Evaluating the Design and Impact of School Sector Development Program (SSDP) Training for 9th and 10th Grade Math and Science Teachers

Baseline Study Methodology and Findings[1]

September 2018

Julie Schaffner (Tufts University)

Uttam Sharma (Independent Consultant)

Paul Glewwe (University of Minnesota)

[1] We thank the International Initiative for Impact Evaluation (3ie) for funding (under Grant PW3.10.NP.IE), the Government of Nepal's National Planning Commission (NPC) for leading this research, and the Government of Nepal's Ministry of Education (MOE), Department of Education (DOE), and National Center for Educational Development (NCED) for their collaboration. We especially acknowledge the efforts of the evaluation's technical committee, which was led by NPC's then Joint Secretary, Dr. Teertha Dhakal, and includes representatives from the above mentioned agencies. We are grateful for research and logistical assistance provided by Deepika Shrestha, Sunil Poudel and Tri Bikram Pandey of the Center for Policy Research and Consultancy (CPReC); for research assistance from Tufts University students Claire McGillem, Mikayla Redding, Jessica Sashihara and Matthew Weinmann, and from University of Minnesota student Matthew Bombyk; for advice on questionnaire design from Dr. Sushan Acharya of Tribhuvan University; and for advice and assistance with implementing Stallings classroom observations from Rashmi Menon and Anuja Venkatachalam of J-PAL South Asia.

I. Introduction

This report describes the methodology for the baseline study associated with the on-going project "Evaluating the Design and Impact of the Secondary School Teacher Training Initiative under the Government of Nepal's School Sector Development Program" (funded by 3ie, the International Initiative for Impact Evaluation), and presents the main empirical findings from the baseline study. Section II offers a brief overview of the larger research project. Section III reviews the objectives of the baseline quantitative study. Section IV describes the baseline study design and reports on the data collection process. Section V offers an empirical description of Nepal's 9th and 10th grade education landscape based on findings from the baseline study. Section VI reviews the theories of change for the teacher training and video assignment interventions to be evaluated in the larger study, and comments more specifically on what the baseline data have to say regarding the assumptions underlying the theories of change. Section VII conducts balance tests, reviews power calculations, and discusses the likely implications for endline impact estimation. Finally, Section VIII discusses the research project's next steps.

II. Overview of the Larger Research Project

Nepal has made great strides in raising school enrolment in recent years, but average student learning in Nepal's public schools remains low. Recognizing that development success requires the country's children and youth to acquire valuable math, science and language skills, the Government of Nepal has prioritized efforts to improve school quality over the seven years (2016 - 2023) covered by the School Sector Development Program (SSDP).

One of the policy initiatives identified by policymakers as having the potential to improve learning is a new wave of teacher training programs. Observing that many teachers, especially at the secondary level, remain weak in subject content and pedagogy, policymakers hope to improve teacher effectiveness by rolling out new teacher training programs and strengthening teachers' accountability for applying new methods in their classrooms.

Focus and Objectives. This study is designed to evaluate the impacts on teacher knowledge, attendance, and pedagogical practices, and on student learning, of the new teacher training (TT) initiative, and to examine its theory of change, with the aim of providing guidance for future policy decisions regarding the scaling up or re-design of these policies.

The study focuses on math and science education in grades 9 and 10 in schools that include at least grades 1 through 10, which are considered models for what most schools will soon be in Nepal.[2] It focuses on one main treatment and one supplementary treatment. The main TT treatment requires all 9th and 10th grade teachers to attend in-service TT modules that are intended to raise their subject knowledge and to motivate and equip them to use methods that develop students' analytical and creative thinking skills, rather than methods focused on memorization. The

[2] Schools that include at least grades 1 through 10 constitute approximately 20% of all schools in Nepal (Government of Nepal, 2016). The fraction of 9th and 10th grade students who are educated in such schools is significantly higher (though the exact fraction is not available in published data).

supplementary Video Assignment (VA) treatment assigns trained teachers to submit videos of themselves competently employing new teaching techniques in their classrooms. The main evaluation questions are:

- What are the impacts on teacher knowledge, teacher attendance, teaching practices, and student learning outcomes of the TT treatment?
- To what extent do TT treatment impacts increase or decrease when the VA treatment is added?
- How do TT and VA treatment impacts differ across schools with stronger and weaker initial school management?
- What are the strengths and weaknesses of the TT and VA treatment implementation processes?

Mixed Methods Approach. By combining qualitative and quantitative research in a "mixed methods approach," the research team aims not only to estimate impacts but also to test theories of change and illuminate governance challenges. The approach mixes qualitative and quantitative methods in two senses. First, it interweaves waves of data collection that traditionally fall under the qualitative label (e.g. small-N studies with open-ended interviews and focus group discussions) with methods that traditionally fall under the quantitative label (e.g. large-N studies involving questionnaires and assessments), in ways that capitalize on the strengths of each and allow each to inform the other. Second, even the "quantitative" data collection waves are designed to incorporate broad ranges of topics and measurement approaches, many of which (e.g. opinion questions and classroom observation) are traditionally more associated with qualitative research.

In broad brush, the research team chose a randomized control trial (RCT) study design for the quantitative research to ensure rigorous estimation of the interventions' causal impacts. The project began with a preliminary qualitative study, which informed the team's choice of evaluation of questions, its understanding of the interventions' theories of change, and the team's development of instruments to be used during the quantitative studies. This was followed by the baseline data collection from the RCT sample, the methodology and findings of which are reported below. The next step will be a wave of qualitative research, which will allow the team to probe more deeply into baseline conditions and theories of change, identify points of strength and weakness in the implementation of the interventions, search for insights about how to improve training and teacher professional development initiatives in the future, and perhaps pursue answers to questions raised by baseline findings. By examining impacts on teacher knowledge, teacher attendance and teaching practices, as well as on learning outcomes, and by comparing impacts for different teacher contract types and different initial levels of school management quality, the team will use endline data to enhance learning about the theory of change, as well as to obtain quantitative estimates of impact.

For the RCT, the team selected 12 schools in each of 15 of Nepal's 75 districts, and 24 schools in a 16th district. Within each district, schools were randomly allocated into two main study arms of equal size: "Phase 1" schools, which are receiving the TT treatment in early to mid- 2018, and "Phase 2" schools, which will receive no treatment at the secondary level until at least June of 2019 and will serve as the controls. The team divided the main TT (or Phase 1) treatment arm into two sub-arms, with one receiving only the TT treatment and the other receiving both the TT treatment and the supplementary VA treatment.

Because there is significant overlap between the content of SSDP training (i.e. the TT intervention) and that of the trainings provided under the School Sector Reform Program (SSRP, the previous wave of government education initiatives), the research collaboration chose to emphasize estimation of SSDP training impacts *for teachers who had not received SSRP training in the past.* This was achieved by employing a stratified random sample design that over-sampled schools in which there was no evidence (in NCED records) of permanent teachers having completed training under the SSRP. As described in more detail below, the study draws two-thirds of the schools from a "priority" stratum in which (according to NCED hard copy records) no permanent teachers (or teachers of unknown contract type) who teach math or science in grades 9 or 10 have completed SSRP training. The aim of this strategy was to (a) build a sample in which there is a large representation of 9th and 10th grade math and science teachers who are untrained at baseline; while still (b) allowing us to estimate characteristics and impacts for the broader population of schools in Nepal (which include both priority and non-priority schools).

The desire to emphasize the impact of SSDP training on teachers without SSRP training also led the team to introduce a departure from standard practice in SSDP training roll-out. Although early SSDP documents suggested that teachers of all contract types would be eligible for SSDP training, in practice budgets are tight and priority tends to be given to inviting for training only teachers under permanent contracts who have not completed SSRP training. In the study districts, the Education Training Centers (which provide SSDP training) have been instructed to invite *all* teachers of 9th and 10th grade math or science, regardless of their contract type. While this means the study intervention is somewhat different from the SSDP intervention in non-study districts at present, it by no means renders the study irrelevant to training policy in Nepal. Historically, there is precedent for the provision of government training to non-permanent teachers; even though SSRP rules also called for training to be provided only to permanent teachers, NCED records indicate that some non-permanent teachers nonetheless received SSRP training. Furthermore, baseline data reveal that approximately 73% of 9th and 10th grade math and science teachers are non-permanent, suggesting that policymakers will face important choices about whether and how much to invest in training non-permanent teachers in the future.

III. Objectives of the Baseline Study

This section describes the objectives that shaped the design of the quantitative baseline study.

Objective 1. To Facilitate Production and Confirmation of High-Quality Impact Estimates at Endline. Impact estimates are "high quality" when researchers have strong reason to believe that they are unbiased as well as statistically precise. In principle, when using randomized control trials (RCTs) it is possible to obtain unbiased and precise estimates using only endline data, as long as sample sizes are large enough. Random allocation of schools to different study arms should (with high probability) lead average conditions to be virtually identical across study arms at baseline. When that is the case, average differences in outcomes across study arms at endline should be unbiased estimates of intervention impacts, because there should be no reason for average outcomes to differ across study arms other than the differences in treatment by the interventions. So, in principle, it is possible to conduct successful RCTs without gathering baseline data. In practice, however, it is highly valuable to gather baseline data on school, teacher and student characteristics and outcomes, because they allow researchers to: (a) confirm that random allocation has indeed produced study arm groups that look virtually identical (using balance tests, as demonstrated below); and (b) increase the precision of impact estimation (i.e. reduce the standard errors of impact estimates) by allowing the estimation of endline regressions that include students' baseline test scores as controls.

Objective 2: To Make Possible Estimation of Heterogeneous Impacts at Endline. Often evaluation studies yield more useful insights when they examine not only the overall average impact of an intervention (e.g. the average impact of SSDP training across all types of schools, teachers and students), but also differences in impacts across groups. For example, as will be argued below, it seems likely that participation in training will have a greater impact on teaching for teachers who work in schools where the Head Teacher or members of the School Management Committee do a good job of holding the teachers accountable for putting what was learned at the training into practice. Training might also have differential impacts on teachers with different contract types or training histories, or across students at higher or lower initial levels of academic performance. For econometric reasons (to be discussed in later documents), it is important to measure the variables potentially associated with this impact heterogeneity at baseline (rather than at endline).

Objective 3: To Gain Insight into the Interventions' Theories of Change. Econometric impact estimates are convincing and enlightening to policymakers only when they are embedded within larger studies that help to establish their plausibility, help to explain why they are large or small, and help to explain why they might differ across groups. It is, therefore, useful to articulate the theories of change underlying the interventions and to gather baseline data that might help to confirm or raise questions about the theories' assumptions (as is done below). More generally, it is useful to gather data that provide researchers with a richer understanding of context, especially in contexts where access to high quality administrative data or data from research projects on related topics is lacking or difficult to access.

Objective 4: To Inform Broader Policy Discussions. Fielding a survey is a time consuming and costly undertaking, but the additional cost of adding questions to a survey that is already heading to the field can be relatively low. It seems prudent, therefore, to take advantage of the opportunity afforded by baseline data collection for the SSDP training impact evaluation to gather additional information that may be useful in broader education policy discussions. With this aim, the research team designed the baseline quantitative survey to provide policymakers with a rich description of 9th and 10th grade math and science education in Nepal, and of the challenges faced by schools and teachers at this level. The team especially hopes that the baseline description of the nature and quality of school management will inform discussion of proposals for Head Teacher Development (another SSDP initiative).

IV. Baseline Study Design and Data Collection

This section describes the sample design and data collection instruments employed at baseline, and reports briefly on the implementation of the field work.

IV.A. Sample Design

Overview. Baseline data were collected from the schools included in the randomized control trial (RCT) of the study interventions. The RCT sample was designed to:

- be large enough (according to power calculations) to yield sufficiently precise impact estimates
- be approximately representative (at least when using sampling weights) of all schools in Nepal that include at least grades 1 through 10
- ensure that most of the sampled schools have teachers who had not completed SSRP training
- be easily divisible into halves, thirds, and quarters within districts, so that (a) one half of the schools in each included district could be allocated to receive the TT treatment during the study period while the other half would receive treatment only after the study period;(b) two thirds of the sample could be drawn from a priority stratum and one-third from a non-priority stratum; and (c) one half of the schools receiving SSDP training (i.e. one quarter of the entire sample within a district) could be assigned to also receive the VA treatment.

The following paragraphs define the sample selection process in more detail.

Power Calculations and Sample Size. Standard formulas allowed the research team to estimate the number of schools required to have at least an 80% chance of detecting (at the 95% significance level using a two-tailed test) an impact of the TT intervention on average student test scores of at least 7 percentage points (judged to be a modest effect size). The calculations depend on two key parameters: the population standard deviation of test scores (sigma) and the test score intra-cluster correlation coefficient (ICC). Estimates of these parameters using pre-existing databases of Nepalese student standardized test scores suggested that a sample of at least 200 schools would be required to achieve this power. (The estimate of sigma was approximately 20 and the estimate of the ICC was approximately 0.65, which is quite high and implies the need for a relatively large sample of schools.) The team anticipates being able to increase precision beyond what is suggested by these calculations by gathering baseline test score data, which can be used as controls in regressions used for impact estimation at endline.

Selection of Districts from which to Sample Schools. The simplest approach to obtaining a sample of schools that is representative of all schools in Nepal would be to randomly sample 200 schools from a list of all schools in Nepal. This would, however, lead to a sample of schools that would be spread across most, if not all, of Nepal's 75 districts. The logistical challenges and cost of collecting data from such a sample would be very high. To reduce data collection costs, the team first randomly selected a representative set of districts, and then sampled schools only within those districts. To reduce costs further, the team also chose to eliminate from consideration some of the most remote districts. The following paragraphs describe this in more detail.

Taking into consideration both statistical and logistical requirements, researchers and policymakers together chose to aim for a sample of approximately 16 districts. Recognizing the potential for administrative problems or lack of necessary administrative data to render work in some districts infeasible, the team decided to begin with a sample of 20 districts, and to reduce this to 16 after attempting to obtain sample frame information for each of these districts.

As a first step toward selecting 20 districts, 10 districts were dropped because of either extreme remoteness or other problems.[3] The districts, and the reasons for their exclusion, are shown in Table 1.

Table 1: Ten Districts Excluded from the Sample Frame, and Reason for Exclusion

District	***Reason for dropping***
Bajura	Remote
Dolpa	Remote
Humla	Remote
Kalikot	Remote
Manang	Remote
Mugu	Remote
Mustang	Remote
Nawalparasi	District straddles two provinces; incomplete records linking schools to districts and provinces
Rukum	District straddles two provinces; incomplete records linking schools to districts and provinces
Taplejung	Area judged too difficult for field work

According to a spreadsheet provided by the Ministry of Education, there are 8,681 community (i.e. government) schools in all of Nepal. The above excluded districts have 497 (or 5.7%) of these schools. Thus the 65 remaining districts account for 94.3% of Nepal's schools.

The second step in selecting districts was to sort the remaining 65 districts by "terrain" (mountains = 1, hills and plains = 2), and then – within each of the two terrain types – by province (of which there are seven in Nepal). Within each of these terrain-province groups, the districts were listed in "serpentine" order (each district on the list having a common border with the next district on the list). These lists of districts for each terrain-province group were then "stacked" into one long list for all 65 districts.

From this list of 65 districts, a random sample of 20 was chosen by first choosing a random "start point" that selects the first district near the top of the list and then using a set "jump" length to identify 19 more districts by moving down the list by a (weighted) count of the districts equal to

[3] At the time of drawing the sample of districts (January of 2017), these districts were 10 districts. More recently, both Nawalparasi and Rukum were divided into two districts, so together these now constitute 12 districts.

the jump length, where the weight was given by the number of eligible schools in the district. The probability of any given district being selected into the sample was, therefore, proportional to the number of eligible schools it contained, where "eligible" schools were government schools that included at least grades 1 through 10. Districts in Province 6 were given a double probability of being selected, as requested by Nepalese policymakers.

The 20 districts that were selected are shown in Table 2.

