**Paper 4**

## Abstract Title Page

**Title:** Power Consideration for Three-Level Growth Models

**Authors and Affiliations:**
Wei Li
University of Missouri

Spyros Konstantopoulos
Michigan State University

## Abstract Body

**Background / Context:**

In recent years, there has been an increased interest in assessing the effects of educational interventions via experimental designs where students, classrooms, or schools are randomly assigned to a treatment and a control condition. An important part of the design phase of an experiment involves power analysis. Statistical power is 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 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 the student, classroom, or school level to ensure a high probability (e.g., > 80 percent) of detecting a treatment effect if it were true (Lipsey 1990).

Education experiments incorporate often times a longitudinal component where students are followed over time and are measured several times. In repeated measures experiments each student has their own trajectory which is a function of time and indicates the rate of change over time (Raudenbush & Bryk, 2002). The change in measurements over time does not always follow a linear trend. Instead, trajectories sometimes point to nonlinearities such as curvilinear trends. 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.

Studies about polynomial change may be viewed as having a nested structure. For example, 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. Raudenbush and Liu (2001) provided power analysis and sample determination methods for repeated measures in two-level models, where repeated measures are nested in individuals. They showed that power is a function of the magnitude of the treatment effect, the sample size of individuals, the study duration, and the frequency of measurements over time. Researchers should take into account all of these parameters in the design phase of the experiment to ensure that treatment effects will be detected.

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 either schools or students randomly to treatment or control groups. For instance, students are assigned to small or regular classes within schools. Or schools are randomly assigned to an assessment program or not. It seems natural to extend methods for power analysis for tests of treatment effects in studies of polynomial change from two to three-levels. Spybrook et al. (2011) reported in the optimal deign manual formulae to calculate power for three-level polynomial change models without covariates, where treatment is at the third level.

**Purpose / Objective / Research Question / Focus of Study:**

The purpose of this study is extend previous methods by Raudenbush and Liu (2001) and 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 and schools) where the treatment is either at the third level (e.g., school intervention) or at the second level (e.g., student intervention). In particular, we provide methods for power analysis for block randomized designs

(BRD) where the treatment is at the second level (e.g., student intervention) and the third level units (e.g., schools) serve as blocks. Please note that methods for power analysis for cluster randomized designs (CRD), where for instance schools are randomly assigned in a treatment and a control group, is omitted in this proposal because of words limits.

**Significance / Novelty of study:**

This study fills the gap in literature through providing methods for power analysis of detecting treatment effects in studies of polynomial change with two levels of nesting.

**Statistical Model: Treatment Assigned at Second Level (Block Design)**

Consider a simple three-level growth design where level-2 units (e.g., students) are randomly assigned to treatment or control conditions. The first level for change over time of level-2 unit $i$ in cluster $j$ can de expressed as a polynomial function, namely

$$Y_{gij} = \alpha_{0ij}c_{0g} + \alpha_{1ij}c_{1g} + \alpha_{2ij}c_{2g} + \ldots + \alpha_{(P-1)ij}c_{(P-1)g} + u_{gij} \tag{1}$$

where $c_{pg}$ represent orthogonal polynomial contrasts of degree $p$ ($p$ = 0, 1, …, $P$-1) at measurement $g$ ($g$ = 1, …, $G$), $\alpha_{pij}$'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 $\sigma_e^2$. We work with orthogonal polynomial contrasts because they facilitate the computations of estimators and their standard errors, and simplify power analysis (see Raudenbush & Liu, 2001). The results apply to studies of any length and for polynomials of any degree (Kirk, 2013).

