IDinsight
DATA. DECISIONS. DEVELOPMENT.

Evaluating the impact of STiR's model on teacher motivation in U.P. and Delhi:

Pre-Analysis Plan

Version 20160525

IDinsight
DATA. DECISIONS. DEVELOPMENT.

Contents

Overview

25th May, 2016

STiR – SIEF

Version 3 – 2016.05.25

This pre-analysis plan is being prepared after the Teacher Motivation midline survey has been administered but before analysis has begun. This document provides a detailed plan for the midline teacher motivation analysis. Prior to this round of midline three baseline components – student testing, classroom observation, and teacher motivation – have been completed in Uttar Pradesh and Delhi. This was done in two rounds – February to April 2015 and July to October 2016.

Project Name: Impact of Non- Financial teacher incentives, India

Client: STiR Education, World Bank/SIEF

Project Team

- Rajkumar Sharma, *Field Manager*
- Tarun Kumar, *Field Manager*
- Girish Tripathi, *Field Manager*
- Richa Verma, *Former Associate*
- Harlan Downs-Tepper, *Former Associate*
- Tauseef Shahidi, *Associate*
- Varun Chakravarthy, *Associate*
- Stuart Shirrell, *Former Manager and current Tech Team Lead*
- Dr. Heather Lanthorn, *Senior Manager*
- Ron Abraham, *Partner*

Objective

STiR has partnered with IDinsight to undertake an evaluation of the effects of STiR's model as well as the efficiency in its implementation in Delhi and Uttar Pradesh.

The study involves two three-arm, stratified cluster-Randomized Controlled Trials (RCTs). Each study is currently planned for two academic years. One study is sited in Affordable Private Schools in East Delhi and the other is in Government schools in Uttar Pradesh.

Evaluation Background

Intervention description

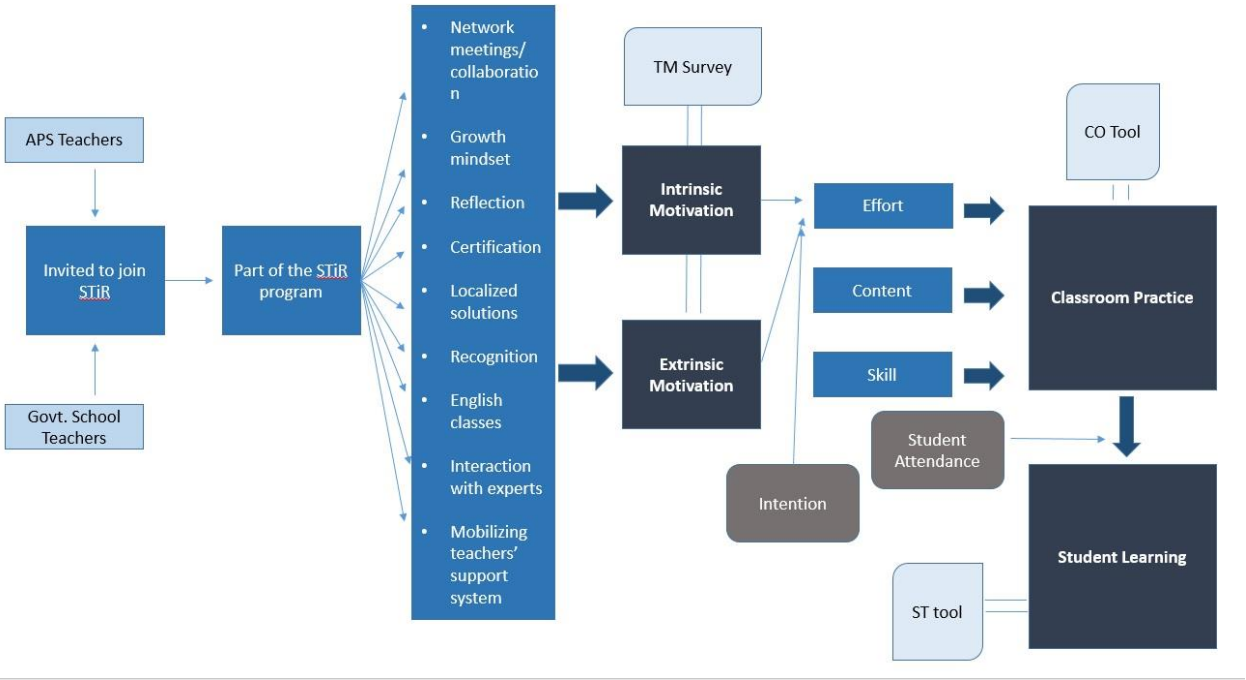
STiR, a non-governmental organization, focuses on helping teachers become central agents of change to overcome a core crisis of learning in low- and middle-income countries: children are increasingly enrolled in school but not learning.

