Power Analysis for Models of Change in Cluster Randomized Designs

Educational and Psychological
Measurement

2016, Vol. 77(1) 119–142

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164416641460

journals.sagepub.com/home/epm

Wei Li¹ and Spyros Konstantopoulos²

Abstract

Field experiments in education frequently assign entire groups such as schools to treatment or control conditions. These experiments incorporate sometimes a longitudinal component where for example students are followed over time to assess differences in the average rate of linear change, or rate of acceleration. In this study, we provide methods for power analysis in three-level polynomial change models for cluster randomized designs (i.e., treatment assigned to units at the third level). Power computations take into account clustering effects at the second and third levels, the number of measurement occasions, the impact of sample sizes at different levels (e.g., number of schools or students), and covariates effects. An illustrative example that shows how power is influenced by the number of measurement occasions, and sample sizes and covariates at the second or third levels is presented.

Keywords

power analysis, longitudinal cluster randomized designs, orthogonal polynomial contrasts

In recent years, there has been an increased interest in assessing the effects of educational interventions via experimental designs where schools are randomly assigned to a treatment or a control condition. An important part of the design phase of an experiment is power analysis. Statistical power is defined as the probability of detecting the treatment effect of interest when it exists (Boruch & Gomez, 1977; Cohen, 1988). A priori power computations are critical in designing experiments because they inform

¹University of Missouri, Columbia, MO, USA

²Michigan State University, East Lansing, MI, USA

Corresponding Author:

Spyros Konstantopoulos, Michigan State University, 461 Erickson Hall, East Lansing, MI 48824, USA.

Email: spyros@msu.edu

empirical researchers about the sampling scheme needed to detect a treatment effect. Specifically, *a priori* power analyses help educational researchers identify how big a sample is needed at each level to ensure a high probability (e.g., $> 80\%$) of detecting a treatment effect if it were true (Konstantopoulos, 2008; Lipsey 1990).

The recent resurgence of experiments in education has been an attempt to establish rigorous research in the field. That is, currently much of the empirical research in education uses randomized experiments that are typically large in scale. These field experiments allow education researchers to examine the effects of school, or student interventions on student performance. In addition, education experiments incorporate often times a longitudinal design where students are followed over time. The main objectives in these studies include assessing whether the treatment effects have lasting benefits or whether they fade over time. For example, the effect of a novel mathematics curriculum is evaluated through an experiment (i.e., novel vs. traditional mathematics curriculum) where measurements of student outcomes (e.g., mathematics achievement) are collected repeatedly over time (e.g., every spring for a few years).

In longitudinal experiments, repeated measures are collected for each individual (e.g., student) and as a result, each individual has his or her own trajectory, which is a function of time and indicates the rate of change over time (see Raudenbush & Bryk, 2002). The central goal in such studies is not only to estimate the treatment effect in the first year of the study (e.g., immediate effects), but also gauge longer term effects over time. For example, a researcher may be interested in the change or growth of mathematics achievement for students who use a novel mathematics curriculum *vis-a-vis* students who use a traditional mathematics curriculum. In this case, it is important for the researcher to compare trajectories of students who received the treatment (i.e., novel curriculum) versus those who did not receive the treatment (i.e., traditional curriculum).

The change in measurements over time does not always follow a linear trend. Instead, trajectories sometimes point to nonlinearities such as curvilinear trends. For example, Huttenlocher, Haight, Bryk, Seltzer, and Lyons (1991) studied how children's vocabulary is accelerated in early years. One way of defining trajectories of change is via polynomial functions (Raudenbush & Liu, 2001). The first-degree polynomial indicates linear rate of change, the second-degree polynomial indicates a quadratic rate of change, the third-degree polynomial indicates a cubic rate of change, and so forth. Treatment effects can be estimated for linear rates or nonlinear rates of change.

Studies about polynomial change may be viewed as having a nested structure. For example, repeated measurements are nested within individuals and this nesting needs to be taken into account in the design phase of the study as well as in the statistical analysis phase. Prior work has used two-level models (e.g., measurements within students) for longitudinal designs with repeated measurements of individuals (Raudenbush & Bryk, 2002). In particular, researchers have presented methods for power analysis of treatment effects in studies of polynomial change with one level of

nesting (see, Raudenbush & Liu, 2001). Power in these designs is a function of the magnitude of the treatment effect, the sample size of individuals, the number of measurement occasions in the study, and the frequency of measurements over time. Researchers should take into account all these parameters in the design phase of the experiment to ensure that the power of the treatment effect is adequate. Besides multilevel models there are alternative statistical methods to modeling growth trajectories such as the latent growth model (see Marcoulides & Schumacker, 2012).

However, populations in education have frequently more complicated structures. For example, students are also nested within classes or schools and so forth. In addition, education interventions typically assign entire groups such as schools randomly to treatment or control conditions. For instance, schools may be randomly assigned to a school assessment program or not. Thus, it seems natural to extend methods for power analysis for tests of treatment effects in studies of polynomial change from two to three levels. Consider for example, a nested structure where measurements are nested within students and students in turn are nested within schools. The first level is repeated measurements, the second level is students, and the third level is schools. Spybrook, Raudenbush, Liu, Congdon, and Martínez (2011) have reported in the optimal design manual formulae to calculate power for three-level polynomial change models without covariates, where the treatment is at the third level. Power analyses for longitudinal multilevel trials have also been discussed in clinical psychology (e.g., Jong, Moerbeek, & van der Leeden, 2010). Moreover, a recently published book has exclusively focused on power analysis in intervention trials for data that follow multilevel structures (Moerbeek & Teerenstra, 2016)

This study extends previous work by Raudenbush and Liu (2001) and formulae provided by Spybrook et al. (2011), and provide methods for power analysis of tests of treatment effects in studies of polynomial change with two levels of nesting (e.g., students at second level and schools at third level) where the treatment is at the third level (e.g., school intervention). In particular, we present methods for power analysis for cluster randomized designs (CRDs) where for instance schools are randomly assigned to a treatment and a control group, students are nested within schools, and repeated measurements are nested within students. This design assumes that third-level units (e.g., schools) are sampled randomly from a larger population at the first stage and then second-level units (e.g., students) within third-level units are randomly sampled. That is, both second- and third-level units are treated as random effects. Our models incorporate covariates at the second and third levels and, thus, also demonstrate the effects of covariates on power estimates. In addition, we provide a detailed description of the models in the three-level case with and without covariates.

