The Matched Pairs Design: Lessons From Several Effectiveness Trials

Qingfeng Zhao, Andrew Jaciw, Akiko Lipton
Empirical Education Inc.

Background.

In the design of group randomized trials, we use matched pairs to increase the precision of the impact estimate. Pairs of similar units are formed and one unit within each pair is randomized to treatment while the other is randomized to control.

Tradeoffs

In a 2-level design, the minimum detectable effect size (MDES) is proportional to the following quantity (Bloom, 2005).

$$MDES(b_0)_{CL} \propto M\sqrt{\frac{\tau^2}{J}+\frac{\sigma^2}{nJ}}$$

- The multiplier, M, reflects the number of degrees of freedom available for the impact estimator. Modeling pairs uses up degrees of freedom and causes this quantity to increase. This quantity also increases as the number of cases randomized decreases.
- The goal of modeling pairs is to discount variation in the outcome at the level of randomization, τ^2, by accounting for between-pair differences that affect the outcome. This will lower the MDES.

We are interested in empirically investigating design tradeoffs, to see whether the discounting of τ^2 that results from pairing more than offsets the increase in M. We expect the benefits of pairing to vary depending on:

(1) the number of cases randomized.

(2) the proportion of total variance in the outcome that τ^2 accounts for (i.e., the intraclass correlation coefficient [ICC]).

Research Questions.

1) Does using matched pairs decrease the standard error of the impact estimator? Does it matter whether pairs are modeled as fixed or random?
2) What is the effect of using matched pairs on the MDES?
3) Does the answer to (2) depend on the number of units randomized, and if so, is there a sample size below which modeling matched pairs increases MDES, and therefore decreases precision?

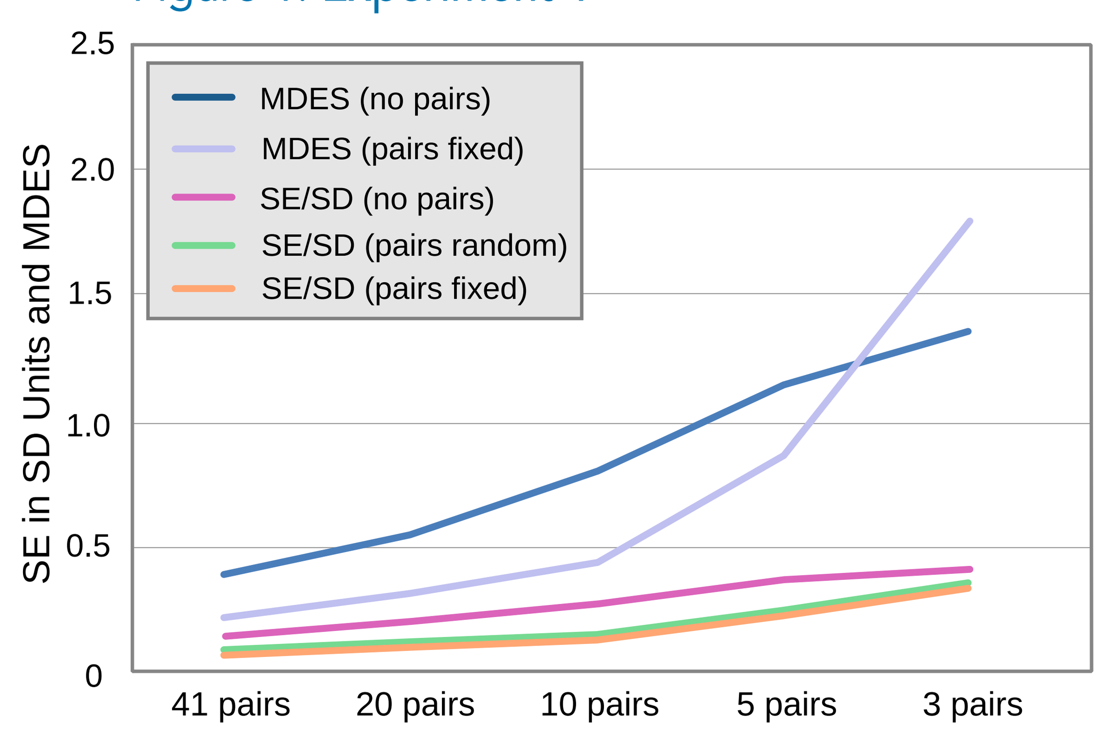
Details of the Experimental Designs.

Trial	Intervention	Unit randomized	No. of pairs	Total no. of students
1	Reform-based science and reading program	Teachers	41	2,638
2	Remedial reading intervention	Classes	13	610
3	Reform-based math intervention	Classes	21	3261
4	Reform-based reading intervention	Classes	21	3370

Main Impacts.

