le cnam
Cnesco
Centre national d'étude des systèmes scolaires

Université Clermont Auvergne

INSPÉ Institut national supérieur du professorat et de l'éducation Clermont Auvergne

CONFÉRENCE DE **CONSENSUS**

L'ÉVALUATION EN CLASSE, AU SERVICE DE L'APPRENTISSAGE DES ÉLÈVES

NOTES DES EXPERTS

#CC_EVALUATION

LES 23 ET 24 NOVEMBRE 2022

En partenariat avec :

ifé INSTITUT FRANÇAIS DE L'ÉDUCATION

Université de Lyon

ENS DE LYON

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

RESEAU-CANOPE.FR
CANOPÉ
LE RÉSEAU DE CRÉATION ET D'ACCOMPAGNEMENT PÉDAGOGIQUES

INSPÉ RÉSEAU DES INSPÉ

NOTES DES EXPERTS

Mars 2023

le cnam
Cnesco

Centre national d'étude des systèmes scolaires

Pour citer ce document, merci d'utiliser la référence suivante :
Cnesco (2023). *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts*. Cnesco-Cnam.

Ce document s'inscrit dans une série de ressources publiées par le Centre national d'étude des systèmes scolaires (Cnesco) sur la thématique : **L'évaluation en classe**.

Disponible sur le site du Cnesco : www.cnesco.fr
Publié en mars 2023.
Centre national d'étude des systèmes scolaires
41 rue Gay-Lussac 75005 Paris

Contact : cnesco@lecnam.net - 06 98 51 82 75

Sommaire

Quels sont les effets psychosociaux de l'évaluation normative ? Fabrizio Butera 5

Quelles sont les nouvelles conceptions de l'évaluation en classe ? Vers des évaluations pour apprendre durablement. Lucie Mottier Lopez 13

La politique évaluative de la France : quelles évolutions engagées depuis 2014 et quels impacts dans les classes ? Brigitte Hazard 23

Quelles sont les pratiques et les représentations de l'évaluation chez les enseignants ? Sophie Genelot 46

Quelles sont les perceptions par les parents et les élèves des pratiques évaluatives dans les classes ? Stéphane Bénit 62

Quelle est la relation entre évaluation et motivation dans le cadre des apprentissages scolaires ? Fabien Fenouillet 75

Quels sont les liens entre l'évaluation, la perception que les élèves ont d'eux-mêmes et leurs progrès ? Pascal Pansu 84

En classe, quels effets peuvent avoir les interactions à dimension évaluative sur les élèves en termes d'inégalités ? Sébastien Goudeau 93

Comment l'évaluation peut-elle soutenir un suivi des progrès qui permette à tous les élèves d'apprendre ? Liesje Coertjens 101

Peut-on évaluer différemment des enfants différents ? Mickaël Jury 115

Comment la planification peut-elle permettre de passer d'une évaluation de l'apprentissage à une évaluation pour l'apprentissage ? Gonzague Yerly 126

À quelles conditions développer des pratiques de notation au service des décisions des enseignants et des apprentissages des élèves ? Raphaël Pasquini 136

Quels sont les apports de la psychologie cognitive pour articuler évaluation et apprentissage ? Daniel Gaonac'h 148

Comment le numérique peut-il aider les enseignants à mettre en œuvre des évaluations formatives ? Franck Silvestre 156

Comment peut-on rendre compte des progrès des élèves aux familles ? L'exemple de l'école maternelle. Christophe Joigneaux 167

Comment peut-on définir les attentes évaluatives en classes ? Une synthèse des réflexions en EPS. Léa Gottsmann 174

Quels apports de la didactique pour penser l'évaluation ? L'exemple des mathématiques. Julie Horoks et Julia Pilet 182

Comment l'évaluation peut-elle favoriser la compréhension de l'oral ? L'exemple des langues vivantes étrangères. Stéphanie Roussel. 193

Comment les enseignants peuvent-ils partager la responsabilité de l'évaluation avec les élèves ? Fernando Morales Villabona 209

Comment la formation des personnels peut-elle favoriser le développement d'une évaluation au service de l'apprentissage ? Sylvie Fontaine 219

Quels sont les effets psychosociaux de l'évaluation normative ?

FABRIZIO BUTERA
UNIVERSITÉ DE LAUSANNE (SUISSE)

Introduction

Traditionnellement, la recherche sur l'évaluation scolaire s'est focalisée sur des questions de précision et de diagnosticité. Un grand nombre de travaux ont développé des méthodes, des programmes et des instruments qui puissent rendre compte au mieux de l'apprentissage et de la performance des élèves, ainsi que de l'atteinte des objectifs pédagogiques. Les travaux qui seront présentés dans cette note portent en revanche sur ce qui se passe « après l'évaluation », en particulier sur les effets psychosociaux qui découlent du système d'évaluation mis en place (Butera *et al*., 2020). En effet, l'évaluation n'est pas un exercice neutre qui consisterait seulement à qualifier et à quantifier les mérites de la production d'un élève : selon le type d'évaluation utilisée (par exemple, normative ou formative) et selon la fonction qu'on lui attribue (sélectionner ou former), les élèves (mais aussi les enseignants) développent des représentations différentes, de leur autonomie, par exemple.

I. Fonction de formation et de sélection, évaluation formative et normative

Pour comprendre pourquoi il peut y avoir des effets psychosociaux différents en fonction du type d'évaluation utilisé, il faut considérer les liens qui existent entre fonctions de l'école, justice distributive et évaluation. En effet, les élèves, leurs parents et les enseignants sont bien conscients que l'école est un lieu où coexistent deux fonctions distinctes (Dornbush *et al*., 1996). D'une part, la formation est la fonction de l'école qui paraît la plus évidente, la plus représentative : l'école a pour mission de transmettre des savoirs et de développer des compétences disciplinaires et sociales. D'autre part, l'école est aussi un lieu de sélection, un lieu où l'évaluation qu'on reçoit a un impact sur le cursus scolaire, sur les opportunités de choix de filières d'études ou de formation, et *in fine* sur l'emploi et les positions sociales auxquelles on peut accéder en fonction des diplômes obtenus (Darnon *et al*., 2009).

Il est important de distinguer ces deux fonctions des systèmes éducatifs lorsqu'on considère l'effet des différentes formes d'évaluation ; dans cette note nous nous concentrerons sur l'évaluation formative et l'évaluation normative. L'évaluation formative consiste à communiquer aux élèves quelles sont les parties de la tâche ou du matériel à apprendre qui nécessitent du travail supplémentaire, ou la nature de l'apprentissage résiduel nécessaire pour atteindre les objectifs fixés, et d'accompagner les élèves dans cette progression (par exemple, Allal, 1991). L'évaluation formative aide donc, tout naturellement, les enseignants à remplir la fonction de formation de l'école. L'évaluation normative consiste à attribuer une valeur à la production des élèves—qu'elle soit représentée par un nombre, une lettre, un jugement—ce qui rend le rang de cette production (et de l'élève) visible dans la comparaison à un standard ou au reste de la classe (Butera, 2011 ; Marshall & Weinstein, 1984). La

simplicité, visibilité et comparabilité de l'évaluation normative rend aisé de classer les élèves et, le moment venu, de les sélectionner (redoublement, orientation, prix, …).

À ce propos, il y a bien des années, Morton Deutsch avait déjà théorisé le lien que l'évaluation normative entretient avec la fonction de sélection de l'école. Ce lien serait basé sur la capacité de l'évaluation normative à attribuer une valeur comparative au travail de l'élève et donc du mérite à l'élève même, mérite qui ensuite est rendu tangible par des récompenses académiques telles que la position hiérarchique dans la classe, la promotion, ou l'orientation vers des filières désirables (Deutsch, 1979). Plus récemment, une étude a illustré cette relation et a montré que l'association entre fonction sélective et évaluation normative n'est pas seulement une construction théorique, mais elle existe en tant que représentation sociale (Autin *et al*, 2015). Dans cette étude, plus d'une centaine d'étudiants universitaires participaient à un exercice qui leur demandait de fonctionner en tant qu'enseignants ; ils devaient ensuite répondre à un questionnaire sur la fonction des systèmes éducatifs et de l'évaluation. Les résultats de cette étude révèlent que plus les participants percevaient le système éducatif comme ayant un rôle de sélection, plus ils percevaient l'évaluation normative comme une méthode utile et à utiliser. De façon intéressante, cette association était expliquée par la perception que l'évaluation normative se fonde sur l'équité en tant que principe de justice distributive (Deutsch, 1979), un principe qui est à la base de l'idéologie méritocratique : donner plus à qui a une performance supérieure (Batruch *et al*., 2019b). En outre, plus les participants percevaient le système éducatif comme ayant un rôle de sélection, plus ils percevaient l'évaluation formative comme une méthode inutile et à ne pas utiliser. Cette association était expliquée par la perception que l'évaluation formative se fonde sur les principes de justice d'égalité et de besoin (Deutsch, 1975) : donner la même chose à tout le monde ou davantage à qui en a le plus besoin. Ces associations se retrouvent, avec d'autres méthodologies, dans plusieurs autres études (Autin *et al.,* 2019 ; Batruch et *al.,* 2019a). En somme, les évaluations normative et formative ne sont pas seulement des méthodes ou des pratiques, mais elles s'insèrent dans un ensemble cohérent de représentations sur les fonctions et les valeurs de justice de l'école. L'utilisation d'une forme d'évaluation plutôt que d'une autre entraîne donc des effets psychosociaux parce qu'elle active une représentation plutôt qu'une autre de l'école.

Les travaux rapportés dans cette note se focalisent sur les effets psychosociaux de l'évaluation normative, car elle est, de loin, la plus utilisée dans tous les systèmes éducatifs (Knight & Yorke, 2003). Cette prévalence est telle que deux études ont montré que les effets psychosociaux de l'évaluation normative sont identiques qu'on confronte les étudiants à une évaluation normative effective ou qu'on active simplement son existence par amorçage (en rendant visible, mais sans y faire référence, le système de notation en vigueur ; Hayek *et al*., 2015 ; voir aussi Hayek *et al.,* 2017 pour une étude menée avec des élèves d'école primaire).

Les effets psychosociaux de l'évaluation normative apparaissent à au moins trois niveaux différents— motivation intrinsèque et autonomie, compétences sociales, et égalité des chances— qui sont détaillés ci-dessous.

II. Effets sur la motivation intrinsèque et l'autonomie

La motivation intrinsèque est l'énergie qui permet d'accomplir une tâche et qui vient de l'intérieur de l'individu, de son plaisir et de son intérêt, par opposition à la motivation extrinsèque, qui vient des incitations et des contraintes externes (Ryan & Deci, 2000). C'est la motivation la plus appréciée et la plus recherchée à l'école, surtout parce qu'elle permet à l'élève de travailler de façon autonome (Ryan & Connell, 1989). Il y a eu un long débat sur les effets de l'évaluation normative (positive) en tant que récompense sur la motivation intrinsèque, avec un camp qui soutenait que la récompense gratifie l'individu et promeut la motivation intrinsèque (Eisenberger *et al*., 1999), et un autre camp qui défendait l'idée que la récompense représente une source externe de motivation et ne peut donc pas promouvoir la motivation intrinsèque (Deci *et al*., 1999)

Des études plus récentes ont opposé ces deux hypothèses spécifiquement sur la question de l'évaluation normative. Les résultats de plusieurs études montrent de façon convergente que l'évaluation normative ne donne pas la possibilité de se sentir suffisamment autonome par rapport aux tâches scolaires pour être motivé à continuer à apprendre indépendamment de la récompense ou de la punition que représente l'évaluation normative. Autrement dit, l'évaluation normative représente une contrainte externe qui réduit le sentiment d'autonomie (Pulfrey *et al*., 2011). De plus, il apparaît que la réduction de l'autonomie perçue induite par l'évaluation normative, réduit aussi la motivation intrinsèque, en particulier la persistance dans le comportement, c'est-à-dire l'engagement dans la tâche même une fois que le devoir est terminé (Pulfrey *et al*., 2013). En résumé, ces études montrent que l'évaluation normative représente pour l'élève une contrainte externe qui réduit le sentiment d'autonomie et la motivation intrinsèque, comparativement à l'absence d'évaluation ou à une évaluation formative. Ces effets ont été trouvés indépendamment du genre des élèves et de leur filière d'études.

III. Effets sur les compétences sociales

Les compétences sociales sont les compétences qui permettent de travailler en groupe et de s'adapter de façon efficace à des environnements sociaux changeants (Smart & Sanson, 2003). Il s'agit de compétences très importantes à l'école—notamment la capacité à collaborer, à partager les informations et les ressources, à faire confiance aux autres— pour assurer le bon fonctionnement de la classe, les bonnes relations entre élèves et l'efficacité du travail collectif (Johnson & Johnson, 2005). Ces compétences sont aussi connues pour prédire la performance scolaire (voir par exemple la méta-analyse de Durlak *et al*., 2011).

Plusieurs études suggèrent que l'évaluation normative entrave les processus de groupe qui devraient amener à l'émergence ou à l'utilisation de compétences sociales fondamentales pour la vie en société. Des études montrent, en effet, que lors d'un travail de groupe, une évaluation qui met en avant la réussite d'un membre par rapport aux autres (plutôt que la réussite du groupe dans son ensemble) amenait les membres du groupe à faire de la rétention d'information, à ne pas partager les éléments qui pourtant auraient été nécessaires à la résolution de la tâche (Toma & Butera, 2009). D'autres études montrent que la présence de l'évaluation normative dans un travail de groupe, ou même sa simple évocation, amenait les membres du groupe à partager les informations de façon stratégique, pour s'avantager personnellement (Hayek *et al*., 2015), et à ne pas se coordonner dans une tâche qui pourtant demandait le concours de tous les membres pour être réussie (Hayek *et al*., 2017).

L'évaluation normative rend saillante la possibilité d'un classement et transforme donc les partenaires en adversaires.

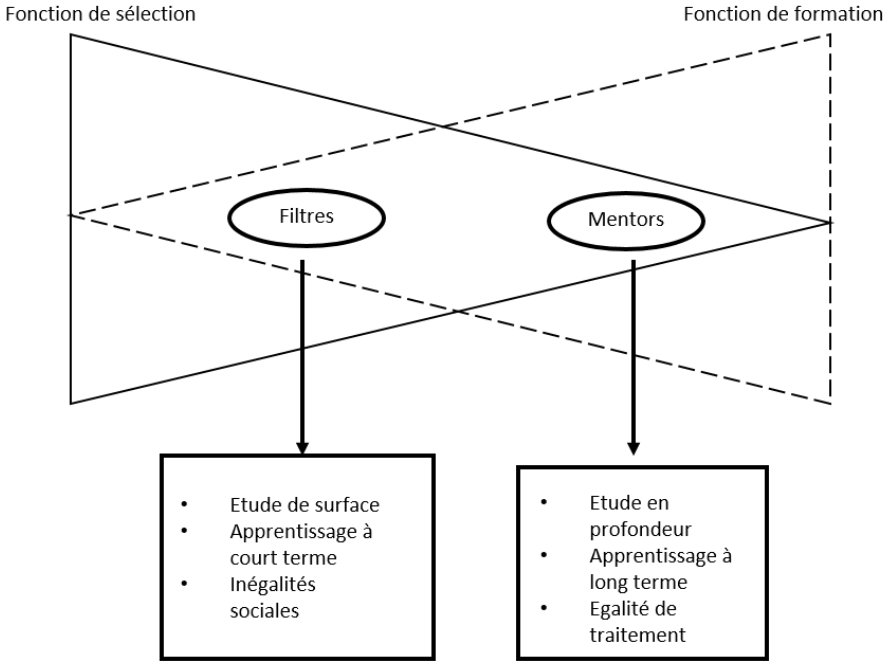
IV. Effets sur l'égalité des chances

Les effets psychosociaux de l'utilisation de l'évaluation normative ne se manifestent pas seulement chez les élèves, mais aussi chez les enseignants. Plusieurs études montrent que lorsque l'évaluation dans le système scolaire est normative et uniquement conçue comme un instrument de sélection, les enseignants sont amenés à contribuer à la reproduction des inégalités scolaires déjà existantes, notamment les inégalités de classe sociale. Une série d'études montrent que les évaluateurs d'une dictée qui devaient utiliser une évaluation normative et sélective trouvaient plus d'erreurs dans la copie d'un élève de classe sociale défavorisée que dans celle d'un élève de classe sociale favorisée, alors que le nombre d'erreurs était strictement le même. Cette différence disparaissait lorsque l'on demandait d'utiliser une évaluation formative (Autin *et al*., 2019). Les mêmes dynamiques apparaissent si la tâche des évaluateurs était de faire de la sélection en orientant un élève : on préconisait davantage une filière moins exigeante—et moins une filière plus exigeante—pour l'élève de classe sociale défavorisée que pour l'élève de classe sociale favorisée, alors que la procédure expérimentale faisait que les résultats scolaires des deux élèves étaient identiques. Cette différence était réduite lorsque la fonction de formation de l'école était rendue saillante, plutôt que la fonction de sélection (Batruch *et al*., 2019a). Ces études montrent donc que la fonction de sélection de l'école, et son association avec l'évaluation normative, amènent à des évaluations qui perpétuent les inégalités qui existent par ailleurs dans le système scolaire (voir par exemple OECD, 2011).

V. Discussion

Ces travaux mettent en évidence que l'évaluation normative suscite les effets psychosociaux documentés ci-dessus car elle favorise la fonction sélective de l'école. Ceci a un impact sur le rôle qui est donné aux enseignants : dans une école sélective qui applique le principe de justice du mérite, les enseignants jouent un rôle de filtre, puisque l'évaluation normative leur permet d'attribuer du mérite aux productions des élèves. Cependant, il faut aussi considérer que dans une école où c'est la fonction formative qui est mise en exergue et qui applique les principes de justice d'égalité et de besoin, les enseignants pourraient avoir un rôle de mentors, puisque l'évaluation formative pourrait leur permettre de se focaliser sur les progrès des élèves (Butera *et al*., 2021, voir Figure 1).

Figure 1 : Le rôle différentiel des enseignants en fonction de la saillance des fonctions formative et sélective de l'école



Adapté et traduit de : Butera *et al.* (2021).

L'évaluation formative existe depuis longtemps et depuis longtemps la recherche a montré ses effets bénéfiques (Allal & Mottier-Lopez, 2007). Cependant, la centration sur l'apprentissage plutôt que sur la performance nécessite une école plus centrée sur sa fonction de formation, pour que l'évaluation formative ne soit pas découragée, voire entravée (Autin *et al*., 2015). Cette fonction peut être favorisée par l'utilisation en classe de l'apprentissage coopératif qui donne des buts communs, des rôles complémentaires et une dignité égale à tous les élèves (Johnson & Johnson, 2009). Cependant, l'apprentissage coopératif est une méthode de travail qui reste encore rare à l'école (Buchs, 2017 ; Buchs *et al*., 2016) et il a besoin d'un développement plus collectif, avec des écoles entières ou même des quartiers, des villes ou des académies qui adoptent des méthodes basées sur l'apprentissage coopératif et l'évaluation formative.

Références

Allal, L. (1991). *Vers une pratique de l'évaluation formative : Matériel de formation continue des enseignants.* De Boeck.

Allal, L. & Mottier-Lopez, L. (dir.) (2007). *Régulation des apprentissages en situation scolaire et en formation*. De Boeck.

Autin, F., Batruch, A. & Butera, F. (2015). Social justice in education: How the function of selection in educational institutions predicts support for (non)egalitarian assessment practices. *Frontiers in Psychology, 6:707.* https://doi.org/10.3389/fpsyg.2015.00707.

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

Batruch, A., Autin, F., Bataillard, F. & Butera, F. (2019a). School selection and the social class divide: How tracking contributes to the reproduction of inequalities. *Personality and Social Psychology Bulletin,* 45, 477–490. https://doi.org/10.1177/0146167218791804

Batruch, A., Autin, F. & Butera, F. (2019b). The paradoxical role of meritocratic selection in the perpetuation of social inequalities at school. Dans J. Jetten et K. Peters (dir.), *The social psychology of inequality* (p. 123-137). Springer Nature.

Buchs, C. (2017). *Comment organiser l'apprentissage des élèves par petits groupes ?* Le CNAM, CNESCO, https://www.cnesco.fr/differenciation-pedagogique/paroles-dexpert/travail-cooperatif-entre-et-avec-les-eleves/

Buchs, C., Gilles, I., Antonietti, J. P. & Butera, F. (2016). Why students need to be prepared to cooperate: A cooperative nudge in statistics learning at university. *Educational Psychology*, *36*, 956-974. https://doi.org/10.1080/01443410.2015.1075963

Butera F. (2011). La menace des notes. Dans F. Butera, C. Buchs, & C. Darnon (Eds.) *L'évaluation, une menace ?* (45-53). Presses Universitaires de France.

Butera, F., Batruch, A., Pulfrey, C., Autin, F. & Toma, C. (2020). Après la note : Effets psychosociaux de l'évaluation normative. Dans N. Younès, C. Gremion et E. Sylvestre (dir.), *Evaluations, sources de synergies ? Entre normalisation, contrôle et développement formatif* (pp. 53-68). Presses de l'ADMEE-Europe.