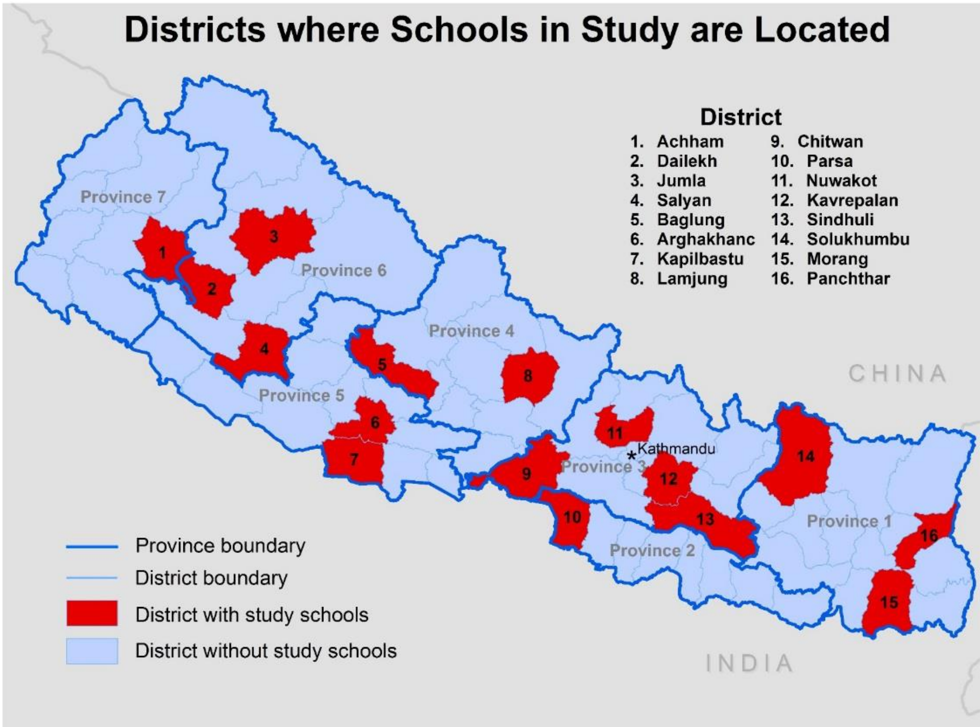
Table 2: Twenty Selected Districts (from 65 Districts in Nepal)

Achham	Lamjung
Arghakhanchi	Mahottari*
Baglung*	Morang
Baitadi	Nuwakot
Chitwan	Panchthar
Dailekh	Parsa
Jumla	Salyan
Kapilvastu (Kapilbastu)	Sindhuli
Kavrepalanchok (Kavre)	Solukhumbu
Khotang*	Tanahun*

* District that is later excluded, as explained in the text.

Finally, the list was reduced to 16 through the exclusion of four districts: Khotang, Mahottari, Tanahun and Baitadi. The choice of these four districts was determined in part by difficulties with obtaining sample frame information and in part by the desire to retain broad geographic representation in the sample. In a meeting held on July 27, 2017 with the National Planning Commission (NPC), the National Center for Educational Development (NCED) and other government officials, it was determined that sample frame data was at that point unavailable for Achham, Baitadi, Khotang and Mahottari. However, Nepalese officials wanted to retain either Achham or Baitadi to represent the far western region. Meeting participants suggested that instead of dropping both Accham and Baitadi, one would be retained and instead one of the following districts could be dropped: Tanahun, Salyan or Sindhuli. (These are in different provinces but in provinces that already had two or three other districts in the sample: Tanahun is in Province 4 and Lamjung and Baglung were already from Province 4; Salyan is in Province 6 and Dailekh and Jumla were already from Province 6; and Sindhuli is in Province 3 and Nuwakot, Kavre and Chitwan were already from Province 3). After further discussions, Tanahun was dropped. The last step was to choose between Accham and Baitadi. This was decided at a meeting with the NCED; as the data from Baitadi indicated that there was very little training taking place in that district, it was decided to keep Accham and drop Baitadi. After further effort, sample frame data were obtained for Accham, making its inclusion feasible. Figure 1 illustrates the geographic spread of the 16 sample districts.

Figure 1: Sixteen Selected Districts



Selection of Schools within Districts. The overall goal in selecting schools (as suggested by power calculations) was to select a random sample of approximately 200 schools, of which half would be randomly assigned to treatment during the study period (Phase I schools) while the other half would be assigned to treatment only after the study period (Phase II schools, serving as the "control group"). The team decided that the sample schools would be randomly allocated to treatment arms within each district, with one half of the schools randomly assigned to the control group, one fourth randomly assigned to the "TT only" treatment group, and one fourth randomly assigned to the "TT plus VA" treatment group. This implied that the number of schools per district had to be divisible by four. For reasons discussed in Section II above, the research team also decided that, within each district, two-thirds of the sample schools would be drawn from the group of "priority" schools, while the other one-third would be drawn from "non-priority" schools. Adhering to this rule implies that the number of schools chosen per district had to be divisible by three. For the number of schools to be divisible by three and by four, the smallest number that would work would be 12.

The above considerations suggested that 12 schools be sampled from each district, which implies a total sample of 192 schools. However, the technical committee which includes NPC and NCED, among others, wanted one or more of the largest districts to have a larger number of schools in the sample, so the team proposed that the largest district have a "double" sample, which is a sample of 24 schools. This would increase the total sample by 12, to 204 schools. Morang district was selected to have the "double" sample because it is the largest district in the sample; in the administrative data used to select the schools, Morang has 154 of the eligible 1,334 schools in the 16 selected districts, which is more than any other district (the next largest is Kavrepalanchok, which has 141).

Although, in principle, the sample should have consisted of 204 schools, in the end it included only 203. One district, Solukhumbu, had only 3 non-priority schools, rendering it impossible to choose 4 non-priority schools there. Thus the final sample of 203 schools consists of 136 priority schools (16×8 + 8 more for Morang) and 67 non-priority schools (16×4 + 4 more for Morang, minus 1 missing from Solukhumbu).

The first step in selecting the sample of schools within any district was to eliminate any schools for which there was evidence that any teachers had already received SSDP training (i.e. the TT treatment to be evaluated in the larger study). This was necessary because, despite initial plans of the research collaboration to prevent roll-out of SSDP training in any study district before the study, these rules were implemented imperfectly. Fortunately, according to these records at least one teacher had received SSDP training in only 83, or 6.2%, of all 1,334 schools with at least grades 1 through 10 in the 16 study districts. The information on SSDP training available at the time of school selection appears to have been imperfect, however; despite eliminating schools with known SSDP training from the sample frame, baseline data reveal that 6.6 % of sampled teachers report having received SSDP training.

The second step when selecting schools within districts was to categorize schools into "priority" and "non-priority" strata. Schools were identified as priority schools if there was no evidence (in hard copy records made available by the NCED) of any permanent teacher, or any teacher for which contract type was unknown, having completed all three modules of the SSRP training (the training under a previous wave of education interventions). The specifics of this rule were dictated largely by the idiosyncrasies of the only existing records identifying which schools and teachers had received SSRP training. Policy implementers would have preferred to prioritize schools in which most or all teachers of 9th or 10th grade math and science both: (a) had permanent contracts; and (b) had not received any SSRP training, because these are the sorts of teachers they would typically prioritize for training on limited budgets. The team initially hoped to combine records from one source indicating which teachers had received SSRP training with records from another source indicating which teachers had permanent contracts. This ultimately proved impossible, and the team ended up relying solely on NCED hard copy records that (in principle) listed all the teachers who had completed all three modules of the SSRP training. The list indicated the teachers' subjects (allowing the team to identify math and science teachers) and levels (allowing identification of "secondary" teachers who taught 9th and/or 10th grade). In some but not all districts the lists indicated whether the teachers were permanent or non-permanent. The list also indicated the teachers' schools at the time of their entry into the list, allowing the team to make tallies of the numbers of math and science teachers from a given school who had completed training, separately for "permanent," "temporary," and "unknown" contract types. Through conversations with policymakers, the team ultimately chose to designate as priority schools those for which the tallies of permanent and unknown contract type teachers who had received SSRP training were zero.

Concerned about potential spillovers of impact from Phase I (treated) schools to untreated schools near to them geographically, the research team chose to select priority and non-priority schools within districts in a way that would reduce the probability of any two sample schools being near to each other geographically. Rather than take simple random samples from among lists of priority and non-priority schools in a given district, therefore, the team first grouped schools by the Village Development Committee (VDC) territories to which they belonged. VDCs are administrative sub-units within districts. In the 16 districts from which the schools were drawn, there were 1,251 "eligible" schools (i.e. schools that had at least grades 1 through 10 and had not apparently received SSDP training) spread over 751 VDCs, so the average VDC had 1.67 eligible schools.[4] The general process to reduce the probability that two sampled schools would be very close to each other was to draw a sample of VDCs, and then within each VDC randomly draw only one school.

The goal for any district was to draw a sample of six non-priority schools (four plus two "spares" in case one or two of the four schools had to be excluded for some reason at a later date). Six VDCs were randomly selected from all the VDCs in the district that contained non-priority schools (except that for Morang district 12 VDCs with non-priority schools were selected), with the probability of selecting a VDC being proportional to the number of non-priority schools in the VDC. (Note that this procedure could sometimes choose the same VDC twice – in this case another VDC with non-priority schools was randomly chosen to get a total of six different VDCs; this is, in effect, sampling without replacement.) In any VDC that had two or more non-priority schools, one of those schools would be randomly selected. This procedure randomly selected six non-priority schools from six different VDCs in a given district, and each non-priority school in a given district had an equal probability of being selected (since VDCs were chosen with probability proportional to the number of non-priority schools they contained, but then only one school was chosen at random per VDC).

The following two-step procedure was used to select 12 priority schools from each district (eight sample schools, plus four "spares" in case one or more of the first eight schools had to be excluded at a later date). First, VDCs that had already been selected in the process of choosing non-priority schools were disqualified. Twelve VDCs with priority schools were randomly selected from the remaining group of VDCs (except that for Morang 24 priority VDCs were selected). Each VDC in this group was given a base weight equal to the number of priority schools in the VDC divided by the total number of priority schools in that district (excluding any priority schools in the six VDCs that were selected as non-priority VDCs). The disqualification of VDCs already selected for non-priority schools implies that VDCs with priority schools that *also* have non-priority schools were less likely to be selected than VDCs with priority schools but no non-priority schools (because the former had to "survive" the possibility of being selected as a non-priority VDC). To ensure that each priority school had a similar probability of being selected, VDCs with both priority schools and non-priority schools were given a greater probability of being selected relative to VDCs with priority schools but no non-priority schools.[5] More specifically, for each VDC that

[4] Of the 751 VDCs, 438 had only one eligible school, 223 had two eligible schools, and 90 had three or more eligible schools.

[5] Of the 564 VDCs that had one or more priority schools, 412 had no non-priority schools, 124 had one non-priority school, 17 had two non-priority schools and 11 had three or more non-priority schools.

had both priority and non-priority schools, the base weights were multiplied by an additional weighting factor; this weighting factor was calculated as the total number of VDCs in the district that had both types of schools divided by the total number of VDCs in the district that had both types of schools *and* were not selected as one of the six non-priority school VDCs. Finally, in each VDC selected to be a priority school that had more than one priority school, one of those priority schools was randomly selected (without any weighting).

For 15 of the 16 districts, we had 6 non-priority schools (of which the last two were "spares") and 12 priority schools (of which the last four were "spares"). Among the 6 non-priority schools, the first four were kept unless there was some reason to exclude any of them (e.g. because it was discovered that some teachers in the school had already received SSDP training). If one school was excluded then the next one on the list was used to replace the excluded school. It was never the case that more than two non-priority schools were excluded. This resulted in a sample of 4 non-priority schools for each of the 15 districts. In Morang an analogous procedure was used to reduce the sample to 8 non-priority schools. Analogous procedures were also used in the 15 districts with 12 selected priority schools to reduce the sample to 8 priority schools, and for Morang to reduce the number of priority schools to 16. As indicated above, the only exception was all three non-priority schools were selected into the sample in Solukhumbu district, yielding a total sample size of 203.

Assignment of Schools to Study Arms. To prepare for the RCT roll-out of interventions, the research team randomly allocated the sample of 203 schools to three study arms: Phase I TT treatment (SSDP training) without Video Assignment (VA); Phase I TT treatment with VA; and Phase II treatment (only after the study period, so this arm functioned as the control group). Within each district-stratum subset of schools, one quarter each were allocated to the Phase I TT without VA and Phase I TT with VA study arms, while one half were allocated to the Phase II study arm. Thus, of the total sample of 203 schools in 16 districts, 51 were randomly assigned to receive the TT treatment without Video Assignment, 51 were randomly assigned to receive the TT treatment with Video Assignment, and 101 were randomly assigned to Phase II.[6]

During baseline data collection, however, neither school personnel nor the survey firm that collected the data knew the study arm to which a school had been assigned. This rule was followed to prevent any differences in thinking among school respondents or in treatment by the survey firm from introducing any difference at baseline in measured characteristics or outcomes across study arms.

Sampling Weights. Two features of the above sampling plan suggest that, in principle, sampling weights should be employed when using the sample data to estimate means, variances or other quantities for the entire population of schools (of eligible type) in the 16 study districts. First, the districts vary in size (i.e. in numbers of schools and students) while the sample includes only the same number of schools (12) in most districts. Second, the shares of schools that met the "priority" requirements differed across districts, but the same shares of priority and non-priority schools were

[6] In Solukhumbu district there were only 3 non-priority schools in the entire district, of these one was randomly assigned to teacher training without video assessment, one was randomly assigned to teacher training with video assessment, and the third was a control school.

selected into the sample in all districts. The research team used Monte Carlo methods to calculate the appropriate weights, but discovered that in most cases in the analysis of baseline statistics, the weighted and unweighted statistics were very similar. For the sake of clarity and simplicity, therefore, this report presents unweighted statistics in what follows.

IV.B Data Collection Instruments

To achieve all the baseline quantitative study objectives described above, we used the following instruments and tools.

- A head teacher questionnaire, administered by an enumerator using a tablet. The questionnaire includes questions about the school, the head teacher, the school management committee and school management practices, as well as questions about teachers and teaching practices in 9^{th} and 10^{th} grade math and science.
- A questionnaire for teachers of 9^{th} and 10^{th} grade math and science, administered by an enumerator using a tablet. The questionnaire includes questions on teacher characteristics and school management practices.
- A student questionnaire, which students in grades 8 and 9 were asked to fill out on their own on paper copies. (Students in grades 8 and 9 at baseline will be finishing grades 9 and 10 at endline.) The questionnaire includes questions on student socioeconomic characteristics and on teaching practices employed in their classrooms.
- Student assessments in 8^{th} and 9^{th} grade math and science. Each student took two one-hour assessments, one in math and one in science. The assessments are standard math and science assessments that were developed for this study by Nepalese academics in the relevant education fields.
- Measures of classroom teaching practices and student engagement derived from use of the Stallings method of classroom observation (*World Bank, 2015*).

The measures of classroom teaching practices using the Stallings method require a trained observer to sit in the back of the classroom for an entire class period. He or she makes 10 observations during a full class period, recording the results of the observations into a tablet (using software provided by the World Bank). The first observation is made 3 minutes into the class period. If M is the total number of minutes in the period (usually 45 in Nepal's secondary schools), then the remaining 9 observations are scheduled every M/10 minutes for the rest of the class period. At each scheduled observation time, the observer scans the classroom, starting from wherever the teacher is, for 15 seconds, and then records information regarding: (a) what type of activity the teacher is engaged in; (b) whether the teacher is interacting with all the students, a small group of students, someone else, or no one; (c) how many of the students are engaged with the whole group or small group activities (versus being unengaged in learning activities); and (d) the teacher's use of various sorts of materials or methods of engagement with students. These observations were carried out during "unannounced" visits. This means that while the schools were informed ahead of time (during the announced visits) that researchers would visit their school to observe classrooms at some time in the next several weeks, the schools did not know the specific day of the visit. The purpose of leaving the date of the visit uncertain is to prevent the teachers, to the extent possible, from planning highly unusual activities for the day of the visit, in order to impress observers. Naturally, teachers will still be on their best behavior with observers in the classroom, but this allowed the team to measure what teaching practices the teachers are capable of, which is an important measurement for understanding students' educational outcomes.

