Evaluation Design Validation Report: Quantitative Assessment of Teacher Motivation, Classroom Practices, and Student Learning

NOVEMBER 2015

Prepared by:
Ronald Abraham
Stuart Shirrell[1]
Harlan Downs-Tepper
Varun Chakravarthy

[1] Please direct all correspondence regarding this report to stuart.shirrell@idinsight.org.

Table of Contents

SIEF Approval

In 2014, SIEF approved an impact evaluation of a new project in India, titled "Impact of Non-financial Teacher Incentives," that seeks to improve student-learning outcomes by increasing teacher motivation via intrinsic and extrinsic motivation packages.

Implementation Partners and Investigators

Proposal

The evaluation and intervention was designed in partnership with STIR education to evaluate its teacher networks as a means of improving teacher motivation in affordable private schools (A.P.S.) in Delhi and government schools in Uttar Pradesh (U.P.). The principal investigators for this study are Dr. Neil Buddy Shah, Andrew Fraker, and Ronald Abraham. The co-principal investigators for this study are Sangeeta Goyal and Dr. Lant Pritchett.

Actual

This has not deviated substantially from the proposal. IDinsight has continued to work with STIR throughout the evaluation and all principal investigators have been involved in the design of the study.

Objective

Proposal

The primary objective of this impact evaluation is to generate evidence on the efficacy and effectiveness of nonmonetary intrinsic and extrinsic motivators as a way to improve teaching and learning.

Actual

The actual objective has not changed from the proposal.

Intervention Design

Proposal

The original proposal included the following sample sizes for Delhi and U.P.:

- Delhi: 180 schools, 3 teachers per school, 10 students per classroom
- Uttar Pradesh: 180 schools, 3 teachers per school, 10 students per classroom.

Each study was to have the following treatment arms, which were to be randomly assigned at the school level:

- **Pure control group**, where STIR would have no contact with teachers or staff.
- **Intrinsic motivation group**, where STIR would run the teacher network program it has been refining for the past three years.
- **Extrinsic motivation group**, where STIR would run its teacher network program, but would also implement other program elements designed to improve teacher motivation through non-network means.

Actual

The actual implementation has been very similar to the proposed design. STIR has implemented its teacher networks in both Delhi and U.P. as planned. There have, however, been some slight modifications to the original proposal:

- **Extrinsic motivators have been broken into "packages"**: On the advice of Lant Pritchett, a co-PI on this study, STIR and IDinsight decided to break the set of extrinsic motivators into smaller packages following common themes.2 These individual packages were assigned to networks in Delhi and U.P. The rationale behind this change was that this allows STIR to learn about what might be working well and potentially recalibrate between the midline and endline surveys, if necessary. Breaking the extrinsic motivators into packages also likely improves implementation quality by decreasing the number of activities implementation staff have to conduct with each network.
- **Sample size has increased in U.P.**: Based on some data obtained from STIR's M&E activities and other sources, IDinsight revised its power calculations for the study in U.P. Based on these revised power calculations, IDinsight decided to increase the sample size in U.P. from 180 schools to 270 schools, split evenly across the three treatment arms.
- **Placebo control schools in Delhi**: In an effort to improve survey access to control schools in Delhi, STIR has designed a placebo intervention for these schools, including things like a newspaper subscription or health clinics. IDinsight does not believe these will substantively impact teacher motivation, but have proved instrumental in providing access to control schools.

---

[2] The intervention now includes the following motivation packages:

- Career and personal development
- Head teacher recognition and development
- Local recognition
- Teacher exposure

- **Some schools have dropped out or refused survey in Delhi**: While the STIR team in U.P. has high-level government buy-in and can work through existing government structures, the STIR team in Delhi has no such luxuries. Instead, the STIR team has had to require APS schools one by one, and APS schools are often reticent to work with outside groups. Moreover, due to recent government pronouncements about the closing of APS schools failing to meet Right to Education (RTE) requirements, schools have become more resistant to enumerators entering their schools.[3,4] As a result of these developments, 11 schools have dropped out of the STIR program and a further 19 schools have refused to allow IDinsight to conduct surveys, bringing the final count for baseline schools to 141. While these schools may not continue with the STIR program, IDinsight will continue to follow up with these schools for the midline and endline surveys.

Data Collection

Proposal

The proposal originally entailed three rounds of data collection: a baseline survey in early 2015, a midline survey in 2016, and an endline survey in 2017.

