

Examining Within-Group Variability in Group Therapy

Mira An and Dennis M. Kivlighan Jr.

Department of Counseling, Higher Education, and Special Education,
University of Maryland College Park

Objective: The present article illustrates a method for examining changes in group cohesion (i.e., group uniformity indicated by within-group variability) across time in group therapy research using multilevel data. **Method:** We used McNeish's (2021) three-level location-scale model for predicting both group-mean and within-group variability of a dependent variable in the framework of a Bayesian multilevel structural equation modeling. We reviewed the existing group research that addressed within-group variability and provided an example by applying this method to archival three-level longitudinal group therapy data. **Results:** We found that the location-scale model allows us to explore how the group-mean and within-group variability of group engagement would differ across time and whether group size (GSZ) would predict the changes in the group-mean and within-group variability of group engagement. **Conclusions:** Future applications in group therapy research as well as the strengths and limitations of the location-scale model were discussed. We hope this article familiarizes group psychotherapy researchers with an effective method for understanding multilevel and complex group uniformity.

Highlights and Implications

- The location-scale model was an effective method for understanding complex within-group variability, which indicates group cohesion.
- The within-group cohesion increased (i.e., variability in group members' perceptions of group engagement decreased) over time in general.
- The group size did not influence the initial level of group cohesion or changes in group cohesion over time.

Keywords: group cohesion, within-group variability, group therapy, group methods

Supplemental materials: <https://doi.org/10.1037/gdn0000198.supp>

Christina is curious about the engagement in the two groups she leads; therefore, she assesses each member's perceptions of group engagement after their first sessions. The two groups exhibit the same average level of engagement, but they are different in the distribution of members' engagement ratings. Members in one group are discrepant

in their engagement ratings, with some rating their climate as very engaged and others rating it as barely engaged. The other group has members who all rated their group climate as moderately engaged. Are the two groups the same in-group engagement because they have the same mean engagement or different because their distributions of engagement ratings differ? Christina as a group leader is curious about what brings about the difference in the distribution of engagement ratings within groups. Would it be possible because of their different group sizes (GSZs)? Near the end of the group, Christina is interested if group members became more consistent (cohesive) or inconsistent in perceptions of engagement over

This article was published Online First January 19, 2023.

Mira An  <https://orcid.org/0000-0002-4859-9919>

Correspondence concerning this article should be addressed to Mira An, Department of Counseling, Higher Education, and Special Education, University of Maryland College Park, 3214 Benjamin Building, College Park, MD 20710, United States. Email: miraan@umd.edu

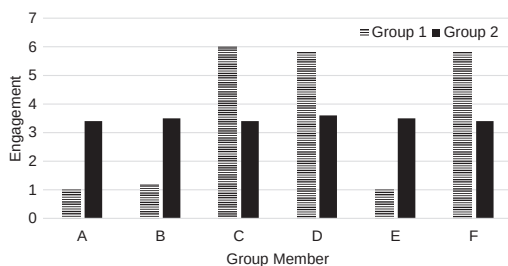
the course of group therapy. In sum, Christina is interested in how her groups differ in both the level (group average) and distribution of engagement ratings and how these differences come about.

Christina's observations about the distribution of engagement ratings are linked to the concept of within-group variance, which is a practically and theoretically important construct. Histograms in Figure 1 show two hypothetical groups' distribution of engagement ratings on the Group Climate Questionnaire (MacKenzie, 1983). Both groups have equivalent means (Group 1's $M = 3.47$; Group 2's $M = 3.47$); however, standard deviations for the two are quite different (Group 1's $SD = 2.40$; Group 2's $SD = .07$). Historically, researchers operationalized the group effect as the mean of the group members' ratings. Therefore, these two groups' engagement is treated as equivalent in analyses due to their identical mean engaged ratings. However, it seems obvious that the two groups do not have the same engagement because of the difference in SD s. This article builds on the theoretical speculations that low within-group variances can be considered as meaningful indices of group cohesion (Deutsch, 1954; Kivlighan, 2021) and introduces an effective way to explore within-group variability to advance group therapy research, which has been hindered by conceptual disagreements about the meaning of cohesion and methodological limitations in measuring

Group Cohesiveness as Low Within-Group Variability

Group cohesiveness is a mutative group factor in group therapy that facilitates greater collective

Figure 1
Groups With a Similar Average but Different Standard Deviations for Engagement



Note. $M = 4.47$, $SD = 2.40$, $n = 6$ for Group 1; $M = 4.47$, $SD = .07$, $n = 6$ for Group 2.

self-esteem, hope, and well-being (Yalom, 1995). It is also the “necessary precondition for effective therapy” (Yalom, 1995, p. 50). Theorists and researchers thus have consistently emphasized the importance of cohesiveness in group therapy (Burlingame et al., 2011, 2018; Kivlighan, 2021; Marmarosh & Van Horn, 2010). For example, Bakali et al.’s (2010) exploration of the role of group cohesion in three stages of group therapy revealed that members of more cohesive groups showed more regular session attendance and were more engaged with the groups in later stages of the group therapy. Similarly, a meta-analytic review of 55 group therapy studies demonstrated the significant influence of cohesion on treatment outcomes (Burlingame et al., 2018). Despite the clinical importance of group cohesion and considerable research on this concept, there has been a lack of consensus on the definition and measurement of group cohesion, which has hindered valid and thorough investigations of this aspect of group process (Budman et al., 1989; Burlingame et al., 2018).

