Title: Design Parameters and Effect Sizes for Designing Two-Level and Three-Level Cluster Randomized Trials to Detect Moderator Effects

Authors: Qi Zhang and Jessaca Spybrook, Western Michigan University

Background: Cluster Randomized Trials (CRTs) are frequently conducted to assess the effectiveness of educational interventions. In a two-level CRT designed to assess the effect of an intervention on student achievement, schools are randomly assigned to treatment and comparison conditions and students are nested within schools. A three-level CRT could be the natural extension of a two-level CRT where students are nested within teachers, teachers are within schools, and random assignment is at the school level. A-priori power analyses are conducted when planning CRTs to determine the capacity of studies to detect meaningful treatment effects. In recent years funding agencies (i.e. IES) are recommending studies to explore the statistical power to answer questions of "for whom" and "under what conditions" interventions work (U.S. DOE, 2018) as an effort to encourage studies to go beyond answering the "what works" question. These context-relevant questions can be answered with planned moderator analyses.

Design parameters such as the intraclass correlation (ICC) and proportion of variance explained by covariates (R^2), as well as effect sizes are essential elements for improving the precision of power analyses. Though there is plenty of literature on the empirical benchmark of main treatment effect sizes, empirical evidence related to the magnitude of effect sizes associated with moderators are rare. In addition, design parameters including the ICCs and R^2 coefficients for planning three-level CRTs are less available compared to two-level CRTs.

Purpose: This study examines three sets of parameters necessary for conducting power analyses for three-level and two-level CRTs that aim to detect the main treatment effect and moderator effects: 1) effect sizes of student-, teacher-, and school-level moderator effects, 2) school-level ICCs (ICC_3 in three-level model, ICC_2 in two-level model) and teacher level ICCs (ICC_2), and 3) R^2 coefficients associated with student-, teacher-, school-level covariates.

Method: In this study, we focus on CRTs that randomly assign schools to treatment conditions and those aim to improve student achievement. Analyses were carried out with datasets of completed impact evaluations. Table 1 lists the datasets of two impact evaluations that we used in our preliminary analysis.

[Insert Table 1 here]

We estimated ICCs, R^2 coefficients, main treatment effect sizes, and moderator effect sizes using three-level and two-level hierarchical linear models (HLM). The three-level model accounts for designs where students are nested within teachers and teachers are within schools. The two-level model accounts for designs where students are nested within schools. The three-level datasets were also used in two-level analyses ignoring the teacher level. This allows us to compare design parameters and effect sizes of three-level and two-level analyses within the same dataset. We used the unconditional model including the treatment indicator to estimate ICC_3 and ICC_2, since the datasets are from impact studies. To estimate the R^2 values, we used fully conditional models that include student-, teacher-, and school-level covariates. Table 1 outlines the covariates we included for each dataset. We aggregated student level (level-1) covariates to the teacher-level (level-2) and school-level (level-3) and teacher covariates to school level explain variations at higher levels.

Table 1 also outlines the moderators we examined at each of the three levels. For the purpose of interpreting subgroup differential treatment effects, we dichotomized continuous or categorical moderators. For instance, we dichotomized years of teaching experience into a dummy variable with 1 indicating experienced teachers (more than 4 years of experience) and 0 indicating new teachers. We also dichotomized school-level moderators based on their population means. For example, schools with greater than 53% (Common Core of Data population average) of FRL students were coded as “low SES schools” and those with less than 53% were “high SES schools”. For consistency, we used the dichotomized variables as covariates in the fully conditional models. Table 2 demonstrates the 2-level fully conditional models for the moderator analysis.

[Insert Table 2 here]

Results: Our preliminary results suggest that there is a high proportion of variance (ICC_2) at the teacher level (Table 3) in 3-level models. For the math and reading outcomes, the ICC_2 and ICC_3 of 3-level models are very similar. As the models change from 3-level to 2-level, we see that 25% of the teacher level variance moves to the school level and 75% goes to the student level except for the CMP2 math outcome. These results are consistent with the patterns observed by Zu et al. (2012) for upper elementary and middle school student outcomes.

[Insert Table 3 here]

Table 4 and Table 5 show the R^2 coefficients for 3-level and 2-level models. As expected the R^2 coefficients associated with student pretest were the highest, which suggests that pretest explained significant amount of variations at all levels. The teacher covariate sets explained very little variations at the teacher level. School-level covariates sets overall explained significant amount of variations at the school level. The R^2 coefficients are similar when comparing estimates of three-level and two-level models, which is consistent with the findings from Zu and colleagues (2012).

