



Published in final edited form as:

*Psychother Res.* 2008 July ; 18(4): 365–376. doi:10.1080/10503300701796992.

## Statistical Analysis of Group-Administered Intervention Data: Re-Analysis of Two Randomized Trials

Paul D. Rohde

### Abstract

Group-administered interventions often create statistical dependencies, which if ignored increase the rate of Type I errors. We analyzed data from two randomized trials involving group interventions to document the impact of statistical dependency on tests of intervention effects and to provide estimates of statistical dependency. Intraclass correlations ranged from .02 to .12. Adjusting for dependencies increased  $p$ -values for the tests of intervention effects. The increase in the  $p$ -values depended upon the magnitude of the statistical dependence and available degrees of freedom. Results suggest that the literature may overstate the efficacy of group interventions and imply that it will be important to study why groups create dependencies. We discuss how dependencies impact statistical power and how researchers can address this concern.

### Keywords

group-administered interventions; within-group dependence; randomized trials; intraclass correlations

---

Group-administered psychotherapeutic and preventive interventions are common (Burlingame, MacKenzie, & Strauss, 2004). Advantages of group-administered interventions include the therapeutic effects of the group (Burlingame, Fuhrman, & Johnson, 2002) and the ability to efficiently treat large numbers of people (e.g., Antonuccio, Thomas, & Danton, 1997). However, trials evaluating group-administered interventions face an important statistical issue. Namely, data from group-administered interventions often violate a key assumption of most statistical analyses used in intervention research—the assumption of independence of observations (Baldwin, Murray, & Shadish, 2005; Burlingame, Kircher, & Honts, 1994; Kenny & Judd, 1986; Kenny, Kashy, & Bolger, 1998; Kenny, Mannetti, Peirro, Livi, & Kashy, 2002; Morgan-Lopez & Fals-Stewart, 2006; Roberts & Roberts, 2005). Analyses of intervention effects that do not account for this statistical dependence increase the risk of Type I errors (Baldwin et al., 2005; Burlingame et al., 1994; Roberts & Roberts, 2005). Problems associated with the analysis of group data, particularly regarding tests of intervention effects, have been documented, although previous research has typically relied on simulated data (cf. Baldwin et al., 2005; Burlingame et al., 1994; Varnell, Murray, Hannan, & Baker, 2001). The purpose of this report is to use actual data from two clinical trials to illustrate the impact of statistical dependencies stemming from group-administered interventions on tests of intervention effects.

Before we summarize the issues involved in the analysis of group-administered intervention data, we define a few terms we use throughout the report. We use the term *group* to refer to the small groups of participants to whom the intervention was delivered and denote group with a lowercase  $g$ . We use *member* to refer to the group members (denoted with a lowercase  $m$ ).

---

Finally, we use the term *condition* to refer to intervention or control conditions (denoted with a lowercase *c*).

A key statistical concept in the analysis of group data is *within-group dependence*, which refers to the fact that during the intervention observations within a group can become correlated, either positively or negatively (Baldwin et al., 2005; Kenny et al., 2002). Positive correlations are thought to develop because group members share a common environment that can homogenize response to the intervention (cf. Kenny et al., 2002). For example, rates of cohesion, attrition, and attendance can impact outcomes in group treatment (Burlingame et al., 2002). Individuals attending groups that are cohesive, retain members, and are consistently attended are more likely to have good outcomes, whereas individuals attending groups that are unfriendly, lose members, and are poorly attended are more likely to have poor outcomes. Negative correlations occur when the group environment differentiates response to the intervention. For example, group members share fixed resources, such as the time and attention of the group leader (Kenny et al., 2002). The differential access to this resource may lead some group members to improve and others to make no change or even deteriorate, which would lead to a negative correlation within groups. Other potentially important variables that may create within-group dependence might be the average motivation level of group members, presence of natural leaders or role models in the group, frequency of scapegoating, presence of negative cliques, presence of a dominating or difficult group member, and the skill of the group leader. Any factor that can vary between groups could be a source of within-group dependence.

Within-group dependence is typically indexed by an intraclass correlation (ICC). An ICC can be either positive or negative—a positive ICC indicates that participants within a group are more similar to one another than to other participants; a negative ICC indicates that participants within a group are more dissimilar from one another than from other participants (Kenny et al. 2002)<sup>1</sup>. When ICCs are positive, which is likely the case in group-administered interventions (e.g., Herzog et al., 2002), the standard error of the intervention effect will be larger than it would have been if observations were independent. Statistical tests that ignore this increased variation underestimate the standard error of the intervention effect, which increases the likelihood that the analysis will incorrectly suggest a significant intervention effect (i.e., Type I error; Baldwin et al., 2005; Burlingame et al., 1994; Murray, 1998; Varnell et al., 2001).

To prevent an increased likelihood of a Type I error, it is important for researchers to use an analytic approach that accounts for the within-group dependencies in the calculation of both the intervention effect and the degrees of freedom for the intervention effect. Specifically, analyses should account for statistical dependencies by (a) including group in the analysis as a nested random effect (Kenny & Judd, 1986; Murray, 1998; Zucker, 1990), which can be accomplished via multi-level or mixed model analyses, and (b) basing the degrees of freedom on the number of groups per condition (Kenny et al., 1998; Murray, Hannan, & Baker, 1996;

---

<sup>1</sup>Some methodologists recommend that negative ICCs be fixed to zero (e.g., Maxwell & Delaney, 2004). However, fixing negative estimates to zero makes the analysis of treatment effects overly conservative (i.e., Type I errors will be below the nominal level, Kenny et al., 1998; Murray et al., 1996). Statistical nonindependence refers to the fact that observations are correlated. That is, given score *X* we know something about score *Y*. Nothing about nonindependence requires that the correlations be positive. Of course, it is possible that as *X* increases, *Y* increases (positive correlation), but it is also possible that as *X* increases, *Y* decreases (negative correlation). Both scenarios reflect nonindependence in the data. One way to model nonindependence is to model the between group variability. If observations are positively correlated then there will be nonzero between group variability. However, a limitation of this approach is that it requires the ICC to be zero or greater than zero. Thus, if the observations are negatively correlated and we follow the practice of setting the ICC to zero, then the model assumes the data are independent when they are not. The methods we describe in the Appendix and those described by Kenny et al. (2002) allow for negative nonindependence in many situations. If the nonindependence is positive, these methods will provide identical model fit to more traditional methods. Readers interested in negative ICCs should consult Kenny et al. (2002) for a readable introduction.

Varnell et al., 2001). Baldwin et al. (2005) discuss potential objections and rebuttals to these two recommendations.