Van Bergen and Koekebakker (1959) stated the essence of cohesiveness is the degree of unification of the group field, which is one of the oldest, but least studied the definition of cohesion. The concept of cohesion comes from the idea of glueyness, which implies that members of a group stick together. Even though theorists and researchers agree that group cohesion is best conceptualized as the degree of unification of a group, unification has been understudied due to an absence of methods to measure the group unification itself (Forsyth, 2021). Instead, researchers have strived to grasp group cohesion by assessing indicators of group cohesion (Forsyth, 2021). For instance, self-report measures of group cohesion, such as the Harvard Group Cohesiveness Scale (Budman et al., 1989), Perceived Cohesion Scale (Bollen & Hoyle, 1990), and the Therapeutic Factor Inventory–Cohesion subscale (Lese & MacNair-Semands, 2000), assess levels of involvement, trust, cooperation, caring, and belonging. Forsyth (2021) criticized this multicomponent approach in that these multiple components might be confounders of group cohesion, which might be different from the degree of group unity itself. The degree of group unification can be operationalized using within-group variability (i.e., low within-group variability indicating high group uniformity/cohesion), but few researchers have

used this measure. To better understand group cohesiveness, we need to examine how group members' perceptions, attitudes, and behaviors are uniform or discrepant across members within a group.

Within-Group Variability to Measure Group Cohesiveness

Only a few researchers have explored within-group variability. One approach used the within-group standard deviation of group members' ratings in addition to the mean of group members' ratings as predictors (Kivlighan et al., 2017; Miles et al., 2011; Schneider et al., 2002). For example, Schneider et al. (2002) used the within-group standard deviation of banking team members' customer service climate ratings to define climate strength. Drawing upon Mischel's (1976) conceptualization of situational strength, Schneider et al. (2002) defined the strength of group climates using within-group standard deviations. Strong climates refer to the cases in that group members perceive climate similarly (low *SD*), with uniform behavioral norms; whereas weak climates indicate group members perceive the climate in different ways (high *SD*), and behavioral norms in the group are inconsistent or even nonexistent (Schneider et al., 2002). Using this conceptualization, Schneider et al. (2002) found a significant interaction between climate level and climate strength (interaction of mean level and standard deviation of services climate within a banking team). That is when teams had strong service climates (high climate strength; low *SDs*), average levels of service climate (*Ms*) had a positive relationship with customer satisfaction.

Similarly, Miles et al. (2011) conceptualized a group behavioral norm composed of two components: the average amount of intimate behaviors shown in a group and the consistency (*SD*) of intimate behaviors across group members. Using the actor-partner interdependence model (Kenny et al., 2002), Miles et al. (2011) explored how a group member's intimate behavior was predicted by (a) the actor effect for the group member's intimate behaviors and (b) partner effects for the average amount (group mean) and (c) consistency (*SD*) of other group members' intimate behaviors and (d) the interaction of the amount and consistency. Findings indicated the interaction for the amount and consistency of a group's intimate

behaviors predicted a group member's intimate behaviors in the next session (Miles et al., 2011). Specifically, when the consistency of intimate behaviors of the other group members was low, there was a significant, negative relationship between group-mean intimate behaviors and the individual group member's intimate behaviors in a session.

Meanwhile, some researchers argued the importance of investigating the effect of uniformity in groups. For instance, González-Romá and Hernández (2014) used observers' ratings to explore the within-group distributions of climate ratings and categorize each group as: (a) completely uniform, (b) nonuniform with relatively low variability, or (c) nonuniform with relatively high variability. González-Romá and Hernández (2014) found that when groups were nonuniform with relatively low variability, compared to uniform climate, members reported more task conflicts and inhibited communication quality. The difference between the nonuniform group with relatively high variability and uniform group with no variability was not significant, and researchers argued that this finding was possibly due to the low number of groups with higher variability in their data set (González-Romá & Hernández, 2014). As a more feasible way of capturing group uniformity, Kivlighan (2016) used skew and kurtosis (DeRue et al., 2010) in examining the effects of convergence, strength, convergence, and uniformity (i.e., mean, *SD*, skew, and kurtosis [skew and kurtosis both operationalize uniformity] of group member working alliance ratings) on changes in students' grade point average in academic enhancement groups.

Although theory and prior findings indicated the importance of exploration of uniformity in groups, methodological issues have limited the utility of the within-group variability construct. One relevant issue is that the traditional method of using mean and *SD* to explore convergence (group *M*), dispersion (*SD*), and convergence by dispersion interaction ($M \times SD$) might not be able to capture each construct effectively. This is because when a group has either a very high or very low mean, dispersion has to be low and uniformity high (Kivlighan, 2021). For instance, in González-Romá and Hernández's (2014) research, mean group climate was highly correlated with within-group dispersion (*SD*; $r = .50$). That is, convergence (indicated by the average level across members measured by group *M*) and

uniformity (indicated by within-group variability measured by *SD* across group members) are constructs hard to separate.

Another issue is that there is a significant gap in literature examining precursors of group cohesion (i.e., group uniformity; Kivlighan, 2021; Lester et al., 2021). With mounting evidence that group uniformity leads to important group outcomes (Lester et al., 2021), there is a growing interest in predicting group uniformity. However, most studies focused on group uniformity as a predictor, and only a handful of studies explored these constructs as an outcome. This seems partial because prior methodological approaches to measure group uniformity did not consider the multilevel structure of group data such that sessions are nested within members, and members are then nested in groups. Indeed, Lester et al. (2021), in their review on organizational research, found that only 13 studies investigated group uniformity as an outcome, and in most of these studies, aggregate measures of dispersion (e.g., *SD* across group members) were used as a single-level outcome that represents group uniformity (Kivlighan, 2021). Additionally, although group researchers and therapists are often interested in groups' change over time, no study has explored whether the within-group dispersion (i.e., an index of the degree of within-group uniformity) changes longitudinally and what would predict the changes in within-group dispersion. Thus, a more sophisticated statistical model is needed to capture the complex nature of within-group variability in multilevel longitudinal data.

