

DIFFÉRENCE DE PERFORMANCE GARÇONS-FILLES À L'ENTRÉE AU PRIMAIRE : COMPLÉMENTARITÉ D'UNE APPROCHE DIDACTIQUE ET PSYCHOSOCIALE

| GRAPIN* NADINE ET SAYAC** NATHALIE

Résumé | En France, les évaluations nationales en mathématiques montrent depuis plusieurs années que les performances des garçons sont meilleures que celles des filles à partir de l'entrée au CE1 (grade 2). Après avoir éclairé ce constat par des travaux combinant approches didactique et psycho-sociale sur les stéréotypes de genre, nous interrogerons la nature de ces évaluations et leur mise en œuvre en classe et montrons en quoi elles sont susceptibles de créer chez les filles des phénomènes de menace de stéréotype.

Mots-clés : stéréotypes de genre, inégalités garçons-filles, validité, didactique des mathématiques, calcul mental

Abstract | In France, national mathematics assessments have shown for several years that boys perform better than girls from the start of CE1 (grade 2). After shedding light on this observation by combining didactic and psycho-social approaches to gender stereotypes, we will examine the nature of these assessments and their implementation in the classroom and show how they are likely to create stereotype threat phenomena in girls.

Keywords: gender stereotypes, gender inequality, validity, mathematics didactics, mental arithmetic

I. INTRODUCTION

Depuis la rentrée 2018, les élèves français sont soumis à des évaluations nationales standardisées en mathématiques à l'entrée au CP (cours préparatoire, grade 1), au CE1 (cours élémentaire 1, grade 2) et 6^e (grade 6). Les résultats aux évaluations de septembre 2023 indiquent :

« En mathématiques, les résultats des garçons sont légèrement inférieurs à ceux des filles en début de CP, à l'exception de la comparaison de nombres et de l'exercice qui consiste à placer un nombre sur une ligne graduée. On assiste à un renversement complet à l'entrée en CE1. On constate alors des écarts importants de réussite, au détriment des filles, en 2023 comme lors des évaluations précédentes. » (Andreu et al., 2023, p. 41)

En tant que didacticiennes des mathématiques préoccupées par les inégalités scolaires dans notre discipline, il nous semble nécessaire de dépasser ce constat et de chercher à l'expliquer en mobilisant des outils propres à notre champ scientifique, mais aussi des travaux en psychologie sociale, notamment ceux relatifs aux stéréotypes de genre. Pour ce faire, nous formulons notre problématique en appui sur des éléments théoriques issus de ces différents champs scientifiques, nous présentons ensuite notre expérimentation et les résultats obtenus et concluons avec quelques pistes pour la recherche et la formation des enseignant·es.

* Laboratoire de Didactique André Revuz – France – nadine.grapin@u-pec.fr

** Laboratoire de Didactique André Revuz – France – nathalie.sayac@univ-rouen.fr

II. PROBLÉMATISATION

1. *Didactique des mathématiques et écart de performance selon le sexe*

Si les résultats aux évaluations à grandes échelles (nationales ou internationales) conduisent à des comparaisons en termes de performances selon le genre, les écarts sont bien souvent éclairés par des éléments d'ordre psycho-social : anxiété, confiance en soi, sentiment d'efficacité personnelle, etc. Peu d'éléments relatifs aux savoirs en jeu en lien avec une analyse *a priori* de la tâche et des erreurs des élèves sont alors proposés. Réciproquement, et à notre connaissance, peu de travaux en didactique des mathématiques prennent en compte la dimension du genre dans leurs questionnements, que ce soit pour documenter les pratiques des enseignants et/ou pour éclairer l'activité de l'élève en contexte d'évaluation. Nous retenons cependant les travaux de Che, Wiegert et Threlkeld (2012) qui montrent des différences dans les types de procédures utilisées par les garçons et les filles lors de la résolution de problème de proportionnalité au grade 3 et ceux de Roditi et Salles (2015) qui, après une analyse des niveaux de mise en fonctionnement des connaissances mathématiques (Robert, 1998) dans les items du PISA 2012, montrent que « les filles sont d'autant plus en difficulté par rapport aux garçons que le niveau requis de mise en fonctionnement des connaissances est un niveau exigeant » (Roditi et Salles, 2015, p. 252). Enfin, nous avons montré dans une recherche antérieure (Sayac et Grapin, 2016) qu'en fin d'école (grade 5), en France, les garçons attribuent un degré de certitude plus élevé que les filles à leur réponse, qu'elle soit correcte ou fausse. Les garçons auraient donc des connaissances plus assurées que les filles, mais aussi, une ignorance ignorée plus fréquente, selon ces indicateurs.

