Analyse de la Qualité Méthodologique de la Littérature sur le Genre et le Format des Questions : une étude de méta-recherche exploratoire

Léonard, F[1] Monseur, C[1] Durieux, N[2]

[1]Université de Liège, Département des sciences de l'éducation et de la formation, RUCHE

[2]Université de Liège, Département de psychologie, RUCHE

09-01-2025

Introduction

Contexte général (1)

- Avec l'augmentation continue de la taille des cohortes dans l'enseignement supérieur, les enseignants sont de plus en plus amenés à utiliser des formats d'évaluation standardisés.
- Les examens composés principalement de questions à choix multiples (QCM) deviennent de plus en plus fréquents, en raison de leur praticité pour évaluer rapidement un grand nombre d'étudiants.
- Cependant, depuis plusieurs décennies, des chercheurs se préoccupent de l'équité des QCM.
- Ces questions d'équité pédagogique sont devenues centrales dans les débats académiques, et sociétaux.

Contexte général (2)

- Les recherches publiées s'interrogeant sur l'équité en lien avec le genre et le format des questions ne parviennent pas à établir un consensus scientifique.
- Ce manque de consensus pourrait en partie s'expliquer par la diversité des niveaux scolaires étudiés, des matières évaluées ou par des lacunes méthodologiques dans les études.

Contexte général (3)

Une étude de qualité voulant investiguer le biais de genre en fonction du format des questions devrait inclure, entre autres :

- une estimation de l'interaction entre le genre et le format des questions,
- une justification de la taille de l'échantillon,
- l'utilisation de statistiques appropriées, comprenant un calcul de l'erreur standard tenant compte du plan d'échantillonnage,
  - Lorsque l'erreur standard n'est pas calculée en adéquation avec le plan d'échantillonnage, cela conduit à une sous-estimation de l'erreur, une surestimation de l'effet et une augmentation des faux positifs (Copas & Li, 1997; Hirschauer et al., 2020, 2021; Nielsen et al., 2009).
- le compte rendu des tailles d'effets accompagnées de leurs intervalles de confiance.

Qualité du compte-rendu et qualité méthodologique d'une étude

- La qualité du compte-rendu d'une étude scientifique constitue un aspect essentiel de la rigueur méthodologique d'une étude. Un compte-rendu insuffisant peut introduire des biais d'interprétation, compromettant la fiabilité et l'utilité des résultats (Fanelli, 2013).
- Évaluer la complétude du compte-rendu permet d'identifier de potentiels axes d'améliorations afin de garantir des études fiables et potentiellement utiles à l'ensemble de la communauté.

Objectifs de l'étude

Cette étude de méta-recherche exploratoire examine la qualité méthodologique au travers de la complétude du compte-rendu statistique d'un échantillon d'études portant sur **l'interaction entre le genre et le format des questions,** et leur impact sur la performance des étudiants.

- Plusieurs critères sont évalués :
  - la justification de la taille de l'échantillon,
  - la présence d'une estimation de l'interaction,
  - la rigueur dans le calcul de l'erreur standard,
  - le compte-rendu des tailles d'effets et de leurs intervalles de confiance.

Méthode

Sélection des études : critères d'inclusion et d'exclusion

Critères d'inclusion

- Études publiées dans des journaux avec relecture par les pairs
- Études sur des étudiants sans handicap
- Études comparant les genres
- Études impliquant à la fois des questions à choix multiple et des questions ouvertes
- Études centrées sur des élèves du grade 6 jusqu'à l'enseignement supérieur
- Études analysant les disciplines académiques couvertes dans le programme d'enseignement.
- Études mesurant la performance académique ou les taux d'omission

Critère d'exclusion

- Études sur les évaluations orales
- Études évaluant des travaux pratiques ou des devoirs

Recherche de publications

Une équation de recherche a été élaborée de manière itérative dans la base de données ERIC (interface Ovid) en novembre 2024, en combinant du langage contrôlé et du langage libre.