STiR seeks to boost teacher motivation and improve teaching practices and classroom culture in order to boost student learning outcomes in government and private schools. STiR provides opportunities for teachers to share their experiences, challenges, and learnings with each other, as well as prospects for special recognition. Broadly, IDinsight's evaluation of STiR tests two possible models: an "intrinsic motivation" treatment arm facilitates teachers' monthly meetings, skill-building and sharing, while a second treatment arm builds on the first treatment arm with additional opportunities for external recognition and sources of motivation.

STiR's Theory of Change

STiR emphasizes the inherent ability of teachers, regardless of the pedagogies they choose: they are the experts in their classrooms, experienced with the types of issues teachers in similar schools may face. STiR seeks to improve teachers' motivation by organizing them as part of local collaborative teacher 'changemaker' networks. By inculcating among teachers the mindset to collaborate with peers and find localized solutions to overcome the challenges they face, STiR believes they can motivate teachers to bring about a change in their classrooms.

This positive motivation, coupled with the pedagogical techniques teachers share with each other, will adjust the ways in which teachers spend their time in the classroom. In turn, with improvements to teaching, student learning outcomes are expected to improve.



Unit of assignment

The unit of treatment is the school, as all teachers within a given school are allocated to either control or treatment. 'Treatment' reflects being invited to join STiR by being part of a selected school; individual teachers then opt to become actively part of the STiR program.

Further, division into treatment arms 1 or 2 is determined at the cluster level.

Evaluation objectives

This evaluation consists of two similar but separate impact evaluations in Delhi and Uttar Pradesh. Both impact evaluations have the same objectives within their respective geographies.

Primary objectives

1. To examine the impact of STiR on teacher motivation in each geography.
2. To examine the impact of STiR on teaching practices in each geography.
3. To examine the impact of STiR on student Hindi competency in each geography.
4. To examine the impact of STiR on student math competency in each geography.

Secondary objectives

5. To examine the spillover impact of STiR on non-participating teachers in each geography.

This pre-analysis plan focusses on only the **first primary objective of the study – the impact of STiR on teacher motivation.**

Data collection

Identification and sampling strategy

Delhi

This study is a stratified, cluster-randomized evaluation. 180 Affordable Private Schools were sampled in East Delhi. These 180 schools were organized into 7 Education Leader-led networks. In Delhi, Education Leaders (ELs) are STiR staff. A third of schools within each EL network were assigned to the control arm. ELs formed the remaining schools into 4 clusters based on geography. Two of these clusters were then assigned randomly to the intrinsic treatment arm and the remaining two were assigned randomly to any one extrinsic arm.

At baseline all teachers in a school were part of the sample for the teacher motivation survey. A total of 1259 teachers were surveyed and formed the sample frame.

This list of all teachers formed the master list for our midline sample. From this list a total of 453 teachers dropped out of the schools during the course of the academic year and hence were not available for surveying at midline. A further 65 teachers refused to participate in the data collection and 84 were not available during the survey period. Given this, the total number of teachers surveyed at midline (for whom baseline data are also available) was 657. These teachers form the sample used for analysis.