Même s'ils sont peu nombreux, les résultats produits par ces travaux montrent qu'une approche didactique, à partir d'une analyse des tâches et/ou des procédures des élèves permet d'éclairer de manière complémentaire les différences de performance entre filles et garçons.

2. *Stéréotype de sexe et performance aux évaluations*

Il est reconnu que les filles sont plus sensibles à la pression évaluative (Marsh, Martin et Cheng, 2008 ; Genoud, Ruiz et Gurtner, 2009) et plus anxieuses que les garçons vis-à-vis de leur travail scolaire (Devine et al., 2012 ; Chen et al., 2013). Même si les travaux ayant attesté de ces différences ont concerné des élèves plus âgés que les élèves de CP et CE1 (grades 1 et 2), on peut tout à fait concevoir que passer des évaluations en mathématiques et en français, la semaine qui suit la rentrée scolaire engendre une certaine pression et de l'anxiété chez des petites filles de 6-7 ans (Sayac, 2023). L'observation d'une quarantaine de passations des évaluations Repères en mathématiques dans des classes de CP en janvier 2023, nous a permis de relever des éléments susceptibles de provoquer cette pression évaluative, comme l'usage du chronomètre pour certains exercices ou le placement d'un grand livre pour séparer l'espace de deux élèves partageant la même table. Par ailleurs, ces évaluations ont un caractère très institutionnel, avec un protocole et des consignes très strictement précisés dans les livrets d'accompagnement (par exemple, MENJS 2023b). Les professeur·es que nous avons observé·es les respectent très majoritairement, mais des différences de ton, de discours et d'étayage dues à la façon dont ils/elles s'approprient ces consignes et à leur personnalité peuvent potentiellement générer de l'anxiété auprès de certains élèves et plus particulièrement des filles.

Marie-Christine Tocsek (2005) a montré combien les stéréotypes de sexe pouvaient influencer les performances scolaires des élèves, notamment à travers la nature des consignes données avant la réalisation d'exercices, lors des évaluations nationales, et qu'ils pouvaient avoir une incidence sur leurs performances en fonction de leur genre en mobilisant le phénomène de menace du stéréotype (Steele et Aronson, 1995 ; Régner, Steele et Huguet, 2014). En croisant les données issues des évaluations

Repères administrées à plus de 2,5 millions d'élèves scolarisé·es en France entre 2018 et 2022 avec des informations sur le contexte familial, les enseignant·es et les caractéristiques générales des écoles (publique, privée, en réseau d'éducation prioritaire — REP — ou à pédagogie alternative), les chercheur·es de l'Institut des Politiques Publiques qui souhaitent élucider la raison de l'émergence des écarts de performance en mathématiques entre les filles et les garçons observés en CP et en CE1 ont conclu qu'aucun de ces facteurs ne pouvait expliquer à lui seul ces écarts (Breda et al., 2024). Il et elles ont donc conclu que ces écarts devaient être dus à « une dynamique diffuse dans la société », sous-entendant que les stéréotypes de sexe étaient le facteur explicatif le plus plausible.

3. *Problématique*

Si les résultats nationaux montrent que les filles réussissent dans l'ensemble moins bien que les garçons aux évaluations à l'entrée au CE1 (grade 2), l'écart est particulièrement marqué sur des compétences de calcul en ligne : pour « additionner », il y aurait 15 % de garçons de plus de filles qui « n'auraient pas de difficultés identifiées » sur cette compétence et 9 % de plus sur la compétence « soustraire » (Andreu et al., 2023). Ces résultats, comme tous ceux produits par la DEPP¹, sont le fruit d'un traitement strictement statistique de données codées à partir des productions des élèves et de seuils prédéfinis pour concevoir les groupes différents groupes auxquels vont être affectés les élèves². Ils ne prennent en compte les types d'erreurs commises par l'élève ou leurs procédures, ni le contexte dans lequel sont passées ces évaluations, ni les effets qu'elles peuvent produire sur les élèves. Or, ces deux éléments (contexte et effets) ne sont pas sans incidence sur les performances des élèves comme de nombreux travaux le présument.

