

mtss-k

Déterminer l'efficacité et l'efficience des interventions Multi-Tiered System of Supports (MTSS) : planification de deux essais contrôlés randomisés dans cinq pays de l'Union européenne

Dachet D., Altinok N., Pellegrini M., Dierendonck C. & Baye A.



Plan de la présentation

1. **Phase pilote** : validation des instruments de mesure
2. **RCT phase 1** : effets de l'intervention SEB MTSS-K
3. **RCT phase 2** : effets des composantes littératie et numératie de l'intervention MTSS-K
4. **Analyse coût-efficacité**
5. **Débat méthodologique** : effets des interventions de niveau 2
 - RCT originelle
 - Appariement par score de propension
 - RCT avec seconde randomisation
 - Design en régression discontinue
 - (Analyses par variable instrumentale)
6. **Recommandations Open Science**
7. **Bibliographie**



1° Phase pilote : validation des instruments de mesure

- **Trois types** de mesure

- Mesure directe sur les enfants, les enseignants et les familles
- Mesure indirecte des enfants par les enseignants
- Mesure indirecte des enfants par les familles

Country	N students
Belgium	125
Luxembourg	125
Germany	125
France	125

Switzerland	125
Italy	125
Total	750

- **Trois fonctions**

- Screening
- Progress monitoring
- Effets de l'intervention

- **Trois variables d'intérêt**

- Littératie précoce
- Numératie précoce
- Compétences SEB

#	Phase	#	Step
1	Item development	1	Identification of the domain and items generation
		2	Assessment of content validity by experts
2	Scale development	3	Qualitative pre-testing
		4	Quantitative pre-testing
		5	Item reduction
		6	Exploratory factor analysis
3	Scale evaluation	7	Dimensionality analysis and creation of scales scores
		8	Tests of reliability
		9	Tests of criterion validity

2° RCT phase 1 : effets de l'intervention SEB MTSS-K

Blocked stratified/cluster randomization

Année 1, cohorte 1 (2025-2026)

Groupe expérimental

Prétest

Posttest

20 classes belges
20 classes allemandes
10 classes luxembourgeoise
20 classes françaises
15 classes suisses

$\delta = 0.26$ (Lee & Gage, 2019)
 $f = 0.13$
ICC = ?
Power [0.54-0.99]

**Analyses multiniveaux
(Finch et al., 2019)**

Groupe contrôle

20 classes belges
20 classes allemandes
10 classes luxembourgeoise
20 classes françaises
15 classes suisses



Ecoles au sein de la population qui respectent les critères d'inclusion dans l'étude

Assignment aléatoire



3° RCT phase 2 : effets de l'intervention LIT/NUM MTSS-K

Blocked stratified/cluster randomization // Année 1

Année 2, cohorte 2 (2026-2027)

Groupe expérimental

Numeracy

20 classes belges
20 classes allemandes
10 classes luxembourgeoise

Literacy

20 classes françaises
15 classes suisses

Groupe contrôle

20 classes belges
20 classes allemandes
10 classes luxembourgeoise
20 classes françaises
15 classes suisses

Prétest

δ numeracy = 0.31
 $\Delta^t_{Covariate} = 0.5722$
 $\Delta^t_{Covariate} = 1.0572$
Power [0.81-0.99]

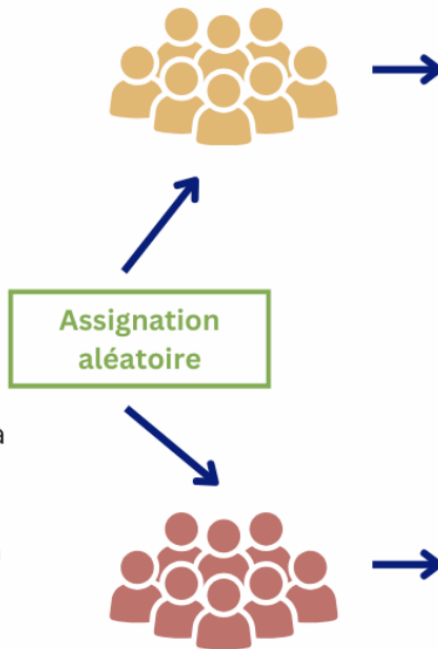
δ literacy = 0.186
 $\Delta^t_{Covariate} = 0.3456$
 $\Delta^t_{Covariate} = 0.7099$
Power [0.31-0.83]