Uttar Pradesh

This study is a stratified, cluster-randomized evaluation. Of three possible districts in Uttar Pradesh non-randomly selected by STiR, IDinsight chose the two districts which were most different from each other to help maximize learning. Within these districts, IDinsight randomly selected 16 clusters, conditional on the cluster having least 15 schools per cluster. Half of all clusters were randomly allocated to control, and half were allocated to treatment. IDinsight then sampled roughly 270 schools from the two districts, allocating one-third of schools from each of the two arms to control. In Uttar Pradesh, education leaders are government school teachers themselves and hence their schools were allocated to treatment.

Enumerators visited schools up to three times during the Teacher Motivation Survey baseline to attempt to survey all teachers affiliated with the school. A total of 1145 teachers were surveyed at baseline.

This list formed the master list for the midline sample. From this list, 288 teachers dropped out of schools during the course of the academic year. These teachers were not available for surveying during the midline

survey. An additional 61 teachers refused to participate in the data collection and a further 41 were not available in schools during the survey period. Given this, out of those teachers for whom baseline data are available, 755 teachers were surveyed successfully at midline.

Unit of analysis

For the teacher motivation survey the primary unit of analysis is the teacher.

Rounds of data collection

Two rounds of teacher motivation surveys have been conducted thus far in both Delhi and U.P.

- **Year 1 baseline Teacher Motivation Survey** was conducted in February 2015 in Delhi and Uttar Pradesh, prior to the beginning of the 2015-2016 school year.
- **Year 2 midline** was conducted in April and May 2016 in both Delhi and Uttar Pradesh just after the beginning of the 2016-2017 school year.

Year 2 endline is tentatively planned for January and February 2017, just before the end of the 2016-17 school year. This may be subject to minor changes depending on program or study needs.

Construction of the Teacher Motivation Index

To capture motivation levels teachers were requested to fill out a teacher motivation survey. Based on an extensive review of the literature, 15 categories that influenced teacher motivation were identified. These were:

1. Recognition by supervisor / colleagues
2. Student performance
3. Availability of good teaching learning material
4. Job security
5. Creative environment
6. Potential for learning new skills
7. Bearing responsibilities related to school
8. Support from students' parents
9. Own family support
10. Student involvement
11. Colleague support
12. Knowledge about policies
13. Salary
14. Supervisor support
15. Sense of mastery of one's job

Based on piloting, and keeping in mind the time burden to respondents, the first 10 categories were included in the final questionnaire used in the impact evaluation[1]. For each of 10 categories, two types of questions were asked – statement questions which helped capture teachers current situation along the categories and situational questions which helped gauge how teachers value these categories. Within each category, each of the different question types were framed in two ways – positively and negatively. The final questionnaire had a total of 40 questions.

[1] Based on experience from the baseline and further piloting 1 of the category from the baseline was dropped and a new category (colleague support) was added to the final 10 category list.

- **Statement 'Likert' questions:** In these questions, teachers were presented with a statement and were required to code their agreement with the statement along a 4 point Likert scale. The options were:
    - Strongly disagree
    - Disagree
    - Agree
    - Strongly agree

  Numerically, values of 1 to 4 were assigned to the different values from low to high level of agreement -- 1 was assigned to strongly disagree and 4 was assigned to strongly agree.
  An example of a positively framed statement question is:

How much do you agree with these statements?	Strongly disagree	Disagree	Agree	Strongly agree
1.1 My principal/manager praises me for my efforts.		✓		

The negatively framed statement for the same category (Supervisor recognition) is:

1.16 I do not receive any praise from my principal/manager for my work as a teacher.			✓	

- **Situational 'vignette' questions:** In these questions, teachers were presented with a hypothetical situation. Teachers selected from a three-point scale, how their motivation would be effected if they were to ever find themselves in the situation. While the wording of the options is specific to the vignette itself, the three options for all questions broadly look at the same three potential responses of how the vignette character would respond:
    - This situation would have no effect on his/her motivation
    - This situation would have an effect on his/her motivation but not his/her teaching
    - This situation would have an effect on both his/her motivation and teaching

  Numeric values of 1 to 3 were assigned to the options, with 1 representing no effect on motivation and 3 representing effect on motivation and teaching.