Butera, F., Batruch, A., Autin, F., Mugny, G., Quiamzade, A. & Pulfrey, C. (2021). Teaching as social influence: Empowering teachers to become agents of social change. *Social Issues and Policy Review, 15,* 323-355. https://doi.org/10.1111/sipr.12072

Darnon, C., Dompnier, B., Delmas, F., Pulfrey, C. & Butera F. (2009). Achievement Goal Promotion at University: Social Desirability and Social Utility of Mastery and Performance Goals. *Journal of Personality and Social Psychology, 96,* 119-134. https://doi.org/10.1037/a0012824

Deci, E. L., Koestner, R. & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*, 627-668. https://doi.org/10.1037/0033-2909.125.6.627

Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social issues*, *31*, 137-149. https://doi.org/10.1111/j.1540-4560.1975.tb01000.x

Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. *American Psychologist*, *34*, 301-401. http://dx.doi.org/10.1037/0003-066X.34.5.391

Dornbusch, S. M., Glasgow, K. L. & Lin, I. C. (1996). The social structure of schooling. *Annual Review of Psychology*, *47*, 401-429. https://doi.org/10.1146/annurev.psych.47.1.401

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*, 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Eisenberger, R., Pierce, W. D. & Cameron, J. (1999). Effects of reward on intrinsic motivation: Negative, neutral and positive: Comment on Deci, Koestner and Ryan, 1999. *Psychological Bulletin, 125*, 677-691. https://doi.org/10.1037/0033-2909.125.6.677

Hayek, A.S., Toma, C., Oberlé, D. & Butera, F. (2015). Grading hamper cooperative information sharing in group problem solving. *Social Psychology, 46*, 121-131. https://doi.org/10.1027/1864-9335/a000232

Hayek, A.S., Toma, C., Guidotti, S., Oberlé, D. & Butera, F. (2017). Grades degrade group coordination: deteriorated interactions and performance in a cooperative motor task. *European Journal of Psychology of Education,* 32*,* 97–112. https://doi.org/10.1007/s10212-016-0286-9

Johnson, D. W. & Johnson, R. T. (2005). New developments in social interdependence theory. *Genetic, Social, and General Psychology Monographs*, *131*, 285–358. https://doi.org/10.3200/MONO.131.4.285-358

Johnson, D. W. & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher,* 38, 365–379. https://doi.org/10.3102/0013189X09339057

Knight, P. & Yorke, M. (2003). *Assessment, learning and employability*, Open University Press.

Marshall, H. H. & Weinstein, R. S. (1984). Classroom factors affecting students' self-evaluations: An interactional model. *Review of Educational Research*, *54*, 301-325. https://doi.org/10.3102/00346543054003301

Organisation for Economic Co-operation and Development. (2011). *Equity and quality in education-supporting disadvantaged students and schools.* Paris, France: Author. https://doi.org/10.1787/9789264130852-en

Pulfrey, C., Buchs, C. & Butera, F. (2011). Why grades engender performance avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology, 103,* 683-700. https://doi.org/10.1037/a0023911

Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist,* 55*,* 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Smart, D. & Sanson, A. (2003). Social competence in young adulthood, its nature and antecedents. *Family Matters*, *64*, 4–9.

Toma, C. & Butera, F. (2009). Hidden profiles and concealed information: Strategic information sharing and use in group decision making. *Personality and Social Psychology Bulletin,* 35, 793-806. https://doi.org/10.1177/0146167209333176

Quelles sont les nouvelles conceptions de l'évaluation en classe ? Vers des évaluations pour apprendre durablement

LUCIE MOTTIER LOPEZ
UNIVERSITÉ DE GENÈVE (SUISSE)

Introduction

« L'évaluation est une activité de recueil de diverses informations considérées comme pertinentes par rapport à des objectifs visés, en vue de porter un jugement et de prendre une décision »
(Amigues & Zerbato-Poudou, 1996)

L'étymologie du mot évaluer est « faire ressortir la valeur de », « extraire la valeur de … ». C'est la fonction pédagogique et institutionnelle qui attribue une connotation sociale à l'évaluation scolaire, notamment au regard du type de décision qu'elle permet de prendre : une décision de régulation, une décision de certification, une décision d'orientation, notamment. Afin d'offrir une ouverture « large » à la problématique de l'évaluation des apprentissages en classe, je vais, dans la première section, rappeler la distinction classique entre évaluation formative et évaluation sommative, et présenter dans les sections suivantes les mises en question de cette distinction et quelques reconfigurations conceptuelles proposées par la littérature scientifique actuelle en sciences de l'éducation et de la formation.

I. À l'origine, une distinction entre évaluation formative et évaluation sommative des apprentissages des élèves

Les travaux internationalement connus de Bloom *et al*. (1971) sur la pédagogie de maîtrise ont amené à différencier l'évaluation *formative* et l'évaluation *sommative* pour ce qui concerne les apprentissages des élèves[1]. Cette distinction a été largement reprise dans tous les continents du monde et elle est, aujourd'hui encore, une référence pour établir une différence entre :

- **Les évaluations dont les résultats sont exploités pour soutenir les apprentissages des élèves**. Dans le *dictionnaire de l'éducation* coordonné par van Zanten, Allal (2008) rappelle que ces évaluations ont une fonction *formative*, c'est-à-dire qu'elles visent « l'adaptation des activités d'enseignement et d'apprentissage afin de favoriser la progression des apprenants vers les objectifs de formation » (p. 312). Les résultats qu'elles produisent fournissent une information à l'enseignant et à l'élève, permettant notamment de différencier les activités didactiques au regard des besoins des élèves (y compris pour les élèves de bon niveau scolaire) et de repérer des problèmes d'apprentissage et/ou d'enseignement. Elles donnent lieu à des *rétroactions* (ou feedbacks) : elles sont exploitées pour réguler (adapter) l'enseignement et soutenir

[1] Introduite par Scriven, cette distinction existait déjà pour l'évaluation des programmes et des dispositifs.

l'autorégulation de l'élève. Elles peuvent être formelles ou informelles, instrumentées (sous forme papier crayon, avec des grilles d'autoévaluation, des listes de vérification, etc.) ou non (dans l'interaction, par l'observation par exemple, mais sans que cela ne donne lieu à des traces consultables par la suite).

- Puis, il y a **les évaluations qui servent à « établir un bilan certifiant les compétences et les connaissances acquises** par l'apprenant à la fin d'une étape intermédiaire (par exemple, semestre) ou d'une étape finale (par exemple, année, cycle pluriannuel) d'un cursus de formation » (Allal, 2008, p. 312). Ces évaluations ont alors une fonction *sommative.* Elles peuvent se dérouler sur toute la durée d'une période de formation (par exemple, par divers « contrôles » ou « tests » ponctuels régulièrement soumis aux élèves donnant lieu à une moyenne numérique ou autre) ou uniquement à la fin (par exemple, un examen final). À noter que les évaluations sont considérées comme sommatives dès que les résultats qu'elles produisent sont pris en compte dans les décisions qui servent à reconnaitre institutionnellement les acquis des élèves dans un document officiel, permanent, public[2]. Elles sont formellement identifiées comme telles : enseignants, élèves, parents des élèves (et tout autre partenaire concerné) savent qu'elles sont associées à des enjeux forts de certification, donc de réussite ou d'échec scolaires, de sélection, d'orientation.

Cette distinction entre évaluation formative et évaluation sommative a fait couler beaucoup d'encre. Certains auteurs considèrent que cette distinction est fonctionnelle car elle permet de reconnaitre que, selon leur fonction, les évaluations des apprentissages des élèves ne répondent pas au même contrat scolaire (voir par exemple, Allal, 2008 ; Crahay, Mottier Lopez & Marcoux, 2019 ; Perrenoud, 1998).

En effet**, de façon certes un peu caricaturale**, l'évaluation formative répond à une logique de formation et d'aide à l'apprentissage. Elle *s'intéresse aux erreurs des élèves et à leurs difficultés*, afin de concevoir des dispositifs didactiques et des situations pour les surmonter. Les élèves sont ainsi invités par exemple à expliciter leurs représentations et compréhensions des activités qu'ils réalisent. L'enseignant analyse avec soin leurs réponses, en utilisant notamment des critères d'évaluation (mais pas exclusivement), dans le but de concevoir des activités de remédiation, d'approfondissement, de prolongement. Toujours dans la perspective de mieux accéder aux apprentissages des élèves (y compris des élèves de bon niveau scolaire), ces derniers sont invités à s'impliquer dans l'évaluation formative, pour s'autoévaluer, s'évaluer entre pairs, confronter leur autoévaluation à l'évaluation de l'enseignant, etc. Les dispositifs évaluatifs peuvent être très variés, y compris être intégrés dans les activités d'enseignement et d'apprentissage, c'est-à-dire sans que l'évaluation apparaisse distincte de ces activités. Leur finalité première est de permettre aux élèves de s'autoréguler et de progresser au regard des attendus de l'enseignement et, plus généralement, des programmes de formation et des enjeux pour la formation générale de la personne de l'élève (incluant des compétences sociales, métacognitives, d'organisation, etc.). Autrement dit, l'évaluation formative engage des régulations de l'enseignement et de l'apprentissage au bénéfice de l'élève avant qu'on lui attribue une note à des fins de certification. L'élève a (encore) le droit de se tromper sans conséquence négative pour lui. C'est même parce qu'il fait des erreurs que l'enseignant pourra lui proposer des dispositifs adaptés et que l'apprenant pourra progresser.

[2] C'est-à-dire destiné à être légitimement consulté par d'autres personnes que l'enseignant et l'élève concerné.

L'évaluation sommative répond, quant à elle, à une tout autre logique, surtout quand le système scolaire exige la création de *classements entre les élèves* pour les orienter dans des filières hiérarchisées : pour restreindre l'accès à certaines formations, pour qualifier et disqualifier les élèves en fonction de leurs résultats, etc. Dans un système scolaire *élitiste* qui notamment privilégie des principes associés à une justice méritocratique plutôt qu'à une justice « corrective » visant une égalité des acquis entre élèves par exemple (Crahay, 2012), la logique qui prévaut pour l'évaluation sommative est celle de la sélection et de la compétition : l'évaluation sommative sanctionne les erreurs des élèves en leur attribuant des plus ou moins bonnes notes ; elle établit des classements normatifs qui ont des conséquences tangibles sur la motivation des élèves (positivement ou négativement), sur le climat de compétition entre eux (entre établissements, etc.), sur leur devenir scolaire et plus généralement sur leur formation future et leur insertion dans le monde du travail. Dans ce cas, l'évaluation sommative est « normative » et non pas « critériée ».

Ce sont ces différents éléments qui amènent un ensemble d'auteurs à justifier la distinction à faire entre évaluation formative et évaluation sommative. À leurs yeux, les contrats scolaires sont clairement différents : l'évaluation formative invite les enseignants, les élèves, et tout partenaire impliqué (dont les parents) à un contrat de confiance, de collaboration, de relation d'aide, de « transparence ». Quant à l'évaluation sommative, comme développé par Perrenoud (1998) par exemple, les intérêts entre enseignants et élèves (et leurs parents) ne peuvent que diverger : pour les premiers, il s'agit d'établir de façon la plus précise et correcte possible les niveaux d'acquisition des connaissances et des compétences des élèves, y compris dans la perspective de trier les « méritants » des « moins méritants ». Pour les deuxièmes, il faut « tenter de faire illusion », afin d'être le moins pénalisés possible, se montrer meilleurs que les pairs. La « transparence » et la collaboration ne peuvent être qu'entachées dans l'évaluation sommative par les enjeux sélectifs.

Les parties suivantes de cette note présentent des mises en question de ce raisonnement binaire, entre la « bonne » évaluation (l'évaluation formative) et la « mauvaise » évaluation en raison des pressions et des conséquences « délétères » qu'elle (l'évaluation sommative *normative*[3], Crahay *et al*., 2019) génère. Il me paraissait important de rappeler ces premiers développements critiques. Une tendance actuelle en effet est de viser à réconcilier l'évaluation formative et l'évaluation sommative (tendance actuelle à laquelle mes travaux et ceux de mon groupe de recherche contribuent également). Sans quoi, une forme d'angélisme des propositions ci-dessous en sciences de l'éducation pourrait certainement leur être reproché.

Toujours sur un plan théorique, cette note poursuit sur des propositions développées par la littérature scientifique, qui problématisent la distinction évaluation formative / évaluation sommative initialement proposée par Bloom et ses collègues pour les apprentissages des élèves en classe. Deux ensembles de travaux en sciences de l'éducation et de la formation peuvent être identifiés : le premier concerne des travaux menés dans le monde francophone essentiellement et dont on observe une visée de spécification accrue des différentes fonctions de l'évaluation. Le deuxième ensemble, quant à lui, réunit des travaux anglo-saxons principalement qui tendent plutôt à reconceptualiser l'évaluation dans

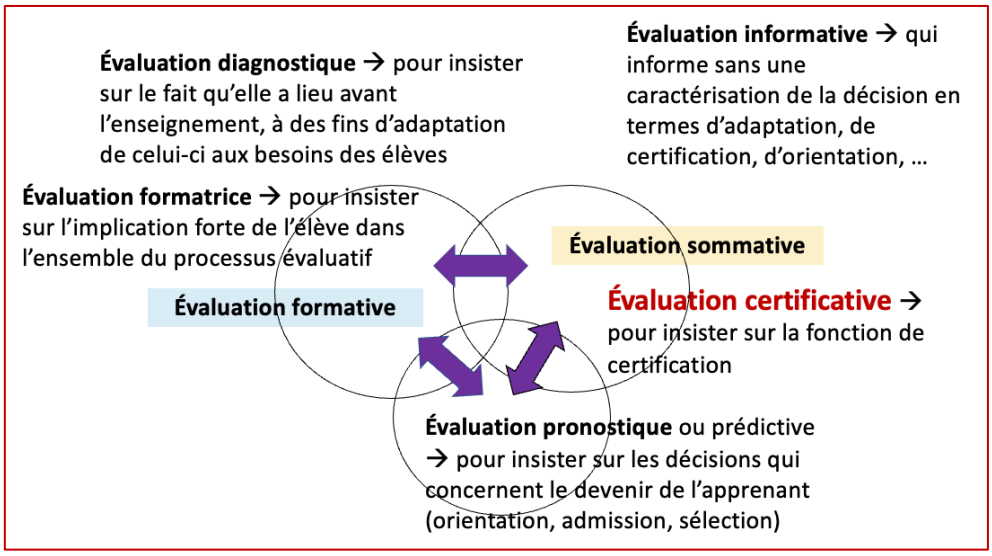
[3] Cette précision est importante, car il s'agit bien ici d'une évaluation sommative dont les résultats se fondent sur une procédure de comparaison entre les lèves. Comme défini par Crahay *et al*., (2019), « paraphrasant de Landsheere (1980, p. 295), on définira l'évaluation normative comme suit : toute performance d'élève est évaluée par comparaison avec celles d'individus placés dans la même situation afin de classer le sujet parmi ceux-ci ».

une perspective intégrative visant à dépasser les débats de fonctions. Les sections ci-dessous présentent brièvement ces deux tendances.

II. Vers une spécification toujours accrue des fonctions plurielles de l'évaluation des apprentissages des élèves

Suite aux propositions initiales de Bloom *et al*. (1971), comme je l'ai analysé dans Mottier Lopez (2015a), les auteurs francophones (voir par exemple, Allal, 2008 ; Bonniol & Vial, 1997 ; De Ketele, 2010 ; Hadji, 1989, 2021 ; Vial, 2012), se sont attelés à « décortiquer » les fonctions spécifiques de l'évaluation des apprentissages. Ils s'accordent sur le constat que les seules deux catégories « formative » et « sommative » sont *réductrices* pour rendre compte de la pluralité des buts et des fonctions de l'évaluation. Il est intéressant de les énumérer (Figure 1 ci-dessous) dans la mesure où certaines désignations se retrouvent parfois dans les discours professionnels du terrain et/ou dans des documents institutionnels actuels. Il est à noter toutefois que des variations de définitions existent entre les auteurs (par exemple pour ce qui concerne l'évaluation diagnostique et l'évaluation pronostique entre les propositions de Hadji, 2021, de De Ketele, 2010 ou encore d'Allal, 2008), observées également dans les terrains professionnels et susceptibles d'entraîner des incompréhensions.

Figure 1 : Vers une spécification des différentes fonctions de l'évaluation



La Figure 1, qui est une proposition personnelle, permet de souligner :

- les différentes fonctions (non exhaustives) introduites par la littérature francophone ;
- les proximités entre (1) évaluation formative, évaluation diagnostique, évaluation formatrice, (2) évaluation sommative et évaluation certificative ;
- les relations que les auteurs établissent entre les fonctions (figurées par les flèches et les intersections entre les cercles dans la figure 1).

Comme souligné par Crahay *et al*. (2019), dès les propositions initiales de Bloom et collaborateurs, la relation entre évaluations formative et évaluation sommative a été conceptualisée : l'évaluation formative vise à « préparer » l'évaluation sommative pour permettre la réussite du plus grand nombre d'élèves dans une perspective de lutte contre l'échec scolaire. Dans la figure 1, les intersections

invitent à considérer que les informations / résultats d'une évaluation peuvent potentiellement servir aussi à une autre évaluation, dépassant, ce faisant, les oppositions énoncées plus haut entre elles.

De Ketele (2010) invite, quant à lui, à reconsidérer plus fondamentalement la terminologie attribuée à l'évaluation, critiquant notamment la confusion qu'il considère être faite entre « fonctions » et « démarches ». Le Tableau 1 reproduit la proposition de l'auteur.

Tableau 1 : Croisement entre fonctions et démarches d'évaluation selon De Ketele (2010)

Fonctions → Démarches ↓	**Orientation** *Préparer une nouvelle action*	**Régulation** *Améliorer une action en cours*	**Certification** *Certifier le résultat d'une action*
Sommative *Mettre une note*	Ex. 1 : Attribuer une note à un test de prérequis langagiers Ex. 2 : Orienter une étudiant sur la base d'un dossier de notes scolaires (comme au terme du collège)	Ex. 1 : Attribuer une note à une interrogation en cours d'apprentissage sur les règles d'accord Ex. 2 : Noter sur 10 points un « quizz test » de 10 questions de compréhension sur un texte en cours d'analyse	Ex. 1 : Attribuer une note à une composition française Ex. 2 : Noter en fin d'année un test composite (analyse, conjugaison, etc.)
Descriptive *Identifier et décrire*	Ex. 1 : Identifier, parmi 10 prérequis de la lecture, ceux qui sont déjà maîtrisés par un élève du préscolaire Ex. 2 : Identifier dans les questions d'examen de français de l'an dernier les questions qui posent problème	Ex. 1 : Repérer dans la classe les élèves qui n'ont pas réussi un exercice individuel de conjugaison pour organiser un tutorat Ex. 2 : Organiser un test diagnostique sur un apprentissage important	Ex. 1 : Remplir un bulletin descriptif (cocher dans une liste d'objectifs du cours de français ceux qui sont atteints de façon stable) Ex. 2 : Rédiger un certificat mentionnant et décrivant les compétences langagières acquises
Herméneutique *Donner un sens*	Ex. 1 : Orienter un étudiant vers une filière littéraire sur la base de nombreux indices (résultats scolaires, observations, entretien) Ex. 2 : Analyser finement en début d'année une première dictée et un premier texte libre des nouveaux élèves en vue de préparer les premières séquences	Ex. 1 : Observer et analyser les productions des élèves pendant une séquence de travail individuel en classe Ex. 2 : Analyser les résultats d'un test diagnostique à la lumière d'une analyse de la façon dont les séquences d'apprentissage antérieures ont été réalisées	Ex. 1 : Préparer une réunion de parents sur la base d'un portfolio des productions significatives de ses élèves Ex. 2 : Certifier le passage de classe, non seulement sur la base des notes obtenues en fin d'année, mais en tenant compte d'autres indices pertinents

De ce tableau, on retiendra notamment que c'est la désignation « évaluation certificative » qui est privilégiée plutôt qu'évaluation sommative. D'une part, elle désigne plus précisément la fonction de cette évaluation (certifier institutionnellement les acquis des élèves), alors que, d'autre part, le terme « sommatif » est associé à l'idée de faire un bilan en tant que démarche pouvant servir les fonctions d'orientation, de régulation, de certification. Quant à l'évaluation formative, elle apparait liée à la

fonction de régulation. La fonction d'orientation est quant à elle associée tant à l'évaluation diagnostique qu'à l'évaluation pronostique (voir Figure 1 pour un rappel de celles-ci).

On retiendra de ce premier ensemble de travaux que les auteurs ont argumenté en faveur de terminologies plus spécifiques et pointues, au point peut-être de risquer un certain *éclatement* entre les différentes fonctions, types, démarches évaluatives. La proposition de De Ketele montre sa préoccupation à penser les liens entre les éléments en faveur d'une approche plus intégrative et systémique, dépassant largement la dichotomie « formatif – sommatif » initiale.

Sur le plan conceptuel, ce travail de spécification théorique rend compte (en partie) de la complexité des pratiques évaluatives et des enjeux scientifiques pour tenter de toujours mieux les comprendre et les caractériser. Des résultats de recherches tant anglophones que francophones nous alertent cependant : ces distinctions conceptuelles restent peu opérationnelles dans les pratiques en classe (voir par exemple, Abernot, 2013 ; Black & Wiliam, 1998 ; Mottier Lopez, 2015b ; Perrenoud, 1991). Par exemple, au regard des cas de figure suivants (non exhaustifs) :

- Les enseignants tendent à ne pas faire la différence entre les fonctions de l'évaluation : par exemple, une évaluation est vue comme pouvant servir à « mettre une note » et à la fois à être utilisée pour envisager des activités de remédiation, pour planifier une prochaine séquence d'enseignement dans un but formatif *a posteriori*. Une même évaluation sert plusieurs fonctions : « pourquoi alors les différencier ? »[4]. Cette indifférenciation dans la pratique prend le dessus sur les conceptualisations théoriques qui font peu sens face à la réalité des mises en œuvre *in situ*.

- L'évaluation reste principalement perçue et représentée au regard de ses enjeux forts de certification, voire d'orientation et de sélection. L'évaluation formative est peu ou n'est pas pratiquée ou, quand elle est interactive et intégrée dans les situations d'enseignement et d'apprentissage (voir par exemple, Allal & Mottier Lopez, 2005 ; Mottier Lopez, 2015b), elle n'est pas vue comme étant une évaluation par les enseignants : « c'est pas évaluer, c'est enseigner ! ». La représentation de l'évaluation ici reste celle d'une pratique formelle, ponctuelle, instrumentée, débouchant nécessairement sur une « note » (qualitative ou quantitative). Tout ce qui s'en écarte n'est pas conçu comme une évaluation. La conséquence est que l'évaluation formative dite « élargie » (intégrant des modalités interactives informelles, impliquant l'élève dans des démarches d'autoévaluation au sens large, etc. voir Allal & Mottier Lopez, 2005 pour une description étayée) n'est pas reconnue en tant que telle.

