J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

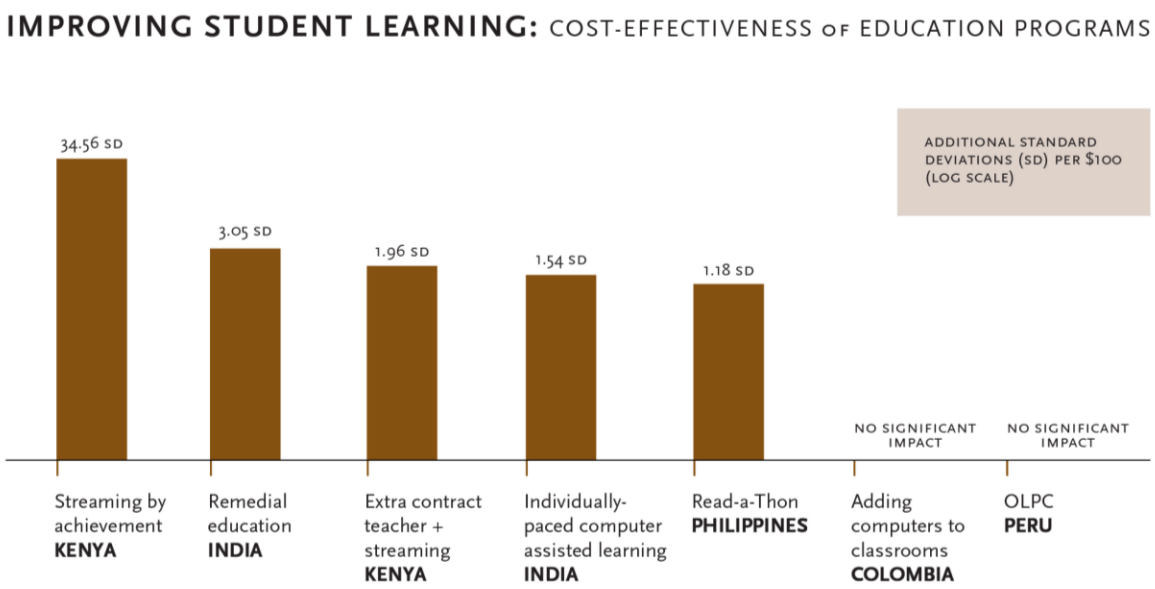
ESTIMATING THE COST-EFFECTIVENESS OF EDUCATION PROGRAMS

A Case Study


Prepared initially for the World Bank Strategic Impact Evaluation Fund (SIEF) by:

Conner Brannen, Meghan Mahoney, and Caitlin Tulloch
of the Abdul Latif Jameel Poverty Action Lab (J-PAL)

with thanks to Innovations for Poverty Action (IPA)



This case study aims to illustrate how to approach cost-effectiveness analysis (CEA) at distinct stages in a program lifecycle, demonstrating how prospective and retrospective CEA differ in terms of the data available and the conclusions that you can draw from the resulting ratios. In the following pages, after some introduction to CEA, you will work through the CEA calculations for the Assistant Teacher Initiative (ATI), a program in West Ghanea that trained assistant teachers (ATs) to provide in-school and after-school remedial classes for the weakest pupils to improve student learning.

I. WHAT IS COST-EFFECTIVENESS ANALYSIS (CEA)?

Calculating the cost-effectiveness of a program—for instance, dollars spent per one unit increase in student test scores—can offer insights into which programs are likely to provide the greatest value for money in given situations. Cost-effectiveness analysis (CEA) summarizes complex programs in terms of a simple ratio of costs to impacts and allows us to use this common measure to compare different programs evaluated in different countries in different years. It may not, by itself, provide sufficient information to inform all policy or investment decisions, but it can be a useful starting point for donors, governments, program implementers, and researchers when choosing between different programs that aim to achieve the same outcome.

METHODOLOGY

To calculate CEA, you must first gather the costs of the program and an estimate of the program's impact. In order to estimate **total program costs**, it is helpful to think about each of the ingredients necessary to implement the program (note that this list should not include any costs associated with evaluation). First, it is important to clearly define what constitutes the program and to understand the situation to which the program is being compared. In cases where a new intervention builds upon an existing program, you should only include the additional marginal costs of running the new components of program.