IV.C Field Work

New Era Pvt. Ltd. was hired to conduct the baseline study. The field work was done in two phases, with a gap in between to accommodate the holidays that most schools took for at least three weeks between September and October. The first phase of the fieldwork, with 54 enumerators and supervisors, started on August 17, 2017, after the conclusion of a 10-day enumerator training. The second phase of fieldwork in the remaining four districts -- Baglung, Morang, Nuwakot and Parsa -- started on October 24, 2017, and was concluded on Nov. 27, 2017. The Stallings classroom observations were conducted from August 30, 2017, to January 18, 2018.

Enumerator trainings were conducted by the principal investigators and New Era staff. Rashmi Menon and Anuja Venkatachalam, Senior Research Associates at J-PAL South Asia, led a three-day training course for 12 enumerators selected for Stallings classroom observations.

In total, 203 Head Teacher questionnaires and 393 Teacher questionnaires were administered. New Era also administered questionnaires and math and science assessments to 7,649 (8,782) students in grade 8 (grade 9). The unannounced Stallings classroom observations were performed in 668 classes for 373 teachers interviewed during baseline in 200 schools. Two separate classroom teaching sessions were observed for most teachers. [7] The enumerators used Survey CTO software to collect the data using their android phones.

The baseline data collection protocol was reviewed and approved by Institutional Review Boards of Tufts University and the University of Minnesota.

V. Baseline Findings on Nepal's 9th and 10th Grade Education Landscape

Conceptual Framework. With the aim of facilitating more informed discussions of education policy in Nepal, this section employs the baseline data to build a description of Nepal's 9th and 10th grade math and science landscape. The description is shaped by a conceptual framework in which the student learning outcomes that a school achieves are shaped not only by (a) the resources available to the school, and (b) the challenges and opportunities the school faces in its student and parent populations and socioeconomic context, but also (c) the quality of "management" that is brought to bear on running the school. Here "management" is a broad reference to any efforts that

[7] Though classroom observation sessions were to be conducted for 393 teachers, only 373 of them were in school during the unannounced visits. Though 29 classes of 19 teachers who were not interviewed during baseline survey were observed, their classroom observations were dropped from the study. Three schools did not have any of the teachers interviewed during baseline present during the unannounced school visit.

people (such as head teachers or School Management Committee members) exert to: (a) assess school needs and challenges; (b) brainstorm possible ways forward; (c) coordinate activities; (d) encourage cooperation among teachers; and (e) motivate teachers to teach well and to apply what they learn during trainings in their classrooms. After offering a basic description of school types, this section first discusses the schools' contexts, student and parent populations, challenges, and resources. It then takes a closer look at various actors who might be involved in managing the schools, seeking to identify who indeed is playing this role for Nepal's schools and what capacity they have to manage well. This is followed by an attempt to describe the quality of management that they achieve and how it varies across schools, albeit using indicators that are sometimes indirect. The section then looks at what schools achieve in the areas of: teacher characteristics and training; teacher attendance and teaching practices; and student attendance and learning outcomes.

Note on Statistical Interpretation. The research team believes that the descriptive statistics below offer a reasonably good description of the education landscape not only in the 203 sample schools, but also in the larger population of relevant schools (i.e. government schools that include at least grades 1 through 10) in the 16 study districts, and in the larger population of relevant schools in most of Nepal (excluding only some of the most remote districts as discussed above). Even though the 203 schools in the sample constitute just 15.2% of all 1,334 government schools with at least grades 1 through 10 in the 16 study districts, and just 3.3% of all 6,163 schools of this type in all of Nepal, the use of random sampling should render it possible to use them for unbiased estimation of characteristics (such as means and proportions of variables) in the larger populations, at least when weighted to account for differences in population sizes and priority school shares across districts.[8] How accurate these estimates are likely to be can be assessed by calculating standard errors and confidence intervals.

Several caveats should be kept in mind when drawing conclusions from the baseline statistics. First, even for describing the 203 study schools, the statistics presented are accurate only if the respondents answered the questions posed to them accurately and candidly. While the research team made many efforts to pose questions in neutral fashion, it remains possible that respondents tended to tilt their answers in directions that might please enumerators or put themselves or their schools in better light. Where possible, we report the similarities and differences between reports from different types of respondents on similar issues, and in other ways attempt to assess how accurate and candid the answers are likely to be.

Second, strictly speaking, to use the sample data to describe the larger population of all relevant schools in the 16 study districts or all relevant schools in Nepal, it would be necessary to calculate statistics using sampling weights equal to the inverse probability of a school's inclusion in the sample. As mentioned above, however, calculations not reported here suggest that in most cases the use of such weights does not tend to alter the statistics very much, so unweighted statistics are reported below.

[8] As mentioned above, while in principle the sample design for the baseline study implies the need to use sampling weights to obtain unbiased estimates of population characteristics, in practice weighted and unweighted estimates very close. Only unweighted statistics are reported below.

Finally, as estimates of population characteristics, the estimates are subject to sampling error. Standard error calculations not reported here (in the interest of reducing clutter) suggest that each mean, percentage or other statistic should be thought of as the center of a confidence interval of modest width. For example, while the percentage of schools reporting having completed a social audit in the previous academic year is 84% when calculated without weights and 85% when using weights, this estimate has a standard error of approximately 2.7 percentage points, implying a 95% confidence interval of approximately [80%, 91%].

School Type and Competitive Environment. As indicated in Table 3 (based on head teacher reports), the vast majority of schools (91%) provide instruction only in Nepali, while 8% provide instruction in both English and Nepali and only 1% provide instruction only in English. Most schools are also relatively large, with 88% having 200 or more students, and 38% have 500 or more students. By construction, all schools in the sample included grades 1 through 10. Most schools (87%) also offer Early Childhood Development (ECD) classes, and 42% also offer grades 11 and 12.

While the typical school in Nepal is relatively remote (compared to schools in many countries), the degree of remoteness varies greatly. The distance to the nearest motorable road is 1 hour or less for 43% of schools, and 15 minutes or less for 22% of schools. On the other hand, for 25% of schools the nearest motorable road was 1-3 hours away, and for 32% it was more than 3 hours away (and more than 5 hours away for 17% of schools).

Despite the high typical degree of remoteness, a little more than half (57%) of the schools in the sample report that there is another government school within three kilometers and 28% report that there is a private school within three kilometers. Yet only 7% of the schools in the sample report that they compete for 9th and 10th grade students with at least one private school. This is especially interesting in light of conversations during preliminary qualitative research which suggested that a number of schools were feeling pressure to introduce English language medium or make other changes to prevent losing students to private schools.

Whether or not schools face competition from private schools, their leaders may feel pressured by parents to achieve high academic standards. A little less than one fourth (21%) of schools report that they face little or no such pressure, and 55% report “some” pressure from a few parents, while just one fourth (24%) report constant pressure from most parents. A closer look reveals that the fraction of head teachers reporting constant parental pressure for high academic standards is not highly correlated with proximity to the nearest all-weather road. It would be interesting to know what forces cause parents in some locations, including remote ones, to become advocates for high academic standards.

Table 3: Basic Characteristics and Resources of Government Schools with Grades 9 and 10 (Percent of schools)

Language of Instruction:	
Nepali only	91
English only	1
Nepali and English	8
Total Number of Students in School:	
Under 200	12
200-499	50
500 or more	38
Grades Offered beyond Grades 1 through 10	
ECD	87
Grades 11 and 12	42
Distance to Nearest All-Weather Motorable Road:	
15 minutes or less	22
15 minutes to 1 hour	21
1-3 hours	25
3-5 hours	15
More than 5 hours	17
Proximity of Other Schools	
One or More Private Schools within 3 Kilometers	28
One or More Government Schools within 3 Kilometers	57
One or more Private Schools that Compete for Students	7
Pressure from Parents to Achieve High Academic Standards:	
Minimal or No Pressure	21
Some Pressure from a Few Parents	55
Constant Pressure from Most Parents	24

School resources. One measure of teaching resources per student is class size, which can be measured by students per section. In Nepal's secondary schools, "sections" are groups of students (subsets of all students in a grade) who follow the same schedule, studying required subjects together with the same teachers at the same time. Most schools have only one section in grade 9 (88%) and one section in grade 10 (89%), while the largest schools have four sections per grade. As indicated in Table 4, the median classes in Nepal are quite large, at 47 for grade 9 and 42 for grade 10. There is a large variation in class size, however, with very large values of 107 (grade 9) and 93 (grade 10) at the 95th percentile.

Another way to gauge the level of resources per student available to a school is to observe its equipment and facilities. By such measures, most schools appear to work with at least somewhat limited physical resources, though the level of resource constraints appears to vary significantly across schools. As indicated in Table 4, most schools (93%) report having at least one computer, though some of these computers may not be very useful, given that only 78% of schools report having electricity and only 32% report having an internet connection. Only 42% report having a science lab and 18% report having computer labs. Calculations not shown indicate that, of the 40% of schools that report having 3 or fewer of the 9 assets listed in Table 4, most have computer, electricity and library, but lack the rest. Only 15% of schools report having 7 to 9 of the 9 assets.

The data in Table 4 also indicate that the vast majority of schools were reasonably effective at providing textbooks to their students in grades 9 and 10 in the previous academic year. Almost all (97%) provided textbooks to their students, and 90% provided a full set of textbooks. The delivery of textbooks was reasonably timely, although perhaps less than ideal for most schools, with over 14% providing them at the beginning of the school year, but 78% providing them only during the first quarter of the academic year.

Table 4: School Resource Information (Reported in Head Teacher Questionnaire)

	5th percentile	**Median**	**95th percentile**
Students per section			
Grade 9	17	47	107
Grade 10	15	42	93
Availability at school of:			**Percent**
Any type of computer			93
Electricity			78
Library			68
Science Lab			42
Projector			41
Internet connection			32
Large room			25
Computer lab			18
Telephone (landline)			18
Provides grade 9 and 10 textbooks to students			97
Provided grade 9 and 10 textbooks at beginning school year			14
Provided grade 9 and 10 textbooks during first quarter of school year			78
Provided **full set** of textbooks to grade 9 and 10 students			90

Student and Parent Populations. Student questionnaire data (Table 5) and head teacher questionnaire data (Table 6) both suggest that, while government schools are doing a good job of attracting girls as well as boys, and of attracting even many children from somewhat impoverished socioeconomic backgrounds, the schools face serious challenges in providing effective quality 9th and 10th grade math and science education because of the socioeconomic characteristics of their student and parent populations.

Table 5: Schools' Student and Parent Populations (from Student Questionnaire)

	Grade 8	Grade 9
Percent female	55	54
Father can read and write (percent)	78	79
Mother can read and write (percent)	51	54
Father's schooling (percent):		
None	10	11
Primary (grades 1-5)	32	32
Lower secondary (grades 6-8)	18	18
Secondary (grades 9 and 10)	18	20
Higher secondary (grades 11-12)	6	6
More than secondary	3	3
Unknown	12	10
Mother attended school (percent):		
None	33	33
Primary (grades 1-5)	27	29
Lower secondary (grades 6-8)	11	11
Secondary (grades 9 and 10)	8	8
Higher secondary (grades 11-12)	2	2
More than secondary	1	1
Unknown	18	16
Father's employment (percent)		
Agriculture	57	57
Operate a small business	14	9
Work for the government	8	20
Work for the private sector	8	8
Work abroad	21	14
Nepalese is the main language spoken at home (percent)	70	73
Family assets (percent whose families own)		
Mobile or landline phone	95	95
Television	54	54
Bicycle	43	40
Motorcycle or scooter	15	15
Refrigerator	12	13
Computer	8	10
Breakfast eaten before school? (percent)		
Breakfast or meal	95	95
Tea only	4	4
Nothing	1	1
Time spent in daily activities (medians, minutes per day)		
Travel time to school (1 way)	20	20
Studying	90	90
Working on family farm or to earn money	60	60
Housework	60	60

More than half of all students in grades 8 and 9 (54-55%) are girls. It is possible that this ratio is high in the sample's government schools because boys are more likely to be sent to private schools.

Several indicators suggest that students enter grades 9 and 10 significantly less well prepared than schools would like. When asked to indicate which items in a list of possible challenges to teaching and learning in 9th and 10th grade are major challenges in their schools, head teacher and teachers provided very similar responses. (Data on some of the Head Teacher responses are shown in Table 6.) Both teachers and head teachers might naturally be biased in the direction of reporting many problems as major challenges that hold their schools back (despite their personal contributions), so the overall frequency of problems reported may be high, but the similarity of patterns across respondents suggests that the data contain some real information about relative importance of problems and about differences across schools. Both groups are nearly unanimous in seeing students entering below grade level as a major challenge.

Low levels of parental education also mean that most secondary students can expect little assistance on their schoolwork at home, and that parents have low capacity for holding teachers accountable for effective teaching practices. Students report that about four fifths of their fathers have attended school (82-83%) and can read and write, but at most only about two thirds of their mothers have attended school, and only about half (51-54%) can read and write. Far fewer than half of fathers and mothers attained a secondary level of schooling (Grades 9 and 10); for fathers the range is from 27-28% (assuming that "don't know" implies less than secondary schooling) to 38-39% (assuming that "don't know" implies secondary or higher education), and for mothers the range is even lower, from 11% to 27-29%.

Head teacher and teacher responses regarding major challenges offer further suggestion that many students' find only weak support for their education at home, though the extent of this problem varies significantly across schools. Over 75% of both groups report poor student attendance, student work requirements, poor student motivation for homework, low parental education, and low parental motivation regarding their children's education as major challenges. Comparing "better off" schools, in which the number of major challenges reported by head teachers was at the 25th percentile of lower, and "worse off" schools, in which the number of major challenges reported by head teachers was at the 75th percentile or higher, suggests that differences across schools are especially great in the prevalence of poor student nutrition and hygiene, and in how much parents resist the use of new teaching practices.

Many students seem to juggle significant demands on their time related to income generating work and housework (Table 5). Most students (57%) report that their fathers work in agriculture, and presumably a majority of these are farmers. Thus it is not surprising that the typical student reports working about one hour per day on the farm. Another interesting aspect of fathers' employment is that fathers often work away from Nepal, in particular this is reported by 21% of grade 8 students and 14% of grade 9 students. The absence of fathers may contribute to the demands on students' time and suggests even less potential to help with schoolwork from fathers.

Asset data suggest that while the typical family of a secondary student in Nepal is not among the poorest households in Nepal, it is still poor by global standards, suggesting that direct and indirect costs may render it difficult for families to send children to school consistently and support their studies. Most families (95%) have mobile phones and 54% have televisions, while slightly over 40% have bicycles. On the other hand, only 15% have motorcycles or scooters, only 12-13% have refrigerators, while 8-10% report having computers.

The data in Table 5 present additional factors that could hinder schooling. First, while Nepali is the exclusive language of instructions in 91% of government schools, and is also used as a language of instruction in another 8% of schools, only 70-73% of students speak Nepali at home, so that 27-30% do not speak the most common language of instruction at home. Second, the typical grade 8 student spends about 2 hours per day working, including housework, and another 40 minutes per day traveling to school. Third, and on a more optimistic note, almost all children come to school well fed, with 95% having had a breakfast meal, and another 4% having had tea (which typically includes milk and sugar), and only 1% not have any breakfast at all.