The level-2 model incorporates the treatment ($T_{ij}$), namely

$$\alpha_{pij} = \beta_{p0j} + \beta_{p1j}T_{ij} + \xi_{pij}\,, \tag{2}$$

where $\beta_{p0j}$'s represent the average polynomial effects within level-3 units, $T_{ij}$ is a dummy variable coded as one if second level unit $i$ in third level unit $j$ is assigned to treatment condition and zero otherwise, $\beta_{p1j}$ is the treatment effect within level-3 units, and the $\xi_{pij}$'s are level-2 random effects within level-3 units for each polynomial change parameter. The random effects follow a multivariate normal distribution with zero means, variances $\tau_{pp}^2$, and covariance $\tau_{pp'}$, between random effects $\xi_{pij}$ and $\xi_{p'ij}$.

The third level equations for the intercept ($\beta_{p0j}$) and the treatment effect ($\beta_{p1j}$) are

$$\begin{aligned} \beta_{p0j} &= \gamma_{p00} + \eta_{p0j} \\ \beta_{p1j} &= \gamma_{p10} + \eta_{p1j} \end{aligned}, \tag{3}$$

where $\gamma_{p00}$'s represent the average polynomial effects across level-3 units, the $\eta_{p0j}$'s are level-3 unit specific random effects for each polynomial change parameter, $\gamma_{p10}$'s represent the average difference between the treatment and the control groups for each polynomial change parameter across level-3 units, and the $\eta_{p1j}$'s are treatment by level-3 unit random effects (interaction effects) for each polynomial change parameter. The $\eta_{p0j}$'s follow a multivariate normal distribution with zero means and variances $\omega_{p0}^2$, whilst the treatment by level-3 unit random effects also follow a normal distribution with a mean of zero and a variance $\omega_{Tpp}^2$, where the subscript *T* indicates treatment at the second level whose effect varies at the third level.

Suppose there are *M* level-3 units and *n* level-2 units within treatment or control condition within each level-3 unit, which means that the total number of level-2 units in each level-3 unit is $N = 2n$. Then, the estimate of the variance of the treatment effect for polynomial *p* is

$$Var(\gamma_{p10}) = \frac{2}{Mn}(n\omega_{Tpp}^2 + \tau_{pp}^2 + \sigma_p^2),\ \sigma_p^2 = \sigma_e^2 / \sum_{g=1}^{G} c_{pg}^2. \tag{4}$$

Suppose that a researcher wants to test the hypothesis that $\gamma_{p10}$ is different from zero and carries out a *t*-test. The test statistic is defined as

$$t = \hat{\gamma}_{p10} / \sqrt{Var(\hat{\gamma}_{p10})}. \tag{5}$$

When the null hypothesis is true, the test statistic *t* has a Student's *t*-distribution with *M-1* degrees of freedom (Konstantopoulos, 2008). When the null hypothesis is false, the test statistic *t* has the non-central *t*-distribution with *M-1* degrees of freedom and non-centrality parameter $\lambda$. 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

$$\lambda = \gamma_{p10}\sqrt{\frac{Mn}{2}}\sqrt{\frac{1}{(n\omega_{Tpp}^2 + \tau_{pp}^2 + \sigma_p^2)}}\ . \tag{6}$$

We define the standardized effect size for a polynomial degree *p* as

$$ES = \frac{\gamma_{p10}}{\sqrt{\omega_{Tpp}^2 + t_{pp}^2}}\ . \tag{7}$$

Then, the non-centrality parameter $\lambda$ of the *t*-test simplifies to

$$\lambda = \sqrt{\frac{Mn}{2}}ES\sqrt{\frac{\omega_{Tpp}^2 + \tau_{pp}^2}{n\omega_{Tpp}^2 + \tau_{pp}^2 + \sigma_p^2}}\ . \tag{8}$$

The power of a two-tailed *t*-test for a specified significance level $\alpha$ is defined as

$$p_1 = 1 - \mathrm{H}\,[c(\alpha/2, M\text{-}1), (M\text{-}1), \lambda] + \mathrm{H}\,[-c(\alpha/2, M\text{-}1), (M\text{-}1), \lambda] \tag{9}$$

where $c(a,v)$ is the level *a* one-tailed critical value of the *t*-distribution with *v* degrees of freedom, and $\mathrm{H}(x, v, \lambda)$ is the cumulative distribution function of the non-central *t*-distribution with *v* degrees of freedom and non-centrality parameter $\lambda$.