Next, you will need an estimate of **program effectiveness**—either from an evaluation of the exact program or an evaluation of a similar program in a similar context. Once you have these two components, a cost-effectiveness ratio can be calculated simply

$$CE\ ratio = \frac{Total\ Cost\ of\ Implementing\ Program}{Total\ Impact\ of\ Program\ on\ Specific\ Outcome}$$

by dividing the total cost of implementing the program by the total impact[1] of the program on the specific outcome of interest (for example, student test scores):

This number, when compared to the cost-effectiveness estimates of other programs, can give implementers a sense of whether the program may (or did) offer good value for its cost.

How this ratio is interpreted and the conclusions that can be drawn from this analysis depend on when the CEA is conducted, in relation to program implementation:

- Prior to the start of a pilot or at-scale program, i.e. "prospective" analysis
- After an evaluation of the program is completed, i.e. "retrospective" analysis

The stage at which CEA is conducted has implications not only in terms of data availability and types of questions asked, but also in terms of conclusions. The following sections aim to discuss the two approaches, and provide guidance for how to interpret the resulting analysis.

II. PROSPECTIVE ANALYSIS

Prior to the start of a program—whether a pilot or at-scale—a donor, government, or other program implementer may wish to know how cost-effective the proposed program could be. To calculate this estimate, you must first gather data about the anticipated costs of the proposed program. Then, if an estimate of the program's effectiveness exists—either from an evaluation of a pilot version of the exact program or an evaluation of a similar program in a similar context—you have the necessary components to estimate program cost-effectiveness. Dividing total program costs by total program impact generates the cost-effectiveness ratio, which allows you to assess *whether it is worth implementing or scaling up*.

These prospective cost-effectiveness ratios are unlikely to be very precise, since they use rough projections of program costs and impacts. As such, program implementers should be conservative about the conclusions drawn from the results. Prospective cost-effectiveness estimates should be treated as general guidelines to identify programs that could be cost-effective or rule out options that are clearly not cost-effective, but they are **not** precise estimates of how cost-effective a program will be, especially when scaling it up.

1 In most impact evaluations, effect size is expressed as the average program impact per participant. To get the total impact, you simply multiply the average program impact by the total number of participants:

$$Total\ impact = Average\ program\ impact\ per\ participant \times Number\ of\ participants$$

III. RETROSPECTIVE ANALYSIS

After a program has been evaluated, a donor, government, or other program implementer may want a more precise estimate of how cost-effective the program was, particularly to compare to other previously evaluated programs. While the methodology involved in conducting retrospective analysis is the same as in prospective analysis, the interpretation of the resulting ratio is slightly different. At this point, more accurate data are available both about the actual costs incurred (and the types of inputs necessary for the program), as well as about impacts. The resulting cost-effectiveness ratio can be interpreted with a much greater degree of certainty, reflecting the exact cost-effectiveness of this program *as it was implemented in this context*.

! The retrospective estimates will still **not** represent the general cost-effectiveness of that type of program or of the same program in another context

II. THE ASSISTANT TEACHER INITIATIVE (ATI) IN WEST GHANEA[2]

While school enrollment rates in developing countries have significantly improved in the last decade, learning is still lagging. Large classes, with pupils of very different levels of preparation, make it difficult for teachers to target instruction appropriately. In West Ghanea, the 2009 National Education Assessment showed that only 20 percent of grade 3 pupils reach expected proficiency levels in English, and only 25 percent in math, despite the fact that the government spends 23 percent of its budget on education.

Previous research has shown that significant improvements can be achieved at relatively low cost by targeting the level of instruction to pupils' abilities. Based on these insights, the West Ghanea Education Service (WGES) and Innovations for Poverty Action (IPA), in partnership with the West Ghanea National Association of Teachers (WGNAT) and the National Youth Employment Program (NYEP)3, piloted and evaluated the Assistant Teacher Initiative (ATI). Under ATI, the following three interventions were piloted for one year in an effort to focus instruction at students' learning levels:

In-School Remedial ATs (ATI-ISR): Assistant teachers (ATs) pulled struggling pupils out of the normal classroom to work on basic literacy and numeracy skills for two hours per day during regular school hours.

