Supplemental Appendix to

Experiments with More Than One Random Factor:

Designs, Analytic Models, and Statistical Power

Charles M. Judd

University of Colorado Boulder

Jacob Westfall

University of Texas at Austin

David A. Kenny

University of Connecticut

Appearing in *Annual Review of Psychology*, 2017

Author emails: Charles.Judd@colorado.edu
Jake.Westfall@utexas.edu
David.Kenny@uconn.edu

Table of Contents

Below is the list of topics discussed in this document. They are hyperlinked and so you can click to go the topic in which you are interested. We refer to the chapter as *JWK* throughout this appendix.

Missing Data

In JWK, there is very little discussion of missing data. However, in virtually every study missing data are bound to happen. Sometimes they happen randomly when there is an equipment malfunction. However, very often missingness occurs non-randomly and is due to a deliberate choice of the researcher. For instance in reaction-time studies, it is a standard practice to treat all very long or very short responses as missing data.

Fortunately, missing data usually do not present analytic problems for mixed model programs. Estimation proceeds without any estimation of the missing values, i.e., imputation. Generally the code that we give later in this appendix works even when there are missing data, unlike traditional repeated measures ANOVA, which generally requires no missing data. However, there are concomitant losses of power, due not only to the lower sample sizes but also due to lack of balance in the data structure created by not having complete data.

Critical to the analysis of studies with missing data is an understanding of the process that creates the missing data. If it is the case that missing data are at random, then the mixed modeling software produces unbiased estimates. If, however, the missing data are not at random, then in principle the results might well be biased. Generally two different types of non-random missingness are defined: missing based on a known or measured variable and missing based on an unknown or unmeasured variable. For the types of studies discussed in JWK, we almost always have missingness due to unknown variables. For instance, when cases are removed because the scores are exceptionally large or small, the missingness is due to unmeasured variables because we do not know what variables made the measurements very small or very large; if missingness is due to unmeasured variables, there would then be concerns about missingness leading to biased parameter estimates and standard errors.

There are modern methods to handle missing observations for these designs. One highly recommended option is multiple imputation (Enders, 2010) which involves a three-step procedure. In the first step, missing values are estimated. These imputed values are random variables which can vary. That is, we estimate the value for the missing case and add a random error to the score. In the second step, multiple datasets are generated, each with a different set of imputed missing values, and the model analysis is conducted on each imputed dataset. In the third step, a standard error is computed for each effect, based on the pooled standard error from each analysis and the variability of the estimates across imputed datasets.

However, as explained by Enders et al. (2016), the first step of estimating missing values is not so simple with mixed models. To estimate the missing values, one needs to have estimates of the variances of the random variables. For instance, to estimate a particular missing value, the intercepts for the participant and the target which is being judged must be known. However, to have a good estimate of those intercepts, the variance of intercepts must be known, but to know that variance, the data must be used to get an estimate. So there is a chicken-egg problem of needing the variance to estimate the missing value, but needing an estimate of the missing value to estimate the variance. As Enders et al. (2016) detail, there are various ways of doing this that are quite complicated and computationally intensive. Fortunately, these methods are available in commonly used software, notably R, SAS, and Mplus. We should note that most of the discussion of these methods is for designs with nested rather than crossed random factors. Thus, if the design is crossed, some adaptation of these methods would likely be required.

We did conduct some simulations with the ***CNC*** design (participants crossed with treatment, targets nested within treatment, and targets and participants crossed). We tried out several different models in which the missing data were generated. The good news is that we found relatively little bias in any of the mean condition effects when the full mixed model analysis was used. The largest bias we found was about a 4 percent underestimation of the condition effect. However, when we computed means across targets (a by-participant analysis) or across participants (a by-target analysis) and analyzed the data by analysis of variance, we found considerably more bias, sometimes nearly seven percent. These simulations, though not at all definitive, suggest that mixed model analyses are far superior to the old-fashioned ANOVA analysis of means. As Enders et al. (2016) note "multilevel imputation techniques are still in their infancy" (p. 12), and so we would expect important developments to emerge in the near future.

Randomizing vs. Controlled Assignment Conditions

One aspect of having either participants or targets nested within condition rather than crossed with condition is that chance fluctuations in the intercepts of the sampled participants and targets add to the variability around the estimated condition difference and thereby detract from statistical power. Statistically speaking, when participants or targets are nested within condition, an additional variance term for the participant intercepts (V_P) or target intercepts (V_T), respectively, appears in the denominator of the noncentrality parameter. The traditional method of avoiding this, when the research context permits it, is to employ the ***Counterbalanced*** design, which effectively crosses both participants and targets with condition without requiring participants to respond to each target multiple times, as in the ***CCC*** design. In more recent years however, flexible computer-based target presentation systems have also made it easy to employ alternative designs involving more complicated randomization. In what we call the ***Random Target Groups*** design, targets are randomly assigned to be half in condition A and the other half in condition B, but this random assignment of targets to condition takes place separately and independently for each participant. The occasional use of this design and related randomized designs in the literature suggests that many researchers believe these to be a statistically superior way of crossing both participants and targets with condition compared to the more traditional ***Counterbalanced*** design. As we demonstrate below, this intuition is mistaken, at least as far as statistical power is concerned.

An important property of the ***Counterbalanced*** design is that every participant responds exactly as often under condition A as under condition B, and likewise every target is responded to exactly as often under condition A as under condition B. This is not true for the ***Random***

Target Groups design. Although the randomization does take place in such a way that every participant responds exactly as often under condition A as under condition B, it is often the case that, by chance, some targets are responded to far more often under condition A than condition B or vice versa. This imbalance in the design is a source of statistical inefficiency.

