Choice of Randomized Design in Large-Scale Education Experiments

Brenda Jenney and Sharon Lohr
Department of Mathematics and Statistics, Arizona State University
Tempe, AZ 85287-1804

Abstract

Educational research often studies subjects that are in naturally clustered groups of classrooms or schools. When designing a randomized experiment to evaluate an intervention directed at teachers, but with effects on teachers and their students, the power for the treatment effect needs to be examined at both levels. A design that is optimal for estimating the treatment effect at one level may be inefficient for estimating the treatment effect at another level. In this paper, we study the efficiency of three designs and their ability to detect a treatment effect: randomize schools to treatment, randomize teachers within schools to treatment, and completely randomize teachers to treatment. The three designs are compared for both the teacher and student level within the mixed model framework, and a simulation study is conducted to compare treatment variances for the three designs with various levels of correlation within and between clusters.

KEY WORDS: experimental design, hierarchical design, multilevel response

1. Introduction

The gold standard for evaluating educational studies is commonly held to be a randomized controlled trial (*What Works Clearinghouse Evidence Standards for Reviewing Studies*, 2006). Boruch (2002, page 38) comments that although randomized trials have "been slow to come to the field of education," they provide the best way of ascertaining which interventions are truly effective for helping students. Randomized studies, the preferred design of evaluators, highlight the impact of interventions (versus simple outcomes of interventions) and are frequently cited by educational reformers (Gueron, 2005). There are many ways, however, of conducting a randomized trial, and it is desired to have a design that provides as much information as possible using the available resources.

In this paper, we study the efficiency of randomized designs for a situation in which teachers at multiple schools are randomized to treatments, and the impact of the educational treatment can be measured at multiple levels: in our case, at both the teacher and the student level. Often, the researchers will not be able to specify the class schedule for students involved in the study, so the design needs to be robust to possible self-selection by students.

The research in this paper was motivated by the problem of evaluating the effects of Project Pathways (CRESMET, Arizona State University, 2007). Project Pathways is a professional development program for secondary STEM (Science, Technology, Engineering and Mathematics) teachers, with the immediate goal of increasing teachers' conceptual and pedagogical knowledge of STEM topics. The primary intervention of Project Pathways is a set of four courses taken by secondary STEM teachers in schools located in Maricopa County, Arizona. The courses are developed around the themes of functions and proportional reasoning; course 1 concentrates on the mathematics, and courses 2-4 integrate the mathematical concepts with biology, physics, chemistry, geology and engineering. It is hypothesized that the teachers' increased understanding will lead to increased knowledge and achievement for the students who take classes from those teachers. Thus, while the treatment is administered to teachers, effects of the treatment need to be evaluated at both the teacher level and the student level. Students will likely take classes from several teachers, but Project Pathways will have no input on assigning students to teachers, so this needs to be taken into account in any analysis of the program. One ultimate goal of Project Pathways is to implement a randomized study of the proposed intervention. This research aims to provide some guidance for choice of experimental design in multi-school studies when the effect of the intervention can be measured at more than one level.

While there have been many small-scale randomized studies performed within schools, and many randomized studies relating to education programs that focus on students' physical and mental health, as well as tobacco, drug and alcohol use prevention programs, there have been relatively few whole school reforms tested with randomized studies (Cook, 2003). As more emphasis is placed on rigorous evaluations in education, the large-scale trials that are common in medical research should become more prevalent when studying educational innovations. Randomized studies have been found to give more reliable results in other fields, such as psychiatry (Johnson, 1998) and criminal justice (Berk, Ladd, Graziano, & Baek, 2003). The designs we explore here also apply to other settings. For further discussion on cluster randomized designs, Cook (2005) provides a synthesis of the most commonly encountered problems, as well as the merits of cluster-based experiments in the social sciences. Reasons to consider cluster randomization are also explored by Gail, Mark, Carroll, Green, and Pee (1996). For a discussion of optimal design when budget constraints are present, see Moerbeek, van Breukelen, and Berger (2000).

A number of researchers have studied the merits of different randomized designs in the hierarchical setting when the response is measured at one level. Moerbeek et al. (2000) derived the relative efficiency of randomize-by school vs. randomize-teacher-within-school designs for evaluating the impact of treatment on teachers. The efficiency of designs for multiple response levels, however, has not been previously explored.

To provide guidance on study design for multilevel-response educational studies, we look at the efficiencies for measuring program impact on teachers and students for three randomization schemes for the intervention: (1) randomly assigning schools, with all their teachers and students, to the experimental or control groups, (2) randomly assigning half of the teachers within each school to the experimental group (a randomized block design for teachers), and (3) randomly assigning teachers regardless of school to the two groups (a completely randomized design for teachers). In Section 2, we construct a unified mixed model framework for responses of teachers and students, and in Section 3 we study the relative efficiencies of the three designs for assessing treatment effects at teacher and student levels. Section 4 describes and presents the results of a simulation study conducted to approximate the power of the various designs for measuring the impact of the intervention on both teachers and students. In Section 5, we discuss implications of the theoretical results for evaluation design choice.

2. Models for Responses

The goal of the study is to evaluate effects of the intervention on teachers and simultaneously on students in their classes. With that goal in mind, we introduce models that could be used at each level of response. We then study efficiencies of the various designs for both models in Section 3. In the following, $\mathbf{I}_k$ is the $k \times k$ identity matrix, $\mathbf{1}_k$ is the k-vector of all ones, and $\mathbf{J}_k = \mathbf{1}_k \mathbf{1}_k'$. We assume that there are a schools available for the study, and that school i has m_i teachers and n_i students who could participate.