Actual

The baseline survey was split into two different parts: the teacher motivation survey and the classroom practice and student testing survey. The teacher motivation survey was conducted in February and March, 2015. This was done first to avoid any contamination from STIR's program, as teacher motivation is likely the first outcome indicator to show any impact from STIR's program. Due to the nature of this survey, it is also much less expensive to administer than the classroom observation or student testing survey. For the teacher motivation questionnaire, all teachers gave written consent.

The classroom observation and student testing took place in July through October 2015. This survey was done after the school break from mid-May to early July to minimize the amount of attrition between baseline and midline surveys, as the break is often when students transfer or teachers transfer or leave the teaching profession altogether. For the classroom observation and student testing survey, teachers gave verbal consent for the

[3] Affordable private schools in Delhi have come under threat of government ordered closure due to inability to meet some of the quality and infrastructure standards of the Right to Education Act. See, for example, http://www.thehindu.com/news/national/300-private-schools-under-delhi-govt-scanner-for-flouting-norms/article7670628.ece. This may cause some schools to be more hesitant in sharing information with external parties.

[4] The Right to Education Act is an act of Indian parliament passed in 2009 that guarantees the rights of all children to free primary education and sets out education standards for nongovernmental schools. For full text of this act, see, for example http://eoc.du.ac.in/RTE%20-%20notified.pdf (retrieved 11 November 2015).

classroom observation and gave *in loco parentis* verbal consent for the student test, though students were also allowed to refuse taking the test. For both Delhi and U.P. studies, in classrooms where there were more than 10 students, 10 students were randomly selected from the set of students present on the day of data collection. If fewer than 10 students were present, then all students were tested.

In U.P., 1,147 teachers completed the teacher motivation questionnaire, 841 teachers were observed using the classroom observation tool, and 7,385 students were tested. The 841 teachers were selected from the original list of 1,147 via stratified random sampling, with stratification at the school level. Schools with fewer than three teachers had all of their teachers selected.

In Delhi, 1,260 teachers completed the teacher motivation questionnaire, 346 teachers were observed using the classroom observation tool, and 3,379 students were tested. From the original list of 1,260 teachers, STIR screened the teachers before knowing if a school was in the treatment or the control group. This gave a list of 811 teachers, from which IDinsight sampled 540 teachers for the classroom observation and student testing baseline survey. Due to the considerations noted above and significant teacher attrition, the number of classroom observation surveys is lower than the originally-planned 540.

Program Assignment to Treatment and Control

Proposal

Schools would be randomly assigned to one of the three treatment arms. No stratification or clustering was specified.

Actual

There have been some slight modifications to the randomization strategy in both Delhi and Uttar Pradesh:

- **Uttar Pradesh**: Schools in U.P. are organized into clusters of roughly 10-25 schools. STIR prefers to have one teacher network per cluster, as this makes working through administrative structures easier. The original plan of randomly allocating schools to treatment arms was thus logistically infeasible. The randomization instead proceeded in two steps: first, clusters were randomly assigned to be either intrinsic or extrinsic clusters; second, one-third of schools in each cluster were assigned to be control schools. This randomization strategy does not negatively impact power when comparing treatment arms to the control arm. It may, however, negatively impact the ability to compare treatment arms to one another, though the proposal noted that the study would likely lack sufficient power to make these comparisons even with school-level randomization.

  As noted above, IDinsight and STIR had decided to break the extrinsic

motivation intervention into smaller bundles. In U.P., these bundles were applied at the school cluster level. Extrinsic motivation clusters were randomly assigned to one of the three motivation packages being implemented in U.P.

In addition to these changes, some schools in U.P. were adjacent to one another. These schools were often primary and upper primary school pairs. In order to minimize contamination or spillovers, IDinsight found all pairs of schools within 300 meters of each other and randomized at the pair level, rather than at the school level. Schools that were not within 300 meters of another school were randomized at the school level. These pairs of schools did not represent a significant fraction of schools in the study sample, and IDinsight does not expect this to have a substantial impact on study power.

- **Delhi**: In Delhi, STIR directly employs the education leaders who run the teacher motivation networks. These education leaders (ELs) each have their own catchment area in east Delhi. While STIR's EL's have the ability to run both intrinsic and extrinsic networks simultaneously, teacher transportation and program spillover was a potential problem in Delhi. As a result, the randomization method was modified from the original plan of randomly assigning schools to different treatment arms. Instead, the randomization followed the following process:

    1. Within each EL catchment area, one third of schools were assigned to the control group.
    2. EL's then took the remaining schools in their respective catchment areas and formed them into four different clusters based on geography.
    3. For each EL, two clusters were randomly assigned to the intrinsic motivation treatment arm and the remaining clusters were randomly assigned to one of the extrinsic treatment arms.

