Why Programs Fail: Lessons for Improving Public Service Quality from a Mixed-Methods Evaluation of an Unsuccessful Teacher Training Program in Nepal

Julie Schaffner, The Fletcher School, Tufts University

Paul Glewwe, Department of Applied Economics, University of Minnesota

Uttam Sharma, Independent Consultant

RISE Annual Conference

Blavatnik School of Government, Oxford University

September 21, 2023

Introduction

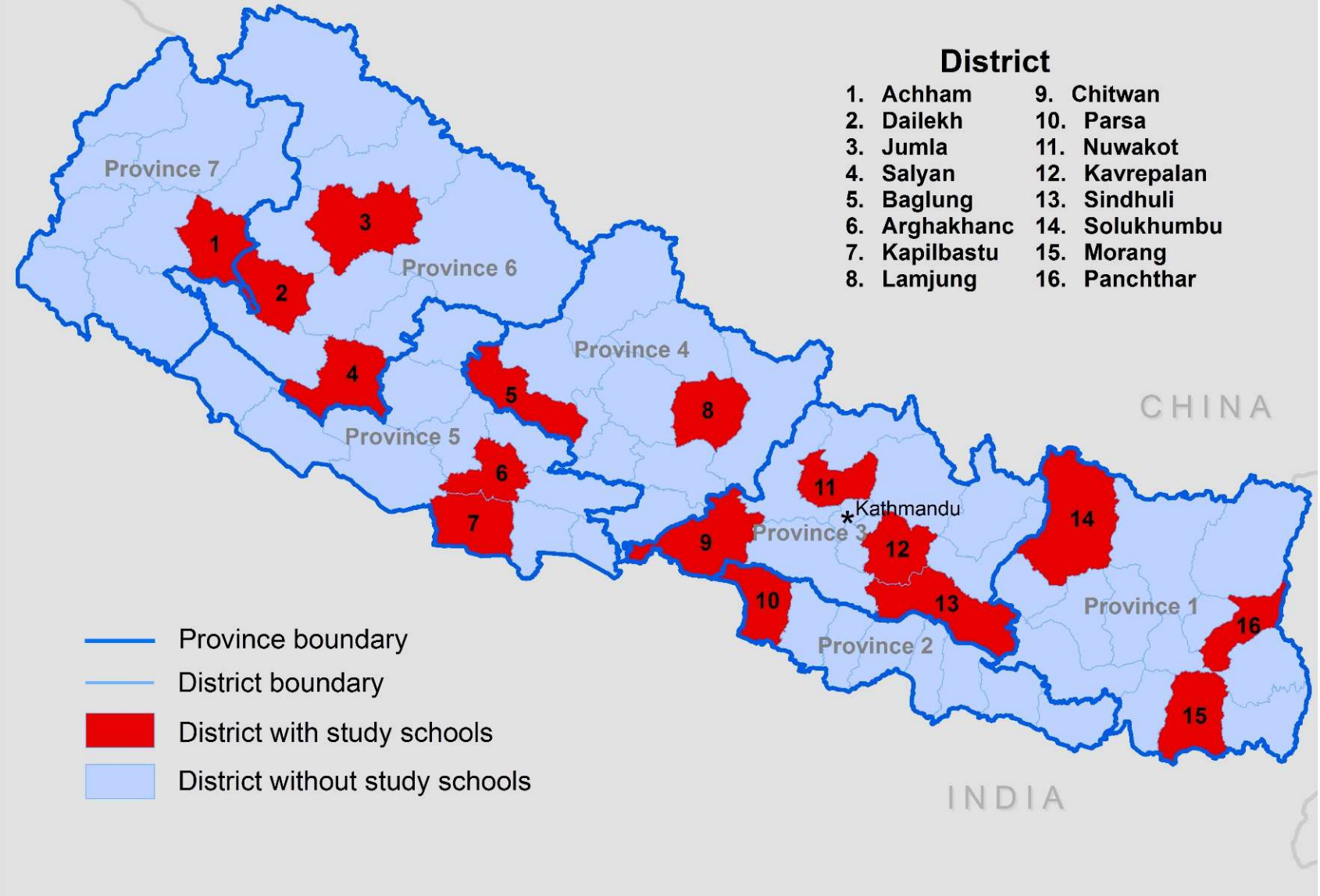
What We Do in this Paper

- We use a randomized control trial to evaluate the impact of a nationwide "in-service" teacher training program in Nepal
- The training lasted 2 weeks, and was for teachers of grade 9 and grade 10 math and science
- We collected a wide variety of data:
  - Student test scores in math and science
  - Teacher subject knowledge, teaching practices, attendance, and satisfaction with the teacher training program
  - Interviews of the teacher trainers
  - Interviews of School Management Committee members
- **The program had no effect on student learning**, so we investigate 5 reasons why the program failed