2.1 Teacher Model

Let T_{ij} be a response of interest for teacher j at school i. T_{ij} might be, for example, the change score on an assessment of content knowledge given before and after the intervention. Then a possible model for the teachers' response is the mixed effects model

$$\mathbf{T}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{1}_{m_i} v_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

where $\mathbf{T}_i = (T_{i1}, T_{i2}, \ldots, T_{im_i})'$ is an m_i-vector of responses, $\mathbf{X}_i = [\mathbf{x}_{i1}\mathbf{x}_{i2}\ldots\mathbf{x}_{im_i}]'$ is an $m_i \times p$ matrix of known covariates for the teachers, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-vector of fixed effects, $v_i \sim N(0, \sigma_v^2)$ is a random effect for the school, and and $\boldsymbol{\varepsilon}_i \sim N(0, \sigma_\varepsilon^2 \mathbf{I})$ is an m_i-vector of random error terms for the teachers, and where v_i and ε_{ij} are independent for all i and j. We assume in this model, then, that teachers from different schools are independent. The last column of $\mathbf{X}_i$ is a vector of indicators for treatment assignment for teacher j from school i:

$$x_{ijp} = \begin{cases} 1 & \text{if in experimental group} \\ -1 & \text{if in control group.} \end{cases}$$

The last element of $\boldsymbol{\beta}$, β_p, is the parameter of interest for assessing the effect of the treatment on the teachers.

Using mixed model theory (Demidenko, 2004), we have for the setup in (1),

$$\text{Cov}\,(\mathbf{T}_i | \mathbf{X}_i) = \mathbf{V}_i = \sigma_v^2 \mathbf{J}_{m_i} + \sigma_\varepsilon^2 \mathbf{I}_{m_i}.$$

Because the data for all schools are independent, the information for β_p is the sum of the information from each school. The information matrix of the generalized least squares estimator $\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{a} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{a} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{T}_i$ is:

$$\mathcal{I}_T(\hat{\boldsymbol{\beta}}, \mathbf{X}) = \sum_{i=1}^{a} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i; \tag{2}$$

where

$$\mathbf{V}_i^{-1} = \frac{1}{\sigma_\varepsilon^2} \mathbf{I}_{m_i} - \frac{\sigma_v^2}{\sigma_\varepsilon^2 (\sigma_\varepsilon^2 + \sigma_v^2 m_i)} \mathbf{J}_{m_i}. \tag{3}$$

We are primarily interested in the (p, p) entry of the matrix $\text{Cov}\,(\hat{\boldsymbol{\beta}})$, the variance for our treatment indicator variable, which we will approximate with $[\mathcal{I}_T(\hat{\boldsymbol{\beta}}, \mathbf{X})]^{-1}$.

2.2 Student Model

In addition to measuring the effect of the intervention on the teachers, we also, more importantly, wish to measure the effect of the intervention on students. During the span of the experiment, student k at school i may have classes from one or more of the teachers in the project. Let Y_{ik} be a response measure for student k at school i for $k = 1, \ldots, n_i$. One choice for Y_{ik} might be a change score for an assessment given before and after the intervention.

A model for Y_{ik} needs to allow students to take multiple classes and account for dependence among students in the same school and among students who take classes from the same teacher. We propose the following mixed model, related to a model in McCaffrey, Koretz, Louis, and Hamilton (2004), for the student's response. The response Y_{ik} depends on characteristics of the student and on characteristics of each teacher who instructs student (i, k):

$$\mathbf{Y}_i = \mathbf{B}_i \boldsymbol{\gamma} + \mathbf{D}_i (\mathbf{X}_i \boldsymbol{\theta} + \mathbf{t}_i) + \mathbf{1}_{n_i} s_i + \boldsymbol{\eta}_i. \tag{4}$$

Here, $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})'$ is an n_i-vector of student-level responses, $\mathbf{B}_i = [\mathbf{b}_{i1}\ \mathbf{b}_{i2} \ldots \mathbf{b}_{in_i}]'$ is an $n_i \times q$ matrix of known covariates for students in school i and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_{q-1})'$ is a q-vector of fixed effects. The $n_i \times m_i$ matrix $\mathbf{D}_i$ describes the assignment of students to teachers: the (k, j) element of $\mathbf{D}_i$ is d_{ikj} = number of classes student k from school i takes with teacher j. As in Section 2.1, $\mathbf{X}_i = [\mathbf{x}_{i1}\ \mathbf{x}_{i2}\ldots\mathbf{x}_{im_i}]$ is an $m_i \times p$ matrix of covariates for teachers at school i, whose last column is a vector of treatment indicators. The p-vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ is a vector of fixed effects for the teachers. Since multiple students take classes from teacher (i,j) we include a random effects vector, $\mathbf{t}_i = (t_{i1}, t_{i2}, \ldots, t_{im_i})'$, such that each $t_{ij} \sim N(0, \sigma_t^2)$. We posit an additive model for the effects of teachers on an individual student with element k of $\mathbf{D}_i(\mathbf{X}_i\boldsymbol{\theta}+\mathbf{t}_i)$ representing the additive effect of all of the teachers taken by student (i,k). A student may take classes from any number of the m_i teachers in the school, and may have the same teacher for multiple classes. The model also includes a random effect for the school, $s_i \sim N(0, \sigma_s^2)$, and random error terms for the students, $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, \ldots, \eta_{in_i})'$ such that $\eta_{ik} \sim N(0, \sigma_\eta^2)$. Assume that t_{ij}, s_i, and η_{ik} are mutually independent.

