Proceedings of the 90th Annual International Meeting of the Psychometric Society, Minneapolis, United States (2025).

Cost-Efficient Sampling Strategies for Cluster-Randomized Trials Detecting Moderation and Main Effects

Z. Shen [*†] B. Kelcey [‡] J. Li [†] W. Leite [¶] and Z. Lu [†]
†Department of Educational Psychology, University of Georgia, Athens, 30602, Georgia, USA
‡School of Education, University of Cincinnati, Cincinnati, 45221, Ohio, USA
¶College of Education, University of Florida, Gainesville, 32611, Florida, USA
*Corresponding author. Email: zuchao.shen@uga.edu

Abstract
Moderation analyses can answer questions about where, for whom, and under what conditions intervention effects are most salient. Together with the main effect analyses, which assess whether interventions work, the results from the moderation and main effect analyses provide the evidence base for effective policy development. The literature has developed statistical power formulas for the main and moderation effects. However, literature has not offered a holistic framework that simultaneously considers both cost efficiency and statistical power for experimental studies investigating these effects. The present study develops a jointly optimal design framework for cluster-randomized trials (CRTs) investigating both moderation and main effects. Such a framework enables researchers to identify optimal sampling allocations by accounting for sampling costs and statistical power for both effects. The results show that the optimal sample allocations in CRTs are the same for the average treatment effect and the moderated treatment effect with cluster-level moderators. The proposed methods, including optimal sample allocation calculations and power analyses for both effects, have been implemented in the R package *odr*.

Keywords: moderation, optimal design, statistical power, experimental design

Moderation analyses can investigate variation in treatment effects across many policy-relevant dimensions, such as groups, geographic locations, and other structural settings. The results from moderation analyses can provide the evidence base for effective policy development and implementation (Raudenbush & Liu, 2000; Spybrook et al., 2016). When designing experimental studies to detect moderation effects, researchers must first consider whether the treatment works, which is usually assessed through main effect analyses. Despite recent developments in statistical power formulas for main and moderation (Dong et al., 2018; Dong, Kelcey, & Spybrook, 2021; Dong, Spybrook, et al., 2021; Raudenbush, 1997; Raudenbush & Liu, 2000), the literature has noted the lack of design tools and frameworks that simultaneously consider cost efficiency and statistical power when designing experimental studies targeting both effects. The present study develops a joint framework for cluster-randomized trials (CRTs) investigating both moderation and main effects. Such a framework enables researchers to identify optimal sampling allocations by accounting for sampling costs and statistical power for both effects.

1. Relevant Literature

1. 1 Varied Sampling Costs in Experiments

Design literature has noted that sampling costs vary across treatment conditions and levels in multilevel experiments (Hedges & Borenstein, 2014; Konstantopoulos, 2009, 2011; Liu, 2003; Nam, 1973; Raudenbush, 1997; Raudenbush & Liu, 2000; Shen & Kelcey, 2020, 2022a, 2022b). Applied literature has documented that the costs in multilevel experiments varied across treatment conditions and levels (Clements et al., 2011; Greenleaf et al., 2011; Hiscock et al., 2008; Jacob et al., 2015; Lewis et al., 2013; May et al., 2015; Springer et al., 2011). The cost of adding a cluster is usually higher than that of an individual. Similarly, the costs of adding a cluster in the experimental group are higher than those in the control group because the delivery and implementation of an intervention will increase the budget request.

1. 2 Sample Allocation and Optimal Design

When the cost of adding a unit varies across levels and treatment conditions, sample size allocation plays a critical role in determining the cost-efficiency and effectiveness of a study design. An optimal design framework aims to identify the sample allocation that maximizes statistical power within a fixed budget. Literature has

developed an optimal design framework for two-, three-, and four-level multisite- randomized trials (MRTs) and cluster-randomized trials (CRTs) detecting main effects (Hedges & Borenstein, 2014; Konstantopoulos, 2009, 2011; Raudenbush, 1997; Raudenbush & Liu, 2000), two-level CRTs testing mediation effects (Cox & Kelcey, 2019; Kelcey & Shen, 2020), and two-level regression discontinuity designs investigating main effects (Rhoads & Dye, 2016). However, the conventional framework assumes equal sampling costs between treatment conditions. It only optimizes sampling ratios across levels, but not between treatment conditions.