The location-scale model is a promising model to predict both group mean and within-group variability of a dependent variable (Hedeker et al., 2008; McNeish, 2021). The location, in the location-scale model, refers to the mean of the construct (e.g., group members' average engagement ratings), and the scale indicates the variability in these ratings (i.e., the heterogeneous residual variance; Lester et al., 2021). Standard multilevel models used in most group research (e.g., Ngai et al., 2015; Taube-Schiff et al., 2007; Uliaszek et al., 2016) are location-only models, which estimate fixed effects of predictors constant across groups. Any deviations from the fixed effects are called random effects, which are unexplained in the location-only models. (i.e., random effects are simply partitioned into different levels; McNeish, 2021). The random effects are assumed to be normally distributed

with a mean of zero and an estimated covariate matrix (McNeish, 2021). That is, location-only models do not allow the within-group residual variance (i.e., scale) to differ across groups, although it is likely that the within-group residual variance varies by group (McNeish, 2021). Thus, using the location-only models, we cannot examine the degree of within-group variability because it does not acknowledge that there can be differences in the variance of members' ratings. The location-scale model addresses this issue by modeling both the group mean and within-group variability so that researchers can investigate how the group mean and within-group variance differ across groups and how the mean and variances within groups change as a function of predictors (Bolger et al., 2003).

For example, Lang et al. (2018) explored the development of group consensus (decreased within-group variation) on job satisfaction in military companies. Lang et al. (2018) also investigated whether military units' initial collective efficacy levels predicted the changes in group consensus on job satisfaction longitudinally. Findings showed a main effect for time, indicating more group consensus on job satisfaction over time (i.e., less within-group variations over time). In addition, Lang et al. (2018) found an interactive effect between initial collective efficacy and change in job satisfaction consensus. When a group was high on initial collective efficacy, the variability in members' job satisfaction ratings within the group did not change over time. However, when a group had a low level of initial collective efficacy, the variability in members' job satisfaction ratings within the group significantly decreased over time.

Theory and prior research suggest that group cohesiveness is a critical aspect of group dynamics. In addition, changes in group aspects over the course of therapy and antecedent factors of the changes have been central interests of the group therapy process and outcome research (Kivlighan & Kivlighan, 2016). Therefore, we believe that researchers interested in how group members are cohesive in certain aspects and how the cohesiveness changes over time would benefit from a sophisticated methodological model that longitudinally predicts the homogeneity of a group (i.e., the heterogeneous residual variance) in addition to the group consensus (i.e., the average). In this article, we introduce a location-scale model example with three-level repeated data

(i.e., including groups, members, and sessions) and discuss how this model can address gaps currently existing in group research. Specifically, we illustrate the method for investigating how the mean and variability of group engagement would differ across time and whether group size would predict the changes in the mean and variability of group engagement in the framework of a Bayesian multilevel structural equation modeling.

Example

The data for this example came from a de-identified archival data set (Kivlighan et al., 1996; Kivlighan & Tarrant, 2001) that is classified as nonhuman subject data. The data set consists of ratings of 596 group members clustered within 66 groups. The groups in Kivlighan et al. (1996) were a combination of personal growth groups for counseling graduate students and interpersonal psychotherapy groups conducted at a midwestern university counseling center. The groups in Kivlighan and Tarrant (2001) were semistructured groups for youth in foster care. Session numbers for each group ranged from 2 to 10, with an average of 7.34 sessions. Groups ranged in size from 3 to 23 members, with an average of nine members per group. As Nich and Carroll (1997) suggested, we did not exclude cases just because they had different numbers of assessments and treated session numbers as a random variable. In each session, members rated their group's climate using the Engaged scale of the Group Climate Questionnaire (MacKenzie & Livesley, 1983). The first three columns of our data set (online Supplementary Material A) are: (a) "group," a unique number representing each of the 66 groups; (b) "member," a unique number representing each of 596 members in groups; (c) "group size," a number referencing to the number of members for each group. Next, the following 10 columns of the data set are each member's engagement ratings for 10 group sessions. As shown in the columns, engagement ratings for each session are placed in the wide format (one column per each session rating), not a long format (one row per each session rating) in our data set. This is because the location-scale modeling in *Mplus* is currently available only for two-level models (McNeish, 2021). According to Asparouhov and Muthén's (2016) suggestion, three-level hierarchies of our data were thus reduced by one by transposing the long format to the wide format. This transposition enables

researchers interested in using three-level repeated measures to exploit the advantages of the location-scale model in *Mplus* (McNeish, 2021).