Table 6: Students' Conditions in Their Homes (from Head Teacher Questionnaire)

Challenges Cited by Head Teachers (from head teacher questionnaire):	
Students spend long hours on housework, farm work or wage work	81%
Parents have too little education to help their children	79%
Parents are not motivated to encourage their children to study	82%
Students are below grade level when entering a new grade	93%
Student have poor attendance	88%
Students lack adequate nutrition, hygiene	71%
Parents resist new teaching practices	43%

Potential "School Managers." In principle, diverse actors might play roles in leading and managing school efforts to achieve the best teaching and learning outcomes possible with available resources, given the challenges faced. The next several sub-sections describe key actors who may play some role in leading or managing school activities: head teachers, School Management Committee members, parents and other actors (including international NGOs) that also provide resources to many schools at the local level in Nepal. The data suggest that: (a) head teachers are the actors most likely to play management and leadership roles, but that high teaching loads and other factors may prevent them from carrying out management activities; (b) parents are at present ill equipped to oversee efforts to improve teaching and learning at the secondary level; and (c) while SMCs are active in improving facilities and equipment, they do not take as active a role in overseeing what happens in classrooms. NGOs and other actors also provide assistance to many government schools, though the degree of their involvement in shaping school management decisions is unclear.

Head Teachers. Basic characteristics of head teachers are shown in Table 7. The overwhelming majority (97%) are men, and the median age is 47 (the overall range in age is from 27 to 60). Most are from the upper castes; 62% are either Brahmin or Chhetri. Almost all of them (93%) were teachers before becoming head teachers. Nearly two thirds (65%) lived in the community before becoming a teacher or head teacher there. Nearly half (47%) began serving as head teacher in their current schools relatively recently (i.e. in last three years). Although they have many years of teaching experience (median of 24 years) they have relatively few years of experience as head teachers (median of 6 years, including years as head teachers in other schools). Over three quarters (78%) are permanent teachers, but only 3% are Class 1 permanent teachers.

Table 7: Basic Characteristics of Head Teachers in 16 Districts

Male	97%
Median Age (in years, range is from 27 to 60)	47
Upper Caste (Brahmin or Chhetri)	62%
Median Years of Experience as Head Teacher in any school	6
Has Previous Experience as Teacher	93%
Median Years of Teaching Experience (range is from 2 to 35)	24
Has permanent contract	78%
Has “Class 1” permanent contract	3%
Lived in the School’s Community Before Becoming a Teacher or Head Teacher here	65%
Highest Degree:	
Intermediate/+2	17%
Bachelor	28%
Master	53%
Higher than Master	2%
*Whether has Degree in Given Subject:**	
School Management	48%
Business or Public Administration	9%
School Leadership	9%
Any of the above	56%
Whether Received Government Training for School Management	73%

* Head teacher asked to report all that apply.

Head teachers have high education levels overall, and many have degrees in fields related to school management. Among head teachers, 83% hold bachelor’s degrees or higher, and 55% hold master’s degrees or higher. Over half of head teachers (56%) hold degrees in school management, business or public administration, or school leadership. Nearly three-quarters (73%) report having had (non-degree) government training for school management and/or school leadership.

Head teachers have substantial teaching duties. As shown in Table 8, the median head teacher teaches 18 periods per week in courses for which he or she is the regular teacher. (One quarter of head teachers teach 24 or more periods per week, and 10% teach 29 or more periods per week.) As will be seen below, the typical math or science teacher for grades 9 and 10 teaches 29 period per week, so 18 periods per week is a teaching load that is almost two thirds (62%) of the teaching load of a typical teacher. In addition, in a typical week a head teacher also teaches three more periods for teachers who are absent, so the overall teaching load is almost three fourths (72%) of what a standard teacher teaches. Quite simply, head teachers spend most of their time teaching.

On a more positive note, most of them (72%) have staff to assist with administration and accounting, but even so this heavy teaching load leaves little time for leadership and management activities, and so it is not surprising (as will be seen below) that only about half of the teachers (49%) think that their head teachers are doing an excellent job of managing/leading the school.

Table 8: Head Teacher Activities and Working Conditions

Periods Taught (per week, as the regular teacher)	
25^{th} percentile	12
median	18
75^{th} percentile	24
90^{th} percentile	29
Median Periods Taught as Substitute for Missing Teacher (per week)	3
Staff to Assist with Administration and Accounting (percent who have)	77%

School Management Committees. Table 9 provides basic information on school management committees (SMCs). Most schools have active SMCs. Nearly all schools (95%) have SMCs, and of those SMCs almost two thirds (66%) meet at least once per month, and another 32% meet 3-6 times per year, leaving only 1% that meet only 1-2 times per year and another 1% that did not meet in the previous (academic) year.

While many SMC members visit homes annually and visit head teachers more frequently than that, few visit classrooms and almost none are a daily presence in schools, suggesting a limited role for overseeing teaching and learning. For example, head teachers in 60% of schools report that SMC members visited them at least once per month in the last year, 70% reported that SMC members visited homes of school-aged children to promote enrollment one or more times in the last year, and 80% report that SMC members visited the school to observe classes at least once per year, but only 3% reported SMC members visiting schools once or more per week. SMCs visit only a small fraction of classrooms, as seen from the finding that only 25% of teachers report that their classrooms were visited even once by an SMC member over the last academic year.

SMCs are more likely to be involved with physical improvements in buildings and equipment than in efforts to improve the quality of teaching and learning. They help monitor construction and repairs to the school building in 96% of schools, are involved in approving school budgets in 94%, help to recruit teachers in 84%, and help with raising cash or in-kind assistance from parents, community members or NGOs in 84%. Head teachers report, possibly generously (in light of statistics cited above), that SMCs are involved in efforts to improve attendance, teaching or learning in 73% of schools.

Head teachers' reports regarding their relationships with SMCs suggest that most SMCs leave day-to-day management of schools to head teachers. Most head teachers (95%) agree or strongly agree that SMCs give them adequate autonomy, while only 13% agree or strongly agree that SMCs try to directly control the decisions they need to make to run the school.

Table 9: Basic Information on School Management Committees (SMCs), as Reported by Head Teacher unless Indicated Otherwise

School has SMC	95%
Frequency of SMC meetings in last academic year:	
More than once per month	21%
Once per month	45%
3-6 times per year	32%
Only one or two times	1%
None	1%
Visits done by SMCs:	
Visited head teachers once or more than once per month	60%
Visited homes of school-age children a least once per year to promote enrollment	70%
Visited school to observe classes one or more times per year	80%
Visited school to observe classes once or more per week	3%
Percent of teachers who report being visited by SMC (in previous academic year)	25%
Enrollment	
Other activities of SMCs (percentage that do activity)	
Overseeing construction/repairs/equipment	96%
Approve school budget	94%
Recruit teachers	84%
Raise resources in cash or in kind from parents, community, NGOs	82%
Overseeing brainstorming to identify, or implementation of, ways to improve attendance, teaching or learning	73%
Head Teacher relationship with SMC:	
Head teacher agrees or strongly agrees that “The SMC has provided me with adequate autonomy in making important decisions regarding school operation.”	95%
Head teacher agrees or strongly agrees that “The SMC tries to directly control daily decisions that I need to take to run the school”	13%

Parents and Community Members. Underlying many decentralizing reforms around the world is the aim of improving teaching and learning outcomes in schools by involving parents and local community members more in running the schools. The hope is that parents and other community members might achieve better outcomes than bureaucrats in capital city offices, both because their being local gives them greater ability to observe and supervise what is going on in schools and because their having children in the schools gives them greater motivation to work at improving school quality.

For greater parental involvement to improve teaching and learning outcomes in 9th and 10th grade math and science, however, parents must know how to observe and evaluate the quality of what is happening in 9th and 10th grade math and science. Statistics reported above, indicating that many parents have never been to school and far fewer than half of parents completed grade 9 or 10, suggest that parents will find it very difficult to evaluate the quality of teaching and learning at that level. Having much less education than the typical teacher or head teacher, they are also likely to be reluctant to criticize these teachers.

Table 10: Involvement of Parents and Community Members

School succeeded in raising resources from parents or community members	54%
School has a Parent-Teacher Association (PTA)	89%
PTA meeting frequency (percent of PTAs)	
3 to 6 times per academic year or more	80%
Once a month or more	13%
School has had a Social Audit	84%
Parental involvement in Social Audit (Head teacher report of number of parents involved)	
Few	17%
More than a few but less than half	44%
Half or more	39%
Teachers' agreement with "When teachers work hard, parents notice and encourage it"	
Strongly agree	13%
Agree	47%
Disagree	29%
Strongly Disagree	10%
Teachers' agreement with "When teachers experiment with using new teaching methods, parents notice and encourage it"	
Strongly agree	7%
Agree	40%
Disagree	42%
Strongly Disagree	11%

The baseline data suggest that parents and community members are frequently involved in contributing resources to schools (whether in cash or in the form of labor or goods), and that many are involved in Parent-Teacher Association (PTA) activities. Even so, teachers do not see parents as playing a strong role in supervising them or encouraging them to use new teaching practices. As shown in Table 10, 54% of head teachers report that schools succeeded in raising resources from parents or community members, and 89% report that their schools have PTAs. While most PTAs meet at least 3 times per year, only 13% meet at least once per month, suggesting that they are not deeply involved in day-to-day school management. Among the 84% of schools that completed school audits in the past year, only 39% report that half or more of the parents were involved in the audit process. According to the teacher questionnaire, 60% of teachers agree or strongly agree that parents provide positive feedback for their hard work, while 40% of teachers disagree or strongly disagree. Somewhat weaker support is reported for feedback on new teacher practices: 47% of teachers agree or strongly agree that parents provide positive feedback for their using new teaching practices, while 53% of teachers disagree or strongly disagree.

Other Organizations and Individuals. Other organizations and individuals may also contribute to school operations, and may help shape school management choices either directly (by requiring schools to undertake specific programs or practices in exchange for the resources they provide) or indirectly (by speaking with diverse stakeholders and influencing the local culture and local understanding of education). A strikingly large share of Nepal's schools receive assistance from national and international non-governmental actors. According to baseline data shown in Table 11), 62% receive financial support from international Non-Governmental Organizations (NGOs) or other international agencies, while 43% report receiving financial assistance from "other" actors (not including families of school children and other community members). Unfortunately, it is unclear in the baseline data who these other actors are. (One example of a possible "other" actor comes from the preliminary qualitative research, where one school reported receiving support from a large local employer.) Many schools also receive support in other forms: 43% of schools receive donations of school uniforms, textbooks and other items for needy children; 11% report donations to pay for needy children's tutoring fees; 42% report donations of materials or equipment, and 11% report support for additional teachers.

Table 11: Involvement of Non-Governmental Organizations (NGOs) and Others Actors

School received financial support from NGO or other international agency	62%
School received funds from "others"	43%
Other types of support from any non-governmental source	
Donations of uniforms, textbooks or other support for needy children	43%
Paying tutoring fees for needy children	11%
Donating materials or equipment	42%
Providing teachers	11%

Management Quality. The baseline statistics presented thus far suggest that of all the actors that might be providing school management inputs to school operations, Head Teachers are the actors most likely to be involved, though the time they have available for management activities is limited. School Management Committees are also involved, though they tend to be more involved in fund raising and physical infrastructure and equipment than with oversight of classroom activities. NGOs and other actors may also influence school management choices in many schools. This sub-section seeks to examine what quality of school management is achieved by these actors. While it is impossible to measure the quality of school management directly, this study seeks to shed light on the level and range of variation in school management quality by looking at the incidence and frequency of diverse practices that are likely to be associated with good management and by eliciting teachers' opinions regarding the leadership roles played by head teachers and other actors.

Consider first the formal school management practices reported in Table 12. Most sample schools have taken formal school management steps required by policy. For example, most schools (97%) have School Improvement Plans (SIPs). Schools are required to submit SIPs to their District Education Offices annually. SIPs are intended to be developed through a participatory process that identifies school needs, so that the government can take these needs into account when providing budgets to the schools. Most schools (84%) also completed Social Audits in the 2016-17 academic year. In the interest of enhancing transparency and accountability to parents and community members, schools are required to undertake annual social audits, in which they present information on school activities and performance in front of parents. How effective these audits are in increasing accountability may be questioned, both by the low levels of education among parents (as reported above) and because in most cases less than half of the parents participated (in only 33% of schools did more than half of the parents participate).

Table 12: School Management Practices

Practices required by policy	
School has a School Improvement Plan	97%
School completed a Social Audit in the last academic year	84%
School completed a Social Audit in the last academic year *and* more than half of parents participated	33%
Other school-wide practices	
When a teacher is absent, his or her class is taught by another teacher	70%
When a teacher is absent, his or her class is canceled or without adult supervision	20%
Offer prizes for high performing students	73%
Offer prizes for teacher attendance or teacher performance	29%
Head teachers report preparing/using annual instructional plans for grade 9 and 10 math and science	89%
Have formal rules regarding how much homework to assign	29%
Have formal rules regarding academic assessments	86%
Have an entrance exam for grade 9	30%
Have an entrance exam for grade 9 *and* require students who fail to start at a lower grade	15%

The remaining indicators of management quality in Table 12 focus on procedures and processes that are part of the daily operation of the school. One basic indicator is what happens when a teacher is absent (see below for information on the frequency of teacher absences). In many cases (70%), another teacher teaches the class, but it is worrisome that in 20% of cases the students are either sent home or left without adult supervision in their classrooms. Covering classes for absent teachers at the 9th and 10th grade math and science level may be particularly difficult, because most schools in the sample have only one teacher who teaches a given subject (i.e. math or science) in both 9th and 10 grade (and often other grades as well).

Evidence on use of prizes to motivate students and teachers for high performance is mixed. While almost three fourths (73%) of schools offer prizes for high performing students, only a little more than one fourth (29%) offer prizes for teachers who perform well and/or have high rates of attendance.

The last five rows in Table 12 pertain to instructional and assessment rules, reflecting schoolwide efforts to implement procedural and quality standards. Nearly 9 out of 10 (89%) head teachers report that their schools prepare and/or use annual instructional plans for mathematics and science in grades 9 and 10, and a similar large majority (86%) have formal rules regarding academic assessments. On a less positive note, only a little more than one fourth (29%) of schools have formal rules on the amount of homework assigned. Given that many head teachers and teachers report that students entering below grade level is a major problem, and given the likely inflow into schools at grade 9 of students who pursued primary and lower secondary grades in other schools, a potentially important management question faced by schools is how to evaluate students' grade levels when they attempt to enter the school at grade 9, and whether and how that information is used for making a grade placement. Baseline data reveal that only 30% of schools require students attempting to enter the school at grade 9 to take an entrance exam, and only 15% have such an exam *and* require that students who fail the exam enter the school at a lower grade. (Another 6% have an entrance exam and deny admission to students who fail.)

The remaining indicators of management quality focus on the head teacher, and attempt to measure the head teachers' involvement in coordinating important collective activities and in motivating, supervising and encourage the professional growth of teachers. The results are shown in Tables 13, 14 and 15. As seen in Table 13, while most head teachers (67%) report holding meetings with teachers one or more times per month to discuss school activities, only 32% report holding meetings with teachers to discuss new teaching methods once or more per month, and 36% of head teachers report that they met less than three times per year to discuss this topic.

Table 13: Management Quality: Frequency of Head Teacher Meetings with Parents and Teachers

How often do you [head teacher]…	**< 3 times/year**	**3-6 times/year**	**≥ 1 time/month**
… meet with teachers to discuss school activities?	5%	27%	67%
… meet with teachers to describe a new teaching method?	36%	32%	32%
… organize meeting with parents, community or donors to solicit financial or other aid?	62%	26%	13%
… gather parents to discuss schools activities or policies?	52%	40%	8%

Head teachers appear to meet less frequently with parents and other community members than with teachers. Only 13% (39%) report organizing meetings at least once a month (at least 3 to 6 times per year) with parents, community members or donors to solicit aid. Only 8% (48%) report gathering parents at least once a month (at least 3 to 6 times per year) to discuss school activities or policies.

