Huit points de progrès, dont deux que personne n'a gagnés

La lecture passe de 53,1 % à 61,8 % des items entre les deux passations. Sur les 585 élèves ayant passé les deux, elle passe de 56,8 % à 63,7 %. L'écart tient à qui s'est présenté, et les scores bruts n'auraient pu donner ni l'un ni l'autre.

Pierre Robentz Cassion

Ce que couvre cette leçon

- La comparaison qui ne peut pas être faite
- Trois choses comparables, et une qui ne l'est pas
- Qui a passé l'épreuve
- Mesurez la sélection au lieu de l'écarter par hypothèse
- Ce que coûte le panel
- Rapportez les trois nombres
- La suite

La comparaison qui ne peut pas être faite — En Python

```
import pandas as pd

assessment = pd.read_csv("learning-assessment-2024.v1.csv")
literacy = assessment[assessment["domain"] == "literacy"]

print(literacy.groupby("assessment_round").agg(
    students=("student_id", "nunique"),
    items=("items", "first"),
    mean_raw=("raw_score", "mean"),
).round(1))
```

La comparaison qui ne peut pas être faite — En R

```r
library(dplyr)

assessment |>
  filter(domain == "literacy") |>
  summarise(students = n_distinct(student_id), items = first(items),
            mean_raw = mean(raw_score), .by = assessment_round)
```

La comparaison qui ne peut pas être faite

Passation	Élèves	Items	Score brut moyen
Initiale	741	**40**	21,2
Finale	658	**50**	30,9

La comparaison qui ne peut pas être faite

- Un score brut est passé de 21,2 à 30,9 et cette comparaison ne signifie rien — car les instruments n'ont pas la même. . .

Trois choses comparables, et une qui ne l'est pas

- Non comparable : le score brut — Dénominateurs différents
- Comparable avec prudence : la proportion d'items réussis — Elle met les deux passations sur une échelle de 0 à 1, ce...

Trois choses comparables, et une qui ne l'est pas — En Python

```
literacy = literacy.assign(proportion=literacy["raw_score"] / literacy["items"])
print(literacy.groupby("assessment_round")["proportion"].mean().round(3))
```

Trois choses comparables, et une qui ne l'est pas — En R

```
assessment |> filter(domain == "literacy") |>
  summarise(proportion = mean(raw_score / items), .by = assessment_round)
```

Trois choses comparables, et une qui ne l'est pas

- 53,1 % au départ et 61,8 % à la fin — Mieux, et reposant encore sur une hypothèse que personne n'a vérifiée
- Comparable par construction : les bandes de compétence — Elles ont été équatées lors de la conception des instruments,...

Trois choses comparables, et une qui ne l'est pas — En Python

```
bands = pd.crosstab(assessment["assessment_round"],
                    assessment["proficiency_band"], normalize="index")
order = ["below-minimum", "minimum", "proficient", "advanced"]
print((bands[order] * 100).round(1))
```

Trois choses comparables, et une qui ne l'est pas — En R

```
assessment |> count(assessment_round, proficiency_band) |>
  mutate(share = n / sum(n), .by = assessment_round)
```

Trois choses comparables, et une qui ne l'est pas

Bande	Initiale	Finale
Sous le minimum	19,3 %	**11,1 %**
Minimum	30,6 %	23,9 %
Compétent	38,2 %	38,9 %
Avancé	11,9 %	**26,1 %**

Trois choses comparables, et une qui ne l'est pas

- La bande sous le minimum passe de 19,3 % à 11,1 % et la bande avancée de 11,9 % à 26,1 % — C'est la comparaison à...

Qui a passé l'épreuve — En Python

```
sat = assessment.groupby("assessment_round")["student_id"].apply(set)
baseline, endline = sat["baseline"], sat["endline"]

print(f"baseline only: {len(baseline - endline)}")
print(f"endline only:  {len(endline - baseline)}")
print(f"both rounds:   {len(baseline & endline)}")
```

Qui a passé l'épreuve — En R

```
# Three groups, and only one of them supports a change estimate.
```

Qui a passé l'épreuve

- 741 ont passé l'initiale, 658 la finale, et 585 les deux — 156 élèves ayant passé le premier test n'ont pas passé le...
- Ce ne sont pas 156 élèves au hasard — Le jour de l'évaluation attrape qui se trouve en classe, et les enfants qui n'y...