Recent literature relaxes the assumption of equal costs across treatment conditions and develops a flexible optimal design framework for detecting main effects in two- and three-level MRTs and CRTs (Shen & Kelcey, 2020, 2022a, 2022b; Shen et al., 2025). This framework allows sampling costs to vary across treatment conditions and levels and identifies optimal sampling ratios across these domains. The flexible framework can identify more powerful and efficient sample allocations than the conventional framework. Thus far, the optimal design framework has been developed primarily for a single effect (e.g., a main or mediation effect). However, an optimal design framework has not been developed to identify optimal sample allocation for moderation effects. The goal of the present study is to develop an optimal design framework for CRTs that investigates moderation effects, along with main effects.

Prior literature has shown that optimal sample allocations differ for main and mediation effects (Shen et al., 2025; Shen, Li, & Leite, in press). In these scenarios, a jointly optimal design framework is needed to simultaneously account for both effects, rather than treating them separately. Next, we present statistical power formulas, develop optimal design parameter expressions for moderation effects, and compare them with those for main effects in CRTs.

2. Statistical Power

2. 1 Statistical Models

In a two-level CRT, whole clusters are randomly assigned to an experimental or a control group. We can estimate the average (or main) treatment effect in a two-level CRT through multilevel linear models (Raudenbush & Bryk, 2002). The individual-level model is

$$Y_{ij} = \beta_{0j} + \beta_{10}^T X_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N(0, \sigma_{1A}^2), \tag{1}$$

where Y_{ij} is the continuous outcome of individual i ($i = 1, 2, \ldots, n$) in cluster j ($j = 1, 2, \ldots, J$), β_{0j} is the mean score of cluster j that varies across clusters, X_{ij} is a vector of level-one covariates, β_{10} is the vector of regression coefficients of covariates with β_{10}^T as its transposed matrix , ε_{ij} is the individual-level error term with a variance of σ_{1A}^2.

The cluster-level model is

$$\beta_{0j} = \gamma_{00} + \delta T_j + \gamma_{01}^T W_j + u_{0j} \qquad u_{0j} \sim N(0, \sigma_{2A}^2). \tag{2}$$

Here, γ_{00} is the overall intercept, T_j is the treatment indicator with $T_j = 1$ for clusters in the experimental group, otherwise $T_j = 0$, and the average treatment effect is δ. W_j is a vector of cluster-level covariates, γ_{01} is the vector of regression coefficients of the covariates, u_{0j} is the random effect with a conditional variance of σ_{2A}^2.

For models without covariates, we can denote the variances at the individual and cluster levels as σ_1^2 and σ_2^2. The standardized treatment effect (Hedges & Hedberg, 2007) is

$$d = \frac{\delta}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \tag{3}$$

The proportion of within-treatment outcome variance explained by covariates at the level i (i = 1, 2; Hedges & Hedberg, 2007) is

$$R_i^2 = \frac{\sigma_i^2 - \sigma_{iA}^2}{\sigma_i^2}. \tag{4}$$

Similarly, we can use multilevel models to estimate moderation effects. Here, we consider moderators at the cluster level; the individual-level model is the same as (1), and the cluster-level model (2) changes to the following one

$$\beta_{0j} = \gamma_{00} + \delta T_j + \gamma_{01}^T W_j + \gamma_{02} M_j + \gamma_{03} M_j T_j + u_{0j} u_{0j} \sim N\left(0, \sigma_{2A(m)}^2\right), \tag{5}$$

where M_j is a moderator, γ_{03} is the moderated treatment effect that captures the differential treatment effect across the value of M_j. We can denote the level-2 residual variance as $\sigma_{2(m)}^2$ in the model without covariates. Then, the proportion of within-treatment and within-moderator outcome variance explained by covariates at the cluster level is

$$R^2_{2(m)} = \frac{\sigma^2_{2(m)} - \sigma^2_{2A(m)}}{\sigma^2_{2(m)}}. \tag{6}$$

In the same vein, we can denote the standardized moderation effect as

$$\gamma = \frac{\gamma_{03}}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \tag{7}$$

2. 2 Statistical Power

The variance of the estimator is a key component in statistical power formulas. The variance of the moderation effect estimator under the maximum likelihood estimation can be expressed as (Dong, Spybrook, et al., 2021)

$$\sigma^2_{\gamma_{03}} = \frac{\sigma^2_{2A} + \sigma^2_{1A}/n}{J\sigma^2_T \sigma^2_M}. \tag{8}$$