After-School Remedial ATs (ATI-ASR): ATs worked with struggling pupils on basic literacy and numeracy skills for two hours after the regular school day.

Normal Curriculum ATs (ATI-NC): This intervention tested the effect of smaller class sizes by pulling pupils out of the classroom at random (a mix of both struggling and high-performing pupils) to work with ATs to review the standard literacy and numeracy curriculum for two hours per day. The group of pupils pulled out to work with the AT was alternated with the teacher's group every other day.

Each ATI intervention was tested in 100 schools, with an average of about 120 pupils from Standards 1-3 per school. All three ATI interventions included the following components:

Training ATs: Both the in-school and after school remedial ATs received one week of instruction in a pedagogy focusing on basic literacy and numeracy skills. The training included instruction in rapid testing methods to easily

2 While based on a real program, this case study uses the example of a fictional program in a fictional country.

3 The National Youth Employment Program (NYEP) is a national program that provides jobs for unemployed college graduates.

identify which children were in need of remedial activities, as well as classroom management, activity-based learning, and how to effectively use the teaching materials provided.

Orientation & Sensitization: District education officials, head teachers, and school management committees were given an orientation to the program, explaining the purpose of the activities and how they could provide support.

III. BEFORE IMPLEMENTING ATI

Imagine that you are a program manager at the West Ghanea Education Service (WGES) and you want to know whether ATI will be a good investment. You specifically want to know whether the pilot will be cost-effective. At this point, the program has not yet been implemented so real-time data on the program's actual costs and impacts does not exist—you will have to estimate cost-effectiveness based on rough projections.

i. Costs

To begin, you must estimate the total cost of the program. In order to calculate the full cost, it is helpful to think about each of the ingredients necessary to implement the program. As discussed in Section I, you should begin by clearly defining the program that is being evaluated, whether it is setting up new systems or building upon existing infrastructure.

Question 1: *Think carefully about all of the ingredients necessary to implement ATI—from materials, to personnel, transportation, and capital investments. Record the list in the space below.*

Your supervisor has asked you to calculate the cost-effectiveness for only the *After-School Remedial ATs* intervention (ATI-ASR). Working with Innovations for Poverty Action (IPA), who provides technical advice on program implementation, you come up with a prospective budget based on IPA's experience implementing similar programs in West Ghanea, India, and Kenya.

Table 1: Estimated Prospective Cost per Child for ATI-ASR	
Total estimated monetary cost per year to implement ATI-ASR	$74,800
Total number of schools/children in ATI-ASR treatment group	12,000 children across 100 schools
Projected cost per child per year to implement ATI-ASR	$6

The ATI-ASR intervention has not yet been evaluated and there are no available estimates of similar programs in the existing literature. However, using the estimated costs for the program, WGES can still estimate the size of the impact ATI-ASR must have on student learning as measured in standard deviations[4] to be a worthwhile investment.

Question 2: *If you define a program as cost-effective if it achieves a total test score gain of at least 1.4 SD per $100 spent—how large does the average per student impact of ATI-ASR need to be for it to be considered a cost-effective investment?*

ii. Impacts

IPA informs you that there have been several evaluations of similar programs from which you can draw impact estimates and has put together a brief report that summarizes the main lessons:

- It is relatively easy to quickly teach children how to read using a simple methodology and low-cost materials such a phonetic charts and story cards (India[5]).
- The key is to focus instruction at the right level e.g., teaching struggling students separately from the rest of the class for a portion of or the entire class time (India and Kenya[6]).

4 An impact expressed in standard deviations shows how much a program shifts the average test score in the treatment group along the distribution of test scores in the comparison group. An intuitive way to think about a standard deviation is the expected difference between the score of a randomly chosen individual in a student population and the average score in that population. For example, 0.2 standard deviations—which is commonly accepted in the education literature as a meaningful impact on student learning—roughly represents the difference between the 50th and the 58th percentile of test scores.