Nous avons donc souhaité éclairer ce constat par une étude combinant approche didactique et psycho-sociale afin de chercher les origines de ces écarts de performance. Comme nous l'avons montré dans les deux paragraphes précédents, chacune de ces approches conduit à des résultats intéressants et il nous semble nécessaire de les mener de façon complémentaire. Nous avons choisi de nous centrer sur une tâche de calcul en ligne issue de l'évaluation nationale menée à l'entrée au CE1 (grade 2) puisque c'est à ce niveau et sur cette tâche que les écarts apparaissent comme étant les plus importants. À partir de cette tâche et dans une approche didactique, il s'agit de déterminer si des erreurs types ou des non-réponses pourraient expliquer ces écarts de performance. Puisque nous considérons que filles et garçons disposent des mêmes capacités, il s'agit également de rechercher les sources de ces écarts dans des origines psycho-sociales, en particulier celles liées à la pression exercée par le test et à la confiance qu'accordent les élèves à leurs réponses.

III. MÉTHODOLOGIE

Notre méthodologie combine deux études complémentaires : la première est quantitative basée uniquement sur l'analyse des réponses des élèves, la seconde est qualitative et prend en compte la réponse de l'élève et d'autres traces de son activité (gestes, temps de réponses, etc.), mais aussi des éléments de contexte et son ressenti face à la tâche. La méthodologie, comme les résultats, sont présentés ici de manière synthétique, le lecteur pourra se référer à Grapin et Sayac (soumis) pour davantage de précisions.

¹ Direction de l'évaluation, de la prospective et de la performance.

² Trois groupes d'élèves sont identifiés : « ceux pour lesquels on peut considérer, dès le début de l'année, qu'ils sont en difficulté [...] ; ceux dont les acquis sont fragiles [...] ; ceux pour lesquels il n'y a pas de difficultés identifiées » (Andreu et al., 2023, p. 24).

1. Tâche retenue pour l'étude

La tâche retenue pour l'étude est celle correspondante à l'exercice 7 de l'évaluation nationale à l'entrée au CE1 en 2023 (Figure 1). L'élève dispose de 7 minutes pour entourer, pour chacun des calculs, une bonne réponse parmi les six proposées. Le premier calcul ($5+3$) figurant sur la première ligne est traité collectivement et vise à faire comprendre aux élèves la tâche qui leur est prescrite. Chaque élève avance individuellement, à son rythme, pour traiter l'ensemble des 15 calculs ; aucune indication ou aide n'est apportée par l'enseignant·e durant ce temps.

$5+3=$	53	9	8	2	5	3
$10+8=$	19	108	10	8	18	2
$9-4=$	94	9	13	5	6	4
$9+9=$	0	99	17	18	90	9
$50+50=$	0	50	79	99	100	5050
$10-2=$	10	8	12	102	9	2
$15-5=$	11	155	10	15	20	5
$20+30=$	20	51	2030	10	50	30
$18+4=$	4	184	18	14	22	21
$19-2=$	19	21	192	18	17	2
$56+10=$	66	56	45	5610	46	10
$21+53=$	2153	32	74	21	53	75
$8+46=$	54	38	8	846	46	55
$50+10+8=$	50108	60	18	68	67	58
$35-15=$	15	3515	21	50	20	35
$33-5=$	28	33	38	5	335	29

(Source : MENJS, 2022b)

Figure 1 – Reproduction des pages 21 et 22 du cahier d'évaluation de l'élève à l'entrée au CE1

L'analyse de la tâche montre que les six choix de réponse sont toujours construits de la même façon (ou presque), mais proposés dans un ordre différent avec : la bonne réponse, la bonne réponse +1 ou la bonne réponse -1, le nombre constitué en accolant les termes de la somme ou de la différence, le résultat de l'opération inverse, chacun des deux termes du calcul.