We conducted a small simulation comparing this design to the ***Counterbalanced*** design to illustrate the power difference. Specifically we empirically computed the ratio of the non-centrality parameter for the ***Counterbalanced*** design over that parameter from the ***Random Target Groups*** design (ncp_{CB}/ncp_{RTG}) for an experiment with total number of targets, $q = 10$, varying the total number of participants (*p*) from 4 to 16 and the proportion of target intercept variance (V_T) from 10% to 50%, with 2000 iterations for each combination of varying parameters. The VPCs other than V_T were set to $V_E = .5(1 - V_T)$, $V_P = .3(1 - V_T)$, $V_{P\times C} = V_{T\times C} = .1(1 - V_T)$, and $V_{P\times T} = 0$. When the number of participants is small (*p* < 10), the ***Random Target Groups*** design is slightly but noticeably less efficient than the ***Counterbalanced*** design, having about 92% to 98% of its efficiency for the parameter values studied. When the number of participants is larger (*p* > 10), the difference between the designs is negligible. There is also some suggestion in the simulation results that the relative efficiency of the ***Counterbalanced*** design is increased when V_T is greater, but if so, the differences are small. Our conclusion is that the ***Counterbalanced*** design dominates the ***Random Target Groups*** design, in the sense that it does at least as well for all combinations of parameter values, and demonstrably better for some combinations.

The ***Random Target Groups*** design is only one of many conceivable randomized designs that one could employ in an experiment with crossed random factors. However, we believe that for any of these randomized designs, there is always some comparable non-randomized design that dominates the randomized design in terms of statistical power. Our recommendation is, therefore, that randomized designs such as the ***Random Target Groups*** design should be avoided whenever possible, and non-randomized counterparts such as the ***Counterbalanced*** design used instead.

Replicates

JWK begins its discussion of the general model for designs with random participants and targets with a fully crossed design with multiple replicates. By this we mean multiple observations of the same participant with the same target in the same condition. (Recall the distinction in JWK between replications and replicates: the former refers to multiple repeated instances of a design and the latter to multiple observations in each participant-by-target cell of a given design.) We denote the number of replicates as *n*.

Classically replicates are treated as nested: Within each cell of the design, more than one observation is measured. Alternatively they can be treated as crossed. When crossed the replicates need to be designated as different variables. Sometimes this is fairly obvious such as when two different measures are used to measure the outcome. For example, to measure how extroverted the target is, each participant is asked how social and how talkative the target is. Alternatively, the replicates are temporally ordered, and the designation is between the first and second replicate. We initially discuss replicates as nested and then turn our attention to them as crossed.

Replicates as Nested

The inclusion of multiple replicates does not add any new variance components. Rather it allows the unconfounding of variance components. Consider the ***CCC*** design. If there are multiple replicates, the variance due to error and PxTxC interaction can be separated.

Not surprisingly, multiple replicates increase power. How much they increase power is a function of how large the residual error variance is and how many participants and targets there are in the study. At the limit, if there were no error variance, then the multiple replicates would have the very same value and there would be no increase in power. Moreover, the smaller the error variance, the less impact on statistical power there is by adding replicates. Likewise, the larger the participant and target sample sizes, the less impact on power there is by adding replicates.

Adding multiple replicates increases the number of responses per participant. In an experiment for which it has been determined that no more than, say, 60 responses can be collected from each participant, there is still a choice of how many unique targets each participant responds to. For example, each participant could respond to 60 targets one time each, 30 targets twice each, 20 targets three times each, and so on, all the way down to responding to a single target 60 times.

What are the advantages of having more than one replicate? Is it better to have an experiment with more targets and fewer replicates, or with fewer targets and more replicates? To answer these questions, we turn to the ***CNC*** design, with participants crossed with condition and targets nested within condition. We compare this design with what we call the ***Multiple Replicates*** design, involving still the same number of responses per participant, with more replicates but necessarily fewer targets. It turns out that, in terms of statistical power, it is always better in terms of power to have more targets and fewer replicates compared to fewer targets and more replicates. In fact, power to detect the condition difference is always highest when the number of replicates is one. This is easiest to see by inspecting the non-centrality parameter for the ***Multiple Replicates*** design and considering a simple change of variables that allows us to examine the effect of varying the number of replicates while holding constant the total number of responses made by each participant. Let *n* be the number of replicates, $\tau = qn$ be the total number of responses made by each participant, and $\pi = q/\tau$ be the proportion of responses by each participant that are made to a unique target. Note that having $\pi = 1$ implies that there is a single replicate and every response is made to a unique target, while lower values of π imply having more replicates and fewer targets. The non-centrality parameter for the ***Multiple Replicates*** design is then

$$ncp_{MR} = \frac{d}{2\sqrt{\frac{V_{P\times C}}{p} + \frac{[V_T + V_{T\times C}]}{q} + \frac{V_E}{pqn} + \frac{V_{P\times T}}{pq}}} = \frac{d}{2\sqrt{\frac{V_{P\times C}}{p} + \frac{[V_T + V_{T\times C}]}{\pi\tau} + \frac{V_E}{p\tau} + \frac{V_{P\times T}}{p\pi\tau}}}.$$

All else being equal, this non-centrality parameter is highest when $\pi = 1$, in which case the ***Multiple Replicates*** design reduces to the ***CNC*** design. The statistical advantage of having more targets with fewer replicates is particularly pronounced when the targets are highly variable ($V_T + V_{T\times C}$ is high), when there is substantial Target × Participant interaction variance ($V_{P\times T}$ is high), when there are few participants (p is small) and when participants make a small number of total responses (τ is low).