1	Test Format/
2	((test* or item* or evaluat* or assess* or exam*) adj3 (format* or type*)).ti,ab.
3	Objective Tests/
4	Multiple Choice Tests/
5	((objective or "true false" or multiple* or close*) adj3 (test* or item* or question*)).ti,ab.
6	3 or 4 or 5
7	Essay Tests/
8	((essay or open* or structur* or construct* or free or short*) adj3 (test* or item* or question*)).ti,ab.
9	7 or 8
10	6 and 9
11	1 or 2 or 10
12	Gender Differences/
13	Gender Issues/
14	gender*.ti,ab.
15	12 or 13 or 14
16	Academic Achievement/
17	Science Achievement/
18	Mathematics Achievement/
19	Reading Achievement/
20	Achievement Gap/
21	Test Results/
22	((academic or schola* or school* or student* or science* or math* or read* or gap* or test*) adj3 (achiev* or perform* or result*)).ti,ab.
23	16 or 17 or 18 or 19 or 20 or 21 or 22
24	11 and 15 and 23
25	(gender* adj3 (achiev* or perform* or result*)).ti,ab.
26	11 and 25
27	24 or 26

Procédure de sélection

Toutes les étapes de la sélection ont été réalisées à l'aide du logiciel Covidence (Veritas Health Innovation, 2024).

Sélection des titres et résumés

- La sélection des titres et résumés a été effectuée par L.F., D.N. et M.C.
- L.F. a examiné l'ensemble des références identifiées, tandis que D.N. et M.C. ont partagé équitablement les références restantes.
- Les conflits ont été résolus par consensus entre les auteurs.

Sélection sur la base des textes intégraux

- La sélection des textes intégraux a été réalisée par L.F.
- En cas de doute, D.N. ou M.C. ont été consultés.

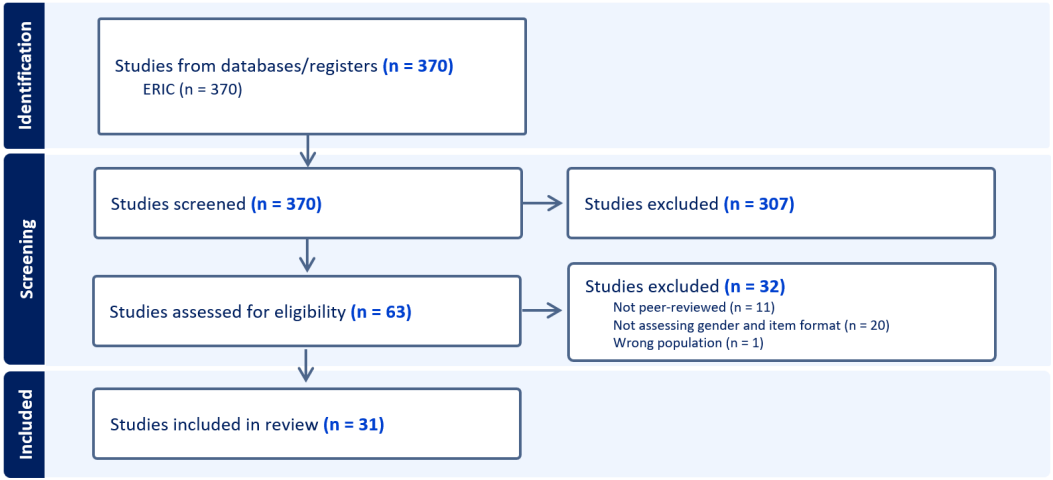
Extraction et analyse des données

- L'extraction a été réalisée avec une grille préétablie dans le logiciel Excel par L.F. En cas de doute, M.C. était consulté.
- Les données extraites ont été synthétisée à l'aide de comptages en raison du faible nombre d'étude et de l'hétérogénéité importante entre les études incluses.

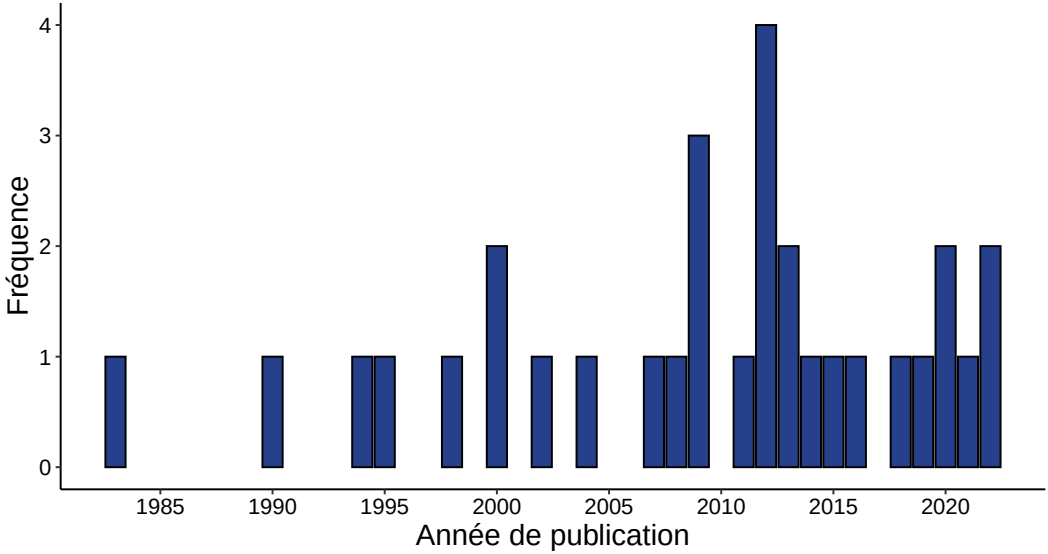
- Matière
- Niveaux scolaires
- Enjeux de l'évaluation
- Type d'étude
- Analyse statistique utilisée
- Test de l'interaction
- Effectif
- Justification de l'effectif
- Présence de taille d'effet
- Présence de l'intervalle de confiance de la taille d'effet
- Calcul de l'erreur type en fonction du plan d'échantillonnage