The three-level location-scale model combines the basic location-scale model structure predicting group means and variance with a growth model that accounts for changes in group mean and variance over time. The statistical representation for our model is as follows:

$$\text{Engagement}_{tij} = \pi_{0ij} + \pi_{1ij}\text{SessionNumber}_t + e_{tij}, \quad (1)$$

$$\pi_{0ij} = \beta_{00j} + r_{0ij}, \quad (2)$$

$$\pi_{1ij} = \beta_{10j}, \quad (3)$$

$$\beta_{00j} = \gamma_{000} + \gamma_{010}\text{GroupSize}_j + u_{00j}, \quad (4)$$

$$\beta_{10j} = \gamma_{100} + \gamma_{110}\text{GroupSize}_j + u_{10j}, \quad (5)$$

$$e_{tij} = N(0, \sigma_j^2), \quad (6)$$

$$\sigma_j^2 = \exp[\omega_0 + \omega_1\text{SessionNumber}_t + \omega_2\text{GroupSize}_j + \omega_3(\text{SessionNumber}_t \times \text{GroupSize}_j)], \quad (7)$$

$$r_{0ij} \sim N(0, \tau_{000}), \quad (8)$$

$$\begin{bmatrix} u_{00j} \\ u_{10j} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} v_{00} & \\ v_{01} & v_{11} \end{bmatrix}\right), \quad (9)$$

where Engagement_{tij} is the engagement rating for t th session of i th member in the j th group and GroupSize is a Z-scored predictor of the size of each group as seen in Equation (1). π_{0ij} represents the engagement score of i th member in the j th group at Session 1, and π_{1ij} indicates an initial status of the group-member-specific mean change as a function of the session number. Also, e_{tij} refers to a session-specific deviation (i.e., random error) for i th member within j th group. At the within-group level, group-member-specific mean across sessions, π_{0ij} , is equal to j th group's mean across members across sessions (β_{00j}) plus the group-member-specific deviation (r_{0ij}). At the between-group level, the group-specific mean (β_{00j}) is equal to the average across all groups (γ_{000}) plus the effect of the size of j th group (γ_{010}) plus some group-specific deviation (u_{00j}). The group-member-specific mean change (π_{1ij}) is equal to the group-specific mean change (β_{10j})

because SessionNumber, is a group-specific variable (i.e., all group members within a group share the same session number). At the between-group level, β_{10j} is the group-specific mean change, which is equal to the average mean change across all groups (γ_{100}) plus the effect of the size of j th group (γ_{110}) on the mean change plus some group-specific deviation (u_{10j}). The average effects denoted by γ are referred to as location fixed effects constant across groups. As a unique aspect of the location-scale model, the within-group residual variance (σ_j^2) in the (e_{ij}) equation has a j subscript, indicating that the residual variance (random effect) is allowed to vary across groups (i.e., different from a traditional location model where σ^2 is assumed to be the same across groups [a normal distribution with a mean of zero]). The group-specific variance (σ_j^2) becomes the exponential of the linear model to predict how the within-group residual variance changes as a function of predictors. The reason the within-group residual variance is modeled from a log-linear model is because variances are not supposed to be negative. Specifically, log-linear models do not permit nonpositive values, which is important because variances must be nonnegative. It is also important to note that the coefficient interpretation is multiplicative in a log-linear model rather than additive as in a linear model. The group-specific variance (σ_j^2) is accounted for by the following scale fixed effects. ω_0 is the log of average within-group residual across all groups at the first session and the average group size. ω_1 is the effect of session number on the log of the residual variance, ω_2 is the effect of group size on the log of the residual variance, and ω_3 is the effect of session number by group size interaction on the log of the residual variance.

The *Mplus* code for the model is in [online Supplemental Material B](#). We can use TWOLEVEL RANDOM as the analysis type because we transform the data set into two levels. For the Bayesian estimation in *Mplus* (ESTIMATOR = BAYES), noninformative priors are chosen by default for computational convenience ([Asparouhov et al., 2018](#)). BITERATIONS is specified as 1,000, indicating that estimation used a minimum of 1,000 iterations with two chains. According to [Gelman and Rubin's \(1992\)](#) potential scale reduction method, iterations ended, and convergence was declared at a potential scale reduction factor of 1.10. In the %WITHIN% portion of the code, member-level component of a growth model is

created by regressing repeated measures constrained to 1 on π_{i0} . Also, to predict within-person residual variance for each measurement by bypassing the programmatic limitation of *Mplus*, we create phantom variables, logveg1–logveg10 ([Rindskopf, 1984](#)). The vertical pipe (|) is used to create variables having the same properties as the residual variance of eg1–eg10 and assign names of logveg1–logveg10. We use logveg labels to remind us that the within-group residual variance will vary across groups in a log-linear model. Now that we make the within-person residual variance into variables (not parameters), the residual variances can be outcomes in submodels. They are latent variables with a single indicator whose loading is constrained to 1 and whose residual variance is constrained to 0 ([McNeish, 2021](#)). In the %BETWEEN% portion of the code, the location fixed effects for the average mean and longitudinal mean change are specified with ($\beta_0 \beta_1$), which corresponds to γ_{000} and γ_{100} in the equation stated earlier. ω_0 and ω_1 are growth factors for phantom variables for scale fixed effects. (ω_0) is the log of the residual variance at Session Number = 0 (ω_0 in the above equation) and at average group size. (ω_1) is the effect of session number on the log of the residual variance (ω_{01} in the above equation). Location and scale predictors are included via regression by placing location and scale variables as the outcome variable: $\beta_0 \beta_1$ on group size and $\omega_0 \omega_1$ on group size. Other studies using the location-scale model present scale submodel predictor effects on the standard deviation scale (e.g., [Lang et al., 2018](#)). According to [McNeish's \(2021\)](#) method, we thus use the MODEL CONSTRAINT statement to make our results on the variance comparable to the studies (i.e., the scale fixed effects divided by 2). The MODEL CONSTRAINT statement also generates the exponential of the intercept of the log of the residual variance, $\exp(\omega_0)$.