The fixed effect vector, including the parameters at both student and teacher levels, is $(\boldsymbol{\gamma}', \boldsymbol{\theta}')'$. The parameter of primary interest is $\boldsymbol{\theta}_p$, which corresponds to the treatment effect. For the model in (4),

$$\boldsymbol{\Sigma}_i = \mathrm{Cov}\,(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{D}_i) = \sigma_s^2\mathbf{J}_{n_i} + \sigma_t^2\mathbf{D}_i\mathbf{D}_i' + \sigma_\eta^2\mathbf{I}_{n_i}.$$

When $(\boldsymbol{\gamma}', \boldsymbol{\theta}')'$ is estimable, the generalized least squares estimator is

$$\begin{bmatrix} \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\theta}} \end{bmatrix} = \left(\sum_{i=1}^{a} \begin{bmatrix} \mathbf{B}_i' \\ (\mathbf{D}_i\mathbf{X}_i)' \end{bmatrix} \boldsymbol{\Sigma}_i^{-1} \begin{bmatrix} \mathbf{B}_i \vdots \mathbf{D}_i\mathbf{X}_i \end{bmatrix} \right)^{-1} \sum_{i=1}^{a} \begin{bmatrix} \mathbf{B}_i' \\ (\mathbf{D}_i\mathbf{X}_i)' \end{bmatrix} \boldsymbol{\Sigma}_i^{-1}\mathbf{Y}_i.$$

The information matrix for a given design $\mathbf{X}$ and student assignment $\mathbf{D}$ is

$$\mathcal{I}_S(\hat{\gamma}, \hat{\boldsymbol{\theta}}, \mathbf{X}, \mathbf{D}) = \sum_{i=1}^{a} \begin{bmatrix} \mathbf{B}_i' \\ (\mathbf{D}_i\mathbf{X}_i)' \end{bmatrix} \boldsymbol{\Sigma}_i^{-1} \begin{bmatrix} \mathbf{B}_i \vdots \mathbf{D}_i\mathbf{X}_i \end{bmatrix}.$$

If there are no student-level covariates, this simplifies to

$$\mathcal{I}_S(\hat{\boldsymbol{\theta}}, \mathbf{X}, \mathbf{D}) = \sum_{i=1}^{a} \mathbf{X}_i'\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i\mathbf{X}_i. \quad (5)$$

For the student model, we are primarily interested in the (p,p) entry of the matrix $\mathrm{Cov}\,(\hat{\boldsymbol{\theta}}|\mathbf{X}, \mathbf{D})$, the variance for our additive treatment effect, which we will approximate with $[\mathcal{I}_S(\hat{\boldsymbol{\theta}}, \mathbf{X}, \mathbf{D})]^{-1}$ (when the inverse exists). For specific values of $\mathbf{D}_i$ we can find the expected value of the information.

3. Efficiencies of Randomization Designs

In this section we examine the variance of the treatment parameter for teachers and students under three possible designs. To facilitate theoretical comparison of the designs, we make several simplifying assumptions. For each design, assume that each of the a schools has the same number of teachers, $m_i = m$, where m is even, and the same number of students, $n_i = n$. Also assume that $\mathbf{x}_{ij}' = [1\ \ x_{ij2}]$, where x_{ij2} is the treatment indicator, and that there are no student covariates available. In practice, any important available covariates should be used to improve the precision of the design, and randomization is employed to remove residual biases (Cox & Reid, 2000, page 33).

In order to examine the variance of the treatment coefficient, we will calculate the information matrix and the expected information for each design, for both the teacher-level and student-level responses. The inverse of the information matrix is the covariance matrix for the estimated fixed effects in each model, when those effects are estimable. We work with the information matrix rather than the covariance matrix because some of the designs can lead to a singular information matrix.

For any randomization that is done at the teacher or school level, the expected information depends on the $\mathbf{X}_i$ matrix. Let $\mathbf{X}_i = [\mathbf{1}_m \vdots \mathbf{R}_i]$ where the jth element of $\mathbf{R}_i$ is the treatment assignment of teacher j from school i:

$$\mathbf{R}_{ij} = \begin{cases} 1 & \text{if in experimental group} \\ -1 & \text{if in control group.} \end{cases}$$

For the teacher model in (1), when randomization is employed, equations (2) and (3) show that the information is:

$$\mathcal{I}_T(\hat{\boldsymbol{\beta}}, \mathbf{X}) = \sum_{i=1}^{a} \left\{ \frac{1}{\sigma_\varepsilon^2} \begin{bmatrix} m & \mathbf{1}'\mathbf{R}_i \\ \mathbf{1}'\mathbf{R}_i & m \end{bmatrix} - \frac{\sigma_v^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + \sigma_v^2 m)} \begin{bmatrix} m^2 & m\mathbf{1}'\mathbf{R}_i \\ m\mathbf{1}'\mathbf{R}_i & \mathbf{1}'\mathbf{R}_i\mathbf{R}_i'\mathbf{1} \end{bmatrix} \right\}. \quad (6)$$

For the student model in (4), the information in (5) depends on how students are assigned to classes within each school. For a given assignment of students to teachers, the information is:

$$\mathcal{I}_S(\hat{\boldsymbol{\theta}}, \mathbf{X}, \mathbf{D}) = \sum_{i=1}^{a} \begin{bmatrix} \mathbf{1}'\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i\mathbf{1} & \mathbf{1}'\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i\mathbf{R}_i \\ \mathbf{R}_i'\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i\mathbf{1} & \mathbf{R}_i'\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i\mathbf{R}_i \end{bmatrix}.$$