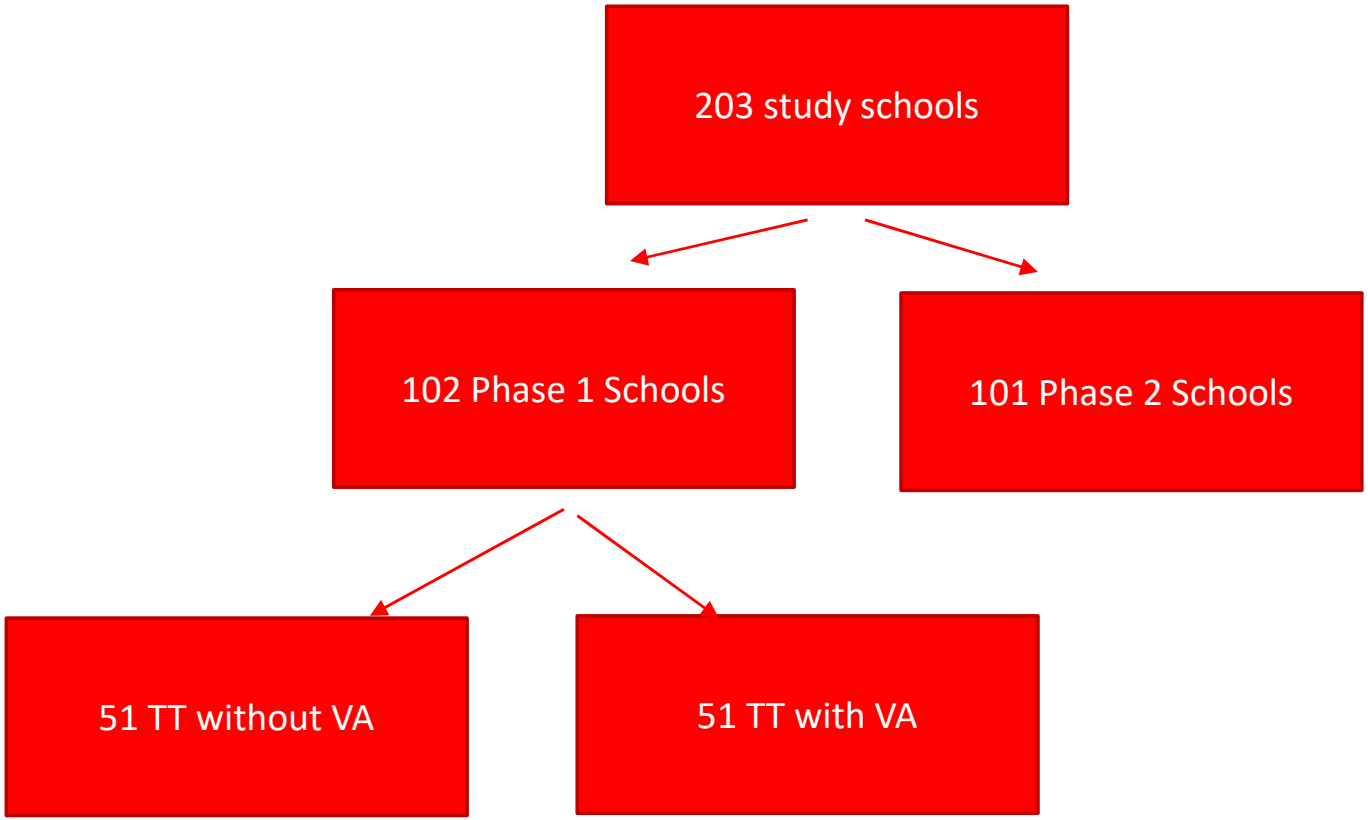
Sample Design

- We collected data from 203 government schools with grades 1-10 or 1-12, which were randomly assigned to three groups:
  - 51 that implemented the standard School Sector Development Program (SSDP) teacher training,
  - 51 that implemented the SSDP teacher training ***and*** a "video assessment" (see figure on slide after the map)
  - 101 control schools
- The schools are representative of 65 of Nepal's 77 districts; from the 65 districts we randomly selected 16 districts (see map on next slide)
- Two-thirds of schools were from a set thought to have teachers with relatively little recent training in grade 9 and 10 math and science (based on administrative records)