There is, however, one potential advantage of having more than one replicate in an experiment. In crossed designs other than the ***CCC*** design, with only a single replicate, random variance due to Participant × Target interactions ($V_{P\times T}$; also Participant × Target × Condition interactions $V_{P\times T\times C}$) is empirically confounded with the residual variance and cannot be uniquely estimated. If the sizes of these variance components are of substantive theoretical interest to an experimenter, then it is necessary to include multiple replicates in order to estimate them. An example of a context where estimating these interactions might be of interest is in studies employing the Social Relations Model (Kenny, Kashy, & Cook, 1994), where the Participants also serve as the “Targets” in the study (Participants both perceive and are perceived by others). Indeed, in the Social Relations Model, what we have called Participant × Target interactions are commonly referred to as “relationship effects.” However, unless an experimenter is specifically interested in Participant × Target interactions or relationship effects, we recommend designing studies with only a single replicate and correspondingly more targets in order to maximize statistical power.

Replicates as Crossed

As was stated above, we can view replicates not as nested but as crossed. That is, the *n* replicates are not interchangeable and arbitrarily ordered, but rather they can be ordered within each cell of the design. One way to order them is by the order of measurement: Replicate one is the first measurement and replicate two is the second measurement. Very often, especially within Generalizability Theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972), replicates are viewed as an additional random factor in the design which we denote as O for occasion.

Consider replicates for the fully crossed or ***CCC*** design. Viewing replicates as nested and not crossed presumes that the O main effect and all of the O interactions are zero. It may indeed be plausible that these terms are zero. For instance, in Social Relations modeling of personality judgments, it is typically found that O x P and O x T sources of variance are very small. As an example consider Study 3 by Gross, Lakey, Edinger, Orehek, and Heffron (2009) who examined how students, the participants, evaluated their teachers, the targets. In this study, they had 74 students evaluate the same 4 teachers, and so participants and targets are crossed. They had no variable like condition, but they did have replicates due to the fact that each teacher was evaluated on 23 items, which Gross et al. treated as crossed and random. The P x O (Student x Occasion, i.e., item) term accounted for only 1.2 percent of the total variance and T x O (Teacher x Occasion) term accounted for less than one percent of the total. Thus, it would be reasonable in this study to treat those terms as zero. Generally, we feel that it might well be implausible that the O main effect is zero (e.g., reaction times might be quicker the second time).

It may, however, be reasonable to assume that all the O interactions are very small or even zero, and we could then treat replicates as nested, removing just the main effect of O as a covariate.

One major difference between treating replicates as fixed, as we did when earlier discussing replicates as nested, versus treating them as random is that if the replicates are fixed, then—assuming there are an equal number of replicates in each cell—it is totally permissible to compute the mean across replicates and analyze the data that way. However, if replicates are random, such a strategy is only permissible if all interactions with the replicates factor are zero.

Viewing the Designs with Two Random Factors as Dyadic Designs

In JWK we discussed briefly the idea that studies with both participants and targets are dyadic designs and one can benefit by thinking of them as such. Here we elaborate on that idea. First, we consider the typology of dyadic designs developed by Kenny et al. (2006) and how those designs map onto the designs considered in JWK. Second, we consider in more detail reciprocal designs, i.e., designs in which each person serves as both a participant and a target, and we show how the JWK designs can be turned into reciprocal designs. And third, we elaborate a re-parameterization of the model used in JWK, adapting an approach used in dyadic modeling.

Typology of Designs

Kenny et al. (2006) define three different types of dyadic designs; these can be mapped onto the JWK designs. One class of dyadic designs is what is called an SRM design. An SRM design is one which could be analyzed by the Social Relations Model. In these designs, each participant judges multiple targets and each target is judged by multiple participants. The JWK designs that are SRM designs are the ***CCC***, ***CNC***, ***NCC***, and ***NNC*** designs, i.e., designs in which Participants and Targets are crossed. All of these designs are called *half-block designs* within the SRM (Kenny et al., 2006) with participants called actors and targets called partners. (Typically in most SRM studies there are the random effects of participant and target, but no fixed effect like condition.) One example of the half-block design is Study 1 in Kenny, Horner, Kashy, and Chu (1992) where 113 participants observed videotapes of 32 targets sitting alone in a room. In actuality, there were two replications with 57 participants judging 32 targets and another 56 judging another 32. Observers were asked to judge the personalities of the targets and the focus was on the extent to which there was target variance.

The second class of dyadic designs considered in Kenny et al. (2006) is a one-with-many design. For this design, the "one" actor is paired with "many" different partners, but the partners are only paired to the one actor. One example of this design is a study in which therapists are paired with clients, with therapists being the "one" actor and clients being the "many" partners. Such designs map onto the ***CCN*$_P$, *CCN*$_T$, *CNN*$_P$, *NCN*$_T$, *NNN*$_P$**, and ***NNN*$_T$** designs in JWK, i.e., designs in which Participants are nested within Targets or vice versa. Note that for the ***CCN*$_P$**, ***NCN*$_P$**, and ***NNN*$_P$** designs, the participant is the one actor (e.g., the therapist) and targets are the many partners (e.g., the clients), whereas for ***CCN*$_T$**, ***CNN*$_T$**, and ***NNN*$_T$** designs, the participant are the many partners and target is the one actor.