Here, σ^2_T is the variance of the treatment indicator, which equals $p(1\text{-}p)$ with p as the proportion of clusters assigned to the treatment condition. σ^2_M is the within-treatment variance of the moderator, which equals 1 in standardization for continuous moderators and equals $Q(1\text{-}Q)$ for binary moderators with Q as the proportion of units in one of the binary groups. The variances of the standardized moderation effect estimator are

$$\sigma^2_{\gamma(b)} = \frac{\sigma^2_{\gamma_{03}}}{\sigma_1^2 + \sigma_2^2} = \frac{\rho\left(1 - R^2_{2(\mathrm{m})}\right) + (1-\rho)(1 - R_1^2)/n}{Jp(1-p)Q(1-Q)}, \tag{9}$$

for binary moderators, and

$$\sigma^2_{\gamma(c)} = \frac{\sigma^2_{\gamma_{03}}}{\sigma_1^2 + \sigma_2^2} = \frac{\rho\left(1 - R^2_{2(\mathrm{m})}\right) + (1-\rho)(1 - R_1^2)/n}{Jp(1-p)}, \tag{10}$$

for continuous moderators. Here, ρ is the unconditional intraclass correlation coefficient (Hedges & Hedberg, 2007) with

$$\rho = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{11}$$

The variance of the standardized main/average effect estimator (Shen & Kelcey, 2020) is

$$\sigma^2_d = \frac{\rho\left(1 - R_2^2\right) + (1-\rho)(1 - R_1^2)/n}{Jp(1-p)}. \tag{12}$$

The statistical power at the significance level a for the two-tailed test (Dong et al., 2021; Shen & Kelcey, 2020) is

$$Power = 1 - \Phi\left[c\left(\frac{\alpha}{2}, df\right),\ df,\ \lambda\right] + \Phi\left[-c\left(\frac{\alpha}{2}, df\right),\ df,\ \lambda\right], \tag{13}$$

where $c\left(\frac{\alpha}{2}, df\right)$ is the critical value in a t-distribution with df degrees of freedom, $\Phi(x,\ df,\ \lambda)$ is the cumulative distribution function of the noncentral t-distribution with df degrees of freedom and a noncentrality parameter λ. For main effects, we have $\lambda = \frac{d}{\sqrt{\sigma_d^2}}$ and $df = J\text{-}q\text{-}2$ with q as the number of covariates at the cluster level. For moderation effects, we have $\lambda = \frac{\gamma}{\sqrt{\sigma_\gamma^2}}$ ($\sigma_\gamma^2 = \sigma^2_{\gamma(b)}$ or $\sigma_\gamma^2 = \sigma^2_{\gamma(c)}$) and $df = J\text{-}q\text{-}4$. Statistical power formulas have been implemented in R package *odr* (Shen & Kelcey, 2026).

3. Optimal Sample Allocation Identification

Suppose the costs of enrolling an individual in the experimental and the control group are c_1^T and c_1, respectively, and the respective costs of each additional cluster are c_2^T and c_2. Then, the total budget is

$$m = Jp(nc_1^T + c_2^T) + J(1-p)(nc_1 + c_2), \tag{14}$$

with n the number of individuals per cluster and J as the total number of clusters. The optimization task is to identify the sample allocation that maximizes statistical power subject to this budget constraint. The literature has developed the optimal sample allocation framework for two-level CRTs investigating main effects (Shen & Kelcey, 2020) using the Lagrange multiplier (first-order derivative) method. The optimization is done by minimizing the variance of the treatment effect estimator with respect to p and n, thereby approximately maximizing statistical power under a fixed budget.

If we use the same method to identify the optimal allocation for the moderation effect, the resulting optimal design parameters are identical to those for main effects. Because the variance of the moderator effect estimator for a continuous moderator in (10) is identical to the variance of the main effect estimator in (12),

except for the proportion of variance explained by covariates at the cluster level ($R^2_{2(m)}$ vs R^2_2), minimizing them with respect to *p* and *n* given the constraint function in (14) results in the same expressions. The variance of the moderator effect estimator for a binary moderator in (9) is proportional to the variance of the main effect estimator in (12), by a factor of 1/Q(1-Q). The resulting minimization with respect to *p* and *n* yields the same results as the main effects.