The results of the three-level location-scale model are presented in [Table 1](#). In the “location fixed effects” ([Table 1](#)), location intercept (γ_{000} ; see the equation above) indicates the average estimated engagement score (across all sessions, members, and groups) at the grand mean of group size is 3.786 (95% credential interval [CI 3.565, 4.008]), which refers to a rating between moderately engaged and quite a bit engaged on the Group Climate Questionnaire's 7-point Likert scale. It should be noted that CI represents a credible

Table 1

Fixed and Random Effects From a Three-Level Location-Scale Model Predicting Mean and Variability of Group Engagement

Effect	Unstandardized estimate	Posterior <i>SD</i>	95% CI		Significance
			Lower 2.5%	Upper 2.5%	
Location fixed effects					
Intercept, γ_{000}	3.786	0.114	3.565	4.008	*
Session number, γ_{100}	0.057	0.015	0.028	0.086	*
Group size, γ_{010}	0.116	0.112	-0.106	0.333	
Session Number $\times$ Group Size, γ_{011}	0.014	0.016	-0.017	0.045	
Scale fixed effects					
Intercept, $\exp(\omega_0)$	0.915	0.044	0.827	1.018	*
Session number, ω_1	-0.033	0.005	-0.042	-0.023	*
Group size, ω_2	-0.033	0.022	-0.070	0.009	
Session Number $\times$ Group Size, ω_3	-0.003	0.008	-0.016	0.011	
Location random effect covariances					
Group intercept variance, v_{00}	0.687	0.155	0.455	1.058	*
Group session number variance, v_{11}	0.009	0.003	0.005	0.016	*
Group session number intercept covariance, v_{10}	-0.014	0.014	-0.046	0.010	
Member intercept variance, τ_{00}	0.444	0.039	0.375	0.526	*

Note. $n = 596$ members within 66 groups.

* $p < .05$.

interval rather than a confidence interval because we used the Bayesian framework. Credible intervals in Bayesian analysis can be used as an analogy for a confidence interval to approximate frequentist inferences (Lee, 1997), even though their philosophical bases differ (VanderPlas, 2014). 95% CI of the location intercept indicates that the subjective probability of the grand mean of group size lying between 3.565 and 4.008 is .95. Also, γ_{100} , corresponding to the initial status of mean engagement change across all members and groups, is .057 (95%CI [.028, .086]). Location group size effect is the effect for group size on group-specific mean ($\gamma_{010} = .116$, 95%CI [-.106, .333]), which was analogously nonsignificant. That is, groups' average engagement scores did not differ significantly depending on group size. The location session number by group size interaction effect was also nonsignificant ($\gamma_{011} = .014$, 95%CI [-.017, .045]), meaning the group size did not have a significant influence on the changes in group mean over time.

In the "scale fixed effects" (Table 1), scale intercept, $\exp(\omega_0)$, indicates that the expected residual variance (within-group variance) for a group after controlling for the effects of other predictors is .915 (95% CI [.827, 1.018]). This is the exponential of the ω_0 estimate from a log-linear

model. Also, the scale fixed effect of session number, ω_1 , is the expected influence of session number on the changes in within-group variance and was significant and negative ($\omega_1 = -.033$, 95% CI [-.042, -.023]). Therefore, within-group variance decreased (i.e., a group became more cohesive in terms of their engagement ratings) over time. However, group size did not significantly influence the average within-group variance or changes in within-group variance ($\omega_2 = -.033$, 95% CI [-.070, .009]; $\omega_3 = -.003$, 95% CI [-.016, .011]).

In "location random effect covariances" (Table 1), group intercept variance (v_{00}) indicates residual variance in group-specific engagement means that are not accounted for the group size. Because the group intercept variance was significant ($v_{00} = .687$, 95% CI [.455, 1.058]), researchers might be interested in exploring other potential factors impacting the group-mean differences. Group session number variance (v_{11}) corresponding to the residual variance in the group-mean change which is not explained by session number was also significant ($v_{11} = .009$, 95% CI [.005, .016]). Conditional on predictors in the model, members also appear to vary in their engagement intercept ($\tau_{00} = .444$, 95% CI [.375, .526]).

Discussion

Group therapy literature emphasizes the importance of group cohesiveness (e.g., Budman et al., 1993; Marziali et al., 1997). However, the actual development of group cohesiveness and antecedent factors to group cohesiveness have been unclear. This is partially because of existing methodological limitations in assessing group cohesiveness (group uniformity) within groups. Specifically, predicting changes in group cohesiveness and exploring a predictor of the changes in group cohesiveness were unavailable for multilevel data. Given that group therapy is a complex and evolving process unfolded in multilevels (Kivlighan & Kivlighan, 2016), sophisticated and proper research designs to explore different group compositions are called for.

In the present article, we introduced readers to McNeish's (2021) method of three-level location-scale model from a multilevel structural equation modeling framework (Li & Hedeker, 2012). We reviewed existing literature that attempted to investigate group dispersion and uniformity and suggested a better alternative by providing an example illustrating the application of the location-scale method to group therapy data. We hope that readers now have a better understanding of a newly introduced methodology for within-group dispersion and uniformity exploration. Through the application of the three-level location-scale model, we expect future group therapy research will be able to explore the development of group cohesiveness over time and its relationship with other variables.