- Les enseignants considèrent que les fonctions formatives et sommatives / certificatives sont inconciliables ; « parfois on doit être coach, parfois on doit être juge ! » et ils ne le vivent pas bien. Faire cohabiter des fonctions différentes de l'évaluation dans la classe exige de parvenir à gérer des tensions complexes, parfois paradoxales, notamment pour ce qui concerne les enjeux de certification et d'orientation (Cardinet, 1984 ; Harlen, 2007 ; Laveault, 2008).

Est-ce alors une bonne « idée » de différencier de façon aussi pointue les différentes fonctions et démarches évaluatives, ou serait-il préférable de concevoir des approches plus intégratives, caractérisée par la *finalité éducative visée* plus que par les différentes décisions qui découlent de

[4] J'insère ici des exemples de phrases souvent entendues dans mes interventions de formation ou de recherches collaboratives. À noter dans le cas de figure rapporté ici que l'aide à l'apprentissage de l'élève se fait alors après la note…

l'évaluation (régulation, certification, orientation) ? C'est l'option tendancielle retenue par les travaux cités ci-dessous.

III. Vers des propositions de reconfiguration conceptuelle visant une « évaluation pour apprendre durablement »

Depuis les années 1990, un ensemble de travaux essentiellement anglophones ont tenté de ne plus opposer les différentes fonctions de l'évaluation, en faveur d'approches plus intégratives et systémiques. De nouvelles « modélisations » ont été proposées, synthétisées dans le Tableau 2 ci-dessous.

Tableau 2 : Propositions conceptuelles visant à reconfigurer l'évaluation des apprentissages (Mottier Lopez, Girardet & Naji, 2021, p. 6, augmenté)[5]

Terme en anglais	Auteurs à l'origine	Traduction en français	Dimensions principales mises en évidence
Assessment for learning	*Assessment Reform Group* (Black & Wiliam, 1998)	Évaluation pour apprendre Évaluation soutien d'apprentissage (Allal & Laveault, 2009)	Englobe toutes formes d'évaluations (formelles et informelles, formatives et sommatives) susceptibles d'encourager les élèves à s'engager dans des apprentissages à des fins de régulation et d'autorégulation ; fortement intégrées aux activités d'enseignement et d'apprentissage.
Authentic assessment	Wiggins (1998)	Évaluation authentique	Les contenus, la structure, les caractéristiques et la validité des évaluations s'associent à des situations complexes du monde « réel », sans distinction entre évaluations formatives et certificatives.
Sustainable assessment	Boud (2000)	Évaluation durable	Les évaluations, quelles que soient leurs formes et leurs fonctions, répondent aux besoins d'apprentissage dans des temporalités courtes et longues (apprentissage tout au long de la vie).
Assessment as learning	Earl (2003)	Évaluation comme apprentissage	Les évaluations sont vues comme susceptibles de représenter des occasions d'apprentissage, particulièrement quand les élèves sont impliqués dans des autoévaluations, des évaluations entre pairs et des évaluations collaboratives formatives. L'évaluation devient objet d'apprentissage.
Learning-oriented assessment	Carless, Joughin et Mok (2006)	Évaluation axée sur l'apprentissage	Des évaluations qui cherchent à réconcilier les fonctions formatives et sommatives pour que toutes orientent l'apprentissage des élèves dans des directions productives.

Un point commun à toutes ces propositions conceptuelles, tant pour les évaluations formatives que sommatives/certificatives, est de les concevoir toutes comme des leviers potentiellement *positifs* pour soutenir des apprentissages de qualité chez les élèves. Quelle que soit sa fonction, l'évaluation se veut « constructive » et « soutenante » pour les apprentissages des élèves et leur motivation à apprendre, sur des temporalités ponctuelles et plus longues. Comme écrit dans Mottier *et al*. (2021), à quelques

[5] Les références citées dans le tableau sont celles, notamment, à l'origine des modèles exposés. D'autres contributions ont ensuite largement développé les propositions y compris dans la littérature scientifique actuelle. Je ne les cite pas ici. Elles sont très nombreuses.

nuances près, les auteurs francophones et anglophones dans la littérature *actuelle*[6] s'accordent alors sur un ensemble de caractéristiques à problématiser. Notamment :

- La cohérence entre ce qui est enseigné, appris et évalué pour construire un « alignement curriculaire » cohérent, y compris avec les didactiques disciplinaires et les prescriptions institutionnelles ;
- La nature des tâches d'apprentissage qui sont susceptibles d'intégrer des évaluations associées à des possibilités de régulation et d'autorégulation des élèves, qui encouragent des collaborations entre eux et avec l'enseignant ;
- Les contrats évaluatifs en jeu, les attentes, les objectifs, les critères d'évaluation, y compris pour renforcer la confiance mutuelle entre apprenants et avec les enseignants au regard des différents enjeux de l'évaluation ;
- L'implication active des élèves dans des démarches d'autoévaluation et d'évaluation entre pairs, y compris dans la perspective de développer chez les élèves des compétences évaluatives, de jugement critique et d'autorégulation ;
- La façon de donner de l'importance et un rôle à la « voix des élèves » pour les considérer comme des partenaires à part entière dans l'évaluation et ses enjeux ;
- Les gestes professionnels de l'enseignant, ses postures évaluatives, son jugement professionnel et les modérations sociales entre professionnels pour construire des interprétations et des décisions évaluatives les plus pertinentes et valides possibles ;
- Les opportunités de rétroaction (feedback) de la part des enseignants et entre élèves dans la perspective de soutenir les apprentissages dans les tâches concernées, incluant les processus de régulation, de co-régulation, d'autorégulation ;
- Les cultures de classe et d'établissement scolaires, construites avec les élèves et l'ensemble des partenaires concernés, en faveur « d'évaluations positives pour l'avenir » (*sustainable assessment*), c'est-à-dire dans la perspective d'une évaluation « pour apprendre tout au long de la vie ».

La force des propositions issues de cette littérature est notamment qu'elles se fondent sur des données empiriques produites avec des protocoles de recherche expérimentaux, des observations, des enquêtes, des analyses comparatives, des interventions.

[6] Très nombreux, je ne suis pas en mesure de les citer dans cette note brève. Je propose par exemple de consulter l'ouvrage collectif de Barroso da Costa, Nizet et Leduc (2021) qui propose des synthèses étayées de quarante ans de travaux en évaluation.

Références

Allal, L. (2008). Évaluation des apprentissages. In A. van Zanten (Ed.), *Dictionnaire de l'éducation* (pp. 311-314). Presses Universitaires de France.

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

Allal, L. & Mottier Lopez, L. (2005). L'évaluation formative de l'apprentissage : revue de publications en langue française. Dans OCDE, *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires* (p. 265-290). Éditions OCDE/CERI.

Amigues, R. & Zerbato-Poudou, M.-T. (1996). *Les pratiques scolaires d'apprentissage et d'évaluation.* Dunod.

Barroso da Costa, C., Nizet, I. & Leduc, D. (Eds.) (2021). *Quarante ans de mesure et d'évaluation*. Presses Universitaires du Québec.

Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-148. http://electronicportfolios.org/afl/InsideBlackBox.pdf

Bloom, B. S., Hasting, J. T. & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. McGraw-Hill.

Bonniol, J.-J. & Vial, M. (1997). *Les modèles de l'évaluation.* De Boeck.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151-167. doi: 10.1080/71369572

Cardinet, J. (1984). *Pour apprécier le travail des élèves.* Institut romand de recherches et de documentation pédagogiques.

Crahay, M. (2012). *L'école peut-elle être juste et efficace ?* De Boeck.

Crahay, M., Mottier Lopez, L. & Marcoux, G. (2019). L'évaluation des élèves : Docteur Jekyll and Mister Hyde de l'enseignement. In M. Crahay (Ed.), *Peut-on lutter contre l'échec scolaire ?* (pp. 358-425). De Boeck.

De Ketele, J. (2010). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée, XV* (1), 25-37. doi:10.3917/rfla.151.0025.

Earl, L. M. (2003). *Assessment as Learning. Using assessment to maximize student learning.* Hawker Brownlow Education.

Hadji, C. (1989). *L'évaluation, règles du jeu.* ESF éditions.

Hadji, C. (2012). Faut-il avoir peur de l'évaluation ? De Boeck.

Hadji, C. (2012). Le défi d'une évaluation à visage humain. ESF.

Harlen, W. (2007). *Assessment of learning.* SAGE publications ltd.

Laveault, D. (2008). Le jugement professionnel : foyer de tensions et de synergies nouvelles en évaluation scolaire. *Revue suisse des sciences de l'éducation*, *30* (3), 483-500.

Mottier Lopez, L. (2015a). *Évaluations formative et certificative : enjeux pour l'enseignement*. De Boeck.

Mottier Lopez, L. (2015b). Évaluation-régulation interactive : étude des structures de participation guidée entre enseignant et élèves dans le problème mathématique 'Enclos de la chèvre'. *Mesure et évaluation en éducation, 38*(1), 89-120.

Mottier Lopez, L. (2015c). L'évaluation formative des apprentissages des élèves : entre innovations, échecs et possibles renouveaux par des recherches participatives. *Questions Vives, 23*, 31-50.

Mottier Lopez, L., Girardet, C. & Naji, T. (2021). L'évaluation continue pour apprendre : enjeux de la pluralité des feedbacks entre pairs dans un cours universitaire. *Mesure et évaluation en éducation*, *44*(2), 1-34.

Perrenoud, P. (1991). Pour une approche pragmatique de l'évaluation formative**.** *Mesure et évaluation en éducation*, *13*(4), 49-81.

Perrenoud, P. (1998). *L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages*. De Boeck.

Vial, M. (2012). *Se repérer dans les modèles de l'évaluation.* De Boeck.

Wiggins*,* G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance.* Jossey-Bass.

La politique évaluative de la France : quelles évolutions engagées depuis 2014 et quels impacts dans les classes ?

BRIGITTE HAZARD

INSPECTION GÉNÉRALE DE L'ÉDUCATION, DU SPORT ET DE LA RECHERCHE (IGÉSR)

Introduction

Définissons le périmètre de notre contribution :

- **Temporel :** une conférence nationale sur l'évaluation des élèves a déjà eu lieu en décembre 2014, l'étude partira donc de cette période, ceci d'autant que c'est concomitant de la loi d'orientation et de programmation pour la refondation de l'école de la République du 8 juillet 2013 dont découlera le nouveau socle commun et de nouvelles modalités d'évaluation et de notation des élèves ;

- **Spatial :** faire de l'évaluation des acquis des élèves une donnée essentielle de l'évaluation de l'École invite à prendre en considération les différentes modalités d'évaluation des élèves : les évaluations standardisées[7], les évaluations dans la classe et les examens, le diplôme national du brevet (DNB) et le certificat de formation générale (CFG) pour le segment obligatoire. Mais ici, nous parlons de l'évaluation en classe donc nous limiterons nos propos aux pratiques d'évaluation des enseignants en classe, en nous concentrant sur l'enseignement obligatoire[8]. Seront donc exclues les évaluations standardisées sauf à les convoquer par l'exploitation de leurs résultats par les enseignants au service des apprentissages des élèves. Nous ferons de fait une incursion au lycée puisque les élèves de 16 ans révolus sont très majoritairement scolarisés en lycées professionnels, généraux et technologiques. *Quid* des examens ? Parler politique des examens en France c'est, au collège comme au lycée, parler de contrôle continu et de contrôle en cours de formation ; et ainsi pouvoir faire des évaluations concernées, que d'aucuns qualifient de « certificatives », des évaluations, elles aussi au service des apprentissages des élèves tant en collège qu'en lycée, dès lors que ces examens ne sont pas des épreuves terminales, épreuves que nous exclurons.

- **Autre délimitation : qui porte le regard ? Qui met en œuvre et exploite ? Qui en bénéficie ?** Nous prendrons comme observateur privilégié l'IGÉSR au sens de sa connaissance de ce qui se passe réellement dans les classes, elle-même nourrie de tous ses interlocuteurs habituels et de tout son

[7] Avec deux familles : les évaluations dites sur échantillons nationales et internationales et les évaluations nationales dites exhaustives développées récemment, dont la fonction est différente puisqu'elles visent à éclairer directement le professeur dans son action.

[8] Depuis la loi Jules Ferry du 28 mars 1882, l'instruction est obligatoire. Cette obligation s'applique à partir de 3 ans depuis la rentrée 2019, pour tous les enfants français ou étrangers résidant en France. À l'origine, l'instruction était obligatoire jusqu'à l'âge de 13 ans, puis 14 ans à partir de la loi du 9 août 1936. Depuis l'ordonnance n°59-45 du 6 janvier 1959, elle a été prolongée jusqu'à l'âge de 16 ans révolus.

état de l'art dès lors qu'il y a une mission et un rapport. Nous évoquerons les acteurs que sont les enseignants mais aussi les cadres, chefs d'établissement et inspecteurs ainsi que les usagers, élèves et leurs familles[9].

- **En termes de définitions**, **nous retenons celles arrêtées dans les textes officiels et par le Conseil d'évaluation de l'école** (CEE, 2022) dont :
    - l'évaluation en classe : ensemble des évaluations réalisées dans le cadre des enseignements. Les finalités (diagnostiques, formatives, sommatives), les modalités (individuelles/collectives, écrites/orales, en classe/à la maison, etc.) et la fréquence sont définies par chaque professeur ou équipe pédagogique.
    - La part des évaluations dans la classe dont les résultats sont pris en compte dans l'attribution d'une évaluation certificative (examen, attestation) correspond au contrôle continu.
- L'évaluation considérée dans cette note est celle portant sur les acquis scolaires des élèves, c'est-à-dire les connaissances et compétences visées par les programmes d'enseignement et les contenus de formation. La question du socle se pose également, en particulier concernant les compétences visées par les domaines 2 et 3 qui sont transversales et qui transcendent souvent les disciplines. Cette évaluation est un enjeu essentiel de l'École qui renvoie à plusieurs finalités majeures des politiques éducatives dont deux que nous considérerons ici, celle de **la formation des élèves (évaluer pour former)** et celle de **l'information des élèves et de leur famille (évaluer pour informer)**. L'évaluation des acquis des élèves est aussi au fondement de la réussite scolaire des élèves et de la capacité de l'École à assurer l'équité scolaire[10] : nous regarderons dans quelle mesure l'évaluation des acquis des élèves pratiquée en classe est un moyen de garantir et de renforcer cette équité.

I. Les politiques évaluatives en France depuis 2014[11]

Le rapport de l'inspection générale *Les acquis des élèves, pierre de touche de la valeur de l'école ?* (IGEN & IGAENR, 2005) montre que « l'évaluation », terme non encore utilisé en tant que tel jusqu'aux années soixante-dix, se confond avec la notation et le classement, pour, ensuite, progressivement s'enrichir et intégrer de nouvelles modalités, des critères d'évaluation, des appréciations, etc. Cependant, le même rapport écrit « Il est peu encourageant d'étudier à nouveau aujourd'hui l'évaluation des élèves tant est faible l'impact sur les pratiques professionnelles des travaux menés par le passé ». Ce constat de 2005 montre que cette question de l'évaluation des élèves (ici de leurs acquis)

[9] Familles qui scolarisent leurs enfants dans un établissement scolaire public ou privé et non celles qui assurent l'instruction des enfants elles-mêmes (avec déclaration préalable).

[10] L'équité scolaire est ainsi définie par le Conseil de l'évaluation de l'École : finalité de politique éducative dans le cadre de laquelle le système éducatif fournit les mêmes chances d'apprendre à tous les élèves. Par équité, on n'entend pas l'obtention des mêmes résultats par tous les élèves, mais plutôt l'absence de lien entre les différences de résultats entre les élèves et le milieu dont ils sont issus ou les facteurs économiques et sociaux sur lesquels ils ne peuvent exercer aucun contrôle (genre, origine sociale, lieu d'habitation, etc.).

[11] En appui sur la loi de la refondation de 2013 puis sur la loi pour une école de la confiance en 2019.

est « un angle mort »[12] qu'il est utile d'analyser de plus près dans les pratiques professionnelles des enseignants.

A. La politique évaluative du cycle 1 à la fin du cycle 4

Depuis 2005, les lois d'orientation de 2005, 2013 et 2019 ont conduit à une évolution de nombreux éléments structurant de l'École : mise en place du socle commun de connaissances, de compétences et de culture, création du Conseil supérieur des programmes (CSP), création du Conseil de l'évaluation de l'École (CEE), création d'une direction du numérique de l'éducation (DNE), redéploiement des évaluations standardisées.

Cette politique ministérielle volontaire relative à l'évaluation des élèves donne un cadre aux pratiques d'enseignement et aux usages de l'évaluation :

- Dans la loi de 2013, on peut lire :

> *« Faire évoluer les modalités d'évaluation et de notation des élèves : Les modalités de la notation des élèves doivent évoluer pour éviter une « notation-sanction » à faible valeur pédagogique et privilégier une évaluation positive, simple et lisible, valorisant les progrès, encourageant les initiatives et compréhensible par les familles. En tout état de cause, l'évaluation doit permettre de mesurer le degré d'acquisition des connaissances et des compétences ainsi que la progression de l'élève. Il faut aussi remédier à la difficulté pour les enseignants d'évaluer les élèves avec des dispositifs lourds et peu coordonnés entre eux. Ainsi, l'évolution des modalités de notation passe notamment par une réforme du livret personnel de compétences actuel, qui est trop complexe, et une diversification des modalités de l'évaluation […] Dans l'enseignement primaire, l'évaluation sert à mesurer la progression de l'acquisition des compétences et des connaissances de chaque élève. Cette logique d'évaluation est aussi encouragée dans l'enseignement secondaire. »* (Loi n°2013-595 du 8 juillet 2013 d'orientation et de programmation pour la refondation de l'école de la République)

- Dans la loi de 2019, le troisième alinéa de l'article L. 311-1 du code de l'éducation est ainsi rédigé : « L'évaluation sert à mesurer et à valoriser la progression de l'acquisition des compétences et des connaissances de chaque élève ». Cet alinéa n'a pas été modifié ensuite (Loi n° 2019-791 du 26 juillet 2019 pour une école de la confiance).

Les enseignements du premier degré et du collège ont ainsi connu des bouleversements importants quant à l'évaluation des élèves, principalement par la création du socle commun de compétences et de connaissances en 2006 (Loi d'orientation de de programme pour l'avenir de l'école du 23 avril 2005). Il a induit très progressivement un nouveau regard sur les apprentissages par l'introduction des compétences et l'établissement de livrets de suivi. Le livret personnel de compétences a été remis en cause du fait de sa complexité opérationnelle en 2012 et remplacé par un livret scolaire unique (LSU) (Arrêté du 31 décembre 2015 fixant le contenu du livret scolaire de l'école élémentaire et du collège ; devenu le livret scolaire numérique (LSN) en 2017) suite à la mise en place d'un nouveau socle de compétences[13], de connaissances et de culture en 2013, avec la loi de refondation de l'école de la

[12] Expression reprise dans le titre d'un article de la revue *Profession Éducation* : « L'évaluation, angle mort des politiques éducatives » (Sgen-Cfdt, 2020).

[13] Compétence : dans le cadre du socle commun de connaissances, de compétences et de culture, « une compétence est l'aptitude à mobiliser ses ressources (connaissances, capacités, attitudes) pour accomplir une tâche ou faire face à une

République de 2013 (Loi n°2013-595 du 8 juillet 2013 d'orientation et de programmation pour la refondation de l'école de la République). La mise en œuvre du LSU repose sur une évaluation avec quatre niveaux de maîtrise appliqués aux cinq domaines du socle[14] et, en fin de parcours, une validation du socle prise en compte dans l'obtention du diplôme national du brevet (lequel est fondé sur des épreuves ponctuelles d'examen et des notes de contrôle continu).

Au moment de l'écriture du socle de 2015 et de sa présentation, le Conseil supérieur des programmes a rédigé sept principes[15] de cadrage général relatifs à l'évaluation des acquis du socle, dont :

- donner un statut clair à l'évaluation, notamment formative, et éviter qu'elle ne connaisse une expansion abusive aux dépens des enseignements eux-mêmes. Proscrire la multiplication excessive des items évalués ;
- définir des procédures garantissant la qualité, la précision et l'équité des évaluations afin que l'élève puisse au mieux progresser à partir de la mise en évidence nuancée de ce qu'il maîtrise et de ce qu'il ne parvient pas encore à réaliser ;
- éviter les calculs artificiels de moyennes, qui font perdre le sens du projet global de formation.

Dans le texte officiel du socle (Décret n°2015-372 du 31 mars 2015 relatif au socle commun de connaissances, de compétences et de culture), on peut lire que la logique du socle commun implique une acquisition progressive et continue des connaissances et des compétences par l'élève, comme le rappelle l'intitulé des cycles d'enseignement de la scolarité obligatoire que le socle commun oriente : cycle 2 des apprentissages fondamentaux, cycle 3 de consolidation, cycle 4 des approfondissements. Ainsi, la maîtrise des acquis du socle commun doit se concevoir dans le cadre du parcours scolaire de l'élève et en référence aux attendus et objectifs de formation présentés par les programmes de chaque cycle. La vérification de cette maîtrise progressive est faite tout au long du parcours scolaire et en particulier à la fin de chaque cycle. Cela contribue à un suivi des apprentissages de l'élève.

Le socle doit permettre aussi l'acquisition des méthodes pour apprendre, la poursuite d'études, la construction d'un avenir personnel et professionnel et préparer à l'exercice de la citoyenneté.

Ce qui complète cette incitation ou cet équipement des enseignants pour encadrer leurs pratiques évaluatives, ce sont les contenus du **livret scolaire** définis dans le décret n°2015-1929 du 31 décembre 2015[16].