Specifically, the optimal *p* and *n* must satisfy the following two equations.

$$p = \frac{\sqrt{(c_1 n + c_2)/(c_1^T n + c_2^T)}}{1 + \sqrt{(c_1 n + c_2)/(c_1^T n + c_2^T)}}. \tag{15}$$

$$n = \sqrt{\frac{(1-\rho)(1-R_1^2)}{\rho(1-R^2_{2(m)})}} \sqrt{\frac{(1-p)c_2 + pc_2^T}{(1-p)c_1 + pc_1^T}}. \tag{16}$$

The optimal p is mainly determined by the sampling cost ratio across treatment conditions ($\sqrt{(c_1 n + c_2)/(c_1^T n + c_2^T)}$). The more expensive the sampling in the treatment condition is, the fewer units should be assigned to it (Shen & Kelcey, 2020). The optimal *n* is influenced by two factors: the outcome residual variance ratio across levels and the sampling cost ratio across levels. The trivial difference is in the optimal *n* expression, where it is $R^2_{2(m)}$ for moderation effects and it is R^2_2 for main effects. Such a difference will result in nearly identical optimal *n*, as optimal design solutions are robust to changes in parameter values (e.g., Shen & Kelcey, 2020). The optimal design solutions have been implemented in the *od.2* function of the R package *odr* (Shen & Kelcey, 2026).

Notice that these solutions are derived through approximation and are not directly maximizing the statistical power under fixed budget or minimizing the budget to achieve a target power. Alternatively, we can use a meta-heuristic optimization method, ant colony optimization (ACO; Dorigo & Stützle, 2004; Socha & Dorigo, 2008), to identify the optimal sample allocation in experiments investigating both moderation and main effects. The ACO algorithm can identify optimal sample allocations with multiple power formulas (Shen et al., 2025). To use the ACO algorithm, we first define the objective function as the required budget to achieve target power levels for the main and moderation effects (e.g., 80% and 80%). Note that we can allow the power for the main and moderation to differ (e.g., by specifying power levels for the main and moderation differently in the arguments). The optimal design parameters for the two-level CRTs are the number of individuals per cluster (*n*) and the proportion of clusters in the experimental group (*p*).

The ACO algorithm has two stages for identifying joint optimization: Initialization and iteration. The four steps in the initialization stage are as follows. (1) The algorithm starts by randomly sampling many (e.g., 50) sets of values of *n* and *p*. We can specify that *n* is sampled from 2 to 50 and *p* is from .10 to .50. (2) For each set of sampled values of *n* and *p*, the algorithm performs power analyses to calculate the number of clusters needed (for the main and moderation effects separately). The larger of the two is retained as the smaller one will not provide enough power for the other effect. (3) The algorithm calculates the total resources needed (*m*), using the budget function. (4) The algorithm forms a probability density function in terms of the required resources (*m*) across *n* and *p*. This probability density function will allow values of *n* and *p* that are associated with fewer resources and their neighborhood areas to have a higher probability of being sampled in the next step.

The steps in the iteration stages are similar to those in the initialization stage, except that in step (1) the algorithm samples *n* and *p* from the latest probability density function, and in step (4) it updates the probability density function. The algorithm iteratively repeats steps (1) to (4) until one of the termination conditions is met (e.g., a maximum number of evaluations or an optimal value of the objective function). Once optimal design parameters (*n* and *p*) are identified, we can perform power analysis using these identified values. Power analysis results in an optimal design will have these features: achieving the maximum statistical power within the same budget and using the least resources to achieve the fixed power. These two optimization methods (i.e., the ACO algorithm and first-order derivative methods) often result in different optimal sample allocations with similar design efficiency. They both successfully perform the optimization tasks (e.g., minimizing the variance and maximizing statistical power). We recommend using the ACO algorithm, as it directly optimizes via the power function and does not rely on approximations via the variance formula. The revision has added this explanation. The algorithm has been implemented in R package *odr* (Shen & Kelcey, 2026).

4. Illustration

Suppose we design a CRT targeting a total effect of d = 0.20 with $\rho = .20$, $R_1^2 = .55$, and $R_2^2 = .75$ (Hedges & Hedberg, 2007). In addition, we target the moderation effect of a binary cluster-level moderator with the following additional parameters: $\gamma = .15$, and Q = .50. The cost structures of sampling are $c_1 = c_1^T = \$10$, $c_2 = \$100$, $c_2^T = \$2{,}000$. The Appendix presents the R code for the illustration.

