Workshop

Workshop on Cluster Randomized Trials

Exercises

Martin Brunner, Sophie E. Stallasch & Larry Hedges

Potsdam

June 9 and 10, 2022

MULTIDES – Multilevel Design Parameters and Effect Size Benchmarks

Project Goals: Using all available German (longitudinal) large-scale assessments to ...

- Develop design parameters for planning cluster-randomized trials
  - Preschool
    - cognitive and socio-emotional outcomes
  - Elementary and Secondary School
    - cognitive outcomes
    - socio-emotional outcomes
- Evaluate covariate characteristics for planning cluster-randomized trials
  - validity degradation over time
  - content alignment between pretest and outcome
  - incremental validity of educational characteristics over pretest
- Provide effect size benchmarks for assessing educational intervention and policy effects
  - learning gains
  - performance gaps between
    - sociodemographic student groups
    - weak and average schools

Project members and funding

- Martin Brunner, Sophie Stallasch, Oliver Lüdtke, Cordula Artelt, Larry Hedges

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Scenario 1

How many schools are required for a two-level cluster-randomized trial?

Exercises: Scenario 1

Research Team 1 would like to conduct a two-level CRT on the effectiveness of a school-wide intervention to improve 4th graders' mathematical achievement in Germany. To improve statistical power, the research team administers a math pretest to each participating student.

How many schools are required for this two-level cluster-randomized trial (CRT)?

We need to specify the following set of parameters

- one- or two-sided testing
- significance level α
- desired power 1 – β
- reasonable value for the minimally detectable effect size *MDES*
- share of schools in the control and intervention condition
- (average) number of students sampled per school
- expected number of missing data / available cases for the analyses
    - school level
    - student level
- an estimate of the intraclass correlation of the outcome
- number of covariates at the school level
- amount of explained variance by covariates at the
- school level
- student level

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 1

> "Two sided tests should be used unless there is a very good reason for doing otherwise. If one sided tests are to be used the direction of the test must be specified in advance. One sided tests should never be used simply as a device to make a conventionally non-significant difference significant. [...] In general a one sided test is appropriate when a large difference in one direction would lead to the same action as no difference at all. Expectation of a difference in a particular direction is not adequate justification. In medicine, things do not always work out as expected, and researchers may be surprised by their results."

Bland, J. M., & Altman, D. G. (1994). Statistics Notes: One and two sided tests of significance. *BMJ*, *309*(6949), 248.

How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- One- or two-sided testing:
    - typically set to two-sided testing

Exercises: Scenario 1

> The risk of mistakenly rejecting the null hypothesis (H_0) and thus of committing a Type I error, α, represents a policy: the maximum risk attending such a rejection. Unless otherwise stated (and it rarely is), it is taken to equal .05 (part of the Fisherian legacy; Cohen, 1990). Other values may of course be selected. For example, in studies testing sev-

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- significance level α
    - typically set to α = .05

Exercises: Scenario 1

In this treatment, the only specification for power is .80 (so $\beta = .20$), a convention proposed for general use. (SPABS provides for 11 levels of power in most of its N tables.) A materially smaller value than .80 would incur too great a risk of a Type II error. A materially larger value would result in a demand for N that is likely to exceed the investigator's resources. Taken with the conventional $\alpha = .05$, power of .80 results in a $\beta:\alpha$ ratio of 4:1 (.20 to .05) of the two kinds of risks. (See SPABS, pp. 53–56.)

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Because statistical power is the probability of making the correct decision when a treatment effect actually exists, high statistical power is desirable. However, in any given design, higher power is achieved only with larger sample sizes. Obtaining larger sample sizes typically requires the commitment of more resources (not only costs but also research staff and burden on schools). Therefore, the benefits associated with higher power must be weighed against the commitment of resources required to achieve these benefits. Technical means alone cannot resolve this cost/benefit judgment. Normative statistical practice seems to be that power of 0.8 or above is considered acceptable, but there is no reason to think that this figure is always appropriate.

Hedges, L. V., & Rhoads, C. (2010). Statistical Power Analysis in Education Research. NCSER 2010-3006. In *National Center for Special Education Research*. National Center for Special Education Research. http://files.eric.ed.gov/fulltext/ED509387.pdf

How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- desired power $1 - \beta$
  - typically set to $1 - \beta = .80$

Exercises: Scenario 1

Table 1
Empirical Distributions of Effect Sizes From Randomized Control Trials of Education Interventions With Standardized Achievement Outcomes

		Subject		Sample Size					Scope of Test		
	Overall	Math	Reading	≤100	101–250	251–500	501–2,000	>2,000	Broad	Narrow	DoE Studies
Mean	0.16	0.11	0.17	0.30	0.16	0.16	0.10	0.05	0.14	0.25	0.03
Standard deviation	0.28	0.22	0.29	0.41	0.29	0.22	0.15	0.11	0.24	0.44	0.16
Mean (weighted)	0.04	0.03	0.05	0.29	0.15	0.16	0.10	0.02	0.04	0.08	0.02
P1	−0.38	−0.34	−0.38	−0.56	−0.42	−0.29	−0.23	−0.22	−0.38	−0.78	−0.38
P10	−0.08	−0.08	−0.08	−0.10	−0.14	−0.07	−0.05	−0.06	−0.08	−0.12	−0.14
P20	−0.01	−0.03	−0.01	0.02	−0.04	0.00	−0.01	−0.03	−0.03	0.00	−0.07
P30	0.02	0.01	0.03	0.10	0.02	0.06	0.03	0.00	0.02	0.05	−0.04
P40	0.06	0.04	0.08	0.16	0.07	0.10	0.06	0.01	0.06	0.11	−0.01
P50	0.10	0.07	0.12	0.24	0.12	0.15	0.09	0.03	0.10	0.17	0.03
P60	0.15	0.11	0.17	0.32	0.17	0.18	0.12	0.05	0.14	0.22	0.05
P70	0.21	0.16	0.23	0.43	0.25	0.22	0.15	0.08	0.20	0.34	0.09
P80	0.30	0.22	0.33	0.55	0.35	0.29	0.19	0.11	0.29	0.47	0.14
P90	0.47	0.37	0.50	0.77	0.49	0.40	0.27	0.17	0.43	0.70	0.23
P99	1.08	0.91	1.14	1.58	0.93	0.91	0.61	0.48	0.93	2.12	0.50
k (number of effect sizes)	1,942	588	1,260	408	452	328	395	327	1,352	243	139
n (number of studies)	747	314	495	202	169	173	181	124	527	91	49

Note. A majority of the standardized achievement outcomes (95%) are based on math and English language art test scores, with the remaining based on science, social studies, or general achievement. Weights are based on sample size for weighted mean estimates. For details about data sources, see Appendix A, available on the journal website. DoE = U.S. Department of Education.