As noted previously, others have addressed the issue of within-group dependence. Nevertheless, our report adds to the methodological work in this area in three ways. First, we document the impact of within-group dependence using real-world data from two group-administered intervention trials. Most research in this area has relied on simulated estimates of the ICC (cf. Baldwin et al., 2005; Burlingame et al., 1994; Varnell et al., 2001). Roberts and Roberts (2005), a notable exception in this regard, documented the impact of within-group dependence on the intervention effect for a single outcome using a mixed-model analysis of covariance (ANCOVA), a common analysis approach for estimating intervention effects. We extend their work by discussing the effects of within-group dependence on tests of the intervention effects in another common analytic approach: repeated measures analysis of variance (ANOVA). We also present a mixed-model formulation for these models, including the interpretation of the parameters and the specification of the ICC affecting the intervention effect. Second, we document ICCs for multiple outcomes in two substantive areas not addressed previously—the treatment of adolescent depression and conduct problems and the prevention of eating disorders. Researchers in these areas can use these estimates when planning future trials. Third, we provide syntax for specifying these models in SAS PROC MIXED and discuss the estimation of mixed-model repeated-measures ANOVA models when there is negative within-group dependence (see the Appendix).

## Methods

### Datasets

**Making a Plan for Success (MAPS)**—The primary focus of the MAPS study (Rohde, Clarke, Mace, Jorgensen, & Seeley, 2004) was to evaluate the efficacy of a group-based CBT depression intervention, the Adolescent Coping With Depression (CWD-A) course (Clarke, Lewinsohn, & Hops, 1990), for depressed adolescents with major depression disorder (MDD) and comorbid conduct disorder (CD). The CWD-A course was compared to a group-administered life skills/tutoring condition. Between 1998 and 2001, 281 adolescents (ages 12–17) were referred to the study by staff of the Department of Youth Services of Lane County, Oregon. A total of 182 adolescents completed intake assessments and were randomized to treatment conditions. Of the 182 assessed adolescents, 93 had current comorbid MDD and CD, 51 had CD only, 21 had MDD only, and 17 had neither diagnosis, although they could have had diagnoses other than MDD or CD or past MDD/CD and all had elevated levels of depressive symptoms that warranted treatment. Previous analyses of the MAPS data have focused on just those adolescents with comorbid MDD and CD (Rohde et al., 2004) or with at least MDD (Rohde, Seeley, Kaufman, Clarke, & Stice, 2006). Given that we are interested in understanding the effects of group-administered interventions on statistical tests of intervention effects, we used the entire sample because the treatment groups consisted of adolescents from all four diagnostic categories. The CWD-A condition included 94 participants treated in nine groups. The Life Skills condition included 88 participants treated in nine groups. Group sizes ranged from 6 – 14 ( $Mdn = 10$ ,  $M = 10.11$ ,  $SD = 1.97$ ). Both interventions consisted of 16 2-hour sessions conducted over an 8-week period. See Rohde et al. (2004) for a complete description of the interventions, participants, and procedures.

We limited the analyses to three primary outcome measures: (a) the Beck Depression Inventory-II (Beck, Steer, & Brown, 1996), a self-report measure of depressive symptoms, (b) the Hamilton Depression Rating Scale (Hamilton, 1960), an interviewer rated measure of depressive symptoms, and (c) the externalizing subscale of the Child Behavior Checklist (Achenbach, 1991), a parent/adult informant rated measure of disruptive behavior. Because

we sought to model intervention effects when there are two time points, we limited the data to baseline and post-treatment data (i.e., data from the follow-up period were not examined).

**Body Project**—The primary focus of the Body Project (Stice, Shaw, Burton, & Wade, 2006) was to evaluate the efficacy of a group-administered eating disorder prevention program involving dissonance-inducing activities that reduce thin-ideal internalization. The dissonance intervention (DI) was compared to a group-administered healthy weight management program, an expressive writing control condition, and an assessment-only control condition. Because the latter two conditions were not conducted in groups, we limited the analyses to the DI condition and healthy weight condition. The DI and healthy weight management program each consisted of 3 weekly 1-hour group sessions.

From 2001 to 2003, 481 female adolescents (ages 14–19) who expressed body-image concerns were recruited from high schools and a university to participate in a study evaluating interventions designed to help young women better accept their bodies. The DI condition included 115 participants treated in 18 groups. The Weight Management condition included 117 participants treated in 18 groups. Group sizes ranged from 1–12 ( $Mdn = 6.5$ ,  $M = 6.44$ ,  $SD = 2.30$ )<sup>2</sup>. See Stice et al. (2006) for a complete description of the interventions, participants, and procedures.

We limited our analyses of the Body Project data to three primary outcome measures: (a) the Ideal-Body Stereotype Scale-Revised (Stice, Fisher, & Martinez, 2004), a self-report measure of internalization of the thin beauty ideal espoused for females in Western cultures, (b) the Satisfaction and Dissatisfaction with Body Parts Scale (Berscheid, Walster, & Bohrnstedt, 1973), specifically items that assessed body parts that often concern females (e.g., stomach, thighs, and hips), a self-report measure of body dissatisfaction, and (c) the Eating Disorder Examination (EDE; Fairburn & Cooper, 1993), a semi-structured interview that measures DSM-IV bulimia nervosa symptoms. We focused on baseline to post-treatment effects.

## Data Analysis

In both datasets we estimated intervention effects using a mixed-model repeated measures ANOVA estimated in SAS (v. 9.1) PROC MIXED, although the problems we discuss in this report affect most analyses used in the intervention trials (e.g., random coefficients models). The Appendix includes syntax for estimating the models in SAS. Although previous reports of the Body Project data used one-tailed tests (Stice et al., 2006), all significance tests in this report were two-tailed.

In repeated measures models, the statistical test of the intervention effect is the time x condition interaction, which tests whether baseline to post-treatment change in the dependent variable is greater for the intervention condition than for the comparison condition. For the dependent variables in each dataset, we estimated two sets of models. First, we estimated intervention effects ignoring group (Model 1a = MAPS; Model 2a = Body Project). The repeated measures model ignoring group is (cf. Littell, Milliken, Stroup, & Wolfinger, 1996):

$$Y_{i;k} = \mu + C_k + T_t + TC_{tk} + [M_{i;k} + e_{i;k}]$$

<sup>2</sup>Because of recruitment problems (e.g., no-shows) during the early stages of the Body Project, one person in the Body Project was seen individually. We conducted analyses that both included and dropped the participant who was seen individually. The analyses differed only slightly and produced similar substantive conclusions. Consequently, we present the results that include her to make our results consistent with other Body Project publications.

Here  $Y_{ti:k}$  is the value of the dependent variable at timepoint  $t$  for individual  $i$  nested within condition  $k$ . Time is modeled as a categorical variable with two levels (cf. Murray, 1998).  $Y_{ti:k}$  is modeled as a function of four fixed effects (outside the brackets) and two random effects (inside the brackets). The four fixed effects are: the grand mean ( $\mu$ ), the main effect for condition ( $C_k$ ), the main effect for time ( $T_t$ ), and the time x condition interaction ( $TC_{tk}$ ). The time x condition interaction is the estimate of the intervention effect, with degrees of freedom equal to  $c(t-1)(n-1)$ , where  $c$  is the number of conditions,  $t$  is the number of time points, and  $n$  is the number of participants per condition. The first random effect,  $M_{i:k}$ , allows for correlation among the repeated observations on the same individual and produces the variance component  $\sigma_{m:c}^2$ , where m:c is participants nested within conditions. The second random effect,  $e_{ti:k}$ , models random variation among individuals and produces the residual variance  $\sigma_e^2$ .