Les procédures pour traiter les calculs dépendent de la taille et de la nature des nombres (Butlen, 2015 ; Grapin et al., 2023). Par exemple, des procédures par comptage (sur-comptage, décomptage) en appui sur la numération orale et avec l'aide ou non des doigts sont possibles notamment pour les calculs dont un des termes est inférieur ou égal à 10 ; un appui sur la numération écrite chiffrée est également possible en particulier pour les calculs dont un des termes (ou les deux) est un nombre entier de dizaines. L'élève peut également mobiliser des procédures passant par une décomposition additive d'un des termes (ou des deux) permettant d'aller à la dizaine supérieure (par exemple : $18 + 4 = 18 + 2 + 2 = 20 + 2 = 22$)

2. Recueil et traitement des données

L'étude quantitative a été conduite à partir des réponses données lors de l'évaluation nationale en septembre 2023 par 265 élèves (128 garçons – 137 filles) de CE1 d'une même circonscription

parisienne. Chacune des réponses a été traitée selon les différentes catégories identifiées lors de l'analyse de la tâche.

Pour l'étude qualitative, des passations individuelles ont été menées par les deux chercheuses auprès de 37 élèves (20 filles / 17 garçons) de fin de CP en juin 2023 résolvant cet exercice dans des conditions similaires à celles des évaluations nationales. A la suite de la réalisation de cet exercice, il a été demandé à l'élève ce qu'il ou elle avait pensé de cet exercice ; il s'agissait avec une question ouverte, de savoir quel était le ressenti de l'élève, s'il ou elle évoquait le caractère complexe ou non de la tâche, certaines des difficultés rencontrées, etc. Nous lui avons également demandé la certitude qu'il ou elle accordait à l'ensemble de ses réponses sur le test. Pour ce faire, nous avons repris de nos travaux précédents (Sayac et Grapin, 2014) une échelle de Likert avec 4 degrés de certitude (pas sûr du tout, un peu sûr, sûr, très très sûr), mais nous l'avons adaptée avec des smileys puisque d'après les recommandations de Houssa et Nader-Grosbois (2018, p. 16), « lorsqu'on utilise une échelle de Likert avec des jeunes enfants, les options de réponse doivent être concrètes et créatives et doivent les aider à comprendre la tâche ».

L'entretien a ensuite porté sur la façon dont l'élève a procédé pour fournir sa réponse à quatre calculs différents : $9 - 4$; $20 + 30$; $18 + 4$; $19 - 2$. Cette étude ayant été réalisée en amont de la première, le choix des calculs ne repose pas sur les observations faites à la suite de l'étude quantitative. Ils ont été retenus parce qu'ils combinent additions et soustractions et que des procédures différentes peuvent être mobilisées pour donner la réponse ; cela permettra ainsi d'étudier si des différences de procédures et/ou d'explications des procédures apparaissent entre garçons et filles. Par exemple, nous avons retenu $9 - 4$ puisqu'il s'agit de la première soustraction du test. Nous supposons que le risque de confusion avec une addition est plus important sur ce calcul et souhaitons étudier si ce type de confusion est plus fréquent chez les garçons ou les filles. En termes de procédures, les élèves peuvent connaître le résultat et le donner rapidement (fait numérique) ou le déterminer en passant par décomptage à partir de 9 ou encore en sachant que $4 + 5 = 9$ (en appui ou non sur les doigts).

Durant la passation, nous avons particulièrement observé si l'élève utilisait ou non ses doigts de façon visible pour donner la réponse et avons pris en note les explications qu'il ou elle nous donnait quant aux procédures qu'il ou elle utilisait pour donner sa réponse ; des enregistrements audio et vidéo de certaines passations ont également été réalisés.

IV. RÉSULTATS

1. Résultats de l'étude quantitative

Les résultats que nous obtenons à l'échelle de la circonscription dans laquelle l'étude a été menée rejoignent ceux obtenus à l'échelle nationale, à savoir que la performance des garçons à cet exercice est meilleure que celle des filles puisque le score moyen des garçons est de 11 réponses justes (sur 15) alors qu'il est de 9 réponses justes (sur 15) pour les filles.