[Insert Table 4 here]

[Insert Table 5 here]

Table 6 shows the main treatment effect size and student-level moderator effect sizes for three-level and two-level analyses of each outcome. The absolute magnitudes of student moderator effect sizes are mostly similar to or smaller than the main treatment effect size. The absolute magnitudes of most teacher moderator effect sizes for the three AMSTI outcomes are also similar to or smaller than their respective main treatment effect sizes (Table 7). In contrary, the absolute magnitudes of most school level moderator effects are larger compared to their respective main treatment effects (Table 8).

[Insert Table 6 here]

[Insert Table 7 here]

[Insert Table 8 here]

Conclusion: Our preliminary results suggest that using the magnitude of the main treatment effect sizes to estimate power for moderator effect is not always precise. For example, the school-level moderator effects appeared to be higher than for the main treatment effect. Standard practice assumes the moderator effect to be the same or smaller than the main treatment effect, making it difficult to power a CRT for a same-level moderator effect. If our finding holds across datasets, it may have important implications for powering a study to detect cluster-level moderators. We will include the results of more datasets and elaborate on their implications in our presentation.

References

Zhu, P.; Jacob, R.; Bloom, H.; & Xu, Z. (2012). Designing and analyzing studies that randomized schools to estimate intervention effects on student achievement outcomes without classroom-level information. *Educational Evaluation and Policy Analysis*, 34(1), 45-68.

U.S. Department of Education. (2018). Request for Applications: Education Research Grants. Washington, DC: Institute of Education Sciences.

Table 1. Description of the Datasets.

Dataset/Study	Outcomes	Student Covariates	Teacher Covariates	School Covariates	Grade Level	Locations	Schools	Teachers	Students
Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMST)	SAT 10 Math Problem Solving Subscale; SAT 10 Reading; SAT 10 Science	Pretest, gender*, minority status*, FRL status*, ELL status*, disability status*, missing value indicators	Dummy for teaching experience > 4 years*, dummy for teaching subject > 4 years*, degree in the subject*, missing value indicators	Urban school indicator*	4-8	Alabama	80	470	2,0730
Effects of the Connected Mathematics Project 2 (CMP2) on Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region	TerraNova CAT2	--	Gender*, minority status*, Dummy for teaching experience > 4 years*, degree in the subject*, advanced degree*, missing value indicators	Urban school indicator*, high minority school indicator*, low SES school indicator*, school gender proportion	6-8	Delaware, Maryland, New Jersey, Pennsylvania, Washington, DC	60	130	5,670

* Covariates used a moderators in estimating moderator effect sizes.
SOURCE: U.S. Department of Education, National Center for Educational Statistics, Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMST), 2012; U.S. Department of Education, National Center for Educational Statistics, Effects of the Connected Mathematics Project 2 (CMP2) on Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region, 2012.

Table 2. 2-level fully conditional models with moderators at level-1 or level-2.

	Model	Parameter
Moderator at level-1	Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j}M_{ij} + \sum_q \beta_{qj}X_{qij} + r_{ij}$	Y_{ij} is the outcome for student i ∈ $(1, 2, 3, \dots, n)$ in school j ∈ $(1, 2, 3, \dots, J)$. β_{0j} is the average student achievement in school J. M_{ij} is the student level moderator and β_{1j} is the coefficient for the moderator. X_{qij} is the q^{th} student-level covariate q ∈ $(2, 3, 4 \dots, Q)$ for student i in school j, β_{qj} is the coefficient for that covariate. r_{ij} is the random error term, conditioned on the moderator and the Q covariates associated with each student, which has a normal distribution with a mean of 0 and homogenous variance σ^2. σ^2 also represents the within school variance after controlling for Q covariates and the moderator.
	Level 2: $\beta_{0j} = \gamma_{00} + \delta_0 T_j + \sum_s \gamma_{0s}W_{sj} + \mu_{0j}$ $\beta_{1j} = \gamma_{10} + \delta_1 T_j + \sum_s \gamma_{1s}W_{sj}$ $\beta_{qj} = \gamma_{qo,}\ \forall q \in (1, 2, 3, \dots, Q)$	γ_{00} is the grand mean of student achievement. δ_0 is the treatment effect and T_j is the indicator of random assignment of schools to conditions. δ_1 is the differential treatment effect associated with the moderator. W_{sj} is the s$^{\text{th}}$ school-level covariate s ∈ $(1, 2, 3, \dots, S)$ and γ_{0s} is the coefficient for that covariate. μ_{0j} is the error term