The third class of dyadic designs is a standard dyadic design. In this design each person interacts with just one person who only interacts with that other person, i.e., participant and target are confounded. An example of such a study is one in which married people are asked how jealous they are about their partner. These designs conform to ***CCN*$_P$**, ***CCN*$_T$**, ***NNN*$_P$**, and ***NNN*$_P$** designs where both the number of targets per participant or participants per target equals just one. (When there is just one participant and target, the ***CNN*$_P$** and ***NCN*$_T$** designs become impossible because the ***CNN*$_P$** requires a minimum of two participants per target and the ***NCN*$_T$** requires a minimum of two targets per participant.) With one per cell, the ***CCN*$_P$** and ***CCN*$_T$** design collapse and become the same design, as do the ***NNN*$_P$** and ***NNN*$_T$** designs. In the ***CCN*** designs, the participant-target combinations are in both conditions, whereas in the ***NNN*** designs, they are in just one.

Reciprocal Designs

A key feature of dyadic research is that designs are often reciprocal in that each participant also serves as a target and each target also serves as a participant Most SRM designs are reciprocal in the sense that when actor A judges partner B, actor B also judges partner A. With reciprocal data there are two SRM correlations. One is the actor-partner correlation and the other is the dyadic correlation. As an example, consider a study of liking of group members of a large sorority. The actor-partner correlation refers to the correlation of how much a member particularly likes others with how much that member is particularly liked by others in the group. The dyadic correlation is the correlation between relationship effects: If Jane particularly likes Sally, does Sally also particularly like Jane?

As discussed in JWK, the ***CCC*** Design can be seen as a reciprocal design when condition refers to whether it is the participants rating the targets or vice versa. The ***CNC*** and ***NCC*** can be combined to form a reciprocal design, but it requires a bit of elaboration. Say we have a speed-dating study, and we have 4 Black men and 4 White men who go on dates with 4 Black women and 4 White women. Looking at the men's judgment of women's attractiveness, the design is ***NCC*** if we consider the race of the participant, but if we look at race of the target, then the design is ***CNN***. Looking at the women's judgment of men's attractiveness, the design is ***NCC*** if we consider the race of the participant, but if we look at race of the target, then the design is ***CNC***. Thus, if the design is reciprocal, there are two Condition variables, race of the participant and race of the target. Moreover, we need to consider the interaction of those two variables, i.e., preference for same race versus different race.

For an example for the ***NNC*** Design as reciprocal, we might have a speed dating study with gays and lesbians. Looking at Table 3 and the ***NNC*** Design in JWK, gay participants 1, 2, 3, go on speed dates with 3 other gays, whom we now denote as Targets 7, 8, 9. Also, lesbian participants 4, 5, 6, go on speed dates with 3 other lesbians, whom we now denote as Targets 10, 11, 12. The study would be reciprocal if we gather data from both members of the speed date. The condition variable in this study would be whether one was gay or lesbian.

We can also view reciprocal dyadic designs as combinations of pairs for the six "one-with-many" or nested designs. Those pairs are $\boldsymbol{CCN_P}$ and $\boldsymbol{CCN_T}$, $\boldsymbol{CNN_P}$ and $\boldsymbol{NCN_T}$, $\boldsymbol{NNN_P}$, and $\boldsymbol{NNN_T}$. The difference between these three pairs is the relationships between the "one" and the "many" with the experimental condition. For the ***CCN*** designs, the pair responds twice, once in each condition; for the ***CNN*** design, the ones are crossed with condition and the many's are nested; and for the ***NNN*** designs, both are nested in condition. So for instance, if the design has students nested in teachers, if condition was enjoyment of math and enjoyment of science, the design would be ***CCN***, if the condition was gender of the student, the design would be ***CNN***, and if condition was gender of the teacher, the design would be ***NNN***.

Re-parameterization of Models

We begin by considering the fully crossed or ***CCC*** design with a single replicate. For that design, we viewed the model as one in which, for the participant, we had three variance terms: We had P variance, P x C variance, and the covariance of the two. The SRM uses an alternative and mathematically equivalent way to parameterize this design. Instead of the thinking of the data as one outcome variable measured in two conditions, we could alternatively think that there are two outcomes, one being the response for condition A and other as the response for condition B. In SRM parlance, we have two half blocks for two different variables. That is, we have a P x T data structure for condition A and a second one for condition B.

It is relatively easy to re-parameterize the model to allow for separate variances by condition and allow for a covariance between conditions. To do so, we create what are called *indicator variables*. We define two new condition variables: C1, a dummy variable which is coded as 0 and 1 and C2 which is coded as 1 and 0. Note that C1 and C2 are perfectly negatively correlated. We then have the random effects interact with both C1 and C2. For instance, the SAS code for CNC design is as follows:

```
proc mixed covtest;
class participant target;
model y = c  / solution ddfm=satt;
random C1 C2/sub=participant type=un;
random C1 C2/sub=target type=vc;
run;
```

In this formulation, consider the two variances and covariance estimated for Participant. The C1 variance is the variance in participants' intercepts in Condition 1, which is denoted as v_1, and the C2 variance is the variance in participants' intercepts in condition 2, which is denoted as v_2. The covariance of the two intercepts is denoted as c_{12}. Considering the three terms involving P in the formulation used in JWK:

P: $[(v_1 + v_2)/2 + c_{12}]/4$

PxC: $[(v_1 + v_2)/2 – c_{12}]/4$

Cov(P,PxC): $(v_1 – v_2)/4$

Note that the P variance in the formulation used in JWK is the average of the two measurements across conditions, whereas the P x C variance measures the variance of the difference divided by two. Interestingly, the covariance reflects the difference in the two P variances, i.e., the heterogeneity of the two condition variances. Note that for T, the variance due to C1 is the variance of the target intercepts for condition 1, and C2 is the variance of the target intercepts for condition 2. There is no covariance as different targets are in each condition. In this formulation, we do not need to force these two variances to be equal as we did in the JWK formulation.