The positively framed vignette for the 'supervisor recognition' category was:

A teacher receives praise from his principal/manager frequently. How will this make her feel?

a. This is important but will not make her happy.
b. She will be happy, but this might not impact her performance.
c. She will be very happy and will work harder.

a

Similarly, the negatively framed vignette was:

Saurabh's principal/manager never praises him. How will Saurabh feel?

a. It won't matter much.
b. He will be disappointed, but his teaching will be not affected.
c. He will be disappointed, and his teaching will be negatively affected.

c

The data from all the questions were collapsed to form one index value which has been termed as the teacher motivation index.

For each teacher the 'index' value is computed using the formula below:

$$\frac{Positive\ Index - Negative\ Index}{20}$$

Where the Positive Index term is computed using the formula $\sum_{10}^{i=1}$ *(Positive statement$_i$ * Positive situation$_i$).* Here, Positive statements refer to the statement (Likert) questions worded positively and Positive situation refers to the situational (vignette) questions worded positively. 'i' ranging from 1 to 10 refer to the 10 categories of the teacher motivation survey.

Similarly, the Negative Index term is computed using the formula $\sum_{10}^{i=1}$ *(Negative statement$_i$ * Negative situation$_i$).* Here again 'i' ranging from 1 to 10 reflects the categories of the teacher motivation survey. Negative statement and Negative situation refer to the negatively worded Likert and negatively worded Vignette questions respectively.

In the examples presented above for the supervisor recognition category, using the above formulas the index would be calculated as follows:

Positive index = (2 * 1)

Negative index = (3 * 3)

Index = 2-9/2= -3.5

The same is done for all 9 other categories.

Research Question 1: What is the impact of STiR on teacher motivation?

Changes in the Teacher Motivation Index will be calculated across the entire index and the nine subcomponents common to both baseline and midline[2].

STiR's program is designed to have different treatment groups – with different emphasis on intrinsic and extrinsically motivating factors. The ten categories of the teacher motivation survey will be classified into 'intrinsic' and 'extrinsic' buckets based on review of the relevant literature. Changes in the Teacher Motivation Index will be calculated across these two buckets as well.

Primary estimate: Intention To Treat (ITT)

The unit of assignment of treatment is the school, so all teachers within a school have the same 'treatment status'. The reason to think of the ITT estimator as the primary estimator is as follows:

- **Non-Compliance:** At the teacher level not all teachers in treatment schools choose to join STiRs program; indeed, STIR values the voluntary nature of the program. Given self-selection of teachers into the changemakers networks, non-compliance at the teacher level is natural.
- **Influence:** An explicit part of STiR's model is for participating teachers to continue to try to influence non-participating teachers in their schools in adopting STiR's approach and programming.

[2] For the 'colleague support' category added during midline, t-tests will be run instead between the different treatment arms

- **Policy relevance:** Given the voluntary nature of the STiR program, the ITT estimate is the more policy relevant estimate as STiR looks to scale their program in the future.
- **The overall effect of the program:** The ITT measure looks at the overall effect of the program. If STiR has a large effect per teacher but only a few teachers participate in the program, their overall impact will still be low.

Secondary estimate: Treatment-On-the-Treated (TOT)

To define 'compliers' to the program, attendance data from the network meetings will be used to classify teachers as those who are active and those who are not active members of STiRs program.

The TOT estimate will be computed by dividing the ITT estimate by the rate of compliance.

$$TOT = \frac{ITT}{Complicance\ Rate}$$

Note that computing the TOT estimate this way does not change the p-values and thus the probability of rejecting the null hypothesis. The only effect would be on the coefficient value.[3]

This secondary estimate would be looked into only after thinking through the definition of 'compliers'. STiR has defined compliers as of now based on different cycles and only towards the end of the year. This may make this estimation tough to interpret.

