Mesure et évaluation en éducation, 2026, vol. 49, n°1, v-viii

ÉDITORIAL
Mesurer autrement. Pluralité des sources de validité et soutien aux apprentissages

Dan Thanh Duong Thi
Directrice de la revue MEE
Université du Québec à Montréal (UQÀM)

C'est avec un grand plaisir que j'ai rédigé l'éditorial du numéro 49(1), qui est le premier numéro que j'ai coordonné à titre de directrice de la revue *Mesure et évaluation en éducation*. Malgré la diversité de leurs objets d'études, les quatre articles de ce numéro partagent un même questionnement : à quelles conditions les informations produites par une évaluation permettent-elles des interprétations valides et des décisions utiles à l'action éducative ? Cette question traverse aujourd'hui le champ de la mesure et de l'évaluation en éducation selon différents angles.

Premièrement, la conception de la validité s'est profondément transformée. Depuis Messick (1989) et Kane (2013), la validité n'est plus considérée comme une propriété de l'instrument mais comme le degré auquel un ensemble de preuves soutient l'interprétation et l'usage des scores. Les *Standards for Educational and Psychological Testing* (AERA et al., 2014) distinguent ainsi cinq sources de preuves : le contenu, les processus de réponse, la structure interne, les relations avec d'autres variables et les conséquences. Construire un argument de validité suppose donc de réunir des preuves de nature différente, ce qu'une seule étude permet rarement.

Deuxièmement, les construits que l'on cherche à développer et à évaluer se sont élargis. Ainsi, l'évaluation de la maturité à la sobriété numérique, de la participation aux activités de plein air et des fonctions exécutives en petite enfance soulève la question de la validité écologique,

Cette oeuvre est distribuée sous licence Creative Commons Attribution 4.0 International
http://creativecommons.org/licences/by/4.0/deed.fr

c'est-à-dire, dans quelle mesure des résultats obtenus permettent l'application et les généralisations aux situations de la vie réelle. Elle oblige donc à choisir avec soin la source d'information comme preuve de validité.

Troisièmement, la frontière entre l'évaluation formative et l'évaluation sommative est réinterrogée. L'*Assessment for Learning ou l'Évaluation-soutien d'Apprentissage* (Allal & Laveault, 2009), postule que l'évaluation, y compris l'évaluation certificative, peut soutenir les apprentissages. Cette ambition se heurte toutefois à des pratiques de notation qui demeurent largement traditionnelles.

Le présent numéro se situe au croisement de ces trois angles. Les contributions réunies sont issues des travaux de 21 autrices et auteurs rattachés à des institutions de Belgique, du Canada, de France et de Suisse. Elles portent sur des publics variés, de la petite enfance à l'âge adulte, et mobilisent à la fois des approches quantitatives et qualitatives.

Trois des contributions relèvent de la psychométrie et portent sur l'élaboration et la validation de différents instruments. Dans la première, Descamps et ses collègues élaborent une échelle destinée à mesurer la maturité à la sobriété numérique des adultes portant sur trois dimensions, soit la maturité numérique, la maturité écologique et la maturité relative à la sobriété numérique. Une version initiale de 165 items, révisée par des experts, a été administrée à 133 adultes. Elle a ensuite été réduite au moyen d'une analyse en composantes principales et de coefficients alpha de Cronbach, puis administrée dans sa forme abrégée à 121 personnes. Malgré quelques limites, l'outil articule ainsi les composantes en écogestes numériques perçus et suggèrent donc des perspectives prometteuses d'implications pratiques éducatives et de recherche auprès de différentes populations.

Dans le domaine de l'activité physique, Top et ses collègues élaborent le QP3A-j, un questionnaire fait pour évaluer la participation des jeunes de 5 à 15 ans à des activités de plein air. L'élaboration du questionnaire est assurée par une démarche rigoureuse du développement et de la validation de l'instrument. Administré auprès de 175 parents provenant de différents milieux du Québec, les analyses de composantes principales et factorielles exploratoires suggèrent de bonnes propriétés psychométriques pour les items, qui sont regroupés en quatre dimensions principales, soit la créativité, l'intérêt et la participation sportive ainsi que le sentiment de sécurité parental. Les résultats obtenus appuient donc l'utilisation du QP3A-j dans différents contextes, grâce à son accessibilité, à son originalité, à sa rapidité

ainsi qu'à sa pertinence, ce qui permet de combler le manque d'outils pour mesurer la participation des enfants aux activités en plein air, notamment en contexte francophone.