One of the most important tasks of a head teacher is to support other teachers in the classroom by observing them teaching and providing feedback on how to improve their teaching. Head teachers' reports about such activities are show in Table 14, while teacher reports are shown in Table 15. Even by head teachers' reports, it is relatively rare for head teachers to observe a teacher for an entire class period – only 15% of head teachers report doing this monthly (or more often), while 57% say that they do not do this at all. When asked about the frequency of visiting the typical 9th or 10th grade teacher's classroom for only part of a class period, head teachers' responses were more encouraging, but still only 59% of head teachers report that they do this one or more than one times per month.

Table 14 Head Teacher Reports of Observing Teachers and Providing Feedback

Think about your interactions with an average teacher of 9th or 10th grade math or science. How often do you [head teacher]…	**Not at all**	**1-2 times/year**	**3-6 times/year**	**≥ 1 time/month**
… observe a full class period for this teacher?	57%	14%	16%	15%
… observe part of a class period for this teacher?	7%	7%	27%	59%
… gather feedback about this teacher from other teachers, students or parents?	7%	23%	33%	36%
… provide informal, verbal feedback to this teacher?	4%	16%	33%	46%
… provide formal, written feedback to this teacher?	83%	7%	4%	5%

Of course, head teachers' observation of teachers is unlikely to have much effect on teaching quality if head teachers do not give teachers feedback. By head teachers' reports, informal verbal feedback from them to other teachers appears to be somewhat common, with 46% of head teachers reporting that they provide such feedback to the average 9th or 10th grade math or science teacher one or more times per month, and another 33% reporting this happening at least three to six times per year. In stark contrast, written feedback is rare. The vast majority of head teachers (83%) state that they do not provide written feedback to teachers.

Teachers' reports suggest that head teacher visits and feedback may be somewhat less common than suggested by head teachers, but paint a similar picture. Table 15 shows answers regarding the frequency of head teacher activities from the *median* teacher for each school. Comparing Tables 14 and 15, both head teachers and other teachers agree that the head teacher rarely observes teachers teach for the whole period; 57% of head teachers say that this never happens, which is only slightly lower than the 65% of the median teachers who says that it never happens. A larger difference is seen in that 15% of head teachers report that they do a full class period observation for the average teacher one or more times per month, while only 6% of median teachers report being observed for a whole class period at least once per month. A similar pattern is seen for the head teacher observing teachers for part of a class period. Only 7% of head teachers report that this never happens, which is slightly lower than the 11% report of the median teacher. Similarly, 59% of head teachers report that they do this at least once per month, while this number is somewhat lower (44%) for the median teacher.

Table 15: Teacher Reports of Being Observed and Receiving Feedback and their Opinions Regarding Management by Head Teacher
(Distribution across schools of median teacher response)

How often does head teacher…	**Not at all**	**1-4 times/year**	**1 time/month**	**≥ 1 time/week**
… observe a full class period?	65%	29%	3%	3%
… observe part of a class period?	11%	45%	31%	13%
… provide informal, verbal feedback?	11%	55%	26%	8%
… provide formal, written feedback?	94%	5%	1%	0%
	Not at all important	**Somewhat important**	**Very important**	
How important is the role of the head teacher in providing school leadership?	1%	11%	87%	
	Not at all effective	**Somewhat effective**	**Very effective**	
How effective is the head teacher in providing school leadership?	2%	49%	49%	

Teacher reports on informal and formal feedback from head teachers also broadly agree with head teacher reports. Comparing Tables 14 and 15, informal/verbal feedback is quite common but formal/written feedback is very rare. Again, head teachers report that both happen somewhat more frequently than the median teacher reports the same.

While most teachers believe that head teachers have a very important role to play in providing leadership for their schools, less than half believe their head teachers are providing excellent leadership. When teachers were asked how important a role the head teacher should play in providing school leadership, 87% said this role was very important, but only 49% indicated that the head teacher was very effective in fulfilling this role.

Instruction Time. Earlier sub-sections described the challenges and resources that schools must work with, the actors who might manage the schools with the aim of achieving high learning outcomes in the face of challenges and resource constraints, and some evidence suggesting that the quality of school management is often weak and varies significantly across schools. This and the remaining sub-sections examine the teaching and learning outcomes that the schools achieve. This sub-section examines instruction time, while subsequent sub-sections examine the teachers and teacher training that schools mobilize, teaching practices, and student learning outcomes.

Total instruction time per year varies significantly across schools, primarily because of variation in the number of days school is held (rather than because of variation in hours of instruction during typical school weeks without holidays). Most school schedules call for class periods of approximately 45 minutes in length (Table 16), and approximately 30 hours of instruction per week (in a week with no holidays). The number of days of instruction (including any days on which at least one class period of instruction took place) ranges from 184 at the 25^{th} percentile to 208 at the 75^{th} percentile for 9^{th} grade, while ranging from 175 at the 25^{th} percentile to 198 at the 75^{th} percentile for grade 10. For both grades, the schools at the 75^{th} percentile had 13 percent more days of instruction in the last academic year than schools at the 25^{th} percentile.

Median values of total instruction time per academic year are on par with that in relatively high performing government school systems. For example, in the state of Massachusetts in the United States, secondary school schedules are required to include at least 990 hours of instruction on at least 180 school days. With median hours of instruction per week of 30 spread over 6 days of instruction, the typical schools in our sample average 5 hours of instruction per day. When multiplied by the median days of instruction (187 for grade 10), this implies 935 hours of instruction per year.

Table 16: Instruction Time

	25^{th} percentile	**Median**	**75^{th} percentile**
How many minutes long is each period?			
Grade 9	40	45	45
Grade 10	40	45	45
When there is no holiday, how many hours of instruction are conducted in a week? (minutes per period multiplied by periods per week)			
Grade 9	29	30	30
Grade 10	29	30	30
On how many days were classes conducted in the last academic year?			
Grade 9	184	195	208
Grade 10	175	187	198

Teacher Characteristics and Training. How instruction time translates into learning depends on the quantity of teachers per student and teacher quality, as well as the nature of teaching practices. Basic characteristics of Grade 9 and 10 math and science teachers, along with their contract status and their educational background, are reported in Table 17. These teachers are overwhelmingly male: 98% of math teachers and 91% of science teachers are men. The average age of these teachers is around 35 years, and the average years of teaching experience is 11. About one third of these teachers lived in the communities where they currently teach before they became teachers, although this figure is somewhat higher for math teachers (41%) than for science teachers (25%). These percentages (of teachers who are from the local community) are lower than for head teachers, perhaps indicating that it is even more difficult to find adequately qualified secondary math and science teachers in many communities than it is to find adequately qualified head teachers (who may teach primary grades).

Table 17: Characteristics of Teachers of Grade 9 and 10 Math and Science

	All teachers	Math teachers	Science teachers
Male	94%	98%	91%
Age (mean, in years)	35	37	34
Teaching experience (mean, in years)	11	12	10
Lived in community before teaching	35%	41%	25%
Contract types (from teacher questionnaire)			
Permanent	27%	30%	25%
Temporary	18%	18%	17%
Rahat	32%	31%	34%
Other government	3%	3%	3%
Locally funded	15%	15%	17%
Other	4%	4%	2%
Highest Degree in Any Field			
I.Ed./PCL	10%	8%	11%
Bachelors	51%	48%	55%
Masters	38%	43%	33%
Above Masters	1%	0%	1%
Highest Degree in Education			
None	24%	18%	37%
I.ED/PCL	6%	5%	7%
Bachelor of Education, 1-year	21%	17%	26%
Bachelor of Education, 2-year	2%	3%	1%
Bachelor of Education, 3-year	25%	29%	17%
Master of Education	22%	29%	15%
Received 10-month Government Pre-Service Teacher Training	17%	18%	16%
Have Completed Neither a Bachelors or Higher Degree in Education Nor the 10-month Government Training	5%	4%	7%
Has Bachelor Degree in Math or Science	73%	71%	78%

Grade 9 and 10 mathematics and science teachers work under a wide variety of contracts, as shown in the middle of Table 17. Slightly more than one in four had permanent contracts, about one in five had temporary contracts, and about one third (32%) had Rahat contracts (literally “relief” contracts, these are temporary contracts for teachers funded by a central government funding stream focused on schools with low teacher-student ratios). Finally, about one in ten was “locally funded” and even fewer had an “other government” or “other” types of contract.

Most (but not all) of these teachers have at least some pre-service training in education, though the levels of such training vary greatly. Nearly half have a Bachelor’s degree in education (some having abbreviated Bachelors’ degrees in education for graduates with Bachelors’ degrees in other fields), and another 22% have Master’s degrees in education, but nearly a quarter have completed no education degree. Many of those with no university degree in education have taken a 10-month training offered by the government as an alternative route to fulfilling education degree requirements; but 5% of secondary math and science teachers have neither a Bachelors in education nor the 10-month training.

Table 18 describes the teaching responsibilities of secondary math and science teachers. Most schools (68%) have one teacher who teaches math for both grades and another teacher who teaches science for both grades. Of the teachers who teach math and science in these two grades, a little less than half (48%) teach math but not science, a somewhat smaller percentage (43%) teach science but not math, and only about one in ten (9%) teach both subjects. In fact, many of these teachers (64%) teach math or science in grade 8 as well, and most of them (80%) also teach other subjects or other grades. Table 18 also shows that the average teacher taught 29 periods per week (with the typical period 45 minutes in length), and that 7% of these secondary math and science teachers were also the head teachers for their schools.

Table 18: Teaching Responsibilities for Grade 9 and 10 Mathematics and Science Teachers

Schools with only 1 math teacher and only 1 science teacher for grades 9 and 10	68%
Of grade 9 and 10 math and science teachers in the survey:	
Teach math but not science in grades 9 and/or 10	48%
Teach science but not math in grades 9 and/or 10	43%
Teach both math and science in grades 9 and/or 10	9%
Teach grade 8 math or science	64%
Teach in other grades or subjects	80%
Average periods taught per week	29
Also are head teachers	7%

One way policymakers and school managers might hope to improve learning outcomes is by helping teachers acquire in-service training. Table 19 indicates that some, but not all, Grade 9 and 10 math and science teachers have received in-service training in their subject areas. More specifically, 30% of math teachers received math training under the SSRP training program (and 8% received science training), and 21% of science teachers received science training under the SSRP training program (and 8% received math training). (Most of the math training reported for science teachers and science training reported for math teachers is accounted for by teachers who teach both math and science, who are included in statistical calculations for both math and science teacher categories.) In addition, a small percentage of both types of teachers report that they have already had SSDP math or science training.[9] Nearly 40% of math teachers and nearly one third of science teachers report having received other government trainings related to math or science. Finally, 25% of math teachers and 22% of science teachers report having government training in other subjects or other types of skills.

Table 19: In-Service Training History of Teachers of Grade 9 and 10 Math and Science

	Math teachers	Science teachers
SSRP Math Training	30%	7%
SSRP Science Training	8%	21%
SSDP Math Training	6%	2%
SSDP Science Training	1%	5%
Other Government Math or Science Training	39%	32%
Other Government Training	25%	22%
NGO Math Training	20%	3%
NGO Science Training	14%	13%
Other Training	9%	9%
Received Any Math or Science Training	68%	53%
Received Any Training	73%	60%

Many teachers report receiving training from an NGO (non-governmental organization) program. That is, 20% of math teachers report that they received mathematics training from an NGO, and 14% report that they have had science training. Turning to science teachers, 13% report having had science training through an NGO, although only 3% report having received math training from an NGO. Finally, 9% of both math and science teachers also report having had "other" training.

Overall, more than half the teachers (and 60% among teachers who have neither university education degrees nor the government 10-month training) have received in-service training for math or science teachers, mostly from the government, but in some cases from non-governmental organizations (NGOs). Still, one third of math teachers and nearly half of science teachers have not received in-service training related to math or science, suggesting significant needs for pedagogical training specific to math and science subjects.

Among teachers who have received government training, a wide majority report being satisfied with the most recent government training they received. As seen in Table 20, about 90% or more of teachers agreed that the most recent government training they attended helped them become more effective teachers, gave them opportunities to develop new lesson plans, and provided them with sufficient time to exchange ideas with other teachers. These teachers also had relatively high opinions of their trainers, with 81% saying they were well prepared and 84% reporting that the

[9] This is a problem for the larger study, the aim of which is to estimate the impact of SSDP training, because the study design is predicated on teachers in both Phase I and Phase II schools *not* having received SSDP training at baseline. Fortunately, the percentage is not high and is very similar across study arms, and it will be possible to adapt impact estimation methods to account for this training.

trainers demonstrated strong competencies in math and science concepts. The one area where these teachers were somewhat less satisfied, but still more positive than negative, is that 64% of the trainers "strongly demonstrated new types of teaching methods".

Table 20: Opinions of Grade 9 and 10 Mathematics and Science Teachers on Recent Government Math or Science Training

Percent of teachers who agreed that their …	
… training helped them become more effective teachers	95
… training gave them an opportunity to develop new lesson plans	93
… training offered them ample time to exchange ideas with other teachers	89
… trainers were well prepared	81
… trainers demonstrated strong competencies in math and science concepts	84
… trainers strongly demonstrated new types of teaching methods	64

Teacher Attendance. Holding constant their preparation and capacity to learn, we expect students to achieve higher learning outcomes when their teachers are present at school more frequently, and when their teachers use more effective teaching practices. This subsection examines teacher attendance, using data from the head teacher questionnaire. As above, the focus is on math and science teachers in Grades 9 and 10.

Head teacher responses to questions about attendance rates and other dimensions of performance for individual teachers of 9th or 10th grade math or science are reported in Table 21. Attendance by these reports is high compared to the low rates that have drawn attention in other developing countries, where reports of 20% or 25% of teachers being absent on a given day seem to be common (Chaudhury, et al. 2006). Approximating average attendance in the top three categories by the category's lower bound, while approximating the average rate in the lowest category by 60%), the weighted average reported attendance rate is 85%. Instead approximating average attendance rates within the top three categories by the midpoint rates (e.g. 95% for the top category) and raising the average for the bottom category to .65, the weighted average attendance rate rises to 90%. True attendance rates may also be somewhat lower if head teachers feel the need to demonstrate management effectiveness to interviewers.

Teaching Practices as Described by Head Teacher, Teacher and Student Questionnaires. Given the perception that even high quality teacher training programs often fail to change actual classroom teaching practices, this study gives particular attention to describing the types of teaching practices that are used in the classroom. A first approach to measuring classroom practices is simply to include questions on them in the head teacher, teacher and student questionnaires.

According to head teacher reports, the assignment of homework for 9th and 10th grade math and science is frequent in most schools. As shown in Table 21, head teachers report that 58% of Grade 9 and 10 math and science teachers assign homework every day, 35% assign homework 2 to 3 days per week and only 6% assign homework once per week or less frequently, so that homework is assigned more than once per week by 93% of the teachers.

Table 21: Teacher Attendance and Teaching Practices, as Reported by Head Teacher

Attendance (% of Grade 9 and 10 math and science teachers in attendance categories):	
Present 90% or more of days	60%
Present 80-90% of days	34%
Present 70-80% of days	5%
Present less than 70% of days	1%
Frequency of assigning homework:	
Every day	58%
2-3 times per week	35%
One time or less per week	6%
Frequency of conducting student assessments (tests):	
At end of every trimester and at end of year	99%
More than once per trimester	83%
Head teacher and teacher use of assessments:	
Determine promotion or retention of students (head teacher)	93%
Provide results to parents (head teacher)	88%
Use to evaluate teachers (head teacher)	85%
Determine which students receive free tutoring	49%
Group students by academic ability	62%
Assign high-performing student to work with a low-performing student	39%
Teachers knowledge of the subject matter taught (reported by head teacher):	
Excellent	27%
Generally good	67%
Partial	6%
Poor	0%
How often teacher uses interactive classroom activities (reported by head teacher):	
Almost every day	11%
Frequently, but not every day	33%
Sometimes	46%
Rarely	9%
Don’t know	1%
How interested teacher is in trying new teaching methods (reported by head teacher):	
Highly interested	42%
Moderately interested	42%
Slightly interested	15%
Completely uninterested	1%

Head teachers also report reasonably frequent use of assessments for 9^{th} and 10^{th} grade math and science teachers. As shown in Table 21, almost all head teachers (99%) report that 9^{th} and 10^{th} grade math and science students are required to take at least one assessment at the end of each trimester and at the end of the school year, and 83% of head teachers report using assessments more frequently than this. When asked whether the assessments were used for various purposes, head teachers responded positively for many uses, such as deciding whether to promote a child to the next grade or retain the child (require the child to repeat) (93%), providing parents with information on their children's academic progress (88%), evaluating teacher performance (85%), and determining which children are provided free tutoring services from the school (49%). By offering the list of uses to head teachers, the questionnaire may have encouraged them to report more positive uses than is true in practice. The fact that 49% of head teachers report that assessments are used to help determine which students receive free tutoring, while only 18% to 26% of schools offer free tutoring to 9^{th} and 10^{th} grade math and science students reinforces this concern.