Non-Centrality Parameters and Degrees of Freedom

Below are the non-centrality parameters (*ncp*) and approximate degrees of freedom (*df*) for each design. The variables *p* and *q* are always the *total* numbers of participants and targets, respectively. For the replication designs, *r* is the number of replications.

To simplify the presentation, we define for each design the residual variance as *E*, using it to indicate the residual variance in each design with all of the random components of variance that are confounded with it. Thus, it is identical to bracketed residual variance components in JWK's Tables 4 and 5.

CCC design

$$E = [\sigma_E^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{E}{2pq} + \frac{\sigma_{P\times C}^2}{p} + \frac{\sigma_{T\times C}^2}{q}}}$$

$$df = \frac{(E + q\sigma_{P\times C}^2 + p\sigma_{T\times C}^2)^2}{\frac{(E)^2}{(p-1)(q-1)} + \frac{(E + q\sigma_{P\times C}^2)^2}{p-1} + \frac{(E + p\sigma_{T\times C}^2)^2}{q-1}}$$

CNC design

$$E = [\sigma_E^2 + \sigma_{P\times T}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{E}{pq} + \frac{\sigma_{P\times C}^2}{p} + \frac{[\sigma_T^2 + \sigma_{T\times C}^2]}{q}}}$$

$$df = \frac{(E + q\sigma_{P\times C}^2/2 + p[\sigma_T^2 + \sigma_{T\times C}^2])^2}{\frac{(E)^2}{(p-1)(q-2)} + \frac{(E + q\sigma_{P\times C}^2/2)^2}{p-1} + \frac{(E + p[\sigma_T^2 + \sigma_{T\times C}^2])^2}{q-2}}$$

NCC design

$$E = [\sigma_E^2 + \sigma_{P\times T}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{E}{pq} + \frac{[\sigma_P^2 + \sigma_{P\times C}^2]}{p} + \frac{\sigma_{T\times C}^2}{q}}}$$

$$df = \frac{(E + q[\sigma_P^2 + \sigma_{P\times C}^2] + p\sigma_{T\times C}^2/2)^2}{\frac{(E)^2}{(p-2)(q-1)} + \frac{(E + q[\sigma_P^2 + \sigma_{P\times C}^2])^2}{p-2} + \frac{(E + p\sigma_{T\times C}^2/2)^2}{q-1}}$$

***NNC* design**

$$E = [\sigma_E^2 + \sigma_{P\times T}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{2E}{pq} + \frac{[\sigma_P^2 + \sigma_{P\times C}^2]}{p} + \frac{[\sigma_T^2 + \sigma_{T\times C}^2]}{q}}}$$

$$df = \frac{(E + q[\sigma_P^2 + \sigma_{P\times C}^2]/2 + p[\sigma_T^2 + \sigma_{T\times C}^2]/2)^2}{\frac{(E)^2}{(p-2)(q-2)} + \frac{(E + q[\sigma_P^2 + \sigma_{P\times C}^2]/2)^2}{p-2} + \frac{(E + p[\sigma_T^2 + \sigma_{T\times C}^2]/2)^2}{q-2}}$$

***CCN*$_P$ design**

$$E = [\sigma_E^2 + \sigma_{T\times C}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{E}{2q} + \frac{\sigma_{P\times C}^2}{p}}}$$

$$df = p - 1$$

***CCN*$_T$ design**

$$E = [\sigma_E^2 + \sigma_{P\times C}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{E}{2p} + \frac{\sigma_{T\times C}^2}{q}}}$$

$$df = q - 1$$

_CNN_*_P_* design

$$E = [\sigma_E^2 + \sigma_T^2 + \sigma_{P\times T}^2 + \sigma_{T\times C}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{E}{q} + \frac{\sigma_{P\times C}^2}{p}}}$$

$$df = p - 1$$

_NCN_*_T_* design

$$E = [\sigma_E^2 + \sigma_P^2 + \sigma_{P\times T}^2 + \sigma_{P\times C}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{E}{p} + \frac{\sigma_{T\times C}^2}{q}}}$$

$$df = q - 1$$

_NNN_*_P_* design

$$E = [\sigma_E^2 + \sigma_T^2 + \sigma_{P\times T}^2 + \sigma_{T\times C}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{E}{q} + \frac{[\sigma_P^2 + \sigma_{P\times C}^2]}{p}}}$$

$$df = p - 2$$

_NNN_*_T_* design

$$E = [\sigma_E^2 + \sigma_P^2 + \sigma_{P\times T}^2 + \sigma_{P\times C}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{E}{p} + \frac{[\sigma_T^2 + \sigma_{T\times C}^2]}{q}}}$$

$$df = q - 2$$

***Counterbalanced* design**

$$E = [\sigma_E^2 + \sigma_{P\times T}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{\sigma^2_{P\times C}}{p} + \frac{\sigma^2_{T\times C}}{q} + \frac{E}{pq}}}$$

$$df = \frac{(q\sigma^2_{P\times C} + p\sigma^2_{T\times C} - 2E)^2}{\frac{(q\sigma^2_{P\times C} + 2E)^2}{p-2} + \frac{(p\sigma^2_{T\times C} + 2E)^2}{q-2} + \frac{4E^2}{(p-2)(q-2)}}$$