5 In an effort to gear instruction towards students' actual learning levels, rather than the expectations of a rigid curriculum, Pratham—a large Indian NGO—piloted a targeted tutoring program, which brought in volunteers to work with the lowest-performing students in rural schools. At the end of two years, students in program schools saw significant test score improvements, with the largest gains for children in the bottom third of the initial distribution (Banerjee, Cole, Duflo, and Linden 2007).

6 An evaluation in Kenya found that placing students in different classes by learning level ("streaming") improved test scores for students at all levels of the distribution, including higher-achieving students (Duflo, Dupas, and Kremer 2011).

Evaluated Program	Duration	Avg. Impact on an Individual's Test Score
Remedial tutoring by volunteers, India	2 years	0.28 SD
Tracking students by ability, Kenya	18 months	0.18 SD

These programs are slightly different from the ATI-ASR intervention and neither was piloted in West Ghanea. Nonetheless, they can provide a rough guide to how effective this intervention could be.

Question 3: *Why do you think the impact estimate from* ***India*** *may or may not be a good approximation of the impact of the ATI-ASR intervention? Why do you think the impact estimate from* ***Kenya*** *may or may not be a good approximation of the impact of the ATI-ASR intervention?*

Now that you have your prospective costs and impact estimate, you are ready to calculate an estimate of the cost-effectiveness of the ATI-ASR intervention.

Question 4: *Based on the impact estimates from the comparable programs presented above, calculate a range of potential cost-effectiveness.*

Table II: Estimated Cost-Effectiveness of ATI-ASR		
Estimated cost per child per year of ATI-ASR	$6	
	Remedial tutoring by volunteers, India	**Tracking students by ability, Kenya**
Estimated impact per child per year of ATI-ASR	0.28	0.18
Estimated cost-effectiveness of ATI-ASR		

Question 5: How would you explain what this estimated range of cost-effectiveness means to your colleagues at the West Ghanea Education Service (WGES)? Does this predict exactly how cost-effective ATI-ASR will be?

Remember, due to the nature of the available data at this stage of the program, your prospective estimate of the cost-effectiveness of ATI-ASR will necessarily be imprecise. Your estimated budget is a best guess of how much the program will likely cost, but the costs may change in practice. You also only had impact estimates from evaluations of similar programs in other contexts, not ATI-ASR itself in West Ghanea. ATI-ASR could have a larger or smaller impact than these programs, as discussed in Question 3. With this level of uncertainty in both components of the cost-effectiveness ratio, the estimate produced by your analysis will not be a promise for how cost-effective your program will be. When discussing the range, be clear about this. You should also be conservative about the conclusions you draw. Based on your prospective estimate, you can conclude that ATI-ASR will likely cost around $XX per standard deviation increase it achieves, if your assumptions about the budget and the impact hold.

IV. REFINING THE ORIGINAL COST-EFFECTIVENESS ESTIMATE FOR ATI

Based on the projections of potential cost-effectiveness, WGES decided to implement the full ATI program. It is now one year later and evaluation data is available. Your supervisor asked you to update the cost-effectiveness estimate using actual costs and impact estimates to get a more precise sense of how cost-effective ATI was, and what factors drove the program's cost-effectiveness.

i. Impacts

Question 6: *Are there any potential risks or uncertainties that that could make ATI-ASR not cost-effective? In your opinion, what are the most important assumptions that must hold true to make ATI-ASR cost-effective?*

IPA conducted a randomized evaluation of ATI and found the following:

Having ATs provide remedial instruction targeted to struggling students, both during and after school, had modest but significant impacts on basic literacy and numeracy skills. Students in schools in the In-school Remedial ATs treatment group (ATI-ISR) saw significant improvements in their math test scores, on average, as well as improvements on certain parts of the local language tests. Students in schools that received after-school remedial ATs (ATI-ASR) saw even larger effects, including improved performance on English tests.

Simply reducing the class size by adding an AT had minimal effects on test scores. Students in schools in the Normal Curriculum AT (ATI-NC) group saw a slight increase in their math scores, though it was smaller than the increase in the ATI-ISR and ATI-ASR schools. This intervention had no effect on average English or local language test scores. This suggests that the significant improvements caused by ATI-ISR and ATI-ASR reflected the combined effects of the smaller class sizes and active basic-skills instruction practices targeted to struggling pupils.