Analytical model

Analysis of Covariance (ANCOVA) will be used to estimate the effect of the STiR treatment on the Teacher Motivation Index.[4]

ANCOVA

STiR's model in Delhi and Uttar Pradesh includes three treatment arms: the control group, the intrinsic motivation group, and the extrinsic motivation group. A treatment indicator will track the teacher's treatment arm – 0 for control; 1 for intrinsic and 2 for extrinsic arms.

The base specification will compare the control, intrinsic, and extrinsic groups without controlling for covariates. An additional specification will include the covariates mentioned in Appendix A.

The specifications mentioned would be run separately for teachers from Delhi and Uttar Pradesh. Given the differences in the two contexts and differences in implementation of the program between Delhi and Uttar Pradesh, the two geographies will be treated as distinct.

Apart from the above, a

Standard errors will be clustered at the school level.

Formula

Base specification:

Midline Teacher Motivation Index = $\beta_0 + \beta_1$*(vector of treatment) + ε_i

[3] Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. "Identification of causal effects using instrumental variables." *Journal of the American statistical Association* 91.434 (1996): 444-455.

[4] Given that we have a baseline with two follow-ups, we expect that ANCOVA will give more power than differences-in-differences, although the difference in power may be smaller because of high autocorrelation. See McKenzie, David (2012) Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics* **99**210-221.

Where, corr(ε_{ij} , ε_{kl}) = 0 if l \neq j and rho if l == j

Controlling for covariates:

Midline Teacher Motivation Index = β_0 + β_1*(vector of treatment) + β_2*(baseline index) + β_3*(teacher sex) + β_4*(teacher age) + β_5*(teacher qualification) + β_6*(number of years of experience) + β_7*(district dummy variable) + β_8*(enumerator dummy variable) +ε_i

Sex, age, qualification, experience, district, enum dummies

Stata code

Base:

```
regress indexMidline i.assignment, vce(cluster schoolCode)
```

Controlling for covariates:

```
regress indexMidline i.assignment indexBaseline sex age qualification
experience district enumerator, vce(cluster schoolCode)
```

The above specifications would be run for three main treatment (assignment) types:

1. STiRs own assignment: This comprises of the control, intrinsic (1.0) and extrinsic (2.0) arm
2. Simple treatment vs control: Intrinsic (1.0) and extrinsic (2.0) will be clubbed
3. Sub groups: Mentioned in detail in subsequent sections.

Sample frame

As noted in previous sections, in both Delhi and Uttar Pradesh, only a subset of teachers who were surveyed at baseline were still in the schools at midline. In Delhi we have a total of 657 teachers and in Uttar Pradesh we have 759 teachers for whom we have both baseline and midline data.

Before finalizing on the analytical model to be used we checked for differential attrition[5] across treatment arms within each geography. The regression fitted on STATA is as below:

```
reg dropoutStatus i.treatmentStatus, vce(cluster schoolCode)
```

where dropoutStatus is a binary variable which is 1 if the teacher has dropped out or 0 if the teacher is still part of the sample. A 5% level of significance was pre-decided. The results were as follows:

Region	Num obs.	Model-df	Reg-df	F-Statistic	Prob>F
Delhi	1,249	2	179	0.4	0.6682
U.P.	1,142	2	270	2.44	0.0893

Similarly, a test was run for checking if drop-out correlated with baseline Teacher Motivation Index scores. The regression run on STATA is below:

```
reg dropoutStatus index, vce(cluster schoolCode)
```

[5] In case of differential attrition the Lee trimming method would be used

where dropoutStatus is a binary indicator, as before, and index is the baseline teacher motivation index. A 5% level of significance was pre-decided as before. The results are shown as below:

Region	Num obs.	Model-df	Reg-df	F-Statistic	Prob>F
Delhi	1,248	1	179	3.86	0.051
U.P.	1,142	1	270	0	0.9851

At the 5% level of significance we did not reject the null hypothesis of no correlation in both cases[6]. Thus, those teachers who dropped out between baseline and midline were dropped to reach our final sampling frame.

Sample and subgroups

The research question of the impact of STiRs model on teacher motivation will be answered for the **whole sample,** and will also be answered for the following subgroups.