Using the methods developed, we can identify the jointly optimal allocation as p = .27 and n = 21 to detect both main and moderation effects with 80% power. The total budget required to achieve 80% power for both effects is $392,791, with a total of 477 clusters (129 in the treated and 348 in the control). Figure 1 shows how statistical power changes across sample allocations under the fixed budget (dashed curves for moderation and solid curves for main). Statistical power decreases when a design departs from the optimal allocation in any direction (Figure 1).

If we use a balanced design with p = .50 (and n = 21), it can achieve only 72% power for the moderation effect under the same budget (Figure 1b). Such a balanced design can only detect a minimum detectable moderation effect of 0.165, 10% higher than the optimal design under the same budget. The balanced design requires 21% more budget than the optimal design to achieve the same 80% power for the moderation effect. Figure 1 shows that the moderation effect is more demanding than the main effect, as the budget can provide almost 100% power for the main effect but only 80% for the moderation.

5. Conclusion

This study develops the expressions for optimal design parameters in CRTs with cluster-level moderators. The results show that researchers can use the optimal sample allocations identified for main effects, which are identical to those for moderation effects with cluster-level moderators. However, researchers should be cautioned that this is a specific case. In general, optimal sample allocations for different effects differ (e.g., Shen et al., 2025; Shen, Li, & Leite, in press). Future studies should develop optimal design expressions for other types of experiments investigating moderation effects with moderators at different levels.

References

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education, 42*(2), 127-166. https://doi.org/10.5951/jresematheduc.42.2.0127

Cox, K., & Kelcey, B. (2019). Optimal sample allocation in group-randomized mediation studies with a group-level mediator. *The Journal of Experimental Education, 87*(4), 616-640. https://doi.org/10.1080/00220973.2018.1496060

Dong, N., Kelcey, B., & Spybrook, J. (2018). Power analyses of moderator effects in three-level cluster randomized trials. *Journal of Experimental Education, 86*(3), 489-514. https://doi.org/10.1080/00220973.2017.1315714

Dong, N., Kelcey, B., & Spybrook, J. (2021). Design considerations in multisite randomized trials to probe moderated treatment effects. *Journal of Educational and Behavioral Statistics, 46*(5):527-559. https://doi.org/10.3102%2F1076998620961492

Dong, N., Spybrook, J., Kelcey, B., & Bulus, M. (2021). Power analyses for moderator effects with (non) randomly varying slopes in cluster randomized trials. *Methodology, 17*(2), 92-110. https://doi.org/10.5964/meth.4003

Dorigo, M., & Stützle, T. (2004). *Ant colony optimization.* Cambridge, MA: MIT Press.

Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., ... & Jones, B. (2011). Integrating literacy and science in biology: Teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal, 48*(3), 647-717. https://doi.org/10.3102/0002831210384839

Hedges, L. V., & Borenstein, M. (2014). Conditional optimal design in three-and four-level experiments. *Journal of Educational and Behavioral Statistics, 39*(4), 257-281. https://doi.org/10.3102/1076998614534897

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87. https://doi.org/10.3102/0162373707299706

Hiscock, H., Bayer, J. K., Price, A., Ukoumunne, O. C., Rogers, S., & Wake, M. (2008). Universal parenting programme to prevent early childhood behavioural problems: cluster randomised trial. *British Medical Journal, 336*(7639), 318-321. https://doi.org/10.1136/bmj.39451.609676.AE

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis, 37*(3), 314-332. https://doi.org/10.3102%2F0162373714549620

Kelcey, B., & Shen, Z. (2020). Strategies for efficient experimental design in studies probing 2-1-1 mediation. *The Journal of Experimental Education, 88*(2), 311–334. https://doi.org/10.1080/00220973.2018.1533796

Konstantopoulos, S. (2009). Incorporating cost in power analysis for three-level cluster-randomized designs. *Evaluation Review, 33*(4), 335-357. https://doi.org/10.1177/0193841X09337991

Konstantopoulos, S. (2011). Optimal sampling of units in three-level cluster randomized designs: An ANCOVA framework. *Educational and Psychological Measurement*, *71*(5), 798-813. https://doi.org/10.1177/0013164410397186