Résultats

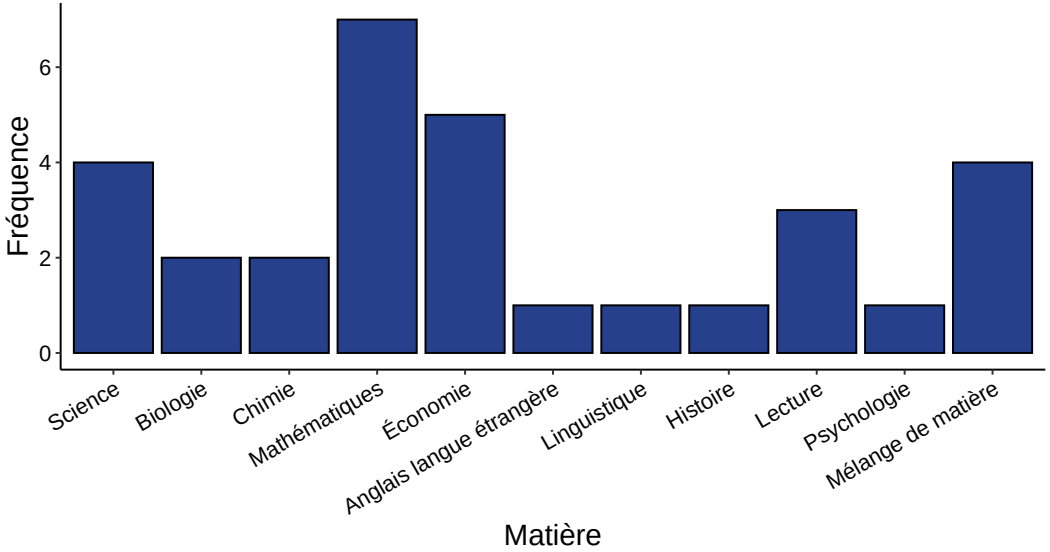
Diagramme de flux



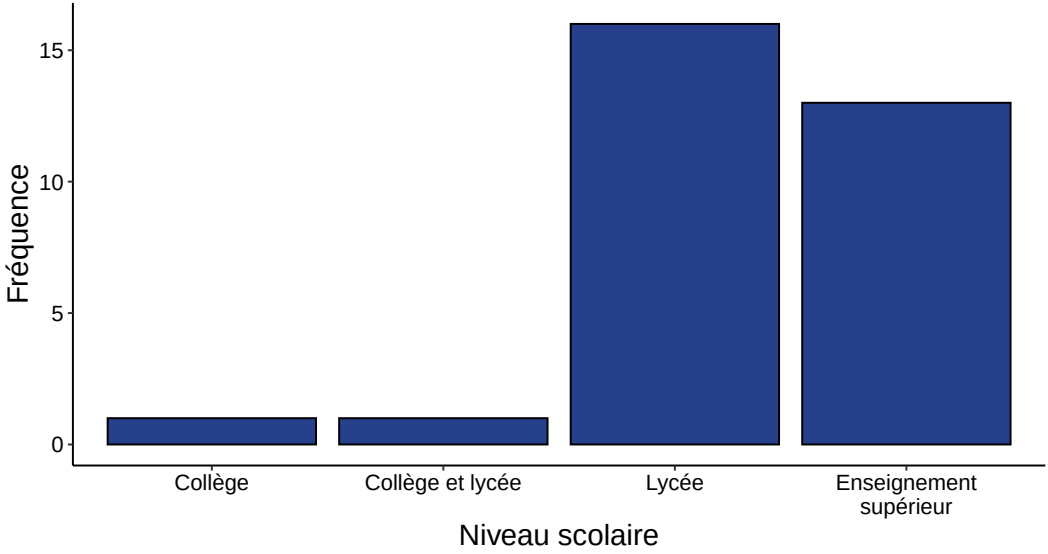
Années de publication



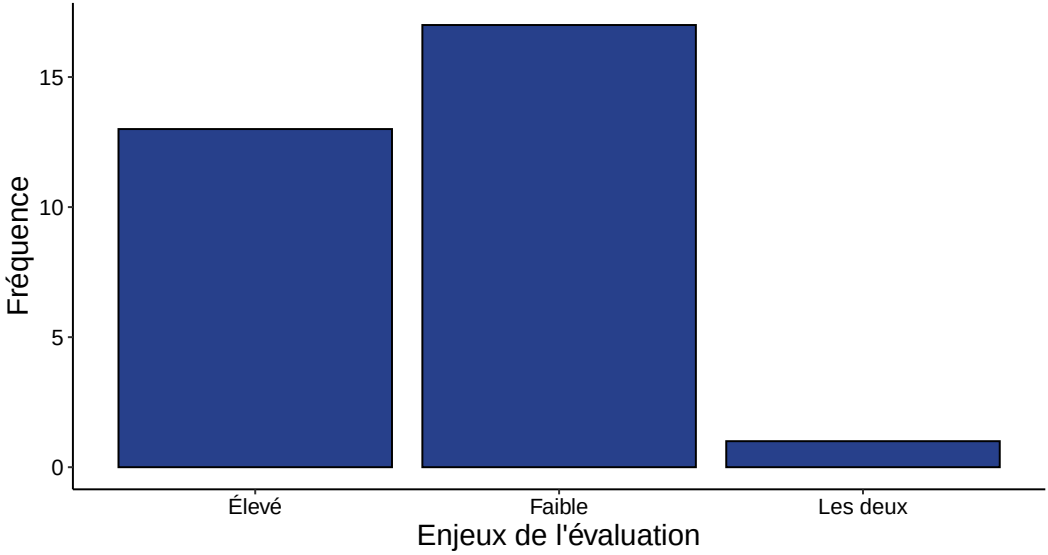
Matière évaluée