Two final observations provide mixed evidence of organized school or teacher efforts to improve teaching and learning by grouping students. As seen in Table 21, 62% of head teachers report that students (within a grade) are grouped by academic ability, either across or within sections. Such grouping could be important for teachers attempting to differentiate instruction across students of different levels or abilities. A much smaller percentage (39%) report the practice of pairing individual low-performing students with high-performing students within classrooms. Such pairing may improve learning for both low-performing students (who benefit from the greater understanding of their partners) and for high-performing students (who might benefit from practice with explaining concepts to their partners), though only if class time is devoted to student discussion in pairs or small groups. As will be seen below, teachers seem to devote little class time to such small group activity.

Head teachers were also asked to assess how knowledgeable Grade 9 and 10 math and science teachers were of the subjects that they teach. Overall, they gave their teachers quite high ratings, with 27% rating those teachers' knowledge as excellent, 67% rated as good, only 6% as "partial", and none rated as poor. Head teachers may be inclined to overstate their teachers' competence in the interest of projecting a good image for their schools. It may be that better measures of teacher subject knowledge can be obtained from administering tests to teachers or asking them to grade student assignments constructed for this purpose.

According to head teachers (Table 21), less than half of teachers use demonstrations, competitions and other interactive classroom activities at least "frequently," but most teachers are at least moderately interested in trying new teaching methods. Head teachers report that 11% of their grade 9 and 10 math and science teachers use demonstrations, competitions or other interactive classroom activities every day, and another 33% use them "frequently, but not every day", while 46% said that teachers used such methods only "sometimes" and 9% said that they were used them "rarely" (and 1% said that they did not know). Head teachers also stated that they thought that 42% of these teachers were highly interested in learning new teaching methods, and another 42%

were moderately interested, while only 15% were "slightly interested" and only 1% were not at all interested.

The time teachers spend preparing for class is of interest for two reasons. First, it sheds light on the effort that teachers devote to their work. Second, it may shed light on a potential obstacle to the spread of new classroom teaching practices. Teachers who are used to teaching without much preparation may be reluctant to work at implementing new, student-centered learning practices, which require significant new pre-class preparation (at least during the first year or two). Table 22 reports that nearly half of teachers (41%) prepare for 15 minutes or less for each class period that they teach. For teachers teaching multiple subjects or grades, it may be difficult to spend more time preparing than this. When developing new curricula and teaching methods, thought should be given to helping teachers overcome demands for large preparation time per class session, perhaps by providing them with lesson plans or scripts. (Somewhat surprisingly, reported preparation does not appear to be systematically higher for teachers who are "new," in the sense of having 2 years or less of teaching experience.)

Table 22: Time Spent Preparing Out of Class for Each Class Period (From Teacher Questionnaire)

None	6%
1-15 minutes	35%
16-30 minutes	26%
31-60 minutes	15%
1-3 hours	17%
More than 3 hours	1%

Head teacher reports regarding teaching practices could be biased in a generous direction. Table 23 examines *students'* reports of those practices among Grade 9 and 10 teachers of math and science. Both math and science students report that teachers use class discussions very frequently, with well over half having discussions every day and another 22-25% of teachers holding discussions 2-3 times per week. According to these students, only about 5% of math classes and 8% of science classes have discussions less than once per week, or not at all. Students' reports of teachers using interactive activities are not very different from the reports of the head teachers in Table 21. For example, students report that 30% (math) to 37% (science) of teachers use such activities either every day or 2-3 times per week, while head teachers report that such activities are used by 44% of teachers almost every day or "frequently".

Table 23: Student Reports of Teacher Teaching Practices

	Math class	Science class
How often does your teacher use class discussion?		
Every day	65%	57%
2-3 times per week	22%	25%
Once per week	6%	9%
Less than once per week	2%	3%
Never	3%	5%
How often does your teacher use contests, games, interactive activities?		
Every day	15%	17%
2-3 times per week	15%	20%
Once per week	27%	27%
Less than once per week	7%	7%
Never	35%	28%

Teaching Practices as Measured Using the Stallings Classroom Observation Method. In the interest of gaining a new perspective on teaching practices and student engagement in Nepal's classrooms, the research team incorporated into the baseline the use of unannounced classroom observations, following the Stalling's "Classroom Snapshot" observation system. As indicated above, trained observers recorded "snapshot" observations of teacher and student activities during actual class periods. The observation days were unannounced, to avoid teachers' preparation of lesson plans and activities specifically for the observation, but teachers are nonetheless aware of being observed during the observation session, and so are likely to be performing to the best of their ability. The observation data should thus be interpreted as describing the best performance of which teachers are capable. Tables 24 and 25 provide statistics derived from the 668 classroom observations pertaining to 373 teachers.

It appears that teachers on their best behavior are indeed actively engaged in learning activities on average for a large fraction of their class time. Table 24 reports the proportions of teacher "time" (as measured by the percentages of "snapshots") devoted to the various categories of activities captured by the Stallings observations. Across all teachers and classes, 77% of teachers' class time is devoted to teaching activities, while 13% is spent on classroom management, and 10% is spent off task.[10] To put these statistics into perspective, high performing primary schools in the United States devote about 85% of the time to learning activities or instruction, while the instructional time in such schools in the Latin American and Caribbean countries between 2009 and 2013 was 65% or less (Bruns and Luque, 2015).

Interestingly, over half (51%) of the snapshots in which teachers are out of the classroom are the first snapshots within class periods, suggesting that teachers are arriving late. This happens in

[10] In four classrooms, the teachers were absent the entire period. The results were largely similar (77.5% of the time for learning activities, 12.5% for classroom management and 10% for off-task) even after removing these four teachers from the analysis.

55% of the class sessions observed. Once teachers arrive in the room and realize they are being observed, they are out of the classroom much less. This is consistent with the suspicion that teachers would be out of the classroom more if they did not know they were being observed.

A troublingly large fraction of teachers' active teaching time is spent on more traditional teaching methods that do not require students to engage actively. While 64% of their time is spent on traditional teaching activities, only 13% is spent on question-and-answer or discussion. It is especially significant that teachers spend little time on interactive activities even while being observed, and even while probably knowing that current best practices encouraged by government authorities and education experts involve more student-centered learning approaches.

Table 24: Teacher's Activities in the Classroom (Stallings Classroom Observation)

	Proportion of Teacher Activity
Traditional teaching activities	64%
Explanation/Demonstration/Lecture	39%
Assignment/Class Work	15%
Reading Out Loud	3%
Copying	3%
Practice and Drill	2%
More interactive teaching activities	13%
Question and Answer/Discussion	13%
Other teacher tasks	13%
Verbal Instruction*	5%
Classroom Management with Students	5%
Classroom Management Alone	2%
Discipline	0%
Time off task	10%
Teacher Out of Classroom	9%
Social Interaction between Teacher and Students	1%
Social Interaction with Another Adult/Teacher not Involved	1%

* This refers to instructions on how students are expected to complete their work, not on academic content.

Table 25 reports observations on use of group work and teaching materials associated with the teacher's main activity. After observing the teacher's main activity and how many students are engaged with the teacher, observers were asked to identify (a) whether the activity involved group work and (b) what materials were used during the activity. (Enumerators were asked to check off as many items as were relevant to the teacher's main activity in a list that included use of cooperative/group work and use of the various materials shown in Table 25.) Given the research team's interest in more interactive student-centered teaching practices, and the frequency with which group work is mentioned as an example of such practices in Nepal education discussions, the team emphasized the primacy of observing group work during enumerator training. In the manual prepared for this classroom observation, the observers were asked to check cooperative/group activity if at least two students were collaborating on an academic topic. Even so, cooperative/group work connected to the teacher's main activity was reported in very few cases (just 0.5% of snapshots).

Table 25: Use of Group Work and Materials in Teacher's Main Activity When Teachers are Teaching*

Percent of snapshots involving cooperative/group work in teacher's main activity	0%	[rounded from 0.47%]
*Percent of snapshots involving teacher's use of materials**		
No materials	19%	
Textbooks	26%	
Notebooks/writing material	15%	
Blackboard/whiteboard	57%	
Learning aids	6%	
Information and Communication Technology (ICT)	0%	[rounded from 0.47%]

*This table employs data only on snapshots during which the teacher's main activity was either traditional teaching or more interactive teaching. The total time using different materials exceeds 100% as teachers sometimes use more than one material at a given time.

Some students might be involved in group work even when teachers' main activities are teaching other students using methods that do not involve groups, classroom management, or off task activities. The data indicate, however, that among snapshots in which any students are reported as engaged in some activity other than the teacher's main activity, students are involved in group work in only 0.25% of cases. Thus the evidence indeed suggests very little use of group work.

The most common materials used by teachers are blackboards or whiteboards. Blackboards or whiteboards are used by teachers more than half of the snapshots (57%), while textbooks are used one fourth of the time. Use of learning aids, information and communication technology (ICT), and cooperative/group activity is limited.

Teachers' choices of materials are constrained by the materials available. Table 26 indicates the percentages of classrooms with various sorts of equipment. Most observed classrooms are equipped with basic resources (e.g., chalk, markers), but interactive and high-tech equipment is rare. Less than 1% of classrooms had a laptop or LCD projector, and no classrooms had a digital whiteboard, television or toys/games displayed.

Given that teachers are making little use of group work, and are primarily using methods intended to involve many students, we can examine their effectiveness at engaging students in learning by examining the numbers of students who appear (to the classroom observers) to be actively engaged with the teacher when the teacher is using traditional or modern teaching methods. Table 27 describes the extent of student involvement in their teachers' main activity, separately by category of teacher activity. While teachers are mostly engaging at least 6 students when they are teaching, they are engaging all students only 36% of the time when using traditional teaching methods, and only 41% of the time when using the more interactive methods of question-and-answer or class discussion.

Table 26: Percentage of Classrooms Having Standard Pedagogical Materials

Chalk/markers	99%
Whiteboard	95%
Textbooks for students	67%
Textbook for teacher	55%
Blackboard	22%
Charts/posters	16%
Science/math equipment	4%
Maps	3%
LCD projector	0.5%
Laptop	0.3%
Digital whiteboard	0%
Television	0%
Toys/games	0%

Table 27: Student Engagement in Teachers' Main Activities

Teacher Activities	**Percent Distribution of Snapshots Across Student Engagement Categories**				
	No student	**One student**	**2 to 5 students**	**6 or more (but not all)**	**All students**
Traditional teaching	1%	1%	6%	57%	36%
More interactive teaching	0%	3%	9%	47%	41%
Classroom management	14%	8%	12%	26%	39%
Off task	92%	0%	2%	3%	3%

The engagement figures of Table 27 may understate students' involvement in learning, if some of the students who are not engaged in the teachers' main activity are nonetheless involved in learning activities on their own. As indicated in Table 28, however, when students are not involved with the teacher's main activity, they are sometimes copying and only rarely doing classwork, but seem mostly to be either engaging with each other socially or are completely unengaged.

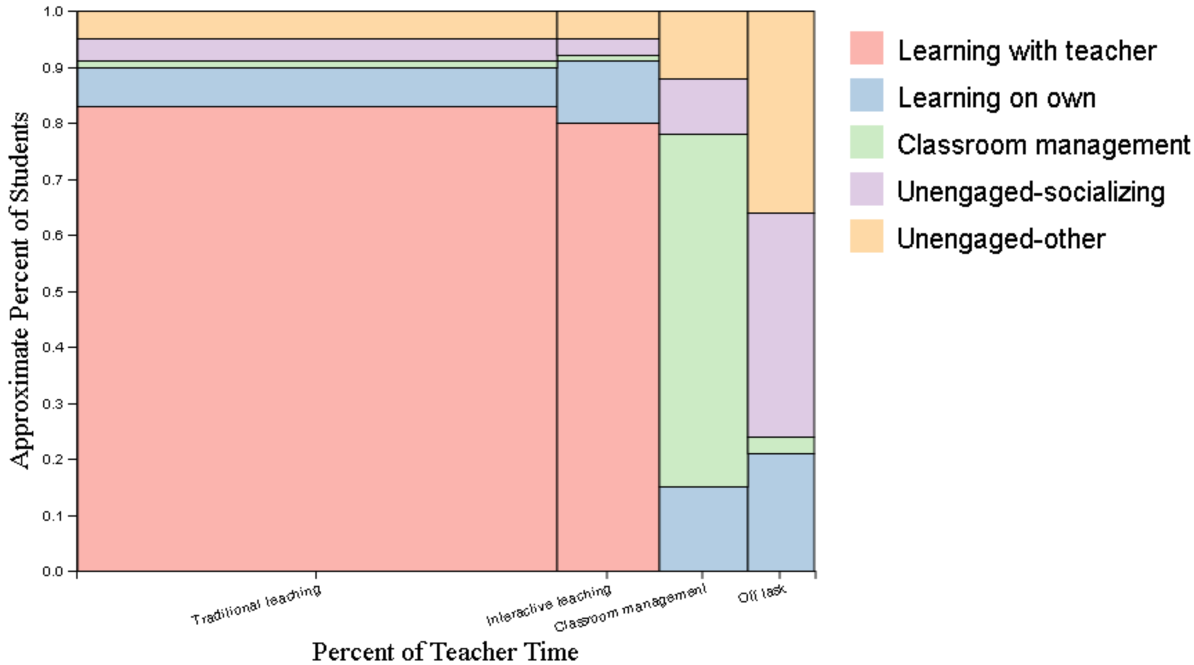
Table 28: Student Engagement in Activities Not Involving the Teacher

Activity	**Among Snapshots in Which Any Students are Not Engaged with Teacher, Percent in Which At Least 2 Students Are Engaged in the Activity**
Reading out loud	1%
Explanation/demonstration/lecture	0%
Question and answer/discussion	1%
Practice and Drill	1%
Assignment/class work	9%
Copying	32%
Classroom management	1%
Social interaction	38%
Fully unengaged (e.g. sleeping, playing alone)	52%

Figure 2 offers an approximate summary description of how teacher and student activities are distributed during the average observed class session.[11] The horizontal axis measures off 100 percent of class time, allowing the width of the vertical bars to indicate the percentages of class time (i.e. the percent of snapshots) that teachers devote to different categories of activity. Here the activities were organized into just four groups: traditional teaching, interactive teaching (i.e. question-and-answer and discussion), classroom management, and off task activities. The vertical axis measures approximate percentages of students, allowing the heights of colored segments within bars to indicate the percentages of students engaged in various activities while the teacher is engaged in the main activity associated with the bar. Here the student activities were organized into five groups: learning with the teacher (i.e. the teacher is teaching and the students are engaged with the teacher), learning on own (i.e. the student is not engaged with the teacher but is nonetheless engaged in a learning activity such as copying or classwork), classroom management (whether with or without the teacher), unengaged-socializing, and unengaged-other (e.g. sleeping, playing, looking out the window).

[11] The authors gratefully acknowledge Thomas Schaffner's assistance with writing a script (available for use at https://afriendlyrobot.github.io/stallings-visualizer/) to produce Figure 1 (instructions forthcoming).

