

STAT 240 Final Project

Rebecca Barter, Andrew Do and Kellie Ottoboni

April 29, 2015

1 Introduction

Over the last two decades developing countries have seen an increase in the number of new primary school entrants, driven in part, by the elimination of school fees. For example, between 1999 and 2004 the number of new entrants to primary school in sub-Saharan Africa increased by more than 30 percent (?). Although this influx of new students is undeniably a positive development, steps need to be taken to ensure that the quality of education is not diminished. For example, by 2005, the average first grade class size in Kenya had swelled to 83, with 28 percent of first grade classes containing more than 100 students (?). Moreover, many of the new students were significantly less prepared than those in the past. Unfortunately, little prior work had been undertaken to evaluate the most effective methods of handling such an influx of students. In a randomized experiment, Duflo et al. aim to answer several questions related to a number of methods of resource allocation in primary education. In particular, the investigators aim to assess the impact of reduction in pupil-teacher ratios, implementing tracking (separating classes into high and low streams based on prior test scores) and different institutional environments (type of teacher, and whether or not the school undertakes teacher monitoring and education). In this report, we investigate the data obtained from the study undertaken by Duflo et al, with a focus on evaluating the impact of tracking on student achievement.

2 Tracking

Tracking involves separating pupils by academic ability within schools. In particular, a student is assigned to the high stream if their prior achievement is above the median, and is assigned to the low stream otherwise. There have been multitudes of studies involving the effects of tracking in developed countries, with no overall consensus on whether it is beneficial or detrimental to future student achievement. However those who claim that tracking is beneficial to students have several arguments. For example the reduced skill differential has the potential to allow for better lesson execution and time allocation by teachers; they can focus on teaching at a level that will benefit all students in the class,

rather than having to cater to a wide range of abilities. Further, it is possible that ensuring that students are placed in classes of the appropriate difficulty levels will reduce behavioral outlashing by students. Finally, proponents argue that the “value-added” is maximized within each group.

In contrast, critics of tracking argue that in the absence of tracking, high performing students are given the opportunity to practice synthesis by teaching low performers. Further, they pose the idea that tracking is a self-fulfilling prophecy; students in the low stream will achieve lower than they otherwise might, simply because they have been put in a class that implies that they are not as able. This idea is reinforced by the argument that teachers may require less of students in the low stream, and thus that the education gap will widen between the high and low achieving groups.

Through our analysis of the data provided by Duflo et al., we will provide data-driven evidence for several of these arguments, and present our position on the effects of tracking on disadvantaged schools in Kenya. In particular, we aim to answer whether 1) there is value-added by introducing tracking, 2) whether tracking has a differential effect on different strata of students, and 3) whether tracking creates a gap between students of approximately average ability, since these students can be considered to be randomly assigned to the low and high streams.

3 Study Design

The study conducted by Duflo et al. involves data from a randomized experiment spanning 18 months involving the first grade class from 210 primary schools in Western Kenya. These schools have a combined 21,000 students and prior to the experiment each school has only a single first grade class taught by a centrally-hired teacher with civil service protection (hereafter referred to as a “civil service teacher”). The study involves several layers of randomization which are summarized in Figure 1. The “Extra Teacher Program” (ETP) provided funds to 140 schools randomly selected from the pool of 210 schools to hire an extra teacher for first grade classes. These teachers were hired locally, and earned approximately a quarter of the civil service teachers, but had the same academic qualifications. In 70 of these 140 ETP schools tracking was introduced (these schools are “tracked” schools), whereby the two classes were divided by initial achievement, and the classes were randomly assigned to either a civil service teacher or a contract teacher. In the other half of the ETP schools (“non tracked” schools), students were randomly assigned to either the local contract teacher or the existing civil service teacher. Finally, half of the 70 non-tracked ETP schools and half of the tracked ETP schools were given funds to empower the local school committee to monitor and train teachers (these schools are referred to as the “monitored” schools).

