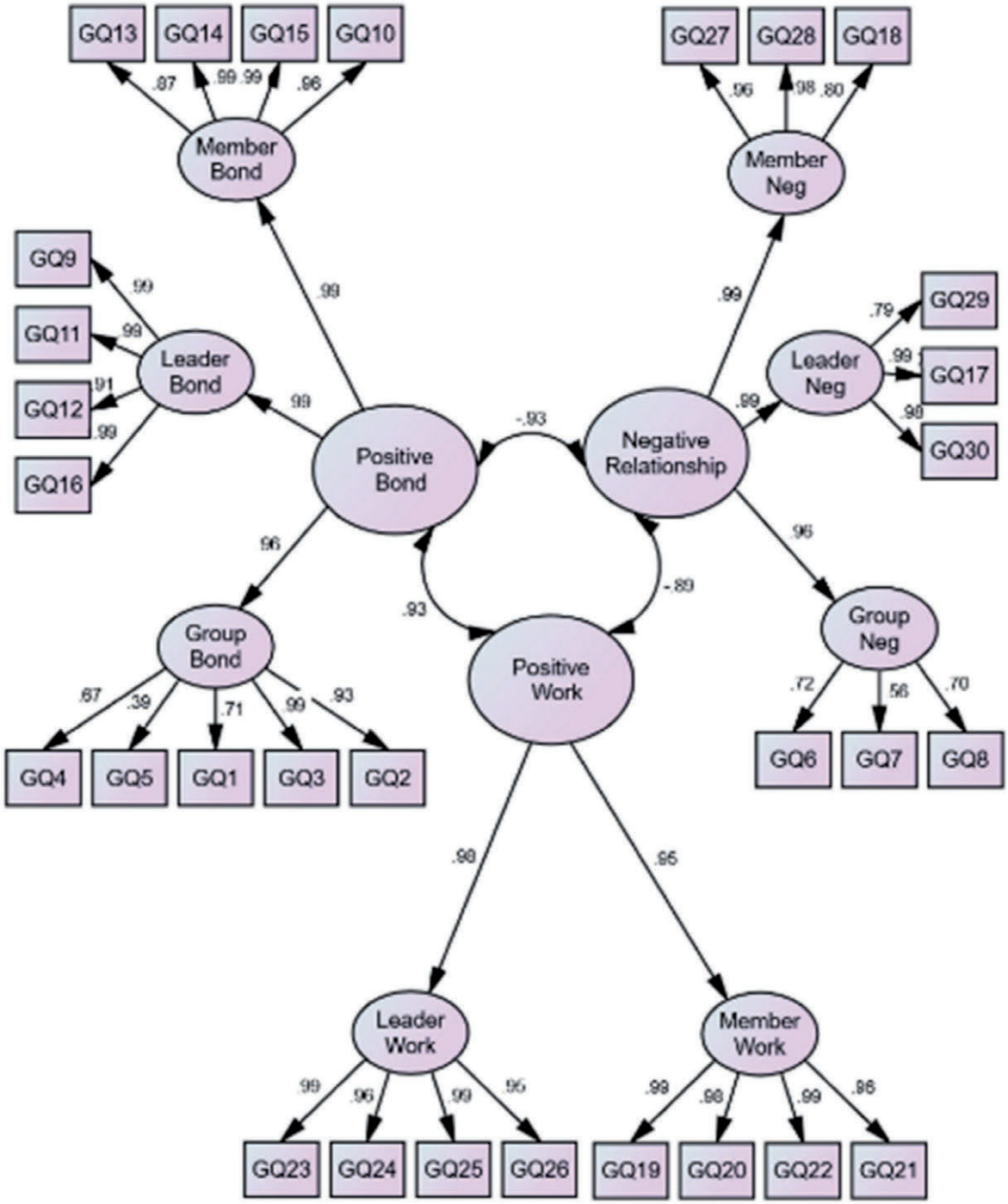
In conclusion, the GQ differs from the other instruments used to evaluate group therapy as it is based on a unified and empirically supported theory about the relationship of the group. The evidence gathered in all recent studies about goodness-of-fit of the factorial structure and validity of the psychometric properties lead to consider the GQ a reliable measure to evaluate the relationships in group therapy. The Italian validation of the GQ further confirms the validity and the factor structure of the GQ questionnaire.