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- set a reasonable value for the minimally detectable effect size *MDES:* ***MDES = .15***
  - **accumulated body of research evidence**

Exercises: Scenario 1

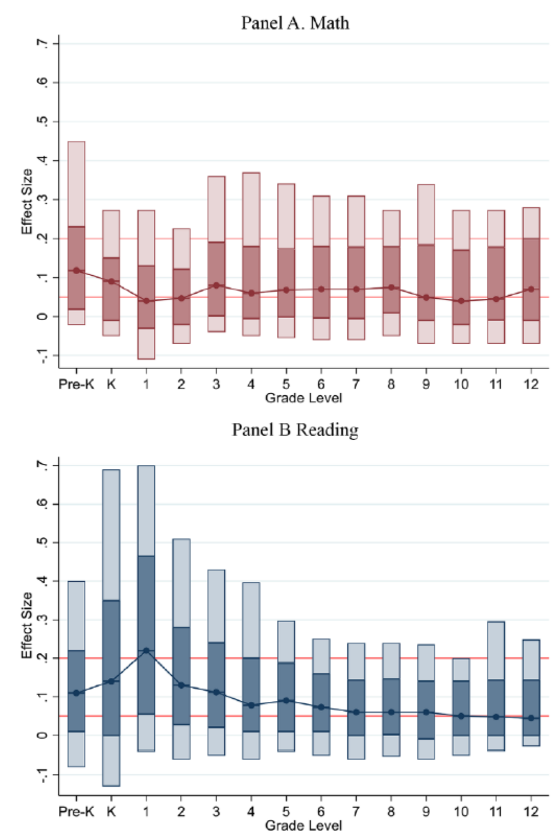


FIGURE 1. *The distribution of effect sizes from randomized control trials of education interventions with student achievement outcomes by subject and grade level.*
Note. Vertical bars represent 10th to 90th percentile ranges with darker shaded interquartile ranges (25th–75th percentiles). Connected line dots illustrate changes in median effect sizes across grade distributions. Red horizontal lines indicate proposed effect size benchmarks.

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- set a reasonable value for the minimally detectable effect size *MDES:* ***MDES = .15***
  - **accumulated body of research evidence**

Exercises: Scenario 1

“The chosen MDES should be considered to be policy relevant.“

- Compare the desired MDES with differences between weak and average schools

  At elementary schools in Germany the standardized mean difference between weak (i.e., at the 10th quantile) and average schools (i.e., at the 50th quantile) in mathematics was 0.24. Thus an intervention effect of d = .15 would close about 63% (.15/.24) of the achievement gap between weak and average schools (Brunner et al., 2022).
- Compare the desired MDES with improving an important criterion

  An intervention effect of d = 0.15 would reduce the number of elementary school students who do not achieve the minimal standard in mathematics from 136 out of 1,000 students in the control condition to about 106 out of 1,000 students in the intervention condition. In other words: Out of 1,000 students, the intervention helps about 30 students successfully continue their school career in secondary level who otherwise would be likely to fail due to their insufficient level of mathematics proficiency (Brunner et al., 2022)

How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- set a reasonable value for the minimally detectable effect size *MDES*: ***MDES = .15***
  - **political perspective**

Brunner, M., Keller, U., Wenger, M., Fischbach, A., & Lüdtke, O. (2018). Between-school variation in students’ achievement, motivation, affect, and learning strategies: Results from 81 countries for planning group-randomized trials in education. *Journal of Research on Educational Effectiveness*, *11*(3), 452–478. https://doi.org/10.1080/19345747.2017.1375584
Brunner, M., Stallasch, S. E., & Lüdtke, O. (2022). *Empirical Benchmarks to Interpret Intervention Effects on Student Achievement in Elementary and Secondary School: Meta-Analytic Results from Germany*. EdArXiv. https://doi.org/10.35542/osf.io/39gbq

Exercises: Scenario 1

"the (monetary) benefits associated with the proposed effect of the educational intervention should outweigh the costs of the intervention itself (see Schochet, 2008, p. 66, for an example)."

"We find that each new high school graduate would yield a public benefit of $209,000 in higher government revenues and lower government spending for an overall investment of $82,000, divided between the costs of powerful educational interventions and additional years of school attendance leading to graduation. The net economic benefit to the public purse is therefore $127,000 per student and the benefits are 2.5 times greater than the costs."

Levin, H., Belfield, C., Muennig, P., & Rouse, C. (2006). The Costs and Benefits of an Excellent Education for America's Children.

*With a grain of salt: Out of 1,000 students, the intervention helps about 30 students successfully continue their school career in secondary level. Thus, the benefits may amount to 30 * 127.000 $ = 3.8 million $ per 1000 students.*

How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- set a reasonable value for the minimally detectable effect size *MDES*: ***MDES = .15***
  - **economic perspective**

Brunner, M., Keller, U., Wenger, M., Fischbach, A., & Lüdtke, O. (2018). Between-school variation in students' achievement, motivation, affect, and learning strategies: Results from 81 countries for planning group-randomized trials in education. *Journal of Research on Educational Effectiveness*, *11*(3), 452–478. https://doi.org/10.1080/19345747.2017.1375584

Exercises: Scenario 1

How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- Share of school in the control and intervention condition
  - set to 50% (i.e., equal number of students in each condition)
- (average) number of students sampled per school
  - set to n = 40 (i.e., most elementary schools in Germany have at least 40 students in a certain grade level)
- available cases for the analyses: sample retention rate
  - school level: set to 100% (which is too optimistic)
  - student level: set to 100% (which is too optimistic)

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 1

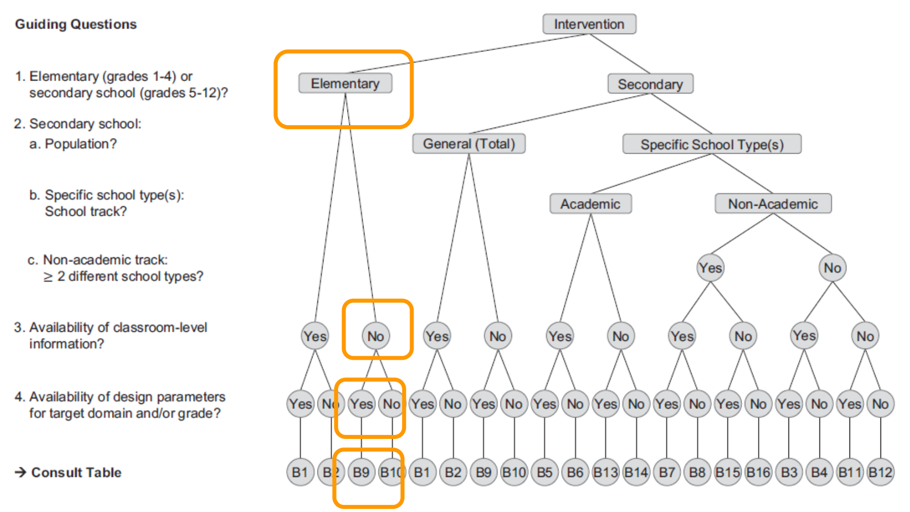


Figure 4. Flow chart to guide the choice of design parameters as a function of key characteristics of the target intervention. *Note.* Tables B1–B16 can be retrieved from Supplemental Online Material B. A comprehensive overview of the achievement measures analyzed in the present study is given in Table A5 in the Supplemental Online Material A. The Supplemental Online Materials are available on the Open Science Framework (https://osf.io/2w8nt).