As we illustrated in our example, modeling the location-scale model is one method for testing the influence of context on group cohesiveness in therapy. Findings from our example showed that the group became more cohesive in terms of their perceptions of the engaged group climate. This suggests that time is one context that we need to consider in understanding group cohesiveness in members' perceptions of group engagement. The group's lack of cohesiveness in the beginning stage of group therapy can be partially accounted for by the time effect. Clinically, it seems important for therapists to know that group cohesiveness as defined as group uniformity might not be a fixed trait-like quality of a group and be able to grow or change over time. Repeated assessments of group cohesiveness rather than one time would be ideal to better understand the group's cohesiveness. In addition, we can consider contextual

factors by examining the effects of other moderators in the location-scale model. Our example used the group size as a moderating factor to see whether group cohesiveness differs at the initial session or grows differently depending on the group size. We hope to find more information about which factors cause a group to be more homogeneous and cohesive, in addition to which factors make groups' means to change, so that we can understand mutative context bringing groups more cohesive in various aspects. For example, researchers might be interested in how group members' amounts of in-session intimate behaviors (Kivlighan et al., 2021) become similar over time as members learn from each other and whether characteristics of group leader (e.g., group therapist self-disclosure; Dies, 1973) could impact changes in group cohesiveness in intimate behaviors. The more knowledge about the predictive factors of group cohesiveness would allow therapists to effectively facilitate noncohesive groups to become more cohesive.

Although we focused on illustrating the application of the three-level location-scale model involving longitudinal exploration, it should be noted that the location-scale model is also available for two-level data. The location-scale model using two-level data (i.e., one-time measurement for members within groups) can tell us how much groups are different in terms of their cohesiveness (i.e., within-group variations) and why these groups' cohesiveness is different from each other (i.e., a predictor of the between-group difference in within-group variations). Researchers interested in one-time measurement in their group cohesiveness analysis can use this two-level approach described in McNeish (2021). In addition, the multilevel structural equation modeling approach used in the location-scale model can be extended to dyadic data using the vector autoregressive model (see McNeish & Hamaker, 2020).

As stated and demonstrated earlier, the primary strength of using the location-scale model is the ability to explore within-group variability for longitudinal multilevel data. Specifically, group therapy researchers can examine the changes in group-mean and within-group variability over time. Additionally, group researchers can examine the effects of various predictors on the changes in the mean and variability.

The high feasibility is another strength of this approach. We introduced McNeish's (2021)

approach to model the location-scale model using *Mplus*. Because of the ubiquity *Mplus* software, many group researchers are familiar with *Mplus*. We thus expect group researchers can learn and use the location-scale model using this software more comfortably and conveniently, compared to other less user-friendly options (e.g., SAS Proc Nlmixed in Hedeker et al., 2008) (McNeish, 2021). McNeish's (2021) method also allows us to circumvent practical and methodological issues (e.g., limitation of sample size) related to other software and computations used in other studies (e.g., Lang et al., 2018).

Because the location-scale model is a relatively new modeling method, there has not been enough literature to document issues with using this method. However, one possible limitation of using the location-scale model would relate to sample size. Because location-scale models can be fitted effectively with small within-group sample size, small sample sizes do not preclude location-scale modeling (Estabrook et al., 2012). However, admittedly, the small sample size can reduce the power of predictors (McNeish, 2021). Thus, three-level models with a too small number of repeated measures might "have difficulty with varying times of observations or a different number of repeated measures per person" (McNeish, 2021, p. 28). It is not clear how big the sample should be and how many measurements are required for using the location-scale model. At this point, the best advice is that probably the more measurements, the better. Another limitation is the significance of predictors in the location-scale model described in this article only indicates the significance of the contextual effects. Because the predictive effects are moderation effects, researchers using the location-scale model are not able to strongly argue the causal influences of predictors on the location (mean) and scale (variability).

We hope that this article can help readers to understand the basic concepts of location-scale modeling, so that they can benefit from the innovation related to examining scale effects. Researchers might be interested in exploring within-group variability as a mediator or predictor of other variables of interest. For example, it might be of interest whether the within-group variability in in-session intimate behaviors predicts group engagement climate, group alliance, or posttherapy outcome. Also, whether the effect of group therapist self-disclosure on group engagement

climate is mediated by the within-group variability in in-session intimate behaviors. Based on the understanding of the location-scale model introduced in this article, researchers can learn more about an example of a mediation location-scale model in Schultzberg and Muthén (2018, p. 427, Model 9). In addition, researchers might want to add random effects to the scale model to allow for unknown sources of between-group differences in their within-group variability. The example used in this study modeled within-group variability deterministically as a function of group size. By adding random effects to the scale model, researchers would be able to assess whether there are additional unmodeled sources. These random effects can also be correlated across the location and scale parts of the model (e.g., see Table 3 in McNeish, 2021). That is, researchers might want to take the correlation between group-mean and within-group variability (e.g., groups with higher means having less variance) into account.

It is also our hope that group researchers can have a better understanding and appreciation for the different group compositions, which have been ignored in research using traditional location-only models. Exploring the development of group cohesiveness and its relationship to other variables using appropriate methods will advance the group therapy theory and practice.

Annotated Bibliography

McNeish, D. (2020). Specifying location-scale models for heterogeneous variances as multilevel SEMs. *Organizational Research Methods*, 24(3), 630–653. <https://doi.org/10.1177/1094428120913083>

This article illustrates the general form of the location-scale model that can be envisioned as a multilevel structural equation modeling and discusses how the model may be able to extend the literature in organizational research.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, 64, 627–634. <https://doi.org/10.1111/j.1541-0420.2007.00924.x>

Hedeker et al. (2008) developed a two-level location scale model that allows for the examination of both the within-subjects and between-subjects variance in addition to means.

Li, X., & Hedeker, D. (2012). A three-level mixed-effects location scale model with an application to ecological momentary assessment data. *Statistics in Medicine*, 31(26), 3192–3210. <https://doi.org/10.1002/sim.5393>

A more flexible three-level mixed-effects location-scale model is introduced in this article.