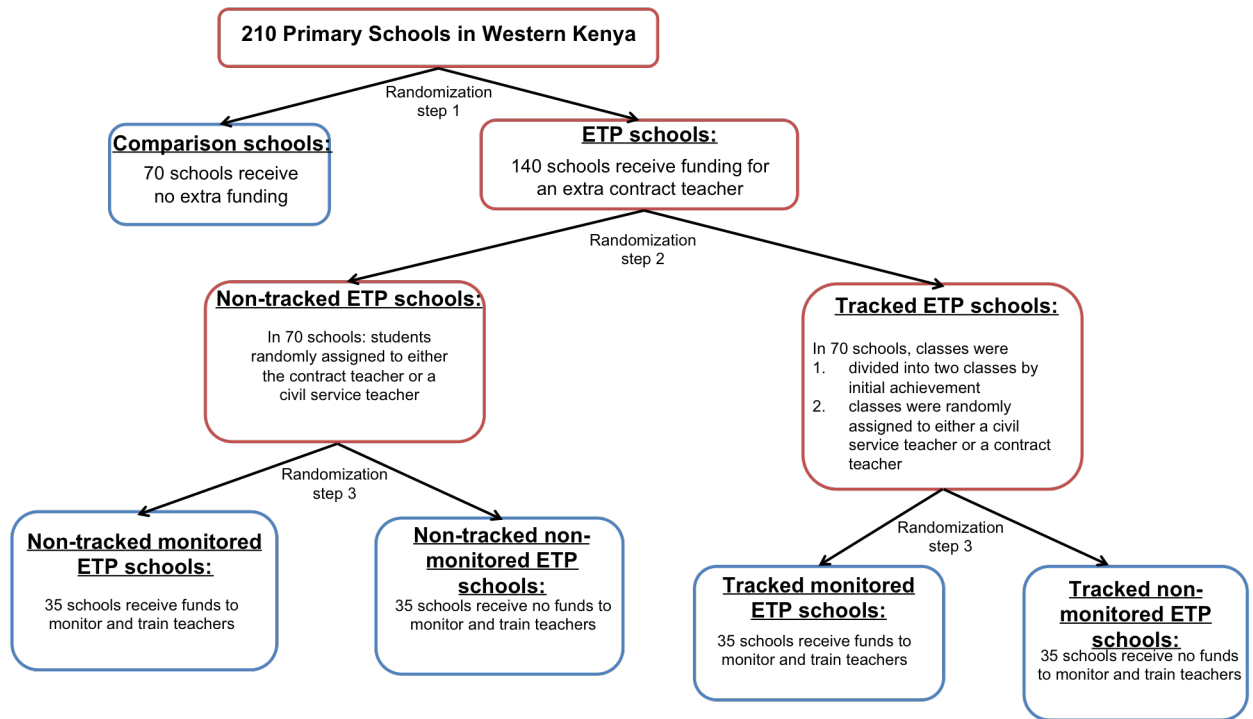


Figure 1: A flowchart describing the randomization steps of the study

Stream assignment in tracked schools was based on initial test scores (baseline test) which were administered locally within schools. Thus these baseline test scores are internally consistent within schools but not comparable across schools. The success of the program was assessed based on scores from a standardized mathematics and language test taken by 60 students from each school (approximately 7000 students all together) after 18 months (the end line test). Another test was also taken by the same students after 24 months (the long-term followup test). We note that the ETP funding ceased after the 18 month endline, so the tracking was no longer in place at the 24 month followup. These tests contained numeracy and literacy questions ranging from counting and identifying letters to subtracting two-digit numbers and writing words.

4 Exploratory Data Analysis

5 Analysis

5.1 Value-added over time by tracking

Unfortunately since the baseline test scores were not comparable across different schools, we were unable to measure the value-added from tracking by comparing baseline with the 18 month endline scores. As a result, we thus chose to measure the long-term value-added by comparing 18 month endline scores with the 24 month follow-up scores. This comparison allows us to see if tracking yielded a lasting effect on students, even after the tracking was no longer in place.