The structure of this article is as follows. First, we present the unconditional model (i.e., no covariates at any level). Second, we expand the model to include covariates in the second and third levels. Third, we discuss how the number of measurement occasions, sample size (number of third- or second-level units), and covariates influence power through an illustrative example.

The Polynomial Change Model

A polynomial is an algebraic expression that contains more than one term and is described as a sum of terms of the same variable (e.g., time) in different powers (Kirk, 2012). For example, student achievement growth can be modeled well through a polynomial equation of the third degree as

$$Y = \beta_0 + \beta_1 a + \beta_2 a^2 + \beta_3 a^3 + \varepsilon, \quad (1)$$

where Y is student achievement, a is a measure of time such as age at each time of measurement, β_0 is a constant, $\beta_1 a$ is a linear component, $\beta_2 a^2$ is a quadratic component, $\beta_3 a^3$ is a cubic component, and ε is an error term. One disadvantage of Equation (1) is that the trend components are highly correlated, which can lead to multicollinearity. To resolve the dependency problem, one can use orthogonal polynomial contrast coefficients, which have been frequently used to fit trends of repeated measures. Equation (1) can then be reconstructed as

$$Y = \alpha_0 c_0 + \alpha_1 c_1 + \alpha_2 c_2 + \alpha_3 c_3 + u, \quad (2)$$

where u is an error term, and c_1 , c_2 , c_3 are orthogonal polynomial coefficients that are independent with each other, and thus enable researchers to independently test a null hypothesis for each of the three components (see Kirk, 2012). Orthogonal polynomial coefficients have been used to fit trends since the early 20th century (e.g., Fisher, 1928). Jennrich and Sampson (1971) provided an algorithm to generate the orthogonal polynomial contrast coefficients. Tables in many experimental design texts provide these contrast coefficients (see Kirk, 2012). Also, the R function `poly` computes orthogonal polynomials easily.

Previous studies had discussed sample size and statistical power considerations for group comparisons using repeated measures, most of this work, however, had been focused on single-level models (e.g., Bloch, 1986; Hedeker, Gibbons, & Waternaux, 1999). Raudenbush and Liu (2001) extended this work and provided power analysis and sample determination methods for repeated measures for CRDs with two levels. They focused on studies in which two groups of individuals (e.g., in two different treatment conditions) were followed over time to assess group differences in the average rate of change, rate of acceleration, or a higher degree polynomial. Through a two-level model combined with orthogonal polynomial contrasts at the first level, the authors examined how the number of measurement occasions, frequency of observation, and number of participants affected statistical power. They found that power increases as the number of measurement occasions or the number of students increase.

Power analysis methods for change models with two levels of nesting have rarely been discussed in prior literature. The optimal design manual by Spybrook et al. (2011) has provided power calculation formulae for three-level models in studies of polynomial change where the treatment is at the third level (e.g., schools), but has not incorporated the effects of covariates. Both Raudenbush and Liu (2001) and Spybrook et al. (2011) discussed unconditional models that did not include any covariates at any level.

However, prior studies have shown that covariates (e.g., students and school characteristics) could increase power significantly. Hedges and Hedberg (2007) documented that prior test scores and demographic covariates such as socioeconomic status account for nearly one third of the variance at the student level. Bloom, Richburg-Hayes, and Black (2007) found that controlling for baseline covariates could improve the precision of CRD studies that examine the impact of school interventions. Konstantopoulos (2012) showed that covariates at different levels of the hierarchy potentially explain a considerable proportion of the variance at the corresponding levels, and that centering of lower level covariates plays an important role in this (see also, Snijders & Bosker, 1999). In this study, we discuss models that incorporate covariates at the second or third levels.

Statistical Models

Unconditional Model

Consider a simple three-level growth design where Level 3 units (e.g., clusters such as schools) are randomly assigned to treatment or control conditions (i.e., clusters are nested within treatment). The first level for change over time of Level 2 unit i in cluster j can be expressed as a polynomial function, namely,

$$Y_{gij} = \alpha_{0ij}c_{0g} + \alpha_{1ij}c_{1g} + \alpha_{2ij}c_{2g} + \cdots + \alpha_{(P-1)ij}c_{(P-1)g} + u_{gij}, \quad (3)$$

where c_{pg} represent orthogonal polynomial contrasts of degree p ($p = 0, 1, \dots, P - 1$) at measurement occasion g ($g = 1, \dots, G$), α_{pij} s represent the mean and the rates of change (linear, quadratic, cubic, etc.), and u_{gij} is the within Level 2 unit random term with variance σ_e^2 . When $p = 0$, $c_{0g} = 1$, and α_{0ij} represents the average outcome for Level 2 unit i in Level 3 unit j . When $p = 1$, c_{1g} is a linear contrast and α_{1ij} is the linear rate of change for Level 2 unit i in Level 3 unit j , and so forth. We work with orthogonal polynomial contrasts because they facilitate the computations of estimators and their standard errors, and simplify power computations (see Raudenbush & Liu, 2001). The results apply to studies of any length and for polynomials of any degree (Kirk, 2012).

Orthogonal polynomial contrast coefficients should satisfy two conditions: (a) the sum of the p th polynomial contrasts is equal to zero and (b) the sum of the products of the p th and p' th polynomial contrasts is equal to zero (see Kirk, 2012):

$$\begin{aligned} \sum_{g=1}^G c_{pg} &= 0 \\ \sum_{g=1}^G c_{pg}c_{p'g} &= 0. \end{aligned} \quad (4)$$

Orthogonal polynomial coefficients that meet conditions shown in Equation (4) are not unique because, any group of orthogonal polynomial coefficients denoted as

$C_{pg} = k_p c_{pg}$, also meet these two conditions, where k_p could be any positive constant. With equally spaced time points, the following formulae could be used to calculate orthogonal polynomial coefficients:

$$\begin{aligned} c_{pg} &= k_p C_{pg} \\ C_{0g} &= 1 \\ C_{1g} &= g - \sum_{g=1}^G \frac{g}{G} \\ C_{p+1,g} &= C_{1g} C_{pg} - \frac{p^2(G^2 - p^2)}{4 \cdot (4p^2 - 1)} C_{p-1,g} \end{aligned} \quad (5)$$

(see Jennrich & Sampson, 1971), where c_{pg} is one possible orthogonal polynomial coefficient of degree p at measurement g as defined before, and k_p could be any positive constant. That is, researchers could choose any k_p to get their own orthogonal polynomial contrast coefficients.