Because Models 1a and 2a ignore groups, they assume that observations taken from individuals within a given group are independent. As we noted above, this assumption is often violated. Consequently, we re-estimated the models, adding two random effects that account for the within-group dependence (Model 1b = MAPS; Model 2b = Body Project). Using the notation of Murray (1998) the repeated measures model accounting for group is:

$$Y_{ti;j:k} = \mu + C_k + T_t + TC_{tk} + [G_{j:k} + TG_{tj;k} + M_{i:k} + e_{ti:k}]$$

Here  $Y_{ti;j:k}$  is the value of the dependent variable at time  $t$  for individual  $i$  nested in group  $j$  nested in condition  $k$ . As before, time is modeled as a categorical variable with two levels (Murray, 1998).  $Y_{ti;j:k}$  is modeled as a function of four fixed effects (outside the brackets), which are the same as Models 1a and 2a. The time x condition interaction remains the test of the intervention effect and has degrees of freedom equal to  $c(g-1)(t-1)$ , where  $c$  is the number of conditions,  $g$  is the number of groups, and  $t$  is the number of time points. Murray et al.'s (1996) simulations showed that these degrees of freedom are necessary to maintain the nominal Type I error rate.  $Y_{ti;j:k}$  is also modeled as function of four random effects (inside the brackets), two of which were included previously. The two additional random effects allow for correlation among group members ( $G_{j:k}$ ) and for correlation among group members within a time x group survey ( $TG_{tj;k}$ ; Murray, 1998). That is,  $TG_{tj;k}$  allows for correlation among members due to being in the same group at the same time. The variance components for the additional random effects are, respectively,  $\sigma_{g:c}^2$ , where g:c is groups nested within conditions, and  $\sigma_{tg:c}^2$ , where tg:c is the time x group interaction nested within condition.

**Intraclass correlation**—The ICC affecting the intervention effect for Models 1b and 2b is estimated as:

$$ICC_{mt:g:c} = \frac{\sigma_{tg:c}^2}{\sigma_e^2 + \sigma_{tg:c}^2}$$

Note that the ICC affecting the intervention effect in a mixed-model repeated measures ANOVA involves the variance component for the time x group interaction ( $\sigma_{tg:c}^2$ ) rather than the variance component for group ( $\sigma_{g:c}^2$ ). Hence,  $ICC_{mt:g:c}$  indexes the extent to which group members are homogenous with respect to change over time. Confidence intervals for  $ICC_{mt:g:c}$  were calculated using formulas from Snedecor and Cochran (1980).

**Intervention effect size**—Wampold and Serlin (2000) recommend using  $\omega^2$  to estimate the proportion of total variability attributable to treatment. Furthermore, Wampold and Serlin

demonstrated that ignoring statistical dependence can inflate the size of  $\omega^2$ . Consequently, we estimated  $\omega^2$  for the models that included groups and those that did not to estimate the degree of inflation.

**Variance inflation**—We also calculated the Variance Inflation Factor (VIF; Donner, Birkett, & Buck, 1981), which represents the amount the variance of the intervention effect (and thus standard error) increased due to the statistical dependencies. The VIF for Models 1b and 2b was calculated as follows:

$$VIF_{mt:g:c} = (1 + (m - 1) ICC_{mt:g:c})$$

where  $m$  is the average number of members per group. Thus, the VIF is a function of the average group size and the ICC; as either value increases, so will the VIF.

**Heterogeneous intraclass correlations**—Within a study it is possible for the random effects and thus the ICC to differ across conditions (Roberts & Roberts, 2005). For example, consider a study that compares a group intervention that fosters intense interactions among group members to a group intervention that is didactic with little interaction among group members. It is conceivable that the ICC would be larger in the former intervention than the latter. To account for this possibility, we re-estimated Models 1b and 2b allowing each of the random effects to vary across condition (see Appendix).

## Results

### Magnitude of the Within-Group Dependence

Table 1 presents the  $ICC_{mt:g:c}$  estimates calculated from the mixed model repeated-measures analyses for each outcome in the MAPS and Body Project studies. For the MAPS data the  $ICC_{mt:g:c}$  for the BDI, HAM-D, and CBCL-E was 0.03, 0.07, and 0.12, respectively. Thus, within-group dependence was moderate for depressive symptoms and relatively high for externalizing behavior. For the Body Project data the  $ICC_{mt:g:c}$  for TII, BD, and ED symptoms was 0.11, 0.02, and 0.02, respectively. Thus, within-group dependence was moderate-high for thin-ideal internalization and low for body dissatisfaction and eating disorder symptoms.

### Magnitude of Variance Inflation

The VIFs in Table 1 indicate how much the variance of the intervention effect (and thus standard error of the intervention effect) was inflated by the within-group dependence. For the MAPS data the  $VIF_{mt:g:c}$  was 1.28, 1.59, and 2.09 for the BDI, HAM-D, and CBCL-E, respectively. For the Body Project data the  $VIF_{mt:g:c}$  was 1.57, 1.12, and 1.13 for the TII, BD, and ED symptoms, respectively. For the CBCL-E, the variance of the intervention effect was 2.09 times larger than it would have been if there was no within-group dependence. Other VIFs can be interpreted similarly. The larger the VIF, the larger the variance of the intervention effect and thus the greater the reduction in the test of the intervention effect (i.e., time x condition  $F$ ) as compared to an analysis that ignores the effects of group.

### Denominator Degrees of Freedom

As noted above, in order to maintain the nominal Type I error rate the analysis of group intervention data includes not only modeling within-group dependence but also basing the degrees of freedom for the test of the intervention effect on the number of groups per condition rather than the number of participants per condition. The denominator degrees of freedom for the time x condition  $F$  is a function of the number of conditions, groups per condition, and time



points. In the MAPS study there were two conditions, nine groups per condition, and two time points. Thus, the test of the intervention effect had 16 denominator degrees of freedom  $(2(9-1)(2-1) = 16)$ . In the Body Project study there were two conditions, 18 groups per condition, and two time points. Thus, the test of the intervention effect had 34 denominator degrees of freedom  $(2(18-1)(2-1) = 34)$ . As can be seen in Table 1, the change in degrees of freedom resulted in a substantial decrease in the denominator degrees of freedom for both studies over the models that ignore the nested nature of the data.