When covariates are included at the second level, the non-centrality parameter $\lambda_A$ becomes

$$\lambda_A = \sqrt{\frac{Mn}{2}}ES\sqrt{\frac{\omega_{Tpp}^2 + \tau_{pp}^2}{nw_3\omega_{Tpp}^2 + w_2\tau_{pp}^2 + \sigma_p^2}}\ . \tag{10}$$

where subscript *A* indicates adjustment because of covariates and

$$w_3 = \omega_{RTpp}^2 / \omega_{Tpp}^2, w_2 = \tau_{Rpp}^2 / \tau_{pp}^2, \tag{11}$$

where the subscript *R* indicates residual variance because of covariates. That is, $w_2$ indicates the proportion of the variance at the second level that is still unexplained, and $w_3$ indicates the proportion of the treatment by level-3 unit variance at the third level that is still unexplained. Then, the power of a two-tailed $t$-test for a specified significance level α is defined as

$$p_2 = 1 - \mathrm{H}\,[c(\alpha/2, M\text{-}q\text{-}1), (M\text{-}q\text{-}1), \lambda_A] + \mathrm{H}\,[-c(\alpha/2, M\text{-}q\text{-}1), (M\text{-}q\text{-}1), \lambda_A], \qquad (12)$$

where $q$ is the number of covariates at the third level.

**Usefulness / Applicability of Method:**

To illustrate the applicability of the methods to assess consequences of the study duration, sample sizes (students and schools), and covariates on power, we utilized data from Project STAR (Student-Teacher Achievement Ratio) in Tennessee (e.g., Finn & Achilles, 1990). This experiment employed a block randomized design, where within each school (the block) and grade, students and their teachers were randomly assigned to small classes and regular-size classes.

We fitted a three level linear change model, where treatment is at the second level, and got variance estimates

$$\sigma_1^2 = 0.060738,\ \tau_{11}^2{=}0.00753,\ \omega_{T11}^2 = 0.02097$$

Using these variance estimates, we explore how powers are influenced by the study duration, sample size, and covariates. Figure 1-3 show how variations of the study duration and sample sizes affect the power to detect the linear rate of change of the treatment effect of interest in block designs, assuming two-tailed $t$-tests at the 0.05 significance level and effect size as 0.40. The power of detecting a linear rate of change does not increase consistently as study duration increases. Similarly, power does not increase consistently as the number of students increases, especially after a certain number of students. However, as the number of schools increases, power increases significantly more. Such results indicate that to boost power it is recommended to sample more schools rather than to sample more students per school.

(please insert figure 1 here)
(please insert figure 2 here)
(please insert figure 3 here)

Figure 4 shows how the power is influenced by the proportion of unexplained variance at the second and third levels. The results show that power increases when covariates are added in the model, as expected.

(please insert figure 4 here)

**Conclusions:**

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 study duration, 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.

Several findings emerged from this study that applies to both CRD and BRD. First, power increases as the study duration, the number of students in each school, or the number of schools increases. Other things being equal, the number of level-3 units (clusters) influences power more than the number of level-2 units (individuals) or the duration of the study. This indicates clearly that researchers should sample more schools rather than students within schools to maximize power. Second, covariates that explain a proportion of variances at the second or third level could increases powers and thus reduce the study duration or sample sizes needed to boost power to a certain level.

**Appendices**


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(1)*, 30-59.

Bloch, D. A. (1986). Sample size requirements and the cost of a randomized clinical trial with repeated measurements. *Statistics in Medicine, 5*(6), 663-667.

Boruch, R. F., & Gomez, H. (1977). Sensitivity, bias, and theory in impact evaluations. *Professional Psychology, 8*(4), 411.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New York: Academic.

Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*(3), 557-577.

Fisher, R. A. (1928). *Statistical methods for research workers* (2d ed.). Edinburgh, London: Oliver and 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*(1), 70-93.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 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*(1), 117-118.