- **Treatment sub-arms:** Apart from the three broader arms, the STiR program has different flavors of their extrinsic treatment arm. There are 4 extrinsic arms in Delhi and 3 in Uttar Pradesh. Two of the extrinsic packages are common to both geographies. Learning about these different approaches to programming was a central goal for STiR.
- **Baseline motivation:** Using the baseline teacher motivation index, teachers will be classified into three categories. This sub-group analysis has potentially important learnings for STiR. Teachers who are inherently more motivated may be more driven to be an active participant in the STiR program. They may also be naturally more eager to adopt what they learn via network meetings in their classrooms. For STiR to achieve their long term targets it is important that they successfully impress upon 'not so' motivated teachers as well.
- **Teacher sex:** While in Delhi more than 90% of the sample of teachers are female, in Uttar Pradesh the proportion of male and female teachers in the sample are similar. Whether their programing has differential impact for male and female teachers has always interested STiR. Differential effects for male and female teachers — who may face different incentives and constraints in participating activievly in STiR and being able to enact ideas from STiR in the classroom – will be examined in U.P.
- **Teacher experience:** Most of the literature suggests that the transformation from a novice teacher into a teacher with 'more experience' happens generally after 3 years of having been a teacher[7]. Creating a binary for teacher experience at the third year threshold, impact would be looked at for 'more' and 'less' experienced teachers. More experienced teachers may be more set in their ways, and therefore less willing to act on STiR's approach, but they may also be more in need of 're-motivation' and may also be better placed to put STiR's ideas into action.
- **Network health:** (WORKING DEFINITION WHILE STiR SENDS FINAL LIST) STiR uses a number of parameters to help define network health, an indicator of how well a particular changemaker network is functioning. These parameters include teacher attendance and retention, teacher engagement, certification, Education leader engagement, Program Manager support to the Education Leaders and support from the government officials. Network health may act as an

[6] For differential attrition -- H_0: There is no differential attrition across treatment groups
For correlation with baseline index numbers: H_0: The attrition is not correlated with baseline index numbers
[7] Bruns, Barbara; Luque, Javier. 2014. *Great teachers : how to raise student learning in Latin America and the Caribbean.* Washington, DC: World Bank Group. http://documents.worldbank.org/curated/en/2014/01/19798994/great-teachers-raise-student-learning-latin-america-caribbean

important indicator to how conducive a particular network is to teachers collaborating, learning and applying those learnings in their classrooms.

Formula

The regression equation will be similar to that done for the main analysis. The assignment variable would now be replaced by an interaction term between the assignment variable and the variables for the above sub-groups.

The covariates would be the same as that for the main regression (Appendix A). The exception in this case would be if the covariate is in itself the sub-group of interest. Eg: When we look at the gender sub-group, we will not use gender as a control in our equation.

Base specification:

Midline Teacher Motivation Index = $\beta_0 + \beta_1$*(i.assignement##i.subgroup) + ε_i

Controlling for covariates:

Midline Teacher Motivation Index = $\beta_0 + \beta_1$*(i.assignement##i.subgroup) + β_2*(baseline index) + β_3*(teacher sex) + β_4*(teacher age) + β_5*(teacher qualification) + β_6*(number of years of experience) + β_7*(district dummy variable) + β_8*(enumerator dummy variable) +ε_i

Stata code

Base:

```
regress indexMidline i.treatment##i.subgroup, vce(cluster schoolCode)
```

Controlling for covariates:

```
regress indexMidline i.treatment##i.subgroup indexBaseline sex age
qualification experience district enumerator, vce(cluster schoolCode)
```

Level of significance

Due to the multiple tests being run, correcting the p-value at which significance is defined:

- **Bonferonni correction:** Given that there will be relatively few outcomes of interest considered here, the Bonferri p-value correction procedure will only be used on the nine subcomponents of the motivation analysis. That is, the nominal alpha value of 0.05 will be used for al analyses except the motivation subcomponents, where the alpha value will be 0.05/9 = 0.0056.
- **Overall significance tests:** For each subgroup analysis, the overall F-test will be presented for the model containing only the interaction terms and the treatment groups. The F-test will help determine the overall impact on the groups in question. In addition, an omnibus randomization inference test of statistical significance will be used to measure the overall impact across treatment indicators.[8]