Remedial classes taught by ATs after school were more effective than those taught during school hours. While in-school programs (ATI-ISR) might be expected to perform better due to better attendance and supervision, the after-school program (ATI-ASR) actually led to larger increases in test scores than ATI-ISR. One partial explanation is that after-school remedial ATs spent significantly more time actually teaching than their in-school counterparts, likely because there were less classroom disruptions.

You are specifically interested in the program's impact on basic literacy skills, which are in Table IV.

Table IV: Impact of ATI on Basic Literacy Skills

Intervention	Duration	Impact on Literacy Test Scores[7]
In-School Remedial ATs (ATI-ISR)	1 year	0.122**
After-School Remedial ATs (ATI-ASR)		0.190***
Normal Curriculum ATs (ATI-NC)		0.092

*** p<0.01, ** p<0.05

ii. Costs

Again, you are interested in estimating the cost-effectiveness of the ATI-ASR intervention. Now that program implementation is complete, there is more data about the actual costs of intervention, but it is still not consolidated in one place. For example, the WGNAT, who trains the ATs, has information on how much the trainings cost. The NYEP has information on the AT salaries, and IPA has the costs for program administration. Because data about program costs can be spread across several organizations (and in some cases includes items that would never appear in a budget, such as volunteer time or the cost of free or donated materials), budgets are insufficient to do a full retrospective CEA. Instead, you can use the "ingredients method" of costing described earlier to list all potential cost categories.

One ingredient you may have specified is instruction material for the ATs to use in teaching their remedial sessions. Below is an example of the form that cost data might come in—a materials delivery receipt for a school in the ATI-ASR treatment group[8]:

7 These numbers are rough approximations of the program's potential impact based on midline results of the ATI evaluation. Delays of several months in the program implementation meant that children were exposed to only around 10 weeks of the interventions before the midline testing began.

8 Costs in Table V are displayed in Western Ghanean Shillings (WGHS)

Table V: Instruction Materials			
Name of School:	Achinad/a Primary		
Number of Students:	120		

Item	Total # Sets School	Unit Price (WGHS)	Total (WGHS)
Language Materials			
Sentence Cards	3	1.73	5.19
Paragraph Cards	3	2.07	6.21
Word Cards	3	1.84	5.52
Letter Cards	2	1.98	3.96
Letter Chart	2	7.80	15.60
Picture Story Cards	2	3.34	6.68
Phonic Charts/Slides (two-letter words)	3	4.03	12.09
Phonic Charts/Slides (three-letter words)	3	4.03	12.09
Phonic Charts/Slides (four-letter words)	3	2.01	6.03
Numeracy Materials			
Sums tables	5	3.51	17.55
Currency notes	20	0.48	9.60
Number chart (1-100)	2	2.00	4.00
Addition chart	3	2.00	6.00
Subtraction chart	3	2.00	6.00
Magic 12	3	4.75	14.25
Magic 15	3	4.75	14.25
Magic 30	3	4.75	14.25
Multiplication chart	3	2.00	6.00
Number cards & shape cards	3	7.00	21.00
Total			**186.27**

Your list of ingredients can then be expanded into a full costing worksheet to allow you to calculate the total cost of ATI-ASR over one year. Under each cost category, it is important to specify for each line item the unit cost and the number of units required to better understand the structure of the program and how its costs were distributed across beneficiaries. Such detailed cost data can allow program implementers—who are interested in scaling the program or replicating it in another context—to vary the specific components of the program, their unit costs, or the number of units needed to reflect their implementation environment.

In the table below, each of the cost categories, except for Instruction Materials, has been collapsed such that only the total is shown. A portion of the Instruction Materials category has been expanded as an example of how to specify the unit cost and number of units needed for each line item.