Figure 2: Average Distribution of Student Class Time by Teacher and Student Activity*



* As described in the text, the percentages of students are approximated based on category responses regarding the numbers of students involved in various activities. The figure is based on 6,483 snapshots.

The percentages of students engaged in different activities in Figure 2 are only roughly approximate, because the data only indicate, for 10 sets of possible activities (i.e. engagement with the teacher, and engagement in nine other sets of activities not involving the teacher), whether 0, 1, "2 to 6", "6 or more" (but not all), or "all" students were engaged in the activity. The approximate percentages engaged in each set of activities were calculated making use of data on the maximum number of students in the classroom during the class period, and approximate counts of students engaged in the activities. The count engaged in an activity was set to 0, 1 or the maximum number of students in the room when the number recorded in the data was 0, 1 or "all." When the reported number was "2 to 6," the count was set to 3.5, unless this would result in a residual number of students in the "6 or more" category that would reduce the average number of students engaged such an activity below 6, in which case the count was set to 2. When the reported number was "6 or more," the count was set to the residual number of students (after accounting for students engaged in activities involving 1, or "2 to 6" students) divided by the number of activities associated with "6 or more" reports. Approximately 7 percent of observations were dropped because of incomplete data or inconsistencies, such as reports of "all" students engaged in two different activities.

Overall, a reasonably large fraction of the area in Figure 1 is pink or blue, indicating that for the average student, a large fraction of time is spent in educational activities (mostly traditional ones), either with or without the teacher. The significant borders of green, purple and orange, however, indicate significant potential to increase learning time for the average student, by improving the efficiency of classroom management, reducing teacher time off task, or improving engagement among students who are currently remaining detached from teaching and learning activities.

Student Attendance. A first student outcome to examine is student attendance. Poor student attendance can be an obstacle to teaching and learning. Making use of head teacher reports regarding the numbers of students enrolled in grades 9 and 10 and the numbers of students present on the most recent regular school day suggest that the average attendance rate is reasonably high (with median rates near 90%), though far from perfect (Table 29). It is possible that Head Teachers' estimates are optimistic. In many cases (73% of schools for grade 8 and 82% for grade 9), however, the number of students present on the day the baseline survey was administered was lower than the number reported by the Head Teacher as present on a the most recent regular school day. The great majorities of head teachers (88%) and teachers (78%) also indicate that poor student attendance is a major problem in their schools.

Table 29: Student Attendance (Head Teacher Reports)

	25th percentile	Median	75th percentile
Student attendance rate (percent)			
Grade 9	75	86	91
Grade 10	79	90	95

Student Performance. The schooling outcomes of ultimate interest in this research are student learning outcomes, as measured by performance on academic achievement tests. Table 30 describes head teachers' reports regarding participation and achievement in the previous academic year on the Secondary Education Examination (SEE), a high stakes exam given to students at the end of grade 10 throughout Nepal. Most 10th grade students in these schools (97% in the median school) take the SEE, but few perform at the highest level (with scores of 3.2 or higher). It should be noted, however, that at the 95th percentile 37% of students achieved these high scores, and at the 99th percentile 70% of students achieved these high scores, so performance on the SEE varies a great deal across schools. Unfortunately, problems with questionnaire administration rendered it impossible measure SEE performance relative to more modest performance thresholds.

Table 30: Secondary Education Examination (SEE) Participation and Performance

	25th percentile	Median	75th percentile
Percent of enrolled students taking the SEE	92%	97%	100%
Percent of SEE takers who achieved high scores of at least 3.2	0%	0%	5%

Given the centrality of student performance measures to the larger evaluation study, the research team implemented original academic achievement tests in 8th and 9th grade math and science in all study schools. The assessments were developed by faculty in the relevant education fields at Tribhuvan University, in consultation with the research team. Because the assessments were to be administered early in the academic year, the research team requested that the assessments for grade 8 cover subject content relevant to grades 6 and 7 and the early months of grade 8, and that the assessments for grade 9 cover subject content relevant to grades 7 and 8 and the early months of grade 9. The team furthermore requested that the assessments include questions at diverse levels of difficulty, from questions that can be answered correctly based on simple memorization to questions that require deeper understanding and ability to apply concepts to new situations. Tables 31 through 34 describe the results for the Grade 8 Mathematics, Grade 8 Science, Grade 9 Mathematics, and Grade 9 Science assessments. Grade 8 assessments were administered to 7,651 students, while the Grade 9 assessments were administered to 8,784 students.

Table 31: Academic Achievement Test Results: 8th Grade Mathematics

	25th percentile	**Median**	**75th percentile**
Multiple Choice Section			
Percent attempted	95%	100%	100%
Score (as percent of total points)	25%	30%	45%
Short and Long Answer Sections			
Percent attempted	38%	63%	88%
Score (as percent of total points)	0%	0%	8%
All Sections			
Score (as percent of total points)	11%	16%	22%

Table 32: Academic Achievement Test Results: 8th Grade Science

	25th percentile	**Median**	**75th percentile**
Multiple Choice Section			
Percent attempted	95%	100%	100%
Score (as percent of total points)	40%	50%	60%
Short and Long Answer Sections			
Percent attempted	45%	73%	91%
Score (as percent of total points)	5%	8%	18%
All Sections			
Score (as percent of total points)	19%	26%	35%

Table 33: Academic Achievement Test Results: 9th Grade Mathematics

	25th percentile	Median	75th percentile
Multiple Choice Section			
Percent attempted	95%	100%	100%
Score (as percent of total points)	35%	45%	60%
Short and Long Answer Sections			
Percent attempted	50%	75%	88%
Score (as percent of total points)	0%	8%	20%
All Sections			
Score (as percent of total points)	18%	24%	36%

Table 34: Academic Achievement Test Results: 9th Grade Science

	25th percentile	Median	75th percentile
Multiple Choice Section			
Percent attempted	90%	100%	100%
Score (as percent of total points)	35%	45%	55%
Short and Long Answer Sections			
Percent attempted	50%	67%	92%
Score (as percent of total points)	7%	15%	24%
All Sections			
Score (as percent of total points)	20%	27%	35%

Test scores in both grades and both subjects are quite low. Median overall scores, expressed as percentages of maximum points possible, are 16%, 26%, 24% and 27% for grade 8 math, grade 8 science, grade 9 math, and grade 9 science, respectively. Twenty-fifth and seventy-fifth percentile scores are also low.

Assessment scores may be low for one or more of the following reasons: (a) student learning was weak, (b) student effort during the assessment was weak, or (c) the assessments targeted inappropriately advanced subject content. Before concluding that student learning was weak, therefore, it is important to consider the possibilities that student effort was weak during the assessment and that the assessments targeted inappropriately advanced subject content.

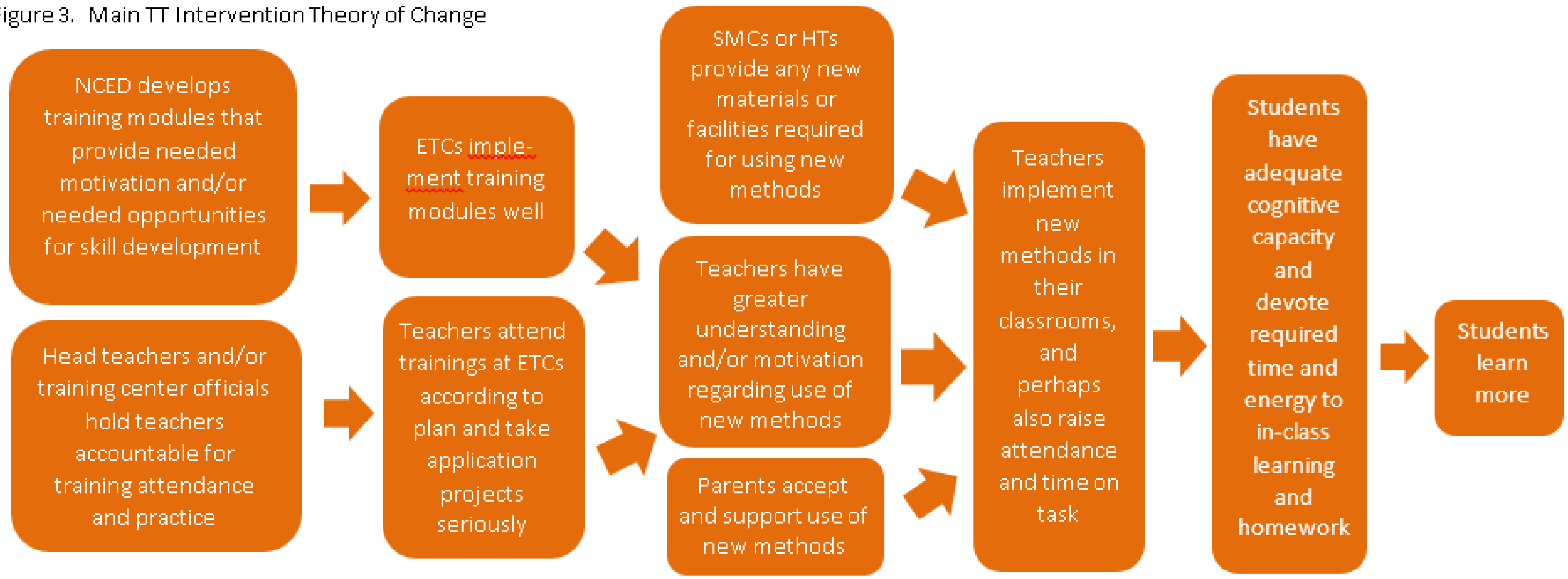
During pre-testing of questionnaires and assessments, the research team observed that in some schools students showed little interest in exerting efforts on the assessments once they were told (during the informed consent process) that these assessments would have no impact on their grades. After further consultation with assessment experts, the team requested that enumerators enlist students' teachers to encourage them to take the assessments seriously. The observation that most students attempted almost all of the multiple choice questions (Tables 31-34) offers some reassurance that students felt compelled to exert some effort on the assessments. The median (and even 25th percentile) scores on the multiple choice sections, though low, are (except in the case of grade 8 math) nonetheless higher than the scores students would have achieved by random guessing (randomly selecting one of four answers for each question). The very low attempt rates on the short and long answer questions are more worrying, but are consistent with the hypothesis that students were willing to exert some effort, but found the short and long answer questions much more taxing and difficult than the multiple choice questions. The research team's tentative conclusion is that low scores are not merely the result of low effort during the assessment periods.

The research team has no reason to believe that the assessments targeted inappropriately difficult subject content. Several teachers and government officials consulted confirmed that the assessment questions were appropriate to the curriculum and not unduly difficult. I, the research team's tentative interpretation, therefore, is that the exams were appropriately targeted and students exerted some effort, but had achieved only poor mastery of the material. This suggests there is great potential to improve learning, possibly by improving teaching.

VI. Baseline Findings on SSDP Training and Video Assignment Theories of Change

TT Intervention Theory of Change. The theory of change for the SSDP training (or TT) treatment is shown in Figure 3. It highlights that the intervention will improve student learning only if the following assumptions hold true. First, secondary school teachers of the relevant subjects must currently lack knowledge of subject content, knowledge of interactive pedagogical practices, knowledge of assessment methods, or motivation to experiment with new practices in their classrooms. Second, the NCED must succeed in developing high quality training curricula (or approving high quality training curricula developed by other actors) that address these needs, and must succeed in hiring and training high quality trainers, so that they can implement valuable training programs. Third, teachers must attend the trainings and participate receptively, so that they learn from the training sessions at the resource centers. Either out of intrinsic motivation (once they have been exposed to good training) or out of desire to get credit for career advancement (for which they must complete required classroom practice), teachers must follow through seriously on the practical application portion of the trainings. For career advancement concerns to motivate application of new skills learned during trainings, it must be the case that head teachers or training personnel will truly be willing and able to assess and report truthfully about whether the teachers have implemented the new methods in their classrooms. Fourth, once teachers have gotten over the initial hurdle to using the new methods in their classrooms, they must continue to teach more effectively for the longer term. For longer-term effectiveness they also must have adequate materials, equipment and facilities, and must not run into prohibitive opposition from parents or other parties. Fifth, students must respond to these new teaching practices either by learning more from a given quantity of their time and energy in learning activities, or by increasing the time and energy spent on those activities, so that they learn more. Students must be neither so poorly nourished nor so tired (from heavy labor requirements outside of schools) that they cannot learn, and must have adequate expectation that by working hard at learning they can improve their lives in some way.

Figure 3. Main TT Intervention Theory of Change



Assumptions:

- Teachers lack know-ledge of subject con-tent and/or pedagogy - NCED correctly identifies what know-ledge teachers lack - NCED develops or approves good training curricula that impart needed knowledge and skills - Teachers' attendance at trainings is observed and they lose points for poor attendance - Head teachers and training center officials accept only practice exercises that reflect real effort	- NCED selects trainers with adequate skills, trains them well and holds them accountable for good performance - Teachers are aware that certification depends on attendance and performance - Teachers are motivated to achieve certification because this will bring monetary or other benefits via process of performance evaluation (and career advance-ment), through social recognition, or through greater satisfaction with teaching experience	- Teachers are not prevented from using new methods by lack of resources/equipment - Teachers' capacity and effort are adequate for them to bring new skills and concepts from the trainings to their classrooms - Teachers may also bring back greater sense of team-work or professionalism - Teachers identify ways of applying concepts or skills learned during training that would address real needs in their classrooms - Teachers are not prevented from using new methods by parents, who believe old methods are better	-Teacher are motivated to make the effort to apply the relevant concepts or skills learned during training on an on-going basis, because this will bring monetary or non-monetary benefits through performance evaluation process (& career advancement), through social recognition, or through greater satisfaction with teaching experience - Application of new skills & concepts leads to changes in classroom practices or homework practices - Greater efficacy in the class-room, or greater sense of professionalism or teamwork may also motivate teachers for higher attendance or time on task	- Students are not absent so frequently that improved teaching practices are lost on them - Students are not too poorly nourished to concentrate and learn - Students do not have to work so much outside of school that they get inadequate rest or have inadequate time for homework - Students have adequate support out-side of school from parents or others	- The match between students' capacities and motivations, on the one hand, and the concepts and skills teachers have acquired through training, on the other hand, is sufficiently good that students learn more

VA intervention theory of change. The VA treatment aims to strengthen teachers' accountability for practicing the new teaching methods (introduced by the TT treatment) in their classrooms. The government sees the integration of serious classroom practice assignments as a critical innovation in this new wave of teacher training. The addition of these assignments may have little impact if teachers are not effectively held accountable for following through on them. It remains possible – even likely – that head teachers will sign off on teachers' reports without much critical assessment. Head teachers are paid little for performing their head teacher responsibilities, over and above their teaching responsibilities, and they may lack training and motivation for school leadership. Many head teachers may be eager to help teachers obtain credentials that can raise those teachers' pay (perhaps to the point of verifying classroom practice reports without conducting an adequate evaluation). It seems possible that teachers could receive credit for classroom practice without putting forth real effort to employ in their classrooms the new methods learned. The Video Assignment is intended to ensure that teachers receive credit for their training courses only if they actually implement the methods in their classrooms. The hope is that this will encourage them to practice the methods at least enough to make videos of themselves implementing the methods in a satisfactory manner. This may help them overcome social, psychological or logistical barriers to implementing the new methods, yielding sustained impact.

For the VA treatment to strengthen the impact of the TT treatment, the following assumptions must hold. First, the government must distribute video cameras to the ETCs, or schools must obtain the use of other video cameras. Second, teachers must care enough about obtaining credit for training courses that they arrange for a video to be taken of their use of new practices in the classroom. They must also understand what exactly they must demonstrate in the videos to receive credit. Third, teachers must have adequate means of delivering the videos to the ETCs and must believe that their use of new teaching methods in the videos will be assessed accurately by the ETCs. Fourth, it must also be the case that, in the absence of the VA intervention, teachers lack adequate motivation to implement new or improved teaching practices gained from training. This condition may fail (meaning that the VA intervention may add little to the TT intervention impact) in schools that are relatively well managed prior to the TT intervention. The VA treatment may, therefore, have greater impact in schools that are relatively poorly managed at baseline. The VA intervention may also add less to the TT intervention impact for teachers hired under temporary contracts or contracts with School Management Committees (who may be fired) than for permanent teachers paid by the national government (who effectively cannot be fired).