For each design, the expected information for θ_2 for a given assignment of students to teachers is:

$$\mathrm{E}\,[\mathcal{I}_S(\hat{\theta}_2, \mathbf{X}, \mathbf{D})|\mathbf{D}] = \sum_{i=1}^{a} \left\{ \mathrm{tr}\,[\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i\mathrm{Cov}\,(\mathbf{R}_i)] + \mathrm{E}\,[\mathbf{R}_i']\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i\mathrm{E}\,[\mathbf{R}_i] \right\}. \quad (7)$$

3.1 Design 1: Randomize Schools

In the first design, we randomize at the school level, assigning half of the available schools to each treatment. If school i is randomized to the treatment group then all teachers in the school i are in the treatment group and $\mathbf{R}_i = \mathbf{1}_m$. Likewise, if school i is randomized to the control group then $\mathbf{R}_i = -\mathbf{1}_m$. Thus, under this design $P(\mathbf{R}_i = \mathbf{1}_m) = P(\mathbf{R}_i = -\mathbf{1}_m) = 1/2$. Also $\sum_{i=1}^{a} \mathbf{1}'\mathbf{R}_i = 0$, $\mathbf{R}_i\mathbf{R}_i' = \mathbf{J}_m$, $\mathrm{E}[\mathbf{R}_i] = 0$, and $\mathrm{Cov}(\mathbf{R}_i) = \mathbf{J}_m$.

For the teacher model, the information matrix from (2) and (6) for any realization of randomization where there are an equal number of schools in each treatment is:

$$\mathcal{I}_{T1}(\hat{\boldsymbol{\beta}}, \mathbf{X}) = \frac{ma}{\sigma_\varepsilon^2 + \sigma_v^2 m}\mathbf{I}_2.$$

Therefore, the variance of the treatment variable for the teacher model when randomization is by school is $(\sigma_\varepsilon^2 + \sigma_v^2 m)/(ma)$.

For the student model, we have from equation (4) that the treatment effect is θ_2 if there are no student covariates and $\boldsymbol{\theta} = (\theta_1, \theta_2)'$. When teachers are randomized by school, the information of the treatment indictor from (7) is:

$$\mathcal{I}_{S1}(\hat{\theta}_2, \mathbf{X}, \mathbf{D}) = \sum_{i=1}^{a} \mathrm{tr}\,[\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i\mathbf{J}_m]. \quad (8)$$

When all teachers in a school are randomized to the same treatment, the information does not depend on $\mathbf{X}$. Consequently, the expected information is attained for every realization of the design.

3.2 Design 2: Randomize Teachers within Schools

In the second design, we randomly assign half of the teachers at each school to the experimental treatment, and the other half to the control treatment. With m teachers at each of the a schools, each school will have the same form for the $\mathbf{X}_i$ matrix, with $x_{ij2} = 1$ for half of the teachers, and $x_{ij2} = -1$ for the other half of the teachers. For any vector $\mathbf{r}_i$ with entries 1 and -1 and with $\mathbf{1}'\mathbf{r}_i = 0$, we have $P(\mathbf{R}_i = \mathbf{r}_i) = \binom{m}{m/2}^{-1}$. Consequently, $\mathbf{1}'\mathbf{R}_i = 0$, $\mathrm{E}[\mathbf{R}_i] = 0$ and $\mathrm{Cov}(\mathbf{R}_i) = (m\mathbf{I}_m - \mathbf{J}_m)/(m-1)$.

For the teacher model, the information matrix from (6) for any realization of this randomization is:

$$\mathcal{I}_{T2}(\hat{\boldsymbol{\beta}}, \mathbf{X}) = ma \begin{bmatrix} \dfrac{1}{(\sigma_\varepsilon^2 + \sigma_v^2 m)} & 0 \\ 0 & \dfrac{1}{\sigma_\varepsilon^2} \end{bmatrix}.$$

For the student model, the information for a specific randomization depends on the assignment of students to classes. The information can be zero in some cases. The expected treatment information over all randomizations for a given assignment of students to classes is, from (7):

$$\mathrm{E}[\mathcal{I}_{S2}(\hat{\theta}_2, \mathbf{X}, \mathbf{D})|\mathbf{D}] = \sum_{i=1}^{a} \frac{1}{m-1}\mathrm{tr}\,[\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i(m\mathbf{I}_m - \mathbf{J}_m)]. \quad (9)$$

3.3 Design 3: Completely Randomized Design of Teachers

In the third design, we randomize half of the ma teachers, regardless of school, to the experimental group, with the other half in the control group. In Sections 3.1 and 3.2, the variance of $\hat{\boldsymbol{\beta}}_p$ was the same for the teacher model regardless of which schools or teachers were randomly assigned to the treatment group. For the completely randomized design, school i may have between 0 and m teachers in the treatment group. Under this design $\mathrm{E}[\mathbf{R}_i] = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{R}_i) = (ma\mathbf{I}_m - \mathbf{J}_m)/(ma-1)$.

For the teacher model, the expected information matrix from (6) for this randomization is:

$$\mathrm{E}[\mathcal{I}_{T3}(\hat{\boldsymbol{\beta}}, \mathbf{X})] = \frac{ma}{\sigma_\varepsilon^2 + \sigma_v^2 m} \begin{bmatrix} 1 & 0 \\ 0 & 1 + \dfrac{(m-1)ma}{ma-1}\dfrac{\sigma_v^2}{\sigma_\varepsilon^2} \end{bmatrix}.$$

For the student model, the expected treatment information from (7) is:

$$\mathrm{E}[\mathcal{I}_{S3}(\hat{\theta}_2, \mathbf{X}, \mathbf{D})|\mathbf{D}] = \frac{1}{ma-1}\sum_{i=1}^{a} \mathrm{tr}\,[\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i(ma\mathbf{I}_m - \mathbf{J}_m)]. \quad (10)$$

3.4 Comparison of Designs

Table 1 gives the expected information for the treatment coefficient for the designs discussed in Sections 3.1-3.3. Since each expected information matrix is diagonal, the anticipated variance of $\hat{\beta}_2$ or $\hat{\theta}_2$ will be the reciprocal of the information (when it is nonzero).