    As with the randomization process in U.P., this does not affect the study's power when comparing treatment arms to the control arm. It will likely negatively impact power when comparing treatment arms to one another, but as noted above, these comparisons were likely underpowered even with school-level randomization.

Program Launch

Proposal

The proposal indicated that the program would launch in early 2015.

Actual

The proposal launched in early 2015 as planned in both Delhi and U.P. Both geographies are currently in the middle of the second network cycle, as per STIR's original plan to finish the first year of the intervention by the end of the academic year in March 2016.

Baseline Balance Checks

The treatment groups in both Delhi and U.P. appear to be well balanced across all treatment groups. 21 variables were tested for Delhi and 20 tested for U.P.[5] F-tests were used to determine the joint significance in difference in means across the three groups. Standard errors were clustered at the school level in all analyses presented.

In the Delhi study, two of 21 variables were significant at the 10% level and no variables tested were significant at the 5% or 1% levels. Theory would predict that roughly two out of 21 variables would be significant at the 10% level by chance, and that one variable at the 5% level would be significant by chance. This seems to indicate that the variables are well-balanced across the various study arms in Delhi.

In the Uttar Pradesh study, three out of the 20 variables were significant at the 5% level and one variable was significant at the 10% level. Two of these variables are fraction of time spend teaching and fraction of time spend off task, which are not independent variables, meaning that imbalance in one is likely responsible for imbalance in both. Along with classroom management, the fractions for each of these variables must add up to 100%. The other significant variables were the fraction of teachers who used learning aids (at the 5% level) and fraction of students engaging in group discussion or Q&A (at the 10% level).[6] While theory would only predict one variable that is significant at the 5% level assuming random distribution, IDinsight believes that the control and treatment groups are nonetheless well-balanced. IDinsight will also take measures to test for bias at endline, for example, by performing robustness checks incorporating baseline covariates.

Future Activities

As noted in the proposal, IDinsight will also be conducting an in-depth process evaluation in late 2015 and early 2016. Thereafter, IDinsight will prepare for the midline survey, which will take place in the second or third quarter of 2016.

[5] Additional teacher qualifications were not recorded at the baseline survey for U.P. These can be collected during the midline survey to take place in 2016.
[6] For the purposes of the survey, a learning aid was defined as anything other than a textbook.

Appendix A: Baseline Balance Checks

Delhi

Variable	Mean Control	Mean Intrinsic	Mean Extrinsic	Num obs.	Model-df	Reg-df	F-statistic	*p*-Value
Teacher Motivation Index	1.9	1.9	2.0	1256	2	178	0.80	0.45
Teacher Age	28.1	28.7	29.5	1252	2	178	0.86	0.42
Teaching Experience (Years)	5.6	5.7	6.5	1248	2	178	1.56	0.21
Female	95%	94%	92%	1257	2	178	1.26	0.29
Teacher Education[1]				1256	2		0.77	0.68
Additional Teacher Qualifications[2]				1251	12		14.08	0.30
Fraction of Time Teaching	73%	64%	71%	1384	2	140	1.92	0.15
Fraction of Time Managing Classroom	26%	33%	28%	1384	2	140	1.51	0.23
Fraction of Time Off Task	1%	2%	2%	1384	2	140	0.96	0.38
Fraction of Students Doing Drills	25%	29%	27%	1383	2	140	0.51	0.60
Fraction of Students Participating in a Group Discussion	17%	16%	15%	1383	2	140	0.16	0.85
Fraction of Students Listening to Lecture*	21%	19%	14%	1383	2	140	2.35	0.10
Fraction of Students Doing Silent Seatwork*	29%	27%	38%	1383	2	140	3.00	0.05
Fraction of Students Off Task	8%	8%	6%	1383	2	140	1.41	0.25
Fraction of Teachers Who Smiled at Least Once	75%	68%	80%	345	2	140	2.21	0.11
Fraction of Classrooms Where At	33%	33%	35%	345	2	140	0.03	0.97

Least One Student Asked a Question								
Fraction of Teachers Who Used Local Information while Teaching	78%	83%	77%	345	2	140	0.84	0.43
Fraction of Teachers Who Used a Learning Aid**	58%	61%	63%	344	2	140	0.34	0.72
Fraction of Teachers Who Asked Students to Work in Small Groups	3%	2%	4%	342	2	139	0.56	0.57
Student's Math ASER Level[3]	3.8	3.6	3.6	3379	2		0.37	0.83
Student's Hindi ASER Level[3]	3.6	3.5	3.5	3379	2		0.67	0.71