Posttest

**Analyses multiniveaux
(Finch et al., 2019)**



Ecoles au sein de la population qui respectent les critères d'inclusion dans l'étude



4° Analyse coût-efficacité

- *Ingredients method* (Levin et al., 2018)
- Données nécessaires
 - Donnée d'efficacité = gain sur la (les) variable(s) d'intérêt dans le GE comparativement au GC
 - Coûts totaux = coût absolu du programme expérimental et du programme contrôle
 - Coûts incrémentiels = coût GE – coût GC

Table 4. Incremental Cost-effectiveness ratios for 1-year MTSS-K implementation

Implementation scenario	Cost per teacher	Cost per student	Average effect size	CER
School-level, one MTSS teacher, 12 students	\$9,493	\$791	0.19	\$4,163
School-level, two MTSS teachers, 12 students	\$7,424	\$619	0.19	\$3,256
District-level, five schools, 12 students	\$6,221	\$518	0.19	\$2,726
School-level, one MTSS teacher, 6 students	\$9,493	\$1,582	0.19	\$8,327
School-level, one MTSS teacher, 24 students	\$9,493	\$396	0.19	\$2,082
District-level, ten schools, 12 students	\$5,816	\$485	0.19	\$2,551

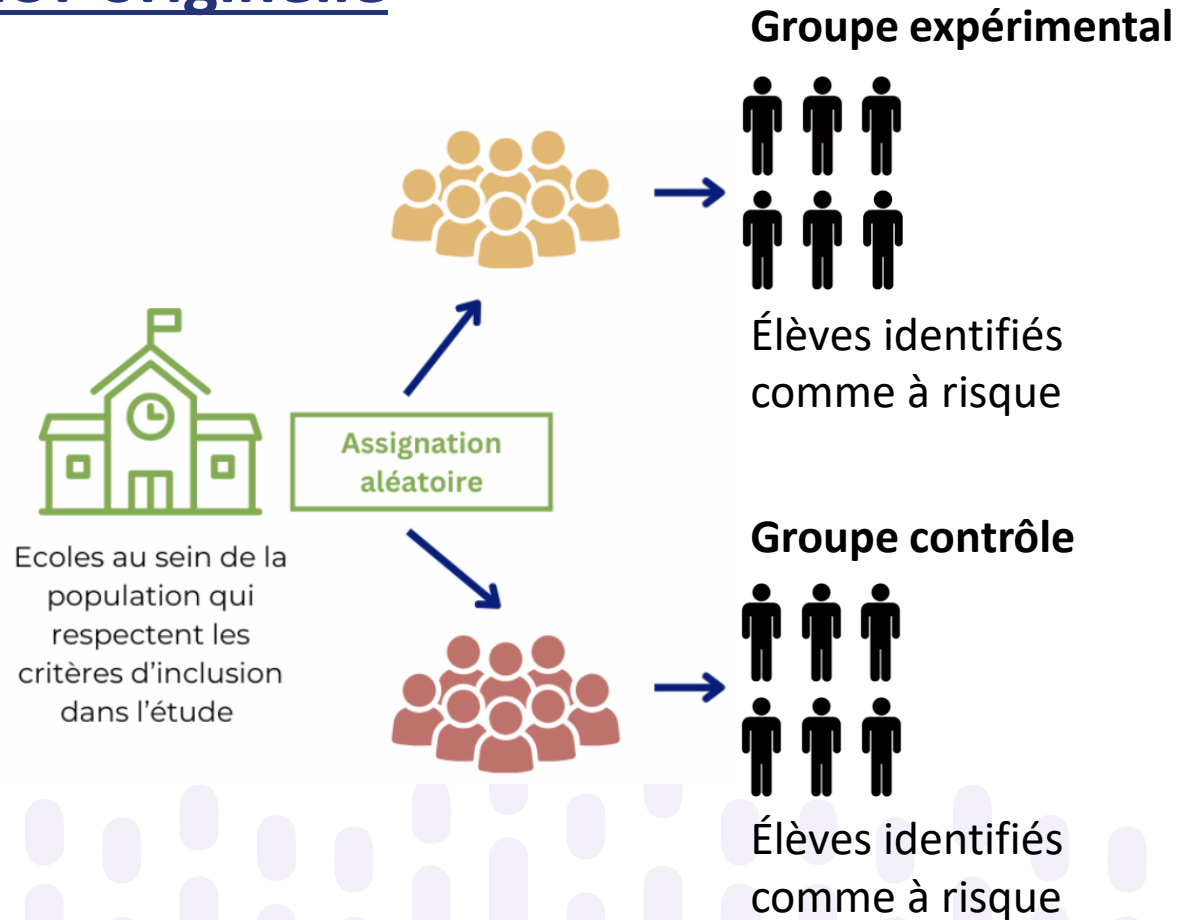
E\$timator 2015-22 Teachers College, Columbia University

[Quickstart](#) [Manual](#) [License](#) [Tutorials](#) [Additional Resources](#)

[Home](#) [Project List](#) [Prices](#) [My Account](#) [Log out](#)

E\$timator

RCT originelle



Mais quid de la valeur ajoutée du niveau 2 d'intervention ?