Lewis, K. M., DuBois, D. L., Bavarian, N., Acock, A., Silverthorn, N., Day, J., ... & Flay, B. R. (2013). Effects of positive action on the emotional health of urban youth: A cluster-randomized trial. *Journal of Adolescent Health, 53*(6), 706-711. https://doi.org/10.1016/j.jadohealth.2013.06.012

Liu, X. (2003). Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *Journal of Educational and Behavioral Statistics, 28*(3), 231-248. https://doi.org/10.3102/10769986028003231

May, H., Gray, A., Sirinides, P., Goldsworthy, H., Armijo, M., Sam, C., Gillespie, J. N., & Tognatta, N. (2015). Year one results from the multisite randomized evaluation of the i3 scale-up of Reading Recovery. American Educational Research Journal, 52(3), 547–581. https://doi.org/10.3102/0002831214565788

Nam, J. M. (1973). Optimum sample sizes for the comparison of the control and treatment. *Biometrics, 29*, 101-108. https://doi.org/10.2307/2529679

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*(2), 173-185. http://doi.org/10.1037/1082-989X.2.2.173

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. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods, 5*(2), 199. https://doi.org/10.1037/1082-989X.5.2.199

Rhoads, C. H., & Dye, C. (2016). Optimal design for two-level random assignment and regression discontinuity studies. *The Journal of Experimental Education, 84*(3), 421-448. https://doi.org/10.1080/00220973.2015.1048846

Shen, Z., & Kelcey, B. (2020). Optimal sample allocation under unequal costs in cluster-randomized trials. *Journal of Educational and Behavioral Statistics, 45*(4), 446–474. https://doi.org/10.3102/1076998620912418

Shen, Z., & Kelcey, B. (2022a). Optimal sample allocation in multisite randomized trials. *The Journal of Experimental Education, 90*(3), 693-711. https://doi.org/10.1080/00220973.2020.1830361

Shen, Z., & Kelcey, B. (2022b). Optimal sample allocation for three-level multisite cluster-randomized trials. *Journal of Research on Educational Effectiveness, 15*(1), 130-150. https://doi.org/10.1080/19345747.2021.1953200

Shen, Z., & Kelcey, B. (2026). *odr: Optimal design and statistical power for experimental studies investigating main, mediation, and moderation effects* (Version 1.8.3) [Software]. https://cran.r-project.org/package=odr

Shen, Z., Leite, W., Zhang, H., Quan, J., & Kuang, H. (2025). Using ant colony optimization to identify optimal sample allocations in cluster-randomized trials. *The Journal of Experimental Education, 93*(1), 167-185.https://doi.org/10.1080/00220973.2024.2306392

Shen, Z., Li, W., & Leite, W. (in press). Statistical power and optimal design for randomized controlled trials investigating mediation effects. *Psychological Methods*. https://doi.org/10.1037/met0000698

Socha, K., & Dorigo, M. (2008). Ant colony optimization for continuous domains. *European Journal of Operational Research, 185*(3), 1155-1173. https://doi.org/10.1016/j.ejor.2006.06.046

Springer, M. G., Ballou, D., Hamilton, L., Le, V. N., Lockwood, J. R., McCaffrey, D. F., ... & Stecher, B. M. (2011). Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching (POINT). *Society for Research on Educational Effectiveness.* https://files.eric.ed.gov/fulltext/ED518378.pdf

Spybrook, J., Kelcey, B., & Dong, N. (2016). Power for detecting treatment by moderator effects in two-and three-level cluster randomized trials. *Journal of Educational and Behavioral Statistics, 41*(6), 605-627. https://doi.org/10.3102%2F1076998616655442