Notes: Unless otherwise noted, F-statistics reflect the model specification statistic for a linear regression model with the outcome variable listed on the leftmost column and the explanatory variables as two binary variables indicating either intrinsic or extrinsic motivation treatment status. Unless otherwise noted, all standard errors in this table are clustered at the school level.
* Indicates significance at the 10% level.
** Indicates significance at the 5% level.
*** Indicates significance at the 1% level.
[1] Highest education level is an ordinal variable, so an ordered logit model was used. Because this model uses maximum likelihood, the model test is a chi-squared test, rather than an F-test. Standard errors were clustered at the school level. Group means have not been listed for this model.
[2] Teacher qualification is an unordered qualitative variable, and a Pearson's chi-squared test was used to determine differences in distributions among the different treatment arms. Standard errors are not clustered at the school level, so the p-value listed may be conservative.
[3] Student Math and Hindi scores are from the ASER test battery, which gives an ordinal value. Because this model uses maximum likelihood, the model test is a chi-squared test, rather than an F-test. Standard errors were clustered at the school level. Readers should use some caution when interpreting group means, as the numerical values from one category to the next are arbitrary.

Uttar Pradesh

Variable	Mean Control	Mean Intrinsic	Mean Extrinsic	Num obs.	Model-df	Reg-df	F-statistic	*p*-Value
Teacher Motivation Index	1.8	1.7	1.8	1145	2	270	0.62	0.54
Teacher Age	38.9	38.7	38.4	1222	2	270	0.19	0.83
Teaching Experience (Years)	11.5	10.9	11.0	1214	2	270	0.46	0.63
Female	57%	52%	54%	1244	2	269	0.37	0.69
Teacher Education[1]				1205	2		0.09	0.96
Fraction of Time Teaching**	77%	81%	71%	3369	2	270	3.50	0.03
Fraction of Time Managing Classroom	7%	7%	8%	3369	2	270	0.22	0.80
Fraction of Time Off Task**	16%	12%	21%	3369	2	270	3.84	0.02
Fraction of Students Doing Drills	32%	27%	28%	3349	2	265	1.53	0.22
Fraction of Students Participating in a Group Discussion*	5%	4%	8%	3349	2	265	2.42	0.09
Fraction of Students Listening to Lecture	28%	34%	25%	3349	2	265	2.07	0.13
Fraction of Students Doing Silent Seatwork	16%	19%	18%	3349	2	265	0.94	0.39
Fraction of Students Off Task	18%	17%	22%	3349	2	265	2.18	0.11
Fraction of Teachers Who Smiled at Least Once	6%	4%	7%	841	2	265	1.62	0.20
Fraction of Classrooms Where At Least One Student Asked a Question	18%	29%	20%	841	2	265	2.09	0.13
Fraction of Teachers Who Used Local Information while Teaching	7%	13%	10%	841	2	265	1.43	0.24

Fraction of Teachers Who Used a Learning Aid**	32%	45%	33%	841	2	265	3.76	0.02
Fraction of Teachers Who Asked Students to Work in Small Groups	6%	6%	5%	841	2	265	0.04	0.96
Student's Math ASER Level[2]	2.3	2.0	2.0	7376	2		2.73	0.25
Student's Hindi ASER Level[2]	2.5	2.3	2.2	7376	2		2.46	0.29

Notes: Unless otherwise noted, F-statistics reflect the model specification statistic for a linear regression model with the outcome variable listed on the leftmost column and the explanatory variables as two binary variables indicating either intrinsic or extrinsic motivation treatment status. Unless otherwise noted, all standard errors in this table are clustered at the school level.
* Indicates significance at the 10% level.
** Indicates significance at the 5% level.
*** Indicates significance at the 1% level.
[1] Highest education level is an ordinal variable, so an ordered logit model was used. Because this model uses maximum likelihood, the model test is a chi-squared test, rather than an F-test. Standard errors were clustered at the school level. Group means have not been listed for this model.
[2] Student Math and Hindi scores are from the ASER test battery, which gives an ordinal value. Because this model uses maximum likelihood, the model test is a chi-squared test, rather than an F-test. Standard errors were clustered at the school level. Readers should use some caution when interpreting group means, as the numerical values from one category to the next are arbitrary.