Districts where Schools in Study are Located



Allocation of study schools to study arms



Sample Design

Calculations for Determining the Sample Size

This sample is large enough to detect, with 90% probability (power), an impact on test scores of 0.20 standard deviations of test scores at 95% significance level

- Note: 0.20 std. dev. ≈ 2/3 of increase in learning from Grade 9 to Grade 10
- Note: 0.20 std. dev. ≈ 7% increase in proportion of test items correct

- Our goal was to detect an increase in test scores of about 7%. The expected higher correlation of test scores within schools (ICC ≈ 0.65) led us to choose a sample size of about 200 schools.

- Given that the effects are not large, and maybe are close to zero, 51 schools was not enough to detect the effect of adding the video assignment to the SSDP TPD program.

Description of the Program

- 10-day trainings at Education Training Centers (ETCs) for teachers of secondary math and science
- Ultimately focused on math/science concepts in current curriculum, and especially ways of teaching those concepts involving demonstrations and use of local materials
- Followed by 5 days of self study to be submitted within 52 days after training (10 lesson plans, action research, 1 of several optional activities)
- NCED (National Center for Educational Development) gave outlines of training curriculum to trainers at ETCs, who were asked to develop specific content and activities
- Teachers were “graded” based on attendance, participation at ETC, performance on test at end of training, self study. Adequate scores required for training credit.

Analysis Using Mixed (Quantitative & Qualitative) Methods

Benefits of mixing methods

- Can use qualitative research to improve the design of quantitative instruments (e.g. questionnaires)
- Can use qualitative data to learn about the rollout of training program, including experiences of teachers and trainers

Lessons from our experience:

- Qualitative and quantitative were indeed complementary.
- The telephone interview study was a valuable and low cost innovation:
  - highly informative, confirming generalizability of the rich small-N study
  - could be used for monitoring as well as evaluation in the future

Assessment (student test) Development

Requirements for our study:

- Focus on mathematics and science for grades 9 and 10
- Some questions emphasizing topics connected to SSDP training
- Some questions from below grade level

Implementation:

- Team combined international test experts with local test experts.
  - Sent international experts relevant local textbooks and SSDP curriculum, so they could target appropriately. They added well-pretested international assessment items.
  - Local experts guaranteed appropriateness to curriculum and test-taking norms
- Pre-tested and eliminated some questions with especially low correct response rates.

Measuring Teaching Practices

We measured teaching practices in three ways:

- Classroom observation by well-trained specialists using standardized protocols (Stallings)
- Self reported answers to simple questions about teaching practices (e.g. use of written lesson plans, use of group work)
- Reports by others (e.g. head teachers and students) on teachers' practices

Measuring Teachers' Subject Knowledge

- We wanted to measure teachers' subject knowledge, but were concerned that teachers would object to taking tests.
- We embedded a check on subject knowledge within forms asking teachers to evaluate 12 of the assessments items found in the student assessments. We asked them to evaluate the quality of the questions and also indicate "which answer they believed the item designer intended as the correct answer."
- Significantly, we made the evaluation forms anonymous.
- Teachers were surprisingly willing to complete these forms, and we believe they made a good faith effort to identify the correct answers.

Process Evaluation

- We used three approaches:
  - Frequent phone calls and data gathering by research assistants
  - Small-N qualitative study (3 districts only, 4 teachers and 1 trainer per district)
  - Larger-N telephone study (part quantitative, part qualitative)
- Repeated phone calls were necessary to obtain even rudimentary information from all ETCs involved.
- (As indicated earlier) Telephone interviews with teachers and trainers were especially enlightening, and were not very costly.

Estimated Impacts on Learning (ITT)

	Point Estimate (Std. Err.)	[95% Confidence Interval]
Normalized assessment score impacts for:		
Grade 9 Math	**-0.110 (0.066)**	**[-0.241, 0.020]**
Grade 10 Math	**-0.044 (0.072)**	**[-0.186, 0.099]**
Grade 9 Science	**-0.109 (0.060)**	**[-0.228, 0.010]**
Grade 9 Math	**0.006 (0.074)**	**[-0.140, 0.152]**

Estimates are not statistically significantly different from zero, and the upper bounds of 95% confident intervals rule out large positive effects.