Nous constatons également une augmentation du taux de non-réponse au fur et à mesure de l'avancée dans les calculs avec systématiquement un nombre plus important de non-réponse chez les filles et un écart entre le pourcentage de non-réponse par calcul entre garçons et filles qui se stabilise autour de 13 % à partir du onzième calcul (Figure 2). En moyenne, le nombre de non-réponse pour les garçons est de 1,1 et de 2,1 pour les filles. Cet écart sur le nombre de non-réponses pourrait être un élément explicatif des écarts de performance entre garçons et filles, mais ces écarts perdurent calcul par calcul si on ne prend en compte que les réponses données, c'est-à-dire en écartant les non-réponses.

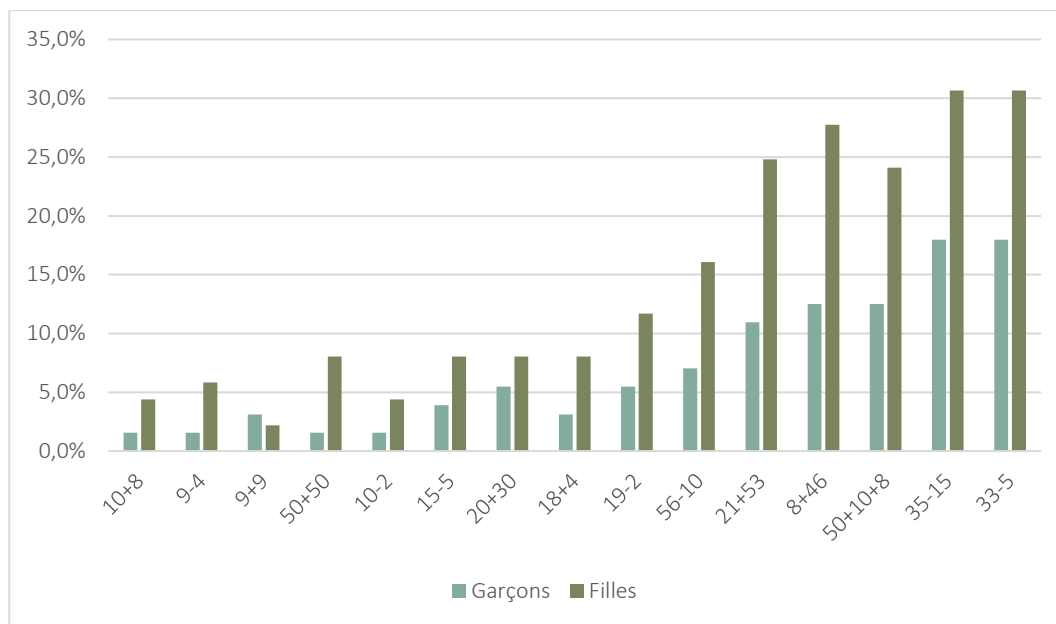


Figure 2 – Pourcentage de non-réponse selon le calcul en fonction du sexe.

Intéressons-nous désormais aux choix de réponses et aux erreurs commises. Les écarts entre garçons et filles sur les types d'erreurs commises sont pour la plupart inférieur à 7 % sauf pour certains calculs. Le nombre composé des deux termes accolés est une erreur peu fréquente (moins de 5 % des réponses) ; l'écart entre garçons et filles est particulièrement marqué pour $50 + 10 + 8$ (10,2 % des garçons et 17,5 % des filles commettent cette erreur).

Les bonnes réponses à $+ ou - 1$ près représentent plus de 10 % des erreurs sur les calculs trois calculs : $56-10$ (10,2 % des erreurs) ; $8+46$ (11,3 %) ; $33-5$ (10,6 %) ; l'écart entre garçons et filles étant particulièrement marqué pour $8+46$ (14,8 % des garçons et 8 % des filles commettent cette erreur).

Enfin, la confusion addition / soustraction représente environ 10 % des erreurs pour chacune des soustractions (elle est inférieure à 4 % pour les additions). La différence entre garçons et filles est particulièrement marquée pour ce type d'erreur sur les trois premières soustractions où la proportion des filles commettant cette erreur est supérieure de plus de 7 % à celle des garçons. Les filles ont davantage tendance à choisir le résultat de l'opération inverse sur les premières soustractions ($9 - 4$; $10 - 2$) et sur la dernière ($33 - 5$) comme le montre l'étude de la répartition des réponses selon les types de distracteurs. L'écart se réduit sur les dernières soustractions, voire s'inverse, ce qui est à relativiser puisque le score de non-réponse est particulièrement élevé pour les filles sur les trois derniers calculs.