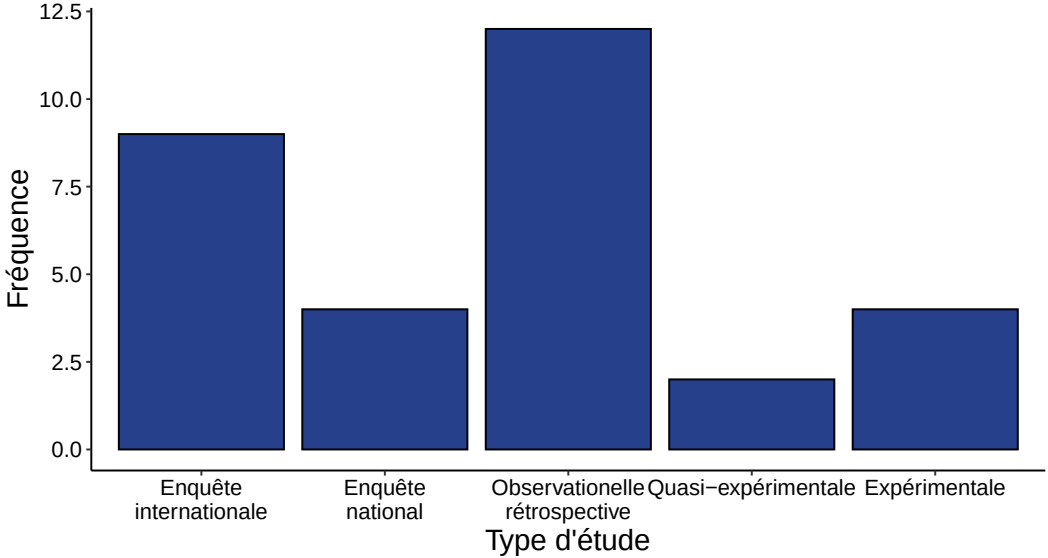
Niveaux scolaires



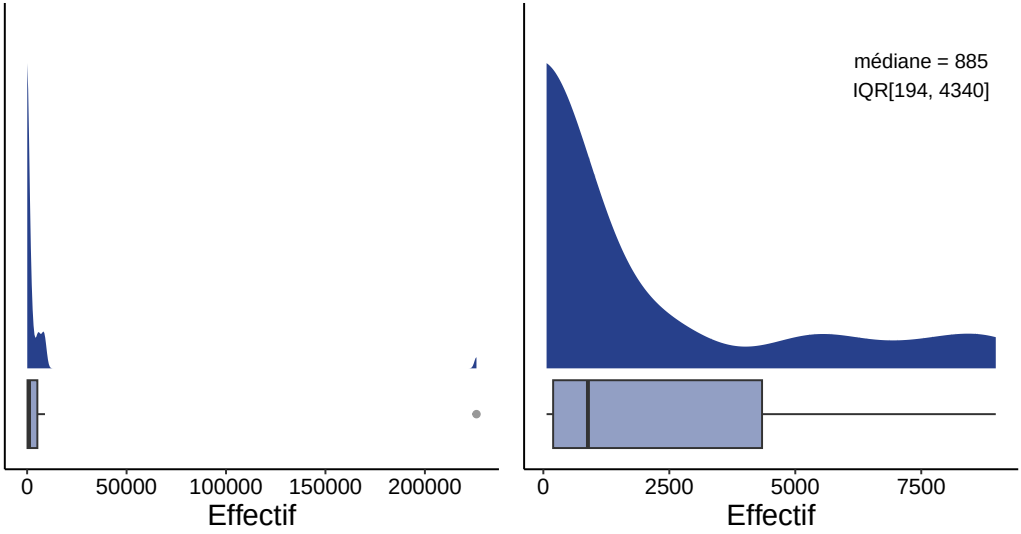
Enjeux de l'évaluation



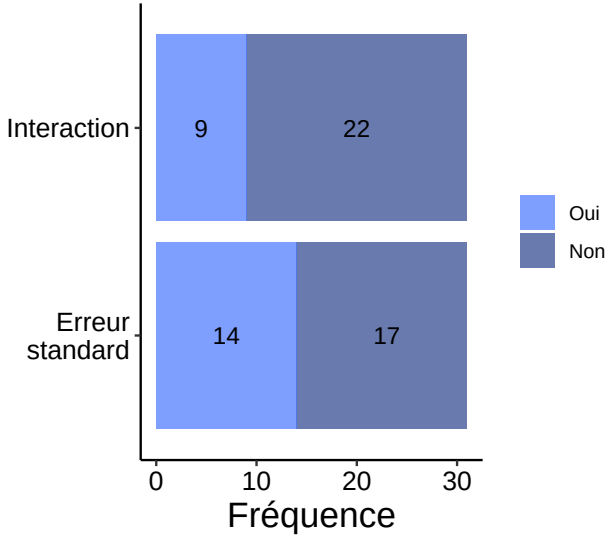
Type d'étude


12.5
10.0
7.5
5.0
2.5
0.0
Fréquence
Enquête internationale
Enquête national
Observationelle rétrospective
Quasi−expérimentale
Expérimentale
Type d'étude

Effectif


médiane = 885
IQR[194, 4340]
0
50000
100000
150000
200000
Effectif
0
2500
5000
7500
Effectif