Question 7: *Use Table V to fill in the blue boxes in Table VI below.*

Table VI: Costs for ATI-ASR

Demographics	Pilot
Number of ATs employed	162
Number of schools	100
Number of students in Standards 1-3	12,000
Duration (in years)	1
Exchange Rate (USD/WGHS)	0.51

Cost Source		Cost Category							Currency Units	Total Cost
IPA Budgets	1	**Start-up & Overhead**							2011 USD	5,152.00
WGES Program Documents & Budgets	2	**Refresher Training for Master Trainers**							2011 USD	591.54
	3	**Training of Circuit Supervisors, WGNAT, NYEP and Other WGES**							2011 USD	2,158.94
	4	**Training of SMCs**							2011 USD	4,962.80
	5	**Master Trainer Training**							2011 USD	3,561.84
	6	**Training of Trainers**							2011 USD	4,191.40
WGNAT Documents & Budgets	7	**Training of Assistants/Teachers**							2011 USD	9,745.24
	8	**Back up Training and Training for replacements**							2011 USD	9,745.22
IPA Budgets	9	**Monitoring Training**							2011 USD	5,819.67
Packing slips from IPA	10	**Instruction Materials**								
			Item	**Unit Cost, Local**	**Local Currency**	**# Sets Per School**	**Total # Sets Needed**	**Total Cost, 2011 WGHS**		**Total Cost, 2011 USD**
		10.1	Sentence Cards		2011 WGHS				2011 USD	
		10.2	Paragraph Cards		2011 WGHS				2011 USD	
		10.3	Word Cards		2011 WGHS				2011 USD	
		10.4	Letter Cards		2011 WGHS				2011 USD	
			…							
			Total Cost of Instructional Materials						2011 USD	20,369.49
WGES Budgets	11	**Supplies & Equipment**							2011 USD	2,830.50
IPA Budgets	12	**Monitoring**							2011 USD	4,833.47
NYEP Budgets	13	**ATs Salaries**							2011 USD	65,520.00
TOTAL COSTS, ATI-ASR*										139,482.13

Question 8: *Now that you have the actual costs of the program, please calculate a more accurate cost per child.*

Table VII: Actual Costs of the ATI-ASR	
What was the total cost of ATI-ASR?	
How many students did ATI-ASR reach?	
What was the cost per child per year?	

Question 9: *With both the cost and impact information, please calculate a retrospective estimate of the cost-effectiveness of the intervention and fill Table VIII.*

Table VIII: Actual Cost-Effectiveness of ATI-ASR	
Actual cost per child per year of ATI-ASR	
Actual impact per child per year of ATI-ASR	
Actual cost-effectiveness of ATI-ASR	

Because you have high quality data from both the impact evaluation and the actual implementation costs of ATI-ASR, your cost-effectiveness estimate is more precise and tells you more about the program. Based on the calculations, you can conclude that the ATI-ASR, as it was implemented, cost $XX per standard deviation change in literacy test scores. However, you should still be careful when discussing this estimate, as the impacts and costs reflect one particular context. This estimate will **not** tell you exactly how cost-effective all remedial education programs that use paid community workers to deliver after-school instruction will be, nor how cost-effective the program will be if scaled up across West Ghanea.

Question 10: *How would you explain this cost-effectiveness estimate to your colleagues at the West Ghanea Education Service (WGES)? How is your explanation and interpretation of this estimate different from the prospective calculation?*

V. CONSIDERING A SCALE-UP OF THE ATI MODEL

Based on the results of the evaluation and your CEA, WGES is now considering a nationwide scale-up of ATI-ASR. However, they are concerned that the cost-effectiveness of the program might change when scaled up. Since your cost and impact estimates are based on a pilot program, there are several considerations that must be taken into account when estimating the cost-effectiveness of ATI-ASR at scale.

Question 11: *How do you expect the impact and costs to change when ATI-ASR is scaled up nationwide? What are some of the factors that you might want to consider when thinking about the impact of the program at scale?*

Based on considerations about how the costs and impact may differ at scale, you can estimate the at-scale cost-effectiveness of ATI-ASR by following the methodology described above. Although you have better cost and impact data than you did before the evaluation, your calculation is still only a projection because you cannot be certain how the program might change when implemented nationwide.

CEA is a useful tool for policymakers and practitioners to decipher which programs will provide the greatest return for their investment. CEA summarizes complex programs in terms of a simple ratio of costs to impacts. How the ratio is interpreted and which conclusions you can draw from the analysis depend on the precision of the impact estimates you use as well as the accuracy of the cost data you have. This exercise aimed to provide you with information on how to approach CEA at different points in the program lifecycle, and on what the implications are for interpreting CEA at each stage.