

Évaluer en mathématiques : une question interdisciplinaire ?

Brigitte Grugeon-Allys, Éric Roditi and Nathalie Sayac

Volume 41, Number 1, 2018

URI: <https://id.erudit.org/iderudit/1055894ar>

DOI: <https://doi.org/10.7202/1055894ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this document

Grugeon-Allys, B., Roditi, É. & Sayac, N. (2018). Évaluer en mathématiques : une question interdisciplinaire ? *Mesure et évaluation en éducation*, 41(1), 1–6.
<https://doi.org/10.7202/1055894ar>

Évaluer en mathématiques : une question interdisciplinaire?

Introduction

Le numéro 1 du volume 41 de la revue *Mesure et évaluation en éducation* est un numéro spécial entièrement consacré à l'évaluation en mathématiques, plus précisément à l'évaluation des apprentissages mathématiques en milieu scolaire. La question posée dans le titre l'indique, ce numéro spécial propose une rencontre scientifique entre deux champs de recherche : celui de la didactique des mathématiques, qui s'intéresse de près aux phénomènes d'enseignement et d'apprentissage des concepts et méthodes de cette discipline scolaire ; et celui de l'évaluation, qui contribue, par ses développements théoriques et méthodologiques, à la production d'outils de mesure des acquis scolaires ainsi qu'à l'étude de la validité de ces outils.

Les échanges entre les acteurs de ces deux champs de recherche sont assez récents et apparaissent prometteurs. La didactique des mathématiques peut en effet contribuer à enrichir les études relatives à la pertinence des outils d'évaluation ; l'évaluation, comme champ scientifique, peut contribuer à faire progresser l'objectivation des apprentissages mathématiques ; et, conjointement, les deux champs pourraient enrichir les analyses des pratiques évaluatives des enseignants.

La rencontre de ces deux champs de recherche a constitué l'objectif majeur du colloque international *Évaluation en mathématiques : dispositifs, validités et pratiques*, qui s'est tenu les 21 et 22 novembre 2016 à l'Université Paris-Est Créteil, en France. La question de l'évaluation des apprentissages des élèves, notamment en mathématiques, est en effet actuellement une préoccupation majeure des différents acteurs de l'éducation, tant à l'échelle nationale qu'internationale. Le colloque s'est également appuyé sur des travaux menés dans le cadre du programme de

recherche Nouveaux outils pour de nouvelles pratiques d'évaluation et d'enseignement des mathématiques (NéOPRÆVAL) financé par l'Agence nationale de la recherche (ANR) en France.

Sur ces questions d'évaluation en mathématiques, le colloque a eu pour ambition d'associer des chercheurs ayant des références différentes (didactique, psychologie, psycho-éducatrice, etc.) et de permettre une rencontre avec des formateurs intervenant en formation initiale ou continue à tous les ordres d'enseignement (primaire, secondaire, supérieur). Trois axes thématiques ont structuré ce colloque.

Axe 1 : les différents dispositifs d'évaluation

Le premier axe thématique a permis d'interroger des dispositifs d'évaluation et leur mise en œuvre : quels sont les dispositifs d'évaluation externe ? Interne ? Des évaluations à portée diagnostique peuvent-elles faire l'objet d'un traitement à grande échelle et quelle serait la valeur informative de ces résultats pour les enseignants ? D'autres questions ont porté sur la place attribuée aux outils numériques dans la conception des évaluations : des tests dans l'environnement papier crayon avec des outils numériques ou des tests dans l'environnement numérique ? Quelles sont leurs potentialités et leurs limites ? Quel rôle peut jouer l'appropriation d'outils développés dans un environnement numérique pour la conception d'évaluations valides ou la production de parcours d'enseignement différencié adaptés aux besoins d'apprentissage de groupes d'élèves ? Une autre entrée a concerné l'accompagnement des enseignants à utiliser de nouvelles ressources et à développer de nouvelles pratiques d'évaluation. L'étude d'exemples de dispositifs d'évaluation issus de plusieurs pays a ainsi permis d'illustrer différentes façons de les concevoir, de les mettre en œuvre et d'interroger leur impact sur les apprentissages des élèves.