How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- an estimate of the intraclass correlation of the outcome
- number of covariates at the school level
- amount of explained variance by covariates
    - school level
    - student level

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 1

Domain	Grade	Study	Model	a. Model Set 1 Intercept-Only Model		b. Model Set 2 Pretest Covariate(s) Model				c. Model Set 3 Sociodemographic[a] Covariates Model				d. Model Set 4 Pretest and Sociodemographic[a] Covariates Model				N		Wave	Name of Outcome Variable	
				ρ_{L2}	SE	R^2_{L1}	SE	R^2_{L2}	SE	R^2_{L1}	SE	R^2_{L2}	SE	R^2_{L1}	SE	R^2_{L2}	SE	L1	L3		(R / Mplus)	(Orig. Data)
Mathematics	1	NEPS-SC2	2l	.17	.01	.40	.02	.22	.09	.11	.01	.41	.04	.42	.02	.45	.05	6731	374	3	mag1_sc1	mag1_sc1
	2	NEPS-SC2	2l	.12	.01	.45	.01	.55	.04	.12	.01	.54	.04	.47	.01	.68	.03	6319	362	4	mag2_sc1	mag2_sc1
	4	NEPS-SC2	2l	.12	.01	.40	.01	.64	.03	.11	.01	.65	.04	.43	.01	.76	.03	5419	349	6	mag4_sc1	mag4_sc1
	5	NEPS-SC3	2l	.42	.03	-	-	-	-	.06	.01	.83	.03	-	-	-	-	5380	225	1	mag5_sc1	mag5_sc1
	7	NEPS-SC3	2l	.42	.02	.33	.01	.95	.01	.06	.01	.88	.04	.35	.01	.97	.01	6279	266	3	mag7_sc1	mag7_sc1
	9	NEPS-SC3	2l	.42	.03	.34	.01	.94	.01	.07	.01	.86	.04	.35	.01	.96	.01	4651	239	5	mag9_sc1	mag9_sc1
	9	NEPS-SC4	2l	.42	.02	-	-	-	-	.08	.00	.78	.02	-	-	-	-	14640	518	1	mag9_sc1	mag9_sc1
	9	PISA-I+	2l	.38	.03	-	-	-	-	.05	.01	.68	.04	-	-	-	-	6020	152	1	m2t1_gt	wle_m2t1_gt
	9	All[b]	2l	.42	.01	.34	.01	.94	.01	.07	.00	.78	.02	.35	.01	.96	.01	25311	909			
	10	PISA-I+	2l	.38	.03	.34	.01	.98	.00	.04	.01	.68	.04	.35	.01	.98	.00	6020	152	2	m2t2_gt	wle_m2t2_gt
Science	1	NEPS-SC2	2l	.19	.01	.31	.02	.44	.11	.11	.01	.54	.04	.35	.02	.67	.08	6731	374	3	scg1_sc1	scg1_sc1
	3	NEPS-SC2	2l	.11	.01	.32	.01	.43	.04	.11	.01	.62	.04	.35	.01	.66	.04	5554	354	5	scg3_sc1	scg3_sc1
	6	NEPS-SC3	2l	.32	.02	-	-	-	-	.06	.01	.90	.02	-	-	-	-	5026	211	2	scg6_sc1	scg6_sc1
	9	NEPS-SC3	2l	.34	.02	.31	.01	.96	.01	.05	.01	.85	.03	.32	.01	.97	.01	4651	239	5	scg9_sc1	scg9_sc1
	9	NEPS-SC4	2l	.36	.02	-	-	-	-	.05	.00	.86	.02	-	-	-	-	14640	518	1	scg9_sc1	scg9_sc1
	9	PISA-I+	2l	.37	.03	-	-	-	-	.04	.01	.75	.04	-	-	-	-	6020	152	1	n2t1	wle_n2t1
	9	All[b]	2l	.36	.01	.31	.01	.96	.01	.05	.00	.84	.01	.32	.01	.97	.01	25311	909			
	10	PISA-I+	2l	.29	.03	.20	.01	.96	.01	.04	.01	.75	.04	.21	.01	.97	.01	6020	152	2	n2t2	wle_n2t2
	2	NEPS-SC2	2l	.11	.01	.15	.01	.42	.04	.06	.01	.29	.05	.17	.01	.45	.04	6319	362	4	rxg2_sc3	rxg2_sc3
	4	NEPS-SC2	2l	.11	.01	.26	.01	.42	.04	.10	.01	.72	.04	.31	.01	.79	.03	5419	349	6	reg4_sc1	reg4_sc1
	5	NEPS-SC3	2l	.34	.02	-	-	-	-	.03	.01	.87	.03	-	-	-	-	5380	225	1	reg5_sc1	reg5_sc1
	7	NEPS-SC3	2l	.33	.02	.22	.01	.92	.01	.02	.00	.81	.03	.22	.01	.95	.02	6279	266	3	reg7_sc1	reg7_sc1

Introduction | Tab.B1_General | Tab.B2_General_ND | Tab.B3_Adjusted | Tab.B4_Adjusted_ND | Tab.B5_Academic | Tab.B6_Academic_ND | Tab.B7_Non-Academic | Tab.B8_Non-Academic_ND | Tab.B9_General-2l | Tab.B10_General-2l_ND

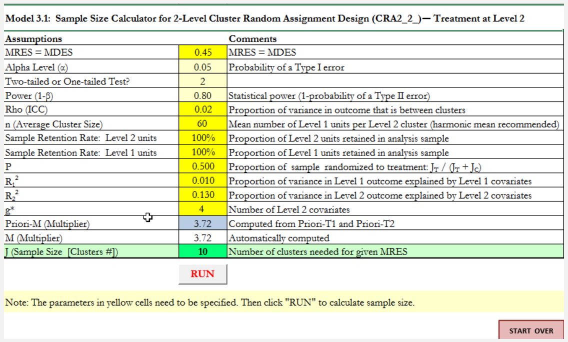
How many schools are required for a two-level cluster-randomized trial (CRT)?

To compute the required number of schools we need to specify the following set of parameters

- an estimate of the intraclass correlation of the outcome: ICC / Rho = 0.12
- number of covariates at the school level: one pretest
- amount of explained variance by covariates
  - school level: R2_L2 = .64
  - student level: R2_L1 = .40

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 1



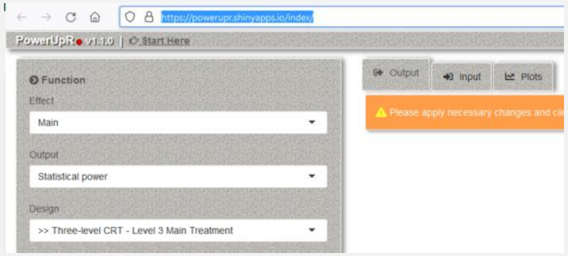
Model 3.1: Sample Size Calculator for 2-Level Cluster Random Assignment Design (CRA2_2_)— Treatment at Level 2

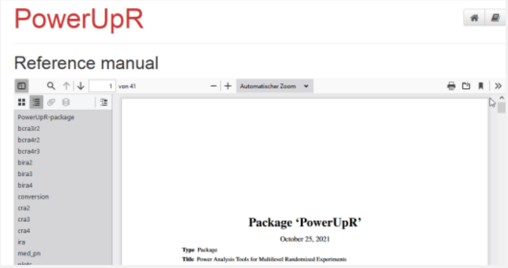
Assumptions		Comments
MRES = MDES	0.45	MRES = MDES
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.02	Proportion of variance in outcome that is between clusters
n (Average Cluster Size)	60	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
Sample Retention Rate: Level 2 units	100%	Proportion of Level 2 units retained in analysis sample
Sample Retention Rate: Level 1 units	100%	Proportion of Level 1 units retained in analysis sample
P	0.500	Proportion of sample randomized to treatment: $J_T / (J_T + J_C)$
R_1^2	0.010	Proportion of variance in Level 1 outcome explained by Level 1 covariates
R_2^2	0.130	Proportion of variance in Level 2 outcome explained by Level 2 covariates
g*	4	Number of Level 2 covariates
Priori-M (Multiplier)	3.72	Computed from Priori-T1 and Priori-T2
M (Multiplier)	3.72	Automatically computed
J (Sample Size [Clusters #])	10	Number of clusters needed for given MRES

RUN

Note: The parameters in yellow cells need to be specified. Then click "RUN" to calculate sample size.