[8] For more details on the omnibus randomization inference test, see: Young, Alwyn. (2016). "Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *Working Paper.* Retreived June 1, 2016 from http://personal.lse.ac.uk/YoungA/ChannellingFisher.pdf

Robustness checks

- **IRT:** Item response theory refers to a family of latent trait models used to establish psychometric properties of items and scales. We will use item analysis to gauge how well the tool measured the underlying ability or trait.
  Given the structure of our questionnaire, the graded response model will be used. Graded response models are the IRT models used to analyze ordinal responses and rating scales. The IRT model will be used to predict the latent 'motivation' for each teacher. Then we will check the correlation between the latent motivation and the index values.
  The irt grm command is the default command in STATA which fits graded response models to ordinal items. In the GRM, items vary in their difficulty and discrimination.

  Stata code:
  Defining the irt model
  ```
  irt grm q11* q12* q21* q22*, vce(cluster schoolCode)
  ```

  Predicting the latent 'motivation'
  ```
  predict motivation, latent
  ```

  Running the correlation between the motivation and the index values
  ```
  corr motivation index
  ```

  Presenting it as a scatted plot
  ```
  scatter motivation index
  ```

Descriptive statistics

The teacher motivation questionnaire is broadly divided into three main sections. The first two sections are common to both the baseline and the midline questionnaire and are used in index computation. The third section was added during the midline survey.

The third section was not meant to be used during the index computation. It was added to cover a few themes important in STiR's program (such as growth mindset) or themes that came up as important indicators of teacher motivation from our experience with the process evaluation (such as benefitting students).

- **Section 1:** This section comprises of all the Likert-scale statement questions used to rate teacher experiences for index computation.
- **Section 2:** This section comprises of all the vignette style situation questions used to rate teacher valuation for the index computation.
- **Section 3:** An additional section with a combination of Likert and vignette questions. The broad themes touched upon in this section are growth mindset, benefit to students, self-reported motivation and additional administrative paperwork (only in Uttar Pradesh).

The descriptive statistics section will look to build upon teacher responses from all sections mentioned above. A broad plan of the descriptive statistics is as follows:

- **Section 1:** Answers to all the questions would be presented in a table format to help showcase the proportion of teachers with different levels of 'agreement'. The table would look as below:

Statement	Strongly disagree (%)	Disagree (%)	Agree (%)	Strongly agree (%)	Refusals (%)	Total
My principal/manager praises me for my efforts						100%

- **Section 2:** For the vignette style questions, the responses will be displayed as stacked bar graph. The x-axis will include each of the sub-categories of the teacher motivation index; there will be 10 bars in total. The y-axis will represent the proportion of teacher responses to each of the three possible answer options (no change in motivation, change in motivation but not teaching, change in motivation and teaching).
- **Section 3:** The Likert statements in section 3 will be presented as tables in the same format as that in section 1. Apart from this, a few questions will be analyzed separately:
    - **Motivation to 'be a teacher':** Teachers were asked how happy they would feel if they were given a 'non-teaching' job or responsibility by their supervisors, when all other considerations, such as pay, stay the same. Teachers were asked to answer along a 4-point scale ranging from very unhappy to very happy. STiR's program looks to rekindle the spark in teachers and make them feel reconnected to their duty as teachers. The responses to this question will be displayed graphically, using stacked bar graphs by treatment arms.
    - **Self-reported motivation:** Teachers were asked two questions of self-reported motivation – how motivated they felt as a teacher and if this is as motivated as they could be. The answers to the two questions would be presented using a spineplot to visualize the conditional analysis.
- **Teacher motivation index:** Apart from the impact analysis, the teacher motivation index will also be used for descriptive analysis.
    - **Distribution:** A histogram would be generated to represent the distribution of the index value across all the data and by treatment arm.
    - **Box-and-whisker plot:** A box-and-whisker plot will be used to show the distribution of the index values for each of the three broad treatment arms. This will depict the data through quartiles and the lines would help represent the variability outside of the highest and lowest quartile.
    - **Group wise average index value over time:** Such a graph is typically presented along with a difference-in-differences analysis framework. We have teacher motivation index numbers for two time points (baseline and midline), one year apart. Point estimates for the average index value for the three broad treatment arms across both time periods will be plotted. A line would be used to connect the points for the same treatment arms to show the trend over the year.
- **Summary statistics of the sample frame**: A table will describe the final sample frame. This would contain quick summary statistics of the demographic characteristics of the teachers in the final sample frame such as average age, median qualification, gender distribution etc.