R(CCC) **design**

$$E = [\sigma^2_E + \sigma^2_{P\times T\times C}]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{rE}{2pq} + \frac{\sigma^2_{P\times C}}{p} + \frac{\sigma^2_{T\times C}}{q}}}$$

$$df = \frac{(E + q\sigma^2_{P\times C} + p\sigma^2_{T\times C})^2}{\frac{(E)^2}{(p-r)(q-r)} + \frac{(E + q\sigma^2_{P\times C})^2}{p-r} + \frac{(E + p\sigma^2_{T\times C})^2}{q-r}}$$

R(CNC) **design**

$$E = [\sigma^2_E + \sigma^2_{P\times T} + \sigma^2_{P\times T\times C}]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{rE}{pq} + \frac{\sigma^2_{P\times C}}{p} + \frac{[\sigma^2_T + \sigma^2_{T\times C}]}{q}}}$$

$$df = \frac{(E + q\sigma^2_{P\times C}/2 + p[\sigma^2_T + \sigma^2_{T\times C}])^2}{\frac{(E)^2}{(p-r)(q-2r)} + \frac{(E + q\sigma^2_{P\times C}/2)^2}{p-r} + \frac{(E + p[\sigma^2_T + \sigma^2_{T\times C}])^2}{q-2r}}$$

R(NCC) **design**

$$E = [\sigma^2_E + \sigma^2_{P\times T} + \sigma^2_{P\times T\times C}]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{rE}{pq} + \frac{[\sigma^2_P + \sigma^2_{P\times C}]}{p} + \frac{\sigma^2_{T\times C}}{q}}}$$

$$df = \frac{(E + q[\sigma_P^2 \ + \sigma_{P\times C}^2] + p\sigma_{T\times C}^2/2)^2}{\frac{(E)^2}{(p-2r)(q-r)} + \frac{(E + q[\sigma_P^2 + \sigma_{P\times C}^2])^2}{p-r} + \frac{(E + p\sigma_{T\times C}^2/2)^2}{q-r}}$$

R(NNC) **design**

$$E = [\sigma_E^2 + \sigma_{P\times T}^2 + \sigma_{P\times T\times C}^2]$$

$$ncp = \frac{\mu_1 - \mu_2}{2\sqrt{\frac{2rE}{pq} + \frac{[\sigma_P^2 + \sigma_{P\times C}^2]}{p} + \frac{[\sigma_T^2 + \sigma_{T\times C}^2]}{q}}}$$

$$df = \frac{(E + q[\sigma_P^2 + \sigma_{P\times C}^2]/2 + p[\sigma_T^2 + \sigma_{T\times C}^2]/2)^2}{\frac{(E)^2}{(p-2r)(q-2r)} + \frac{(E + q[\sigma_P^2 + \sigma_{P\times C}^2]/2)^2}{p-2r} + \frac{(E + p[\sigma_T^2 + \sigma_{T\times C}^2]/2)^2}{q-2r}}$$

SAS, SPSS, and R Code for Analysis

For R the “lme4” and the “lmerTest” packages must be installed and loaded. And the datafile is called “MyData.”

List of variables:

c: condition – a contrast or effect coded (-1 and +1) variable

y: outcome variable

participant: Each participant is given a unique code.

target: Each target is given a unique code.

r: For the replication designs, a variable denoting replication number. For these analyses, r must be a “factor”: `r <- factor(r)`

b: contrast or effect coded (-1 and +1) variable that splits participants into two groups with different sets of targets (only used in ***Counterbalanced*** design).

CCC design

```
SAS:    proc mixed covtest;
        class participant target;
        model Y = c;
        random intercept c/sub=participant type=un;
        random intercept c/sub=target type=un;
        random intercept /sub=participant*target;
        run;
```

Alternatively, only within SAS, the covariance between PxT with Error can be estimated using the following code:

```
        proc mixed covtest;
        class participant target;
        model Y = c / solution ddfm = satt;
        random intercept c/sub=participant type=un;
        random intercept c/sub=target type=un;
        random intercept c/sub=participant*target type=un;
        parms .3 .0 .3 .3 .0 .3  .3  .0.3 0.000001 /hold = 10;
        run;
```

What the above code does is fixes or “holds” the estimated error variance (the tenth parameter) to a very small value of 0.000001. The actual error variance is given by variance in the c*participant*target variance times two. The other 9 values in the “parms” statement are starting values of the nine variance and covariances of the model.

```
SPSS:   mixed y with c
        /fixed = c
        /print=solution testcov
        /random = intercept c | subject(participant) covtype(un)
        /random = intercept c | subject(target) covtype(un)
        /random = intercept | subject(participant*target).
        execute.

R:      model <- lmer(y ~ c + (c|participant) + (c|target) +
            (1|participant:target), data=myData)
        summary(model)
```

CNC design

```
SAS:    proc mixed covtest;
        class participant target;
        model Y = c/ solution ddfm=satt;
        random intercept c/sub=participant type=un;
        random intercept /sub=target;
        run;

SPSS:   mixed y with c
        /fixed = c
        /print=solution testcov
        /random = intercept c | subject(participant) covtype(un)
        /random = intercept | subject(target).
        execute.

R:      model <- lmer(y ~ c + (c|participant) + (1|target), data=myData)
        summary(model)
```

