

BILAN DU GROUPE DE TRAVAIL 11

ÉVALUATIONS DANS L'ENSEIGNEMENT ET L'APPRENTISSAGE DES MATHÉMATIQUES

HOROKS* JULIE, OURAHAY** MUSTAPHA, RUMINOT*** CAROLINA ET CHANUDET****

MAUD

L'évaluation des apprentissages est devenue un aspect crucial de l'éducation contemporaine, mettant l'accent sur le soutien continu de l'apprentissage des élèves et sur l'adaptation des méthodes d'enseignement pour répondre à leurs besoins. L'implantation de nouvelles approches pédagogiques et d'évaluations s'inscrivant dans un processus dynamique visant à améliorer l'apprentissage tout au long du processus scolaire, est la base d'un modèle souhaité dans plusieurs systèmes éducatifs. Ce modèle complexe a commencé à se concrétiser d'un point de vue à la fois théorique et pratique (Black, 2013 ; Earl, 2012 ; Allal, 2008). Les colloques EMF ont enrichi ces travaux, explorant divers contextes éducatifs, abordant les évaluations internationales, les transitions scolaires, et questionnant les pratiques évaluatives enseignantes avec une attention portée aux spécificités culturelles et scientifiques. En 2022, le focus s'est porté sur l'évaluation située, les pratiques enseignantes, la formation et l'usage des outils numériques pour l'évaluation en mathématiques. Le GT11 en 2025 s'inscrit dans cette continuité thématique, mettant l'accent sur des enjeux et des opportunités de l'éducation mathématique face à un monde en accélération, à travers la question directrice : *peut-on envisager des évaluations à visage humain, prenant en compte la véritable nature des mathématiques ?* Nous interrogeons les effets des évaluations sur l'élève, des liens entre évaluation, enseignement des mathématiques et technologies au service de l'évaluation des apprentissages, selon les trois axes de travail développés ci-dessous :

- *Une évaluation au service de l'apprentissage et la place des élèves.* L'évaluation ne se limite plus à certifier les acquis : elle devient un levier de soutien à l'apprentissage, engageant l'élève de manière active et réflexive dans son propre parcours.
- *Des pratiques enseignantes en mutation.* Le GT11 s'intéresse à la manière dont les enseignants conçoivent et mettent en œuvre l'évaluation en mathématiques, en lien avec la didactique et les spécificités disciplinaires.
- *Technologies et innovations évaluatives.* Avec l'essor du numérique et de l'IA, de nouvelles formes d'évaluation apparaissent. Le GT11 explore les opportunités et défis liés à leur intégration dans les pratiques éducatives.

Onze contributions ont fait l'objet de discussion dans le groupe (dix communications et un poster) qui a accueilli jusqu'à 20 participants, de différents pays (Belgique, Suisse, Canada, France, Maroc) et de statut variés (chercheuses, formateurs/trices, enseignant.es). Ces contributions ont approfondi principalement deux des trois axes initialement proposés. Les questions traitées dans ces présentations,

* Coresponsable du groupe de travail n° 11 – France – julie.horoks@u-pec.fr

** Coresponsable du groupe de travail n° 11 – Maroc – mu.ourahay@uca.ac.ma

*** Coresponsable du groupe de travail n° 11 – Canada – cruminot@uottawa.ca

**** Correspondante du comité scientifique – Suisse – maud.chanudet@unige.ch

avec des études centrées sur l'évaluation sommative, les évaluations à grande échelle, ainsi que l'évaluation formative en formation initiale tournaient principalement autour des points suivants :

- Évaluer en résolution de problèmes (2 communications). Analyse des pratiques évaluatives en résolution de problèmes pour mieux réguler et dévoluer les apprentissages, notamment chez les élèves en difficulté.
- Analyse des évaluations à grande échelle (4 communications). Étude critique des dispositifs nationaux et internationaux (PISA, tests ministériels) pour questionner leur validité, leur conception et leurs enjeux didactiques.
- Représentations, pratiques et formation à l'évaluation (3 communications). Exploration des conceptions et difficultés des enseignants (débutants ou en formation) pour mieux former à une évaluation mathématique pertinente.
- Genre et évaluation (2 communications). Mise en évidence de l'impact des stéréotypes de genre sur les performances et pratiques évaluatives différenciées dès l'école primaire.