### Consequences for the Statistical Significance of Intervention Effects

Adjusting for within-group dependence affects  $p$ -values associated with tests of the intervention effect. As we have seen, the recommended analysis reduces both  $F$ -values and has fewer degrees of freedom, which together increase the  $p$ -value for the test statistic. The size of the increase varies as a function of the magnitude of the ICC and the number of groups per condition. In the MAPS data all  $p$ -values increased when group was included in the analysis (see Table 1). In the analysis that ignored group (Model 1a) the intervention effects were nonsignificant for all outcome measures. Because all ICCs from Model 1b were positive, all effects remained nonsignificant after we accounted for group. In the Body Project data all  $p$ -values increased when group was included in the analysis (see Table 1). In Model 2a, the intervention effects for TII and BD were statistically significant (both  $p = .01$ ) whereas the intervention effect for ED symptoms was marginally significant ( $p = .09$ ). Because  $ICC_{mt:g:c}$  for TII was substantially larger than  $ICC_{mt:g:c}$  for BD (0.11 vs. 0.02), in Model 2b the intervention effect for TII became marginally significant ( $p = .06$ ) whereas the intervention effect of BD remained significant ( $p = .02$ ). The intervention effect for ED symptoms remained nonsignificant.

### Consequence for the Intervention Effect Size

As can be seen in Table 1, adjusting for within-group dependence did not affect the estimates of  $\omega^2$  much. In the MAPS data, the BDI and CBCL-E had  $\omega^2$  values of zero, making it impossible for  $\omega^2$  to become any smaller. For the HAM-D  $\omega^2$  was 0.02 in Model 1a and 0.01 in Model 1b. In the Body Project data,  $\omega^2$  for the TII and BD did not change from Model 2a to Model 2b ( $\omega^2 = 0.04$  for TII and  $\omega^2 = 0.05$  for BD, respectively). However, for ED symptoms,  $\omega^2$  went from 0.02 in Model 2a to 0.01 in Model 2b. When  $\omega^2$  changed, the magnitude of the reductions in  $\omega^2$  were similar to the magnitude of reductions observed by Wampold and Serlin (2000) for positive ICCs less than 0.10 (the size of the majority of ICCs in this study).

One way to interpret an ICC is the proportion of variance accounted for by groups. Consequently, Kim, Wampold, and Bolt (2006) have suggested that contrasting the magnitude of the ICC and  $\omega^2$  from the models that include group provides a heuristic comparison of the relative importance of interventions and groups. For the MAPS data, the magnitude of the ICC exceeded  $\omega^2$  for all outcomes. For the Body Project data, the ICC exceeded  $\omega^2$  for TII and ED symptoms but  $\omega^2$  exceeded the ICC for BD. Together these results suggest that between group differences were likely at least as important if not more important than between intervention differences.

### Heterogeneous Intraclass Correlations

Because it is possible for ICCs to differ across intervention conditions, we re-estimated Models 1b and 2b allowing  $ICC_{mt:g:c}$  to differ. We refer to the original models as homoscedastic because they assume that within-group dependence is constant across conditions. We refer to the models that allow  $ICC_{mt:g:c}$  to differ across intervention conditions as heteroscedastic because they allow within-group dependence to vary across conditions. Because the homoscedastic models estimate fewer parameters, they are usually preferred over the heteroscedastic models, unless the heteroscedastic model significantly improves model fit. We

compared the fit of the heteroscedastic and homoscedastic models using a  $-2$  Log Likelihood test, which has a  $\chi^2$  distribution and degrees of freedom equal to the difference in random effects between the two models.

Table 2 includes  $ICC_{mr:g:c}$  for each outcome variable stratified by condition. As can be seen, within-group dependence differed across condition for most outcomes. Nevertheless, the heteroscedastic models did not improve model fit for any of the outcome variables in the MAPS study nor for TII and BD in the Body Project. In contrast, the heteroscedastic models did improve model fit for ED symptoms,  $\chi^2(4) = 24.9, p < .001$ . The  $ICC_{mr:g:c}$  for ED symptoms from the homoscedastic model was 0.02 (95% CI =  $-0.05 - 0.14$ ). The heteroscedastic model indicated that for the ED symptoms  $ICC_{mr:g:c}$  was larger in the Healthy Weight condition than the Dissonance condition (Healthy Weight  $ICC_{mr:g:c} = 0.18$  [95% CI =  $0.04 - 0.42$ ], Dissonance  $ICC_{mr:g:c} = 0$  [95% CI =  $-0.06 - 0.11$ ]<sup>3</sup>).

The homoscedastic and heteroscedastic models produced similar tests of the intervention effect for ED symptoms (Homoscedastic,  $F(1, 34) = 2.47, p = 0.13$ ; Heteroscedastic,  $F(1, 34) = 1.84, p = 0.18$ ). The heteroscedastic models reduced the  $F$  values and increased the  $p$ -values. However, the heteroscedastic model did not alter the conclusions we drew about the efficacy of these interventions.

## Discussion

Our primary purpose in re-analyzing data from the MAPS and Body Project studies was to discuss how to account for group in a repeated measures analysis and illustrate the effects of within-group dependence on statistical tests of intervention effects. Accounting for within-group dependence requires researchers to change their analysis of intervention effects in two ways: (a) account for the ICC in the statistical test of the intervention effect and (b) base the degrees of freedom on the number of groups rather than number of individuals. As our results demonstrated, in repeated measures analyses these changes will, respectively, reduce the test statistic for the intervention effect and the available degrees of freedom. These changes had minimal effects on the intervention effect size, although that may be due to the fact that intervention effect sizes were small even in the analyses that ignored groups. If not planned for, these reductions can leave investigators with little power to detect intervention effects.

A key issue in dealing with within-group dependence is identifying its source. Kenny et al. (2002) described three sources of within-group dependence. The first source is group composition. Observations within a group may be correlated if group members are homogenous with respect to sex or psychiatric diagnosis, or if groups are primarily composed of members comfortable with disclosure and other group dynamics (Burlingame et al., 2002). The second source is a common fate. Group members become more alike because they share a common experience, such as working with the same group leader. The third source is mutual influence, which means that within-group dependence develops because group members interact and influence one another throughout the course of the intervention. Mutual influence is likely to be the most important source of within-group dependence (cf. Burlingame et al., 2002).

<sup>3</sup>PROC MIXED fixed the random effect for the time x group interaction to zero, suggesting the possibility of negative within-group dependence in the dissonance condition. Consequently, we re-estimated the intervention effects by analyzing posttest data adjusted for baseline and modeled the within-group dependence as a covariance rather than a variance. We estimated both a homoscedastic and heteroscedastic model. As before the heteroscedastic model improved model fit ( $\chi^2(2) = 33.1, p < .001$ ) but did not significantly alter the conclusion about the intervention effect. The ICC for ED symptoms from the homoscedastic model was 0.01. The heteroscedastic models indicated that for the ED symptoms the ICC was positive in the Healthy Weight condition (0.09) and negative in the Dissonance condition ( $-0.02$ ).