For example, when $k_0 = 1$, $k_1 = 1$, and $k_2 = 1$, one can compute the first three orthogonal coefficients as

$$\begin{aligned} c_{0g} &= 1 \\ c_{1g} &= g - \sum_{g=1}^G \frac{g}{G} \\ c_{2g} &= \left(g - \sum_{g=1}^G \frac{g}{G} \right)^2 - \frac{G^2 - 1}{12}. \end{aligned} \quad (6)$$

When $G = 4$, then the values of the orthogonal coefficients are

$$\begin{aligned} c_0 &= (1, 1, 1, 1) \\ c_1 &= (-1.5, -0.5, 0.5, 1.5) \\ c_2 &= (1, -1, -1, 1). \end{aligned} \quad (7)$$

The least squares estimates for each of the change parameters and their variances can be computed as

$$\begin{aligned} \hat{\alpha}_{pji} &= \frac{\sum_{g=1}^G c_{pg} Y_{gij}}{\sum_{g=1}^G c_{pg}^2} \\ Var(\hat{\alpha}_{pji}) &= \frac{\sigma_e^2}{\sum_{g=1}^G c_{pg}^2} \end{aligned} \quad (8)$$

(see Seber & Lee, 2003), where

$$\sum_{g=1}^G c_{pg}^2 = k_p^2 \cdot \frac{(p!)^4}{(2p)! \cdot (2p+1)!} \cdot \frac{(G+p)!}{(G-p-1)!} \quad (9)$$

(see Plackett, 1960).¹

The growth model described above assumes equally spaced repeated measurements. For example, suppose students participate in a 4-year study (first through fourth grade) and a measure of student achievement such as mathematics is collected every April in each grade. Then, there are four occasions of measurements altogether or $G = 4$. In addition, we assume that the frequency of each measurement occasion is equal to 1 (i.e., one observation per measurement occasion).

In the second-level model, each of the parameters α_{pij} (e.g., the average polynomial change for each individual) from the first-level equation varies between Level 2 units (e.g., individuals) within Level 3 units (e.g., schools), namely,

$$\alpha_{pij} = \beta_{p0j} + \xi_{pij}, \quad (10)$$

where β_{p0j} s represent the average polynomial effects within Level 3 units such as schools and the ξ_{pij} s are individual specific random effects between Level 2 units within Level 3 units for each polynomial change parameter. These random effects follow a multivariate normal distribution with zero means, variances τ_{pp}^2 , and covariances $\tau_{pp'}$, between the random effects ξ_{pij} and $\xi_{p'ij}$.

At the third level each of the parameters, β_{p0j} s (average polynomial change for each Level 3 unit) vary across third level units such as schools, namely

$$\beta_{p0j} = \gamma_{p00} + \gamma_{p01} T_j + \eta_{p0j}, \quad (11)$$

where T is a binary indicator for a treatment, γ_{p00} s represent the average polynomial effects across Level 3 units, γ_{p01} s represent the average difference between the treatment and the control group for each polynomial change parameter, and the η_{p0j} s are Level 3 unit specific random effects for each polynomial change parameter. These random effects follow a multivariate normal distribution with zero means, variances ω_{pp}^2 , and covariances $\omega_{pp'}$, between the random effects η_{p0j} and $\eta_{p'0j}$.

Suppose there are N Level 2 units within each Level 3 unit and m Level 3 units within each treatment condition, which means that the total number of Level 3 units is $M = 2m$ and, thus, the total number of Level 2 units is MN . Then, the estimate of the variance of the treatment effect for polynomial p is

$$Var(\gamma_{p01}) = \frac{2}{mN} (N\omega_{pp}^2 + \tau_{pp}^2 + \sigma_p^2), \quad \sigma_p^2 = \frac{\sigma_e^2}{\sum_{g=1}^G c_{pg}^2} \quad (12)$$

and $\sum_{g=1}^G c_{pg}^2$ is defined in Equation (9) (see Konstantopoulos, 2008; Raudenbush & Liu, 2001; Spybrook et al., 2011).²

Covariates at Second and Third Levels

Adding Level 2 covariates in the model changes the second-level equation to

$$\alpha_{pij} = \beta_{p0j} + \mathbf{X}_{ij}\mathbf{B}_{p2j} + \xi_{Apij} \quad (13)$$

where $\mathbf{X}_{ij}$ is a row vector of k Level 2 unit characteristics, and $\mathbf{B}_{p2j}$ is a column vector of k coefficients of Level 2 unit characteristics. The ξ_{Apij} s are Level 2 random effects within Level 3 units for each polynomial change parameter, and subscript A indicates an adjustment in the error term (i.e., variance reduction) because of covariates. The random effects follow a multivariate normal distribution with zero means, variances τ_{Rpp}^2 , covariances $\tau_{Rpp'}$ between random effects ξ_{pij} and ξ_p , and subscript R indicates residual variance because of covariates. All other terms have been defined previously.

Adding Level 3 covariates in the model changes the third-level equation to

$$\beta_{p0j} = \gamma_{Ap00} + \gamma_{Ap01}T_j + \mathbf{Z}_{P02}\mathbf{\Gamma}_j + \eta_{Ap0j}, \quad (14)$$

where $\mathbf{Z}_{P02}$ is a row vector of q Level 3 unit characteristics, and $\mathbf{\Gamma}_j$ is a column vector of coefficients of level-3 unit characteristics. The η_{Ap0j} s are Level 3 specific random effects for each polynomial change parameter, where subscript A indicates adjustment because of covariates (see Konstantopoulos, 2008). These random effects follow a multivariate normal distribution with zero means, variances ω_{Rpp}^2 , covariance $\omega_{Rpp'}$ between random effects η_{p0j} and $\eta_{p'0j}$, and subscript R indicates residual variance because of covariates. All other terms have been defined previously.

We define the proportions of the unexplained variances at the second and third levels as

$$w_2 = \tau_{Rpp}^2 / \tau_{pp}^2, \quad w_3 = \omega_{Rpp}^2 / \omega_{pp}^2, \quad (15)$$

that is, w_2 indicates the proportion of the variance at the second level that is still unexplained; while w_3 indicates the proportion of the variance at the third level that is still unexplained. For example, when $w_3 = 0.80$, it indicates that the variance at the third level decreased by 20% because of the inclusion of covariates at the third level (i.e., the covariates at the third level explain 20% of the variance at that level). Similarly, when $w_2 = 0.80$, it indicates that the variance at the second level decreased by 20% because of the inclusion of covariates at the second level assuming a centering approach where covariates can explain variance in the outcome only at their corresponding levels.