The main advantage of this instrument is to overcome the traditional fragmentation between constructs, typical of the group psychotherapy researches, by proposing a unified view of the group process based on the concept of *relationship*. The facility of administration of the GQ, the useful information for clinician that it provides and the possibility to analyze in detail group relationships looking at the quality and structure (*i.e.*, Kivlighan et al., 2017; Lo Coco et al., 2016a), are the key elements that led to using the GQ in a number of researches that monitor group process and patient's changes during treatment as well as in training practices for group psychotherapists (Marmarosh, 2018). Moreover, several studies have also used member's GQ feedback to support clinical judgment in enhancing psychotherapy outcomes (Griner et al., 2018) showing reduced rates of relationship deterioration and failure when progress feedback was held constant (Burlingame et al., 2018a). Through the GQ, group members report their group experiences, making information that could facilitate the conduction in order to achieve greater effectiveness of interventions.

Therefore, the creation of GQ and its adaptation in

various cultural contexts represent a significant contribution for both researchers involved in research on psychotherapy and clinicians who lead psychotherapy groups, supporting the possibility to increase the effectiveness of the therapeutic work and the knowledge of group process into psychotherapy training programs.

Although this study contributes to better understanding of quality and structure of groups and processes into groups, some limitations need to be mentioned. The first limitation is given by the reference population: we used a mainly community sample, so replications with groups of clinical people are needed to examine whether observed relationships can be generalized to more severe clinical populations. The inclusion of further normative data from patient population could also help to determine the clinical interpretation of GQ scores. A second limitation was