Axe 2 : l'étude de la validité des dispositifs d'évaluation et de leur contenu

Les travaux présentés et discutés dans cet axe thématique ont porté sur la qualité des informations que les évaluations permettent de recueillir ainsi que sur les moyens utilisés pour apprécier cette qualité. C'est bien sûr la question classique de la validité qui est ici posée : savoir, d'une part, si les dispositifs d'évaluation permettent d'évaluer ce que l'on cherche vraiment à évaluer ; savoir, d'autre part, quels outils mettre en œuvre pour estimer cette validité, depuis la conception du dispositif d'évaluation jus-

qu'à l'analyse des performances des élèves. Ces questions ont été posées tant au sujet d'enquêtes nationales ou internationales menées à grande échelle qu'à propos d'évaluations menées en classe de mathématiques, y compris dans le cas d'un usage de questionnaires issus d'évaluations externes pour construire des outils d'évaluations internes.

Axe 3 : l'analyse des pratiques d'évaluation en classe

Le troisième axe s'est centré sur les pratiques d'évaluation des enseignants. Quelles sont ces pratiques et quelles fonctions leur sont attribuées? Comment les étudier? Selon les ordres d'enseignement (primaire, secondaire, supérieur), quelles catégories d'évaluation (diagnostique, formative, sommative) peut-on observer? Ces catégories construites par la recherche sur l'évaluation conviennent-elles pour décrire les pratiques évaluatives des enseignants? Comment l'évaluation contribue-t-elle à articuler la programmation de l'enseignement, son ajustement aux élèves et la régulation des apprentissages? En quoi différentes approches de didactique disciplinaire, de psychologie ergonomique et de psychologie du développement peuvent-elles favoriser ces études?

Cinq articles sont proposés dans ce numéro spécial de la revue. Ils s'inscrivent dans les axes décrits ci-dessus et correspondent plus particulièrement à des conférences plénières données lors de ce colloque international qui a rassemblé plus de 150 participants.

Dans sa conférence d'ouverture, Sylvie Coppé (Université de Genève) a décrit le contexte scientifique dans lequel s'inscrit cette rencontre entre spécialistes de l'évaluation et didacticiens des mathématiques. L'auteure indique d'abord la variété des travaux sur l'évaluation, la diversité des approches (docimologie, évaluations internationales, évaluation formative, effets sur la motivation, liens avec l'orientation des élèves, etc.). Elle se centre sur les liens entre évaluations et apprentissages dans le cadre de l'enseignement primaire ou secondaire en prenant le point de vue de la didactique des mathématiques. Plus précisément, elle cherche à montrer qu'une entrée par la didactique des mathématiques avec ses cadres théoriques, ses outils et ses méthodes peut éclairer d'un jour nouveau les résultats produits dans le cadre des recherches sur l'évaluation, mais aussi que des recherches sur l'évaluation peuvent être développées en didactique des mathématiques et ainsi renouveler certains questionnements didactiques. Après un bref historique sur les recherches portant sur l'évaluation

formative, l'auteure fait le point sur les travaux actuels en didactique des mathématiques dans les pays francophones. Enfin, elle présente les résultats d'une recherche internationale sur l'évaluation entre pairs.

L'article de Paul Drijvers (Université d'Utrecht) s'inscrit principalement dans l'axe 1 et s'intéresse à l'évaluation numérique en mathématiques, à ses limites et à ses contraintes. La question centrale abordée est de savoir comment concevoir des tests numériques valides, c'est-à-dire qui évaluent les connaissances mathématiques visées. En s'appuyant sur des études sur la validité et en distinguant évaluation papier crayon avec outils numériques et évaluation dans l'environnement numérique, l'auteur présente des arguments en faveur d'une évaluation numérique (p. ex., richesse d'items numériques dynamiques et interactifs, grande adaptation possible des items, génération de *feedback* automatique, etc.) et ses potentialités. Il spécifie aussi ses principales difficultés. À partir de trois exemples d'évaluation numérique, l'auteur explore différentes façons de concevoir des tests numériques. En conclusion, il plaide en faveur du développement d'environnements numériques d'évaluation qui offrent de riches occasions aux élèves de « faire » des mathématiques et aux concepteurs de tests de concevoir des items très riches. Il insiste sur la nécessité d'améliorer le calcul automatisé des points obtenus par item, en particulier en ce qui concerne la notation des étapes intermédiaires dans les stratégies de résolution de problèmes.