Again, suppose there are N Level 2 units within each Level 3 unit and m Level 3 units within the treatment or control condition, and thus the total number of Level 2 units is mN . Then, the estimate of the variance of the treatment effect for polynomial p is

$$Var(\gamma_{Ap01}) = \frac{2}{mN} (Nw_3\omega_{pp}^2 + w_2\tau_{pp}^2 + \sigma_p^2), \quad \sigma_p^2 = \frac{\sigma_e^2}{\sum_{g=1}^G c_{pg}^2}. \quad (16)$$

Power Analysis

Suppose that a researcher wants to test the hypothesis that γ_{p01} or γ_{Ap01} are different from zero and carries out the usual t test. When covariates are included at the second and third levels the test statistic is defined as

$$t = \hat{\gamma}_{Ap01} / \sqrt{Var(\hat{\gamma}_{Ap01})}. \quad (17)$$

When the null hypothesis is false, the test statistic t follows a noncentral t -distribution with $2m - q - 2$ degrees of freedom and non-centrality parameter λ_A . The non-centrality parameter is defined as the expected value of the estimate of the treatment effect divided by the square root of the variance of the estimate of the treatment effect, namely

$$\lambda_A = \gamma_{Ap01} \sqrt{\frac{mN}{2}} \sqrt{\frac{1}{Nw_3\omega_{pp}^2 + w_2\tau_{pp}^2 + \sigma_p^2}}, \quad (18)$$

where subscript A indicates adjustment because of covariates. The power of a two-tailed t test for a specified significance level α is defined as

$$p_1 = 1 - H[c(\alpha/2, 2m - q - 2), (2m - q - 2), \lambda_A] + H[-c(\alpha/2, 2m - q - 2), (2m - q - 2), \lambda_A], \quad (19)$$

where $c(\alpha, \nu)$ is the Level α one-tailed critical value of the t distribution with ν degrees of freedom (e.g., $c(0.05, 20) = 1.72$), $H(x, \nu, \lambda_A)$ is the cumulative distribution function of the noncentral t distribution with ν degrees of freedom and non-centrality parameter λ_A (or λ), and q is the number of the third level covariates. Alternatively, one can use an F test with 1, $2m - q - 2$ degrees of freedom (or $2m - 2$ degrees of freedom in the unconditional case) and a noncentrality parameter λ_A^2 or λ^2 .

In an unconditional model (i.e., no covariates at any level) the noncentrality parameter is

$$\lambda = \gamma_{p01} \sqrt{\frac{mN}{2}} \sqrt{\frac{1}{N\omega_{pp}^2 + \tau_{pp}^2 + \sigma_p^2}}, \quad (20)$$

and the power of the t test in the unconditional case is computed as

$$p_2 = 1 - H[c(\alpha/2, 2m - 2), (2m - 2), \lambda] + H[-c(\alpha/2, 2m - 2), (2m - 2), \lambda]. \quad (21)$$

Effect Size Considerations

The power computations are facilitated by defining a standardized effect size instead of a mean difference. Prior literature has provided three definitions of effect sizes for three-level models (e.g., Hedges, 2011; Konstantopoulos, 2008). The first option of defining an effect size for a polynomial degree p in three-level models is to standardize (i.e., divide) the group difference by the square root of the total variance (Hedges, 2011; Konstantopoulos, 2008)

$$ES_1 = \frac{\gamma_{p01}}{\sqrt{\omega_{pp}^2 + \tau_{pp}^2 + \sigma_p^2}}. \quad (22)$$

However, the denominator of ES_1 depends on σ_p^2 , which is a function of the number of measurement occasions as shown in Equation (8). In other words, ES_1 changes as the number of measurement occasions varies. Because this study evaluates various designs with a different number of measurement occasions, but with a fixed effect size, ES_1 is not appropriate.

The remaining two ways of defining an effect size are

$$ES_2 = \frac{\gamma_{p01}}{\omega_{pp}}, \quad (23)$$

or

$$ES_3 = \frac{\gamma_{p01}}{\sqrt{\omega_{pp}^2 + \tau_{pp}^2}}. \quad (24)$$

In ES_3 the group difference (i.e., numerator of Equation 24) is standardized by the square root of the sum of the Level 2 and Level 3 variances; whereas in ES_2 the group difference is divided by the square root of the Level 3 variance. Both ES_2 and ES_3 can be used as effect sizes in three-level models (see Hedges, 2011).

When ES_2 is used, the noncentrality parameter λ in an unconditional model becomes

$$\lambda = \frac{ES_2}{\sqrt{N}} \sqrt{\frac{mN}{2}} \sqrt{r}, \quad (25)$$

where $r = \frac{N\omega_{pp}^2}{(N\omega_{pp}^2 + \tau_{pp}^2 + \sigma_p^2)}$ is the reliability of third-level units (e.g., schools; see Raudenbush & Bryk, 2002).

Now, when ES_3 is used, the noncentrality parameter λ in an unconditional model becomes

$$\lambda = ES_3 \sqrt{\frac{mN}{2}} \sqrt{\frac{\omega_{pp}^2 + \tau_{pp}^2}{N\omega_{pp}^2 + \tau_{pp}^2 + \sigma_p^2}}. \quad (26)$$

The underlying assumption is that the treatment coefficient does not change after adding covariates at the second or third levels ($\gamma_{Ap01} = \gamma_{p01}$), which is reasonable because in experimental designs the treatment (T_j) should be independent of any covariates, observed or unobserved.