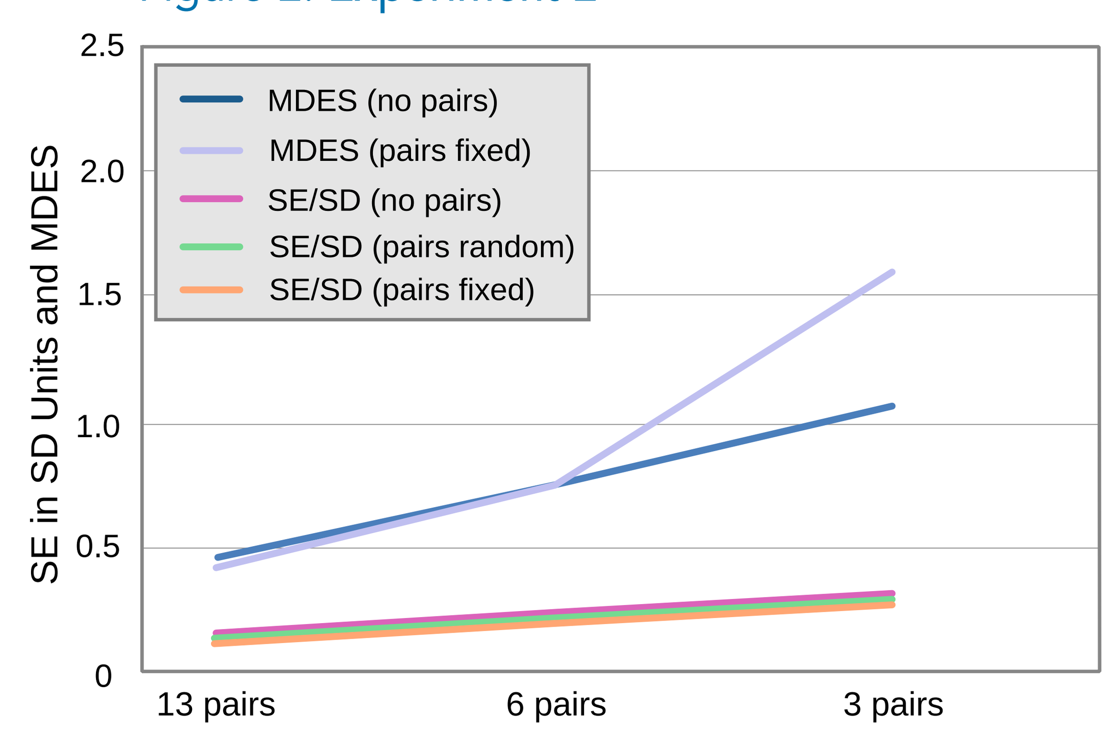
Standard Errors (SEs) in Standard Deviation (SD) Units and Minimum Detectable Effect Sizes

Figure 1. Experiment 1



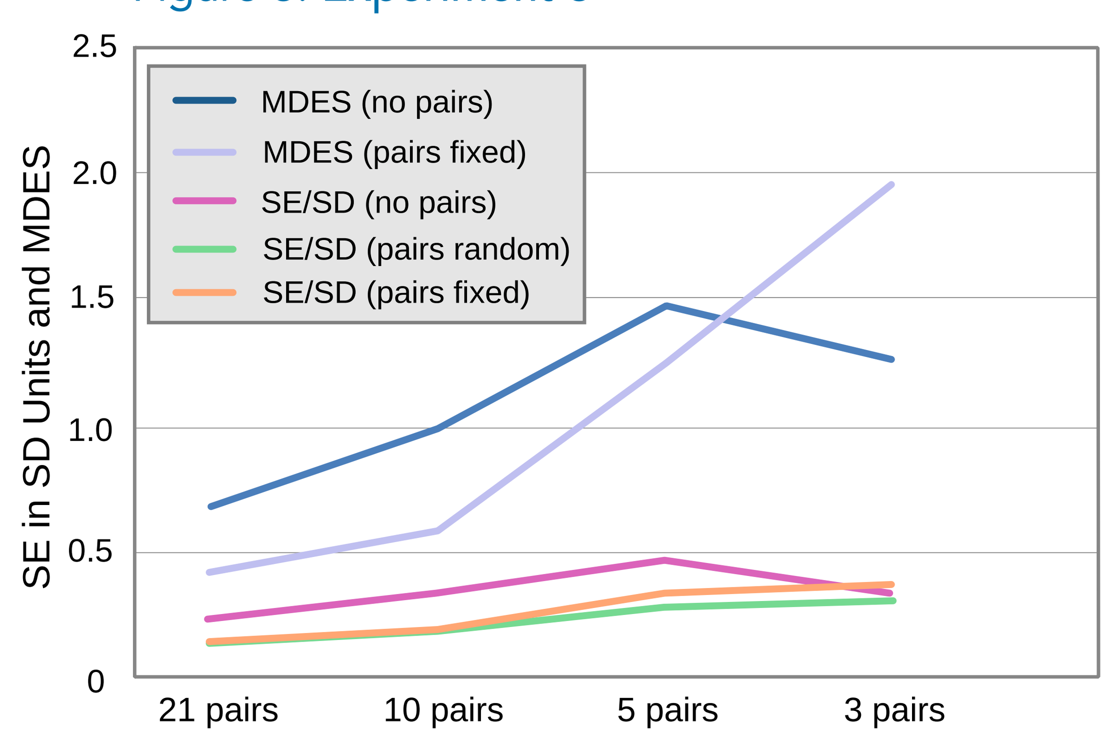
*The ICC for this experiment is .35.

