Experimental Power for Three-Level School-Randomized Studies Probing Mediation

Authors and Affiliations:

Kyle Cox (presenting)
University of North Carolina at Charlotte
kyle.cox@uncc.edu

Ben Kelcey
University of Cincinnati
ben.kelcey@gmail.com

Yanli Xie
xieyl@mail.uc.edu

Fangxing Bai
baifg@mail.uc.edu

Purpose

We develop closed-form expressions to estimate the statistical power to detect mediation effects in three-level school-randomized studies. Mapping the sensitivity of three-level designs is of critical importance in education because such design sensitivity directly governs the types of evidence researchers can bring to bear on theories of action under common sample sizes. The results provide a set of formulas and software tools that guide researchers in the planning multilevel studies incorporating mediation. We extend tests of mediation that can be used both in the planning and analysis phases including the asymptotic Sobel test, component-wise joint test, the resampling-based Monte Carlo interval test, and the partial posterior predictive distribution test. These tests represent classical and modern approaches, capture important differences among tests in terms type one error rates and power levels, can each be implemented in the design phase, and are collectively representative of the range of tests most commonly used in the literature.

Background

Three-level designs are quite widespread in education research because they align well with the core organizational structure of schooling—students nested within classrooms nested within schools (e.g., Spybrook, Shi, & Kelcey, 2016). An important design consideration in three-level designs is the power with which we can detect effects if they exist. Historically, studies have been exclusively designed with a focus on detecting main effects. Recent literature has, however, expanded that consideration to include mediation effects because the comprehensive study of the components of a theory of action is critical to advancing scientific theories (e.g., Desimone, 2009; US DoE & NSF, 2013). Although prior literature has detailed the power to detect main and moderation effects in three-level designs, literature regarding the calculation of power for mediation in three-level designs has been largely absent (Raudenbush, 1997; Dong, Kelcey, & Spybrook, 2018).

Model

For brevity, we outline the development and application of power analyses using one simple (albeit sub-optimal) example—the Sobel test with students nested within classrooms nested within schools. The Sobel test has received due criticism for shortcomings; however, for the purposes of an abstract, the simplicity of its derivation provides a good conceptual description of the more general analyses and results. We again note that although we outline the Sobel test under a 3-2-1 mediation example, our full study derives power formulas for the aforementioned tests and for a broader range multilevel mediation effects as detailed through the potential outcomes framework (e.g., Kelcey, Dong, Spybrook, & Cox, 2017).

To conceptually outline our analysis, we consider one example estimand—the mediation effect that describes the extent to which a school-assigned treatment (T) impacts an individual-level outcome (Y) through a classroom-level mediator (M) conditional upon covariates. The corresponding path model is

Mediator model

$$M_{jk} = \pi_{0k} + \pi_1(\bar{X}_{jk} - \bar{X}_k) + \pi_2(W_{jk} - \bar{W}_k) + \pi_3 U_{jk} + \varepsilon_{jk}^M \qquad \varepsilon_{jk}^M \sim N(0, \sigma_{M|}^2) \tag{1}$$

$$\pi_{0k} = \zeta_{00} + aT_k + \zeta_{01}\bar{X}_k + \zeta_{02}\bar{W}_k + \zeta_{03}Z_k + u_{0k}^M \qquad u_{0k}^M \sim N(0, \tau_{M|}^2)$$

Outcome model

$$Y_{ijk} = \beta_{0jk} + \beta_1 (X_{ijk} - \bar{X}_{jk}) + \beta_2 V_{ijk} + \varepsilon_{ijk}^Y \qquad \varepsilon_{ijk}^Y \sim N(0, \sigma_{Y|}^2)$$
$$\beta_{0jk} = \gamma_{00k} + b_{2k}(M_{jk} - \bar{M}_k) + \gamma_{01}(\bar{X}_{jk} - \bar{X}_k) + \gamma_{02}(W_{jk} - \bar{W}_k) + \gamma_{03} U_{jk} u_{0jk}^Y \qquad u_{0jk}^Y \sim N(0, \tau_{Y|}^2) \tag{2}$$
$$\gamma_{00k} = \zeta_0 + B\bar{M}_k + \Delta\bar{M}T_k + c' T_k + \xi_1 \bar{X}_k + \xi_2 \bar{W}_k + \xi_3 Z_k + \upsilon_{00k}^Y \qquad \upsilon_{00k}^Y \sim N(0, \nu_{Y|}^2)$$

Here we use Y_{ijk} as the outcome for student *i* in class *j* in school *k,* and add V_{ijk} as a student-level covariate that varies only among students, $M_{jk} - \bar{M}_k$ as the group-centered class-level mediator with coefficient b_2 , $\bar{M}_k$ as the mean of the mediator in school *k* with path coefficient *B*, *c'* as the treatment-outcome conditional path coefficient, and υ_{00k}^Y , u_{0jk}^Y and ε_{ijk}^Y as the school, class, and student error terms. Given this formulation, our example analysis investigates the power to detect the mediation effect defined by *aB* (we assume no interactions for ease of presentation).