Mesurez la sélection au lieu de l'écarter par hypothèse — En Python

```
panel = literacy.pivot_table(index="student_id", columns="assessment_round",
                             values="proportion")
panel = panel.dropna()

print(f"panel of {len(panel)} students")
print(f"  baseline {panel['baseline'].mean():.1%} → endline {panel['endline'].mean():.1%}")

all_comers = literacy.groupby("assessment_round")["proportion"].mean()
print(f"all comers")
print(f"  baseline {all_comers['baseline']:.1%} → endline {all_comers['endline']:.1%}")
```

Mesurez la sélection au lieu de l'écarter par hypothèse — En R

```
assessment |> filter(domain == "literacy") |>
  mutate(proportion = raw_score / items) |>
  select(student_id, assessment_round, proportion) |>
  tidyr::pivot_wider(names_from = assessment_round, values_from = proportion) |>
  filter(!is.na(baseline), !is.na(endline)) |>
  summarise(n = n(), baseline = mean(baseline), endline = mean(endline))
```

Mesurez la sélection au lieu de l'écarter par hypothèse

	Initiale	Finale	Évolution
Tous présents	53,1 %	61,8 %	**+8,7 points**
Panel de 585	56,8 %	63,7 %	**+6,9 points**

Mesurez la sélection au lieu de l'écarter par hypothèse

- Deux des 8,7 points tiennent à qui s'est présenté — La valeur initiale du panel dépasse de 3,7 points celle de tous les…
- Rapportez l'estimation du panel et montrez le chiffre tous-présents à côté — Le panel est le nombre défendable ;…

Ce que coûte le panel

- Le panel n'est pas la population — 585 élèves sur 741 à l'initiale, et ce sont les plus assidus
- Il ne dit rien des 156 partis — Leurs apprentissages ont pu aller n'importe où, et le rapport honnête dit que…

Ce que coûte le panel — En Python

```
left = literacy[literacy["student_id"].isin(baseline - endline)
                & (literacy["assessment_round"] == "baseline")]
print(f"the 156 who did not return scored "
      f"{left['proportion'].mean():.1%} at baseline")
print(f"the 585 who did scored "
      f"{panel['baseline'].mean():.1%}")
```

Ce que coûte le panel — En R

```
# Quantify the difference rather than asserting it.
```

Ce que coûte le panel

- Les 156 qui ne sont pas revenus obtenaient 39,3 % à l'initiale contre 56,8 % pour le panel — dix-sept points en dessous…
- Chiffrer la sélection est le livrable — non l'éliminer

Rapportez les trois nombres — Exemple (suite)

```
Literacy, baseline to endline

  Proficiency bands, all who sat each round
    Below minimum        19.3% → 11.1%
    Advanced             11.9% → 26.1%

  Proportion of items correct
    All comers           53.1% → 61.8%   +8.7 points   n=741, n=658
    Panel of both rounds 56.8% → 63.7%   +6.9 points   n=585

  Instruments differed: 40 items at baseline, 50 at endline. Raw scores are
  not comparable and are not reported. Proficiency bands were equated at
  design; the proportion of items assumes equivalent difficulty.

  156 baseline students did not sit the endline. They scored 39.3% at
  baseline against 56.8% for those who returned, so the all-comers change
```

Rapportez les trois nombres — Exemple (suite)

```
overstates learning by about 1.8 points. The panel estimate covers
children who sat both rounds and does not cover those 156.
```

La suite

- Vous disposez désormais de tous les indicateurs éducatifs que ce cours peut produire — inscription, présence, rétention et apprentissages — chacun avec son dénominateur et sa réserve.

Lire la leçon complète, avec le code exécutable

https://data-analysis.cassion.dev/fr/cours/education-programme-analysis/edu-07-comparing-two-rounds