When covariates are included at the second and third levels the noncentrality parameter λ_A simplifies to

$$\lambda_A = \frac{ES_2}{\sqrt{N}} \sqrt{\frac{mN}{2}} \sqrt{r_A}, \quad r_A = \frac{N\omega_{pp}^2}{(Nw_3\omega_{pp}^2 + w_2\tau_{pp}^2 + \sigma_p^2)} \quad (27)$$

or

$$\lambda_A = ES_3 \sqrt{\frac{mN}{2}} \sqrt{\frac{\omega_{pp}^2 + \tau_{pp}^2}{Nw_3\omega_{pp}^2 + w_2\tau_{pp}^2 + \sigma_p^2}}. \quad (28)$$

Illustrative Example

Linear Growth

To illustrate the applicability of the methods to assess the impact of the number of measurement occasions, sample sizes (at the second or third levels), and covariates on power, we use an example from education. Specifically, we use data from a large-scale experiment conducted in the state of Indiana during the 2009-2010 school year. This experiment used a CRD, where students were nested within schools, and schools were nested within treatment (an assessment program) and control groups. Random assignment took place at the school level, that is, schools were randomly assigned to treatment and control conditions. Schools in the treatment group adopted interim assessment tools to measure student learning, while control schools did not use any specific assessment program (business as usual). The study incorporated a longitudinal component and thus mathematics and reading achievement scores were measured three times overall in the spring of 2010, 2011, and 2012 (see Konstantopoulos, Miller, & van der Ploeg, 2013, for a more detailed discussion of the experiment). The total number of participating schools was 50 with 32 schools in the treatment group and nearly 20,000 students participated in the study during the 2009-2010 school year.

In our example, the outcome is standardized mathematics achievement. Because the number of measurement occasions was only three, we used a linear rate of change model at Level 1 (repeated measures), namely

$$Math_{gij} = \alpha_{0ij}c_{0g} + \alpha_{1ij}c_{1g} + u_{gij}, u_{gij} \sim N(0, \sigma_e^2)$$

where $Math_{gij}$ is mathematics achievement in year g , $c_{0g} = (1, 1, 1)$ and $c_{1g} = (-1, 0, 1)$ at $g = 1, 2, 3$ according to the orthogonal polynomials in Equation (6). This model defines α_{0ij} as the mean mathematics achievement for student i in school j when $g = 2$, and α_{1ij} is the average rate of linear change of mathematics achievement for student i in school j .

The second-level model (e.g., student level) is

$$\alpha_{0ij} = \beta_{00j} + \xi_{0ij}, \xi_{0ij} \sim N(0, \tau_{00}^2)$$

$$\alpha_{1ij} = \beta_{10j} + \xi_{1ij}, \xi_{1ij} \sim N(0, \tau_{11}^2),$$

where β_{00j} is the mean mathematics achievement in school j when $g = 2$, and β_{10j} is the average growth rate in school j .

The third-level model (e.g., school level) is

$$\beta_{00j} = \gamma_{000} + \gamma_{001}T_j + \eta_{00j}, \eta_{00j} \sim N(0, \omega_{00}^2)$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}T_j + \eta_{10j}, \eta_{10j} \sim N(0, \omega_{11}^2)$$

where γ_{000} is the mean mathematics achievement across students and schools when $g = 2$, γ_{001} is the main average treatment effect when $g = 2$, T_j is a binary indicator coded as one for treatment schools and zero for control schools, γ_{100} is the average rate of change, and γ_{101} is the main effect of treatment for the linear rate of change, which is the primary coefficient of interest. The estimates of the relevant variances are

$$\sigma_e^2 = 0.00092, \quad \tau_{11}^2 = 0.00091, \quad \omega_{11}^2 = 0.00012.$$

To demonstrate the computation of power, suppose that $ES_2 = 0.80$ or $ES_3 = 0.30$ standard deviations and the significance level is .05. Because our methodology works for balanced designs suppose $m = 20$ and $N = 30$, which indicates 20 schools in the treatment group (40 schools in total) and 30 students in each school (treatment or control).

According to Equation (8) and Equation (9) with $G = 3$, $p = 1$ and $k_I = 1$, first we calculate σ_1^2 as

$$\sigma_1^2 = \frac{12 \cdot 0.00092}{4 \cdot 3 \cdot 2} = 0.00046.$$

The noncentrality parameter λ of the t test for ES_2 (Equation 25) is

$$\begin{aligned}\lambda &= ES_2 \sqrt{\frac{mN}{2}} \sqrt{\frac{\omega_{pp}^2}{N\omega_{pp}^2 + \tau_{pp}^2 + \sigma_p^2}} \\ &= 0.80 \cdot \sqrt{\frac{20 \cdot 30}{2}} \cdot \sqrt{\frac{0.00012}{30 \cdot 0.00012 + 0.00091 + 0.00046}} = 2.153,\end{aligned}$$

and for ES_3 (Equation 26) is

$$\begin{aligned}\lambda &= ES_3 \sqrt{\frac{mN}{2}} \sqrt{\frac{\omega_{pp}^2 + \tau_{pp}^2}{N\omega_{pp}^2 + \tau_{pp}^2 + \sigma_p^2}} \\ &= 0.30 \cdot \sqrt{\frac{20 \cdot 30}{2}} \cdot \sqrt{\frac{0.00012 + 0.00091}{30 \cdot 0.00012 + 0.00091 + 0.00046}} = 2.366.\end{aligned}$$

The critical value of the test using the t distribution with $(2 \times 20) - 2 = 38$ degrees of freedom is $c(0.05, 38) = 2.024$. To compute power we use Equation (21), namely,

$$p_2 = 1 - H[2.024, 38, 2.153] + H[-2.024, 38, 2.153] = 0.555,$$

or

$$p_3 = 1 - H[2.024, 38, 2.366] + H[-2.024, 38, 2.366] = 0.635.$$

Tables 1 through 3 and Figures 1 through 3 show how variations in the number of measurement occasions and sample sizes at the second and third levels affect the power of the t test for the treatment effect of the linear rate of change in CRD. The t tests are two-tailed, the significance level is at .05, and the effect size is either 0.80 (ES_2) or 0.30 (ES_3) standard deviations. Table 1 and Figure 1 provide power estimates for designs that vary the number of measurement occasions and the number of schools (M), holding the number of students (N) in each school constant at 30. The estimates of the left panels correspond to $ES_2 = 0.80$ standard deviations and the estimates of the right panels correspond to $ES_3 = 0.30$ standard deviations. As the number of measurement occasions or the number of schools increase, power increases. When $ES_3 = 0.30$, the number of measurement occasions is 3, and the number of schools is 60, the power is 0.813, which is greater than the lower bound of power (i.e., 0.80). Note that, power only changes marginally as the number of measurement occasions increases from 3 to 6 for a fixed number of schools; whereas the number of schools has a bigger effect on power for a fixed number of measurement occasions. For example, assuming $ES_2 = 0.80$, when the number of measurement occasions is doubled from 3 to 6, the power changes only slightly; whereas when the number of schools increases from 40 to 70, power increases more than one third.