		associated with schools, conditioned on the S covariates and the treatment indicator. τ is the between school variance after controlling for S covariates and the treatment indicator. γ_{qo} is the fixed effect associated with q[th] student-level covariate.
Moderator at level-2	Level 1: $Y_{ij} = \beta_{0j} + \sum_q \beta_{qj} X_{qij} + r_{ij}$	Y_{ij} is the outcome for student i $\in (1, 2, 3, \ldots, n)$ in school j $\in (1, 2, 3, \ldots, J)$. β_{0j} is the average student achievement in school J. X_{qij} is the q^{th} student-level covariate q $\in (1, 2, 3, \ldots, Q)$ for student i in school j, β_{qj} is the coefficient for that covariate. r_{ij} is the random error term, conditioned on the moderator and the Q covariates associated with each student, which has a normal distribution with a mean of 0 and homogenous variance σ^2. σ^2 also represents the within school variance after controlling for Q covariates.
	Level 2: $\beta_{0j} = \gamma_{00} + \delta_0 T_j + \gamma_{01} M_j + \delta_1 T_j M_j + \sum_s \gamma_{0s} W_{sj} + \mu_{0j}$ $\beta_{qj} = \gamma_{qo,} \forall q \in (1, 2, 3, \ldots, Q)$	γ_{00} is the grand mean of student achievement. δ_j is the treatment effect and T_j is the indicator of random assignment of schools to conditions. M_j is the school level moderator and γ_{01} is the effect associated with the moderator. $T_j M_j$ is the interaction term of the treatment and the moderator and δ_1 is the differential treatment effect associated with the moderator. W_{sj} is the s[th] school-level covariate s $\in (3, 4, 5 \ldots, S)$ and γ_{0s} is the coefficient for that covariate. μ_{0j} is the error term associated with schools, conditioned on the S covariates and the moderator. τ is the between school variance after controlling for S covariates and the moderator. γ_{qo} is the fixed effect associated with q[th] student-level covariate.

NOTE: The 3-level fully conditional models are simply the natural extensions of 2-level models.
SOURCE: U.S. Department of Education, National Center for Educational Statistics, Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMST), 2012; U.S. Department of Education, National Center for Educational Statistics, Effects of the Connected Mathematics Project 2 (CMP2) on Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region, 2012.

Table 3. ICCs.

Study	Student Achievement Outcome	3-Level Model		2-Level Model
		ICC_3	ICC_2	ICC_2
AMSTI	Mathematics	0.15	0.17	0.19
	Reading	0.16	0.14	0.19
	Science	0.16	0.07	0.18
CMP2	Mathematics	0.12	0.13	0.24

NOTE: ICC_2 in a 3-level model represents proportion of the outcome variation at the teacher level and ICC_2 in a 2-level model represents proportion of the outcome variation at the school level.
SOURCE: U.S. Department of Education, National Center for Educational Statistics, Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMST), 2012; U.S. Department of Education, National Center for Educational Statistics, Effects of the Connected Mathematics Project 2 (CMP2) on Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region, 2012.

Table 4. R^2s for three-level models.

Data	Student Achievement Outcome	Model 1.0 Student pretest			Model 1.1 Student Demographics			Model 1.2 Model 1.0 + Model 1.1			Model 2 Teacher Demographics			Model 3 School Demographics			Model 4 Model 1.2 + Model 2 + Model 3		
		R^2_{L1}	R^2_{L2}	R^2_{L3}	R^2_{L1}	R^2_{L2}	R^2_{L3}	R^2_{L1}	R^2_{L2}	R^2_{L3}	R^2_{L1}	R^2_{L2}	R^2_{L3}	R^2_{L1}	R^2_{L2}	R^2_{L3}	R^2_{L1}	R^2_{L2}	R^2_{L3}
AMSTI	Mathematics	0.37	0.81	0.94	0.09	0.21	0.42	0.38	0.81	0.95	0	0.01	0.13	0	0	0.03	0.38	0.81	0.95
	Reading	0.61	0.81	0.96	0.18	0.2d6	0.53	0.62	0.81	0.98	0	0	0.11	0	0	0.03	0.62	0.81	0.99
	Science	0.46	0	0.29	0.16	0.27	0.76	0.48	0.13	0.76	0	0	0	0	0	0	0.49	0.10	0.76
CMP2	Mathematic	0.57	0.92	0.83	--	--	--	--	--	--	0	0	0.40	0	0	0.80	0.57	0.92	0.80

NOTE: L1 is the student level, L2 is the teacher level, and L3 is the school level.
SOURCE: U.S. Department of Education, National Center for Educational Statistics, Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMST), 2012; U.S. Department of Education, National Center for Educational Statistics, Effects of the Connected Mathematics Project 2 (CMP2) on Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region, 2012.