For teacher-level assessments, as shown in Moerbeek et al. (2000) the expected information is highest when randomization is done within the schools. The information is lowest for design 1, when teachers are randomized by school. Depending on the realization of randomization, the efficiency of the completely randomized design will be between that of randomizing by teacher within schools and randomizing by school.

For student-level assessments, the expected information of $\hat{\theta}_2$ given $\mathbf{D}_i$ is of the form

$$\sum_{i=1}^{a} \mathrm{tr}\,[\mathbf{D}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i(k_1\mathbf{I}_m + k_2\mathbf{J}_m)],$$

for k_1, k_2 specified in Table 1. In general $\mathbf{D}_i$ depends on the assignment of students to classes in each school, and a design that is most efficient for teacher-level assessments may not be efficient for student-level assessments. Because $\boldsymbol{\Sigma}_i$ depends on $\mathbf{D}_i$, in general the expressions for expected treatment information in the student model need to be computed numerically. Simplifications are possible for some special cases of the student model when $\mathbf{D}_i$ has specified characteristics, and we examine some of these in the next section.

Table 1: Comparison of Expected Information for Treatment Variable

Randomization	Expected Treatment Information: Teacher Model	Student Model
By School	$\frac{ma}{\sigma_\varepsilon^2 + \sigma_v^2 m}$	$\sum_{i=1}^{a} \mathrm{tr}\,[\mathbf{D}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i \mathbf{J}_m]$
Within School	$\frac{ma}{\sigma_\varepsilon^2}$	$\frac{1}{m-1} \sum_{i=1}^{a} \mathrm{tr}\,[\mathbf{D}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i (m\mathbf{I}_m - \mathbf{J}_m)]$
CRD	$\frac{ma}{(\sigma_\varepsilon^2 + \sigma_v^2 m)} \left[1 + \frac{(m-1)ma}{ma-1} \frac{\sigma_v^2}{\sigma_\varepsilon^2} \right]$	$\frac{1}{ma-1} \sum_{i=1}^{a} \mathrm{tr}\,[\mathbf{D}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i (ma\mathbf{I}_m - \mathbf{J}_m)]$

3.5 Information for student model when $\mathbf{D}_i$ is balanced

In this section, we examine the special case when the assignment of students to classes is done in a balanced way. We define a balanced assignment to be one in which each of the possible $\binom{m}{c}$ assignments of c teachers to a student occurs with the same number of students. Consequently, in a balanced design, each class has nc/m students. We show in the appendix that for a balanced assignment $\mathbf{D}_i$,

$$\mathrm{tr}\,(\mathbf{D}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i \mathbf{J}_m) = \frac{mnc^2}{mn\sigma_s^2 + nc^2\sigma_t^2 + m\sigma_\eta^2}, \tag{11}$$

and

$$m\mathrm{tr}\,(\mathbf{D}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i) = \mathrm{tr}\,(\mathbf{D}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i \mathbf{J}_m) + \frac{m(m-1)nc(m-c)}{m(m-1)\sigma_\eta^2 + nc(m-c)\sigma_t^2}. \tag{12}$$

From (8) and (11), then, for any realization of the randomize-by-school design,

$$\mathcal{I}_{S1}(\hat{\theta_2}, \mathbf{X}, \mathbf{D}) = \frac{ac^2n}{n\sigma_s^2 + c^2\sigma_t^2 n/m + \sigma_\eta^2}.$$

From (9), (11) and (12) the expected information for the randomize within schools design, given $\mathbf{D}$ is:

$$\mathrm{E}\,[\mathcal{I}_{S2}(\hat{\theta_2}, \mathbf{X}, \mathbf{D}|\mathbf{D})] = \frac{amnc(m-c)}{nc(m-c)\sigma_t^2 + m(m-1)\sigma_\eta^2}.$$

From (10), (11) and (12) the expected information for the completely randomized design, given $\mathbf{D}$ is:

$$\mathrm{E}\,[\mathcal{I}_{S3}(\hat{\theta_2}, \mathbf{X}, \mathbf{D}|\mathbf{D})] = \frac{1}{ma-1}\left(\frac{a(a-1)mnc^2}{mn\sigma_s^2 + nc^2\sigma_t^2 + m\sigma_\eta^2}\right) + \frac{1}{ma-1}\left(\frac{a^2m(m-1)nc(m-c)}{nc(m-c)\sigma_t^2 + m(m-1)\sigma_\eta^2}\right).$$

3.5.1 *Information when each student takes one class from each teacher*

When $c = m$, that is, each student takes a class from each of the m teachers, then

$$\mathcal{I}_{S1}(\hat{\theta_2}, \mathbf{X}, \mathbf{D}) = \frac{am^2n}{\sigma_\eta^2 + n(\sigma_s^2 + m\sigma_t^2)},$$

$$\mathrm{E}\,[\mathcal{I}_{S2}(\hat{\theta_2}, \mathbf{X}, \mathbf{D}|\mathbf{D})] = 0,$$

and

$$\mathrm{E}\,[\mathcal{I}_{S3}(\hat{\theta_2}, \mathbf{X}, \mathbf{D}|\mathbf{D})] = \frac{a(a-1)m^2n}{(ma-1)(n\sigma_s^2 + nm\sigma_t^2 + \sigma_\eta^2)}.$$

Thus, in the design in which teachers are randomized within schools to the treatment, we cannot even estimate the desired treatment effect at the student level. In design 3, the expected information is positive, but it is possible to have a randomization in which θ_2 is not estimable. The expected information from design 3 is less than that from design 1. When $c = m$, the most efficient design for estimating θ_2 is design 1, in which all the teachers in a school are randomized to the same treatment. In this situation, then, design 1 is the most efficient design for estimating the effect of the intervention on the students. When examining the effect on teachers, we found that design 1 was the least efficient design.