In fact, we can observe the predicted outcomes of mutual influence by examining the largest ICCs in the MAPS and Body Project studies. In MAPS, the largest ICC was for CBCL-E ( $ICC_{mr:g:c} = 0.12$ ), a measure of externalizing or disruptive behavior. This is consistent with literature on the effects of peer-to-peer reinforcement of externalizing behavior in group-based interventions. However, whereas some researchers have argued that this reinforcement process is iatrogenic (Dishion, McCord, & Poulin, 1999), the majority of groups in the present study improved, suggesting the reinforcement process was helpful. In the Body Project, the largest ICC was for thin-ideal internalization ( $ICC_{mr:g:c} = 0.11$ ). Mutual influence is also a compelling explanation for this result. Much of the intervention time, especially in the dissonance condition, was spent discussing drawbacks associated with the thin-ideal. Given this focus, it is not surprising that groups became similar on this attitudinal dimension.

Researchers should consider design features that will affect mutual influence when designing their studies, such as the duration and intensity of interaction among group members. If the group intervention is highly interactive, with numerous opportunities for group members to interact and influence each other, we would expect within-group dependence to be high. Likewise, lengthy groups that meet over the course of many months will provide group members many more opportunities to interact with one another than groups that meet a few times over the course of a month. Group size might also impact mutual influence. Interaction is more frequent in small groups, increasing the likelihood of mutual influence and thus within-group dependence. We are not arguing that researchers simply aim to eliminate or minimize mutual influence. Indeed, mutual influence is often at the heart of group interventions. Rather we are arguing that researchers attend to processes that lead to between group differences in mutual influence and attempt to maximize processes that produce good outcomes.

The preceding point illustrates that although within-group dependence complicates research design and statistical analyses, it is not simply a methodological nuisance (Kenny et al., 2002; Roberts & Roberts, 2005). Rather within-group dependence likely reflects psychological and social processes central to group-based interventions (Yalom & Leszcz, 2005). Future research needs to address important substantive issues related to within-group dependence, such as optimal group composition, characteristics of successful group leaders, and the processes by which mutual influence occurs. Such research would benefit clinicians as well as researchers. Kenny et al. (2002) reviewed methods for studying group influence that could be adapted for group-administered intervention research.

As noted throughout, power is often low in group research. When comparing group-administered interventions, statistical power is a function of four variables: (a) the number of groups per condition, (b) the average group size, (c) the magnitude of the within-group dependence, and (d) the effect size. To illustrate this we computed the sample sizes needed to achieve 80% power to detect a medium effect size in a study with two conditions. Table 3 provides the results for studies using individual interventions and studies using group-administered interventions. For the group interventions we varied the group size ( $m = 5, 10,$  and  $15$ ) and the ICC ( $-.05, .00, .05, .15,$  and  $.30$ ). As can be seen in Table 3, when the ICC is positive, as is typical in group-administered intervention research, delivering interventions in a group format requires larger samples than delivering interventions individually. However, when the ICC is negative, delivering interventions in a group format decreases the overall sample size requirements. Furthermore, when the ICC is positive, increasing the number of groups will have a greater impact on power than increasing group size (Baldwin et al., 2005; Kenny et al., 1998). Just the opposite is true when the ICC is negative.

The magnitude of the within-group dependence (i.e., the ICC) impacts power the most (see Table 3). Consequently, statistical techniques that reduce the size of the ICC also increase power (Murray & Blinstead, 2003). For example, including covariates that may account for the

within-group dependence, such as group cohesion, attendance, or group size, can increase power provided the covariates are independent of the intervention (i.e., not correlated with intervention condition). Covariates measured at the individual-level rather than the group-level are preferable because group-level covariates will reduce the group-level degrees of freedom, which are often low (Baldwin et al., 2005).

One might argue that to conserve power we should ignore within-group dependence for dependent variables with small and non-statistically significant ICCs. This line of reasoning is problematic for three reasons. First, small amounts of within-group dependence increase Type I error rates. For example, Kenny et al. (2002) noted that when the average group size is 12, an ICC of 0.04 will raise the Type I error rate to 10%. Second, the statistical problems created by within-group dependence do not depend upon the statistical significance of the within-group dependence. Variance inflation is present whether statistically significant or not. Third, power to determine whether ICCs differ from zero is typically very low (Kenny et al., 1998). Consequently, Kenny et al. (2002) recommended that “for small-group studies with four or more persons, the correct strategy is to assume the data are nonindependent and not to bother to test whether the nonindependence is statistically significant” (p. 925; see also Murray, 1998, p. 232 and Roberts & Roberts, 2005).

One might also argue that basing degrees of freedom on the number of groups is too conservative and unnecessarily reduces power. However, basing degrees of freedom on the number of groups has little effect on power, if there are sufficient groups (20 or more groups across all conditions; Kenny et al., 1998). This principle is illustrated in the present study from the Body Project where there were 36 total groups and thus considerable degrees of freedom at the group level. The change in degrees of freedom had little impact on  $p$ -values for Model 2b (see Table 1). For example, adjusting for the within-group dependence for eating disorder symptoms resulted in a time  $\times$  condition  $F = 2.47$  with 34 denominator degrees of freedom ( $p = 0.13$ ). If we raise the denominator degrees of freedom to 230 (the degrees of freedom from the model that ignores group), the  $p$ -value changes to 0.12. Applying the same logic to the Thin Ideal Internalization variable, the  $p$ -value changes from 0.06 (34 degrees of freedom, group model) to 0.05 (230 degrees of freedom, non-group model). Only if we strictly adhere to conventions of statistical significance would we consider that a meaningful difference. Therefore, because the Body Project had many groups, basing degrees of freedom on the number of groups had little impact.

Of course, if there are not very many groups in the study, there will be fewer degrees of freedom at the group level and power will be lower. However, this may be a reflection of the fact that many intervention studies are under-powered to begin with (Kazdin & Bass, 1989) and adjusting for within-group dependence only exacerbates the problem. For example, Baldwin et al. (2005) reviewed group-administered treatment studies from the American Psychological Association Division 12’s empirically supported treatments list (Task Force, 1995, 1998). Among the 33 studies they reviewed, the median number of groups per condition was three and the median group size was about five. Thus if we designed an intervention study with the three conditions and each condition had three intervention groups of five people, we would have a total sample size of 45 ( $3 \times 3 \times 5 = 45$ ). Even if we ignored within-group dependence and treated person as the unit of analysis, this hypothetical study would have low power to detect anything but a large effect size ( $d = .80$ ). Adjusting for within-group dependence would reduce the power further. Therefore, given that properly calculating degrees of freedom in group intervention studies tends not to create *new* power problems and that not adjusting degrees of freedom tends to increase the probability of Type I errors (Murray et al., 1996; Varnell et al., 2001), most methodologists recommend basing degrees of freedom on the number of groups.

The best way to increase power is to plan for within-group dependence when designing a group-based intervention study. Planning for within-group dependence will allow researchers to measure covariates that may account for within-group dependence and ensure that they include a sufficient number of groups per condition. Murray (1998) provides formulas that can assist group researchers in their sample size calculations, although these formulas require ICC estimates. Our results provide ICC estimates for six outcome variables. Thus, researchers using similar outcome measures can use these estimates in planning their group studies. Given that very few ICC estimates have been published, it would be very helpful if group intervention researchers report ICCs along with the results of their study or make their data available so that ICCs could be computed.