Ces constats peuvent, en partie, expliquer l'écart de score entre les filles et les garçons souligné précédemment, mais ils engendrent d'autres interrogations : pourquoi les filles sont-elles plus nombreuses que les garçons à produire des non-réponses, notamment concentrées sur les derniers calculs ? pourquoi choisissent-elles de manière plus marquée l'opération inverse de celle attendue ?

Ces différences constatées entre les productions des filles et des garçons ne peuvent s'expliquer d'un point de vue cognitif, mais la piste d'une pression évaluative plus grandement ressentie par les filles ou celle de l'effet de la menace du stéréotype restent à explorer.

2. Résultats de l'enquête qualitative

Les scores que nous obtenons sur l'échantillon de 37 élèves en fin de CP diffèrent de ceux de CE1 à l'échelle nationale ou à celle de la circonscription. Non seulement l'écart de score entre garçons et filles est plus réduit que celui à l'échelle de la circonscription, mais il est à la faveur des filles : leur score moyen étant de 8,3 bonnes réponses sur 15 et celui des garçons de 7,5. En revanche, comme pour l'enquête précédente, le nombre de non-réponse moyen reste plus élevé chez les filles que chez les garçons et ce dans les mêmes proportions (2,2 *vs* 1,2).

Concernant les procédures, nous remarquons que les élèves utilisent régulièrement leurs doigts de façon visible lorsqu'ils ou elles résolvent la tâche durant le temps imparti pour répondre : les filles les utilisent davantage, en moyenne pour 4 calculs et les garçons pour 3. Même si nous ne le développons pas ici, ce résultat interroge la validité de cette tâche (Grapin et al., 2022) puisque l'objectif qui lui est assigné selon le guide de l'enseignant (MENJ, 2023a) est d'évaluer des connaissances en calcul en ligne des élèves. Les explications données par les élèves lors des échanges avec les chercheuses montrent que si les filles et les garçons utilisent dans une proportion similaire une procédure de comptage, les filles semblent davantage mentionner explicitement l'utilisation des doigts pour expliquer la façon dont elles procèdent. La différence est particulièrement marquée pour $18 + 4$ où par ailleurs aucune fille ne propose une décomposition d'un des deux nombres pour atteindre une dizaine supérieure, alors qu'ils sont trois garçons à le faire.

Les observations réalisées durant le temps de réponse ne permettent pas d'apporter d'élément explicatif sur la présence d'un taux de non-réponse plus important chez les filles. Sur les quatre calculs retenus pour l'entretien, le nombre de non-réponses est faible : 6 non-réponses pour $20 + 30$ (5 filles et 1 garçon) et 1 non-réponse (une fille à chaque fois) pour les trois autres. Lors de l'entretien, 4 filles sur les 5 qui n'avaient pas répondu ont été capables, lors de l'entretien, de choisir une réponse et de proposer une procédure pour produire une réponse. On pourrait alors supposer qu'elles ont préféré ne pas répondre plutôt que de fournir une réponse dont elles n'étaient pas sûres.

Les entretiens menés à la suite de la passation ne permettent pas de montrer des différences notoires dans le ressenti des élèves face à la difficulté de l'exercice. Le degré de certitude moyen accordé aux réponses sur l'ensemble du test est légèrement supérieur chez les filles (3,7/5 chez les filles *vs* 3,2/5 chez les garçons). Signalons cependant que spontanément, la durée de l'exercice n'a été évoquée que par deux garçons (« *c'était dur, mais quand même, j'ai arrivé à faire le minuteur* », « *mon record dans tout ça c'était 3'50 si je ne l'avais pas battu, j'aurais pas aimé* ») et seules trois filles ont fait référence à ce qu'elles avaient travaillé au préalable avec leur enseignante pour expliquer leur ressenti face à la difficulté de certains calculs (« *On n'a pas appris avec la maitresse* » ; « *y'en a qu'on avait appris avec la maitresse* »).

