Analysis Plan for Principal Leadership

December 28, 2010

Contents

1 Description of Study

Principal leadership is a program designed to test the effectiveness of giving structured feedback to school principals by way of measuring teachers' perceptions of the principals effectiveness.

Year 1 will consists of 4 waves of data. The first, Spring 2011, will be pre-randomization and thus includes baseline measurements to be included as covariates in the analysis. Waves 2-4 will include Fall 2011, Winter 2011, and Spring 2012. Starting with Fall 2011, principals will be randomized to one of two groups (A or B). In year 1, group A receives feedback and group B receives no feedback. In year 2 (Fall 12, Winter 12, Spring 13), group A is split into two subgroups, feedback or feedback plus coaching, and group B receives feedback. In year 3 (Fall 13, Winter 13, Spring 14), data will only be collected on group B, in which principals receive feedback or feedback plus coaching.

The primary measurement will be the teacher's evaluation of the principal's effectiveness. At baseline (Spring 2011), there will be 72 items in the survey, in which each item includes six yes/no questions pertaining to source of their observation. In Fall 2012, the survey will be reduced to 36 items, with yes/no questions regarding source after every 6 items. A survey will also given to the principals, which essentially rates how the principal thinks the teachers will rate him/her. Additionally, two other measures will be administered, TRUST and two scales of Academic Press. We expect to recruit 90 principals in the study.

2 Data Management

We first focus on an analysis of the first year data. This is a reasonable approach given that the treatments are modified in the second year of the study. For consistency, we will only look at the 36-item based summary scores for all analyses, including those in year 1. Because we don't have unique identifiers at the teacher level, but do have unique identifiers at the principal level, it makes more sense to summarize the data at the principal level in order to proceed with the analysis.

For the main outcome, TGMean_36, this involves taking the mean across teachers for a given principal, so that there is only one row of data per principal for a given wave. Similar averages can be taken for other variables of interest, including "Teacher's don't know ratio" (are there other variables of interest?). Covariates of interest (including treatment condition) should be principal-level variables, and will need to be merged into the data set for consideration in the analysis.

3 Considerations

Descriptive statistics are given in descriptives.pdf for the 59 schools that agreed to participate. In general, there do not appear to be evidence of covariate imbalance at baseline between the treatment groups. The covariates considered are at the school or principal level, and most of them have about 10 missing values (17%). Missing data may be problematic when fitting models, especially if multiple covariates are used and the missing values from various covariates are from different individuals thus we will consider imputation.

Another issue is that many of the outcomes are collected at the teacher level, but there is no unique identifier for teacher in this longitudinal design. Although one could account for the clustering at the school level in a multilevel model, there is no way to account for the clustering at the individual teacher level. An analysis with the teacher-level data that only clusters at the school level makes the assumption that two observations from the same teacher at 2 different time points are independent after adjusting for the school effect. However, this would be an invalid assumption and would lead to smaller standard errors than what you should get. In addition, because randomization is at the school level, failing to properly account for clustering at a level lower than the randomization has been shown to lead to very misleading results.

Hence, the most straightforward approach is to summarize the teacher data at the principal level and fit the models at the school/principal level. Such an analysis loses some information, e.g. some schools will contribute more data (through more teachers) at the teacher level, but when summarized at the school level all schools will contributed an equal amount of information. Although not ideal, this would be a less serious mistake than mishandling the clustering. In addition, because randomization is at the school level, and the measurement outcome is targeting the school, I think it makes perfectly good sense to analyze the data with this type of summary score. For those principals joining the study after baseline we will evaluate the estimate and consider omitting the intercept term for treatment to avoid extrapolating a non-zero difference at baseline.

4 Primary Analysis

We will focus first on an intent to treat analysis that does not adjust for any covariates, and fit a model for each of the 9 main outcomes (ignoring subscales for now). The variables of interest include pgmean.36.80m, pprcval.80m, pselfeff.80m, pt.trust.80m, pacadempress.80m, m.tgmean.36.80m, m.tacadempress.80m, m.ttt.trust.80m. For these primary outcome variables, we will use a linear mixed model on the principal level data. Instead of adjusting for baseline, we will model the baseline (time 0) and post randomization waves (time 1,2,3) measures. This will allow us to keep all of the principals in the analysis even when there are missing baseline values. We will include a random intercept and slope for principal, with an unstructured covariance matrix of the random effects. We will include the following predictors:

1. time
2. intervention
3. time∗intervention

We will produce the standard output from the model, differences in means at 12 months between the intervention groups, and graphs of the predicted means across time. The efficacy of the intervention will be assessed with the time∗intervention interaction.

5 Comparing Intervention Groups A vs. B

An additional question is to compare the two types of intervention: Type A received general feedback and group B received general feedback plus information on the two lowest core sub-scales. We can re-fit the models identified for the primary analysis, but divide the intervention group into 3 classifications: intervention type A, intervention type B, and control, so there will be 2 main effects for treatment and 2 interactions with intervention and time. With Type A as the reference level, the intervention B * time interaction will provide an assessment comparing the two intervention groups.