Power in any intervention study, group-based or not, is affected by the intervention effect size. No matter how large and rigorous our design, if an intervention is weak, we are not likely to observe meaningful effects. Thus, we can increase statistical power by increasing the potency of our interventions. We may be able to increase the potency of interventions by improving intervention delivery, receipt, and adherence (Lichstein, Riedel, & Grieve, 1994; Shadish, Cook, & Campbell 2002). Furthermore, a better understanding of the behaviors we hope to change or prevent and of how people change, including barriers to change, will help researchers refine their interventions (cf. Susser, 1995). This increased understanding together with strengthening implementation should increase the potency of our interventions, which will increase the likelihood that researchers will observe meaningful effects in their intervention trials.

Another source of nonindependence in group-administered research is therapists (Wampold & Serlin, 2000). Groups may become similar (dissimilar) to one another because they are facilitated by the same therapist. Thus, accounting for nonindependence due to groups may not be sufficient to protect against Type I errors. A key issue is what proportion of the nonindependence is accounted for by groups and therapists, respectively (cf. Murray et al., 1996). If therapists account for a minimal amount of the nonindependence, including groups in the analysis will account for the majority of the nonindependence and the inflation of Type I errors will be minimal. The opposite is true if therapists account for a large proportion of the nonindependence. Unfortunately, we were unable to address this issue with our data because some groups were co-facilitated and some conditions had a single therapist who co-facilitated all groups in the intervention. We are unaware of any research that directly addresses whether therapists produce nonindependence in the data above and beyond the group. Future research should address this issue.

## Conclusions and Implications

We recognize that the implications of this study may initially appear daunting to researchers of group-administered interventions. Our goal is to recognize the importance of these issues and offer real-world data on the magnitude of these effects, as well as illustrate how to account for within-group dependence in a repeated measures analysis and provide recommendations for ways group researchers can plan for within-group dependence in future research. Researchers in a number of disciplines where within-group dependence is an issue (e.g., education, public health, industrial/organizational psychology) have begun to routinely adapt their research design and analysis to account for within-group dependence (e.g., Varnell, Murray, Janega & Blitstein, 2004). We are optimistic that group-administered psychosocial intervention researchers will also be successful. To that end, we encourage future group prevention and intervention research to report the ICCs for primary outcomes and to account for within-group dependence in their inferential tests.

This research also has implications for evidenced based practice (Baldwin et al., 2005). Many efforts to identify standards of evidence based practice have focused on identifying treatment or intervention packages shown to be effective in methodologically rigorous research (Task Force, 1995, 1998). As we have shown, group-administered interventions add an extra layer of complexity to these efforts. Consequently, future efforts to identify evidence based interventions should consider whether within-group dependence was accounted for in the analysis of group data.

Additionally, expanding our conception of evidence based practice to include group processes known to predict good outcomes (e.g., cohesion; Burlingame et al., 2002) seems warranted. Indeed, these results provide evidence for the power of the group—for the majority of outcomes between group differences were at least as important as between intervention differences. An important step in group-based intervention research is to understand when and how within-group dependence develops so that we may foster processes that lead to good group outcomes. Given rising health care costs and the potential of group-based interventions to mitigate those costs, attending to within-group dependence in the evaluation of group-based interventions and in understanding group dynamics is an important research priority.

## Appendix

### Mixed Model Time x Condition Analysis - SAS PROC MIXED

```
proc mixed;
  class cond group id time;
  model y=cond time cond*time/ddf=x,x,x;
  repeated time/type=cs subject=id(group*cond);
  random time/type=cs subject=group(cond) g;
```

The proc mixed statement calls up the mixed-model routine in SAS. The class statement specifies categorical variables, in this case cond (intervention condition), group (intervention group), id (participant id), and time (time point). The model statement first specifies the dependent variable (y) and then the fixed effects in the model. In a time x condition analysis we include main effects for condition, time, and the time x condition interaction. The ddf statement allows the user to specify the denominator degrees of freedom for the fixed effects.

In this syntax, the repeated statement allows for correlation among the repeat observations on the same group members ( $\sigma_{m:g;c}^2$ ), which is labeled CS id(group\*cond) in the SAS covariance parameter output. Including time on this the repeated line specifies what variable identifies the repeat observations (i.e., the repeated measures were over time). The subject=id(group\*cond) specifies that repeated measures were taken on participants and that those participants were nested within groups and condition. The type=cs specifies that the structure of the within-person covariance matrix is compound symmetry. If more than two time points are available, other structures may be more appropriate (Littell et al., 1996). The residual variance ( $\sigma_e^2$ ) is estimated by default and is labeled residual in the SAS output.

The random statement allows for the correlation among group members. The subject=group(cond) options identifies that there were repeated observations on each group. The time/type=cs specification creates two dummy variables for time, one for baseline and one for post-treatment and specifies that the structure of the between-person covariance matrix is compound symmetry. The g on the random line requests that SAS produce the between-person covariance matrix. The off-diagonal element in this matrix is the covariance between the two dummy

variables and corresponds to random effect for group ( $\sigma_{g:c}^2$ ). SAS labels this parameter as CS group(cond) in the covariance parameter output. The diagonal element in the between-person covariance matrix is the sum of the random effect for group and random effect for the time x group interaction. Rather than reporting this sum in the covariance parameter output, SAS reports only the random effect for the time x group interaction ( $\sigma_{tg:c}^2$ ) which is labeled variance group(cond).

Occasionally the random effects for group and the time x group interaction will be estimated as negative. This can occur for three reasons. First, the group environment or group assignment process may produce differentiation among the group members. Second, the model may be misspecified. That is, the data are independent but we include a parameter that models nonindependence. A third reason for negative within-group dependence is the negative bias for variance components (Murray et al. 1996). When the true value of a variance component is positive but close to zero or zero, the probability that the variance component will be estimated as negative exceeds 50%. The probability will increase as the number of groups per condition decreases or the number of members per group decreases. It is difficult to differentiate between these three sources of negative within-group dependence. Furthermore, ignoring negative within-group dependence often makes the test of the intervention effect overly conservative (Murray et al., 1996). Consequently, we recommend that when possible researchers model the negative within-group dependence.

Modeling negative within-group dependence in repeated-measures models can be challenging in PROC MIXED and other mixed-model regression software. One option in PROC MIXED is to relax the non-negativity constraint for variance components using the nobound option. Many researchers find nobound dissatisfying because it allows variance components, which by definition are greater than or equal to zero, to be negative. However, models that relax the non-negativity constraint typically produce identical model fit and identical estimates of the within-group dependence as models that use theoretically legitimate methods for modeling negative within-group dependence.