Figure 2 and Table 2 provide power estimates for designs that vary the number of measurement occasions and the number of students (N) in each school, holding the number of schools (M) constant at 40. As the number of measurement occasions or the number of students in each school increases, power increases as well. However,

Table 1. Effects of Number of Measurement Occasions (D) and Number of Schools (M) on Power: Linear Rate of Change.

D	$ES_2 = 0.80$				$ES_3 = 0.30$			
	M				M			
	40	50	60	70	40	50	60	70
3	0.555	0.655	0.737	0.802	0.635	0.736	0.813	0.870
4	0.579	0.680	0.761	0.824	0.660	0.760	0.834	0.888
5	0.587	0.688	0.769	0.831	0.669	0.768	0.841	0.894
6	0.591	0.692	0.772	0.834	0.672	0.772	0.845	0.896

Note. The significance level is .05. The number of students per school is fixed at 30.

Figure 1. Effect of the number of measurement occasions and number of schools (M) on power, holding number of students (N) in each school fixed at 30.

Note. The significance level is .05.

as the number of students in each school increases (e.g., from 60 to 70) power does not increase that much (from 0.71 to 0.72), which indicates that increasing the number of students per school is not the most effective way of boosting power. In contrast, increasing the number of schools is important.

Figure 3 and Table 3 provide power estimates for designs that vary the number of students (N) in each school and the number of schools (M), holding the number of measurement occasions constant at 3. As the number of students per school or the number of schools increases, power increases. However, the number of schools affects power more than the number of students holding the number of measurement

Table 2. Effects of Number of Measurement Occasions (*D*) and Number of Students (*N*) on Power: Linear Rate of Change.

<i>D</i>	<i>ES</i> ₂ = 0.80				<i>ES</i> ₃ = 0.30			
	<i>N</i>				<i>N</i>			
	30	40	50	60	30	40	50	60
3	0.555	0.585	0.604	0.618	0.635	0.666	0.686	0.699
4	0.579	0.605	0.621	0.632	0.660	0.686	0.702	0.714
5	0.587	0.611	0.626	0.637	0.669	0.693	0.708	0.718
6	0.591	0.614	0.629	0.639	0.672	0.696	0.710	0.720

Note. The significance level is .05. The number of schools is fixed at 40.

Figure 2. Effect of the number of measurement occasions (*D*) and number of students (*N*) on power holding number of schools (*M*) fixed at 40.

Note. The significance level is .05.

occasions fixed. Power reaches to 0.80 with various combinations of the number of schools and the number of students in each school (e.g., $M = 70$ and $N = 30$ when $ES_2 = 0.80$; $M = 60$ and $N = 30$ when $ES_3 = 0.30$).

Covariates at different levels also influence power under the assumption that they explain a certain proportion of the second- or third-level variances. Figure 4 and Table 4 show how power changes as the proportion of the explained variance at the second and third levels changes for a design with $M = 40$ (or $m = 20$), $N = 30$, $D = 3$, and $ES_2 = 0.80$ or $ES_3 = 0.30$. The estimates of the left panels correspond to $ES_2 =$

Table 3. Effects of Number of Schools (M) and Number of Students (N) on Power: Linear Rate of Change.

M	$ES_2 = 0.80$				$ES_3 = 0.30$			
	N				N			
	30	40	50	60	30	40	50	60
40	0.555	0.585	0.604	0.618	0.635	0.666	0.686	0.699
50	0.655	0.686	0.706	0.719	0.736	0.766	0.784	0.797
60	0.737	0.767	0.785	0.797	0.813	0.839	0.855	0.866
70	0.802	0.829	0.845	0.856	0.870	0.892	0.905	0.913

Note. The significance level is .05. The number of measurement occasions (D) is fixed at 3.

Figure 3. Effects of number of schools (M) and number of students (N) on power holding the number of measurement occasions (D) fixed at 3.

Note. The significance level is .05.

0.80 and the estimates of the right panels correspond to $ES_3 = 0.30$. The degrees of freedom decrease when we add covariates at the third level. Suppose two covariates (e.g., prior school achievement and school socioeconomic status) are added at the third level ($q = 2$); the degrees of freedom are then reduced to $(2 \times 20) - 2 - 2 = 36$. As the explained variance increases because of covariates, power increases. For example, when $ES_2 = 0.80$, and 30% of the variances at the second or third levels are explained (i.e., $w_3 = w_2 = 0.7$), the power is 0.690, which is larger than the power without covariates (0.555). In addition, in our example, covariates at the third level affect power significantly more than covariates at the second level. For example, when $ES_2 = 0.80$, and the variance explained at the second level increases from 30% to 60%, while the variance explained at the third level stays

Table 4. Effects of Covariates on Power: Linear Rate of Change.

Percent of variance explained: Level 3	$ES_2 = 0.80$				$ES_3 = 0.30$			
	Percent of variance explained: Level 2							
	30	40	50	60	30	40	50	60
30	0.690	0.701	0.712	0.724	0.770	0.780	0.791	0.801
40	0.735	0.747	0.759	0.772	0.811	0.822	0.833	0.844
50	0.784	0.796	0.809	0.822	0.854	0.865	0.876	0.886
60	0.835	0.848	0.861	0.874	0.896	0.907	0.917	0.926

Note. The significance level is .05. The number of measurement occasions (D) is fixed at 3, 40 schools, 30 students per school.

Figure 4. Effects of covariates on power.

Note. The number of measurement occasions (D) is 3 with 40 schools and 30 students in each school; two covariates included at the third level; the significance level is .05.

fixed at 30%, power increases only slightly from 0.690 to 0.724. However, when $ES_2 = 0.80$, and the variance explained at the third level increases from 30% to 60%, while the variance explained at the second level stays fixed at 30%, power increases significantly from 0.690 to 0.835.

Finally, we compute power estimates for designs that vary the number of students (N) in each school and the number of schools (M), assuming 40% of the second- or third-level variances is explained (i.e., $w_2 = w_3 = 0.60$), holding the number of measurement occasions constant at 3. These estimates are presented in Figure 5 and Table 5. The estimates of the left panels correspond to $ES_2 = 0.80$ and the estimates

Table 5. Effects of Number of Schools (M) and Number of Students (N) on Power: Covariates Added at Levels 2 and 3; Linear Rate of Change.