<u>En application des dispositions de la loi n 2013-595 du 8 juillet 2013 d'orientation et de programmation pour la refondation de l'école de la République</u>, le décret vise à **faire évoluer et à diversifier les**

situation complexe ou inédite » (Décret n° 2015-372 du 31 mars 2015 relatif au socle commun de connaissances, de compétences et de culture).

[14] Dans le domaine de formation intitulé « les langages pour penser et communiquer », l'évaluation distingue quatre composantes : langue française ; langues étrangères et, le cas échéant, langues régionales ; langages mathématiques, scientifiques et informatiques ; langages des arts et du corps. Il y a donc en réalité 8 composantes à évaluer.

[15] Les autres ne sont pas énoncés ici car ils ont trouvé une place dans la rédaction même du livret scolaire (privilégier les mécanismes d'évaluation qui constituent en même temps un outil favorisant une réflexion collégiale dans les écoles et les établissements) et du diplôme national du brevet (DNB) (4 niveaux de maîtrise et non une évaluation binaire, pas de compensation entre les domaines, une prise en considération de toutes les compétences définies dans le socle commun et les programmes sans se contenter de celles évaluées dans les épreuves classiques).

[16] Le décret n° 2015-1929 du 31 décembre 2015 relatif à l'évaluation des acquis scolaires des élèves et au livret scolaire, à l'école et au collège précise le cadre du livret scolaire et l'échelle d'évaluation des composantes du socle commun de connaissances, de compétences et de culture à partir de la rentrée 2016.

modalités de notation et d'évaluation des élèves de l'école primaire et du collège pour éviter une « notation-sanction » à faible valeur pédagogique et privilégier une évaluation positive, simple et lisible, valorisant les progrès, encourageant les initiatives et compréhensible par les familles. Cet article relatif aux familles précise également que « cette information se fait notamment par l'intermédiaire du carnet de suivi des apprentissages à l'école maternelle, du livret scolaire à l'école élémentaire et au collège, ainsi que du bulletin et du livret scolaires dans les lycées *».* Enfin le décret précise que l'évaluation doit aussi permettre de **mesurer le degré d'acquisition des connaissances et des compétences ainsi que la progression de l'élève.**

L'arrêté du 31 décembre 2015 modifié fixant le contenu du livret scolaire de l'école élémentaire et du collège précise le contenu des bilans et l'emploi de l'application nationale[17].

Ainsi, le livret scolaire de la scolarité obligatoire, unique du CP à la troisième comprend des **bilans périodiques**, qui se substituent aux anciens livrets des écoles et aux bulletins des collèges, **des bilans de fin de cycle** et les **attestations** officielles obtenues par l'élève à l'école ou au collège. **L'application nationale « Livret scolaire unique du CP à la troisième »** permet, pour la première fois, de disposer d'un outil numérique national de suivi des acquis de l'élève tout au long de chaque cycle et de l'ensemble de sa scolarité à l'école et au collège.

Il ne faut pas omettre ici, en termes de politique évaluative, celle qui passe par l'examen, ici le DNB qui atteste en particulier de la maîtrise du socle :

> *« L'obtention du diplôme national du brevet (DNB) dépend à la fois des points obtenus par l'élève dans la maîtrise du socle commun de connaissances, de compétences et de culture, appréciée par les équipes pédagogiques à la fin du cycle 4, et de ses résultats aux épreuves finales de l'examen ».* (Site Éduscol, 2020) [18]

Tout ce qui relève des conseils ou des principes d'action pour évaluer les acquis des élèves ne se lit que dans des documents proposés sur Éduscol[19] ou des circulaires, dont les circulaires de rentrée qui ont accompagné la mise en œuvre du socle, du livret et des nouvelles modalités du DNB[20].

En effet, si la maîtrise des compétences du socle commun s'évalue sur la base des connaissances et compétences fixées par les programmes d'enseignement, et si les attendus de fin de cycle précisés dans les programmes donnent des repères nécessaires pour apprécier le degré d'acquisition des connaissances et des compétences ainsi que la progression de chaque élève au cours du cycle, les modalités de l'évaluation sont laissées à l'appréciation des équipes :

- au cours du cycle, celles-ci constituant un objet de travail essentiel pour les conseils de cycle, à l'école primaire, ou pour le conseil pédagogique, au collège ;

[17] Article 8 : Les éléments constitutifs du livret scolaire, définis à l'article 1er, sont numérisés dans une application informatique nationale, dénommée livret scolaire unique numérique.

[18] Pour plus de détails, voir le décret n°2015-1929 du 31-12-2015 relatif à l'évaluation des acquis scolaires des élèves et l'arrêté du 31 décembre 2015 modifié relatif aux modalités d'attribution du DNB.

[19] Deux documents sont notamment disponibles depuis le site https://eduscol.education.fr/141/modalites-d-evaluation-des-acquis-scolaires-des-eleves : « Principes d'action pour évaluer les acquis des élèves » ; « Évaluer la maîtrise du socle commun aux cycles 2, 3 et 4 ».

[20] Circulaires de rentrée de 2016 et de 2017.

- en fin de cycle, celles-ci permettant d'apprécier le niveau de maîtrise de chacune des composantes du socle commun.

En effet, les premières observations réalisées par l'Inspection générale dès 2016-2017 ont montré le grand désarroi des enseignants face à l'expertise non accompagnée qui leur était demandée, celle de mesurer des acquisitions de compétences et de connaissances formulées dans leurs programmes sans repères de niveaux de maîtrise et dont l'attendu en termes de niveaux de maîtrise était référé à des domaines du socle et non des contenus de programmes d'enseignement.

La Dgesco a permis de mettre rapidement au travail des équipes d'enseignants et d'inspecteurs pour produire des documents d'accompagnement pour l'évaluation des acquis du socle commun de connaissances, de compétences et de culture, éléments pour l'appréciation du niveau de maîtrise satisfaisant en fin de cycle 2, 3 et 4[21].

Encadré : le cycle 1

Au cycle 1, conformément aux recommandations du programme[22], l'évaluation se fonde sur l'observation et vise avant tout les progrès des élèves.

Extrait de la page Éduscol : https://eduscol.education.fr/107/suivi-et-evaluation-des-apprentissages-des-eleves-l-ecole-maternelle

« À l'école maternelle, deux outils permettent d'assurer le suivi des apprentissages et des progrès des élèves : le carnet de suivi des apprentissages, renseigné tout au long du cycle, et la synthèse des acquis de l'élève, établie à la fin de la dernière année du cycle 1. Le carnet de suivi des apprentissages est communiqué aux parents ou au responsable légal de l'élève selon une fréquence adaptée à l'âge de l'enfant. »

Le document de synthèse des acquis scolaires renseigné par l'équipe pédagogique du cycle 1 est transmis à l'école élémentaire lors de l'admission de l'élève en première année du cycle 2, et communiqué aux parents ou au responsable légal de l'élève.

Ainsi **la loi de refondation de l'école de la République de 2013** a ouvert à partir de l'année 2015 un grand chantier pédagogique associant conjointement l'évolution des pratiques d'évaluation des élèves, la définition du nouveau socle commun et la refonte des programmes de la scolarité obligatoire ainsi qu'une formation repensée des enseignants. Elle engage à une évaluation positive, simple et lisible, compréhensible des familles, permettant de mesurer le degré d'acquisition des connaissances et des compétences ainsi que la progression de chaque élève et allégeant les dispositifs d'évaluation des enseignants.

[21] Voir notamment le document d'accompagnement disponible à cette adresse : https://eduscol.education.fr/document/17320/download

[22] Extrait du programme du cycle 1 : « Une école qui pratique une évaluation positive. L'évaluation constitue un outil de régulation dans l'activité professionnelle des enseignants ; elle n'est pas un instrument de prédiction ni de sélection. Elle repose sur une observation attentive et une interprétation de ce que chaque enfant dit ou fait. Chaque enseignant s'attache à mettre en valeur, au-delà du résultat obtenu, le cheminement de l'enfant et les progrès qu'il fait par rapport à lui-même. »

Le référentiel des compétences des métiers du professorat de 2013 (Arrêté du 1-7-2013 relatif au référentiel des compétences professionnelles des métiers du professorat et de l'éducation), par la compétence à « évaluer les progrès et les acquisitions des élèves » formalise qu'évaluer est un métier.

B. La politique évaluative en lycée

Si on quitte l'école et le collège pour évoquer <u>le lycée général et technologique</u> il faut attendre la réforme engagée dès la rentrée 2018 pour parler d'évaluation mais à travers l'ambition affichée de la réforme du baccalauréat, qui est de valoriser le travail des lycéens et sa régularité, de mieux accompagner les élèves dans la conception de leur projet d'orientation et de servir de tremplin vers la réussite dans le supérieur. C'est pour servir ce dessein que, dans la voie générale, les séries ont été abandonnées et qu'une part inédite de contrôle continu (40 % de la note finale) a été introduite pour l'obtention du diplôme.

L'introduction d'un « grand oral » au rang des épreuves finales en fin de terminale vient parachever une réforme qui vise à donner du sens à la formation dispensée au lycée.

L'introduction du contrôle continu qui incite chaque élève à s'impliquer toute l'année dans ses apprentissages et celle d'un grand oral impliquent une évolution dans les approches évaluatives des enseignants et la formation des élèves, en particulier une réflexion collective sur les pratiques d'évaluation au lycée.

La note de service de la Dgesco du 28 juillet 2021 (Note de service du 28-7-2021 portant sur les modalités d'évaluation des candidats à compter de la session 2022), en recommandant que les conditions d'organisation du contrôle continu du baccalauréat soient définies par un projet d'évaluation pour l'établissement, marque une évolution vers une concertation renforcée – au sein et entre les disciplines – et une définition collective des principes d'évaluation des élèves au lycée. **Le projet d'évaluation** ne définit pas de quelle manière (note, grille de compétences…) il convient d'évaluer les acquis et les travaux des élèves mais vise à ce que les professeurs expliquent leur évaluation aux élèves et à leurs familles. Ce cadre vise à garantir la valeur du diplôme du baccalauréat, en portant une attention collective aux moyennes qui ont une valeur certificative**.**

Jusqu'à la parution au printemps 2021 du guide pour le contrôle continu en terminale en vue du baccalauréat 2021[23], fruit d'un travail de l'IGÉSR[24], le contrôle continu était institutionnellement assez peu prescrit dans le voie générale et technologique alors qu'il joue un rôle majeur dans les procédures d'orientation et dans la vie des élèves. Le guide de l'évaluation au service des apprentissages des lycéens qui s'inscrit dans le prolongement et en complément de la note de service, et élaboré par le collège « expertise disciplinaire et pédagogique » de l'IGÉSR, apporte des recommandations, **dans le respect de la liberté pédagogique de chaque enseignant[25]**. Celles-ci complètent les programmes officiels qui énoncent uniquement les compétences et les connaissances à enseigner.

---

[23] Disponible à cette adresse : https://eduscol.education.fr/2688/nouveau-lycee-general-et-technologique-guide-de-l-evaluation

[24] Une mission sur l'évaluation au lycée (général, technologique et professionnel) était également inscrite au programme de travail 2020-2021 de l'IGÉSR. Ce document a été complété en septembre 2021.

[25] La liberté pédagogique « s'exerce dans le respect des programmes et des instructions du ministre chargé de l'éducation nationale et dans le cadre du projet d'école ou d'établissement avec le conseil et sous le contrôle des membres des corps d'inspection » (Code de l'éducation, article L912-1-1).

La transformation de la **voie professionnelle** interroge également les pratiques des enseignants et le parcours des élèves. La certification intermédiaire disparaît (Décret n° 2020-1277 du 20 octobre 2020 relatif aux conditions de certification des candidats à l'examen du baccalauréat professionnel et portant suppression du brevet d'études professionnelles) au profit d'une autre modalité d'évaluation. De nombreuses nouvelles modalités pédagogiques (co-intervention, accompagnement renforcé, réalisation du chef-d'œuvre, etc.) s'accompagnent d'une nouvelle approche de l'évaluation notamment avec la prise en compte du contrôle continu dans l'évaluation du chef-d'œuvre.

Enfin, un Conseil d'évaluation de l'école (CEE) a été installé en 2019 et a notamment pour mission de mettre en cohérence toutes les évaluations de l'enseignement scolaire.

II. Ce que l'on observe

A. Tentative d'approche des pratiques évaluatives des enseignants à partir des enquêtes menées par la Depp

1. L'enquête Talis 2018 : 3 000 enseignants de collège en France

L'enquête internationale Talis (OECD, 2019) aborde la question des pratiques d'évaluation en se basant sur les réponses des enseignants du premier cycle du secondaire et de l'enseignement primaire.

Talis 2018 donne **4,7 heures** à l'Item « Corriger le travail de l'élève », temps consacré par semaine à la notation et à la correction[26] par les enseignants de collège en France.

Talis demande aux enseignants d'indiquer à quelle fréquence ils utilisent un ensemble de quatre pratiques pour évaluer les apprentissages des élèves dans leur classe spécifique. Pour le collège, les résultats montrent une position singulière de la France sur 3 des 4 items envisagés : 96,2 % des enseignants *versus* 79,6 % en moyenne pour les pays de l'Union européenne participants à l'enquête, déclarent administrer fréquemment ou toujours leur propre évaluation aux élèves ; 77,4 % *versus* 62,6 % pour la moyenne UE déclarent fournir des commentaires écrits sur le travail des élèves en plus de mettre une note ; 20,5 % *versus* 36,2 % pour la moyenne UE indiquent qu'ils laissent les élèves évaluer leurs propres progrès ; observer les élèves pendant qu'ils effectuent en classe une tâche particulière et leur apporter immédiatement des commentaires sur leur travail (78,3 % *versus* 80,3 % pour la moyenne UE).

Parallèlement, dans un autre champ de questionnement, celui du niveau de stress des enseignants français, parmi les sources de stress identifiées comme importantes les plus citées figure le fait « d'avoir trop de devoirs à corriger » (24 % contre 18 % pour la moyenne UE)[27]. Les disciplines artistiques, les lettres et la technologie sont celles où les taux d'enseignants se disant stressés sont les plus élevés. À l'opposé, les sciences humaines, les maths et l'EPS affichent des taux plus bas.

[26] En moyenne dans les pays de l'OCDE, les enseignants consacrent 4,2 heures par semaine à la notation et à la correction (l'équivalent de 11 % de leur temps de travail total).

[27] Les autres sources les plus citées sont le fait de devoir respecter les nouvelles exigences des autorités nationales, académiques ou locales (24 % contre 16 %) et le maintien de la discipline (22 % contre 13 %).

2. Épode 2018 auprès d'enseignants de collège

Les pratiques visant l'implication active des élèves dans les apprentissages (dimensions « Pédagogie active » et « Évaluation formative ») sont moins fréquentes et considérées comme moins faisables et moins prioritaires d'après l'enquête Enquête PériOdique sur l'Enseignement (Épode) (Note d'information de la Depp n°20.23, 2020) : seuls 54 % des professeurs interrogés déclarent proposer fréquemment des démarches d'investigation (ou des démarches inductives) aux élèves et 33 % déclarent proposer fréquemment aux élèves des outils d'auto-évaluation (grilles d'indicateurs sur les capacités, connaissances et attitudes) pour qu'ils évaluent leurs progrès, leurs réussites et leurs difficultés.

Les pratiques visant le développement de l'autonomie des élèves, la collaboration avec l'équipe à finalité pédagogique et l'évaluation formative sont beaucoup plus souvent réalisées par les enseignants décrivant un profil de classe très favorable aux apprentissages. Des écarts de plus de 10 points de pourcentage sont observés s'agissant de la proportion d'enseignants rapportant fréquemment organiser la collaboration entre élèves, leur proposer des outils d'auto-évaluation ou les impliquer dans la régulation de la vie de classe. Dans les classes décrites comme présentant un profil peu propice aux apprentissages et où le pourcentage d'élèves ayant un comportement considéré comme adapté est le plus faible, le lien École-famille est moins développé et jugé moins faisable. Malgré les difficultés scolaires de ces classes, les enseignants déclarent y faire également un peu moins de remédiation et d'explicitation. Par exemple, 63 % d'entre eux disent adapter fréquemment leur pédagogie pour répondre aux difficultés d'un élève lorsqu'il présente des signes de décrochage, contre 74 % pour les professeurs rapportant un profil de classe très favorable aux apprentissages. Ils sont par ailleurs 39 % à déclarer faire fréquemment un bilan avec leurs élèves pour savoir ce qui a favorisé ou empêché l'apprentissage visé (contre 56 %).

3. Épode 2018 auprès des enseignants des écoles

Jugées moins faisables et moins prioritaires par les enseignants, les pratiques visant l'implication active des élèves dans les apprentissages (dimensions « Pédagogie active » et « Évaluation formative ») sont pour la plupart moins adoptées. Moins de trois enseignants sur dix déclarent s'engager fréquemment dans des démarches d'expérimentation ou d'innovation, attribuer fréquemment des rôles aux élèves lors des débats, ou proposer fréquemment des outils d'auto-évaluation. Cependant quelques pratiques font exception : plus de sept enseignants sur dix disent confronter fréquemment les élèves à des situations-problèmes, prévoir fréquemment différentes modalités d'apprentissage au cours d'une séance, ou proposent fréquemment des activités mobilisant les connaissances dans des situations nouvelles. Alors que l'explicitation est une pratique très fortement adoptée à tous les niveaux d'enseignement, l'évaluation formative et la pédagogie active sont progressivement mises en place du CP au CM2, niveau où leur score de fréquence se rapproche de celui du collège. Au CP, 13 % des professeurs déclarent attribuer à certains élèves des rôles lors des débats (contre 33 % au CM2). Une proportion similaire déclare donner une grille d'évaluation aux élèves pour qu'ils comprennent leurs réussites et leurs erreurs (contre 30 % au CM2) (Note d'information de la Depp n°20.23, 2020).

B. Tentative d'approche des pratiques évaluatives des enseignants à partir des observations et enquêtes de l'IGÉSR[28]

1. Les pratiques évaluatives en école primaire

L'évaluation du niveau de maîtrise de compétences, alliée à la suppression des notes, étant déjà assez bien installée dans le premier degré avant 2015[29], nombre de professeurs des écoles n'ont guère modifié leurs pratiques d'évaluation des élèves au sein de leur classe. L'évaluation est avant tout perçue comme un préalable à la différenciation et à l'accompagnement pédagogique. Le souci d'une évaluation positive est régulièrement présent dans les projets d'école.

Les enseignants du premier degré ont pu développer une réflexion collective sur le rôle des différents outils de suivi des acquis des élèves selon leurs destinataires. Lorsque le LSU est bien perçu comme un outil de communication aux familles, certaines équipes améliorent, dans les bilans, la représentation visuelle de la progression de l'élève et trouvent des appréciations compréhensibles pour l'aider à s'approprier le bilan. Ils continuent pour autant à utiliser leurs outils professionnels de suivi des apprentissages, sur des supports variés, propres à chacun (carnets de progrès, classeurs de réussite, etc.) jusqu'aux outils numériques privés (comme par exemple Edumoov[30]) et renseignent directement les bilans périodiques dans le LSU. Quand une analyse des bilans est menée dans une même école par les enseignants, cela crée un foisonnement d'idées d'amélioration et d'échanges de pratiques dépassant largement le cadre de l'évaluation. Dans certaines circonscriptions, une politique volontariste a permis de mettre en place, voire de généraliser, des modalités d'évaluation nouvelles visant par exemple à associer l'élève à son évaluation[31] et à lui laisser le choix des moments auquel il se sent prêt pour être évalué, dans une dynamique progressive et positive.

Certains enseignants du premier degré sont désemparés par le changement de culture induit par les bilans volontairement synthétiques, alors qu'ils étaient habitués à détailler plus finement les compétences des élèves. Par exemple, en français, pour un positionnement dans la composante « lecture et compréhension de l'écrit », il ne leur est pas possible de rendre compte, d'une part, du fait que les objectifs sont atteints pour ce qui concerne le décodage des mots et, d'autre part, du fait qu'ils ne le sont pas pour ce qui concerne la compréhension. Les professionnels et les usagers y voient une perte en qualité d'informations.

À propos des fins de cycle, si la question du positionnement des élèves en fin du cycle 2 ne suscite pas de problèmes particuliers, il n'en est pas de même du cycle 3 qui suscite des problèmes spécifiques qui

---

[28] Certaines de ces enquêtes et observations ont donné lieu à des rapports rendus publics comme le rapport « L'enseignement en cours moyen : état des lieux et besoins n°2022-048 – avril 2022. D'autres ne l'ont pas été mais des éléments pourront être ici partagés. Il s'agit de : la note « Les premières applications des nouvelles formes et des nouveaux outils d'évaluation et de validation du socle (phases 1 à 3) n°N2017-06 de septembre 2017, qui a été remise à monsieur le ministre de l'éducation nationale ; le rapport « L'outil livret scolaire unique et son impact sur les modalités d'évaluation » n°2017-092, remis à monsieur le ministre ; le rapport à monsieur le ministre de l'Éducation nationale, de la jeunesse et des sports sur « Le parcours de l'élève au collège, formation, engagement, éducation formelle et non formelle » n°2021-178, en octobre 2021 ; le rapport à monsieur le ministre de l'Éducation nationale, de la jeunesse et des sports sur « L'évaluation des élèves au lycée (général, technologique et professionnel) n°2021-159, en octobre 2021.

[29] Une note de l'inspection générale relative à l'évaluation du socle commun en mars 2010 précise que la mise en œuvre d'une évaluation continue des élèves, organisée selon des échelles d'appréciation multiples rendant compte des différents niveaux d'acquisition possible, est habituelle dans les écoles.

[30] Ce logiciel privé est validé par le ministère et peut être interfacé avec l'application LSU.