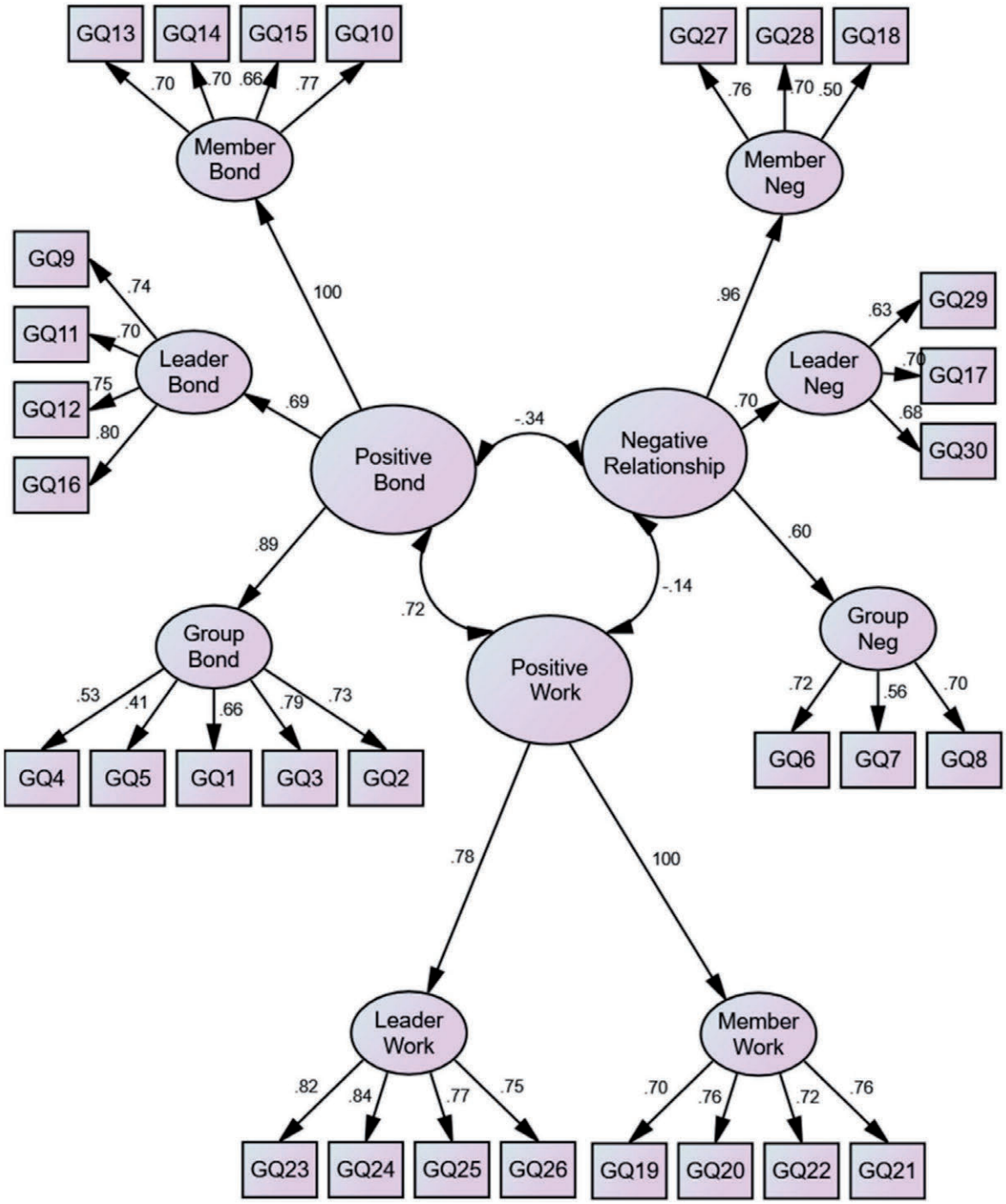


$\chi 2$ (798) = 1098.22, CFI=.926, TLI=.920 RMSEA=.037; SRMRbetween=.245; N= 310; all standardized coefficients are significant ($p < .001$).

Figure 2. Factor Structure of GQ. Between groups.

that all the data was gathered through self-reporting measures and the exclusive use of this type of measures should be complemented with other sources of information to produce conclusive findings, thus future replication using multi-informant methods is warranted.

Its limitations notwithstanding, the Italian validation of the GQ can provide group clinicians and trainees with a valid measurement of the quality of group member's therapeutic relationships offering reliable and non-redundant information across the three main structural dimensions of the group relationships (member-member, member-leader, and member-group).



$\chi 2$ (798) = 1098.22, CFI=.926, TLI=.920 RMSEA=.037; SRMRwithin =.081; N= 310; all standardized coefficients are significant (p< .001).

Figure 3. Factor Structure of GQ. Within groups.

References

Arbuckle, J.L., & Wothke, W. (1999). *AMOS 4.0 user's guide*. Chicago: Smallwaters.

Bakali, J.V., Baldwin, S.A., & Lorentzen, S. (2009). Modeling group process constructs at three stages in group psychotherapy. *Psychotherapy Research*, *19*(3), 332-343. doi:10.1080/10503300902894430

Bakali, J.V., Wilberg, T., Hagtvet, K. A., & Lorentzen, S. (2010). Sources accounting for alliance and cohesion at three stages in group psychotherapy: Variance component analyses. *Group Dynamics: Theory, Research, and Practice*, *14*(4), 368. doi:10.1037/a0019170

Barbaranelli, C. (2006), *Analisi dei Dati: Tecniche Multivariate per la Ricerca Sociale*, Milan: LED Edizioni.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bullettin, 107,* 238-246.

Bormann, B., Burlingame, G. & Strauss, B. (2011) Der Gruppenfragebogen (GQ-D) Instrument zur Messung von therapeutischen Beziehungen in der Gruppenpsychotherapie. *Psychotherapeut 56,* 297-309. doi:10.1007/s00278-011-0841-4

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen, J.S. Long (eds.), *Testing structural equation models.* Newbury Park, CA: Sage, pp.136-162.

Burlingame, G. M., Fuhriman, A., & Johnson, J. E. (2001). Cohesion in group psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, *38*(4), 373. doi:10.1037/0033-3204.38.4.373

Burlingame, G. M., MacKenzie, K. R., & Strauss, B. (2004). Small group treatment: Evidence for effectiveness and mechanisms of change. In M. J. Lambert (Eds), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change (5th ed., pp.* 647-696).

Burlingame, G. M., McClendon, D. T., & Alonso, J. (2011). Cohesion in group therapy. *Psychotherapy*, *48*(1), 34. doi:10.1037/a0022063

Burlingame, G. M., Strauss, B., & Joyce, A. S. (2013). Change mechanisms and effectiveness of small group treatments. In M. J. Lambert (Eds), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (6th ed., pp. 640-689).