Kirk, R. E. (2013). *Experimental design : procedures for the behavioral sciences* (4th ed.). Thousand Oaks: Sage Publications.

Konstantopoulos, S. (2008). The power of the test for treatment effects in three-level block randomized designs. *Journal of Research on Educational Effectiveness*, 1(4), 265-288.

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(4),* 481-499.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114*(2), 497-532.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research* (Vol. 19): Sage.

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2*(4), 371-402.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*(1), 123-151.

Plackett, R. L. (1960). *Principles of regression analysis.* Oxford: Clarendon Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models : applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.

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*(4), 387-401.

Raudenbush, S. W., Martinez, A., & Spybrook, J. (2007). Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis, 29(1)*, 5-29.

Seber, G. A. F., & Lee, A. J. (2003). *Linear regression analysis* (2nd ed.). Hoboken, N.J.: Wiley-Interscience.

Spybrook, J., Raudenbush, S. W., Liu, X. F., Congdon, R., & Martínez, A. (2011). Optimal Design for Longitudinal and Multilevel Research: Documentation for the Optimal Design Software Version 3.0. Available from www.wtgrantfoundation.org.

## Appendix B. Figures

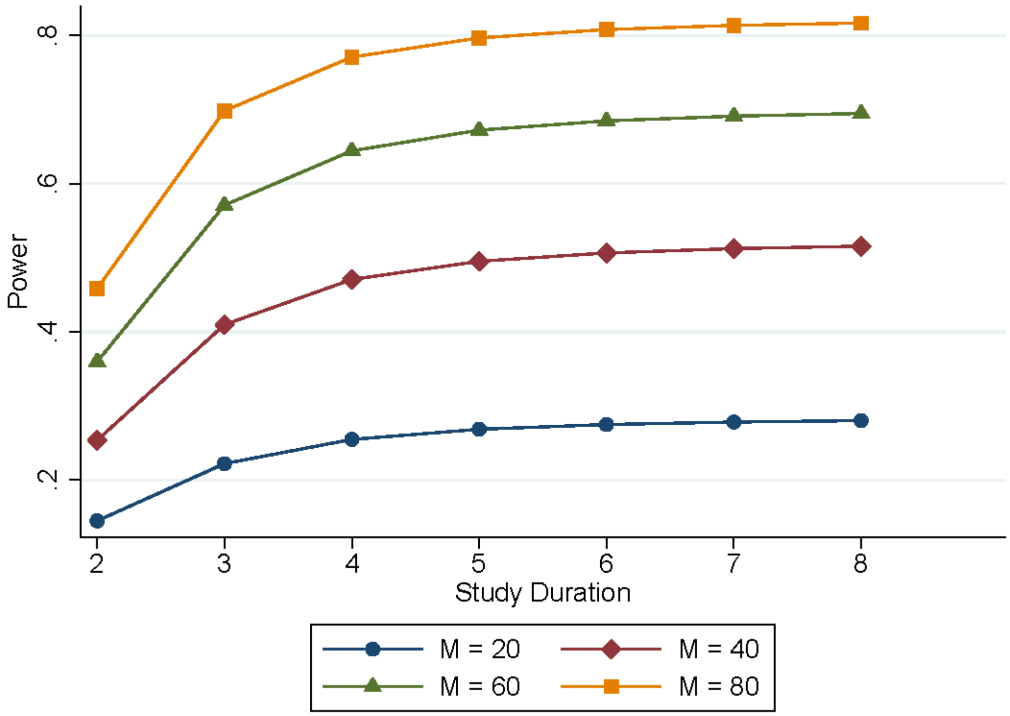


Figure 1. Effect of Study Duration (*D*) and Number of Schools (*M*) on Power Holding Number of Students (*N*) in Each School Constant at 40: BRD, Linear Rate of Change
Note: Effect size is 0.4 with a significance level of 0.05.