START OVER





How many schools are required for a two-level cluster-randomized trial (CRT)?

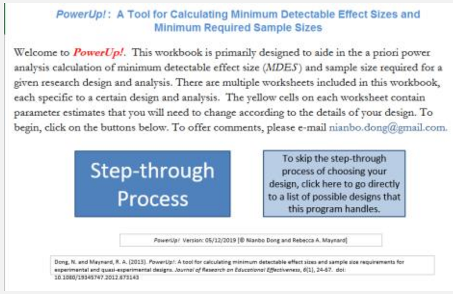
Three ways to do the power analyses

- Excel-Sheets: PowerUp, ... (https://www.causalevaluation.org/power-analysis.html)
- Shiny-App: https://powerupr.shinyapps.io/index/
- R-packages: e.g., PowerUpR (https://www.r-pkg.org/pkg/PowerUpR)
  - R syntax for all exercises can be accessed at: **XXX**

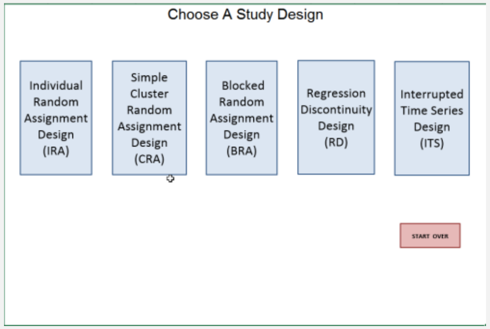
Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 1

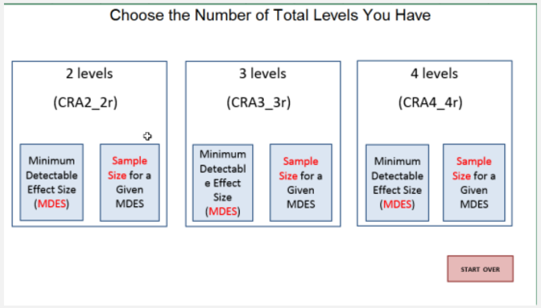
1. First Sheet „INTRO“



2. Choose a Design



3. Choose number of levels



4. Plug in an informed set of parameters

Model 3.1: Sample Size Calculator for 2-Level Cluster Random Assignment Design (CRA2_2_)— Treatment at Level 2

Assumptions		Comments
MRES = MDES	0.45	MRES = MDES
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.02	Proportion of variance in outcome that is between clusters
n (Average Cluster Size)	60	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
Sample Retention Rate: Level 2 units	100%	Proportion of Level 2 units retained in analysis sample
Sample Retention Rate: Level 1 units	100%	Proportion of Level 1 units retained in analysis sample
P	0.500	Proportion of sample randomized to treatment: J_T / (J_T + J_C)
R_1^2	0.010	Proportion of variance in Level 1 outcome explained by Level 1 covariates
R_2^2	0.130	Proportion of variance in Level 2 outcome explained by Level 2 covariates
g*	4	Number of Level 2 covariates
Priori-M (Multiplier)	3.72	Computed from Priori-T1 and Priori-T2
M (Multiplier)	3.72	Automatically computed
J (Sample Size [Clusters #])	10	Number of clusters needed for given MRES

RUN

Note: The parameters in yellow cells need to be specified. Then click "RUN" to calculate sample size.

START OVER

How many schools are required for a two-level cluster-randomized trial (CRT)?

- Excel-Sheet: PowerUP

Exercises: Scenario 1

Model 3.1: Sample Size Calculator for 2-Level Cluster Random Assignment Design (CRA2_2_)— Treatment at Level 2		
Assumptions		**Comments**
MRES = MDES	0.15	MRES = MDES
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.12	Proportion of variance in outcome that is between clusters
n (Average Cluster Size)	40	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
Sample Retention Rate: Level 2 units	100%	Proportion of Level 2 units retained in analysis sample
Sample Retention Rate: Level 1 units	100%	Proportion of Level 1 units retained in analysis sample
P	0.500	Proportion of sample randomized to treatment: J_T / (J_T + J_C)
R_1^2	0.400	Proportion of variance in Level 1 outcome explained by Level 1 covariates
R_2^2	0.640	Proportion of variance in Level 2 outcome explained by Level 2 covariates
g*	1	Number of Level 2 covariates
Priori-M (Multiplier)	2.84	Computed from Priori-T1 and Priori-T2
M (Multiplier)	2.84	Automatically computed
J (Sample Size [Clusters #])	81	Number of clusters needed for given MRES
	RUN	

How many schools are required for a two-level cluster-randomized trial (CRT)?

Research Team 1 would like to conduct a two-level CRT on the effectiveness of a school-wide intervention to improve 4th graders mathematical achievement in Germany. The research team needs **81 schools** (with 40 students per school) to achieve MDES = 0.15 (with α = .05 and power 1 – β = .80)

Exercises: Scenario 1 (continued)

Research Team 1 would like to conduct a two-level CRT on the effectiveness of a school-wide intervention to improve 4th graders' mathematical achievement in Germany.

How many schools are required for a two-level cluster-randomized trial (CRT)?

- Everything else being equal and using the design with the pretest measure as baseline, what do you expect how the number of required schools changes, when ...
  - ... power increases from .80 to .90?
  - ... the set of four socio-demographic covariates are used—in addition to the pretest —to explain variance at
    - at Level 1?
    - at Level 2?
  - ... no covariates are used
  - ... the intraclass correlation (Rho / ICC) increases from .12 to .20?

Check your statistical intuition by using the information provided in Table Tab_B.9_General_2l and the parameters as set before.

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 1 (continued)

Research Team 1 would like to conduct a two-level CRT on the effectiveness of a school-wide intervention to improve 4th graders' mathematical achievement in Germany.

How many schools are required for a two-level cluster-randomized trial (CRT)?

- Everything else being equal and using the design with the pretest measure as baseline, what do you expect how the number of required schools changes, when ...
  - ... power increases from .80 to .90? J = 81 vs. J = 108
  - ... the set of four socio-demographic covariates are used—in addition to the pretest —to explain variance at
    - at Level 1? J = 81 (with R^2_{L1} = .40) vs. J = 80 (R^2_{L1} = .43)
    - at Level 2? J = 81 (with R^2_{L2} = .64 and g* = 1) vs. J = 61 (with R^2_{L2} = .76 and g* = 5)
  - ... no covariates are used at Level 1 and Level 2: J = 81 vs. J = 201
  - ... the intraclass correlation (Rho / ICC) increases from .12 to .20? J = 81 vs. J = 120

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 1 (continued)

How many schools are required for a two-level cluster-randomized trial (CRT)?

Let‘s assume that research Team 1 wants to use the baseline design with the pretest measure as covariates at Level 1 and 2

The design parameters (ICC, R^2_{L1}, and R^2_{L2}) are estimates of population quantities, and, thus, associated with statistical uncertainty.