Burlingame, G., Gleave, R., Beecher, M., Griner, D., Hansen, K., & Jensen, J. (2017). *Administration and Scoring Manual for the Group Questionnaire—GQ*. Salt Lake City, UT: OQ Measures.

Burlingame, G., McClendon, D., & Yang, C. (2018a). Cohesion in Group Therapy: A Meta-Analysis. *Psychotherapy,* 55, 4, 384–398. doi:10.1037/pst0000173

Burlingame, G., Whitcomb, K., Woodland, S., Olsen, J., Beecher, M., & Gleave, R. (2018b). The Effects of Relationship and Progress Feedback in Group Psychotherapy Using the Group Questionnaire and Outcome Questionnaire– 45: A Randomized Clinical Trial. *Psychotherapy 55*, 2, 116–131. doi:10.1037/pst0000133

Carmines, E., & McIver, J. (1981). Analyzing Models with Unobserved Variables: Analysis of Covariance Structures. *Social Measurement: Current Issues*, eds. G. Bohrnstedt and E. Borgatta, Beverly Hills, CA: Sage, pp. 61–73.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319. doi:10.1037/1040-3590.7.3.309

Compare, A., Tasca, G., Lo Coco, G., Kivlighan, D.M. (2016). Congruence of Group

Therapist and Group Member Alliance Judgments in Emotionally Focused Group Therapy

for Binge Eating Disorder. *Psychotherapy, 53*, 163–173. doi:10.1037/pst0000042

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*(1), 68. doi:10.1037/1040-3590.3.1.68

Gaston, L., & Marmar, C. (1994). The California psychotherapy alliance scales. *The working alliance: Theory, research and practice*, 85-108.

Giannone, F., Giordano, C., & Di Blasi, M. (2015). Group Psychotherapy in Italy. *International Journal of Group Psychotherapy, 65*, 501–511. doi:10.1521/ijgp.2015.65.4.501

Giannone, F., Lo Cascio, M., Gullo, S., Ferraro, A.M., Infurna, M.R., Guarnaccia, C. (2019). *La Relazione Terapeutica nei Gruppi. I Principali Costrutti del Processo Terapeutico: Definizioni Empiriche e Strumenti.* F. N. Vasta, S. Gullo, R. Girelli (eds). Psicoterapia Psicodinamica di Gruppo e Ricerca Empirica. Una Guida per il Clinico. Roma: Alpes.

Griner, D., Beecher, M., Brown, L., Millet, A., Worthen, V., Boardman, R., Hansen, K. Cox, J., & Gleave, R. (2018). Practice-Based Evidence Can Help! Using the Group Questionnaire to Enhance Clinical Practice. *Psychotherapy, 55,* 2, 196–202. doi:10.1037/pst0000136

Gullo, S., Lo Coco, G., Di Fratello, C., Giannone, F., Mannino, G., & Burlingame, G. (2015). Group climate, cohesion and curative climate. A study on the common factors in group process and their relation with members attachment dimensions. *Research in Psychotherapy: Psychopathology, Process and Outcome, 18*(1), 10-20. doi:10.4081/rippo. 2015.160

Hofstede, G. (2001). Culture's consequences: Comparing values, behaviors, institutions and organizations across nations (2nd ed.). Beverly Hills, CA: Sage.

Hofstede, G., & Minkov, M. (2010). Long-versus short-term orientation: new perspectives. *Asia Pacific business review*, *16*(4), 493-504. doi:10.1080/13602381003637609

Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling*, *6,* 1-55. doi:10.1080/10705519909540119

Janis, R. A., Burlingame, G. M., & Olsen, J. A. (2016). Evaluating Factor Structures of Measures in Group Research: Looking Between and Within. *Group Dynamics: Theory, Research, and Practice. 20*, (3), 165-180. doi:10.1037/gdn0000043

Janis, R. A., Burlingame, G. M., & Olsen, J. A. (2018). Developing a therapeutic relationship monitoring system for group treatment. *Psychotherapy, 55,* 105–115. doi:10.1037/pst0000139

Johnson, J. E., Burlingame, G. M., Strauß, B., & Bormann, B. (2008). Die therapeutischen Beziehungen in der Gruppenpsychotherapie. *Gruppenpsychotherapie und Gruppendynamik*, *44*(1), 52-89. doi:10.13109/grup.2008.44.1.52