Lang, J. W., Bliese, P. D., & de Voogt, A. (2018). Modeling consensus emergence in groups using longitudinal multilevel methods. *Personnel Psychology*, 71, 255–281. <https://doi.org/10.1111/peps.12260>

This article introduced a three-level model to explore consensus emergence in which group members' job satisfaction in military companies becomes uniform over time and the predictors of the rate of consensus emergence.

Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction*. <https://www.statmodel.com/download/IntroBayesVersion%203.pdf>

This article can provide background knowledge of how Bayesian analysis is carried out in *Mplus*. A series of examples, including a structural equation model, a two-part growth model, and a mixture model, are presented.

References

- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling*, 25(3), 359–388. <https://doi.org/10.1080/10705511.2017.1406803>
- Asparouhov, T., & Muthén, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 163–192). Information Age.
- 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: Theory, Research, and Practice*, 14(4), 368–383. <https://doi.org/10.1037/a0019170>
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54(1), 579–616. <https://doi.org/10.1146/annurev.psych.54.101601.145030>
- Bollen, K. A., & Hoyle, R. H. (1990). Perceived cohesion: A conceptual and empirical examination. *Social Forces*, 69(2), 479–504. <https://doi.org/10.2307/2579670>
- Budman, S. H., Soldz, S., Demby, A., Davis, M., & Merry, J. (1993). What is cohesiveness? An empirical investigation. *Small Group Research*, 24(2), 199–216. <https://doi.org/10.1177/1046496493242003>
- Budman, S. H., Soldz, S., Demby, A., Feldstein, M., Springer, T., & Davis, M. S. (1989). Cohesion, alliance and outcome in group psychotherapy. *Psychiatry*, 52(3), 339–350. <https://doi.org/10.1080/00332747.1989.11024456>
- Burlingame, G. M., McClendon, D. T., & Alonso, J. (2011). Cohesion in group therapy. *Psychotherapy*, 48(1), 34–42. <https://doi.org/10.1037/a0022063>
- Burlingame, G. M., McClendon, D. T., & Yang, C. (2018). Cohesion in group therapy: A meta-analysis. *Psychotherapy*, 55(4), 384–398. <https://doi.org/10.1037/pst0000173>
- DeRue, D. S., Hollenbeck, J., Ilgen, D., & Feltz, D. (2010). Efficacy dispersion in teams: Moving beyond agreement and aggregation. *Personnel Psychology*, 63(1), 1–40. <https://doi.org/10.1111/j.1744-6570.2009.01161.x>
- Deutsch, M. (1954). Field theory in social psychology. In G. Lindzey (Ed.), *Handbook of social psychology* (pp. 181–222). Addison-Wesley.
- Dies, R. R. (1973). Group therapist self-disclosure: An evaluation by clients. *Journal of Counseling Psychology*, 20(4), 344–348. <https://doi.org/10.1037/h0034808>
- Estabrook, R., Grimm, K. J., & Bowles, R. P. (2012). A Monte Carlo simulation study of the reliability of intraindividual variability. *Psychology and Aging*, 27(3), 560–576. <https://doi.org/10.1037/a0026669>
- Forsyth, D. R. (2021). Recent advances in the study of group cohesion. *Group Dynamics: Theory, Research, and Practice: Theory, Research, and Practice*, 25(3), 213–228. <https://doi.org/10.1037/gdn0000163>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- González-Romá, V., & Hernández, A. (2014). Climate uniformity: Its influence on team communication quality, task conflict, and team performance. *Journal of Applied Psychology*, 99(6), 1042–1058. <https://doi.org/10.1037/a0037868>
- Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, 64(2), 627–634. <https://doi.org/10.1111/j.1541-0420.2007.00924.x>
- Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, 83(1), 126–137. <https://doi.org/10.1037/0022-3514.83.1.126>
- Kivlighan, D. M., III. (2016). *Convergence, strength, and uniformity of group members' perceptions of the working alliance in academic enhancement*

