

L'évaluation qui décline la question

La commande demandait si la cantine scolaire améliore les apprentissages. La réponse honnête est que ce protocole n'aurait pas pu détecter un effet de l'ampleur qui intéresse le programme, que ce qu'il a détecté est nul, et que les deux affirmations doivent figurer dans le résumé et non en annexe.

Pierre Robentz Cassion

Leçon 8 sur 8

Méthodes d'évaluation d'impact — Ce que l'évaluation affirme

Ce que couvre cette leçon

- Rassemblez ce que le cours a produit
- Le résumé honnête
- Quatre phrases qu'une évaluation doit pouvoir écrire
- Ce que vaut « aucun effet détecté »
- Écrire pour le lecteur qui citera une seule phrase
- Ce dont ce cours a traité
- Rapportez-le en entier
- La suite

Rassemblez ce que le cours a produit

Question	Réponse	D'où elle vient
Les scores ont-ils monté ?	+6,97 pts, IC à 95 % +6,1 à +7,9	Avant-après, leçon 1
Est-ce le programme ?	Non — les écoles témoins ont monté de 7,62	Leçon 1
Les groupes sont-ils comparables ?	Non — six caractéristiques sur six déséquilibrées	Tableau d'équilibre, leçon 2
Quel est l'effet ?	−0,96 pt, IC à 95 % −3,5 à +1,6	Différences de différences, leçon 3
L'appariement change-t-il cela ?	−0,71 pt, sur 9 des 15 écoles traitées	Leçon 4
Un meilleur protocole pourrait-il tourner ?	Pas sur ces données — aucun résultat au-delà du seuil	Leçon 5
Qu'aurait-il pu détecter ?	Rien en dessous de 4,9 points	Puissance, leçon 6

Rassemblez ce que le cours a produit

- Deux de ces sept lignes sont le rapport et les cinq autres sont l'annexe — Le constat est la quatrième ligne et la...

Le résumé honnête — Exemple (suite)

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

Le résumé honnête — Exemple (suite)

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.

- Le premier paragraphe refuse la question et le dernier dit ce qui y répondrait — Entre les deux il y a un constat, un. . .

Quatre phrases qu'une évaluation doit pouvoir écrire

- « Le contrefactuel est X. » — Ici : les écoles sans programme, supposées avoir suivi la même trajectoire
- « L'estimation vaut Y, avec l'intervalle Z, dans les unités qu'emploie le programme. » — Ici : $-0,96$ point de...
- « Le plus petit effet que ce protocole pouvait détecter vaut W. » — Ici : 4,9 points
- « Cette analyse ne peut pas écarter V. » — Ici : tout effet entre zéro et environ trois points, c'est-à-dire toute la...

Ce que vaut « aucun effet détecté »

Lecture	Étayée ?
« Le programme ne marche pas »	Non — le protocole ne pouvait pas détecter l'effet pertinent
« Le programme n'a pas de grand effet sur la littératie »	Oui — au-delà de 3,5 points est écarté
« Le gain de 7 points n'est pas attribuable au programme »	Oui — les écoles témoins ont monté davantage
« Il faut réorienter le financement »	Non — cela exige une estimation d'effet que cette étude n'a pas
« La prochaine évaluation a besoin de 60 écoles »	Oui — calculé, et actionnable

Ce que vaut « aucun effet détecté »

- Les deux lignes « oui » du milieu sont de vrais constats et ce sont des constats importuns — ils retirent un chiffre...

Écrire pour le lecteur qui citera une seule phrase

- Pas — « Aucun effet statistiquement significatif de la cantine scolaire sur la littératie n'a été trouvé »
- Pas — « La littératie a monté de 7 points dans les écoles du programme »
- Ceci — « La littératie a monté de 7 points dans les écoles avec et sans programme ; cette évaluation n'était pas assez. . . »
- Mettez cette phrase dans le résumé, dans la conclusion et dans le courriel d'accompagnement — La phrase citable est. . .

Ce dont ce cours a traité

- Ce qui signifie que les décisions les plus lourdes de conséquences ont été prises avant l'existence de la moindre. . .
- La chose utile que ce cours demande n'est donc pas une technique — C'est d'être dans la pièce en janvier, avec une. . .

Rapportez-le en entier — Exemple (suite)

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

Rapportez-le en entier — Exemple (suite)

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.

- Huit lignes, et un relecteur peut reconstituer chaque décision à partir d'elles —
C'est le livrable vers lequel ce. . .

- Le module 5 est complet : l'incertitude, puis les modèles, puis les protocoles.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/impact-evaluation-methods/
ie-08-the-evaluation-that-says-no](https://data-analysis.cassion.dev/fr/cours/impact-evaluation-methods/ie-08-the-evaluation-that-says-no)