[31] Talis 2018 pour l'école primaire indique que 18,1% des enseignants en France versus 35,2% dans l'UE déclarent laisser les élèves s'évaluer eux-mêmes.

ne sont toujours pas résolus à cette date. Ce positionnement concerne les professeurs de 6e mais également les professeurs de CM1 et de CM2. Au-delà des difficultés organisationnelles non abordées ici, restent les difficultés pédagogiques : comment les professeurs de CM1-CM2 peuvent-ils se prononcer sur la maîtrise d'attendus de fin de cycle concernant des élèves qu'ils n'ont pas suivis depuis une année ? La réponse consiste à « prépositionner » dans les domaines du socle en fin de CM2, ce qui est louable mais alerte : seuls sont définis par l'institution des attendus de fin de cycle, revient donc aux équipes de définir les 4 niveaux de maîtrise des domaines du socle de fin de CM2. Subsiste aussi la question des destinataires de ce pré-positionnement. En effet, ce peut être un outil diagnostique utile pour les professeurs de 6e, mais aussi générer des incompréhensions vis-à-vis des familles : comment expliquer qu'un élève positionné en niveau 3 dans un domaine en fin de CM2 soit au niveau 2 un an plus tard ?

Il convient de toute façon que les bilans périodiques de la classe de CM2 soient suffisamment clairs et précis tant au niveau des éléments de programme travaillés que des appréciations pour faciliter la prise en compte par les professeurs de 6e.

Les ressources publiées sur Éduscol « Documents d'accompagnement pour l'évaluation du socle » sont quasiment non utilisées. Rares sont les situations où les instances comme les conseils de cycle 2 ou de cycle 3 ou le conseil école-collège sont des lieux de réflexion sur les pratiques évaluatives.

Selon les observations faites en cours moyen (IGÉSR, 2022), la situation n'est pas la même selon les disciplines.

Pour **les enseignements artistiques**, **l'histoire des arts** et l'**EMC**, les élèves ne sont pas évalués.

Pour **les langues vivantes**, quelques évaluations ont été rencontrées dans environ un quart des classes observées. Il s'agit généralement d'une évaluation par période, en fin de séquence, portant sur des thématiques pouvant être rencontrées au cycle 2 comme au cycle 3, comme : « Se présenter », « Les jours et les mois », « Les nombres », « Les couleurs », « Les parties du corps », etc. Ces évaluations portent sur du lexique avec, la plupart du temps, des réponses à donner sous forme de mots isolés.

Pour **l'histoire**, **la géographie**, **les sciences et la technologie**, des évaluations sont mises en place, souvent au rythme d'une évaluation par période et en fin de séquence. Elles évaluent principalement des connaissances qui doivent être restituées sous forme de groupes nominaux ou de phrases courtes. Ces évaluations sont relativement longues : sans doute entre 30 minutes et une heure de travail selon les cas. Elles portent sur une thématique unique ou sur les différents chapitres étudiés pendant la période.

En mathématiques et en français, les évaluations se limitent souvent à des tests bilans de fin de séquence, les interrogations écrites très courtes sont excessivement rares.

Pour **le français**, la situation est assez hétérogène selon les classes. Dans de très nombreuses classes, l'orthographe est évaluée à un rythme soutenu sous forme de dictées (une dictée chaque semaine avec généralement un travail quotidien pour la préparer). Le second type d'évaluation le plus fréquent en français, rencontré dans plus de la moitié des classes, concerne les conjugaisons. Ces évaluations de conjugaison portent principalement sur le présent et l'imparfait de l'indicatif ; dans de rares classes, le futur et le passé composé sont également interrogés dans les évaluations. Dans un peu moins d'une classe sur deux, des évaluations de grammaire sont également présentes. Elles portent principalement sur la distinction des classes de mots et le repérage de constituants d'une phrase plus rarement le type

de phrases (déclaratives, interrogatives ou impératives) ou la forme négative sont au cœur des évaluations. Dans un peu moins de 10 % des classes, les évaluations de français comprennent des productions écrites, généralement qualifiées de « rédaction » : « description d'une image », « bilan de lecture », « rédaction sur Charlot », etc.

Pour **les mathématiques**, les évaluations sont plus fréquentes et plusieurs évaluations par période sont mises en place dans presque toutes les classes. Ces évaluations sont très majoritairement de longues évaluations de fin de séquence ; de courtes évaluations en cours de séquence pour permettre aux élèves et aux enseignants de faire un point avant de poursuivre dans la séquence ne sont pour ainsi dire jamais proposées.

Les compétences en calcul mental ne sont pratiquement jamais supports d'évaluation. Quand elles le sont, les évaluations sont soit trop générales, et cela ne permet pas d'en tirer un véritable enseignement sur les compétences des élèves en calcul mental.

L'évaluation est donc le plus souvent un moment distinct de la leçon dramatisant parfois excessivement l'évaluation.

Une série de petites évaluations dans le fil du cours sont souvent bien plus efficaces pour la structuration des connaissances et la vérification par le professeur des acquis de chaque élève (Roediger & Karpicke, 2006).

L'absence de petites interrogations régulières peut s'expliquer par la prise d'information de certains enseignants par une lecture quotidienne et systématique de l'ensemble des cahiers du jour des élèves. Cette lecture permet au professeur de prendre connaissance avec précision des réussites et des difficultés de chaque élève. Cependant, cette lecture des cahiers du jour ne semble pas, ou plus, systématique dans de nombreuses classes. Si la lecture régulière des cahiers du jour permet sans doute de prendre des informations quant aux progrès et aux difficultés des élèves, elle ne permet cependant pas de remplacer les courtes évaluations pour réguler les apprentissages des élèves, en facilitant la prise de conscience, par les élèves, de leurs éventuelles difficultés en cours de séquence afin de pouvoir mieux les surmonter.

2. Les pratiques évaluatives en collège

Sur cette question, comme d'ailleurs sur d'autres aspects pédagogiques, on constate un écart culturel important entre professeurs du premier et du second degré.

Des observations conduites par l'IGÉSR dès 2016-2017 ont montré des évolutions positives dont une prise de conscience, chez les équipes enseignantes, de la nécessité de faire évoluer leurs pratiques évaluatives. La conception même de l'évaluation a évolué, celle-ci devenant de plus en plus intégrée à l'acte d'enseignement et de moins en moins perçue comme une finalité en soi.

Les situations d'évaluation sommative se sont diversifiées, se limitant de moins en moins au traditionnel devoir de contrôle réalisé individuellement par écrit, et s'étendent de plus en plus à l'évaluation de l'oral, de démarches de projets et de travaux de groupes. En parallèle, les dimensions diagnostique et formative de l'évaluation ont été de plus en plus prises en compte, notamment à travers le dispositif d'accompagnement personnalisé.

Dans certains établissements, un travail élaboré a pu être engagé comme dans les établissements relevant de l'éducation prioritaire bénéficiant à la fois de temps dédiés à la concertation et de la

présence de coordonnateurs ou de formateurs particulièrement engagés dans la réflexion sur l'évaluation, ou encore des classes de troisième « prépa-métiers » dans lesquelles les équipes sont depuis longtemps engagées dans une dynamique autour de l'évaluation.

Le travail a alors abouti à des échelles descriptives spécifiant à la fois des observables communes à plusieurs disciplines et, pour chacune d'elles, les attendus correspondant à chacun des quatre niveaux de maîtrise. L'élaboration collective de telles grilles a le mérite de faire réfléchir les équipes à des exigences communes et de les objectiver, ce qui leur permet notamment, en conseil de classe, d'être en capacité d'argumenter l'attribution de tel ou tel niveau de maîtrise. Cependant, l'appropriation collective de ce type d'outils, surtout s'ils ont été fabriqués par d'autres, s'avère délicate et particulièrement chronophage. S'il semble difficile d'en généraliser l'usage, c'est bien en s'inspirant de ce type de démarche que les équipes peuvent arrêter et partager collectivement des observables.

Concernant les disciplines, il ressort que l'EPS, les sciences expérimentales et les mathématiques ont été assez naturellement acteurs du changement quant aux nouvelles modalités d'évaluation, tandis que les sciences humaines ont eu plus de difficultés. Il est à noter que toutes les disciplines se trouvent confrontées au *risque d'une atomisation des compétences en micro-compétences*, risque amplifié par l'utilisation de certains logiciels privés de suivi des élèves. Certains inspecteurs, comme certains chefs d'établissement, très impliqués dans l'évaluation par compétences du socle 2006, ont également eu l'impression d'une régression, notamment à travers la mise en avant des disciplines dans les bilans périodiques.

Si les nombreux travaux dès 2006 puis à compter de 2015 ont permis aux professeurs comme aux disciplines scolaires de s'emparer des attendus du socle, si la notion de compétence a pris aujourd'hui une place significative dans les constructions pédagogiques comme dans les évaluations des apprentissages, les interrogations demeurent. Elles sont alimentées par l'écart de développement très significatif entre un socle commun particulièrement rédigé et développé et des programmes qui, parallèlement, ont eu tendance à privilégier une rédaction plus resserrée. La base théorique du socle n'apparaît pas clairement et n'a pas été suffisamment explicitée et portée par l'institution.

Cela se traduit également par la place occupée par l'évaluation par compétences qui reste le plus souvent opposée à la notation. Les professeurs utilisent des notes et parallèlement positionnent au fil de l'eau les élèves sur les niveaux de maîtrise des compétences du socle. De plus, au moment du DNB, il n'y a pas de croisement entre le socle et les épreuves ponctuelles, les élèves ne se préoccupant plus des compétences. Il est aussi relevé qu'il y a deux types de compétences : celles des disciplines et celles du socle. Les professeurs semblent s'accorder sur un régime mixte qui permet un compromis vertueux entre deux tendances opposées : les compétences permettent aux élèves de comprendre la note et la note leur permet de se situer par rapport aux autres.

Pour tous les professeurs, la capacité à prélever des informations à partir de situations d'évaluation ou d'observations disciplinaires, puis à les synthétiser collégialement pour aboutir au positionnement final dans les domaines ou composantes du socle (dont certains sont éloignés du cœur même des disciplines) est un exercice extrêmement complexe. Les deux ne se recoupant pas parfaitement, des processus locaux relativement opaques de conversion, de convergence sont mis en œuvre pour proposer une évaluation du socle commun.

Cela s'explique à la fois par la tradition d'une évaluation globale basée sur le calcul de moyennes d'évaluations disciplinaires, l'absence de culture de la délibération collective et la dissymétrie intrinsèque des contributions des disciplines aux domaines du socle.

En l'absence de cadrage institutionnel sur un sujet dont ils ne détenaient pas les clés, certains chefs d'établissement ont choisi de ne pas trop bousculer les habitudes de leurs enseignants en matière d'évaluation. Celles-ci, souvent réduites à des calculs de moyennes disciplinaires, ont simplement été étendues à des calculs de moyennes entre des évaluations effectuées dans toutes les disciplines contribuant à l'évaluation d'un même domaine.

Le « document d'accompagnement pour l'évaluation des acquis du socle commun de connaissances, de compétences et de culture », publié sur Éduscol, a été conçu dans le but d'aider les équipes à identifier des repères communs pour l'appréciation collective du niveau de maîtrise satisfaisant en fin de cycle 4. Si les « éléments signifiants » introduits dans ce document composite ont pu être utilisés par certains chefs d'établissements comme échelon intermédiaire entre les compétences travaillées dans les différentes disciplines et les domaines du socle, force est de constater que leur usage est très complexe et que le document, dans son ensemble, manque d'opérationnalité.

Les difficultés auxquelles se sont heurtées les équipes pour arrêter les positionnements dans les domaines du socle sont mises en évidence dans le décalage entre les niveaux de maîtrise attribués dans les domaines du socle et les appréciations portées sur le niveau des élèves par oral lors des conseils, ou par écrit sur les bilans de fin de cycle[32].

De tels écarts prouvent qu'il y a manifestement des incohérences dans l'attribution des niveaux de maîtrise des domaines du socle[33].

De plus, aucun consensus n'existe entre les professeurs sur l'opportunité d'arrêter la notation chiffrée. La plupart des professeurs souhaitent garder la double évaluation considérant, comme nous l'avons vu *supra*, qu'elle apporte des informations de nature différente. Les professeurs continuent à confondre ce que l'on évalue et comment on exprime la valeur de ce que l'on a évalué. Cette différence de fond n'est pas claire non plus pour tous les chefs d'établissement qui opposent notation sur 20 et évaluation par compétences, et exercent parfois sur leurs équipes de fortes pressions pour qu'elles se rallient à cette dernière modalité et suppriment les notes.

Les parents désapprouvent majoritairement l'évaluation par compétences jugeant de manière parfois péremptoire qu'elle est illisible et incompréhensible. Les élèves quant à eux sont parfois dépassés par la finesse des items explicitant les compétences évaluées et utilisés dans certains bulletins.

Quoiqu'il en soit, la complexité actuelle de l'évaluation des élèves écorne sans aucun doute la relation à l'usager.

---

[32] Ainsi, lors de l'un des conseils, il a été dit d'un élève à qui ont été attribués 340 points sur 400 qu'il rencontre de sérieuses difficultés dans le passage à l'écrit et qu'il manque de rigueur ou encore, d'un élève à qui ont été attribués 350 points qu'il est fragile en SVT, en espagnol et en mathématiques. Mérite également d'être signalée la situation d'un élève crédité de 350 points alors qu'il est dit être fragile en mathématiques et en anglais.

[33] Notons que d'après la loi du 8 juillet 2013, « Le Gouvernement présente tous les trois ans au Parlement un rapport sur la manière dont les programmes prennent en compte le socle commun et sur la maîtrise de celui-ci par les élèves au cours de leur scolarité ». Pour répondre à cette obligation (La DEPP procède sous la forme d'une enquête sur échantillon), l'indicateur RAP (rapport annuel de performance) utilisé est la « proportion d'élèves maîtrisant, en fin de CE2, 6e et 3e les principales composantes du domaine 1 « les langages pour penser et communiquer » du socle commun ». L'institution prévoit donc un dispositif spécifique d'évaluation des acquis des élèves appuyés sur une évaluation standardisée.

Les établissements du second degré qui renseignent le LSU par saisie directe sont extrêmement rares. En effet, les établissements utilisent depuis de nombreuses années des suites logicielles privées payantes ou des outils libres et gratuits spécialement conçus pour l'évaluation par compétences. Ces outils restituent les performances des élèves sous la forme d'un positionnement alphanumérique, la concaténation des différentes évaluations (par discipline pour positionner l'élève en fin de période, ou issues de plusieurs disciplines pour le positionnement de fin de cycle) se fait par l'intermédiaire d'un algorithme reposant sur des paramétrages laissés au libre choix de l'administrateur du logiciel, et donc variables d'un établissement à l'autre et souvent totalement inconnus des utilisateurs. Les référentiels figurant par défaut dans ces outils sont constitués d'items très fins, détaillés à l'extrême, correspondant à des « micro-compétences » et ne renvoient pas une vision synthétique du travail réalisé durant une période.

L'algorithme sous-jacent ne pondère souvent ni les disciplines, ni les différentes évaluations à l'intérieur d'une discipline. Ainsi, l'interprétation d'une œuvre littéraire n'aura pas plus de poids pour le positionnement dans la composante 1 du domaine 1 (comprendre, s'exprimer en utilisant la langue française à l'écrit et à l'oral) que la compréhension d'un énoncé dès lors qu'un professeur de mathématiques ou de sciences aura choisi de relier une évaluation consistant en la résolution d'un problème à la composante D1-1. Souvent, seul est pris en compte le nombre de situations permettant d'évaluer toutes les compétences rattachées à un domaine donné. Donnant l'illusion de la rigueur et de l'objectivité (puisque fournis par l'ordinateur, ils ne relèveraient pas de la subjectivité humaine), l'utilisation d'un tel outil a pu également avoir pour effet de dispenser les équipes de discussions délicates et houleuses et le chef d'établissement de sa responsabilité.

En fin de cycle 4, la note du socle peut résulter d'une moyenne pondérée des notes issues des disciplines mais également du calcul automatisé réalisé par ce type d'outil, le résultat étant parfois à la base d'un conseil de socle visant à valider ces calculs automatiques. Il n'est alors pas étonnant que les professeurs déplorent bien souvent que les élèves s'y retrouvent difficilement entre le socle et les disciplines. La prise en charge du socle par des logiciels interroge quand on paramètre le logiciel et que de façon automatique, celui-ci renseigne les éléments d'un profil de l'élève. Les pistes actuelles sont dans des collèges qui font évoluer leurs conseils de classe vers de véritables conseils de socle où chaque professeur met son expertise d'évaluateur disciplinaire au service du positionnement des élèves sur les compétences transversales posées par le socle commun.

Si on relie les moyennes des résultats obtenus aux épreuves écrites de français et de mathématiques au DNB pour la session 2019 au score moyen obtenu au socle, on constate une corrélation de plus en plus forte entre les deux lorsque l'on examine successivement les collèges REP+, les collèges REP et les collèges publics hors éducation prioritaire, attestant en cela d'une politique assurément différente de l'évaluation du socle. Dans les collèges REP+, les politiques d'évaluation du socle sont très hétérogènes mais on observe tout de même dans de nombreux REP+ que l'évaluation du socle semble être utilisée comme « régulateur » des résultats au diplôme national du brevet (DNB).

La variabilité des politiques évaluatives concerne à la fois :

- la répartition des domaines et composantes du socle entre les différentes disciplines ;
- les modalités d'attribution des niveaux de maîtrise de ces domaines, y compris le travail préparatoire (par le chef d'établissement et/ou le professeur principal).

Cette variabilité a une incidence sur les niveaux de maîtrise attribués à chaque élève dans les domaines du socle, ce qui soulève la question de l'équité d'une évaluation qui impacte à la fois la note obtenue à un examen national (le DNB, séries générales et professionnelles) et l'affectation post-troisième.

Nombreux sont ceux qui s'interrogent sur la pertinence qu'il y a, en parallèle, à discuter de manière fine sur le barème des épreuves écrites terminales au vu du manque de rigueur qui préside à l'attribution des 400 points relatifs aux domaines du socle.

Quant aux chefs d'établissement, ils regrettent l'absence de cadrage national pour mettre en œuvre une réforme de l'évaluation sans doute séduisante au plan intellectuel, mais dont l'opérationnalisation ne leur semble pas avoir été suffisamment pensée.

Si l'article D. 122-2 du code de l'éducation énonce que « chacun de ces domaines requiert la contribution transversale et conjointe de toutes les disciplines et démarches éducatives », à aucun moment il n'est stipulé que cette contribution doit se faire à part égale, tant dans la formation que dans l'évaluation des élèves.

Enfin, certains propos des élèves concernant la différence entre l'école primaire et la classe de sixième sont particulièrement intéressants revendiquant en creux une clarification des pratiques : « à l'école primaire, le professeur souligne nos forces alors qu'en sixième, les professeurs soulignent plus souvent nos faiblesses » ; « la note permet de se comparer aux autres alors que l'évaluation par compétence nous comparent à nous-mêmes. L'idéal serait les deux systèmes en même temps ». Ils considèrent que les professeurs doivent concilier « bienveillance » et « exigence ».

L'accompagnement à l'orientation, la qualité de la relation avec le professeur principal, le diplôme national du brevet (DNB) – qui « fait peur » mais auquel l'attachement est réel – constituent des sujets-clefs abordés par les élèves de troisième.

3. Les pratiques évaluatives en lycée

Précisons que le propos concerne ici la scolarité obligatoire, donc, dans les lycées, plus spécifiquement la classe de seconde. Mais dans les lycées, tous les enseignants ou presque accompagnent les élèves au-delà de celle-ci et donc vers les certifications. Ils déclarent d'ailleurs que leurs pratiques d'enseignement et surtout évaluatives sont très liées à leurs programmes d'enseignement et aux attendus certificatifs. De plus, le contexte dans lequel les enseignants ont été rencontrés au moment où la mission à l'origine du rapport support de ce propos, a travaillé, celui de la crise sanitaire et de toutes les modifications induites en termes d'évaluation dont la mise en place du contrôle continu et la place de l'évaluation dans Parcoursup, a influencé les réponses tant au niveau des enseignants que des élèves et de leurs familles, particulièrement pour le lycée général. Nous serons donc plus brefs sur ces points et en nous limitant aux apports de l'enquête par questionnaire, réalisée dans le cadre de cette mission auprès de 400 professeurs et de 4 900 élèves.

Dans le code de l'éducation, le mot « note » au singulier et au pluriel apparaît plus de 200 fois. Il n'est donc pas étonnant que les professeurs évoquent souvent le caractère institutionnel de l'évaluation, et justifient d'emblée l'usage de la notation pour évaluer[34]. De même, les élèves et les parents pensent directement à l'évaluation sommative quand on demande ce qu'est l'évaluation pour eux. Cependant, très vite dans les échanges avec les acteurs de l'évaluation, la distinction entre notation et évaluation

[34] Les notes chiffrées sont citées à la hauteur de 95,5 % alors que les niveaux de maîtrise le sont pour 31,9 %.

se fait, notamment par la fonction donnée à celles-ci (progresser, apprendre, motiver, classer, s'orienter, mesurer, certifier, etc.). Cette observation est plus marquée pour les enseignants que pour les élèves[35]. Cette appréhension de la notation comme un moyen objectif (sans biais) conduit les enseignants à proposer :

- des évaluations sommatives[36] en cours de formation, régulières et tournées vers les attendus de l'examen ;
- relativement peu de différenciation dans l'évaluation, et une évaluation formative très peu mise en œuvre en tant que telle (ce qui est moins vrai en lycée professionnel et en tout début de cycle de formation)[37] ;
- des approches qualitatives de l'évaluation peu développées (explicitation, appréciation mettant en lumière les progrès, les acquis du processus d'apprentissage, etc.)[38] ;
- relativement peu d'approches collectives de l'évaluation des élèves[39], et donc des pratiques d'évaluation très diverses et peu remises en question sous le couvert de la liberté pédagogique et individuelle de l'enseignant ;
- une appropriation difficile des résultats des évaluations par les élèves[40].