Duval et ses collègues se penchent, pour leur part, sur l'évaluation les fonctions exécutives des jeunes enfants de 3 à 6 ans, telles qu'elles se manifestent dans les milieux éducatifs. Les autrices proposent donc la GOFE-R, une grille d'observation des comportements spontanés de l'enfant en situation de jeu libre, cotés à partir de vidéos par des observatrices formées et certifiées. Menée auprès de 312 enfants fréquentant des services de garde ou des classes de préscolaire de la région de la ville de Québec, l'étude permet de réunir plusieurs sources de preuves telles que la fidélité interjuges, la structure interne, la validité convergente et divergente et la validité concourante. Les analyses factorielles exploratoires suggèrent que la structure de l'instrument comporte deux composantes, soit les fonctions d'ajustement et de régulation comportementale et la fonction de régulation cognitive et métacognitive. Possédant une bonne qualité de fidélité et de validité, l'outil soutient la possibilité d'observer les comportements liés aux fonctions exécutives en contexte de service de garde à l'enfance et en contexte d'éducation préscolaire.

La contribution de Pasquini et ses collègues adopte, quant à elle, une démarche qualitative. Menée dans une classe de 11e année du canton de Vaud en Suisse, en collaboration avec un enseignant de mathématiques, cette étude de cas cherche à cerner ce qui permet à une épreuve sommative de certifier les acquis tout en soutenant leur régulation à travers les interactions entre l'enseignant et les élèves pendant la passation. L'analyse des données (artefacts, interactions entre l'enseignant et des élèves, entretiens auprès de l'enseignant et de certains élèves) se base sur trois aspects, à savoir l'alignement curriculaire élargi (Pasquini, 2019) et de critères d'évaluation explicites, qui servent de leviers aux échanges, et les interactions durant la passation comme soutien de l'apprentissage. Les résultats soulignent la possibilité de remplir la fonction de l'évaluation pour le soutien de l'apprentissage des élèves dans une situation d'évaluation sommative comme celle qui est présentée dans la recherche. Les auteurs soulèvent aussi des réflexions pertinentes sur la nature des interactions, notamment en ce qui concerne le type de rétroactions fournies aux élèves et le maintien de l'équilibre entre le rôle de l'enseignant et celui de l'évaluateur lors de la passation de l'épreuve.

Sur le plan scientifique, le numéro apporte au champ trois instruments francophones dans des domaines où les auteurs constataient un manque d'outil : la maturité à la sobriété numérique des adultes, la participation des jeunes aux activités de plein air et les fonctions exécutives observées en contexte éducatif. Ces études illustrent concrètement la construction progressive d'un argument de validité, fondé sur des preuves de contenu, de structure interne et, dans le cas de la GOFE-R, de relations avec d'autres variables. L'article de Pasquini et ses collègues contribue, pour sa part, à la réflexion sur l'Évaluation-soutien d'Apprentissage en contexte certificatif. Il montre, dans le cas étudié, qu'un alignement curriculaire élargi et des critères explicites peuvent faire de l'épreuve sommative un moment de régulation.

Pour conclure, ce numéro invite à considérer la validité non pas comme un label acquis une fois pour toutes, mais comme un argument qui se construit dans la durée. Il intéresse aussi les personnes enseignantes et formatrices. Expliciter les critères et le seuil de réussite, diversifier les sources d'information et observer les apprentissages en situation sont autant de principes transférables à l'évaluation en classe.

RÉFÉRENCES

Allal, L., & Laveault, D. (2009). Assessment for learning: évaluation-soutien d'apprentissage. *Mesure et évaluation en éducation*, *32*(2), 99-106. https://doi.org/10.7202/1024956ar

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Messick, S. (1989). Validity. Dans R. L. Linn (dir.), *Educational measurement* (3e éd., p. 13-103). American Council on Education, Macmillan.

Pasquini, R. (2019). Élargir conceptuellement le modèle de l'alignement curriculaire pour comprendre la cohérence des pratiques évaluatives sommatives notées des enseignants : enjeux et perspectives. *Mesure et évaluation en éducation*, *42*(1), 63–92. https://doi.org/10.7202/1066598ar