NCC design

```
SAS:    proc mixed covtest;
        class participant target;
        model Y = c/ solution ddfm=satt;
        random intercept /sub=participant;
        random intercept c/sub=target type=un;
        run;

SPSS:   mixed y with c
```

```
         /fixed = c
         /print=solution testcov
         /random = intercept | subject(participant)
         /random = intercept c | subject(target) covtype(un).
         execute.

R:       model <- lmer(y ~ c + (1|participant) + (c|target), data=myData)
         summary(model)
```

NNC design

```
SAS:     proc mixed covtest;
         class participant target;
         model Y = c/ solution ddfm=satt;
         random intercept /sub=participant;
         random intercept /sub=target;
         run;

SPSS:    mixed y with c
         /fixed = c
         /print=solution
         /random = intercept | subject(participant)
         /random = intercept | subject(target).
         execute.

R:       model <- lmer(y ~ c + (1|participant) + (1|target), data=myData)
         summary(model)
```

CCN$_P$ design

```
SAS:     proc mixed covtest;
         class participant target;
         model Y = c/ solution ddfm=satt;
         random intercept c/sub=participant type=un;
         random intercept /sub=target;
         run;

SPSS:    mixed y with c
         /fixed = c
         /print=solution testcov
         /random = intercept c | subject(participant) covtype(un)
         /random = intercept | subject(target).
         execute.

R:       model <- lmer(y ~ c + (c|participant) + (1|target), data=myData)
         summary(model)
```

CCN$_T$ design

```
SAS:    proc mixed covtest;
        class participant target;
        model Y = c/ solution ddfm=satt;
        random intercept /sub=participant;
        random intercept c/sub=target type=un;
        run;

SPSS:   mixed y with c
        /fixed = c
        /print=solution testcov
        /random = intercept | subject(participant) covtype(un)
        /random = intercept c | subject(target) covtype(un).
        execute.


R:      model <- lmer(y ~ c + (1|participant) + (c|target), data=myData)
        summary(model)
```

CNN$_P$ design

```
SAS:    proc mixed covtest;
        class participant target;
        model Y = c/ solution ddfm=satt;
        random intercept c/sub=participant type=un;
        run;

SPSS:   mixed y with c
        /fixed = c
        /print=solution testcov
        /random = intercept c | subject(participant) covtype(un).
        execute.

R:      model <- lmer(y ~ c + (c|participant), data=myData)
        summary(model)
```

NCN$_T$ design

```
SAS:    proc mixed covtest;
        class participant target;
        model Y = c/ solution ddfm=satt;
        random intercept c/sub=target type=un;
        run;

SPSS:   mixed y with c
        /fixed = c
        /print=solution testcov
        /random = intercept c | subject(target) covtype(un).
        execute.

R:      model <- lmer(y ~ c + (c|target), data=myData)
        summary(model)
```

NNN_P design

```
SAS:    proc mixed covtest;
        class participant target;
        model Y = c/ solution ddfm=satt;
        random intercept /sub=participant;
        run;

SPSS:   mixed y with c
        /fixed = c
        /print=solution
        /random = intercept | subject(participant).
        execute.

R:      model <- lmer(y ~ c + (1|participant), data=myData)
        summary(model)
```

NNN_T design

```
SAS:    proc mixed covtest;
        class participant target;
        model y = c / solution ddfm=satt;
        random intercept /sub=target;
        run;

SPSS:   mixed y with c
        /fixed = c
        /print=solution
        /random = intercept | subject(target).
        execute.

R:      model <- lmer(y ~ c + (1|target), data=myData)
        summary(model)
```

Counterbalanced design

```
SAS:    proc mixed covtest;
        class participant target;
        model y = c b b*c/ solution ddfm=satt;
        random intercept c/sub=participant type=un;
        random intercept c/sub=target type=un;
        run;

SPSS:   mixed y with c
        /fixed = c b b*c
        /print=solution testcov
        /random = intercept c | subject(participant) covtype(un)
        /random = intercept c | subject(target) covtype(un).
        execute.

R:      model <- lmer(y ~ c + b + b:c + (c|participant) + (c|target),
        data=myData)
        summary(model)
```

R(CCC) design

```
SAS:    proc mixed covtest;
        class participant target r;
        model Y = c r r*c/ solution ddfm=satt;
        random intercept c/sub=participant type=un;
        random intercept c/sub=target type=un;
        random intercept /sub=participant*target;
        run;

SPSS:   mixed y with c by r
        /fixed = c r r*c
        /print=solution testcov
        /random = intercept c | subject(participant) covtype(un)
        /random = intercept c | subject(target) covtype(un)
        /random = intercept | subject(participant*target).
        execute.

R:      model <- lmer(y ~ c + r + r:c + (c|participant) + (c|target) +
            (1|participant:target), data=myData)
        summary(model)
```

R(CNC) design

```
SAS:    proc mixed covtest;
        class participant target r;
        model y = c r r*c/ solution ddfm=satt;
        random intercept c/sub=participant type=un;
        random intercept /sub=target;
        run;
```

```
SPSS:   mixed y with c by r
        /fixed = c r r*c
        /print=solution testcov
        /random = intercept c | subject(participant) covtype(un)
        /random = intercept | subject(target).
        execute.
```

```
R:      model <- lmer(y ~ c + r + r:c + (c|participant) + (1|target),
        data=myData)
        summary(model)
```

R(NCC) design

```
SAS:    proc mixed covtest;
        class participant target r;
        model y = c r r*c/ solution ddfm=satt;
        random intercept /sub=participant;
        random intercept c/sub=target type=un;
        run;
```

```
SPSS:   mixed y with c by r
        /fixed = c r r*c
        /print=solution testcov
        /random = intercept | subject(participant)
        /random = intercept c | subject(target) covtype(un)
```

```
R:      model <- lmer(y ~ c + r + r:c + (1|participant) + (c|target),
        data=myData)
        summary(model)
```

