

Fichier à compléter et intégrer dans votre dépôt sur *Sciencesconf* :

Communication individuelle

Merci de pas inscrire les noms et prénoms des auteurs et intervenants dans ce document pour l'anonymat de l'évaluation.

1- Titre de la communication :

Les performances des élèves redoublants sont-elles sous-estimées ?

2- Axe :

Axe 1 : Evaluation des apprentissages tout au long de la vie : continuités et ruptures

3- Résumé court de la communication :

Consigne : 300 mots maximum¹

Depuis les années 50, la docimologie met en évidence les inégalités de traitement évaluatif dont sont victimes les élèves porteurs de stéréotypes (Dutrévis, 2015). Des études expérimentales indiquent, à titre d'exemple, qu'à copie équivalente, un élève issu d'un contexte de vie plus précaire (Autin et al., 2019 ; Rangvid, 2015 ; ...) sera sous-évalué comparativement à un élève issu d'un contexte plus favorable. Dans le système éducatif de Belgique francophone, le redoublement est massivement utilisé (Lafontaine et al., 2019) dans le but de faire face aux difficultés et à l'hétérogénéité scolaires (Mons, 2007). L'analyse des croyances des enseignants atteste que les élèves redoublants sont associés à une série de stéréotypes négatifs (Dardenne et al., 2019). Dans la lignée des études docimologiques, des inquiétudes peuvent dès lors être formulées quant à l'éventualité que ces élèves puissent être victimes de biais de l'évaluation menant à une dépréciation de leur production. Afin de tester cette hypothèse, notre recherche suit un design expérimental. Cent-quarante enseignants de l'enseignement secondaire ordinaire en Fédération Wallonie-Bruxelles ont été assignés aléatoirement à différentes conditions dans lesquelles ils ont été amenés à évaluer une copie strictement identique attribuée soit à un élève redoublant, soit à un élève non redoublant en recourant ou non à une grille d'évaluation. De façon cohérente avec les résultats antérieurs, nos résultats ont mis en évidence une sous-évaluation des copies attribuées à un élève redoublant comparativement aux copies attribuées à un élève non redoublant. Bien que ces effets ne soient pas significatifs, leur magnitude n'est pas sans importance d'un point de vue pragmatique. En effet, ces effets sont inquiétants quand on les analyse à l'échelle de la scolarité. L'omniprésence des pratiques évaluatives ainsi que la répétition des biais qu'elles génèrent en défaveur des élèves stigmatisés pourraient avoir des impacts importants sur leur parcours scolaire.

4- Résumé long de la communication :

Consigne : 1000 mots maximum¹

Depuis les années 50, la docimologie met en évidence les inégalités de traitement évaluatif dont sont victimes les élèves porteurs de stéréotypes (Dutrévis, 2015). Des études expérimentales indiquent, à titre d'exemple, qu'à copie équivalente, un élève issu d'un contexte de vie plus précaire, c'est-à-dire un élève défavorisé (Autin et al., 2019),

¹ Nous vous demandons de veiller à respecter les limites imposées dans les différentes consignes sous peine de rejet de la communication.

