The evaluation that declines the question

The commission asked whether school feeding raises learning. The honest answer is that this design could not have detected an effect of the size the programme cares about, that what it did detect is nothing, and that both statements have to appear in the summary rather than the annex.

Pierre Robentz Cassion

Lesson 8 of 8

What this lesson covers

- Assemble what the course produced
- The summary that is honest
- Four sentences an evaluation must be able to write
- What "no effect detected" is worth
- Writing for the reader who will quote one sentence
- What the course has been about
- Report it whole
- What comes next

Assemble what the course produced

Question	Answer	Where it came from
Did scores rise?	+6.97 pts, 95% CI +6.1 to +7.9	Before-after, lesson 1
Is that the programme?	No — comparison schools rose 7.62	Lesson 1
Are the groups comparable?	No — six of six characteristics imbalanced	Balance table, lesson 2
What is the effect?	−0.96 pts, 95% CI −3.5 to +1.6	Difference-in-differences, lesson 3
Does matching change it?	−0.71 pts, on 9 of 15 treated schools	Lesson 4
Could a better design run?	Not on this data — no outcome past the cut-off	Lesson 5
What could it have detected?	Nothing below 4.9 points	Power, lesson 6

Assemble what the course produced

- Two of those seven rows are the report and the other five are the annex — The finding is the fourth row and the. . .

The summary that is honest — Example (cont.)

```
School feeding and learning outcomes: evaluation summary

  FINDING

  This evaluation cannot answer whether the school feeding programme
  improves literacy, and the reason is in its design rather than its
  results.

  The programme is delivered in 15 schools and compared against 9. With
  fifty children per school and the observed clustering, the smallest
  effect this comparison could have detected is 4.9 percentage points. The
  programme team's threshold for expansion is 3 points. The evaluation was
  therefore incapable of answering the question it was commissioned to
  answer, and this was knowable before it began.

  What it does establish: literacy rose 7.0 points over the school year in
```

The summary that is honest — Example (cont.)

```
both programme and comparison schools, with no difference between them
(-0.96 points, 95% CI -3.5 to +1.6). Effects larger than 3.5 points in
either direction are ruled out. Effects between 0 and 3 points -- the
range the programme cares about -- are not.

RECOMMENDATION

Do not discontinue the programme on the basis of this evaluation. Do not
expand it on the basis of the 7-point gain, which is the school year and
is present in schools without the programme.

To answer the question, an evaluation needs about 60 schools rather than
24. If the programme is expanding to new schools, randomising the order of
the roll-out costs nothing and produces a comparison group that does not
have to be argued for.
```

The summary that is honest

- The first paragraph refuses the question and the last paragraph says what would answer it — Between them there is a...

Four sentences an evaluation must be able to write

- "The counterfactual is X." — Here: schools without the programme, assumed to have been on the same trajectory
- "The estimate is Y, with interval Z, in units the programme uses." — Here: -0.96 percentage points of literacy, -3.5 to...
- "The smallest effect this design could detect is W." — Here: 4.9 points
- "This analysis cannot rule out V." — Here: any effect between zero and about three points, which is the whole range the...

What "no effect detected" is worth

Reading	Supported?
"The programme does not work"	**No** — the design could not detect the relevant effect
"The programme has no large effect on literacy"	**Yes** — above 3.5 points is ruled out
"The 7-point gain is not attributable to the programme"	**Yes** — comparison schools gained more
"Funding should be redirected"	**No** — that requires an effect estimate this study does not have
"The next evaluation needs 60 schools"	**Yes** — computed, and actionable

What "no effect detected" is worth

- The two "yes" rows in the middle are real findings and they are unwelcome ones — they take away a headline number a...

Writing for the reader who will quote one sentence

- Not — "No statistically significant effect of school feeding on literacy was found." True, and it will be read as "the...
- Not — "Literacy rose 7 percentage points in programme schools." True, and it will be read as the effect
- This — "Literacy rose 7 points in both programme and comparison schools; this evaluation was not large enough to detect...
- Put that sentence in the summary, the conclusion and the covering email — The quotable line is chosen by you or it is...

What the course has been about

- Which means the most consequential decisions were made before any data existed — who got the programme and in what...
- So the useful thing this course asks for is not a technique — It is being in the room in January, with one page, asking...

Report it whole — Example (cont.)

Evaluation of the school feeding programme: methods and limitations

Design	Difference-in-differences, 15 programme and 9 comparison schools, baseline and endline literacy on 585 students with both rounds.
Counterfactual	Comparison schools, assumed to have been on the same trajectory. Two rounds exist, so parallel trends is argued (numeracy shows the same null; no single school drives the result) and not tested.
Estimate	-0.96 percentage points, 95% CI -3.5 to +1.6, clustered on 24 schools.
Power	Minimum detectable effect 4.9 points at 80% power, against a 3-point threshold of interest. Underpowered

Report it whole — Example (cont.)

	for the relevant effect size.
Balance	All six baseline characteristics imbalanced beyond 0.25 standardised. Adjusted for baseline and district; unmeasured selection is not addressed.
Attrition	585 of 741 baseline students have an endline. Those who left scored 39.3% at baseline against 56.8% for those who stayed.
Not claimed	Any effect on attendance, enrolment or nutrition. Any causal effect beyond the difference-in-differences assumption. Any conclusion about effects below 3 points.
Plan	Written 2024-01-15, one deviation listed in Annex A, three exploratory analyses in Annex B.

Report it whole

- Eight rows, and a reviewer can reconstruct every decision from them — That is the deliverable this course has been...

What comes next

- Module 5 is complete: uncertainty, then models, then designs.

Read the full lesson, with runnable code

https://data-analysis.cassion.dev/courses/impact-evaluation-methods/ie-08-the-evaluation-that-says-no