The uncertainty can be taken into account in the planning process by using the standard errors (SE) for each design parameter as given in Table B9_General_2l. Assuming large-sample properties, we use the standard normal distribution to compute the limits for 95% confidence intervals

- ICC = .12 (SE = 0.013)
  - lower bound of the 95% CI = .12 – 1.96 * 0.013 = .095
  - upper bound of the 95% CI = .12 + 1.96 * 0.013 = .145
- R^2_{L1} = .40 (SE = 0.011)
  - lower bound of the 95% CI = .40 – 1.96 * 0.011 = .378
  - upper bound of the 95% CI = .40 + 1.96 * 0.011 = .422
- R^2_{L2} = .64 (SE = 0.033)
  - lower bound of the 95% CI = .64 – 1.96 * 0.033 = .575
  - upper bound of the 95% CI = .64 + 1.96 * 0.033 = .705

A **conservative approach** for planning the sample size is to use the upper bound of the 95% CI for the ICC = .145, and the lower bound estimates for R^2_{L1} = .378 and R^2_{L2} = .575. Using this set of parameters shows that the team requires J = 107 schools to achieve an MDES = .15. Recall, when using the point estimates of the design parameters the team required J = 81 schools.

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Scenario 2

How many schools are required for a three-level cluster-randomized trial?

Exercises: Scenario 2

Research Team 2 would like to conduct a three-level CRT on the effectiveness of a school-wide intervention to improve 4th graders’ mathematical achievement. The team plans to sample 3 classrooms with 20 students per classroom from every school and to use a pretest measures as obtained for all students as covariate. The researchers are interested in K, the number of schools necessary to detect a standardized intervention effect on student achievement of d = 0.25.

How many schools are required for a three-level cluster-randomized trial (CRT)?

- **Can you, please, assist the team with the planning process by**
  - (a) using the point estimates of the design parameters,
  - (b) developing a conversative estimate, and
  - (c) showing how the combined application of a pretest and sociodemographic characteristics as covariates reduces the number of required schools?

Recall, that consulting Figure 4 in Stallasch et al. (2021) may be helpful to find the corresponding design parameters.

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 2

Research Team 2 would like to conduct a three-level CRT on the effectiveness of a school-wide intervention to improve 4th graders' mathematical achievement. The team plans to sample 3 classrooms with 20 students per classroom from every school and to use a pretest measures as obtained for all students as covariate. The researchers are interested in K, the number of schools necessary to detect a standardized intervention effect on student achievement of d = 0.25.

How many schools are required for a three-level cluster-randomized trial (CRT)?

- **Can you, please, assist the team with the planning process by**
  - **(a) using the point estimates of the design parameters,**
    - point estimates: K = 24 schools
  - **(b) developing a conversative estimate, and**
    - conservative approach: K = 38 schools
  - **(c) showing how the combined application of a pretest and sociodemographic characteristics as covariates reduces the number of required schools?**
    - using pretest and sociodemographic characteristics as covariates: K = 15 / 26 schools when using the point estimates / a conservative approach

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Scenario 3

Which MDES is attainable for a three-level cluster-randomized trial?

Exercises: Scenario 3

Suppose that research Team 2 now plans a three-level CRT to study the impact of an intervention that is intended to affect students’ history achievement in comprehensive schools (grades 5–12). Due to budgetary constraints, a fixed maximum number of 40 schools (with 2 classrooms, and 20 students each) are at the researchers’ disposal. Given these limits, the primary concern of Team 3 is to ensure that the attainable MDES lies within the range of effects on students’ achievement (i.e., $0.20 \leq d \leq 0.30$) that is typically observed for this kind of intervention.

Which MDES Is attainable for a CRT?

- Can you, please, help the research team to figure out whether the planned sample size is sufficient to detect the typical intervention effect with confidence?
    - To this end, you may want to develope your solution strategies in a group of two to three people.
    - Recall, that consulting Figure 4 in Stallasch et al. (2021) and the Sheet „Introduction“ in the „Supplemental Online Material B“ may be helpful to find the corresponding design parameters.

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 3

Suppose that research Team 2 now plans a three-level CRT to study the impact of an intervention that is intended to affect students' history achievement in comprehensive schools (grades 5–12). Due to budgetary constraints, a fixed maximum number of 40 schools (with 2 classrooms, and 20 students each) are at the researchers' disposal. Given these limits, the primary concern of Team 3 is to ensure that the attainable MDES lies within the range of effects on students' achievement (i.e., 0.20 ≤ d ≤ 0.30) that is typically observed for this kind of intervention.

Which MDES Is attainable for a CRT?

Team 2 consults Figure 4 to find the suitable table of design parameters. Since the intervention is targeted at a single, specific school type within the non-academic track, Team 2 uses the design parameters that are adjusted for mean-level differences between school types. Moreover, since design parameters for history are not available, Team 2 consults Table B4 outlining the normative distributions across the various achievement domains to determine small (i.e., 25th percentile [P25]), medium (i.e., median), and large values (i.e., 75th percentile [P75]) of the design parameters. Entering the respective values for the intraclass correlations (P25: $\rho_{L2} = 0.04$, $\rho_{L3} = 0.08$; median: $\rho_{L2} = 0.06$, $\rho_{L3} = 0.10$; P75: $\rho_{L2} = 0.09$, $\rho_{L3} = 0.12$; see Table B4), Team 2 learns that the attainable *MDES* is 0.32/0.35/0.39 for small/medium/large values of ρ_{L2} and ρ_{L3}. Including both pretests and sociodemographics as covariates (P25: $R^2_{L1} = 0.17$, $R^2_{L2} = 0.70$, $R^2_{L3} = 0.84$; median: $R^2_{L1} = 0.22$, $R^2_{L2} = 0.77$, $R^2_{L3} = 0.90$; P75: $R^2_{L1} = 0.31$, $R^2_{L2} = 0.86$, $R^2_{L3} = 0.97$; see Table B4) and using the 75th percentiles of the values for ρ_{L2} and ρ_{L3} (as more conservative upper bounds), the respective values for the *MDES* reduce to 0.20/0.18/0.14 for small/medium/large values of R^2 at the various levels. Consequently, Team 2 can be quite confident that their CRT design offers sufficient sensitivity to detect a true intervention effect within the desired range when including both pretests and sociodemographics.

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Scenario 4

Which MDES is attainable for a two-level MSIRT?

Exercises: Scenario 4

Now consider an MST for estimating the cross-site mean effect size of an educational intervention on third-grade reading scores with: 30 elementary schools (J); 50 third-graders per school (n); 0.60 of the third-graders at each school randomized to treatment ($\overline{T}$), and individual second-grade reading scores as a baseline covariate (X). For third-grade reading test scores, Table 8 in Bloom et al. (2007) reports estimated values of ρ_C ranging from 0.15 to 0.22 across the five school districts examined, with a mean of 0.18. The table also reports corresponding estimates of $R^2_{C(within)}$ that range from 0.22 to 0.52, with a mean of 0.38.[15] So let's assume these mean values for ρ_C and $R^2_{C(within)}$.

Now we must assume a value for τ_*. Toward the high end of their findings, Weiss et al. (2017)

Which MDES is attainable for a two-level MSIRT?