d'origine immigrée (Bonefeld et Dickhäuser, 2018 ; Rangvid, 2015 ; Sprietsma, 2013) ou dont les parents ont un faible niveau d'éducation (Rangvid, 2015) sera sous-évalué comparativement à un élève issu d'un contexte plus favorable. Dans une même mesure, les filles performantes ont tendance à être sous-évaluées en mathématiques comparativement aux garçons qui démontrent pourtant un même niveau de compétence (Lafontaine & Monseur, 2009). Le recours à des grilles d'évaluation est considéré comme une méthode prometteuse afin de réduire ces biais de l'évaluation (Brown & Glasner, 1999) ; sans jamais, pour autant, parvenir à les faire entièrement disparaître. Dans le système éducatif belge francophone, le redoublement est massivement utilisé (Lafontaine et al., 2019) dans le but de faire face tant aux difficultés scolaires qu'à l'hétérogénéité des élèves (Mons, 2007). Or, les effets délétères du redoublement sont largement documentés (Allen et al, 2009 ; Goos et al., 2021). En dehors de son inefficacité dans sa visée originelle de résorber les difficultés des élèves, les effets nocifs du point de vue psychosocio-affectif du redoublement sont nombreux (Jimerson, 2001). De tels effets sont observés dans la majorité des contextes éducatifs permettant le maintien des élèves (Goos et al., 2013 ; Matthys, 2017 ; Troncin, 2005). En outre, l'analyse fine des croyances des enseignants atteste que, de manière générale, les élèves redoublants sont perçus négativement et, partant, sont associés à une série de stéréotypes négatifs (Fresson & Daredenne, 2019) : fainéants, paresseux, dissipés, bagarreurs, distraits, malpolis... Dans la lignée des travaux réalisés en docimologie, des inquiétudes peuvent dès lors être formulées quant à l'éventualité que ces élèves puissent être victimes de biais de l'évaluation menant à une dépréciation de leur production. Notre recherche vise ainsi à répondre à la question suivante : les performances des élèves redoublants sont-elles sous-estimées lors d'une évaluation sommative dans un contexte éducatif qui recourt massivement aux pratiques de redoublement ? Autrement dit, nous posons la question des stéréotypes portés par les élèves redoublants et de leurs effets sur le jugement porté à leur égard. Parallèlement, nous abordons la question des effets des grilles d'évaluation dans la réduction de ce biais. Notre recherche suit un design expérimental. Cent-quarante enseignants de l'enseignement secondaire ordinaire en Fédération Wallonie-Bruxelles ont été assignés aléatoirement à différentes conditions dans lesquelles ils ont été amenés à évaluer une copie strictement identique tantôt attribuée à un élève redoublant, tantôt attribuée à un élève n'ayant pas connu le redoublement et ce, en recourant ou non à une grille d'évaluation. Ainsi, quatre conditions (N=35) ont été constituées : une « copie redoublant » évaluée à l'aide d'une grille d'évaluation (groupe 1), une « copie non-redoublant » évaluée à l'aide d'une grille d'évaluation (groupe 2), une « copie redoublant » évaluée sans une grille d'évaluation (groupe 3), et enfin une « copie non-redoublant » évaluée sans l'aide d'une grille d'évaluation (groupe 4). La copie évaluée est issue d'un exercice d'écriture d'un texte argumentatif et a été rédigée par un élève de troisième année secondaire dans le cadre d'un cours de français. Afin de contrôler l'ensemble des paramètres de l'étude, il a été décidé de retranscrire en version informatique cette copie avant de la soumettre pour évaluation aux participants. L'étude a été menée de janvier à mars 2022. De façon analogue aux résultats des études antérieures portant sur d'autres biais évaluatifs, nos résultats ont permis de confirmer notre hypothèse initiale en démontrant une sous-évaluation des copies présentées comme étant réalisées par un élève redoublant (D de Cohen = -0,33) et ce, comparativement aux copies présentées comme étant réalisées par un élève non redoublant. Par ailleurs, si le recours à une grille d'évaluation permet, comme postulé, de diminuer l'impact de ce biais, la

différence entre les deux groupes reste importante, malgré l'utilisation cet outil (D de cohen = -0.21). Bien que ces effets ne soient pas significatifs, leur ampleur n'est pas sans importance d'un point de vue pragmatique. En effet, en reprenant l'argument formulé antérieurement par Autin et al. (2019), ces effets sont inquiétants quand on les analyse à l'échelle de la scolarité. Les pratiques évaluatives ponctuant le quotidien des élèves de l'enseignement obligatoire, la répétition des biais qu'elles génèrent en défaveur des élèves stigmatisés pourraient avoir des impacts considérables sur leur parcours scolaire (échec, désengagement, etc.). Ces résultats plaident en faveur d'un recours plus systématique aux méthodologies d'évaluation à la chaîne et à l'aveugle. Toutefois, nous nous devons de rester prudents dans le cadre de l'interprétation des résultats, car l'une des principales limites de notre étude réside dans la petite taille de notre échantillon ($N=140$). Or, on sait désormais que la taille des effectifs est négativement corrélée aux amplitudes de l'effet observés au sein d'une étude (Cheung & Slavin, 2016 ; Slavin & Smith, 2009) et positivement corrélée à la puissance statistique d'une étude (Barker Bausell & Li, 2002). Dès lors, la fiabilité des résultats obtenus est toujours intimement dépendante de la taille des effectifs concernés au sein d'une étude (Ajar et al., 1983). Cette étude, bien qu'elle mériterait d'être répliquée dans une perspective d'Open Science (Nosek et al., 2015) afin de garantir la validité de ses conclusions, ouvre des perspectives intéressantes en termes de recherche docimologique sur les biais touchant, plus généralement, les élèves en difficultés.

5- Bibliographie

Consigne : Indiquer les références utilisées dans les résumés en utilisant les normes APA

Nous sommes conscients que ce *template* empêche toute mise en forme, ce qui ne permet pas de respecter strictement la mise en forme selon les normes APA (ex. mise en italique)

Allen, C. S., Chen, Q., Wilson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: a meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, 31, 480-499.