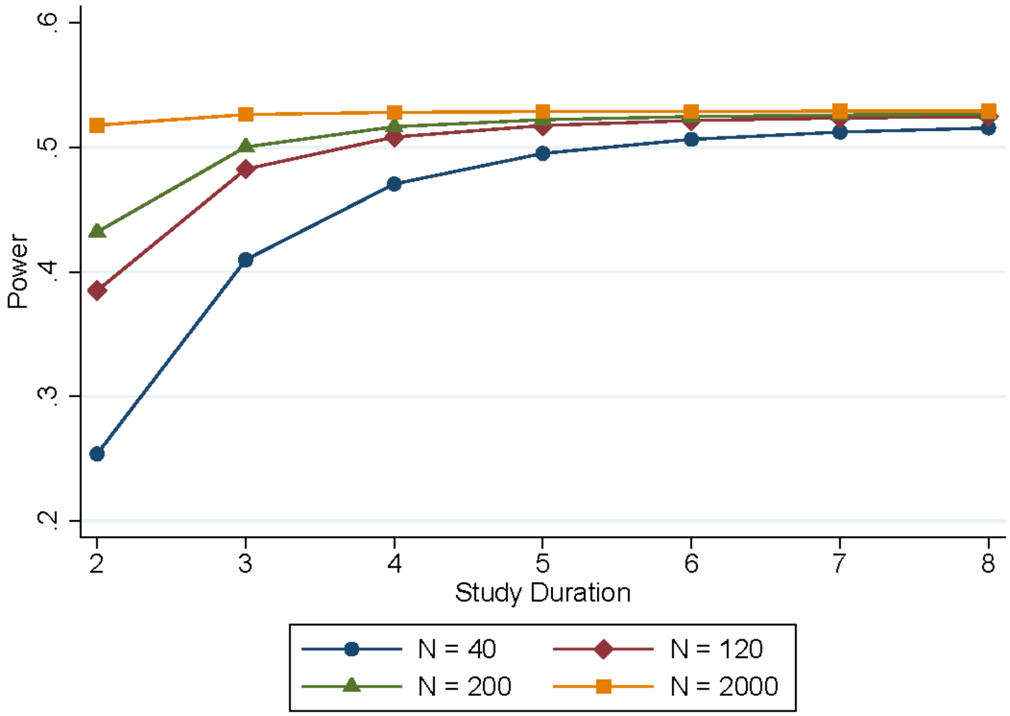


Figure 2. Effect of Study Duration (*D*) and Number of Students (*N*) on Power Holding Number of Schools (*M*) Constant at 40: BRD, Linear Rate of Change
Note: Effect size is 0.4 with a significance level of 0.05.