M	$ES_2 = 0.80$				$ES_3 = 0.30$			
	N				N			
	30	40	50	60	30	40	50	60
40	0.747	0.781	0.802	0.817	0.822	0.852	0.870	0.882
50	0.840	0.868	0.885	0.896	0.900	0.922	0.935	0.943
60	0.901	0.923	0.935	0.943	0.946	0.961	0.968	0.973
70	0.940	0.956	0.964	0.970	0.971	0.981	0.985	0.988

Note. The significance level is .05; the number of measurement occasions (D) is fixed at 3. Forty percent of the variance is explained at Levels 2 or 3.

Figure 5. Effects of covariates, number of schools (M), and number of students (N) in each school on power holding the number of measurement occasions (D) fixed at 3.

Note. Two covariates included at Level 3; forty percent of the variance is explained at Levels 2 or 3; the significance level is .05.

of the right panels correspond to $ES_3 = 0.30$. Compared with estimates in Table 3, the power estimates in Table 5 are always larger, which shows that when covariates explain a considerable percentage (e.g., 40%) of the variances at the second or the third levels, power increases significantly (other things being equal).

Quadratic Model

We also illustrate the applicability of the methodology in the case of a quadratic rate of change. To achieve this we used data from Project STAR (Student–Teacher

Achievement Ratio), an experiment conducted in Tennessee in the mid-1980s (see Nye, Hedges, & Konstantopoulos, 2000). In this experiment, a cohort of kindergarten students was followed for 4 years (through third grade). Our data included students who had all four measurements in outcomes of interest (e.g., mathematics). Within a grade, students were randomly assigned to smaller or larger size classes. The nesting structure was repeated measures that were nested within students who were nested within classrooms, and the treatment was at the classroom level. For the purposes of the computation, schools were not taken into account.

The outcome is standardized student mathematics achievement, which was measured four times at the end of kindergarten, first grade, second grade, and third grade. A model with a quadratic rate of change at Level 1 (repeated measures) was used, namely,

$$Math_{gij} = \alpha_{0ij}c_{0g} + \alpha_{1ij}c_{1g} + \alpha_{2ij}c_{2g} + u_{gij}, u_{gij} \sim N(0, \sigma_e^2),$$

where $Math_{gij}$ is student mathematics achievement in year g , $c_{0g} = (1, 1, 1, 1)$, $c_{1g} = (-1.5, -0.5, 0.5, 1.5)$, and $c_{2g} = (1, -1, -1, 1)$ at $g = 1, 2, 3, 4$ following Equation (6). This model defines α_{2ij} as the average quadratic rate of change of mathematics achievement for student i in school j . All the other terms have been defined previously.

The second-level model (student level) is

$$\alpha_{0ij} = \beta_{00j} + \xi_{0ij}, \xi_{0ij} \sim N(0, \tau_{00}^2)$$

$$\alpha_{1ij} = \beta_{10j} + \xi_{1ij}, \xi_{1ij} \sim N(0, \tau_{11}^2)$$

$$\alpha_{2ij} = \beta_{20j} + \xi_{2ij}, \xi_{2ij} \sim N(0, \tau_{22}^2),$$

where β_{20j} is the average quadratic growth rate in school j . All the other terms have been defined previously.

The third-level model (school level) is then

$$\beta_{00j} = \gamma_{000} + \gamma_{001}T_j + \eta_{00j}, \eta_{00j} \sim N(0, \omega_{00}^2)$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}T_j + \eta_{10j}, \eta_{10j} \sim N(0, \omega_{11}^2)$$

$$\beta_{20j} = \gamma_{200} + \gamma_{201}T_j + \eta_{20j}, \eta_{20j} \sim N(0, \omega_{22}^2)$$

where γ_{200} is the average quadratic rate of change, and γ_{210} is the main effect of treatment for the quadratic rate of change, which is of primary interest. The variance estimates we computed are $\sigma_e^2 = 0.24138$, $\tau_{22}^2 = 0.00117$, and $\omega_{22}^2 = 0.03013$.

For the purposes of the power computations suppose $ES_2 = ES_3 = 0.80$ standard deviations and the significance level is set at .05. Suppose $m = 20$, and $N = 30$. According to Equations (7) and (8) with $G = 4$, $p = 2$, and $k_2 = 1$, first we compute

$$\sigma_2^2 = \frac{0.24138}{4} = 0.060345.$$

Then, the noncentrality parameter λ of the t test for ES_2 (Equation 25) or ES_3 (Equation 26) is

$$\begin{aligned}\lambda &= ES_2 \sqrt{\frac{mN}{2}} \sqrt{\frac{\omega_{pp}^2}{N\omega_{pp}^2 + \tau_{pp}^2 + \sigma_p^2}} \\ &= 0.80 \cdot \sqrt{\frac{20 \cdot 30}{2}} \cdot \sqrt{\frac{0.03013}{30 \cdot 0.03013 + 0.00117 + 0.060345}} = 2.448\end{aligned}$$

or

$$\begin{aligned}\lambda &= ES_3 \sqrt{\frac{mN}{2}} \sqrt{\frac{\omega_{pp}^2 + \tau_{pp}^2}{N\omega_{pp}^2 + \tau_{pp}^2 + \sigma_p^2}} \\ &= 0.80 \cdot \sqrt{\frac{20 \cdot 30}{2}} \cdot \sqrt{\frac{0.03013 + 0.00117}{30 \cdot 0.03013 + 0.00117 + 0.060345}} = 2.495.\end{aligned}$$

The critical value of the test using the t distribution with $(2 \times 20) - 2 = 38$ degrees of freedom is $c(0.05, 39) = 2.024$. Finally, we computed power as

$$p = 1 - H[2.024, 38, 2.448] + H[-2.024, 38, 2.448] = 0.665$$

and then as

$$p = 1 - H[2.024, 38, 2.495] + H[-2.024, 38, 2.495] = 0.681.$$

Figure 6 and Table 6 provide power estimates for quadratic rate of change in designs that vary the number of students (N) in each school and the number of schools (M), holding the number of measurement occasions constant at 4. Overall,

Table 6. Effects of Number of Schools (M) and Number of Students (N) per School on Power: Quadratic Rate of Change.