The syntax we provide allows the random effect for group to be negative by modeling it as a covariance rather than a variance. The time x group interaction continues to be modeled as a variance and is thus assumed to be positive. This assumption was met for all outcomes in this report. This specification also assumes that the sum of the random effect for group and the random effect for the time x group interaction is greater than the absolute value of the random effect for group. This assumption was met for four of the six outcomes. When the latter assumption was violated (BDI and CBCL-E), PROC MIXED produced identical model fit and estimates as a model that relaxed the non-negativity constraint. However, the model produced theoretically impossible values, such as correlations less than  $-1$ . Thus, for the BDI and CBCL-E we dropped the random effect for group and re-estimated the intervention effect. The results were similar to and produced the same substantive conclusion about the intervention effect as a model that included a random effect for group but relaxed the non-negativity constraint. Consequently, we presented the results for the model that dropped the random effect for group.

In cases where it does not seem justifiable to drop the random effect for group or if the time x group interaction is negative, one option is to estimate intervention effects with the post-treatment data and include the baseline value of the dependent variable as a covariate. Negative within-group dependence can be easily accommodated in those models by the methods outlined by Kenny et al. (2002). We estimated the treatment effects for the BDI and CBCL-E using the adjusted posttest data and obtained similar results to the repeated-measures analysis. However, we reported the results of the repeated measures analyses as that is the focus of our report.



To allow for heterogeneous random effects across conditions the repeated and random lines are changed to:

repeated time/type=cs subject=id(group\*cond) group=cond;

random time/type=cs subject=group(cond) group=cond;

The group=cond option allows the random effects to vary across conditions.

## References

- Achenbach, TM. Manual for the Child Behavior Checklist/4–18 and 1991 profile. University of Vermont Department of Psychology; Burlington, VT: 1991.
- Antonuccio DO, Thomas M, Danton WG. A cost-effectiveness analysis of cognitive behavior therapy and fluoxetine (prozac) in the treatment of depression. *Behavior Therapy* 1997;28:187–210.
- Baldwin SA, Murray DM, Shadish WR. Empirically supported treatments or Type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology* 2005;73:924–935. [PubMed: 16287392]
- Beck, AT.; Steer, RA.; Brown, GK. Manual for Beck Depression Inventory-II. Psychological Corporation; San Antonio, TX: 1996.
- Berscheid E, Walster W, Bohrnstedt G. The happy American body: A survey report. *Psychology Today* 1973;7:119–131.
- Burlingame, GM.; Fuhrman, A.; Johnson, JE. Cohesion in group psychotherapy. In: Norcross, JC., editor. *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients*. Oxford University Press; New York: 2002. p. 71-87.
- Burlingame GM, Kircher JC, Honts CR. Analysis of variance versus bootstrap procedures for analyzing dependent observations in small group research. *Small Group Research* 1994;25:486–501.
- Burlingame, GM.; MacKenzie, KR.; Strauss, B. Small-group treatment: Evidence for effectiveness and mechanisms of change. In: Lambert, MJ., editor. *Bergin and Garfield's Handbook of psychotherapy and behavior change*. Vol. 5th ed.. Wiley; New York: 2004. p. 647-696.
- Clarke, G.N.; Lewinsohn, PM.; Hops, H. Adolescent Coping With Depression Course. 1990. The therapist manual and the adolescent workbook may be downloaded for free from the Internet at <http://www.kpchr.org/acwd/acwd.html>
- Dishion TJ, McCord J, Poulin F. When interventions harm: Peer groups and problem behavior. *American Psychologist* 1999;54:755–764. [PubMed: 10510665]
- Donner A, Birkett N, Buck C. Randomization by cluster: Sample size requirements and analysis. *American Journal of Epidemiology* 1981;114:322–326.
- Fairburn, CG.; Cooper, Z. The Eating Disorder Examination. In: Fairburn, CG.; Wilson, GT., editors. *Binge eating: Nature, assessment, and treatment*. Vol. 12th ed.. Stanford University Press; Stanford, CA: 1993. p. 317-360.
- Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry* 1960;23:56–61.
- Herzog TA, Lazev AB, Irvin JE, Juliano LM, Greenbaum PE, Brandon TH. Testing for group membership effects during and after treatment: The example of group therapy for smoking cessation. *Behavior Therapy* 2002;33:29–43.
- Kazdin AE, Bass D. Power to detect differences between alternative treatments in comparative psychotherapy outcomes. *Journal of Consulting and Clinical Psychology* 1989;57:138–147. [PubMed: 2925964]
- Kenny DA, Judd CM. Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin* 1986;99:422–431.
- Kenny, DA.; Kashy, DA.; Bolger, N. Data analysis in social psychology. In: Gilbert, DT.; Fiske, ST.; Lindzey, G., editors. *The handbook of social psychology*. Vol. 1. Oxford Press; New York: 1998. p. 233-265.



- Kenny DA, Mannetti L, Peirro A, Livi S, Kashy DA. The statistical analysis of data from small groups. *Journal of Personality and Social Psychology* 2002;83:126–137. [PubMed: 12088122]
- Kim D-M, Wampold BE, Bolt DM. Therapist effects in the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research* 2006;16:161–172.
- Lichstein KL, Riedel BW, Grieve R. Fair tests of clinical trials: A treatment implementation model. *Advances in Behavior Research and Therapy* 1994;16:1–29.
- Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD. SAS system for mixed models. SAS Institute; Cary, NC: 1996.
- Maxwell, SE.; Delaney, HD. Designing experiments and analyzing data: A model comparison approach. Vol. 2nd ed.. Lawrence Erlbaum Associates; Mahwah, NJ: 2004.
- Morgan-Lopez AA, Fals-Stewart W. Analytic complexities associated with group therapy in substance abuse treatment research: Problems, recommendations, and future directions. *Experimental and Clinical Psychopharmacology* 2006;14:265–273. [PubMed: 16756430]
- Murray, DM. Design and analysis of group-randomized trials. Oxford University Press; New York: 1998.
- Murray DM, Blitstein JL. Methods to reduce the impact of intraclass correlation in group-randomized trials. *Evaluation Review* 2003;27:79–103. [PubMed: 12568061]
- Murray DM, Hannan PJ, Baker WL. A Monte Carlo study of alternative responses to intraclass correlation in community trials. Is it ever possible to avoid Cornfield's penalties? *Evaluation Review* 1996;20:313–337. [PubMed: 10182207]
- Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials* 2005;2:152–162. [PubMed: 16279137]
- Rohde P, Clarke GN, Mace DE, Jorgensen JS, Seeley JR. An efficacy/effectiveness study of cognitive-behavioral treatment for adolescents with comorbid major depression and conduct disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* 2004;43:660–668. [PubMed: 15167082]
- Rohde P, Seeley JR, Kaufman NK, Clarke GN, Stice E. Predicting time to recovery among depressed adolescents treated in two psychosocial group interventions. *Journal of Consulting and Clinical Psychology* 2006;74:80–88. [PubMed: 16551145]
- Shadish, WR.; Cook, TD.; Campbell, DT. Experimental and quasi-experimental designs. Houghton Mifflin; Boston, MA: 2002.
- Snedecor, GW.; Cochran, WM. Statistical methods. The Iowa State University Press; Ames, IA: 1980.
- Stice E, Fisher M, Martinez E. Eating disorder diagnostic scale: Additional evidence of reliability and validity. *Psychological Assessment* 2004;16:60–71. [PubMed: 15023093]
- Stice E, Shaw H, Burton E, Wade E. Dissonance and healthy weight eating disorder prevention programs: a randomized efficacy trial. *Journal of Consulting and Clinical Psychology* 2006;74:263–275. [PubMed: 16649871]
- Susser M. Editorial: The tribulations of trials--Intervention in communities. *American Journal of Public Health* 1995;85:156–158. [PubMed: 7856769]
- Task Force on Promotion and Dissemination of Psychological Procedures. Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *Clinical Psychologist* 1995;48:3–23.
- Task Force on Promotion and Dissemination of Psychological Procedures. Update on empirically supported therapies II. *Clinical Psychologist* 1998;51:3–16.
- Varnell, S.; Murray, DM.; Hannan, PJ.; Baker, WL. Intraclass correlation at the level of the unit of intervention in a randomized clinical trial: Implications for analysis; Paper presented at the Annual Meeting of the American Evaluation Association; St. Louis, MO. 2001;
- Varnell S, Murray DM, Janega JB, Blitstein JL. Design and analysis of group-randomized trials: A review of recent practices. *American Journal of Public Health* 2004;94:393–399. [PubMed: 14998802]
- Wampold BE, Serlin RC. The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods* 2000;5:425–433. [PubMed: 11194206]
- Yalom, ID.; Leszcz, M. The theory and practice of group psychotherapy. Vol. 5th ed.. Basic Books; New York: 2005.