5.2 Stratified permutation tests for comparing tracking with non tracking schools

In comparison to the previous analysis, which involved comparing scores over time for the students in tracking schools, we now turn to a comparison of scores achieved by students in tracking schools with students in non-tracking schools.

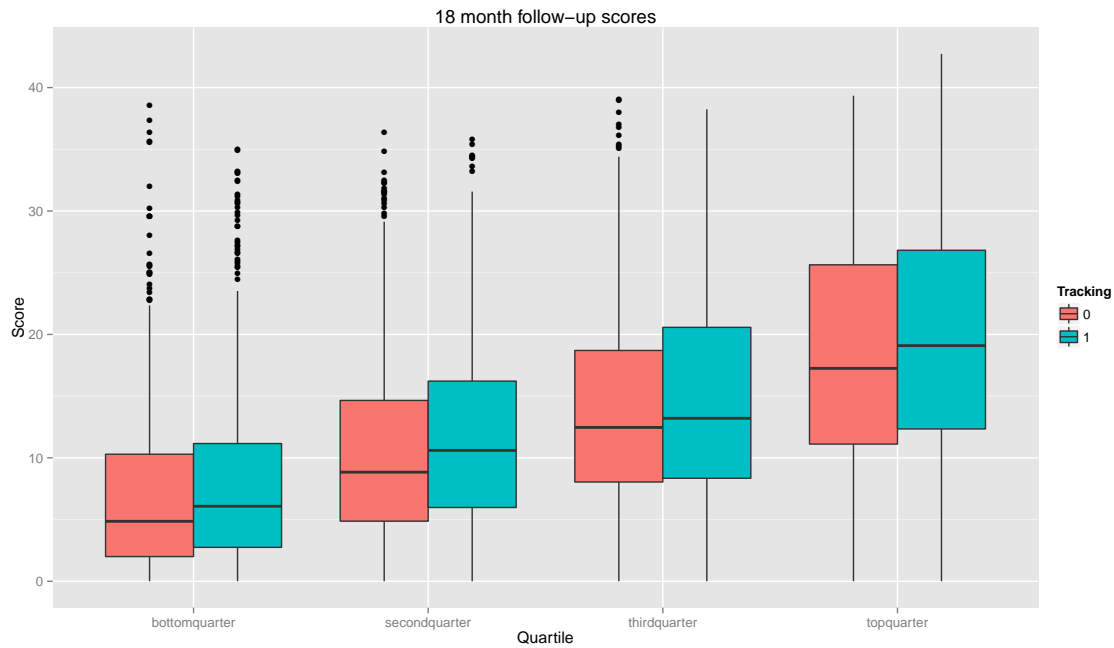


Figure 2:

	Bottom quarter	Second quarter	Third quarter	Top quarter	Overall
Difference in means	0.963	1.318	0.798	1.425	1.124
t	2.466	2.998	1.690	2.810	2.493
P-value	0.013	0.003	0.091	0.005	0.000

Table 1: Test for differences in final score, stratified by baseline quartile and overall.

	Bottom quarter	Second quarter	Third quarter	Top quarter	Overall
Word Score	0.048 (0.963)	0.542 (0.597)	1.017 (0.319)	1.084 (0.272)	0.661 (0.188)
Sentence Score	-0.297 (0.770)	-0.761 (0.450)	0.572 (0.560)	1.173 (0.235)	0.161 (0.746)
Letter Score	2.988 (0.002)	3.389 (0.001)	1.924 (0.050)	4.466 (0.000)	3.191 (0.000)
Spelling Score	1.290 (0.197)	1.111 (0.268)	0.740 (0.450)	1.689 (0.093)	1.210 (0.015)
Literacy Score	1.504 (0.135)	1.530 (0.131)	1.291 (0.202)	2.405 (0.015)	1.680 (0.001)
Math Score	2.869 (0.005)	3.878 (0.000)	1.718 (0.086)	2.508 (0.011)	2.750 (0.000)

Table 2: t-statistics and p-values for the test of differences in subject-level final score between tracking and non-tracking schools, stratified by baseline quartile and overall.