We need to specify the following set of parameters

- one- or two-sided testing
- significance level α
- desired power 1 – β
- number of schools
- (average) number of students sampled per school
- share of students within schools in the control and intervention condition
- an estimate of the intraclass correlation of the outcome
- an estimate of the cross-site impact variation τ_*
- amount of explained variance by covariates
    - at the student level
    - in the cross-site impact variation with a certain number of covariates

Bloom, H. S., & Spybrook, J. (2017). Assessing the Precision of Multisite Trials for Estimating the Parameters of a Cross-Site Population Distribution of Program Effects. *Journal of Research on Educational Effectiveness*, *10*(4), 877–902. https://doi.org/10.1080/19345747.2016.1271069

Scenario 4: Heterogeneity in intervention effects

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
2017, VOL. 10, NO. 4, 843–876
https://doi.org/10.1080/19345747.2017.1300719

METHODOLOGICAL STUDIES

Check for updates

How Much Do the Effects of Education and Training Programs Vary Across Sites? Evidence From Past Multisite Randomized Trials

Michael J. Weiss[a], Howard S. Bloom[a], Natalya Verbitsky-Savitz[b], Himani Gupta[a], Alma E. Vigil[b], and Daniel N. Cullinan[a]

nature human behaviour

PERSPECTIVE
https://doi.org/10.1038/s41562-021-01143-3

Check for updates

Behavioural science is unlikely to change the world without a heterogeneity revolution

Christopher J. Bryan[1], Elizabeth Tipton[2] and David S. Yeager[1]

Scenario 4: Cross-Site Impact Variation (Weiss et al., 2017)

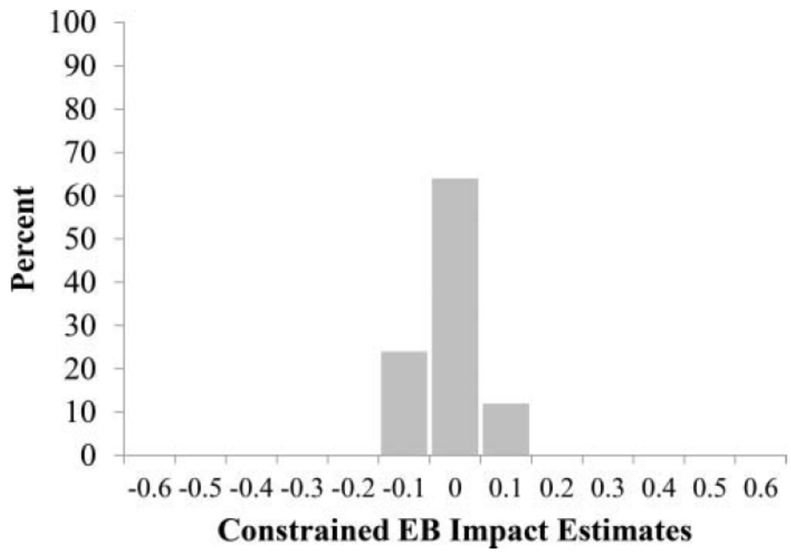


Figure 1. Histogram of site-level constrained empirical Bayes impact estimates—After School Reading Program, Year 1 reading outcome.

Consistent Near-Zero Impacts for Sites. The After School Reading Program has an estimated cross-site mean reading-achievement effect size near zero ($\hat{\beta}_Z = -0.02\sigma$) and a very small estimated cross-site standard deviation of effect sizes ($\hat{\tau}_Z = 0.05\sigma$ with a 90% confidence interval of 0.00σ to 0.10σ). These findings suggest that most program sites had little effect on reading achievement relative to "business as usual" in the same after-school center. Figure 1 graphically illustrates this situation by summarizing the cross-site distribution of constrained empirical Bayes effect-size estimates for this study.[18]

Weiss, M. J., Bloom, H. S., Verbitsky-Savitz, N., Gupta, H., Vigil, A. E., & Cullinan, D. N. (2017). How Much Do the Effects of Education and Training Programs Vary Across Sites? Evidence From Past Multisite Randomized Trials. *Journal of Research on Educational Effectiveness*, *10*(4), 843–876. https://doi.org/10.1080/19345747.2017.1300719

Why is it important to estimate the cross-site impact variation?

Scenario 4: Cross-Site Impact Variation (Weiss et al., 2017)

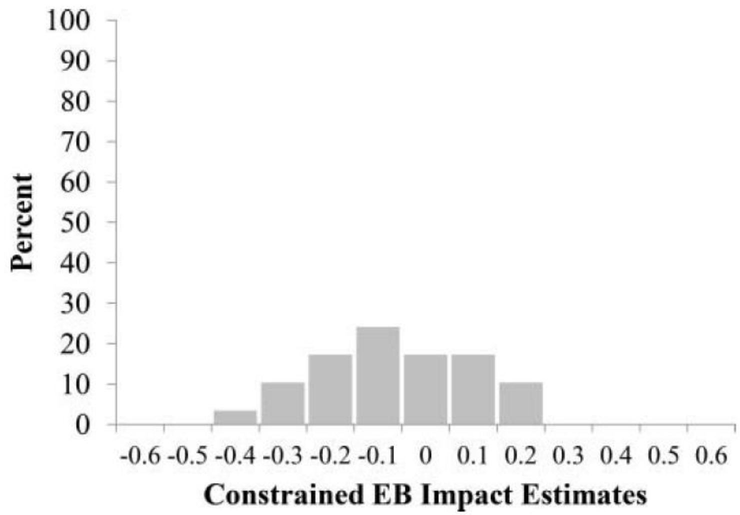


Figure 2. Histogram of site-level constrained empirical Bayes impact estimates—Charter Middle School, Year 2 reading outcome.

Near Zero Average Impact With Substantial Cross-Site Variation. Charter middle schools also have an estimated cross-site mean reading-achievement effect size ($\hat{\beta}_Z$) that is near zero (–0.02σ for the first follow-up year and –0.07σ for the second follow-up year), relative to their counterfactual schools). However, the estimated cross-site standard deviation of effect sizes for charter middle schools is substantial ($\hat{\tau}_Z = 0.15\sigma$ for the first follow-up year and $\hat{\tau}_Z = 0.16\sigma$ for the second follow-up year). Hence, the near-zero mean effect of these charter schools masks substantial cross-site impact variation, a fact that is illustrated in Figure 2. It should be noted, however, that although $\hat{\tau}_Z$ is 0.16σ for the second follow-up year, its 90% confidence interval is 0.08σ to 0.24σ. This highlights the considerable uncertainty that exists with respect to this estimate.

Why is it important to estimate the cross-site impact variation?

Scenario 4: Cross-Site Impact Variation (Weiss et al., 2017)

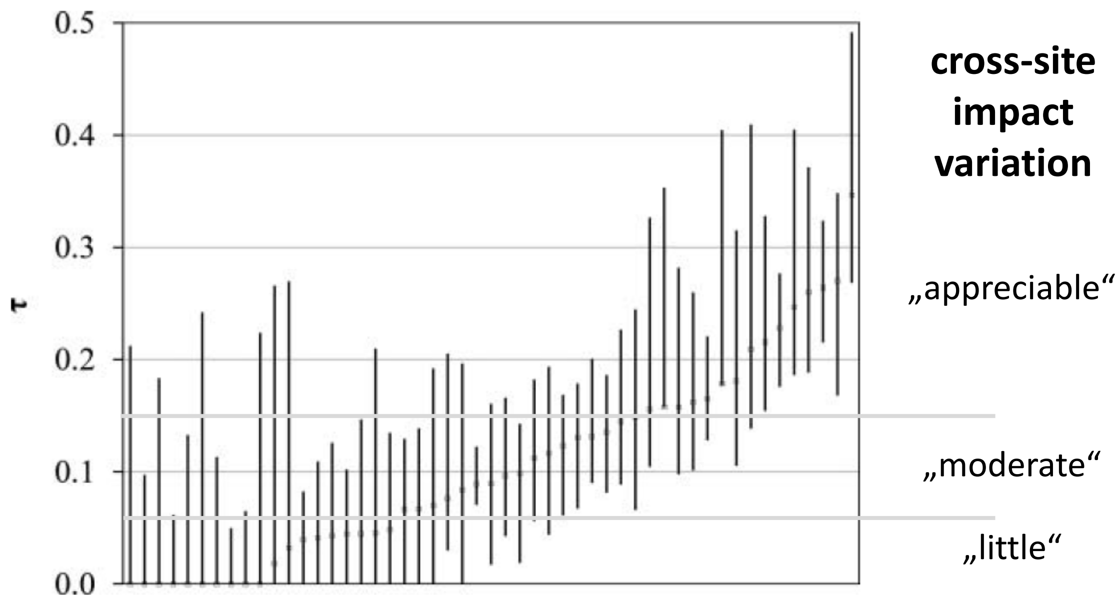


Figure 3. Caterpillar plot of $\hat{\tau}_Z$ and corresponding 90% confidence intervals for each study by outcome.