V. DISCUSSION ET CONCLUSION

Nos résultats montrent, si besoin en était, toute la complexité du phénomène à investiguer. L'approche quantitative a permis de renseigner le type d'erreur commis et surtout de faire apparaître que les taux de non-réponses étaient plus importants chez les filles. L'étude qualitative, menée sur un échantillon restreint d'élèves permet quant à elle de formuler quelques hypothèses quant à l'origine des écarts de performance. D'une part, ces écarts semblent ne pas être présents lorsque les conditions de passation ne sont pas identiques à celles des évaluations nationales et que la pression évaluative est moindre : ce résultat rejoint celui de Sayac (2023) sur l'anxiété générée par les conditions de passation de telles évaluation chez les petites filles, au début de l'école primaire. D'autre part, l'évocation plus marquée de l'utilisation des doigts chez les filles lorsqu'elles expliquent les procédures utilisées pour donner le résultat d'un calcul pourrait être interprétée comme étant issue d'une confiance plus grande

dans cette procédure qu'elles utilisent depuis la maternelle ; elles ne prendraient pas le risque d'exposer une procédure qui est plus récente pour elles et qu'elles pensent peut-être moins bien maîtriser. La répétition d'une telle expérimentation serait nécessaire pour investiguer ces hypothèses.

Plus largement, c'est également en menant des recherches portant conjointement sur les pratiques de l'enseignant·e (notamment ses pratiques évaluatives), l'activité en mathématiques de l'élève lors d'épisodes évaluatifs et leurs affects mathématiques que nous pourrions avoir une meilleure compréhension de la façon dont se construisent, s'intensifient ou se réduisent les inégalités de genre en mathématiques en classe. Enfin, former les enseignant·es, mais aussi les formateurs et formatrices, à prendre conscience des inégalités de genre qui peuvent advenir dans les classes et qui se manifestent lors d'épisodes évaluatifs, les amener à « chausser des lunettes de genre » (Plateau, 2011), est une orientation nécessaire. Le protocole FÉM (formation à l'égalité en mathématiques), développé par Sayac (soumis, EMF 2025) est un exemple de formation s'appuyant sur des résultats similaires à ceux que nous avons proposés et mobilisant des outils élaborés dans le cadre de recherches antérieures (degrés de certitude, analyse des tâches, degré de certitudes, composition des groupes, type d'interactions, etc.) afin de sensibiliser les enseignants aux inégalités filles-garçons et à la façon dont elles se manifestent en classe.

RÉFÉRENCES

- Andreu, S., Conceição, P., Desclaux, J., Eteve, Y., Gill-Sotty, C., Loi, M., Laskowski, C., Le Breton, S., Neirac, L., Persem, E., Raffy, G., Rocher, T., Rue, G., Thumerelle, J. et Vourc'h, R. (2023, novembre). *Évaluation 2023. Repères CP, CE1 : premiers résultats* (Document de travail n° 2023-E05). DEPP. <https://www.education.gouv.fr/evaluations-2023-reperes-cp-ce1-premiers-resultats-379863>
- Breda, T., Sultan Parraud, J. et Touitou, L. (2024). Le décrochage des filles en mathématiques dès le CP : une dynamique diffuse dans la société. *Notes de l'Institut des Politiques Publiques*, 101. <https://www.ipp.eu/publication/le-decrochage-des-filles-en-mathematiques-des-le-cp-une-dynamique-diffuse-dans-la-societe/>
- Butlen, D. (2005). *Le calcul mental entre sens et technique*. Presses universitaires de Franche Comté.
- Che, M., Wiegert, E. et Threlkeld, K. (2012). Problem solving strategies of girls and boys in single-sex mathematics classrooms. *Educational Studies in Mathematics*, 79(2), 311-326.
- Chen, H. Y., Chen, M. F., Lee, Y. S., Chen, H. P. et Keith, T. Z. (2013). Gender reality regarding mathematic outcomes of students aged 9 to 15 years in Taiwan. *Learning and Individual Differences*, 26, 55-63. <https://doi.org/10.1016/j.lindif.2013.04.009>
- Devine, A., Fawcett, K., Szűcs, D. et Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and brain functions*, 8, Article 33.
- Genoud, P. A., Ruiz, G. et Gurtner, J. L. (2009). Évolution de la motivation scolaire des adolescents. Différences selon la filière et le genre. *Schweizerische Zeitschrift für Bildungswissenschaften*, 31(2), 377-395.
- Grabin N., Vantourout, M. et Grugeon-Allys B. (2022). S'assurer de la validité des évaluations en milieu scolaire. *Recherches en didactique*, 33(1), 59-83. <https://doi.org/10.3917/rdid1.033.0059>