3.5.2 *Information when each student takes one class*

For the special case in which each student takes a class with one teacher and each class has the same number, n/m, of students, then the information under any realization of design 1, and the expected information under

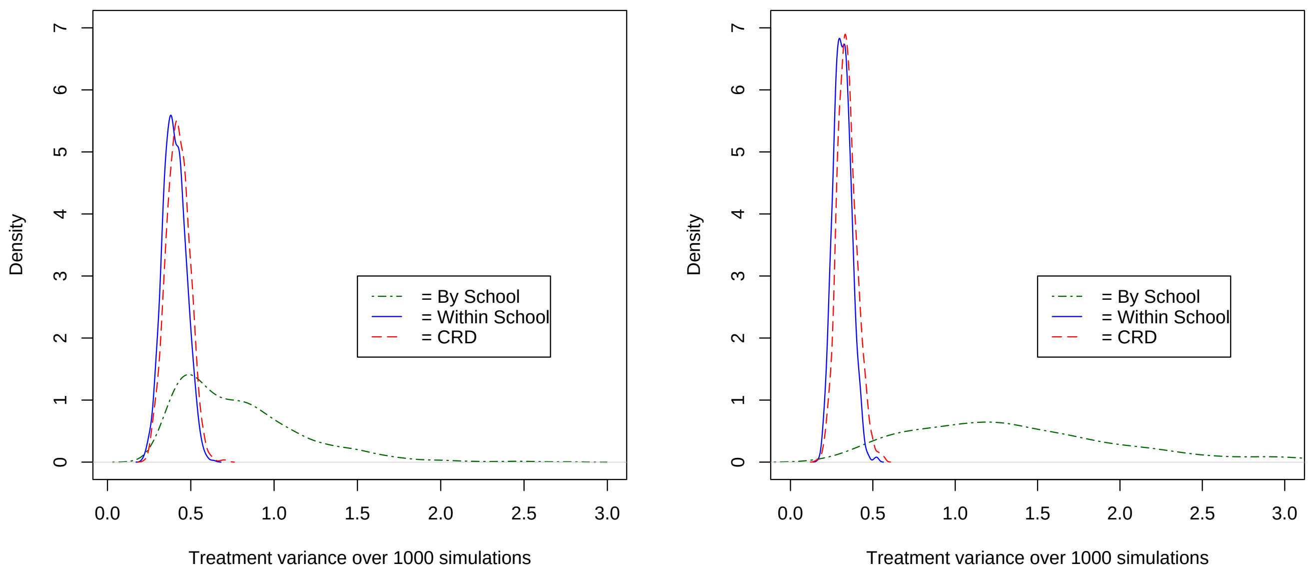

Figure 1: Teacher model: compare smoothed density of treatment variance for the three randomization designs. Left panel: $\sigma_v^2 = 0.9$, $\sigma_\varepsilon^2 = 8.1$. Right panel: $\sigma_v^2 = 2.7$, $\sigma_\varepsilon^2 = 6.3$. The randomize-schools design is less efficient when the between-school variance is a higher proportion of total variance (right panel).

designs 2 and 3 are:

$$\begin{aligned}\mathcal{I}_{S1}(\hat{\theta}_2, \mathbf{X}, \mathbf{D}) &= \frac{an}{\sigma_\eta^2 + \sigma_t^2 n/m + n\sigma_s^2},\\ \mathrm{E}\left[\mathcal{I}_{S2}(\hat{\theta}_2, \mathbf{X}, \mathbf{D}|\mathbf{D})\right] &= \frac{an}{\sigma_t^2 n/m + \sigma_\eta^2},\\ \mathrm{E}\left[\mathcal{I}_{S3}(\hat{\theta}_2, \mathbf{X}, \mathbf{D}|\mathbf{D})\right] &= \frac{1}{ma-1}\left(\frac{a(a-1)n}{n\sigma_s^2 + \sigma_t^2 n/m + \sigma_\eta^2}\right)\\ &\quad + \frac{1}{ma-1}\left(\frac{a^2 n(m-1)}{\sigma_t^2 n/m + \sigma_\eta^2}\right).\end{aligned}$$

In this case, design 2, when teachers are randomized within schools to the treatment, is the most efficient for estimating the treatment effect at the student level. Design 2 is also most efficient when examining the treatment effect at the teacher level.

4. Examining the Distribution of the Anticipated Sample Variance

In Section 3 we derived expected information matrices for the teacher and student models. For the student model, the information matrix depends in a complex way on the assignment of students to classes, and is analytically tractable only for special cases. In addition, it is possible for a specific realization of a design to have information that is far from its expected value. In this section, we examine the distribution of the anticipated variance of the treatment effect computationally under various scenarios.

The SAS macro *multleveldesign*, which is available from the authors, estimates the expected value and distribution of the anticipated variance of the treatment effect for inputted values of n, m, a and the variance components. Unlike other programs such as Optimal Design (Raudenbush, 2006), *multleveldesign* handles multiple response levels and displays the distribution of the sample variance of the treatment effect for simulated data.

