
Étude de l'effet du mode de passation d'évaluations standardisées sur la performance des élèves : l'exemple de CEDRE mathématiques collège

Reinaldo DOS SANTOS ()*

() Ministère de l'Éducation Nationale, DEPP*

`reinaldo.dos-santos@education.gouv.fr`

Mots-clés. (6 maximum) : Modèles mixtes, multimode, psychométrie, séries temporelles.

Domaines. 4.4 Effets de mode ; 19.1 Évaluation des élèves

Résumé (entre 350 et 900 mots environ)

La Direction de l'Évaluation, de la Prospective et de la Performance (DEPP) du Ministère de l'Éducation Nationale s'est engagée dans la transition d'enquêtes réalisées sur papier à des enquêtes au format numérique. Cette transition pose certaines questions d'ordre méthodologique, notamment en termes de comparabilité entre les cycles des séries temporelles disciplinaires (enquêtes CEDRE).

Afin de pouvoir observer l'impact du changement de mode d'attribution des évaluations, il faut pouvoir quantifier l'effet du mode sur l'item. Une fois cet « effet mode » pris en compte, il faut s'assurer que la performance des élèves demeure identique quel que soit le mode d'attribution. La théorie psychométrique permet de décrire les caractéristiques des items et la performance des élèves à une évaluation. Pour ce faire, il est nécessaire de créer une échelle abstraite, déterminée par la distribution de la performance des élèves.

Nous avons construit une évaluation composée d'items dans leur format papier et leur format numérique. Celle-ci a été présentée à un échantillon représentatif d'élèves de 3^{ème}. À l'aide d'un plan en blocs incomplet équilibré, nous avons garanti qu'un élève ne croise pas deux fois le même item. L'utilisation de ce design expérimental a permis de calculer, pour chaque élève de notre étude, un score obtenu sur l'enquête « papier crayon » et un score obtenu sur l'enquête au format numérique. L'utilisation des modèles mixtes multiniveaux permettent alors de décomposer les scores des élèves entre effets aléatoires à chacun des niveaux (intra-individuel, intra-établissement, intra-académique) et effets fixes, dont notamment le mode d'attribution de l'évaluation.

Cette étude démontre que la transition numérique ne remet pas en cause la qualité d'une série longitudinale en termes de comparaison dans le temps, ni en termes d'égalité de traitement entre les différents groupes d'élèves. Toutefois, il n'est pas possible de considérer que les items demeurent invariants quand on les adapte au format numérique. Il est donc nécessaire de maintenir

un ancrage « papier crayon » avant de basculer l'ensemble de l'évaluation au format numérique.

Bibliographie

Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *The Journal of Technology, Learning and Assessment*, 6(9).

Bressoux, P., Coustère, P., & Leroy-Audouin, C. (1997). Les modèles multiniveaux dans l'analyse écologique : le cas de la recherche en éducation. *Revue française de sociologie*, 67-96.

Fishbein, B., Martin, M. O., Mullis, I. V., & Foy, P. (2018). The TIMSS 2019 item equivalence study : examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education*, 6(1), 11.

Givord, P., & Guillerm, M. (2016). Les modèles multiniveaux (No. m2016-05). Institut National de la Statistique et des Etudes Economiques.

Jerrim, J., Micklewright, J., Heine, J. H., Salzer, C., & McKeown, C. (2018). PISA 2015 : how big is the mode effect? and what has been done about it? *Oxford Review of Education*, 44(4), 476-493.

Khoshsima, H., & Hashemi Toroujeni, S. M. (2017). Comparability of Computer-Based Testing and Paper-Based Testing : Testing mode effect, testing mode order, computer attitudes and testing mode preference. *International Journal of Computer (IJC)*, 24(1), 80-99.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests : A meta-analysis. *Psychological bulletin*, 114(3), 449.

Ninnin, L. M., & Salles, F. (2020). Cedre 2008-2014-2019. Mathématiques en fin de collège : des résultats en baisse.

Rocher, T. (2015). Quelles méthodes pour l'évaluation standardisée des compétences des élèves? *Statistique et Société*, 3(2), 59-66.

Sautory, O. (1991). Redressements d'échantillons d'enquêtes auprès des ménages par calage sur marges. *Proceedings, Journée de Méthodologie Statistique*, 299-326.

Thionet, P. (1961). La méthode des sondages. *Revue de Statistique Appliquée*, 9(1), 7-52.

Tillé, Y. (1998). Estimation in surveys using conditional inclusion probabilities : Simple random sampling. *International Statistical Review*, 66(3), 303-322.