(In Figure 3 the reference group was the control group of a certain multi-site randomized trial.)

How much does the impact of interventions vary across sites?

- Weiss et al. estimated some 50 values for the cross-site impact variation with data from 16 large MSCRT‘s
- τ_Z depicts the magnitude of variation (i.e., the standard deviation) in site-average treatment effects relative to the outcome variability of the reference group.
  - 37% values of τ_Z ranged between $0 \leq \tau_Z \leq 0.05$: „little cross-site impact variation“
  - 33% values of τ_Z ranged between $0.05 < \tau_Z \leq 0.15$: „moderate cross-site impact variation“
  - 29% values of τ_Z ranged between $0.15 < \tau_Z \leq 0.35$: „appreciable cross-site impact variation“

Exercises: Scenario 4 (Bloom & Spybrook, 2017)

Now consider an MST for estimating the cross-site mean effect size of an educational intervention on third-grade reading scores with: 30 elementary schools (J); 50 third-graders per school (n); 0.60 of the third-graders at each school randomized to treatment ($\overline{T}$), and individual second-grade reading scores as a baseline covariate (X). For third-grade reading test scores, Table 8 in Bloom et al. (2007) reports estimated values of ρ_C ranging from 0.15 to 0.22 across the five school districts examined, with a mean of 0.18. The table also reports corresponding estimates of $R^2_{C(within)}$ that range from 0.22 to 0.52, with a mean of 0.38.[15] So let's assume these mean values for ρ_C and $R^2_{C(within)}$.

Now we must assume a value for τ_*. Toward the high end of their findings, Weiss et al. (2017) report an estimate near $0.25\sigma_C$ for the cross-school standard deviation of small-class effects on reading and math achievement for elementary-school students in the Tennessee STAR class size experiment (Word et al., 1990). At the lowest extreme of their findings, Weiss et al. (2017) report an estimate of τ_* equal to zero from data for a multisite study of an after-school math program (Black, Somers, Doolittle, Unterman, & Grossman, 2009). Given the preceding assumptions for $J,\ n,\ \overline{T}, \rho_C$ *and* $R^2_{C(within)}$ plus a value of τ_* equal to 0.25, Equation 15 implies that our MDES would be $0.17\sigma_C$. If instead we assume that τ_*equals zero, Equation 15 implies that our MDES would be $0.11\sigma_C$. This comparison illustrates the influence of τ_* on the MDES. The comparison also provides insight into the potential trade-off between the precision of a cross-site mean impact estimate obtained from site super population inference framework (which must account for sampling error due to cross-site impact variation) and that obtained from a fixed-site population inference framework, which need not account for this sampling error.[16]

Which MDES is attainable for a two-level MSIRT?

- Worked example by Bloom & Spybrook (2017)

Bloom, H. S., & Spybrook, J. (2017). Assessing the Precision of Multisite Trials for Estimating the Parameters of a Cross-Site Population Distribution of Program Effects. *Journal of Research on Educational Effectiveness*, *10*(4), 877–902. https://doi.org/10.1080/19345747.2016.1271069

Exercises: Scenario 4 (Bloom & Spybrook, 2017)

Weiss et al. (2017)

τ_Z The standard deviation of the cross-site distribution of site mean effect sizes in the Z metric. This parameter indicates the magnitude of variation in site-average treatment effects relative to the outcome variability of the reference group.

PowerUp (Dong & Maynard, 2013)

$\rho = \frac{\tau_2^2}{\tau_2^2+\sigma^2}$, unconditional intraclass coefficient (ICC).

$\omega = \frac{\tau_{T2}^2}{\tau_2^2}$ indicates treatment effect heterogeneity, which is the ratio of the variance of the treatment effect between blocks to the between-block residual variance. Note that $\rho\omega = \frac{\tau_{T2}^2}{\tau_2^2+\sigma^2}$, which is effect size variability.

Conversion

$$\tau_Z = \sqrt{\frac{\tau_{T2}^2}{\tau_2^2+\sigma^2}}$$

with τ^2_2 as the (unconditional) between-cluster/block variance (e.g., variance in the outcome between schools) and σ^2 as the (unconditional) within-cluster/block variance (e.g.,. variance in the outcome within schools).
When the outcome is z-standardized (e.g., for the total student population), it follows that
$\tau_2^2 + \sigma^2 = 1$
$\tau_Z^2 = \tau_{T2}^2$
$\rho = \tau_2^2$
$\Rightarrow \omega = \tau_Z^2 / \rho$

Which MDES is attainable for a two-level MSIRT?

- We need to convert the information provided in the study by Weiss et al. (2017) on the cross-site impact variation into ω that is requested by PowerUp

Exercises: Scenario 4 (Bloom & Spybrook, 2017)

$\omega = \tau_Z^2 \;/\; \rho = 0.25^2 \;/\; 0.18 = 0.3472$

Model 2.3: MDES Calculator for 2-Level Random Effects Blocked Individual Random Assignment (BIRA2_1r) Designs—Individuals Randomized within Blocks

Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.18	Proportion of variance in outcome between clusters
ω	0.35	Treatment effect heterogeneity: variability in treatment effects across Level 2 units, standardized by the variability in the Level-2 outcome
P	0.60	Proportion of Level 1 units randomized to treatment: n_T / (n_T + n_C)
R_1^2	0.38	Proportion of variance in the Level 1 outcome explained by Level 1 covariates
R_{2T}^2	0.00	Proportion of between block variance in treatment effect explained by Level 2 covariates
g*	0	Number of Level 2 covariates
n (Average Block Size)	50	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Blocks])	30	Number of Level 2 units in the sample
M (Multiplier)	2.90	Computed from T_1 and T_2
T_1 (Precision)	2.05	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.85	Determined from given power level
MDES	**0.171**	Minimum Detectable Effect Size

Which MDES is attainable for a two-level MSIRT?

- Worked example by Bloom & Spybrook (2017)

Bloom, H. S., & Spybrook, J. (2017). Assessing the Precision of Multisite Trials for Estimating the Parameters of a Cross-Site Population Distribution of Program Effects. *Journal of Research on Educational Effectiveness*, *10*(4), 877–902. https://doi.org/10.1080/19345747.2016.1271069

Exercises: Scenario 4 (Bloom & Spybrook, 2017)

$\omega = \tau_Z^2 / \rho = 0.0^2 / 0.18 = 0$

Model 2.3: MDES Calculator for 2-Level Random Effects Blocked Individual Random Assignment (BIRA2_1r) Designs—Individuals Randomized within Blocks		
Assumptions		**Comments**
Alpha Level (α)	0.05	Probability of a Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1-probability of a Type II error)
Rho (ICC)	0.18	Proportion of variance in outcome between clusters
ω	0.00	Treatment effect heterogeneity: variability in treatment effects across Level 2 units, standardized by the variability in the Level-2 outcome
P	0.60	Proportion of Level 1 units randomized to treatment: $n_T / (n_T + n_C)$
R_1^2	0.38	Proportion of variance in the Level 1 outcome explained by Level 1 covariates
R_{2T}^2	0.00	Proportion of between block variance in treatment effect explained by Level 2 covariates
g*	0	Number of Level 2 covariates
n (Average Block Size)	50	Mean number of Level 1 units per Level 2 cluster (harmonic mean recommended)
J (Sample Size [# of Blocks])	30	Number of Level 2 units in the sample
M (Multiplier)	2.90	Computed from T_1 and T_2
T_1 (Precision)	2.05	Determined from alpha level, given two-tailed or one-tailed test
T_2 (Power)	0.85	Determined from given power level
MDES	**0.109**	Minimum Detectable Effect Size

Which MDES is attainable for a two-level MSIRT?