Les élèves, comme les parents, restent également très attachés à la note qui constitue pour ces derniers en particulier, un repère simple[41]. Elle participe à leur positionnement social et scolaire. Elle est une forme de reconnaissance entre pairs et au-delà. Elle est un passeport pour leur avenir. Cette appréhension identitaire est d'autant plus graduée entre le lycée professionnel (moins forte, mais très présente également), le lycée technologique et le lycée général, que l'élève appartient aux catégories socioprofessionnelles dites supérieures. Et pourtant, 46 % des lycéens interrogés pensent qu'il y a trop d'évaluations notées et que les notes sont des éléments de stress[42] (cités à la hauteur de 63,8 %).

L'évaluation en voie professionnelle présente un paysage plus varié qu'en voie générale et technologique. Le fait qu'une partie importante des élèves ont rencontré des difficultés dans leur scolarité antérieure d'une part, et d'autre part, le fait qu'en voie professionnelle le parcours de l'élève s'inscrit et de se décrit sans doute plus aisément dans une logique de « compétences », incitent les

[35] Pour les élèves, à la hauteur de 62,3 % parmi la liste de propositions faites, l'évaluation est du contrôle. Préparer des examens : 42,7 % ; connaître son niveau dans les disciplines par rapport à d'autres : 41,7 % ; connaître sa progression : 42,3 %. Pour leurs professeurs, l'évaluation c'est diagnostiquer (49 %), vérifier (39 %), progresser (31 %), contrôler (24 %), noter (18 %). Les propositions « aider », apprendre, encourager, valoriser recueillent respectivement 9 %, 12 %, 13 % et 18 %.

[36] Sous la forme de devoirs surveillés (90,3 %), de contrôles brefs, d'interrogations courtes, d'exercices écrits (86,1 %). Ce que confortent les élèves, évoquant les devoirs surveillés, programmés et notés (95,4 %) et les contrôles courts et les interrogations écrites notés et programmés (48,1 %).

[37] Pour autant les tâches en cours d'apprentissage et/ou les observations des élèves pendant les situations d'apprentissage sont citées par les enseignants à la hauteur de 81,6 % ; ce qui n'est pas du tout conforté par les élèves (15,5 %).

[38] 72,3 % des enseignants disent que dans leurs évaluations, ils recourent à des appréciations, mais pour 42,2 %, en termes de progression de l'élève, pour 35,7 % en termes de compétences.

[39] 63,5 % disent ne jamais ou rarement construire des évaluations avec leurs collègues, 80,2 % jamais ou rarement pratiqué des évaluations croisées avec des collègues.

[40] Voir la note de bas de page suivante.

[41] Ils plébiscitent les notes dans les bulletins à la hauteur de 89,9 % et seulement 58,5 % les commentaires sur les notes dans les copies. Et pourtant ils trouvent ces commentaires pas assez nombreux et explicites pour savoir comment s'améliorer.

[42] La proposition « les évaluations notées sont des éléments de motivation pour travailler les cours » ne recueillent que 30,1 % de suffrage.

enseignants à s'interroger de façon plus active sur les modalités d'évaluation et à expérimenter davantage qu'en voie générale et technologique. La caractéristique la plus visible est la part plus importante consacrée à l'évaluation par compétences, à la fois dans les disciplines professionnelles et dans celles d'enseignement général.

Ainsi, le sondage réalisé par la mission montre que, par rapport aux voies générale et technologique, l'évaluation des élèves de la voie professionnelle repose beaucoup moins sur des devoirs surveillés programmés et moins sur des interrogations orales notées. En revanche, en voie professionnelle, l'évaluation s'appuie davantage sur des questions posées par les professeurs pendant les cours, sur l'auto-évaluation par les élèves ou sur l'évaluation par les pairs.

L'enquête confirme la place importante de l'évaluation de la maîtrise de compétences chez les enseignants de la voie professionnelle : 58 % des enseignants de spécialités professionnelles déclarent la pratiquer de manière élevée ou très élevée. Cette tendance est moins marquée chez les enseignants de disciplines générales en voie professionnelle : ils ne sont que 48 % à déclarer y recourir, ce qui reste cependant très supérieur aux 36 % d'enseignants de la voie générale et technologique qui y font appel.

De nombreux outils de suivi et d'évaluation sont utilisés par les équipes (grilles officielles d'évaluation, livrets de suivi et/ou d'évaluation conçus à différents niveaux : par le professeur, par une équipe de professeurs d'une discipline ou relevant d'un diplôme au niveau national). Ces outils, sont parfois complexes dans leur utilisation ou réducteurs dans leur résultat (il suffit de cocher des cases pour générer une note).

Malgré ces contraintes, cette évaluation dont les critères se rapportent explicitement aux compétences a induit des pratiques d'évaluation formative répondant bien aux objectifs des enseignements, même si beaucoup d'entre elles restent encore très liées à l'objectif terminal du diplôme, voire à des approches plus traditionnelles en termes de savoirs. Des portfolios, des livrets de compétences ont été élaborés pour suivre le développement des compétences des élèves. Certains de ces livrets ont une dimension formative très aboutie, quand d'autres restent très axés sur le résultat. La dimension certificative de l'évaluation reste très importante au lycée professionnel, ce qui nécessite de poser des notes chiffrées. Le passage de l'évaluation de la compétence (niveau) à la note reste problématique. Ainsi à chaque rénovation de diplômes, les premières questions des professeurs portent souvent sur les épreuves et les futures grilles d'évaluation, même si les référentiels intègrent de plus en plus souvent les indicateurs de performance ou les critères d'évaluation qui seront repris dans les grilles d'évaluation.

Il en va différemment des disciplines générales enseignées en lycée professionnel, pour lesquelles les compétences mentionnées (quand elles le sont) dans les programmes sont plus abstraites. Les élèves, les parents et, dans de nombreux cas, les enseignants eux-mêmes ont souvent des difficultés à bien comprendre ce que recouvrent certaines de ces compétences. Par ailleurs, cette abstraction rend toute autoévaluation très délicate puisqu'elle exige, pour avoir du sens, que l'élève maîtrise correctement la signification des compétences de la discipline et sache identifier dans son travail les capacités qui relèvent, ou pas, de telle ou telle compétence tout en évaluant son degré de maîtrise de ces capacités.

Souhaitant promouvoir l'autoévaluation des élèves, de nombreux enseignants font figurer en marge des questionnaires la nature des compétences évaluées par chaque question et demandent à l'élève d'identifier lui-même son niveau de maîtrise (souvent symbolisé par des couleurs ou des symboles de

type « smileys »). Ce type d'évaluation nécessite une expertise, qui est propre à chaque discipline et que les élèves n'ont pas ; certains enseignants s'attèlent à la leur faire acquérir, le risque étant alors de créer une surcharge cognitive importante chez les élèves tout en brouillant quelque peu les objectifs de formation.

La très grande variété de formulation des compétences évaluées dans les différentes disciplines, ou au sein d'une même discipline, dans différents contextes, est une source de confusion importante chez les élèves et leurs parents. De plus, pour une même discipline, les compétences mentionnées dans le livret scolaire du baccalauréat professionnel (LSL pro) ne sont pas toujours les mêmes que celles qui figurent dans les bulletins ou même dans les programmes. L'échelle des degrés de maîtrise des compétences elle-même est souvent différente dans le LSL pro et dans les grilles d'évaluation utilisées en CCF, qui servent souvent de référence aux enseignants pour leurs évaluations pendant l'année scolaire.

D'une discipline à l'autre, le paysage est encore plus compliqué ; les disciplines font un usage différent des notions de capacités et de compétences et utilisent des formulations parfois subtilement différentes.

Au lycée général et technologique, comme indiqué supra, la réforme du baccalauréat ayant beaucoup d'effet sur les pratiques évaluatives et sur la façon dont elles sont vécues lors du cycle terminal, nous ne les évoquerons pas ici. Mais la mission a pu observer des pratiques évaluatives dans certains établissements rencontrés où les chefs et les équipes pédagogiques souhaitent maintenir et organiser une distinction entre évaluation formative et sommative, où l'on cherche à pratiquer des évaluations par compétences qui comptent et qui soient intégrées dans la restitution formelle du niveau des élèves, et où les équipes pratiquent l'auto-évaluation des élèves par leurs pairs, en particulier pour l'évaluation à l'oral.

Pour résoudre à la fois la façon d'évaluer les compétences des élèves et celle de les noter, certaines équipes ont développé ce qu'on a nommé « les classes sans note », essentiellement en classe de seconde.

L'entrée au lycée est souvent vécue par les élèves et qualifiée par les enseignants comme un moment de rupture avec le collège, plus qu'une continuité pédagogique. C'est vrai pour l'évaluation et le rapport aux apprentissages. On peut être surpris du peu de cas qui est fait du LSU par les professeurs de lycée. Ce que dit le livret sur les compétences acquises par les collégiens n'est pas pris en compte dans le parcours de l'élève lors de son passage au lycée, alors qu'il est un des éléments d'évaluation diagnostique intéressants pour préparer le parcours personnel de l'élève au lycée.

Les tests de positionnement à l'entrée en seconde, proposés ces dernières années, ont été peu utilisés par les professeurs[43]. Les pratiques d'évaluation diagnostique et de positionnement des élèves étant mieux installées au lycée professionnel par rapport aux lycées généraux et technologiques, l'utilisation des tests de positionnement y est plus soutenue.

Pour terminer, il convient de préciser que les enseignants de lycée disent rencontrer des difficultés dans la pratique de l'évaluation, plus spécifiquement ceux du lycée général et ceux qui ont la charge des enseignements généraux des voies technologiques et professionnelles. Ils disent manquer de formation.

[43] 74,5 % disent ne pas être destinataires des résultats et la moyenne est de 2,11 quand on les interroge sur leur utilité dans une échelle de 1 à 4.

Conclusion

Il existe donc dans notre système éducatif des boussoles comme le socle, les livrets, les programmes, les examens, les contenus des évaluations standardisées qui fixent des directions évaluatives. Des circulaires, des documents d'accompagnement viennent à un instant *t* aider à la compréhension du sens et à la mise en œuvre. Dans tous les cas, l'attention porte sur le fait de ne pas isoler l'enjeu de l'évaluation d'autres enjeux liés à l'orientation des élèves, la certification, le rôle des diplômes, la pédagogie, la mise en place de dispositifs d'accompagnement, de personnalisation et la formation. Le national impulse et accompagne, avec le relais des pilotes en académies, en mettant en œuvre des formations dès lors que ces politiques évaluatives se mettent en place et des frémissements se font jour dans les évolutions souhaitées. En attestent des évolutions positives dans les classes, en particulier en maternelle ou dans le premier degré plus généralement. Mais au collège et dans la première année du lycée, si le frémissement a parfois fait place à un épanouissement, ce n'est pas le cas le plus fréquent.

Les modalités d'évaluation restent assez différentes entre l'école et le collège, de même qu'entre le collège et le lycée.

Malgré les attendus de la validation du socle et la mise en œuvre du contrôle continu en lycée, encadrée par le projet d'évaluation, porteur d'espoirs[44], l'évaluation dans la classe reste peu concertée au sein des écoles et établissements et parfois peu cohérente (finalités, types, modalités, fréquence). L'évaluation est parfois aussi une source de tension au sein de l'équipe éducative notamment au collège à cause du socle qui au contraire devrait rapprocher les enseignants. Elle est bien souvent considérée comme relevant exclusivement de la liberté pédagogique de chaque professeur.

L'évaluation reste focalisée sur la note et la moyenne, notamment dans le second degré, obérant la pertinence de l'évaluation. L'évaluation formative reste peu développée, l'évaluation des apprentissages l'emportant sur l'évaluation pour les apprentissages. Les évaluations formative et sommative sont fréquemment confondues au sein d'une seule et même note.

Finalement, alors qu'il joue un rôle essentiel dans les procédures d'orientation (à la fin du collège et du lycée), le contrôle continu est peu harmonisé. Le guide de l'IGÉSR représente un véritable progrès en ce sens qu'il apporte des recommandations disciplinaires aidant les équipes, dès lors qu'elles s'en emparent au-delà d'un nombre de notes à respecter[45].

Mais proposer des recommandations qui comme indiqué *supra* font craindre que la liberté pédagogique soit mise à mal ainsi que définir des exigences institutionnelles vécues comme des injonctions contradictoires (notes et compétences dans les livrets ; notation pour le contrôle continu et suivi des apprentissages) font souvent du sujet de l'évaluation un objet de crispation et de blocage favorable au *statu quo*.

[44] La note de service de la Dgesco du 28 juillet 2021, en recommandant que les conditions d'organisation du contrôle continu du baccalauréat soient définies par un projet d'évaluation pour l'établissement, marque une évolution vers une concertation renforcée – au sein et entre les disciplines – et une définition collective des principes d'évaluation des élèves au lycée.

[45] Lors de l'enquête de la mission, 51,6 % des enseignants interrogés ont dit ne pas avoir connaissance de ce guide mais il peut s'agir des professeurs des lycées professionnels non concernés par celui-ci. Parmi ceux l'ayant utilisé, 47,4 % lui ont reconnu une utilité dans leurs pratiques.

Références

Article L912-1-1, Code de l'éducation (2005).
https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000006525569/

Arrêté du 1-7-2013 relatif au référentiel des compétences professionnelles des métiers du professorat et de l'éducation, NOR : MENE1315928A, BO n°30 (2013).
https://www.education.gouv.fr/bo/13/Hebdo30/MENE1315928A.htm?cid_bo=73066

Arrêté du 31-12-2015 relatif aux modalités d'attribution du diplôme national du brevet, NOR : MENE1531424A, JORF n°0002 (2016).
https://www.legifrance.gouv.fr/loda/id/JORFTEXT000031742288/

Arrêté du 31 décembre 2015 fixant le contenu du livret scolaire de l'école élémentaire et du collège, NOR : MENE1531425A (2015). https://www.legifrance.gouv.fr/loda/id/JORFTEXT000031742317/

Benhaim, J., Longhi, L., Monseur, C., Solnon, A., Verdon, R., Charpentier, A., Raffaelli, C. & Raffaelli, C. (2020). Premiers résultats de l'enquête sur les pratiques d'enseignement, EPODE, en 2018 au collège. *Note d'information de la Depp*, *20.23*, 4.

Benhaim-Grosse, J., Longhi, L., Monseur, C., Solnon, A., Verdon, R., Charpentier, A. & Raffaelli, C. (2020). Premiers résultats de l'enquête sur les pratiques d'enseignement des professeurs des écoles, EPODE, en 2018. *Note d'information de la Depp*, *20.32*, 4.

Circulaire de rentrée 2016, NOR : MENE1608893C, BO n°15 (2016).
https://www.education.gouv.fr/bo/16/Hebdo14/MENE1608893C.htm

Circulaire de rentrée 2017, NOR : MENE1707568C, BO n°10 (2017).
https://www.education.gouv.fr/bo/17/Hebdo10/MENE1707568C.htm

Conseil d'évaluation de l'École. (2022). Renforcer la cohérence des évaluations des acquis des élèves : Un enjeu de réussite et d'équité. *Note d'information du CEE*, *22-01*, 4.

Décret n° 2015-372 du 31 mars 2015 relatif au socle commun de connaissances, de compétences et de culture, NOR : MENE1506516D, BO n°17 2015-372 (2015).
https://www.education.gouv.fr/bo/15/Hebdo17/MENE1506516D.htm?cid_bo=87834

Décret n° 2015-1929 du 31 décembre 2015 relatif à l'évaluation des acquis scolaires des élèves et au livret scolaire, à l'école et au collège, NOR : MENE1531422D (2015).
https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000031742252/

Décret n° 2020-1277 du 20 octobre 2020 relatif aux conditions de certification des candidats à l'examen du baccalauréat professionnel et portant suppression du brevet d'études professionnelles, NOR : MENE2018661D, JORF n°0257 (2020).
https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000042452152

Évaluer la maîtrise du socle commun aux cycles 2, 3 et 4. (2016). Éduscol.
https://cache.media.eduscol.education.fr/file/socle_commun/42/9/Evaluer-la-maitrise-du-socle-commun-aux-cycles-2-3-4_1314429.pdf

Guide de l'évaluation des apprentissages et des acquis des élèves dans le cadre de la réforme du lycée général et technologique. (s. d.). Éduscol | Ministère de l'Éducation nationale, de la Jeunesse et des Sports - Direction générale de l'enseignement scolaire. https://eduscol.education.fr/2688/nouveau-lycee-general-et-technologique-guide-de-l-evaluation

IGEN & IGAENR. (2005). *Les acquis des élèves, pierre de touche de la valeur de l'école ?* (N° 2005-079; p. 85). https://www.education.gouv.fr/media/70472/download

IGÉSR. (2017). *Les premières applications des nouvelles formes et des nouveaux outils d'évaluation et de validation du socle (phases 1 à 3)* (N°N2017-06), note remise à monsieur le ministre de l'Éducation nationale.

IGÉSR. (2017). *L'outil livret scolaire unique et son impact sur les modalités d'évaluation* (N°2017-092).

IGÉSR. (2021). *L'évaluation des élèves au lycée (général, technologique et professionnel)* (N°2021-159).

IGÉSR. (2021). *Le parcours de l'élève au collège, formation, engagement, éducation formelle et non formelle* (N°2021-178).

IGÉSR. (2022). *L'enseignement en cours moyen : État des lieux et besoins* (N° 2022-048 ; p. 81). https://www.education.gouv.fr/media/113766/download

Loi n°2005-380 du 23 avril 2005 d'orientation et de programme pour l'avenir de l'école, NOR : MENX0400282L (2005). https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000259787

Loi n°2013-595 du 8 juillet 2013 d'orientation et de programmation pour la refondation de l'école de la République, NOR : MENX1241105L, JORF n°0157 (2013). https://www.legifrance.gouv.fr/loda/id/JORFTEXT000027677984/

Loi n° 2019-791 du 26 juillet 2019 pour une école de la confiance, NOR : MENX1828765L, JORF n°0174 (2019). https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000038829065/

Ministère de l'Éducation nationale, de la Jeunesse et des Sports. (2016). *Document d'accompagnement pour l'évaluation des acquis du socle commun de connaissances, de compétences et de culture. Éléments pour l'appréciation du niveau de maîtrise satisfaisant en fin de cycle 4*. Éduscol. https://eduscol.education.fr/document/17320/download

Ministère de l'Éducation nationale, de la Jeunesse et des Sports. (2020). *Programme d'enseignement de l'école maternelle (cycle 1) (Annexe de l'arrêté du 17/07/2020)*. https://cache.media.education.gouv.fr/file/31/89/4/ensel712_annexe_1312894.pdf

Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. (s. d.). *Évaluer à l'école maternelle. De l'observation instrumentée au carnet de suivi*. https://eduscol.education.fr/document/5640/download?attachment

Modalités d'évaluation des acquis scolaires des élèves. (2020). [Site institutionnel]. Éduscol. https://eduscol.education.fr/141/modalites-d-evaluation-des-acquis-scolaires-des-eleves

Note de service du 28-7-2021 portant sur les modalités d'évaluation des candidats à compter de la session 2022, NOR : MENE2121270N, BO n°30 (2021). https://www.education.gouv.fr/bo/21/Hebdo30/MENE2121270N.htm

OECD. (2019). *Résultats de TALIS 2018 (Volume I) Des enseignants et chefs d'établissement en formation à vie*. OECD Publishing. https://doi.org/10.1787/5bb21b3a-fr

Principes d'action pour évaluer les acquis des élèves. (s. d.). Éduscol. https://cache.media.eduscol.education.fr/file/socle_commun/43/1/Principes-action-evaluation-du-socle_1314431.pdf

Roediger, H. L. & Karpicke, J. D. (2006). The Power of Testing Memory : Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, *1*(3), 181-210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Sgen-Cfdt. (2020). L'évaluation, angle mort des politiques éducatives. *Profession Éducation*, *276*, 24. https://sgen-cfdt.fr/contenu/uploads/2020/10/PE276_compressed.pdf

Suivi et évaluation des apprentissages des élèves à l'école maternelle. (2022). [Site institutionnel]. Éduscol. https://eduscol.education.fr/107/suivi-et-evaluation-des-apprentissages-des-eleves-l-ecole-maternelle

Quelles sont les pratiques et les représentations de l'évaluation chez les enseignants ?

SOPHIE GENELOT

UNIVERSITÉ DE BOURGOGNE (FRANCE), INSTITUT DE RECHERCHE SUR L'ÉDUCATION (IREDU)

Introduction

Dans le cadre de ses activités, le Cnesco a souhaité s'associer à l'Institut de recherche sur l'éducation (Iredu) en 2017 pour mener une recherche visant à documenter factuellement l'activité évaluative dans l'enseignement élémentaire et secondaire inférieur (collège) en France. Cette recherche a concerné un nombre limité de classes (voir détail dans l'encadré méthodologique ci-dessous), mais s'est appuyée sur une observation longue de celles-ci, en évitant de catégoriser l'activité évaluative *a priori* préférant la saisir par le repérage, au cours du déroulement quotidien de l'activité des classes, de situations impliquant dans une dynamique interactive, l'action des enseignants et celle des élèves.

C'est en référence à une acception large de l'activité évaluative qui tient compte de ses formes les plus diffuses, notamment celles constituées par les interactions évaluatives (ou dialogues évaluatifs comme les nomme Mottier Lopez, 2015) entre enseignants et élèves apparaissant au gré de leur activité quotidienne que nous avons organisé cette observation. Elle s'est structurée autour de la notion de situation évaluative, définie comme suit : « toute situation dans laquelle l'enseignant et/ou les élèves prélèvent des informations à propos des apprentissages en cours, les interprètent en fonction de critères (plus ou moins explicites) pour formuler une appréciation qui est communiquée et fonde une prise de décision (plus ou moins explicite ; différée ou immédiate) » (Genelot, 2019). Sur l'ensemble des classes observées nous avons identifié et caractérisé un corpus de 351 situations évaluatives. Des entretiens avec les enseignants réalisés à l'issue de la période d'observation ont permis d'accéder aux représentations de leur pratique évaluative que nous avons pu confronter aux pratiques observées dans leurs classes.