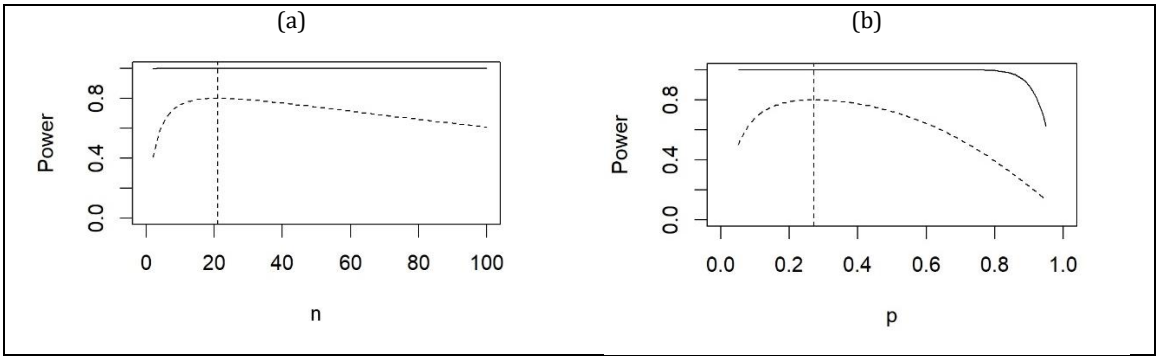

Figure 1. Statistical Power under a fixed budget for main (black curves) and moderation (dashed curves) effects with d = 0.20 and γ = 0.15

Appendix. R Code for the Illustration

```
library("odr")
# OD for main----
myod <- od.2(icc = .2, r12 = .55, r22 = .75,
                c1 = 10, c1t = 10, c2 = 100, c2t = 2000)
myod$out # n  = 21, p = .27
# OD for moderation and main using ACO algorithm----
myod1 <- od.2.mod(d = 0.2, gamma = .15, Q = .5,
                        icc = .2, r12 = .55, r22 = .75,
                        c1 = 10, c1t = 10, c2 = 100, c2t = 2000)
myod1$out # n  = 21, p = .27, m = 392791

# Required sample size for moderation----
J.mod <- power.2.mod(expr = myod1, gamma = .15, q.mod = 1, power.mod = .80)
J.mod$out  # m = 392797, J  = 477

# Balanced design (with constrained p = .50)----
power.2.mod(expr = myod1, constraint = list(p = 0.5),
                m = myod1$out$m, gamma = 0.15, q.mod = 1)$out$power.mod
# power: 72%

# Required sample size and budget for the balanced design
J.mod.balanced <- power.2.mod(expr = myod1, gamma = .15, q = 1, power.mod = .80,
                                constraint = list(p = .5))
J.mod.balanced$out  # m = 474641, J  = 377

# Additional budget
474641/392766
# 21% more budget to achieve 80% power

# The minimum detectable moderation effect under the same budget
power.2.mod(expr = myod1,  constraint = list(p = 0.5), q.mod = 1,
      power.mod = .8, m = myod1$out$m)$out
# gamma = .165

# Plot the power curves
d = 0.2; gamma = .15
m <- m.od <- myod1$out$m
p <- p.od <-myod1$out$p
n <- n.od <-myod1$out$n

## n plot ====
p = p.od
nrange = c(2:100)
plot.main <- plot.mod <- NULL
for (n in nrange) {
  powr.main <- power.2(expr = myod, constraint = list(n = n),
                        m = m, d = d, q = 1)$out$power
  plot.main <- c(plot.main, powr.main)
  powr.mod <- power.2.mod(expr = myod1, constraint = list(n = n), m = m, gamma = gamma,
                        q.mod = 1)$out$power.mod
  plot.mod <- c(plot.mod, powr.mod)
}
```

```
nlim = c(0, 100)
plim = c(0, 1)
varlim = c(0.55, .9)
nlab = "n"
plab = "p"
varlab = "Power"

graphics::plot(nrange, plot.main, type = "l", lty = 1, xlim = nlim, ylim = c(0, 1),  xlab = nlab, ylab = varlab,
                    main = "", col = "black"  #axes = FALSE)
graphics::abline(v = n.od, lty = 2, col = "Black")
lines(nrange, plot.mod, col = "black", type = "l", lty = 2)

## p plot ====
n = n.od
plot.main <- plot.mod <- NULL
prange <- seq(plim[1] + 0.05, plim[2] - 0.05, by = 0.01)
for (p in prange) {
  powr.main <- power.2(expr = myod, constraint = list(p = p), m = m, d = d, q = 1)$out$power
  plot.main <- c(plot.main, powr.main)
  powr.mod <- power.2.mod(expr = myod1, constraint = list(p = p), m = m, gamma = gamma,
                                  q.mod = 1)$out$power.mod
  plot.mod <- c(plot.mod, powr.mod)
}

graphics::plot(prange, plot.main, type = "l", lty = 1, xlim = plim, ylim = c(0, 1), xlab = plab, ylab = varlab,
                    main = "", col = "black"  #axes = FALSE)
graphics::abline(v = p.od, lty = 2, col = "black")
lines(prange, plot.mod, col = "black", type = "l", lty = 2)
```