R(NNC) design

```
SAS:    proc mixed covtest;
        class participant target r;
        model Y = c r/ solution ddfm=satt;
        random intercept /sub=participant;
        random intercept /sub=target;
        run;
```

```
SPSS:   mixed y with c by r
        /fixed = c r
        /print=solution
        /random = intercept | subject(participant)
        /random = intercept | subject(target).
        execute.

R:      model <- lmer(y ~ c + r + r:c + (1|participant) + (1|target),
data=myData)
        summary(model)
```

SAS, SPSS and R Sample Output

Below is the sample output with from the ***CNC*** design with 24 Participants and 48 Targets. At the end of the output we present the estimated condition effect size (***d***) for these data using the effect size formulas of Table 4 of JWK.

SAS Output

```
                         Model Information
        Data Set                     WORK.B
        Dependent Variable           y
        Covariance Structures        Variance Components, Unstructured
        Subject Effects              target, participant
        Estimation Method            REML
        Residual Variance Method     Profile
        Fixed Effects SE Method      Model-Based
        Degrees of Freedom Method    Satterthwaite

                      Class Level Information
    Class        Levels  Values
    participant      24  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
    target           24  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

                            Dimensions
                  Covariance Parameters        5
                  Columns in X                 2
                  Columns in Z per Subject    72
                  Subjects                     1
                  Max Obs per Subject        576

                      Number of Observations
               Number of Observations Read        1152
```

Number of Observations

Number of Observations Used	576
Number of Observations Not Used	576

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Intercept	**target**	25.8035	8.0832	3.19	0.0007
UN(1,1)	**participant**	15.4497	4.8525	3.18	0.0007
UN(2,1)	**participant**	-2.1766	2.7969	-0.78	0.4364
UN(2,2)	**participant**	9.6421	3.1402	3.07	0.0011
Residual		24.1110	1.5158	15.91	<.0001

Fit Statistics

-2 Res Log Likelihood	3658.7
AIC (Smaller is Better)	3668.7
AICC (Smaller is Better)	3668.8
BIC (Smaller is Better)	3658.7

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	6.2921	1.3269	40.2	4.74	<.0001
C	2.4041	1.2324	35.3	1.95	0.0591

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
C	1	35.3	3.81	0.0591

SPSS Output

Mixed Model Analysis

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	c	1		1	
Random Effects	Intercept + c[b]	2	Unstructured	3	participant
	Intercept	1	Variance Components	1	target
Residual				1	
Total		5		7	

a. Dependent Variable: y.

b. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.

Information Criteria[a]

-2 Restricted Log Likelihood	3658.727
Akaike's Information Criterion (AIC)	3668.727
Hurvich and Tsai's Criterion (AICC)	3668.833
Bozdogan's Criterion (CAIC)	3695.490
Schwarz's Bayesian Criterion (BIC)	3690.490

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: y.

Fixed Effects

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	40.182	22.485	.000
c	1	35.338	3.805	.059

a. Dependent Variable: y.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	6.292077	1.326929	40.182	4.742	.000	3.610632	8.973522
c	2.404083	1.232379	35.338	1.951	.059	-.096925	4.905091

a. Dependent Variable: y.

Covariance Parameters

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		24.111024	1.515847	15.906	.000	21.315772	27.272832
	UN (1,1)	15.449697	4.852526	3.184	.001	8.347759	28.593680
Intercept + c [subject = participant]	UN (2,1)	-2.176587	2.796912	-.778	.436	-7.658433	3.305260
	UN (2,2)	9.642117	3.140188	3.071	.002	5.092816	18.255206
Intercept [subject = target]	Variance	25.803473	8.083193	3.192	.001	13.964676	47.678815

a. Dependent Variable: y.

R Output

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ c + (c | participant) + (1 | target)
   Data: MyData

REML criterion at convergence: 3658.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2200 -0.5926 -0.0025  0.6057  2.9600

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 participant (Intercept) 15.450   3.931
             c            9.642   3.105    -0.18
 target      (Intercept) 25.803   5.080
 Residual                24.111   4.910
```

```
Number of obs: 576, groups:  participant, 24; target, 24

Fixed effects:
            Estimate Std. Error     df t value Pr(>|t|)    
(Intercept)    6.292      1.327 40.180   4.742 2.66e-05 ***
c              2.404      1.232 35.340   1.951   0.0591 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
  (Intr)
c -0.055
```

Using the estimated values, the estimate of effect size (***d***) is:

$$0.27759 = \frac{2(2.4041)}{\sqrt{15.4497 + 9.6421 + 25.8035 + 24.1110}}$$

References

Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. 1972. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: John Wiley.

Enders CK. 2010. *Applied Missing Data Analysis*. New York: Guilford Press.

Enders CK, Mistler SA, Keller, BT. 2016. Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychol. Meth.*, in press.

Gross JA, Lakey B, Edinger K, Orehek E, & Heffron D. 2009. Person perception in the college classroom: Accounting for tastes in students' evaluations of teaching effectiveness. *J. App. Soc. Psychol.* 39: 1609–1638.

Kenny DA, Horner C, Kashy DA, & Chu, L. 1992. Consensus at zero acquaintance: Replication, behavioral cues, and stability. *J. Pers.Soc. Psychol.* 62: 88-97.

Kenny DA, Kashy DA, Cook WL. 2006. *Dyadic Data Analysis*. New York: Guilford Press.