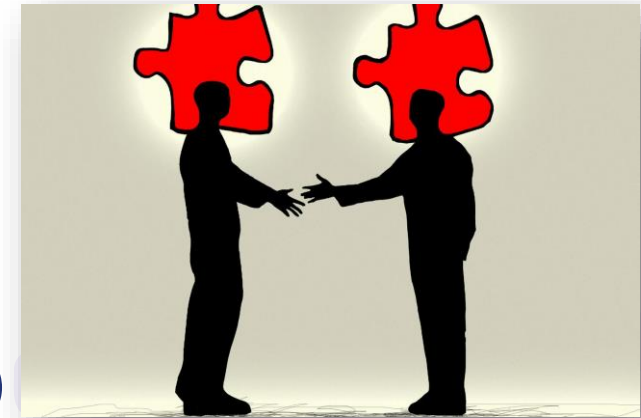
Mais quelle équivalence des groupes « à risque » au prétest ?

Appariement par score de propension

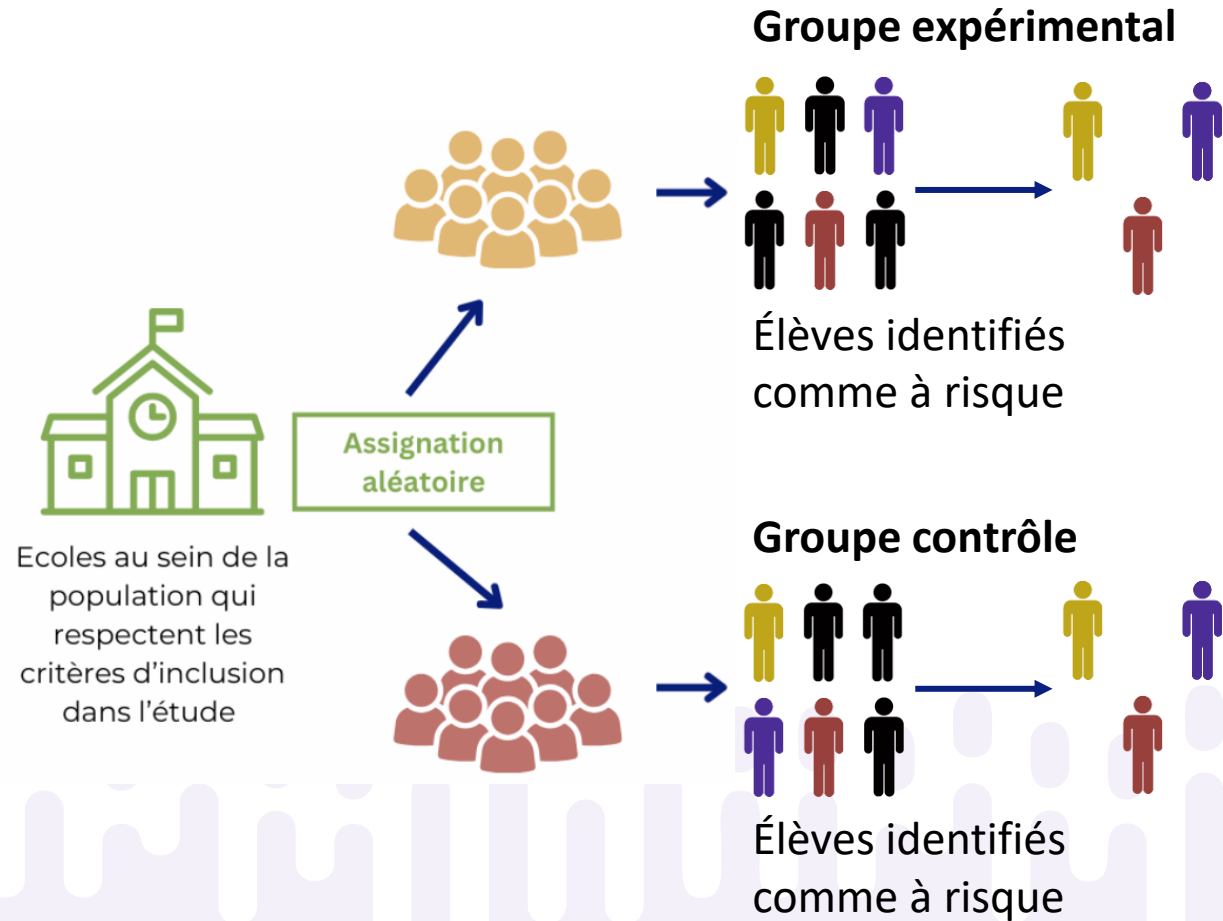
Quelques préalables statistiques/méthodologiques (Austin, 2011)