One of the most common tests of statistical significance for indirect effects is the Sobel test that compares the ratio of the product of the *a* and *B* coefficients to the standard error (σ_{aB}) of this product

$$z^{Sobel} = aB / \sqrt{\sigma_{aB}^2} \tag{3}$$

Our omitted derivations extend this test to three-level settings such that the resulting Sobel test for a mediation effect is

$$z_{3L}^{Sobel} = \frac{aB}{\sqrt{\sigma_{aB}^2}} = \frac{aB}{\sqrt{\left(\frac{\tau_M^2(1 - R_{M_\omega^{L3}}^2) + (1 - R_{M_\omega^{L2}}^2)\sigma_M^2 / n_2}{n_3 \sigma_T^2 (1 - R_T^2)}\right)B^2 + a^2\left(\frac{\nu_Y^2(1 - R_{Y^{L3}}^2) + \tau_Y^2(1 - R_{Y^{L2}}^2)/n_2 + (1 - R_{Y^{L1}}^2)\sigma_Y^2 / n_2 n_1}{n_3(\tau_M^2(1 - R_{M_\Omega^{L3}}^2) + (1 - R_{M_\Omega^{L2}}^2)\sigma_M^2 / n_2)}\right)}} \tag{4}$$

We use τ_M^2 and σ_M^2 as the unconditional school- and class-level variances of the mediator, σ_T^2 as variance of the treatment, $R_{M_\omega^{L3}}^2$ and $R_{M_\omega^{L2}}^2$ as the school- and class-level mediator variance explained by predictors in the mediation model, R_T^2 as treatment variance explained by predictors, and n_3 and n_2 as the school- and class-level sample sizes. Similarly we introduce ν_Y^2, τ_Y^2 , σ_Y^2 , τ_{MT}^2 , and σ_{MT}^2 as the unconditional school-, class- and student-level outcome and the unconditional school- and class-level treatment-by-mediator variance and $R_{MT_\Omega^{L3}}^2$, $R_{MT_\Omega^{L2}}^2$, $R_{M_\Omega^{L3}}^2$, and $R_{M_\Omega^{L2}}^2$ as the school- and class-level interaction and mediator variances explained by other predictors in the outcome model.

Assuming the alternative hypothesis is true, the power of a two-sided test to detect the multilevel mediation effect can be approximated with

$$P(z_{3L}^{Sobel} > z_{critical}) = 1 - \Phi(z_{critical} - z_{3L}^{Sobel}) + \Phi(-z_{critical} - z_{3L}^{Sobel}) \tag{5}$$

where Φ is the normal distribution with $z_{critical}$ as the chosen critical value (e.g., 1.96) corresponding to a nominal type one error rate.

Findings

To illustrate the utility of the results, consider an example in which schools are randomly assigned to treatment conditions (e.g., professional development), we observe a classroom-level mediator (e.g., teacher instructional quality), and record a student-level outcome (achievement). Assume that the theory of action suggests that exposure to professional development (treatment) improves student achievement (outcome) by improving on teacher instruction (mediator). In planning a study, we might inquire as to approximately how many schools we need to have a reasonably high chance of detect the mediation effect if it exists. Previously, formulas were not available to determine such sample sizes and estimate power. Using the resulting formulas, let us assume we anticipate the following parameter values for our study:

$a = 0.50$ (treatment-mediator relationship [Cohen's d scale])

$B = 0.30$ (mediator-outcome relationship [Standardized regression scale])

$v_Y^2 = 0.10$ (unconditional outcome variance at school-level)

$\tau_Y^2 = 0.10$ (unconditional outcome variance at class-level)

$\sigma_Y^2 = 0.80$ (unconditional outcome variance at individual-level)

$\tau_M^2 = 0.20$ (unconditional outcome variance at school-level)

$\sigma_M^2 = 0.80$ (unconditional outcome variance at class-level)

$R^2_{Y L3} = R^2_{Y L2} = R^2_{Y L1} = 0.50$ (outcome variance explained at each level)

$R^2_{M L3} = R^2_{M L2} = 0.50$ (mediator variance explained at each level)

$P = 0.50$ (proportion of schools receiving treatment)

$n_2 = 4$ (classrooms/school)

$n_1 = 20$ (students/classroom)

The resulting power curve as a function of the school sample size is plotted in Figure 1. Based on our derivations, we would expect that sampling about 40 schools would yield about a 0.80 chance of detecting the mediation effect under a three-level group-randomized design.

Conclusions

Designing studies with the capacity to test whether or not a program works and to examine the mechanisms underlying the program theory has become a prominent and critical aim of research studies. Our work bridges that gap by establishing a flexible framework from which to estimate power under many common three-level designs and implements them in easy to use software. With these tools, we hope to streamline multilevel mediation power analyses in ways that help researchers understand how to carefully plan studies.

References

Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181–199.

Institute of Education Sciences, U.S. Department of Education & National Science Foundation (2013). *Common Guidelines for Education Research and Development* (NSF 13-126). Retrieved February 15, 2014, from http://ies.ed.gov/pdf/CommonGuidelines.pdf

Dong, N., Kelcey, B., & Spybrook, J. (2018). Power analyses for moderator effects in three-level cluster randomized trials. *Journal of Experimental Education*, *86*, 3, 489-514.

Kelcey, B., Dong, N., Spybrook, J., & Cox, K. (2017). Statistical power for causally-defined indirect effects in group-randomized trials with individual-level mediators. *Journal of Educational and Behavioral Statistics, 42,* 499-530.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. Psychological Methods, 2, 173-185.

Spybrook, J., Shi, R., & Kelcey, B. (2016). Progress in the Past Decade: An Examination of the Precision of Cluster Randomized Trials Funded by the U.S. Institute of Education Sciences. *International Journal of Research & Method in Education, 39, 3, 255-267.*

Figure 1
Power to detect mediation effect in a three-level group-randomized design