Encadré méthodologique

Cette recherche a été conduite à la demande du Cnesco par une équipe coordonnée par une chercheuse de l'Iredu (Université de Bourgogne), le volet qualitatif a consisté à observer *in situ* 13 classes de façon indirecte (via la captation vidéo des séances de classe en continu réalisée par les enseignants eux-mêmes) : sept à l'école élémentaire (4 de cycle 2 et 3 de cycle 3 ; 3 en REP, une en milieu rural et 3 en urbain non REP) et 6 classes de collège (3 en 6e et 3 en 3e ; 2 en français, 2 en mathématiques et 2 en anglais ; 2 en REP, une en milieu rural et 3 en urbain non REP) sur un temps long (une semaine de classe pour l'école élémentaire, une séquence de cours pour le collège).

Le recueil de données s'est déroulé entre début mars et mi-avril 2017 et a permis de collecter environ 135 heures d'enregistrement vidéo et audio, des entretiens post-observation auprès de chaque enseignant observé (pratiques auto rapportées), et un grand nombre de documents confiés par les enseignants (« que vous jugez utiles à la compréhension de vos pratiques évaluatives ») : ces deux derniers types de données donnant accès aux représentations des enseignants.

Cette recherche a été conduite selon une approche collaborative grâce à la constitution d'un groupe composé de différents professionnels de l'Éducation nationale qui a été chargé de la coordination recherche/terrain, de l'élaboration et la mise en œuvre du protocole d'observation et a participé à l'élaboration des outils d'analyse des données ainsi qu'à leur codage.

Cette recherche a fait l'objet d'un rapport remis au Cnesco en mars 2019.

Les enseignants des classes observées qui le souhaitaient ont participé entre 2017 et 2019 à quatre journées de formation au cours de laquelle les résultats ont été présentés et discutés, complétés par la présentation des travaux de trois autres chercheurs. Cette recherche a été également intégrée à la formation initiale des enseignants : elle fait l'objet depuis 2017 de séminaires d'initiation à la recherche dans les masters MEEF premier et second degré de l'Inspé de Bourgogne. Deux mémoires de recherche du M2 MEEF mention PIF (Formation de formateurs) soutenus en 2020 ont porté sur l'analyse secondaire d'une partie du corpus des données.

I. L'activité évaluative : une part importante du temps d'enseignement

Nos analyses permettent tout d'abord de constater que dans l'ensemble des 13 classes observées et sur la période considérée, l'activité évaluative (au sens entendu dans cette recherche) occupe un peu moins de la moitié du temps d'enseignement (48,5 %). Néanmoins, cette part varie de façon plus ou moins importante selon les classes : elle apparaît, en moyenne, plus importante dans les classes de collège (55,2 %) que dans celles de niveau élémentaire (45,3 %) et plus importante en CM1-CM2 (entre 55 et 59 %) qu'en cycle 2 (entre 29 % et 45 %). Au collège, semble s'ajouter également, une variabilité en fonction des disciplines concernées : c'est dans les deux classes de mathématiques que le temps passé à l'évaluation est le plus important (61 % et 82 %), il l'est un peu moins dans les deux classes d'anglais (aux environs de 59 %) et représente moins de la moitié du temps d'enseignement dans celles de français.

II. Trois principales catégories de situations évaluatives

L'analyse du corpus des 351 situations évaluatives initialement repérées, conduite selon une démarche inductive, a opéré par regroupements progressifs et successifs selon deux dimensions descriptives principales :

1. le caractère plus ou moins explicite de leur dimension évaluative ;
2. le caractère plus ou moins formel de leurs différentes caractéristiques.

On aboutit à l'élaboration d'une typologie en trois catégories principales décrites dans le Tableau ci-dessous :

Tableau 1 : Typologie des situations évaluatives observées dans les classes de l'échantillon

Situations évaluatives formalisées explicites	**Situations évaluatives interactives**					
	Différées Interaction enseignant/élève(s) engagée **à l'issue** de l'activité des élèves			**Simultanées** Interaction enseignant/élève(s)engagée **au cours** de la réalisation de la tâche		
	Assez formelles, présentées plus ou moins souvent de façon explicitement évaluative aux élèves			**Les plus informelles,** pas toujours présentées aux élèves, le plus souvent sans référence explicite à l'évaluation		
	Individuelle privée	**Individuelle publique**	**Collective**	**Collective**	**Individuelle**	**Mixte**
Modalités (organisation matérielle, temporelle, formes d'appréciations, etc.) **très formelles** et **présentation explicitement évaluative aux élèves** - Évaluations finales (fin de séquence) - Évaluations intermédiaires (en cours de séquence) - Évaluations ritualisées (fréquentes et régulières) ➔ Intitulés variés et nombreux	L'intervention de l'enseignant concerne la production d'un seul élève à la fois. Rétroactions adressées uniquement à cet élève, pas ou peu audibles du reste de la classe.	L'intervention de l'enseignant concerne la production d'un seul élève à la fois. Rétroactions adressées principalement à cet élève devant la classe audibles par les autres élèves.	L'enseignant réalise la tâche devant l'ensemble de la classe sous la dictée des élèves qu'il sollicite successivement, ou sollicite les commentaires des élèves à propos d'une (ou plusieurs) production(s) d'élève(s).	Interaction entre l'enseignant et l'ensemble de la classe (ou un groupe d'élèves).	Interaction entre l'enseignant et un ou successivement plusieurs élève(s).	L'enseignant en interaction individuelle avec un (ou successivement plusieurs) élève(s), puis formule un commentaire à l'ensemble de la classe (rappel reformulation consigne, explication, etc.).

A. Les situations évaluatives formalisées explicites

Des trois catégories construites ce sont, sans conteste, les situations évaluatives formalisées explicites qui semblent, dans un premier temps, constituer le « cœur » de l'activité évaluative des classes : ce sont celles que les enseignants évoquent prioritairement (quelquefois les seules) et le plus spontanément lors des entretiens, celles qu'ils nous décrivent le plus longuement et le plus précisément. Elles font presque toujours l'objet d'une présentation aux élèves (seules 2,4 % d'entre elles ne sont pas présentées aux élèves) sous une forme en grande majorité explicitement évaluative (près des trois quarts) ou ritualisée (17 %).

Ce sont également celles dont la dénomination (en classe ou au cours de l'entretien) fait l'objet du plus grand nombre d'intitulés : on trouve à la fois des termes assez génériques qui se réfèrent très explicitement à l'évaluation ou à des outils d'évaluation (exemples : « contrôle », « évaluation », « cahier de réussite », « évaluation formelle », « évaluation institutionnelle », « évaluation sommative », « autoévaluation », « évaluation informelle »)[46]. Certains autres termes, sans faire explicitement référence à l'évaluation, y sont cependant explicitement associés dans la culture didactique de certaines disciplines (« tâches finales », « tâches intermédiaires », en anglais ; « DS/devoir surveillé », « DM/devoir à la maison », « calcul mental » en mathématiques ; « dictée » en français) ou dans la culture scolaire en général (« révision », « interrogation », « exposé »). D'autres intitulés enfin désignent des dispositifs d'évaluation particuliers (« contrôle flash », « quiz-quiz trade », « question flash », « teacher assistant », « activité mentale », « îlots bonifiés ») dont certains prennent une forme compétitive teintée d'une dimension ludique (« Défi-table », « Défi du jour », « Challenge group », « Joker », « Twictée »).

Nous pourrions donc qualifier ces situations comme de « l'évaluation qui dit son nom », même si, par ailleurs certains enseignants « édulcorent » un peu cette dénomination en employant des termes plus « neutres » devant leurs élèves. Tel cet enseignant de cycle 3 qui décrit, lors de l'entretien, ses « évaluations formelles » en fin de séquence de mathématiques et de français mais qui, lorsqu'il les présente aux élèves, parle « d'exercices à réaliser en autonomie » ou annonce « passer aux choses sérieuses ». C'est également le cas d'une enseignante de cycle 2 qui lors de l'évaluation en connaissance du monde, présente cette activité aux élèves comme étant « des exercices pour savoir ce que vous avez retenu ». Car en effet, certains enseignants expriment une certaine prudence, voire de la méfiance vis-à-vis de ce type de pratique ; à cet égard on peut repérer trois positionnements différents parmi les enseignants de notre échantillon :

- Un positionnement « sans états d'âme » : exprimé par des enseignants qui présentent des choix affirmés, émettent peu de doutes à l'égard de ces pratiques qui semblent relativement stables. On trouve cette attitude surtout chez trois enseignantes de collège (les deux en anglais et celle de mathématiques en sixième) et une enseignante de CM1/CM2.
- Un positionnement « en recherche » : chez des enseignants qui expriment de nombreuses interrogations vis-à-vis de leurs pratiques évaluatives de ce type, notamment à l'égard des élèves en difficulté : ils multiplient les dispositifs différents et leurs pratiques ne semblent pas très stabilisées. On trouve parmi eux deux enseignantes de collège (celle de français en sixième

[46] Les propos et expressions retranscrits entre guillemets et en italique sont ceux employés par les enseignants soit en classe, soit lors de leurs entretiens.

et celle de mathématiques en troisième) et trois enseignantes d'école élémentaire (CP ; CP/CE1 et CM1/CM2).

- Un positionnement de « distance critique » : chez des enseignants adoptant une attitude très prudente, voire défiante, vis-à-vis de cette forme d'évaluation : ils la pratiquent peu et la questionnent beaucoup. On trouve ce positionnement chez une seule enseignante de collège (en français en troisième) et trois enseignants d'école élémentaire (CP ; CP/CE1 et CM2, tous en réseau d'éducation prioritaire (REP)).

B. Les situations évaluatives interactives différées

La catégorie **«** interactive différée **»** est principalement constituée de dialogues évaluatifs entre l'enseignant et un ou plusieurs élèves, engagés à l'issue de l'activité des élèves. Les interactions interviennent donc postérieurement à la production par les élèves de la tâche prescrite, réalisée en autonomie : l'enseignant « s'abstenant » d'intervenir pendant cette phase. Ce « retrait » de l'enseignant pendant la phase de production n'est pas sans rappeler les situations évaluatives formalisées explicites mais ces situations-ci s'en distinguent, cependant, par le caractère non explicitement évaluatif de la présentation des tâches par l'enseignant (exemple : « on va faire un exercice, on va s'entraîner, on va réviser… »). L'enseignant s'englobant souvent, par l'emploi du « on », dans le processus de production, indiquant par là même, que l'autonomie de l'élève dans la réalisation de la tâche est relative.

Le fait que l'enseignant n'intervienne qu'*a minima* lors de la phase de production des élèves ne veut pas dire qu'il soit pour autant inactif. Le plus souvent il adopte une posture d'observation des élèves, quelquefois discrète, d'autres fois plus marquée en circulant et en regardant les élèves réaliser les tâches prescrites mais il ne formule pas d'appréciation à ce moment-là. On repère également des situations où l'enseignant semble rester volontairement très en retrait pendant la phase de production des élèves attendant qu'ils aient terminé pour prendre connaissance de ce qu'ils ont produit[47].

Ces situations évaluatives interactives différées s'apparentent, pour partie, à l'une des activités la plus courante de la forme scolaire : la « correction » des productions des élèves. L'analyse des entretiens montre combien ce terme est polysémique dans l'usage qu'en font les enseignants et nos observations révèlent le caractère protéiforme de ces situations, aussi bien du point de vue des modalités de participation des élèves, des supports mobilisés, des formes d'appréciations ou encore des types d'intervention des enseignants. De ce point de vue, nous en avons repéré trois modalités typiques (Tableau 1) selon que l'interaction avec l'enseignant concerne toute la classe ou bien un seul élève (ou plusieurs successivement), de façon publique (exemple typique de « l'élève au tableau ») ou privée (en « tête à tête » avec l'enseignant).

Seuls sept enseignants de l'échantillon évoquent ce type de situation lorsqu'ils décrivent leurs pratiques évaluatives lors des entretiens (deux enseignants de collège et cinq enseignants de classe élémentaire). La plupart de ces derniers expriment à cet égard une conception très précise de cette pratique, qui apparaît très conscientisée et qu'ils relient assez explicitement à leurs pratiques évaluatives. C'est le cas, par exemple de cet enseignant d'école élémentaire :

[47] C'est également souvent le cas dans les classes élémentaires à cours doubles dans lesquelles les enseignants sont souvent occupés par un groupe d'élèves, tandis que l'autre réalise une tâche en autonomie.

« (…) C'est une évidence, pour moi, mais je sais pas, elle (l'évaluation) doit d'abord être utile pour les élèves, donc **pour moi, une correction collective, c'est déjà une évaluation**[48], c'est-à-dire que si l'on repasse en revue une activité ou un exercice que l'on a pu faire individuellement ou en binôme, peu importe le mode de travail, si on repasse collectivement derrière pour voir qui a pu faire des erreurs, pourquoi, échanger sur les procédures, les outils que l'on a pu utiliser, c'est déjà une évaluation et ça apporte beaucoup d'informations sur qui sait pertinemment et est autonome, sait utiliser ces outils là, au final, c'est ce que l'on demande aux élèves, c'est d'accéder à l'autonomie, donc voilà, je suis plutôt… C'est plutôt mon fonctionnement « naturel » … »
(Enseignant de CP/CE1, en REP)

C. Les situations évaluatives interactives simultanées

La catégorie « interactive simultanée » regroupe les situations évaluatives les plus informelles : la moitié d'entre-elles ne font l'objet d'aucune présentation aux élèves. Lorsqu'elles sont présentées aux élèves, c'est le plus souvent sans référence explicite à l'évaluation (22 %) ou sous une forme ritualisée (14,5 %), moins souvent de façon explicitement évaluative (13 %).

Ces situations sont constituées de dialogues évaluatifs entre l'enseignant et un (ou plusieurs) élève(s), tout comme les interactions évaluatives différées, mais elles s'en distinguent du point de vue du moment auquel intervient l'interaction avec l'enseignant par rapport à la réalisation de la tâche par le ou les élève(s). Une partie d'entre elles concernent des « exercices d'application » faisant suite à une phase de découverte de procédures ou de notions, que les élèves réalisent par exemple individuellement à l'écrit. Dans ce cas, la seule différence avec les interactions évaluatives de type différé réside dans le fait qu'au lieu d'attendre que les élèves aient terminé leurs exercices, l'enseignant interagit avec eux pendant qu'ils sont au travail. Pour ce faire le plus souvent, il circule entre les rangs et interagit successivement avec un, ou plusieurs, élèves. Mais on peut également trouver ces situations sous forme collective, comme par exemple dans cette classe de CM1/CM2, lors de « débats interprétatifs », par petits groupes, au cours desquels les élèves interagissent, entre eux et avec l'enseignante, pour argumenter leur compréhension du texte littéraire qu'ils sont en train d'étudier.

C'est dans cette catégorie d'interactions évaluatives simultanées que l'on observe d'ailleurs le plus d'activités orales, comme par exemple dans la classe de troisième en anglais avec une activité assez systématique du début des séances (« évaluation informelle ») au cours de laquelle par une série de questions/réponses orales très rapides, l'enseignante amène les élèves à réactiver le lexique et les structures syntaxiques étudiés dans les séances précédentes.

À la différence des situations évaluatives interactives différées, celles de type simultané ne se réduisent pas aux seules mises en exercice des élèves mais concernent une gamme beaucoup plus étendue d'activités scolaires : par exemple lors de l'élaboration d'une « trace écrite » collective, destinée à synthétiser un savoir commun à la classe (observé par exemple dans les deux classes de français au collège mais également dans certaines classes de l'école élémentaire en sciences, histoire, etc.). Lors de ces situations, les interventions des enseignants vont consister, le plus souvent, à faire rappeler, par interactions sous forme de questions/réponses avec la classe, les acquis de séances précédentes. Leurs rétroactions vont servir à faire préciser les propositions des élèves, en rapport avec des critères quelquefois très explicites, afin de valider des acquis à titre collectif.

[48] Les propos retranscrits en gras sont soulignés par nous.

Seuls cinq enseignants se sont exprimés au sujet de ce type de situations lors des entretiens. En croisant leurs propos avec les observations de leurs classes, il apparaît assez clairement que cette pratique n'est explicitement conscientisée comme évaluative que par les enseignants des classes où elle est la plus fréquente, comme par exemple dans celle de 6e en mathématiques dont l'enseignante décrit très explicitement sa démarche :

> « Alors, en général, je construis mon cours pour les faire s'entraîner sur chaque notion, je leur fais faire des exercices donc tout au long du chapitre, ils ont des exercices d'application puis des exercices qui vont être un peu plus complexes où ils vont devoir mettre en corrélation plusieurs notions différentes et on passe toujours par une phase d'entraînement, ces exercices sont des exercices d'application « pur et dur ». Lors de ces séances d'exercices qui peuvent… Ça peut durer 30 minutes, 40 minutes, peu importe, je passe dans les rangs régulièrement, je vérifie ce qui est écrit, ils posent leurs questions librement, je passe auprès d'eux pour vérifier, soit pour vérifier leurs résultats, soit parce qu'ils ont des difficultés… **Donc ça, c'est une première façon d'évaluer leur niveau d'acquisition, où est-ce qu'ils en sont**… »
> (Enseignante de mathématiques en 6e)

Ou encore cette enseignante de cette classe de CP (REP) dans laquelle ces situations évaluatives interactives simultanées sont les plus fréquentes de tout l'échantillon :

> Enseignante de CP : « J'essaie, quand je peux et je pense que je l'ai fait dans toutes les vidéos, de corriger avec les enfants les fichiers. Je passe dans les rangs et quand l'enfant a bon, je dis que c'est bon, s'il se trompe, j'essaie de lui expliquer où il s'est trompé, pourquoi il s'est trompé et je refais verbaliser, donc… » (…) « Ouais, parce que je me dis que corriger quand ils sont pas là, ils ne reviennent pas dessus, ils ne comprennent pas, cela n'a pas de sens et je pense que le sens, c'est important pour la compréhension, donc j'essaie de faire ça (…). Donc c'est un peu long, cela demande du temps, souvent, les enfants s'impatientent, c'est l'aspect un peu négatif, mais c'est formateur pour les élèves, donc c'est ce que j'essaie de faire quoi. »
> Enquêtrice : « D'accord. Donc, tu passes à côté d'eux pendant qu'ils sont en train de faire et tu interviens même s'ils ont pas terminé, par exemple, un exercice… »
> Enseignante de CP : « Oui, je lui dis « oh regarde ! Tu as écrit ça là, pourquoi tu as écrit ça comme ça ? » donc voilà… Des fois, c'est bon aussi ! Je n'interviens pas seulement quand ils se trompent, j'interviens aussi quand c'est bon pour qu'ils verbalisent et pour qu'ils m'expliquent. »
> Enquêtrice : « Donc, en fait, systématiquement, tu circules quoi, en fait ? Dans ces périodes d'exercice, tu circules systématiquement et tu interviens… »
> Enseignante de CP : « Voilà, et **c'est une manière pour moi d'évaluer leur travail.** »

III. Différentes catégories de situations évaluatives inégalement représentées selon les niveaux scolaires et les disciplines concernés

Si, dans l'ensemble de l'échantillon, ces différentes catégories de situations évaluatives se répartissent de façon sensiblement égale (35 % du corpus pour les situations évaluatives interactives simultanées, 35 % également pour les situations évaluatives formalisées explicites et un peu moins de 30 % pour les situations évaluatives interactives différées), cette représentativité des trois catégories de situations évaluative diffère selon le niveau de scolarisation (voir Tableau 2).

Tableau 2 : Répartition des situations évaluatives (en % du nombre de situations évaluatives) selon les niveaux scolaires

« Profil » classes d'école élémentaire	« Profil » classes de collège	Ensemble des classes
Situations évaluatives formalisées explicites 28 %	Situations évaluatives formalisées explicites 47 %	Situations évaluatives formalisées explicites 35 %
Situations évaluatives interactives simultanées 44 %	Situations évaluatives interactives simultanées 21 %	Situations évaluatives interactives simultanées 35 %
Situations évaluatives interactives différées 28 %	Situations évaluatives interactives différées 32 %	Situations évaluatives interactives différées 30 %

Note : Les situations évaluatives les plus fréquentes pour chaque colonne ont été soulignées.

En effet, si dans les propos des enseignants, les situations évaluatives formalisées explicites semblent prépondérantes, au plan des pratiques elles ne constituent cependant pas, dans toutes les classes, la part la plus importante de leur activité évaluative.

C'est surtout dans les classes de collège qu'elles apparaissent nettement majoritaires (voir Tableau 2), spécialement dans les classes de troisième où elles représentent plus de 60 % des situations évaluatives observées (voir Tableau 3), mais elles ne sont majoritaires que dans les classes d'anglais (en 6e comme en 3e) et de mathématiques (en 3e).

Dans les classes élémentaires, en revanche, c'est sous la forme d'interactions évaluatives entre les enseignants et les élèves que se déroule, très majoritairement, l'activité évaluative (44 % pour les formes simultanées et 28 % pour les formes différées). Dans la plupart des classes (5 sur 7), ces interactions évaluatives se déroulent principalement sous une forme simultanée plutôt que sous une forme différée : c'est particulièrement le cas dans les deux classes de cycle 2 situées en REP (plus de 60 % de situations de ce type). Sur l'ensemble des trois classes de cycle 3, même si les formes interactives simultanées sont également les plus représentées (39 %), le profil des situations évaluatives est plus équilibré entre les trois catégories (33% pour les situations interactives différées et 28% pour les situations formalisées explicites).

Les situations évaluatives interactives s'observent également au niveau du collège (voir Tableau 3), spécialement en 6e (majoritaires dans les classes de français et de mathématiques à ce niveau) mais sous la forme différée principalement (39 % en 6e et 32 % en 3e); les interactions évaluatives simultanées n'apparaissent que de façon très anecdotique au niveau du collège (14%), excepté dans la classe de mathématiques en 6e REP (près de 40 % des situations observées dans cette classe).