- groups [Conference session]. American Psychological Association Annual Convention, Denver, CO.
- Kivlighan, D. M., Jr. (2021). From where is the group? To what is the group? Contributions of actor-partner interdependence modeling. *Group Dynamics: Theory, Research, and Practice*, 25(3), 229–237. <https://doi.org/10.1037/gdn0000164>
- Kivlighan, D. M., Jr., Gullo, S., Giordano, C., Di Blasi, M., Giannone, F., & Lo Coco, G. (2021). Group as a social microcosm: The reciprocal relationship between intersession intimate behaviors and in-session intimate behaviors. *Journal of Counseling Psychology*, 68(2), 208–218. <https://doi.org/10.1037/cou0000495>
- Kivlighan, D. M., III, & Kivlighan, D. M., Jr. (2016). Examining between-leader and within-leader processes in group therapy. *Group Dynamics: Theory, Research, and Practice*, 20(3), 144–164. <https://doi.org/10.1037/gdn0000050>
- Kivlighan, D. M., Jr., Multon, K. D., & Brossart, D. F. (1996). Helpful impacts in group counseling: Development of a multidimensional rating system. *Journal of Counseling Psychology*, 43(3), 347–355. <https://doi.org/10.1037/0022-0167.43.3.347>
- Kivlighan, D. M., III, Owen, J., & Antle, B. (2017). Members' attendance rates and outcomes of relationship education groups: A consensus-dispersion analysis. *Journal of Family Psychology*, 31(3), 358–366. <https://doi.org/10.1037/fam0000287>
- Kivlighan, D. M., Jr., & Tarrant, J. M. (2001). Does group climate mediate the group leadership-group member outcome relationship? A test of Yalom's hypotheses about leadership priorities. *Group Dynamics: Theory, Research, and Practice: Theory, Research, and Practice*, 5(3), 220–234. <https://doi.org/10.1037/1089-2699.5.3.220>
- Lang, J. W., Bliese, P. D., & de Voogt, A. (2018). Modeling consensus emergence in groups using longitudinal multilevel methods. *Personnel Psychology*, 71(2), 255–281. <https://doi.org/10.1111/peps.12260>
- Lee, P. M. (1997). *Bayesian statistics: An introduction*. Edward Arnold.
- Lese, K. P., & MacNair-Semands, R. R. (2000). The therapeutic factors inventory: development of a scale. *Group*, 24(4), 303–317. <https://doi.org/10.1023/A:1026616626780>
- Lester, H. F., Cullen-Lester, K. L., & Walters, R. W. (2021). From nuisance to novel research questions: Using multilevel models to predict heterogeneous variances. *Organizational Research Methods*, 24(2), 342–388. <https://doi.org/10.1177/1094428119887434>
- Li, X., & Hedeker, D. (2012). A three-level mixed-effects location scale model with an application to ecological momentary assessment data. *Statistics in Medicine*, 31(26), 3192–3210. <https://doi.org/10.1002/sim.5393>
- MacKenzie, K. R. (1983). The clinical application of a group climate measure. In R. R. Dies & K. R. MacKenzie (Eds.), *Advances in group psychotherapy: Integrating research and practice* (pp. 159–170). International Universities Press.
- MacKenzie, K. R., & Livesley, W. J. (1983). A developmental model for brief group therapy. In R. R. Dies & K. R. MacKenzie (Eds.), *Advances in group psychotherapy: Integrating theory and research* (pp. 101–135). International Universities Press.
- Marmarosh, C. L., & Van Horn, S. M. (2010). Cohesion in counseling and psychotherapy groups. In R. K. Conyne (Ed.), *The Oxford handbook of group counseling* (pp. 137–163). Oxford University Press.
- 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(4), 475–497. <https://doi.org/10.1080/00207284.1997.11490846>
- McNeish, D. (2021). Specifying location-scale models for heterogeneous variances as multilevel SEMs. *Organizational Research Methods*, 24(3), 630–653. <https://doi.org/10.1177/1094428120913083>
- McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*, 25(5), 610–635. <https://doi.org/10.1037/met0000250>
- Miles, J. R., Paquin, J. D., & Kivlighan, D. M., Jr. (2011). Amount and consistency, two components of group norms: An actor partner interdependence analysis of intimate behaviors in groups. *Group Dynamics: Theory, Research, and Practice: Theory, Research, and Practice*, 15(4), 326–342. <https://doi.org/10.1037/a0024676>
- Mischel, W. (1976). Towards a cognitive social model learning reconceptualization of personality. In N. S. Endler & D. Magnusson (Eds.), *Interactional psychology and personality* (pp. 166–207). Wiley.
- Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction*. <https://www.statmodel.com/download/IntroBayesVersion%203.pdf>
- Ngai, I., Tully, E. C., & Anderson, P. L. (2015). The course of the working alliance during virtual reality and exposure group therapy for social anxiety disorder. *Behavioural and Cognitive Psychotherapy*, 43(2), 167–181. <https://doi.org/10.1017/S135246581300088X>
- Nich, C., & Carroll, K. (1997). Now you see it, now you don't: A comparison of traditional versus random-effects regression models in the analysis of longitudinal follow-up data from a clinical trial. *Journal of Consulting and Clinical Psychology*, 65(2), 252–261. <https://doi.org/10.1037/0022-006X.65.2.252>

- Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, 49(1), 37–47. <https://doi.org/10.1007/BF02294204>
- Schneider, B., Salvaggio, A. N., & Subirats, M. (2002). Climate strength: A new direction for climate research. *Journal of Applied Psychology*, 87(2), 220–229. <https://doi.org/10.1037/0021-9010.87.2.220>
- Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling*, 25(4), 495–515. <https://doi.org/10.1080/10705511.2017.1392862>
- Taube-Schiff, M., Suvak, M. K., Antony, M. M., Bieling, P. J., & McCabe, R. E. (2007). Group cohesion in cognitive-behavioral group therapy for social phobia. *Behaviour Research and Therapy*, 45(4), 687–698. <https://doi.org/10.1016/j.brat.2006.06.004>
- Uliaszek, A. A., Rashid, T., Williams, G. E., & Gulamani, T. (2016). Group therapy for university students: A randomized control trial of dialectical behavior therapy and positive psychotherapy. *Behaviour Research and Therapy*, 77, 78–85. <https://doi.org/10.1016/j.brat.2015.12.003>
- Van Bergen, A., & Koekebakker, J. (1959). “Group cohesiveness” in laboratory experiments. *Acta Psychologica*, 16, 81–98. [https://doi.org/10.1016/0001-6918\(59\)90087-3](https://doi.org/10.1016/0001-6918(59)90087-3)
- VanderPlas, J. (2014). *Frequentism and Bayesianism: A python-driven primer*. arXiv preprint arXiv:1411.5018. <https://doi.org/10.48550/arXiv.1411.5018>
- Yalom, I. D. (1995). *The theory and practice of group psychotherapy* (4th ed.). Basic Books.

Received January 31, 2022

Revision received December 17, 2022

Accepted December 19, 2022 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <https://my.apa.org/portal/alerts/> and you will be notified by e-mail when issues of interest to you become available!