M	$ES_2 = 0.80$				$ES_3 = 0.80$			
	N				N			
	30	40	50	60	30	40	50	60
40	0.549	0.580	0.600	0.614	0.565	0.596	0.616	0.631
50	0.648	0.681	0.701	0.715	0.665	0.697	0.718	0.732
60	0.730	0.761	0.781	0.794	0.746	0.777	0.796	0.809
70	0.796	0.824	0.842	0.853	0.811	0.838	0.855	0.866

Note. The significance level is .05; the number of measurement occasions (D) is fixed at 4.

Figure 6. Effects of number of schools (M) and number of students (N) on power holding the number of measurement occasions (D) fixed at 4: A quadratic rate of change. Note. The significance level is .05.

the results are similar to those reported in Figure 3 and Table 3 for the linear rate of change. As the number of students per school or the number of schools increases, power increases. However, the number of schools affects power more than the number of students holding the number of measurement occasions fixed.

Conclusion

Multilevel experimental designs are becoming increasingly common in education. Frequently, these designs assign randomly entire clusters such as schools to a treatment or a control group and follow them over time. In such designs, researchers face the challenge of choosing the number of measurement occasions and sample sizes at the second and third levels to ensure that treatment effects will be detected with a high enough probability (e.g., $> .80$). The present study extended previous work on power analysis for two-level CRDs in studies of polynomial change and presented methods to compute power for three-level CRDs that incorporate polynomial change in the first level.

The power of the test of the treatment effect in studies of polynomial change with two levels of nesting is a function of the magnitude of the treatment effect, the number of measurement occasions, the sample size of individuals, the sample size of clusters, and the proportion of the variances that covariates at the second or third levels explain.

Two main findings emerged from this study. First, power increases as the number of measurement occasions, the number of students in each school, and the number of

schools increase. Other things being equal, the number of Level 3 units (clusters) influences power more than the number of Level 2 units (individuals) or the number of measurement occasions of the study. In particular, the number of students and the number of measurement occasions of the study have a rather limited influence on power. This indicates clearly that researchers should sample more schools rather than students within schools to maximize power. The number of schools affects power also through the degrees of freedom of the t test and thus a smaller number of covariates should be included at the third level. In addition, the higher the order of the polynomials a growth model includes, the longer the number of measurement occasions needed. For example, to fit a linear rate of change model, a minimum of three measurements is needed; to fit a quadratic rate of change model, a minimum of four measurements is needed. This finding holds for linear rate of change models and for higher order polynomials (e.g., quadratic) which shows the robustness of our results.

Second, covariates that explain variance at the second or third levels boost power estimates considerably and can help reduce the number of measurement occasions or sample sizes needed at the second or third levels. For instance, when $ES_3 = 0.30$ and covariates explain 40% of the variance at the second or third levels, the required number of schools to achieve power greater than 0.80 drops from 60 to 40 holding the number of students in each school constant at 30 and the number of measurement occasions fixed at three. Researchers should be aware that there is a trade-off in power when third-level covariates are included in the model. Specifically, adding third-level covariates will reduce the degrees of freedom of the test and therefore power may be reduced. At the same time the covariates will explain additional variance at the third level and therefore power may be increased.

One potential limitation of the study is that the methods presented assume equally spaced occasions of measurements. Thus, the methods reported here do not hold exactly for unequally spaced repeated measurements. In addition, our methods hold only for longitudinal designs where one observation is collected in each measurement occasion.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Our Equation (9) is based on equation (8) on page 103 in Plackett (1960).
2. Our equation is similar to equation (11.10) in Spybrook et al. (2011). However, the notation in Spybrook et al. (2011) is different. In addition, equation (11.10) in Spybrook et al.

(2011) differs in that it also includes the frequency of observation at each time point, whereas we do not.

References

- Bloch, D. A. (1986). Sample size requirements and the cost of a randomized clinical trial with repeated measurements. *Statistics in Medicine*, 5, 663-667.
- Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, 29, 30-59.
- Boruch, R. F., & Gomez, H. (1977). Sensitivity, bias, and theory in impact evaluations. *Professional Psychology*, 8, 411-434.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge.
- Fisher, R. A. (1928). *Statistical methods for research workers* (2nd ed.). Edinburgh, Scotland: Oliver & Boyd.
- Hedeker, D., Gibbons, R. D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: Comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics*, 24, 70-79.
- Hedges, L. V. (2011). Effect sizes in three level designs. *Journal of Educational and Behavioral Statistics*, 36, 346-380.
- Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlations for planning group-randomized experiments in education. *Educational Evaluation and Policy Analysis*, 29, 60-87.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27, 236.
- Jennrich, R. I., & Sampson, P. I. (1971). Remark as R3: A remark on algorithm AS 10. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 20, 117-118.
- Jong, K. D., Moerbeek, M., & van der Leeden, R. (2010). A priori power analysis in longitudinal three-level multilevel models: An example with therapist effects. *Psychotherapy Research*, 20, 273-284.
- Kirk, R. E. (2012). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Thousand Oaks, CA: Sage.
- Konstantopoulos, S. (2008). The power of the test for treatment effects in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness*, 1, 66-88.
- Konstantopoulos, S. (2012). The impact of covariates on statistical power in cluster randomized designs: Which level matters more? *Multivariate Behavioral Research*, 47, 392-420.
- Konstantopoulos, S., Miller, S., & van der Ploeg, A. (2013). The impact of Indiana's system of interim assessments on mathematics and reading achievement. *Educational Evaluation and Policy Analysis*, 35, 481-499.
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research* (Vol. 19). Thousand Oaks, CA: Sage.
- Marcoulides, G. A., & Schumacker, R. E. (2012). *New developments and techniques in structural equation modeling*. New York, NY: Taylor & Francis.
- Moerbeek, M., & Teerenstra, S. (2016). *Power analysis of trials with multilevel data*. Boca Raton, FL: CRC Press.

- Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). Effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal*, 37, 123-151.
- Plackett, R. L. (1960). *Principles of regression analysis*. Oxford, England: Clarendon Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S. W., & Liu, X. F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6, 387-401.
- Seber, G. A. F., & Lee, A. J. (2003). *Linear regression analysis* (2nd ed.). Hoboken, NJ: Wiley-Interscience.
- Snijders, T., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Spybrook, J., Raudenbush, S. W., Liu, X. F., Congdon, R., & Martinez, A. (2011). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" Software Version 3.0*. Retrieved from <http://www.wtgrantfoundation.org>