- Estimation d'une probabilité d'appartenir au GE ou au GC // régression logistique avec pour VD l'appartenance au groupe et pour VI l'ensemble des co-variables ayant un effet sur la (les) variable(s) d'intérêt de l'étude
 - Identifiables à partir d'approches *step wise*
- Plusieurs méthodes → 1:1 ; 1:K ; tolérance (caliper)
- Effet de l'intervention est déterminé à partir de l'échantillon apparié
- Mais risque d'attrition plus ou moins important

- Développement d'algorithmes et de méthodes plus robustes (Zhu et al., 2015)

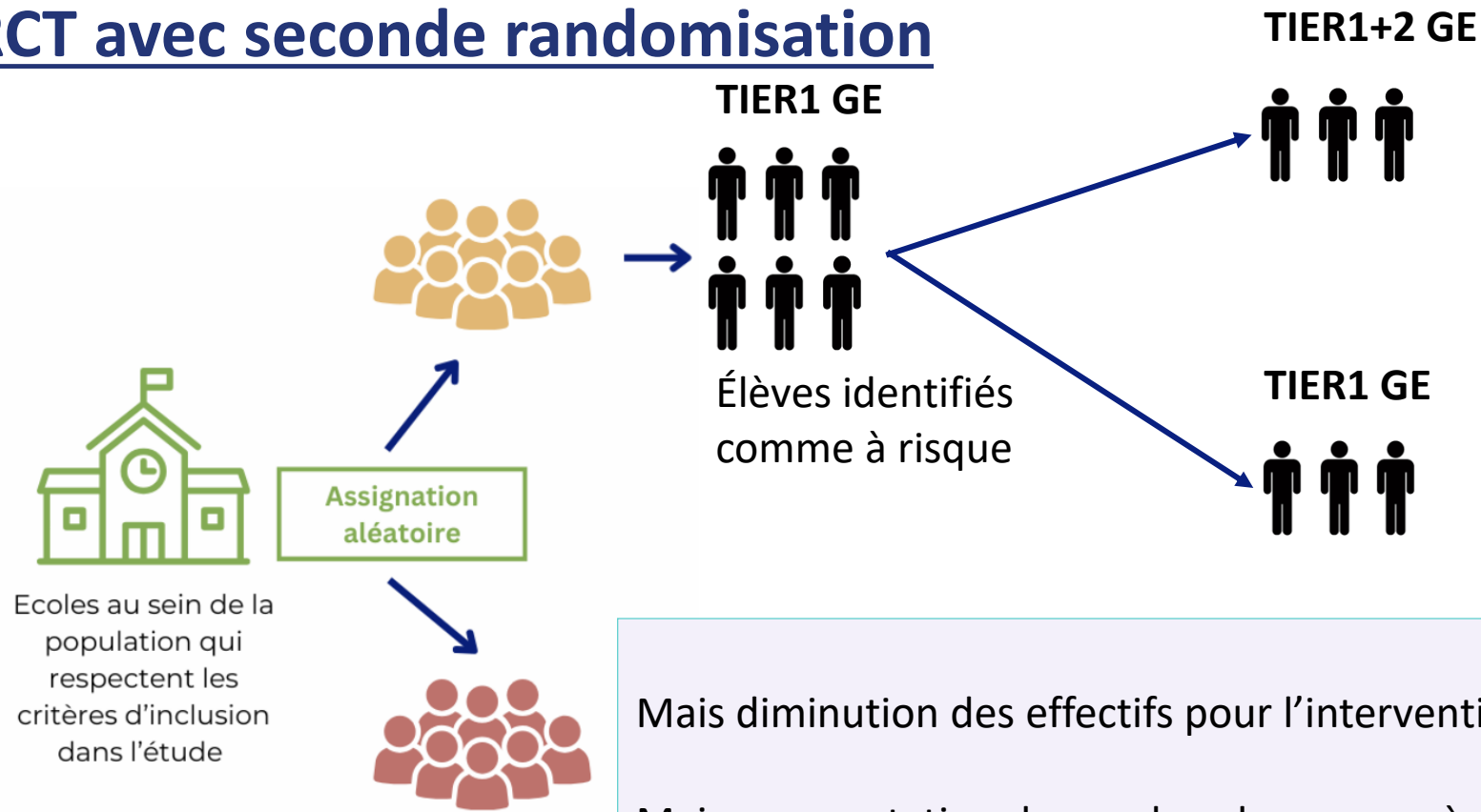


Appariement par score de propension



Mais quid de la valeur ajoutée du niveau 2 d'intervention ?
Mais analyse post-hoc souvent critiquée (Slavin et al., 2019)

RCT avec seconde randomisation



Mais diminution des effectifs pour l'intervention de niveau 2

Mais augmentation du nombre de groupes à comparer

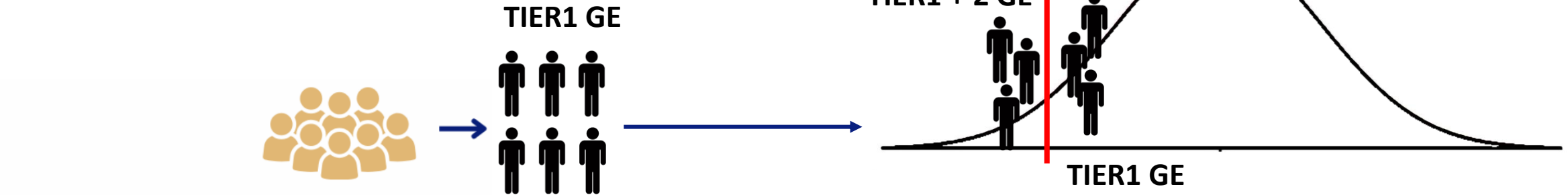
➔ Diminution de la puissance statistique (Barker Bausell & Li, 2002)

Design en régression discontinue

Quelques préalables statistiques/méthodologiques (Jacob & Zhu, 2012)

- *Running variable* sur laquelle un seuil (*cut-off*) est utilisé pour déterminer l'éligibilité d'un individu à l'obtention (ou non) d'une intervention
- Individus juste au-dessus et juste en-dessous du seuil sont jugés comparables sur toutes les dimensions sauf sur l'assignement à l'intervention → discontinuité
- Empan (*bandwidth*) d'individus sélectionnés de part et d'autre du seuil
- Standard/casual/sharp RDD = l'assignement au traitement est strict
- Fuzzy RDD = l'assignement n'est pas strict, même si la probabilité d'assignement au traitement change significativement au niveau du seuil

Design en régression discontinue



Ecoles au sein de la population qui respectent les critères d'inclusion dans l'étude

Assignation aléatoire

Mais les deux groupes comparés restent quantitativement différents, sauf pour la faible population autour du cut-off

→ Quel empan définir/sélectionner ?

→ Quelle puissance statistique pour les analyses ?

Mais dans l'idéal un seul cut-off pour toutes les classes, toutes les écoles et tous les pays et un assignement le plus strict possible ...

Mais on considère une erreur de mesure au cut-off // question éthique

Mais la causalité n'est que « locale »