- Worked example by Bloom & Spybrook (2017)

Bloom, H. S., & Spybrook, J. (2017). Assessing the Precision of Multisite Trials for Estimating the Parameters of a Cross-Site Population Distribution of Program Effects. *Journal of Research on Educational Effectiveness*, *10*(4), 877–902. https://doi.org/10.1080/19345747.2016.1271069

Scenario 5

How many schools are required for a three-level MSCRT?

Exercises: Scenario 5

Research Team 5 would like to conduct a three-level MSCRT to study the effects of a new teaching method involving learning software developed to enhance grade 9 students' English listening comprehension skills in the academic track. In this MSCRT, students are nested within classrooms, and classrooms are nested within schools, where classrooms within schools are randomly assigned to experimental conditions. Team 5 plans to have J = 4 and n = 20. The researchers consider an intervention effect of d = 0.10 policy-relevant. Since the goal of Team 5 is to generalize the study findings to the population of German academic track schools beyond those sampled for their MSCRT, they treat the school effects as random. Team 5 assumes that τ_Z= .15.

How many schools are required for a three-level MSCRT?

- Can you, please, help the research team to figure out how many schools are required to carry out this MSCRT using the assumptions made above?
- To this end, you also study how the of use socio-demographic characteristics at Level 1, 2, and 3 may improve the MDES. Here you assume that these covariates explain 9% of the variance of the intervention effect across schools (at L3).

$$\omega_3 = \tau_Z^2 / \rho_3$$

ρ_3: intraclass correlation at Level 3 (i.e., unconditional between-school variance in the outcome)

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 5

Research Team 5 would like to conduct a three-level MSCRT to study the effects of a new teaching method involving learning software developed to enhance grade 9 students' English listening comprehension skills in the academic track. In this MSCRT, students are nested within classrooms, and classrooms are nested within schools, where classrooms within schools are randomly assigned to experimental conditions. Team 5 plans to have J = 4 and n = 20. The researchers consider an intervention effect of d = 0.10 policy-relevant. Since the goal of Team 5 is to generalize the study findings to the population of German academic track schools beyond those sampled for their MSCRT, they treat the school effects as random. Team 5 assumes that τ_Z= .15.

How many schools are required for a three-level MSCRT?

- **Can you, please, help the research team to figure out how many schools are required to carry out this MSCRT using the assumptions made above?**
  - K = 198 schools
- **To this end, you also study how the of use socio-demographic characteristics at Level 1, 2, and 3 may improve the MDES. Here you assume that these covariates explain 9% of the variance of the intervention effect across schools (at L3).**
  - K = 89 schools

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 5

How many schools are required for an MSCRT? Without Covariates

Model 4.2: Sample Size Calculator for 3-Level Random Effects Blocked Cluster Random Assignment Designs (BCRA3_2r)— Treatment at Level 2

Assumptions		Comments
MRES = MDES	0.10	Minimum Relevant Effect Size = Minimum Detectable Effect Size
Alpha Level (α)	0.05	Probability of Type I error
Two-tailed or One-tailed Test?	2	
Power (1-β)	0.80	Statistical power (1 - probability of Type II error)
Rho_3 (ICC_3)	0.07	Proportion of variance in outcome between Level 3 units: V3/(V1+V2+V3)
Rho_2 (ICC_2)	0.19	Proportion of variance in outcome between Level 2 units: V2/(V1+V2+V3)
ω_3	0.32	Level 3 teatment effect heterogeneity: variance in treatment effect across Level 3 units, standardized by Level-3 outcome variance: $\omega = \xi_{\eta}^2 / \xi^2$
P	0.50	Proportion of level 2 units randomized to treatment: J_T / (J_T + J_C)
R_1^2	0.00	Proportion of variance in Level 1 outcome explained by Level 1 covariates
R_2^2	0.00	Proportion of variance in Level 2 mean outcomes explained by Level 2 covariates
R_{3T}^2	0.00	Proportion of between block variance in treatment effect explained by Level-3 covariates
g_3*	0	Number of Level 3 covariates
n (Average Sample Size for Level 1)	20	Mean number of Level 1 units per Level 2 unit (harmonic mean recommended)
J (Average Sample Size for Level 2)	4	Mean number of Level 2 units per Level 3 unit (harmonic mean recommended)
M (Multiplier)	2.82	Automatically computed
K (Sample Size [# of Blocks])	198	Number of Level 3 blocks needed for given MRES

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539

Exercises: Scenario 5

How many schools are required for an MSCRT? With Socio-Demographic Covariates

Model 4.2: Sample Size Calculator for 3-Level Random Effects Blocked Cluster Random Assignment Designs (BCRA3_2r)— Treatment at Level 2

Assumptions		Comments
MRES = MDES	0.10	Minimum Relevant Effect Size = Minimum Detectable Effect Size
Alpha Level (α)	0.05	Probability of Type I error
Two-tailed or One-tailed Test?	2	
Power ($1-\beta$)	0.80	Statistical power (1 - probability of Type II error)
Rho_3 (ICC_3)	0.07	Proportion of variance in outcome between Level 3 units: V3/(V1+V2+V3)
Rho_2 (ICC_2)	0.19	Proportion of variance in outcome between Level 2 units: V2/(V1+V2+V3)
ω_3	0.32	Level 3 teatment effect heterogeneity: variance in treatment effect across Level 3 units, standardized by Level-3 outcome variance: $\omega=\xi_{T}^{2}/\xi^{2}$
P	0.50	Proportion of level 2 units randomized to treatment: J_T / (J_T + J_C)
R_1^2	0.01	Proportion of variance in Level 1 outcome explained by Level 1 covariates
R_2^2	0.72	Proportion of variance in Level 2 mean outcomes explained by Level 2 covariates
R_{3T}^2	0.09	Proportion of between block variance in treatment effect explained by Level-3 covariates
g_3*	0	Number of Level 3 covariates
n (Average Sample Size for Level 1)	20	Mean number of Level 1 units per Level 2 unit (harmonic mean recommended)
J (Average Sample Size for Level 2)	4	Mean number of Level 2 units per Level 3 unit (harmonic mean recommended)
M (Multiplier)	2.83	Automatically computed
K (Sample Size [# of Blocks])	89	Number of Level 3 blocks needed for given MRES

Stallasch, S. E., Lüdtke, O., Artelt, C., & Brunner, M. (2021). Multilevel Design Parameters to Plan Cluster-Randomized Intervention Studies on Student Achievement in Elementary and Secondary School. *Journal of Research on Educational Effectiveness*, *0*(0), 1–35. https://doi.org/10.1080/19345747.2020.1823539