Nous avons choisi d'organiser nos sessions avec des temps de discussion après les présentations, avec un.e réactant.e assigné.e pour chaque communication. La suite de ce bilan porte sur une analyse croisée dans chaque des quatre thèmes et illustre les apports des différentes communications.

Évaluer en résolution de problèmes (2 Communications). Dans leurs travaux, **Beylot, Blanchouin et Grapin** étudient, au CP, l'activité évaluative d'une enseignante expérimentée face à un élève en grande difficulté pour résoudre des problèmes arithmétiques verbaux. Leur étude met en lumière comment ses gestes évaluatifs — questions, rétroactions et soutien matériel — visent à équilibrer aide et autonomie, mais révèlent aussi des limites quand l'enseignante prend trop en charge la tâche. De leur côté, **Chanudet et Favier** s'intéressent au secondaire I à Genève et explorent comment les enseignants peuvent réguler l'activité de recherche des élèves grâce à des discussions évaluatives qui ciblent à la fois leurs actions et leurs représentations mentales. Leur étude de cas montre que les élèves restent parfois bloqués dans de mauvaises représentations (ici la proportionnalité) malgré des contre-exemples, soulignant l'importance d'intervenir sur la compréhension du problème et les heuristiques mobilisées. Ensemble, ces recherches rappellent que l'évaluation formative, qu'elle passe par des gestes ou des régulations, doit ajuster en permanence guidance, questionnement et autonomie pour réellement soutenir l'apprentissage.

Analyse des évaluations à grande échelle (4 communications). Les communications de Chenevotot, Grapin, Grugeon-Allys, Mounier et Pilet, de Demonty, Bilodeau, Barallobres et Roiné, ainsi que de Bénard, Tremblay et Delobbe interrogent toutes, à leur manière, les tensions et enjeux de validité dans l'évaluation des mathématiques, du diagnostic précoce jusqu'à la certification finale. Le projet DiToM (**Chenevotot et al.**) développe un outil européen de diagnostic en mathématiques pour dépister très tôt les difficultés des élèves en numération, calcul et algèbre, posant la question de la validité didactique et de l'adaptation aux contextes nationaux. **Demonty**, de son côté, interroge les écarts entre les compétences sollicitées par PISA 2022 et celles réellement travaillées dans le curriculum belge francophone, soulignant la place de la résolution de problèmes ancrés dans la vie réelle et l'enjeu du numérique. **Bilodeau et ses collègues** adoptent une perspective anthropologique pour analyser l'activité de correction des épreuves au Québec : ils montrent que cette correction engage un travail interprétatif complexe, entre standardisation et reconnaissance de la diversité des raisonnements des élèves. Enfin, **Bénard, Tremblay et Delobbe** questionnent la validité des épreuves certificatives au Québec, notamment leur capacité à évaluer réellement le raisonnement mathématique, à partir d'une analyse croisée avec les cadres internationaux (PISA, TIMSS) et un modèle conceptuel québécois. Ensemble, ces travaux soulignent combien l'évaluation en mathématiques, qu'elle soit diagnostique,

formative ou certificative, constitue un terrain sensible où se jouent la reconnaissance des compétences complexes, la justice scolaire et la crédibilité des dispositifs institutionnels.

Représentations, pratiques et formation à l'évaluation (3 communications). Dans les recherches de Ruminot, de Horoks, ou de Mortaqi et Ourahay, la question de l'évaluation des apprentissages en mathématiques dans la formation des enseignant.es du primaire est explorée de différentes manières. **Ruminot** met en évidence que les futur.es enseignant.es reconnaissent l'importance de l'évaluation formative pour ajuster l'enseignement, mais restent marqué.es par une conception encore centrée sur la mesure sommative, soulignant ainsi la tension entre diagnostic et certification. **Horoks** s'intéresse aux représentations initiales des enseignant.es débutant.es et montre que celles-ci, influencées par leurs propres souvenirs d'élèves, orientent fortement leurs pratiques évaluatives ; elle propose de transformer ces représentations en leviers pour enrichir la formation et développer une vision plus formative et réflexive de l'évaluation. De leur côté, **Mortaqi et Ourahay** analysent, au Maroc, la performance des futur.es enseignant.es à une épreuve sommative en mathématiques et révèlent des lacunes importantes, en partie liées à une tradition pédagogique centrée sur la répétition d'exercices standardisés au détriment de la compréhension et de la résolution de problèmes. Ensemble, ces études mettent en lumière le défi commun de faire évoluer les conceptions et les pratiques évaluatives des enseignant.es en formation initiale, afin de mieux articuler évaluation formative et sommative et d'ancrer la résolution de problèmes au cœur de l'enseignement des mathématiques.