- Grapin, N., Chenevotot, F., Ledan, L., Beylot, D., Mounier, E. et Blanchouin, A. (2022). Étude exploratoire de procédures d'élèves de 7-8 ans en calcul mental aditif. *Revue Math-École*, 238, 29-40. <https://doi.org/10.26034/vd.rm.2022.3570>
- Grapin, N. et Sayac, N. (soumis). Etudier les différences de performance garçons-filles en mathématiques à l'entrée au primaire : complémentarité d'approches didactique et psycho-sociale. *Education et didactique*.
- Houssa, M. et Nader-Grosbois, N. (2018). *Réalisation d'enquêtes quantitatives par questionnaire auprès de jeunes enfants (5-8ans). Étape 1 : rapport final de recherche*. Institut des sciences psychologiques (IPSY, UCL) ; Université catholique de Louvain. https://oejaj.cfwb.be/fileadmin/sites/oejaj/uploads/Hors_PublicationsTravaux/Documents/CIDE/Participation_des_enfants_en_FWB/outils_pratiques/rapport_final_interroge_enfant.pdf
- Marsh, H. W., Martin, A. J. et Cheng, J. H. (2008). A multilevel perspective on gender in classroom motivation and climate: Potential benefits of male teachers for boys? *Journal of Educational Psychology*, 100(1), 78-95. <https://dx.doi.org/10.1037/0022-0663.100.1.78>
- Ministère de l'Éducation nationale, de la Jeunesse et des Sports [MENJS]. (2023a). *Repères CE1 2023. Guide pour le professeur*. <https://eduscol.education.fr/document/51744/download>
- Ministère de l'Éducation nationale, de la Jeunesse et des Sports [MENJS]. (2023b). *Repères CE1 2023. Cahier de l'élève*. <https://eduscol.education.fr/document/51723/download>
- Plateau, N. (2011). Chaussons nos lunettes de genre. *Cahiers pédagogiques*, (487). <https://www.cahiers-pedagogiques.com/chaussons-nos-lunettes-de-genre/>
- Régner, I., Steele, J. R. et Huguet, P. (2014). Stereotype threat in children: Past and present. *Revue internationale de psychologie sociale*, 27(3), 5-12.
- Robert, A. (1998). Outil d'analyse des contenus mathématiques à enseigner au lycée et à l'université. *Recherches en Didactique des Mathématiques*, 18(2), 139-190.
- Roditi, E. et Salles, F. (2012). Nouvelles analyses de l'enquête PISA 2012 en mathématiques : un autre regard sur les résultats. *Éducation & formations*, 86-87(11), 235-255.
- Sayac, N. et Grapin, N. (2014). Évaluer par QCM en fin d'école : stratégies et degrés de certitude. *Annales de Didactique et de Sciences Cognitives*, 19, 169-199. <https://doi.org/10.4000/128a7>
- Sayac, N. et Grapin, N. (2016). Stratégies et degrés de certitude des filles et des garçons en mathématiques. *Repères IREM*, (104), 43-57.
- Sayac, N. (2023). Comment expliquer les écarts de performance entre les filles et les garçons en mathématiques aux évaluations nationales de CP ? *Évaluer. Journal international de recherche en éducation et formation*, 9(3). <https://doi.org/10.48782/e-jiref-9-3-31>
- Sayac N. (soumis). Expérimentation d'un protocole de formation à l'égalité en mathématiques (FÉM), de la maternelle au lycée. *Recherche et formation*.
- Steele, C. M. et Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811. <https://doi.org/10.1037/0022-3514.69.5.797>
- Toczek, M. C. (2005). Réduire les différences de performances selon le genre lors des évaluations institutionnelles, est-ce possible ? Une première étude expérimentale... *L'orientation scolaire et professionnelle*, 34(4), 439-460.