Johnson, J. E., Burlingame, G. M., Olsen, J. A., Davies, D. R., & Gleave, R. L. (2005). Group Climate, Cohesion, Alliance, and Empathy in Group Psychotherapy: Multilevel Structural Equation Models. *Journal of Counseling Psychology*, *52*(3), 310. doi:10.1037/0022-0167.52.3.310

Kivlighan, D. M., Lo Coco, G., Gullo, S. (2015). Is there a group

effect? It depends on how you ask the question: Intraclass correlations for California Psychotherapy Alliance Scale–Group items. *Journal of Counseling Psychology, 62*(1), 73-78. doi.10.1037/cou0000046

Kivlighan, D., Lo Coco, G., Oieni, V., Gullo, S., Pazzagli, C., & Mazzeschi, C. (2017). All Bonds Are Not the Same: A Response Surface Analysis of the Perceptions of Positive Bonding Relationships in Therapy Groups. *Group Dynamics: Theory, Research, and Practice, 21*, 159-177. doi:10.1037/gdn0000071

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

Lo Coco, G., Gullo, S., & Kivlighan, D. M., Jr. (2012). Examining patients' and other group members' agreement about their alliance to the group as a whole and changes in patient symptoms using response surface analysis. *Journal of Counseling Psychology, 59*, 197–207 .doi:10.1037/a0027560

Lo Coco, G., Gullo, S., Di Fratello, C. Giordano, C. & Kivlighan, D. M. Jr. (2016a). Group Relationships in Early and Late Sessions and Improvement in Interpersonal Problems. *Journal of Counseling Psychology, 63*(4), pp. 419-428. https://doi.org/10.1037/cou0000153

Lo Coco, G., Gullo, S., Oieni, V., Giannone, F., Di Blasi, M., & Kivlighan Jr, D. M. (2016b). The relationship between attachment dimensions and perceptions of group relationships over time: An actor–partner interdependence analysis. *Group Dynamics: Theory, Research, and Practice*, *20*(4), 276–293. doi:10.1037/gdn0000056

Lo Coco, G., Gullo, S., Profita, G., Pazzagli, C., Mazzeschi, C., & Kivlighan, D. M., Jr. (2019a). The Codevelopment of Group Relationships: The Role of Individual Group Member's and Other Group Members' Mutual Influence and Shared Group Environment. *Journal of Counseling Psychology*, *66*, 640-649. doi:10.1037/cou0000349

Lo Coco, G., Tasca, G. A., Hewitt, P., Mikail, S. F., & Kivlighan, Jr., D. M. (2019b). Ruptures and repairs of group therapy alliance. An untold story in psychotherapy research. *Research in Psychotherapy: Psychopathology, Process and Outcome, 22*(1), 58-70. doi:10.4081/ripppo.2019.352

Marogna C., & Caccamo F. (2014). "Analysis of the process in brief psychotherapy group: the role of therapeutic factors". *Research in Psychotherapy, Psychopathology, Process and Outcome, 17*(1), 43-51. doi:10.4081/ripppo.2014.161

Marmarosh, C. (2018). Introduction to Special Issue: Feedback in Group Psychotherapy, *Psychotherapy, 55,* 101–106. doi:10.1037/pst0000178

Marziali, E., Munroe-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*, 475-497. doi:10.1080/00207284.1997.11490846

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. doi:10.1177/001872678303600201

Popolo, R., MacBeth, A., Canfora, F., Rebecchi, D., Toselli, C., Salvatore, G., & Dimaggio, G. (2018). Metacognitive Interpersonal Therapy in group (MIT□G) for young adults with personality disorders: A pilot randomized controlled trial. *Psychology and Psychotherapy: Theory, Research and Practice*, *92*(3), 342-358. doi:10.1111/papt.12182

Thayer, S. D., & Burlingame, G. (2014). The Validity of the Group Questionnaire: Construct Clarity or Construct Drift? *Group Dynamics: Theory, Research, and Practice*. *18*, 318–332. doi:10.1037/gdn0000015

van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*, 89-99. doi:10.1027/1016-9040.1.2.89

Yalom, I. D., & Leszcz, M. (2005). *The theory and practice of group psychotherapy (5th ed.).* New York: Basic Books.