L'article d'Antoine Bodin (Université de la Méditerranée Aix-Marseille) et de Nadine Grapin (Université Paris-Est Créteil) s'inscrit également dans l'axe 1. Il mobilise des théories issues de la didactique des mathématiques pour analyser les deux enquêtes internationales majeures évaluant les acquis des élèves dans cette discipline scolaire : le PISA et la TIMSS. Les auteurs prennent acte de l'abondante littérature à leur sujet, dans diverses disciplines relatives à l'éducation, et proposent de renouveler le regard porté sur elles grâce à une étude approfondie des questions d'évaluation. Pour l'analyse du PISA, les auteurs conjuguent une analyse cognitive des items et des outils de statistique implicative sur les performances des élèves. Ils mettent ainsi au jour des réseaux de connaissances et de savoir-faire, puis examinent comment ces réseaux varient d'un pays à l'autre. Par une analyse épistémologique et didactique des items de la TIMSS, les auteurs montrent que certaines notions ne sont évaluées que très partiellement, tant en ce qui concerne les connaissances à acquérir que leur mise en œuvre pour résoudre des tâches mathématiques.

L'article de Nathalie Loye (Université de Montréal) correspond à une conférence plénière donnée au sein de l'axe 2. Dans ce texte, l'auteure revient sur la question de la validité des instruments de mesure, en particulier des questionnaires ou des instruments d'évaluation. Dans un premier temps, l'auteure développe l'idée selon laquelle, globalement, les études portant sur la validité sont de deux types distincts : soit elles reposent sur des procédures quantitatives psychométriques, soit elles s'appuient essentiellement sur des considérations qualitatives émises par des experts du domaine. Dans un second temps, après avoir retracé l'évolution du concept de validité en évaluation, l'auteure s'appuie sur les deux références que sont le modèle de Kane et celui de Mislevy pour montrer la nécessité de conjuguer des démarches quantitative et qualitative si l'on souhaite produire des arguments pour soutenir la validation d'un instrument d'évaluation. L'auteure suggère en outre que la validation d'une épreuve, en tant que processus, doit se réaliser en même temps que sa conception, en se basant sur une bonne compréhension de ce qui est à évaluer.

L'article de Lucie Mottier Lopez et de Fernando Morales Villabona (Université de Genève) présente une étude que Lucie Mottier Lopez a présentée dans le cadre de la conférence plénière qu'elle a donnée au sein de l'axe 3 pour illustrer son propos. Cette étude examine le jugement évaluatif d'enseignants de l'école primaire genevoise lorsqu'ils attribuent des notes certificatives, plus particulièrement dans des contrôles écrits proposant des exercices de résolution de problèmes additifs. Les analyses entreprises, notamment au sujet d'incidents critiques observés pendant la correction des travaux des élèves, débouchent sur la conceptualisation de différents cas de figure entre des jugements analytiques et holistiques, dont la mise en tension constructive peut être vue comme un enjeu majeur de développement pour l'évaluation d'apprentissages complexes.

Les cinq articles de ce numéro spécial visent ainsi à aborder différentes questions relatives à l'évaluation en mathématiques. Chacun à sa manière montre comment les deux champs de recherche que sont la didactique des mathématiques et l'évaluation peuvent contribuer à la production de savoirs scientifiques originaux et fertiles pour chacun des deux champs. Cette production contribuera, à n'en pas douter, à la naissance de vocations interdisciplinaires sur cette question de l'évaluation en mathématiques. Par leur diversité et la richesse de leurs arguments, nous espérons que ces textes contribueront à en convaincre le lecteur.

Brigitte Grugeon-Allys

*Université Paris-Est Créteil
Laboratoire de Didactique André Revuz*

Éric Roditi

*Université Paris Descartes
Laboratoire Éducation Discours Apprentissages*

Nathalie Sayac

*Université Paris-Est Créteil
Laboratoire de Didactique André Revuz*