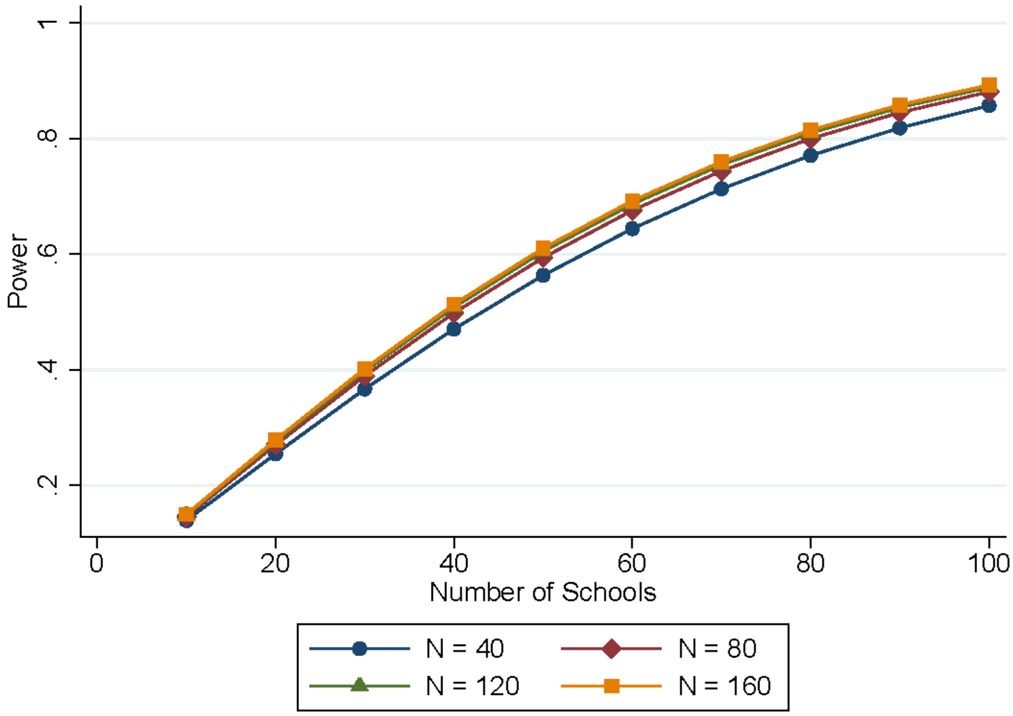


Figure 3. Effects of Number of Schools (*M*) and Number of Students (*N*) on Power Holding Study Duration (*D*) Constant at 4: BRD, Linear Rate of Change
Note: Effect size is 0.4 with a significance level of 0.05.

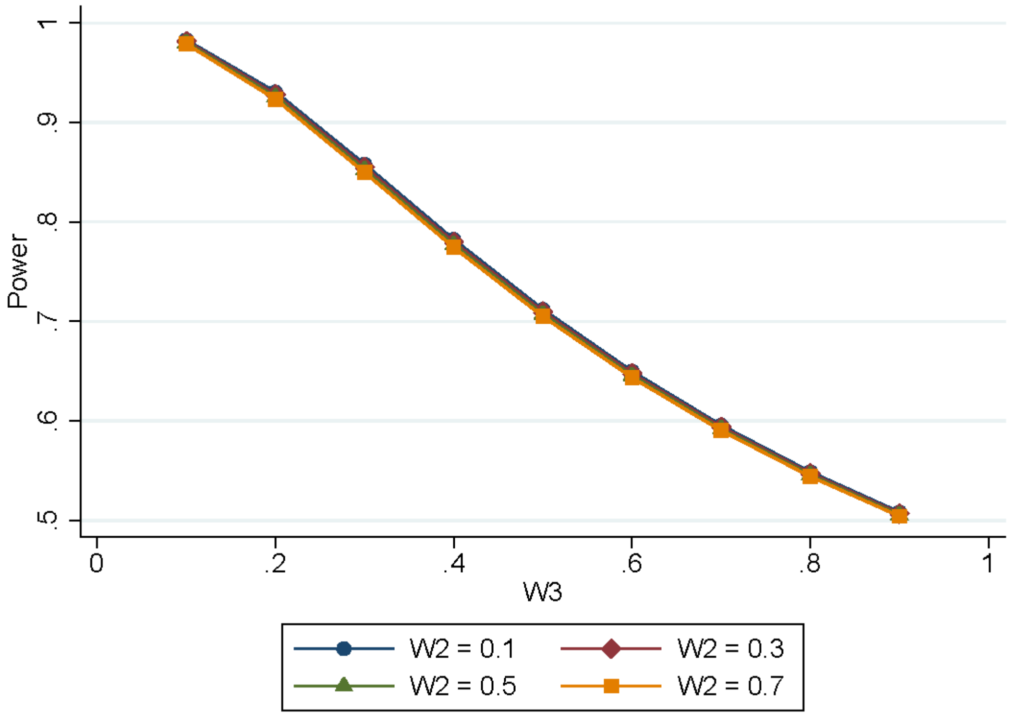


Figure 4. Effect of Covariates on Power: BRD, Linear Rate of Change
Note: The study duration is 4 with 40 schools and 40 students in each school; significance level is 0.05.