Genre et évaluation (2 communications). Dans leurs recherches, Brismontier d'une part, et Grapin et Sayac d'autre part, questionnent les inégalités de genre en mathématiques à l'école primaire à travers le prisme de l'évaluation. **Brismontier** s'intéresse à l'influence des stéréotypes de genre sur les interactions évaluatives en classe, en analysant comment les régulations didactiques, injonctives et affectives des enseignant.es peuvent renforcer ou atténuer les écarts de performance entre filles et garçons. De leur côté, **Grapin et Sayac** combinent une approche didactique et psycho-sociale pour expliquer pourquoi les filles, souvent meilleures en début de CP, deviennent moins performantes que les garçons dès le CE1, notamment sur le calcul mental. Leur étude montre que la pression évaluative et la menace du stéréotype jouent un rôle central : les filles produisent plus de non-réponses, s'appuient davantage sur des procédures visibles comme le comptage sur les doigts et manifestent une confiance moindre dans leurs réponses. Ensemble, ces travaux soulignent que les écarts de performance ne sont pas une fatalité mais le fruit de pratiques évaluatives et de contextes qu'il est possible de transformer pour limiter l'effet du genre dès les premières années de scolarité.

Conclusion

Les communications présentées lors de ce GT11 confirment que l'évaluation des apprentissages en mathématiques demeure un chantier essentiel et complexe, qui interroge sans cesse la place de l'élève, le rôle de l'enseignant, la validité des dispositifs et l'impact des contextes socioculturels. Les travaux discutés illustrent la diversité des échelles et des angles d'analyse : du geste évaluatif individuel à l'épreuve nationale standardisée, de l'influence des stéréotypes de genre aux enjeux de formation initiale, du diagnostic précoce aux pratiques de correction et de certification. Tous convergent vers une même exigence : penser une évaluation réellement au service de l'apprentissage, qui articule rigueur didactique, prise en compte de la complexité du raisonnement mathématique et attention aux dimensions humaines et sociales du geste évaluatif.

Pour prolonger ces dynamiques et enrichir les débats lors du prochain GT11, plusieurs pistes se dessinent :

- approfondir la question de l'évaluation des compétences spécifiques comme le raisonnement mathématique sous ses multiples formes (modélisation, géométrie, reformulation) ;
- interroger la mise en œuvre de pratiques différenciées, notamment sous l'angle du genre ou des besoins particuliers ;
- comparer les cultures évaluatives, leurs injonctions institutionnelles et leurs effets sur les apprentissages ;
- intégrer davantage le point de vue des élèves et leur rôle actif dans l'évaluation ;
- ouvrir la réflexion aux spécificités de l'enseignement supérieur et à la formation des futur·es enseignant·es, notamment quant à l'exemplarité et à l'harmonisation des discours entre théorie et pratique ;
- enfin, saisir les opportunités et les limites offertes par le développement des technologies numériques et de l'intelligence artificielle pour penser des formes d'évaluation plus adaptatives, plus transparentes et toujours plus centrées sur l'apprentissage.

Autant de questions vives qui invitent à poursuivre, ensemble, la construction d'évaluations à visage humain, respectueuses de la complexité des mathématiques et de la diversité des apprenant.es.

RÉFÉRENCES

- Allal, L. (2008). Évaluation des apprentissages. Dans A. Van Zanten (dir.), *Dictionnaire de l'éducation* (p. 311-314). Presses universitaires de France.
- Black, P. (2013). Formative and summative aspects of assessment: Theoretical and research foundations in the context of pedagogy. Dans J. H. McMillan (dir.), *SAGE handbook of research on classroom assessment* (p. 167-178). SAGE.
- Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.