Analyse par variable instrumentale

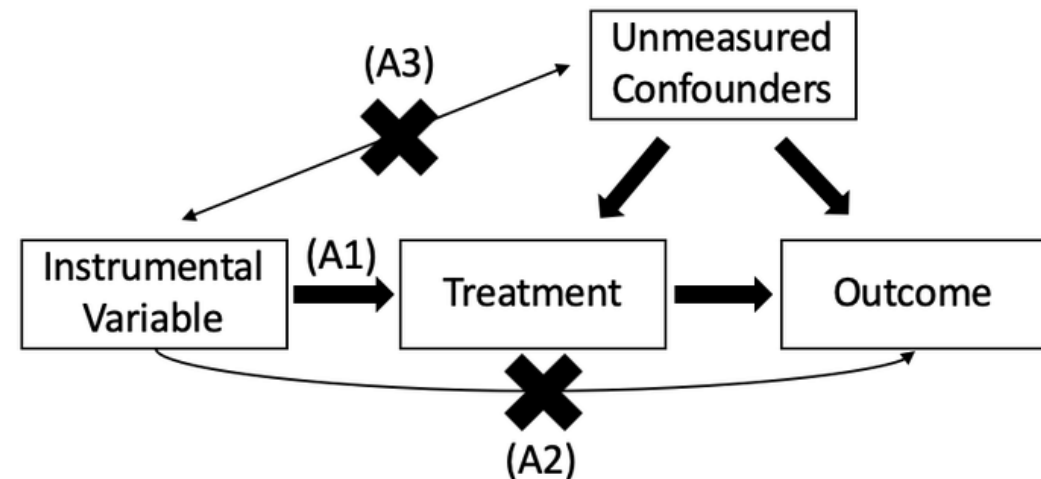
Quelques préalables statistiques/méthodologiques (Angrist & Krueger, 2001)

- Variable instrumentale = variable extérieure qui permet d'étudier l'effet causal de la variable explicative principale sur la variable dépendante


Z (variable instrumentale) permet d'identifier l'effet de X (variable explicative) sur Y (variable dépendante)

→ Z est corrélé avec X

→ Z affecte Y par X (et donc n'est pas corrélé avec les facteurs non observés)



6° Recommandations Open Science

- Respect des standards méthodologiques de l'IES (WWC, 2022)
 - les procédures de calcul a priori des tailles d'échantillons recommandées par Hedges et Rhoads (2010)
 - la méthode d'assignement des participants aux conditions ;
 - le taux (individuel ou par cluster ; général ou différentiel) d'attrition expérimentale ;
 - le taux (individuel ou par cluster ; général ou différentiel) de nouveaux arrivants ;
 - l'équivalence au prétest ;
 - la qualité des instruments de mesure.
 - Préenregistrement des protocoles(Nosek et al., 20)
 - Trois méta-analyses
 - Deux RCTs
 - FAIR (GoFair, n.d.) & Open Data :
 - Findable = faciles à retrouver par tous (identificateur pérenne et unique, métadonnées riches)
 - Accessible = authentification et autorisation
 - Interoperable = intégration avec d'autres données et utilisation avec des applications d'analyse, de stockage et de traitement
 - Reusable = description précise et complète
- 

7° Bibliographie

- Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives*, 15, 68-85. <https://doi.org/10.1257/jep.15.4.69>
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399-424. <https://doi.org/10.1080/00273171.2011.5>
- Barker Bausell, R., & Li, Y.-F. (2002). *Power Analysis for Experimental Research: a practical guide for the biological, medical and social sciences*. Cambridge University Press.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in public health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2020). *A practical introduction to regression discontinuity designs : foundations*. Cambridge Elements.
- Finch, W. H., Bolin, J. E., & Kelley, K. (2019). *Multilevel modeling using R* (2nd ed.). CRC Press (Taylor & Francis Group)
- GoFair. (n.d.). *FAIR Principles*. Retrieved from: <https://www.go-fair.org/fair-principles/>
- Hedges, L. V., & Rhoads, C. (2010). *Statistical power analysis in education research*. Institute of Education Sciences.
- Jacob, R., & Zhu, P. (2012). *A practical guide to regression discontinuity*. New York, NY: mdc
- Johnson, M., Cao, J., & Kang, H. (2019). Detecting heterogeneous treatment effect with instrumental variables. *arXiv:1908.03652*.
<https://doi.org/10.48550/arXiv.1908.0365>
- Nosek, B. A., Alter, G., Bancks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Gorog, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T. A., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Levy Paluck, E., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R. & Yarkoni, T. (2015). Promoting an open research culture : Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*, 348(6242), 1422-1425.
<https://doi.org/10.1126/science.aab2374>
- Slavin, R.E., Lake, C., Inns, A., Baye, A., Dachtel, D., & Haslam, J. (2019). *A quantitative synthesis of research on writing approaches in grades 2 to 12*. Education Endowment Foundation.
- Zhu, Y., Coffman, D. L., & Ghosh, D. (2015). A Boosting Algorithm for Estimating Generalized Propensity Scores with Continuous Treatments. *Journal of causal inference*, 3(1), 25-40. <https://doi.org/10.1515/jci-2014-0022>
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). This report is available on the What Works Clearinghouse website at <https://ies.ed.gov/ncee/wwc/Handbooks>