Baseline Lessons Regarding the Theories of Change. Baseline data shed light on some, but far from all, of the assumptions underlying these theories of change. This section lists the questions regarding the theories of change for which the baseline data contain relevant information.

- *Do 9th and 10th grade math and science teachers lack knowledge of subject content, interactive pedagogical practices or assessment methods of the sort they might learn during SSDP trainings?*

The research team's tentative conclusion is that yes, it is likely that many teachers lack sufficient knowledge or practice with likely SSDP content. The baseline data provide only somewhat weak evidence regarding teachers' need for training focused on math and science subject content. Many head teachers speak reasonably positively about teachers' subject knowledge, and many teachers have at least bachelor's degrees in math or science. Having high degrees in math or science, however, does not necessarily mean that a teacher is well-trained in the explanations and exercises that are helpful for teaching secondary students important math and science concepts. Baseline data show that that many teachers who have attended government in-service trainings in the past respond positively to the subject content they were exposed to. The data also show that many teachers have not attended previous trainings of this sort. So it seems likely that many teachers could benefit from appropriate SSDP subject content training.

The baseline data provide stronger evidence that teachers are not making frequent use of student-centered, more interactive learning activities in their classrooms. Whether this is because they do not know enough about the practices or because they know about them but do not want to work at implementing them is unknown.

- *Are NCED trainers and trainings of high enough quality that they might help 9th and 10th grade math and science teachers acquire knowledge or skills that they lack?*

The research team will learn more about the current round of trainings in up-coming qualitative research, but baseline teacher reports about the quality of past trainings are quite positive, suggesting institutional capacity to run adequate trainings.

- *Are teachers likely to be well motivated to implement in their classrooms new teaching practices that they learn during SSDP trainings?*

The baseline data suggest a variety of reasons to worry that many teachers may have little motivation to work at implementing new teaching methods in their classrooms. First, they are unlikely to feel pressure from parents to implement improved methods, because parents with low education are probably unable or reluctant to observe and provide feedback on such things. Second, if anyone holds them accountable for trying to improve their teaching success, it is probably their head teacher, but head teachers have very little time for monitoring and providing feedback on teaching practices. It is also unclear how motivated and well-equipped head teachers are to work at promoting improved teaching practices in their schools. In addition, teachers have high workloads, and many are used to spending only a small amount of time preparing for each class. The need to spend significantly longer amounts of time preparing each class (at least during the first year or two) to make the transition to using the new teaching methods may be too daunting.

The data also suggest, however, that schools may differ greatly in the extent to which head teachers and School Management Committees succeed in creating environments that motivate teachers to learn new teaching methods. In some schools head teachers visit teachers' classroom frequently and provide feedback, and some schools offer prizes for high performing teachers.

- *Might teachers run into opposition to the use of new teaching methods from parents or others?*

The baseline data suggest quite strongly that the answer to this is "no" for most schools. Head teachers and teachers were presented with lists of possible barriers to learning, and were asked to indicate which of these were major barriers to learning for their 9th and 10th grade math and science students. Head teachers and teachers nearly unanimously indicated that most listed problems were major problems in their schools. One of the few exceptions to this was that relatively few indicated serious resistance to the use of new teaching methods from parents. (Anecdotes had suggested this could be an important barrier, so the research team found this result somewhat surprising and striking.)

- *Might improved teaching practices be infeasible or fail to produce improved student learning because other critical inputs to learning – such as adequate facilities and equipment – are lacking in their schools?*

In general, the learning environment appears somewhat lacking in resources in many schools, but not to the point of making learning look impossible. In a few schools it will be important to address lack of classroom space before real progress can be made, but it is at least possible that teacher training could make a difference in most schools.

- *Might improved teaching practices fail to produce improved learning because students of low socio-economic status lack energy, nourishment or other pre-requites for learning?*

The baseline data suggest that secondary students and their families, while certainly not rich by global standards, are also not among the poorest of the poor. Many students face significant work and housework requirements, but not so much that they cannot do homework, and almost all eat some sort of meal before coming to school.

Two other obstacles to secondary math and science learning might matter for significant fractions of students. First, approximately one third of students are studying in a language that is not the language they speak at home. Second, the vast majority of head teachers and teachers report that students entering grades 9 or 10 with knowledge below grade level is a major problem. SSDP trainings focused on grade-level content may fail to prepare teachers for teaching students who lack adequate academic preparation.

- *Might teachers lack accountability for following through on classroom practice (post SSDP training), suggesting the potential for the Video Assignment to improve their motivation for practice?*

It does indeed seem possible that the Video Assignment could improve motivation for practicing new teaching activities, because head teacher efforts to hold teachers accountable for such practice seem weak at best.

Summary. The baseline data paint a picture of the 9th and 10th grade math and science environment in Nepal in which it is certainly possible that SSDP trainings could provide teachers with valuable training in math and science content and pedagogical practices, and that as teachers put into practice what they learn in such trainings, their students' learning would improve. Naturally, the baseline data do not guarantee that the training will have this effect. The baseline data thus support the continued pursuit of the larger evaluation study. The baseline data also provide support for interest in studying differences in training impacts across schools with stronger and weaker management, and across schools where teachers are and are not subject to the video assignment.

VII. Balance Tests and Review of Power Calculations

This section addresses several technical issues related to the larger impact evaluation for which the baseline data are informative. It provides reason to believe that the endline impact assessment will indeed deliver unbiased and adequately precise estimates of the interventions' impacts.

Balance Tests. One important use of the baseline data is to check whether the random allocation of schools to different study arms did in fact lead to groups of schools that are nearly identical in terms of the means and distributions of their observable characteristics. Randomized assignment may fail to create comparable treatment groups just by chance, so checking for "balance," or near equality of means across study arms, is important for establishing the validity of endline impact estimation results.

Recall that the 203 schools were randomly divided into 102 program (Phase I) schools and 101 control (Phase II) schools. In addition, the 102 program schools were randomly divided into 51 schools that would receive the SSDP training without the video assignment (VA) and 51 schools that would receive the SSDP training with the VA. This leads to three groups of schools that should be very similar in terms of the observed characteristics if the random assignment worked as hoped. Table 35 examines the math and science test scores, separately for Grades 8 and 9, to investigate whether they are significantly different from each other.

Table 35: Balance Checks for the Three Groups of Randomly Assigned Schools

Subject and Grade	**Training without VA**	**Training with VA**	**Control Schools**	**Test of Differences in Means (*p*-value)**
Math, grade 8 (maximum = 45)	8.27	7.72	8.20	0.494
	(0.58)	(0.33)	(0.29)	
Science, grade 8 (maximum = 50)	13.6	13.58	13.82	0.940
	(0.73)	(0.66)	(0.39)	
Math, grade 9 (maximum = 45)	12.57	12.08	13.09	0.490
	(0.67)	(0.61)	(0.59)	
Science, grade 9 (maximum = 50)	13.62	13.55	14.32	0.450
	(0.44)	(0.49)	(0.48)	

Note: Standard errors, in parentheses, are clustered at the school level.

For each of the four tests, the differences across the three types of schools are relatively small. In fact, none of the difference across the three types of schools is statistically significant. Thus it appears the randomization was successful in generating three groups of schools that are comparable to each other. Future documents will report on additional balance tests, related to basic school, student and teacher characteristics, Stallings classroom observation outcomes, and SEE performance.

Re-visiting Power Calculations. The balance test results just reported suggest that endline impact estimates will be unbiased. To assess whether they are also likely to be precise enough to deliver definitive conclusions, it is useful to revisit the power calculations that led the research team to choose a sample of approximately 200 schools. The original Minimum Detectable Effect (MDE) calculations (for MDEs expressed as differences in un-normalized percentage test scores) were based on the assumptions that "sigma" (the standard deviation of an un-normalized percentage test score) was 20 and that the ICC (intra-class correlation) was 0.65. Table 36 presents estimates of sigma and the ICC for the four assessments employed at baseline. In all cases, the estimates of both sigma and the ICC are smaller than the values assumed in the initial calculations. This is good news, implying that endline impact estimates are likely to be more precise, and that endline MDEs (in un-normalized percentage score terms) are likely to be smaller, than initially expected.

Table 36: Baseline Evidence on Power Calculation Assumptions

	Estimates of Sigma	**Estimates of ICC**
Grade 8 math assessment		
Multiple choice only	14	0.25
All sections	11	0.32
Grade 8 science assessment		
Multiple choice only	17	0.29
All sections	12	0.44
Grade 9 math assessment		
Multiple choice only	18	0.30
All sections	16	0.37
Grade 9 science assessment		
Multiple choice only	14	0.24
All sections	11	0.35

VIII. Next Steps

This section closes the report with a discussion of next steps.

Further Analysis of Baseline Data. The research team plans several additional analyses involving the baseline data in the near term. They are:

- A more disaggregated study of student outcomes (requested by policymakers), which examines how time use, consumption of meals before school, class size and test scores differ across districts, remoteness, ethnicity, whether Nepali is spoken at home, and gender.

- Construction of a school management quality index (possibly based on estimation of an Item Response Theory Model in which school management quality is the latent variable) and examination of how teaching practice and test score outcomes vary with management quality.
- Further analysis of the properties of the baseline academic assessments, employing both Item Response Theory and consultations with experts regarding appropriateness of examination content.
- Additional balance tests, covering a wider range of school, teacher and student characteristics, and teaching practices as well as assessment scores.

Much of the analysis will inform the development of a pre-analysis plan, describing all the methods that will be employed when estimating intervention impacts using endline data.

Upcoming Qualitative Research. The research team plans a round of qualitative research in August or September, which will be led by Dr. Sushan Acharya. The main aim is to learn about the strengths and weaknesses of the rollout of the study interventions, while also deepening understanding of teachers' perspectives on training and professional development and evaluating deeper dimensions of the interventions' theories of change. The research will include in-depth open-ended interviews with teachers, trainers, focal persons and Teachers' Federation representatives in three districts (Achham, Lamjung and Morang). It will also include shorter phone interviews with teachers and trainers in all 16 study districts.

Fixing the Timing of Endline Data Collection. The original proposal for the larger evaluation study called for rollout of the TT and VA interventions early in the 2017-2018 academic year and endline data collection near the end of the 2018-2019 school year. This would have given the TT and VA interventions at least one and a half academic years to work out their impacts in the treated schools. Allowing such time to elapse between rollout and endline data collection is important, because it may take time for teachers to become effective at implementing new material and new teaching practices in their classrooms, and because the effects of new teaching practices on student knowledge and skill are expected to grow over time. In practice, however, the rollout of the interventions was not completed until the end of the 2017-2018 school year. This means that if endline data collection proceeds as planned, the interventions will have been in place for less than a full academic year when impact is estimated.

Having only a short gap between roll-out and endline data collection reduces the chance of detecting positive impacts, increases the chance of detecting only a fraction of impacts that will ultimately be larger, and complicated interpretation of results. For example, if estimates show little or no effect, it will be unclear whether this is because the interventions were ineffective or because the impacts (which may grow over time) are being estimated too soon. On the other hand, if the study estimates large effects even after a short period, it will be a striking result.

The delay in roll-out raises a question of whether the endline data collection should be postponed for one year. The main benefit of doing so would be to allow the interventions more time to work out their impacts, thereby facilitating stronger interpretations of impact estimates. The risks and costs of doing so, however, are great. First, postponing the endline would require policymakers to prevent the treatment of Phase II schools for an additional year. In a period of decentralizing reforms, this may prove impossible. Thus postponing the endline increases the risk of losing the ability to estimate any impacts at all. Second, even if Phase II schools are suitably protected until a delayed endline, many students who were in grade 9 at baseline will have graduated out of the study. Rather than estimate impacts for students in two grades, therefore, the study would have to estimate impacts only for students in grade 8 at baseline. The tentative conclusion of the research collaboration is that the endline should proceed as scheduled in early 2019.

Fielding the endline study as originally planned in early 2019, and thus the decision to work with a shorter span between roll-out and endline, renders it especially important to consider two additional actions. First, it will be especially important to identify specific teaching practices and specific math and science concepts addressed during SSDP trainings, so that the research team can develop measurement instruments that are well focused on those practices and concepts. This will increase the study's potential to detect impacts. (More on this below.)

Second, the research team recommends that the larger research collaboration seek funding for a follow-up study one year after the end of this study's endline. A relatively low-cost study that focuses only on students who were in grade 8 at baseline, and focused only on a few key teaching and learning outcomes, might add significantly to understanding of how the interventions' impacts grow over time.

Next Steps for DOE[12] and Other Government Partners. Regardless of when endline data collection takes place, it is vital that policy implementers prevent the roll-out of SSDP training to Phase II schools for the duration of the study (i.e. at least until June of 2019). If Phase II schools receive SSDP training in the near term, this could raise teaching and learning outcomes in the schools serving as "controls," creating a bias toward under-estimation of intervention impacts.

It is also vital for policy implementers to exercise caution in rolling out any secondary school interventions other than the SSDP teacher training interventions. From the perspective of obtaining good impact estimates in the current study, it would be beneficial to prevent the rollout of any additional interventions affecting secondary school teaching and learning during the study period. The research team understands, however, that this may be impossible. If some roll-out of additional interventions does take place during the study period, they should be rolled out in a way that is unrelated to whether a school received the TT treatment, with or without the VA treatment, so that schools in different study arms are equally likely to receive treatment by the other interventions, and the percentages of schools affected by a new intervention are very similar across study arms. If, for example, policy implementers instead rolled out new interventions disproportionately in Phase II schools (perhaps because they seek to compensate for the lack of SSDP training treatment by providing these schools with a different valuable service), this could improve outcomes in the schools serving as controls, thereby causing the study to under-estimate TT and VA impacts.

Preparations for Endline Data Collection. The research team plans to revise all of its data collection instruments – questionnaires, Stallings classroom observations, and academic

[12] After reorganization, the DOE is now part of the Center for Education and Human Resource Development.

assessments -- before endline data collection. There are several reasons for revision. First, analysis of baseline data revealed flaws in the way a few questions were asked or in how answers to some questions were recorded. Second, analysis of baseline data revealed some puzzles, which can be resolved by adding new questions. Third, on-going decentralizing reforms suggest the importance of learning more about the role of local government bodies in managing schools. Fourth, some information collected at baseline does not need to be collected again at endline. Fifth, analysis of baseline data and conversations with policymakers suggest the usefulness of learning more about the use in classrooms of pedagogical practices that are promoted during SSDP trainings. (This will be possible only once the research team obtains a detailed list of the pedagogical practices that are promoted during SSDP trainings.) Sixth, further analysis of baseline data may reveal flaws in the academic assessments, and even if the assessments are not flawed, they must be revised to account for the progression of students across grades and might benefit from greater focus on math and science concepts specifically addressed during SSDP trainings.

The research team welcomes suggestions regarding how to revise the data collection plans and instruments. The team urges interested parties to share their suggestions by October 31, 2018.

References

Bruns, Barbara, and Javier Luque. 2015. *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean*. doi:10.1596/978-1-4648-0151-8. Washington, DC: World Bank.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muraldhiran, and F. Halsey Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*. 20(1):91-116.

Government of Nepal. Ministry of Education and Department of Education. 2016. *School Level Educational Statistics of Nepal. Consolidated Report 2015 (2072).* Sanothimi, Bhaktapur, Nepal.

World Bank. 2015. *Conducting classroom observations: analyzing classrooms dynamics and instructional time, using the Stallings 'classroom snapshot' observation system. User guide.* Washington, DC: World Bank.