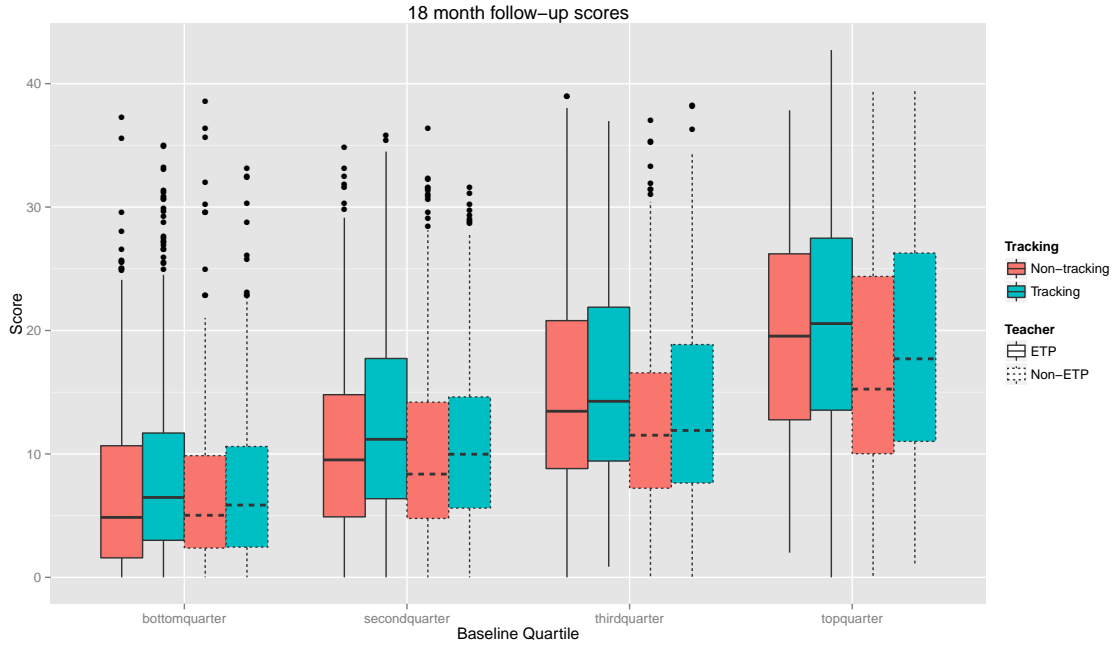


Figure 3:

	Bottom quarter	Second quarter	Third quarter	Top quarter	Overall
ETP	2.506 (0.011)	3.695 (0.000)	0.759 (0.459)	1.918 (0.061)	2.254 (0.000)
Non-ETP	0.817 (0.421)	0.522 (0.611)	1.906 (0.061)	2.404 (0.017)	1.423 (0.004)

Table 3: t-statistics and p-values for the test of difference in mean follow-up exam score between students in tracking vs non-tracking schools, stratified by ETP vs civil servant teachers.

5.3 Regression discontinuity for comparing students in the high stream with those in the low stream

We now aim to analyze the effect of stream assignment on students in tracked schools who initially achieved near the cutoff point. This analysis is based on the idea that students who achieved near the cutoff point have approximately equivalent abilities and it was random noise that lead to their assignment into either the high or low stream. For example, if a student slept poorly the night before the baseline test, they may have scored a few points below the cutoff point, whereas they may otherwise have scored a few points above the cutoff point. As a result, we can consider the students whose initial scores fall within a small window of the cutoff point as being randomly assigned into either the high or low stream.

6 Conclusion