Each simulation in Figure 1 and Figure 2 generates data for 10 schools, with 8 teachers and 200 students at each school, according to the assumptions for the teacher model in equation (1), and the student model in equation (4). We also assume that each student takes two classes from teachers at their school, and allow the possibility that a student would have the same teacher for two classes, in order to generate the $\mathbf{D}$ matrix. The variances for the random terms in the model were based on a preliminary study of Project Pathways data where total variance, $\sigma_v^2 + \sigma_\varepsilon^2 = 9$. A factorial experiment was conducted which set the intra-class correlation for the teacher model, $\rho = \sigma_v^2/(\sigma_v^2 + \sigma_\varepsilon^2)$, to be 0.1 or 0.3. All simulations assume that student and teacher scores are distributed the same; the variance components for the student model simulations are specified in the caption of Figure 2. The simulated data includes no covariates for teachers or students.

Although we did not include teacher or student covariates in our simulations, they would likely be available to most researchers conducting an educational study. The simplifying assumptions about the covariates make no difference; we take σ_v^2 and σ_ε^2 to be variance components of residuals after adjusting for known covariates. This

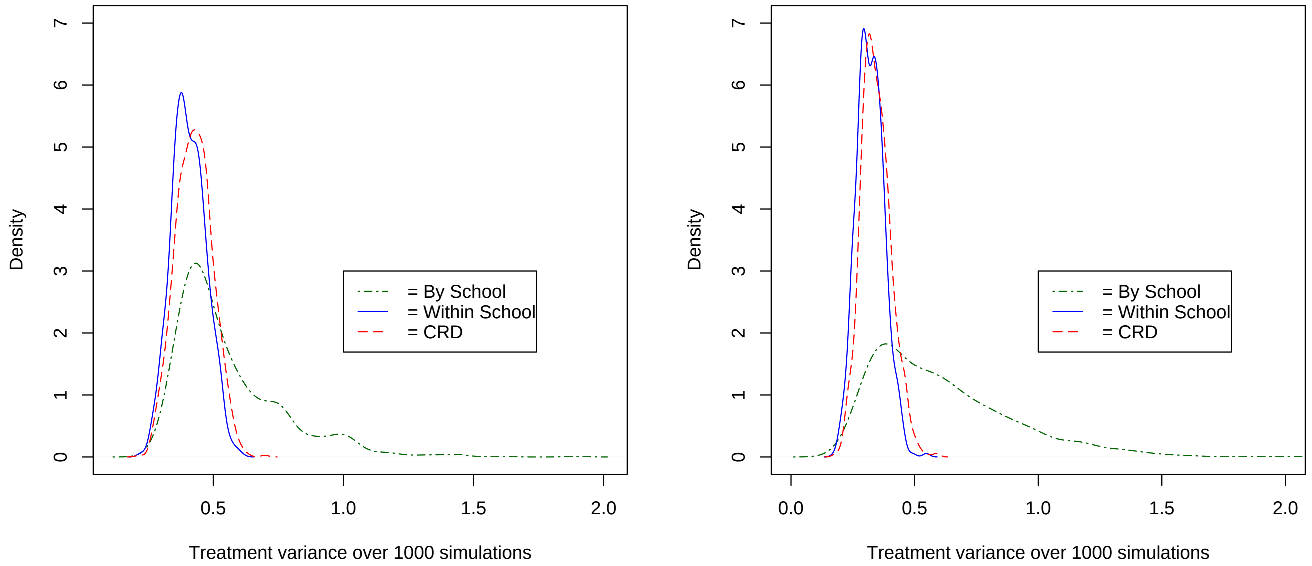

Figure 2: Student model: compare smoothed density of treatment variance for the three randomization designs. Left panel: $\sigma_s^2 = 0.9$, $\sigma_t^2 = 8.1$, $\sigma_\eta^2 = 8.1$. Right panel: $\sigma_s^2 = 2.7$, $\sigma_t^2 = 6.3$, $\sigma_\eta^2 = 6.3$. The randomize-schools design is nearly as efficient as the other two designs for student model when the between-school variance is expected to be low (left panel).

has the effect of reducing the cluster effects, since some of the school-to-school and teacher-to-teacher variability can be explained by covariates such as socio-economic status, years of teacher experience, and other variables. The inclusion of covariates in the model should increase the power for detecting treatment differences (Bloom, Richburg-Hayes, & Black, 2005).

5. Conclusion

The teacher-student conduit is crucial to the success of Project Pathways, making it particularly important to evaluate the effect of the intervention on both teachers and students. We desire to measure the impact of the intervention on teachers, since the intervention is a professional development program. We also need to measure whether the intervention is effective in improving student knowledge and achievement in mathematics and science.

There are clearly issues other than efficiency involved in the choice of evaluation design for Project Pathways. Among these are ease of implementation and school and teacher compliance, as well as the perceived fairness of the mechanism of randomization itself.

Because many interventions such as Project Pathways encourage community-based activities, there can be a risk of contamination – when some teachers in the control group adopt the intervention method – if the randomization is performed within schools or if a completely randomized design is used (Moerbeek, 2005). These concerns, together with the need to measure impact on students, suggest that it may be beneficial for such studies to consider randomization at the school level even though that design is less efficient for measuring the effect of the intervention on teachers.

If the interest is mostly in the effect of the intervention on teachers, the most efficient designs would be to randomize assignment of treatments within schools or to match teachers on variables such as experience or educational background. But these designs may not be optimal for estimating the effect of the intervention on students—indeed, in certain cases when teachers are randomized within schools, the effect of the intervention on students is not even estimable. To estimate effects on students, it is better to use a design in which randomization is performed at the school level when a student takes classes from multiple teachers at the school.