Interaction et erreur standard



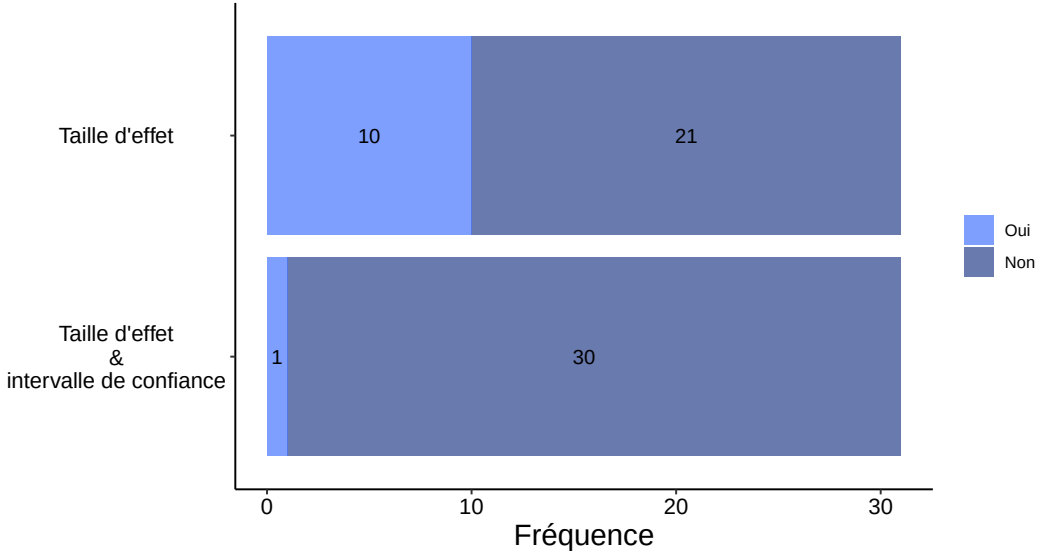
Interaction

- L'étude avec le plus grand effectif rapporte un coefficient de regression de 0.06 (0.02) pour l'interaction.

Méthode de calcul de l'erreur de mesure

- Balanced *repeated replications* : 3 études
- *Jackknife* : 1 étude
- Modèle multi-niveaux : 6 études
- Échantillonnage aléatoire : 4 études

Taille d'effet et intervalle de confiance



Discussion

Observations principales

- **Répartition thématique et méthodologique :**
  - Matières majoritairement représentées : les mathématiques et les sciences.
  - Les études observationnelles rétrospectives et les enquêtes internationales, dominent, peu d'études expérimentales.
- **Impact des enjeux d'évaluation :**
  - Légère majorité des études avec des enjeux faibles.
- **Problèmes de qualité méthodologique :**
  - Taille d'échantillon souvent non justifiée.
  - Peu d'études incluent des intervalles de confiance pour les tailles d'effet.
  - Moins de la moitié des études testent l'interaction entre le genre et le format.
  - Calculs d'erreurs standards souvent inappropriés, sous-estimant ainsi l'erreur standard et augmentant le taux de faux positif.

Limites

- Les études incluses ont été sélectionnées uniquement via la base de données ERIC. L'exclusion des études non publiées, telles que les thèses ou communications lors de conférences.
  - Cependant, l'utilisation d'échantillon d'étude est une pratique courante en méta-recherche (Mbuagbaw et al., 2020).
- La sélection sur les textes intégraux et l'extraction des données a été réalisée par un seul chercheur.
- La diversité des méthodologies, formats d'évaluation et matières rend difficile une synthèse cohérente des résultats.

Recommandations pour des recherches futures

- **Standardisation des pratiques méthodologiques :**
  - Justification des tailles d'échantillons (quand cela est pertinent).
  - Inclusion des tailles d'effet, des intervalles de confiance, et des calculs d'interactions.
- **Focus sur des approches expérimentales :**
  - Augmenter la proportion d'études expérimentales pour tester les interactions dans des contextes contrôlés.
- **Encourager le partage des codes et données pour une science transparente et reproductible.**

Références

Références I

Copas, J. B., & Li, H. G. (1997). Inference for non-random samples. *Journal of the Royal Statistical Society: Series B (Methodological)*, *59*(1), 55–95. https://doi.org/10.1111/1467-9868.00055

Fanelli, D. (2013). Redefine misconduct as distorted reporting. *Nature News*, *494*(7436), 149. https://doi.org/10.1038/494149a

Hirschauer, N., Grüner, S., Mußhoff, O., Becker, C., & Jantsch, A. (2020). Can p-values be meaningfully interpreted without random sampling? *Statistics Surveys*, *14*(none). https://doi.org/10.1214/20-SS129

Hirschauer, N., Grüner, S., Mußhoff, O., Becker, C., & Jantsch, A. (2021). Inference using non-random samples? Stop right there! *Significance*, *18*(5), 20–24. https://doi.org/10.1111/1740-9713.01568

Références II

Mbuagbaw, L., Lawson, D. O., Puljak, L., Allison, D. B., & Thabane, L. (2020). A tutorial on methodological studies: The what, when, how and why. *BMC Medical Research Methodology*, *20*(1), 226. https://doi.org/10.1186/s12874-020-01107-7

Nielsen, R. B., Davern, M., Jones Jr., A., & Boies, J. L. (2009). Complex Sample Design Effects and Health Insurance Variance Estimation. *Journal of Consumer Affairs*, *43*(2), 346–366. https://doi.org/10.1111/j.1745-6606.2009.01143.x

Veritas Health Innovation. (2024). *Covidence*. Veritas Health Innovation.