Presentation of results

Presentation of the descriptive statistics has been described in detailed above. The results of the impact analysis would be presented as a combination of tables and graphs as appropriate:

- **Main regression results:** The main regression results would be displayed in tabular form.
- **Sub-group analysis:** The sub-group impact analysis would be presented in tabular form, with one table for all categories of a particular 'group'.

- **Category-wise index:** The movement in category wise index would be presented in a table will as the other regression results. The results will also be displayed graphically using a forest plot.

Appendix A: Covariates panel A

The control variables mentioned here in Appendix A are those variables which we are not interested in directly but we expect are related to the dependent variable (in this case teacher motivation) or may have differential experiences as part of the STiR program. We will include these in the regression equation as the independent variable to 'control' for their effect.

1. **Teacher sex:** Male and female teachers may be motivated by different factors or conversely different things may motivate male and female teachers differently. Further the access to, experience with the STiR program and the interest and ability to act on STiR's ideas may vary with sex.
2. **Age:** Teachers of different ages may find different barriers in being active members of STiR. They may be more susceptible to pressures from family, colleagues and supervisors. They may also have less decision making power while trying to influence the classroom culture or practice. At the same time the desire to collaborate with peers may be more exciting to younger teachers.
3. **Teacher qualification level:** Teachers in schools have varied backgrounds and training. This may influence a teachers' ability to influence her classrooms in general, and specifically with regards to learnings from STiRs programs. A teacher with higher training or qualification may also be better at finding localized solutions to the challenges in the classrooms and may be more 'active' participants in network meetings.
4. **Total number of years teaching:** A teacher's experience may influence how they 'value' STiRs program. It may be the case that teachers younger in their teaching career are more incentivized to be a part of the STiR program that someone who is further ahead in their career or closer to retirement. At the same time, it may be that slightly more experienced teachers are better aware of the challenges specifically in their classrooms or in general. They may be able to use the STiR experience in a more fruitful manner.
5. **District:** Different districts have different education officials; district officials play an important role in providing space (including physical meeting space) for STiR to operate and in encouraging ELs and teachers to participate and maintain enthusiasm. Different officials, therefore, may not only influence the experience of an individual as a teacher but would also impact a teachers' experience with STiR. If the district officials are more supportive of STiR, the teachers may have a better experience as part of the program by participating more actively in STiR as well as being able to enact STiR's idea in the classroom.
6. **Enumerator dummies:** It is important to control for enumerator dummies, to prevent for any enumerator specific biases during data collection. While the motivation questionnaire is self-administered, bias may creep in due to an enumerators communication and explanation skills.
7. **Baseline teacher motivation index:** A teacher's inherent motivation may be an important determinant of how actively they are a part of STiRs network. STiRs program requires teachers to spend time and effort outside of classrooms in network meetings and to find solutions to existing problems. Along the way they may have to face barriers of different kinds which may act as disincentives eg: travel, pressure from head teachers, family etc. If a teacher is highly motivated, she would be more likely to overcome these hurdles or may be more excited by the opportunity to learn.
8. **Dummy variable for network:** Teachers are organized into local changemaker networks in which they interact with teachers from other schools in the same network. Each network is led by one EL and contains schools with geographical proximity to one another.