Sampling methods

Five sampling methods were compared: stratified and unstratified random sampling, stratified and unstratified convenience sampling, and stratified balanced sampling (SBS). We model the sample selection process in two stages. In the first stage, we generated indicators of *potential participation*—that is, whether a school would participate in the trial *if approached for recruitment*. Let N be the total number of schools in the population, and let schools be indexed by $j = 1, ..., N$. We define E_j as a binary indicator that school j will agree to participate if contacted by recruiters, where $E_j = 1$ if the school agrees, and $E_j = 0$ if the school refuses. Each school was checked for approval by sampling from a Bernoulli distribution with probability equal to π_j^P for each school j

$$E_j \sim B(\pi_j^D) \tag{1}$$

The potential participation indicators were generated at the begining of each iteration and were therefore common across the five sampling methods examined.

We modeled the sampling process as observing the potential participation indicator for schools in a ranked list, where the order in which schools are contacted is determined by a score $S = S_1, ..., S_N$. Let $Z_j(S)$ be an indicator whether school i is sampled based on the score S. For the unstratified sampling methods, we determined $Z_1(S), ..., Z_N(S)$ by sorting schools according to S and selecting the first 60 schools where $E_j = 1$. Specifically,

$$Z_j(S) = I \left[60 \geq \sum_{k=1}^{N} E_k I \left(S_k \leq S_j \right) \right] \tag{2}$$

where $I(C)$ denotes the indicator function, equal to 1 if C is true and otherwise equal to 0. Based on the sample selection indicators, we calculated the number of schools contacted as

$$R(S) = \sum_{j=1}^{N} I(S_j \leq S_{max}), \tag{3}$$

where $S_{max} = \max\{S_1 Z_1, S_2 Z_2, ..., S_N Z_N\}$.

For the stratified sampling methods, the above process was applied within each stratum. We determined target sample sizes for each stratum based on a proportional allocation. Letting N_k denote the total number of schools in stratum k, we set a target sample size of $n_k = [60 \times N_k/N]$ for stratum k, where $[x]$ is the integer nearest to x.

Random Sampling

In the context of educational MRTs, unstratified random sampling is typically highly impractical. Large subsets of schools are likely to be overlooked if samples are too small. Further, random sampling from a broad population (such as an entire state) is likely to produce samples that are geographically dispersed), creating logistic difficulties for data collection and treatment implementation. However, unstratified random sampling is nonetheless interesting as a theoretically simple ideal against which to compare other sampling methods.

We simulated unstratified simple random sampling by ranking each school according to a random number from a uniform distribution, so that their order of recruitment was determined by the score $S_j \sim U(0, 1)$.

In practice, methods such as cluster sampling, stratified sampling, or a combination of both would likely offer advantages over unstratified random sampling. We therefore also simulated a stratified random sample, with strata determined by the results of the cluster analysis and using a proportional allocation.

Convenience Sampling

We also simulated two convenience sampling methods: unstratified (UCS) and stratified convenience sampling (SCS). Here, we assumed that recruiters have some knowledge of each schools' probability of participating, if

approached, and that they prioritize higher probability schools in order to minimize effort. To operationalize unstratified convenience sampling, we assumed that schools are approached for recruitment one at a time, with their order determined by sampling without replacement, with probability proportional to participation propensity scores $\pi_1^P, ..., \pi_N^P$. Once a school was selected and assigned a rank, the next school was selected with a probability proportional to the weights of the remaining schools. Once all ranks were assigned, schools were again approached until 60 schools agreed to be in the sample We implemented stratified convenience sampling using the same process, but with ranks determined independently within each stratum, and using a proportional allocation across strata.

Balanced Sampling

SBS is unique in that rankings are directly related to school characteristics and do not change across iterations. Scores within strata were based on equation @ref(eq:euclid) (i.e., $S_j = d_{jk}$), where schools that are closer to the "center" of the stratum are more representative of it. Though extremely unlikely, it is possible that several schools could be equidistant from the center of the stratum; in such cases, schools were ordered randomly. Because Tipton (REF) proposed balanced sampling in connection with stratification based on a cluster analysis, we only consider the stratified version, SBS.