In this paper, we discussed design issues in the context of an educational study. The results, however, are general and can be applied to any setting in which data are collected at multiple levels. One application, for example, would be clinical trials in which patients are treated by several health care practitioners.

Appendix

Proof of equations (11) and (12):
First note that for any $\mathbf{D}_i$,

$$\mathbf{\Sigma}_i^{-1} = \mathbf{A} - \sigma_t^2 \mathbf{A}\mathbf{D}_i(\mathbf{I}_m + \sigma_t^2 \mathbf{D}_i' \mathbf{A}\mathbf{D}_i)^{-1}\mathbf{D}_i'\mathbf{A},$$

where

$$\mathbf{A} = \frac{1}{\sigma_\eta^2}\mathbf{I}_n - \frac{\sigma_s^2}{\sigma_\eta^2(\sigma_\eta^2 + n\sigma_s^2)\mathbf{J}_n}.$$

When $\mathbf{D}_i$ is balanced, i.e., each of the $\binom{m}{c}$ possible assignments of students to teachers occurs $k = n/\binom{m}{c}$ times, then each class has nc/m students and exactly $k\binom{m-2}{c-2} = nc(c-1)/m(m-1)$ students take a class from teacher j and a class from teacher l. Thus,

$$\begin{aligned}\mathbf{D}_i'\mathbf{D}_i &= \frac{nc(m-c)}{m(m-1)}\mathbf{I}_m + \frac{nc(c-1)}{m(m-1)}\mathbf{J}_m, \\ \mathbf{D}_i'\mathbf{1}_n &= \frac{nc}{m}\mathbf{1}_m,\end{aligned}$$

and

$$\mathbf{D}_i'\mathbf{J}_n\mathbf{D}_i = \left(\frac{nc}{m}\right)^2\mathbf{J}_m.$$

Then

$$\begin{aligned}\mathbf{D}_i'\mathbf{A}\mathbf{D}_i &= \frac{nc}{m\sigma_\eta^2}\left(\frac{m-c}{m-1}\right)\mathbf{I}_m \\ &\quad + \frac{nc}{m\sigma_\eta^2}\left(\frac{c-1}{m-1} - \frac{nc\sigma_s^2}{m(\sigma_\eta^2 + n\sigma_s^2)}\right)\mathbf{J}_m\end{aligned}$$

and

$$\begin{aligned}\mathbf{D}_i'\mathbf{\Sigma}_i^{-1}\mathbf{D}_i &= (\mathbf{I}_m + \sigma_t^2\mathbf{D}_i'\mathbf{A}\mathbf{D}_i)^{-1}\mathbf{D}_i'\mathbf{A}\mathbf{D}_i \\ &= \frac{1}{b\sigma_t^2}\left((b-1)\mathbf{I}_m + \frac{d}{b+md}\mathbf{J}_m\right)\end{aligned}$$

where

$$b = 1 + \frac{nc(m-c)\sigma_t^2}{m(m-1)\sigma_\eta^2}$$

and

$$d = \frac{nc\sigma_t^2}{m\sigma_\eta^2}\left(\frac{c-1}{m-1} - \frac{nc\sigma_s^2}{m(\sigma_\eta^2 + n\sigma_s^2)}\right).$$

Acknowledgements

This research was partially supported by the National Science Foundation under grants EHR-0412537 and SES-0604373.

References

Berk, R. A., Ladd, H., Graziano, H., & Baek, J.-H. (2003). A randomized experiment testing inmate classification systems. *Criminology & Public Policy*, *2*(2), 215–242.

CRESMET, Arizona State University. (2007). *Project Pathways (MSP).* Retrieved January 30, 2007 from http://cresmet.asu.edu/proj_res/msp.shtml.

Bloom, H., Richburg-Hayes, L., & Black, A. R. (2005). Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions. *MDRC Working Papers on Research Methodology.*

Boruch, R. (2002). The virtues of randomness. *Education Next*, *Fall*, 37–41.

Cook, T. D. (2003). Why have educational evaluators chosen not to do randomized experiments? *Annals of the American Academy of Political and Social Science*, *589*, 114–149.

Cook, T. D. (2005). Emergent principles for the design, implementation, and analysis of cluster-based experiments in social science. *Annals of the American Academy of Political and Social Science*, *599*(1), 176–198.

Cox, D., & Reid, N. (2000). *The theory of the design of experiments.* Boca Raton, FL: Chapman and Hall/CRC.

Demidenko, E. (2004). *Mixed models: Theory and applications.* Hoboken, NJ: Wiley.

Gail, M. H., Mark, S. D., Carroll, R. J., Green, S. B., & Pee, D. (1996). On design considerations and randomization-based inference for community intervention trials. *Statistics in Medicine*, *15*, 1069–1092.

Gueron, J. (2005). Throwing good money after bad: A common error misleads foundations and policymakers. *Stanford Social Innovation Review*, *Fall*, 69–71.

Johnson, T. (1998). Clinical trials in psychiatry: Background and statistical perspective. *Statistical Methods in Medical Research*, *7*, 209–234.

McCaffrey, D. F., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*(1), 67–101.

Moerbeek, M. (2005). Randomization of clusters versus randomization of persons within clusters: Which is preferable? *The American Statistician*, *59*(1), 72–78.

Moerbeek, M., van Breukelen, G. J. P., & Berger, M. P. F. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics*, *25*(3), 271–284.

Raudenbush, S. W. (2006). *Optimal design software v1.76.* Retrieved August 16, 2007 from http://sitemaker.umich.edu/group-based/optimal_design_software.

What Works Clearinghouse evidence standards for reviewing studies. (2006, September). Retrieved January 30, 2007 from http://www.whatworks.ed.gov/reviewprocess/study_standards_final.pdf.