Zucker D. An analysis of variance pitfall: The fixed effect analysis in a nested design. *Educational and Psychological Measurement* 1990;50:731–738.

**Table 1**  
**Estimated Intervention Effects from the Mixed-Model Repeated Measures ANOVA Analyses**

	Time x Condition <i>F</i>				$\omega^2$		<i>ICC<sub>mt:g;c</sub></i> (95% CI)	<i>VIF<sub>mt:g;c</sub></i>
	Model without Group	Model with Group	Model without Group	Model with Group	Model 1a	Model 1b		
MAPS	Model 1a	Model 1b	Model 1a	Model 1b	Model 1a	Model 1b		
Beck Depression Inventory	<i>F</i> (1, 164) = 0.17 <i>p</i> = 0.68	<i>F</i> (1, 16) = 0.14 <i>p</i> = 0.71	0.00	0.00	0.00	0.00	0.03 (-0.03 – 0.17)	1.28
Hamilton Rating Scale for Depression	<i>F</i> (1, 175) = 2.78 <i>p</i> = 0.10	<i>F</i> (1, 16) = 1.61 <i>p</i> = 0.22	0.02	0.00	0.02	0.01	0.07 (-0.01 – 0.23)	1.59
Child Behavior Checklist - Externalizing	<i>F</i> (1, 122) = 0.03 <i>p</i> = 0.86	<i>F</i> (1, 16) = 0.001 <i>p</i> = 0.98	0.00	0.00	0.00	0.00	0.12 (0.03 – 0.32)	2.09
Body Project	Model 2a	Model 2b	Model 1a	Model 1b	Model 1a	Model 1b		
Thin-Ideal Internalization	<i>F</i> (1, 230) = 6.21 <i>p</i> = 0.01	<i>F</i> (1, 34) = 3.94 <i>p</i> = 0.06	0.04	0.04	0.04	0.04	0.11 (0.01 – 0.25)	1.57
Body Dissatisfaction	<i>F</i> (1, 230) = 7.44 <i>p</i> = 0.01	<i>F</i> (1, 34) = 6.34 <i>p</i> = 0.02	0.05	0.05	0.05	0.05	0.02 (-0.05 – 0.14)	1.12
Eating Disorder Symptoms	<i>F</i> (1, 230) = 2.92 <i>p</i> = 0.09	<i>F</i> (1, 34) = 2.47 <i>p</i> = 0.13	0.02	0.01	0.02	0.01	0.02 (-0.05 – 0.14)	1.13

MAPS = Making a Plan for Success;  $\omega^2$  = proportion of total variance accounted for by the intervention condition; *ICC<sub>mt:g;c</sub>* = intraclass correlation; CI = Confidence Interval; *VIF<sub>mt:g;c</sub>* = Variance Inflation Factor

**Table 2**  
**Intraclass Correlations and 95% Confidence Intervals**  
**for Each Outcome Measure by Intervention Condition**

	Intervention Condition	
	<i>ICC<sub>mt:g;c</sub></i> (95% CI)	<i>ICC<sub>mt:g;c</sub></i> (95% CI)
MAPS	Coping with Depression	Life Skills
Beck Depression Inventory	< .001 (-0.06 – 0.21)	0.10 (-0.01 – 0.42)
Hamilton Rating Scale for Depression	0.11 (-0.005 – 0.42)	0.03 (-0.05 – 0.29)
Child Behavior Checklist - Externalizing	0.18 (0.04 – 0.52)	0.09 (-0.02 – 0.39)
Body Project	Dissonance	Healthy Weight
Thin-Ideal Internalization	0.08 (-0.03 – 0.30)	0.12 (-0.002 – 0.34)
Body Dissatisfaction	0.01 (-0.08 – 0.18)	0.05 (-0.05 – 0.25)
Eating Disorder Symptoms	0 <sup>a</sup> (-0.06 – 0.11)	0.18 (0.04 – 0.42)

MAPS = Making a Plan for Success; *ICC<sub>mt:g;c</sub>* = intraclass correlation; CI = Confidence Interval.

<sup>a</sup> see footnote two regarding this ICC estimate.

**Table 3**  
**Sample sizes required to maintain 80% power for individual and group interventions for a medium effect size ( $d = .50$ )**

$d = .50$

Individual Intervention	Group Intervention	ICC = -.05	ICC = .00	ICC = .05	ICC = .15	ICC = .30
$N = 128$	$m = 5$	120 ( $G = 24$ )	140 ( $G = 28$ )	170 ( $G = 34$ )	220 ( $G = 44$ )	290 ( $G = 58$ )
	$m = 10$	100 ( $G = 10$ )	160 ( $G = 16$ )	220 ( $G = 22$ )	320 ( $G = 32$ )	500 ( $G = 50$ )
	$m = 15$	90 ( $G = 6$ )	180 ( $G = 12$ )	270 ( $G = 18$ )	450 ( $G = 30$ )	690 ( $G = 46$ )

The sample size calculations are for two conditions and assume an alpha level of .05 and a two-tailed test. Values in the intraclass correlation (ICC) columns are the total number of participants ( $N$ ) in the entire study required to maintain 80% power. The total number of groups ( $G$ ) included in the study was calculated by dividing  $N$  by number of members per group ( $m$ ).