Estimated Impacts on Learning

We obtain similar results (though with wider confidence intervals) if we:

- Focus only on assessment questions most closely tied to SSDP curriculum
- Focus only on teachers who did not report having SSRP training prior to the study
- Apply LATE/IV estimates to study the impact of having a trained teacher (rather than the ITT impact of being in a school where all teachers are invited to training).

Why Did the SSDP Program Fail?

The SSDP program clearly failed to increase the skills of grade 9 and 10 students in math and science.

But why did this happen?

We identify five weaknesses in the program that led it to be ineffective.

We cannot say which of these five weaknesses are the most important, but none of them is likely to be the sole cause.

Why Did the SSDP Program Fail?

Weakness #1: Low rates of teacher participation

Only 61% of the math teachers and 42% of the science teachers at endline had actually participated in the training.

- Although some of this reflects teacher turnover, it mostly reflects non-participation among those invited.
- The training took place when schools were in session, and many schools did not want to send their teachers: 78% had only 1 math teacher and 83% had only 1 science teacher.
- Not the entire explanation: IV/LATE estimates show that having a treated teacher did not increase test scores.

Why Did the SSDP Program Fail?

Weakness #2: Weak governance of ETC training sessions

The ETCs provided little guidance to teacher trainers, and gave them little time to prepare for the training.

- 19 of the 23 trainers we interviewed said that the ETCs gave them no training to conduct the SSDP trainings.
- Although most were given documents describing the training, many said that the documents were lacking in detail, or that they received them too late to use them.
- Some (about 20%) teachers said that their trainers had inadequate content knowledge and/or preparation.
- Unlikely to be the only reason for failure; 80% of teachers state that at least half of the content was “very valuable”.

Why Did the SSDP Program Fail?

Weakness #3: Mismatch of training content and teacher skill

While most teachers did well on the math and science items that we asked them to check, some did not.

- 28% of math teachers score 9 or lower out of 11 questions, and 39% of science teachers scored 8 or lower out of 11.
- Thus, “teaching at the right level” (TaRL) applies not only to students but also to some of the teachers.
- But this is unlikely to be the only reason why the program failed, since most teachers did scored well on the 11-question assessments.

Why Did the SSDP Program Fail?

Weakness #4: Weak motivation and support for post-training changes in teaching practices

After the training, most teachers did little to change their teaching practices, and no one else followed up with them.

- Teachers should have done 5 days of a “self-study project”, but few of them did this.
- Few head teachers (school principals) asked teachers how they were teaching differently based on the training.
- Teacher attendance was unchanged, and students & head teachers reported no changes in teaching practices.
- Unlikely to be the sole cause of failure; some evidence that teachers changed classroom practices, but not for the better.

Why Did the SSDP Program Fail?

Weakness #5: Mismatch of training content and student skills

Many (most?) 9th & 10th graders had weak skills, so they were unprepared for the advanced subject matter of the training.

- About 56% of 9th graders and 46% of 10th graders incorrectly answered a grade 8 math question, and 21% of 9th graders did the same for a grade 3 (!) math question.
- Student performance was worse on the science questions.
- Yet even strong students showed little improvement in performance, so this is not the only reason why the program failed.

Implications for Designing and Implementing Teacher Training Programs

Typical teacher training programs bring teachers to centrally located training centers, provide instruction but no lesson plans or teaching materials, and have no follow up.

- Recent discussions of "best practices" suggest that programs should provide "structured curriculum support" (e.g. lesson plans) and should include subsequent monitoring/coaching.
- Our results are consistent with this advice, but we also raise other problems:
  - Weak governance of training programs
  - Many or most students' skills are below grade level
  - Some teachers' skills are also inadequate
  - Trainings should not be held when schools are in session

Implications for Governance of Public Services

Reports and studies by the World Bank and others focus on service provider accountability to *increase their motivation*, not just for education but also for other public services.

- While we agree that service provider *motivation* is essential, **we would also argue that service providers need adequate:**
  - *Time* (e.g. to prepare lesson plans, for teachers & trainers)
  - *Capacity* (e.g. pedagogical skills and subject knowledge)
  - *Information* (e.g. trainers need adequate documents describing how they should train teachers)
  - *Budgets* (e.g to buy materials needed for "hands-on" student learning)

Thank you!

Please let us know what questions, comments and suggestions you have!

Julie Schaffner (Julie.Schaffner@tufts.edu)

Paul Glewwe (pglewwe@umn.edu)

Uttam Sharma (uttamsharma@gmail.com