Table 5. R^2s for two-level models.

Data	Student Achievement Outcome	Model 1.0 Student pretest		Model 1.1 Student Demographics		Model 1.2 Model 1.0 + Model 1.1		Model 2 School Demographics		Model 3 Model 1.2 + Model 2	
		R^2_{L1}	R^2_{L2}	R^2_{L1}	R^2_{L2}	R^2_{L1}	R^2_{L2}	R^2_{L1}	R^2_{L2}	R^2_{L1}	R^2_{L2}
AMSTI	Mathematics	0.42	0.90	0.09	0.41	0.43	0.91	0	0.05	0.43	0.91
	Reading	0.63	0.94	0.17	0.51	0.64	0.95	0	0.04	0.64	0.95
	Science	0.38	0.29	0.15	0.70	0.41	0.71	0	0	0.41	0.71
CMP2	Mathematic	0.58	0.88	--	--	--	--	0	0.53	0.58	0.88

NOTE: L1 is the student level and L2 is the school level.
SOURCE: U.S. Department of Education, National Center for Educational Statistics, Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMST), 2012; U.S. Department of Education, National Center for Educational Statistics, Effects of the Connected Mathematics Project 2 (CMP2) on Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region, 2012.

Table 6. Main treatment effect sizes and effect sizes of L-1 (student level) moderators.

Data	Student Achievement Outcomes	Main Treatment Effect Size (ES)		Level 1 Moderator Effect Size (DES)									
				Gender		FRL status		Minority status		ELL status		Disability status	
		3L	2L	3L	2L	3L	2L	3L	2L	3L	2L	3L	2L
AMSTI	Mathematic	0.03	0.03	-0.01	0	-0.02	-0.02	-0.05	-0.04	-0.08	-0.10	-0.02	-0.01
	Reading	0.05	0.06	-0.01	0	-0.02	-0.02	-0.06	-0.06	0.09	0.09	-0.04	-0.02
	Science	-0.03	-0.02	0.01	0.01	-0.07	-0.06	-0.05	-0.05	-0.01	0.03	0.03	-0.01
CMP2	Mathematics	-0.01	-0.02	--	--	--	--	--	--	--	--	--	--

SOURCE: U.S. Department of Education, National Center for Educational Statistics, Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMST), 2012; U.S. Department of Education, National Center for Educational Statistics, Effects of the Connected Mathematics Project 2 (CMP2) on Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region, 2012.

Table 7. Main treatment effect sizes and effect sizes of L-2 (teacher level) moderators.

Data	Student Achievement Outcomes	Main Treatment Effect Size (ES)	Level 2 Moderator Effect size (DES) for 3-Level Model						
			Tch. gender	Tch. minority status	Yrs of teaching experience	Yrs of teaching subject	Earned advanced degree	Earned degree in subject	Teaching certificate in subject
AMSTI	Mathematic	0.03	--	--	0.01	0.03	--	0.01	--
	Reading	0.05	--	--	0	0.03	--	0.02	--
	Science	-0.03	--	--	0.14	0.03	--	-0.02	--
CMP2	Mathematics	-0.01	0.06	0.24	-0.10	--	0.01	0.03	--

SOURCE: U.S. Department of Education, National Center for Educational Statistics, Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMST), 2012; U.S. Department of Education, National Center for Educational Statistics, Effects of the Connected Mathematics Project 2 (CMP2) on Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region, 2012.

Table 8. Main treatment effect sizes and effect sizes of L-3 (school level) moderators.

Data	Student Achievement Outcomes	Main Treatment Effect Size (ES)		Level 1 Moderator Effect Size (DES)					
				Urban school		High minority school		Low SES school	
		3L	2L	3L	2L	3L	2L	3L	2L
AMSTI	Mathematic	0.03	0.03	0.18	0.12	--	--	--	--
	Reading	0.05	0.06	0.11	0.09	--	--	--	--
	Science	-0.03	-0.02	-0.41	-0.41	--	--	--	--
CMP2	Mathematics	-0.01	-0.02	0.18	0.18	0.02	0.02	0.22	0.17

SOURCE: U.S. Department of Education, National Center for Educational Statistics, Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMST), 2012; U.S. Department of Education, National Center for Educational Statistics, Effects of the Connected Mathematics Project 2 (CMP2) on Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region, 2012.