<https://doi.org/10.3102/0162373709352239>

Ajar, D., Dassa, C., & Gougeon, H. (1983). L'échantillonnage et le problème de la validité externe de la recherche en éducation. *Revue des sciences de l'éducation*, 9(1), 3-21.

<https://doi.org/10.7202/900396ar>

Autin, F., Batruch, A., & Butera, F. (2019). The function of selection of assessment leads evaluators to artificially create the social class achievement gap. *Journal of Educational Psychology*, 111, 717-735. <https://doi.org/10.1037/edu0000307>

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

Bonfeld, M., Dickhäuser, O. (2018) (Biased) Grading of Students' Performance: Students' Names, Performance Level, and Implicit Attitudes. *Frontiers in psychology*, 481 (9), <https://doi.org/10.3389/fpsyg.2018.00481>

Brown, S. & Glasner, A. (Ed.) (1999). *Assesment matters in higher education : choosing and using diverse approaches*. Society for Research into Higher Education and Open University Press.

Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45, 283-292. <https://doi.org/10.3102/0013189x16656615>

Dachet, D. & Baye, A. (2021). Evidence-Based Education: The (not so simple) case of French-speaking Belgium. *ECNU Review of Education*, 4(1), 164-189.

<http://dx.doi.org/10.1177/2096531120928086>

Dutrévis, M. (2015). Menace du stéréotype et groupes ethno-raciaux minoritaires. Quel poids sur les performances des élèves ?. *Revue française de pédagogie*, 191, 61-72. Retrieved from:

<http://rfp.revues.org/4751>

Fresson, M., & Dardenne, B. (2019). Les redoublants, victims de stéréotypes dévalorisants ? In M. Crahay (eds.), *Peut-on lutter contre l'échec scolaire* (pp.231-260), De Boeck.

Goos, M., Pipa, J., & Peixoto, F. (2021). Effectiveness of grade retention: a systematic review and meta-analysis. *Educational Research Review*, 34(100401), 1-14.

<https://doi.org/10.1016/j.edurev.2021.100401>

Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade retention in the Flemish educational context: effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology*, 51, 323-347.

<https://doi.org/10.1016/j.jsp.2013.03.002>

Jimerson, S. R. (2001). Meta-analysis of grade retention research: implications for practice in the 21st century. *School Psychology Review*, 30, 420-437. Retrieved from:

<http://web.a.ebscohost.com/ehost/detail/detail?vid=0&sid=0a25bece-734a-4c84-87ea->

[e42d8069bf5b%40sessionmgr4007&bdata=JnNpdGU9ZWwhvc3QtbGl2ZQ%3d%3d#AN=5508140&db=aph](http://web.a.ebscohost.com/ehost/detail/detail?vid=0&sid=0a25bece-734a-4c84-87ea-e42d8069bf5b%40sessionmgr4007&bdata=JnNpdGU9ZWwhvc3QtbGl2ZQ%3d%3d#AN=5508140&db=aph)

Lafontaine, D., Baye, A., & Monseur, C. (2019). Redoublement, efficacité et équité : l'éclairage des enquêtes internationales. In M. Crahay (eds.), *Peut-on lutter contre l'échec scolaire ?* (4th ed.) (pp. 110-154). Louvain-la-Neuve, Belgique : de Boeck supérieur

Lafontaine, D., & Monseur, C. (2009). Les évaluations des performances en mathématiques sont-elles influencées par le sexe de l'élève ? *Mesure et Evaluation en Education*, 32 (2), 1-106.

<https://doi.org/10.7202/1024955ar>

Mathys, C., Véronneau, M.-H., & Lecoq, A. (2017). Grade Retention at the Transition to Secondary School: Using Propensity Score Matching to Identify Consequences on Psychosocial Adjustment. *The Journal of Early Adolescence*, 39(1), 97-133.

<https://doi.org/10.1177/0272431617735651>

Mons, N. (2007). *Les nouvelles politiques éducatives : la France fait-elle les bons choix ?* Presses universitaires de France.

Nosek, B. A., Alter, G., Banks, 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>

Rangvid, B. S. (2015). Systematic differences across evaluation schemes and educational choice. *Economics of Education Review*, 48, 41–55. <http://dx.doi.org/10.1016/j.econedurev.2015.05.003>

Slavin, R. E., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Education Evaluation and Policy Analysis*, 31, 500-506.

<https://doi.org/10.3102/0162373709352369>

Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, 45, 523–538. <http://dx.doi.org/10.1007/s00181-012-0609-x>

Troncin T. (2005). *Le redoublement : radiographie d'une décision à la recherche de sa légitimité*. [Doctoral dissertation, University of Bourgogne]. Retrieved from: <https://tel.archives-ouvertes.fr/tel-00140531/>