Figure 2. Experiment 2



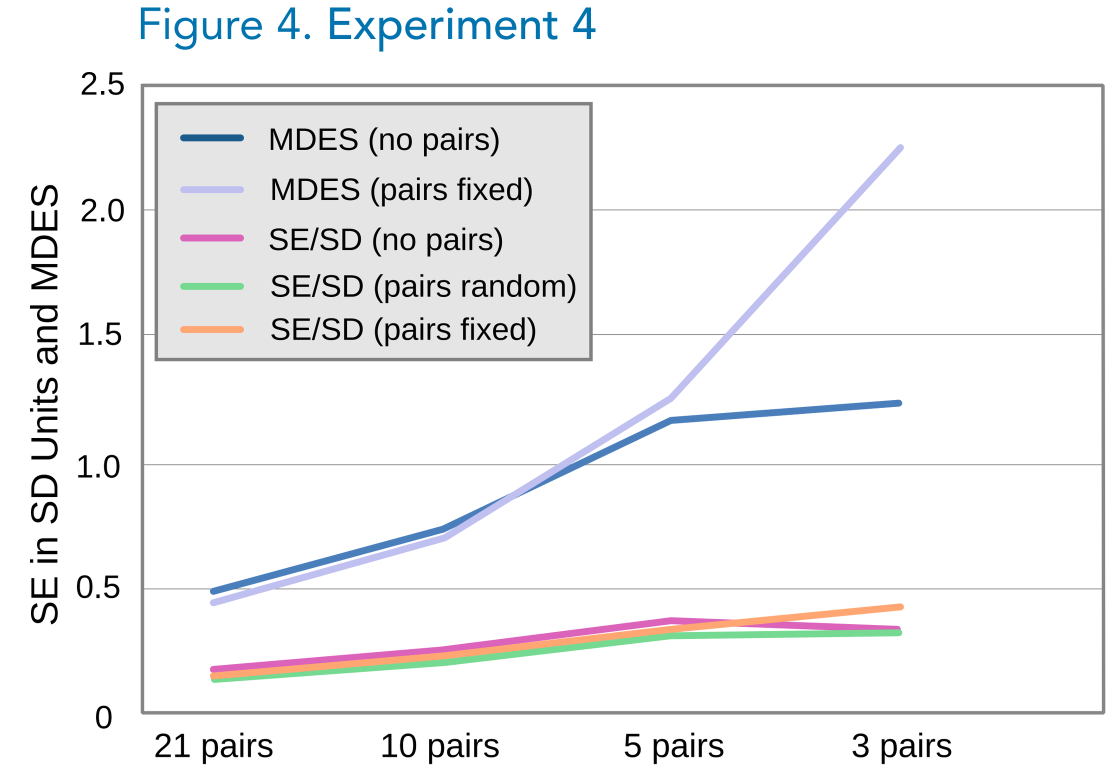
*The ICC for this experiment is .10.

Figure 3. Experiment 3



*The ICC for this experiment is .50.

Figure 4. Experiment 4



*The ICC for this experiment is .22.

References:
Bloom, H. S., (2005). Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning More From Social Experiments.* New York: Russell Sage Foundation.

Methods.

We analyzed four experiments that used a matched pairs design. In each experiment we estimated the impact for the full sample using a mixed model. We then randomly selected smaller subsets of pairs from the full sample of pairs to create progressively smaller sub-experiments, and for each subset, reran the analysis. (We drew multiple subsets of a given size and averaged the results from the separate analyses of those subsets.) We repeated the analyses with the pair levels modeled in the following ways.

1) Pairs were ignored.
2) Pairs were modeled as random.
3) Pairs were modeled as fixed.

We computed the MDES for the cases where (1) pairs are ignored, and (2) pairs are modeled as fixed.

Results.

(a) The standard error and MDES generally increase as the sample size and number of matched pairs decrease.

(b) For the most part, the standard error decreases when we model matched pairs for the two experiments with the larger ICCs (.50, .35), but not for the two experiments with the smaller ICCs (.22., .10).

(c) The standard error for the impact estimate appears unaffected by whether pairs are modeled as fixed or random.

(d) In the two experiments where modeling pairs reduces the standard error, the MDES also decreases when we model pair effects. This improvement is substantial especially with larger numbers of pairs. The benefit from modeling pairs ceases with four or fewer pairs.

(e) In the two experiments where modeling pairs does not lower the standard error, the MDES is either unaffected or increases when we model pair effects. This increase happens with six or fewer pairs.

Conclusion.

When pairing is successful at reducing the standard error of the impact estimate, there is a sizeable benefit to precision from modeling pair effects, so long as the number of pairs is four or greater. If pairing is unsuccessful, then modeling pairs can only hurt precision and this occurs when the number of pairs is approximately six or fewer. In the studies considered here, there was no benefit to using a matched pairs design for experiments with smaller ICCs. More studies of this kind are needed to substantiate the results found here. Similar tradeoffs to modeling matched pairs should be examined when the pretest and other covariates are modeled.