Tableau 3 : Répartition des situations évaluatives (en % du nombre de situations évaluatives) dans les six classes de collège

« Profil » 6e	« Profil » 3e	Dans les classes de collège
Situations évaluatives formalisées explicites 34 %	Situations évaluatives formalisées explicites 61 %	Situations évaluatives formalisées explicites 47 %
Situations évaluatives interactives différées 39 %	Situations évaluatives interactives différées 24 %	Situations évaluatives interactives différées 32 %
Situations évaluatives interactives simultanées 27 %	Situations évaluatives interactives simultanées 14 %	Situations évaluatives interactives simultanées 21 %

Note : Les situations évaluatives les plus fréquentes pour chaque colonne ont été soulignées.

Enfin, sur l'ensemble des classes de l'échantillon, l'analyse des différentes formes typiques repérées dans chacune de ces deux catégories de situations évaluatives interactives, révèle une sorte de « continuum » dans l'usage différencié de celles-ci au cours de la scolarité obligatoire. À la forme individuelle, privilégiée dans les petites classes de l'école élémentaire de notre échantillon, semble se substituer la forme collective, majoritaire au collège, en passant par des modalités intermédiaires, plus « mixtes », qui s'observent plutôt à la « charnière » des deux degrés d'enseignement (en cycle3).

IV. Différentes catégories de situations évaluatives inégalement représentées selon les classes

Au-delà des tendances générales décrites ci-dessus pour chaque degré scolaire (qu'il faut prendre avec prudence au vu de la petite taille de l'échantillon de classes concernées), l'analyse plus précise classe par classe permet, par exemple, de relativiser certaines observations formulées précédemment. La prédominance des situations interactives simultanées repérée, par exemple, sur l'ensemble du corpus de l'école élémentaire ne s'observe en fait que dans quatre des sept classes. Peut-être est-ce là une caractéristique davantage spécifique à certains enseignants et/ou à certains contextes de ces classes : dans ce groupe on trouve, par exemple, les trois classes de REP et la seule de milieu rural.

Cette approche classe par classe au niveau élémentaire permet de faire émerger trois « profils d'activité évaluative » différents :

- ***Une activité évaluative réduite et fortement formalisée et explicite*** (dans deux classes de CP et CP-CE1) : l'activité évaluative y représente une part assez faible du temps de classe (moins de 40 %) mais sous une forme, majoritairement, formalisée explicite (40 et 62 %).
- ***Une activité évaluative modérée*** principalement ***sous une forme interactive simultanée*** (dans les deux autres classes de CP et CP-CE1 de REP) : la part consacrée à l'activité évaluative y est moyenne (entre 43 et 45 % du temps de classe) réalisée très majoritairement sous une forme interactive simultanée (à plus de 60 %) avec très peu d'évaluations formalisées explicites
- ***Une activité évaluative importante*** (plus de la moitié du temps d'enseignement) dans les trois classes de CM : ***soit sous une forme assez équilibrée*** entre les trois catégories d'évaluation (pour une classe), ***soit*** privilégiant ***majoritairement la forme interactive simultanée*** (la classe de REP et la classe rurale) et accordant une part réduite à l'évaluation formalisée explicite.

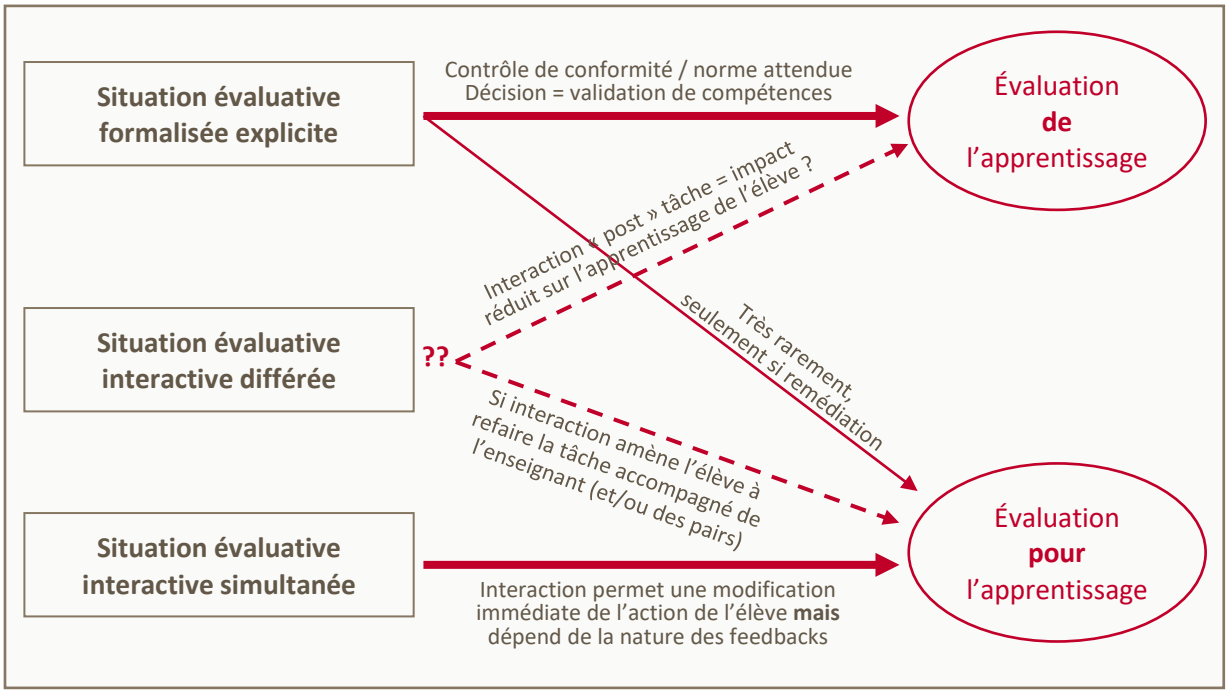
Au niveau du collège, la prédominance des évaluations formalisées explicites repérée globalement sur l'ensemble des six classes, ne s'observe en réalité que dans la moitié d'entre elles : les deux classes d'anglais et la classe de 3e en mathématiques. Cette dernière présente la plus forte présence de cette catégorie de tout l'échantillon et celle dans laquelle l'activité évaluative est également la plus fréquente (82 % du temps), ce qui constitue un profil évaluatif très particulier. Contrairement à l'école élémentaire, il est difficile de repérer des groupes de profils particuliers. Les régularités éventuelles semblent s'opérer, comme décrit plus haut, davantage au plan des disciplines concernées ou par niveau (6e/3e) que par un profil spécifique par classe.

Tout au plus, peut-on signaler la particularité de la classe de mathématique en 6e REP qui est la seule de l'échantillon (tous niveaux confondus) dans laquelle les situations évaluatives interactives différées sont majoritaires et celle où les formes interactives simultanées sont les plus fréquentes au niveau du collège : ces caractéristiques semblent la rapprocher davantage des profils évaluatifs des classes élémentaires (notamment celles de CM) que celles du collège. Là encore s'agit-il d'un effet du contexte REP (mais on n'observe pas le même profil dans l'autre classe de collège REP) ou bien d'une pratique spécifique à un enseignant ? Nos données ne permettent pas de trancher.

V. Quels liens avec les apprentissages des élèves ?

Pour faire le lien entre ces données et la thématique de la conférence de consensus, nous proposons d'analyser les relations potentielles entre chacun des trois types de situations évaluatives observées et les deux principales fonctions de l'évaluation décrites par la littérature scientifique anglosaxonne : « *assessment of learning* » *vs* « *assessment for learning* », traduites en français par « évaluation de l'apprentissage »[49] *vs* « évaluation pour l'apprentissage ou évaluation-soutien d'apprentissage » (EsA) (Allal & Laveault, 2009). Cette analyse est résumée par le schéma ci-dessous :

Figure 1 : Liens entre typologie des situations évaluatives observées et fonctions de l'évaluation



[49] Pour une discussion plus approfondie sur ces différents termes, voir Fagnant, 2023 (rapport produit dans le cadre de la conférence de consensus du Cnesco sur l'évaluation en classe au service de l'apprentissage des élèves).

Dans nos observations il apparaît assez clairement que les situations évaluatives formalisées explicites relèvent principalement d'une évaluation de l'apprentissage visant prioritairement la validation des compétences des élèves en référence aux normes visées par les programmes. C'est particulièrement clair pour les situations qui se déroulent en fin de séquence, mais également pour celles qui interviennent en cours de celles-ci, y compris sous des formes plus régulières, quelquefois même ritualisées, qui s'apparentent davantage à une évaluation plus continue, comme celles décrites, par exemple, par deux enseignantes de collège :

> Enseignante de mathématique en 3e : « Ce que j'appelle des activités mentales, ce sont des petites capsules que je mets au tableau, qui sont vidéo-projetées, qui sont des documents Powerpoint où il va y avoir trois ou quatre questions rapides auxquelles ils vont devoir répondre, où je n'exige aucune justification, aucune rédaction, je veux juste la réponse. (…) Sur ces activités mentales, je les fais par séries de trois, donc une par semaine et les trois activités mentales d'une même série vont porter sur les mêmes notions. Donc la première et la deuxième sont des activités mentales d'entraînement et la troisième est évaluée ».

> Enseignante de français en 6e : « Le contrôle flash, (…) on pose cinq questions, je leur ai fait apporter des ardoises et ils répondent « vrai » ou « faux », c'est un vrai/faux quoi sur les points essentiels et après, on y revient, cela prend vraiment deux minutes, on revient sur chacune des propositions, c'est vrai ou c'est faux puis ça permet de réactiver ce qui était fait à la séance précédente »

Bien que, comme le souligne l'une de ces enseignantes, ces situations évaluatives les renseignent sur l'avancée de leurs élèves dans les apprentissages visés (« le contrôle flash, l'avantage, c'est que cela permet d'avoir une idée de l'état de la classe à ce moment-là de la séquence. » Enseignante de français, 6e), nous n'observons pas qu'elles provoquent pour autant une régulation de leurs enseignements : nous n'avons constaté, par exemple, aucune modification substantielle de la progression pédagogique initialement programmée par les enseignants[50], pas plus que des pratiques de différenciation pédagogique suite à des évaluations de ce type.

Dans de rares cas cependant, on observe une action de remédiation leur faisant suite ; par exemple, lors du rendu d'un devoir sur table de maths, l'enseignante de 3e, constatant qu'un des exercices n'a été que très peu réussi, propose aux élèves de s'entraîner à nouveau sur le même type de tâche avant de retenter la validation de la compétence. Dans ce cas cette évaluation ne se contente donc pas du constat de réussite ou d'échec : l'enseignante remet les élèves sur le même type de tâche et apporte, par ses interactions avec eux, l'étayage nécessaire à leur progression dans la compétence visée. Mais, ces cas-ci sont extrêmement rares dans les situations évaluatives formalisées explicites, qui se terminent généralement par la communication aux élèves de l'appréciation (traduite par une note chiffrée, un niveau d'acquisition, une couleur ou une appréciation écrite), validant ou pas la (les) compétence(s) visée(s), et par une correction collective rapide où l'enseignant cherche avant tout à communiquer « les bonnes réponses » aux élèves. Nous sommes donc là dans une évaluation à visée plus informative, comme la dénomme Mottier Lopez (2015), que formative.

À l'inverse et tout aussi clairement, il nous semble que les situations évaluatives interactives simultanées que nous avons observées dans les classes relèvent très majoritairement d'une évaluation

[50] Dans les documents confiés par les enseignants, il y avait souvent leurs fiches de préparation ou de programmation : en confrontant ces documents et les pratiques effectives filmées, nous n'avons constaté aucun cas de modification du déroulement des séances initialement prévu suite à une situation évaluative formalisée explicite.

pour l'apprentissage. En effet, le fait que l'interaction avec l'enseignant intervienne simultanément à la production de l'élève est particulièrement propice à la modification de l'action de celui-ci et impacte donc nécessairement son apprentissage (sa compréhension de la consigne, des procédures, le rappel des savoirs nécessaires à la réalisation de la tâche, etc.). L'interaction avec l'élève au cours de la réalisation de la tâche permet également une adaptation très souple de celle-ci en fonction des performances de ceux-ci comme nous avons pu l'observer, par exemple, dans le cas du travail sur ardoise. À la différence d'exercices du même type proposés à l'écrit (par exemple, quatre verbes à conjuguer au cahier de brouillon), la visibilité simultanée de toutes les réponses des élèves, permet à l'enseignant une adaptation immédiate des tâches proposées : selon que celles-ci sont plus ou moins réussies par l'ensemble des élèves, il va, par exemple, apporter l'aide nécessaire à la compréhension de leurs erreurs par les élèves et proposer à nouveau une tâche de même difficulté, pour conforter leurs acquis. Ou bien, s'il constate après une ou deux tâches du même type, que celles-ci sont réussies par l'ensemble des élèves, il va passer à des tâches plus variées ou plus complexes. Cette adaptation immédiate de l'action de l'enseignant aux difficultés rencontrées par les élèves s'observe particulièrement dans les situations évaluatives interactives simultanées de type mixte : l'enseignant intervient d'abord sous une forme individuelle en se déplaçant par exemple vers les élèves en train d'effectuer une tâche écrite et, souvent après avoir constaté une difficulté récurrente, stoppe le travail des élèves et régule leur activité en apportant à l'ensemble du groupe des précisions nécessaires pour la compréhension de la consigne, ou bien une aide méthodologique, ou encore en faisant réaliser un exemple au tableau, etc.

Cependant, comme le souligne Mottier Lopez (2015), toute interaction avec l'enseignant n'entraîne pas systématiquement une régulation de l'activité de l'élève à visée formative, pour que ce soit le cas il faut que cette « évaluation-régulation interactive » constitue « une forme de participation guidée qui a une fonction d'étayage et de médiation à l'autorégulation de l'élève » (p. 79). De ce point de vue, on pourrait également convoquer les activités d'auto-évaluation et d'évaluation entre pairs, que de nombreux auteurs (Laveault, 1999, 2012 ; Coen & Bélair 2015), décrivent comme particulièrement favorables à la régulation des apprentissages des élèves. Or, bien qu'évoquées par la plupart des enseignants de notre échantillon lors des entretiens, ces pratiques restent des formes relativement marginales dans l'activité évaluative des classes que nous avons observées : nous avons relevé 23 situations de ce type (sur 351) dans seulement six classes de l'échantillon et dans quatre d'entre elles (3 de niveau élémentaire et la classe de 3e en mathématiques) de façon très anecdotique (une seule situation par classe). Seules deux classes (une de CM1/CM2 et la classe d'anglais en 6e) font exception par la fréquence plus importante de ce type de pratique (sous forme ritualisée pour la classe d'anglais : « *quiz quiz trade* », « *teacher assistant* ») qui semble relever d'un choix explicite de la part des deux enseignantes de ces classes et dont elles évoquent les effets positifs sur les apprentissages des élèves au cours de l'entretien.

Ainsi il ressort de nos observations que **les situations évaluatives interactives simultanées constituent, potentiellement, une forme d'évaluation favorable à la régulation de l'activité des élèves, mais que c'est la nature des feedbacks qui leurs sont adressés lors de ces situations qui est déterminante** pour qu'elles constituent un réel soutien à l'apprentissage des élèves.

Enfin, le cas des situations évaluatives interactives différées est plus délicat à analyser du point de vue de son impact potentiel sur les apprentissages des élèves. En effet, le fait que l'interaction avec l'enseignant n'intervienne qu'une fois la tâche terminée a comme conséquence que l'activité de l'élève ne s'en trouve

que faiblement impactée. Puisque ces situations recouvrent le plus souvent des activités de corrections des productions des élèves, prenons deux cas typiques « extrêmes » rencontrés dans notre corpus.

- Si la correction consiste à fournir à l'élève les réponses justes aux exercices réalisés, comme nous l'avons fréquemment observé lors de situations évaluatives interactives différées collectives (aussi bien dans les classes élémentaires que dans celles de collège), notamment à l'occasion de la correction d'exercices réalisés à la maison : alors la « plus-value » formative de la situation est faible. Elle relève, comme indiqué plus haut à propos des situations évaluatives formalisées explicites, d'une évaluation essentiellement informative : les élèves savent s'ils ont « juste ou faux » et prennent connaissance de la réponse juste, au mieux (mais ce n'est pas toujours le cas) assortie d'un commentaire démonstratif/justificatif de la part de l'enseignant. Il n'est pas certain cependant que tous les élèves ayant échoué initialement à la tâche soient en capacité de mieux la réussir suite à cette activité de correction.
- Si la correction consiste pour l'enseignant, comme nous l'avons vu pratiquer, par exemple dans une classe de CM1/CM2, à faire comparer par l'ensemble de la classe plusieurs productions réalisées par différents groupes d'élèves (il s'agissait de schémas expliquant une éclipse solaire) en favorisant par l'expression d'arguments contradictoires (« qui pense que cette production est correcte ? pourquoi oui ? pourquoi non ? ») l'émergence ou l'appropriation des critères de réussite de la tâche : il nous semble que ce type de situation soutient plus nettement le processus d'apprentissage des élèves.

Dans d'autres cas assez typiques de cette catégorie de situations évaluatives, « l'élève au tableau », la relation à l'apprentissage dépend du type d'élèves que l'enseignant choisit d'interroger. S'il choisit un élève en réussite, l'interaction se résume le plus souvent à la validation de la réponse (qui est souvent juste), mais si au contraire, il interroge un élève en difficulté l'amenant par un jeu de questions/réponses à devoir expliquer sa démarche, analyser ses erreurs et/ou à refaire la tâche avec l'accompagnement de l'enseignant et/ou des autres élèves : alors le soutien à l'apprentissage est plus net. Ces observations valent autant pour les situations évaluatives interactives différées collectives, comme dans les exemples cités plus haut, que lorsque l'interaction est adressée individuellement : dans les deux cas c'est bien la nature des feedbacks adressés aux élèves soit individuellement, soit collectivement, qui va déterminer le soutien potentiel à l'apprentissage.

Dans nos analyses des formes typiques des situations évaluatives interactives différées, nous avons tenu à distinguer cependant les formes individuelles privées (lorsque l'interaction de l'enseignant s'adresse à un élève en particulier mais n'est pas ou peu audible par les autres élèves) des formes individuelles publiques lorsque les interactions avec l'enseignant se déroulent devant le reste de la classe (l'élève interrogé à sa place ou au tableau). Dans ce cas en effet, la situation, en rendant publique les réputations d'infériorité de certains élèves (comme nous l'avons vu pratiquer par exemple en envoyant au tableau systématiquement les élèves qui répondent faux), devient potentiellement menaçante pour les apprentissages de ces élèves, *via* les perturbations attentionnelles qu'entraîne cette situation de comparaison sociale défavorable – comme l'ont montré les travaux sur la motivation (Ames & Archer, 1988 ; Sarrazin *et al.*, 2006) et de nombreuses recherches expérimentales (Butera & Darnon, 2011 ; Toczek & Souchal, 2017).

Ainsi donc, plus que la situation en elle-même, ce sont les différents choix que vont opérer les enseignants, dans leurs interactions avec les élèves, qui vont rendre celle-ci plus ou moins favorable à l'apprentissage des élèves.

Pour conclure

Ainsi donc, au plan des représentations, ce sont les modalités les plus formellement et explicitement évaluatives, l'évaluation « qui dit son nom », qui occupent la part prépondérante des réflexions des enseignants. Dans la réalité des pratiques telles que nous avons pu les observer dans cette recherche, ce sont cependant les formes les plus diffuses et les plus informelles de l'évaluation, celle qui « ne dit pas son nom », qui se déroulent lors d'activités ordinaires, voire banales, qui apparaissent constituer la modalité principale de l'activité évaluative de la plupart des classes de notre échantillon (9 classes sur 13).

Il nous semble que c'est là l'un des intérêts principaux de cette recherche empirique que de montrer **que l'activité évaluative des classes de la scolarité obligatoire ne se réduit pas, loin s'en faut, aux seules tâches d'évaluation choisies, élaborées et mises en œuvre consciemment et explicitement à cette fin par les enseignants. Elle transite également, pour une part non négligeable, par des pratiques plus informelles, pas nécessairement conscientisées par les enseignants sous leur aspect évaluatif**. Ce résultat contribue, nous semble-t-il, à contredire la représentation la plus courante de l'évaluation chez les enseignants, du moins ceux de notre échantillon, et à conclure qu'un certain nombre d'entre eux font sans doute souvent « de l'évaluation sans le savoir ».

C'est une particularité sur laquelle nous attirons l'attention des acteurs qui s'emploient à vouloir faire évoluer les pratiques évaluatives des enseignants : car comment faire évoluer des pratiques qui ne sont qu'en partie conscientisées comme évaluatives par les enseignants eux-mêmes ? D'autant que ce sont celles-ci qui, d'après nos observations, semblent potentiellement les plus à même d'impacter les apprentissages des élèves. Ces constats devraient, nous semble-t-il, inciter les enseignants à apporter une attention toute particulière aux choix qu'ils opèrent (pas toujours consciemment ?) dans leurs interactions avec leurs élèves (qui vais-je interroger ? comment dois-je intervenir ? à quel moment ? auprès de quels élèves ? selon quelle modalité ? quelles questions poser ? …).

Références

Allal, L. & Laveault, D. (2009). Assessment for Learning: Évaluation-soutien d'Apprentissage. *Mesure et évaluation en éducation*, *32*(2), 99-107. https://doi.org/10.7202/1024956ar

Ames, C. & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of educational psychology*, *80*(3), (p 260-267).

Butera, F. & Darnon, C. (Éd.). (2011). *L'évaluation, une menace ?* Presses Universitaires de France.

Coen, P. F. & Bélair, L. M. (2015). *Evaluation et autoévaluation, Quels espaces de formation ?* De Boeck Supérieur.

Fagnant, A. (2023). Les pratiques d'évaluation en classe : des compétences professionnelles pour soutenir l'apprentissage des élèves. Cnesco-Cnam.

Genelot, S. (2019). Observer les pratiques évaluatives des enseignants. Rapport de recherche pour le Cnesco. IREDU. Dijon

Laveault, D. (1999). Autoévaluation et régulation des apprentissages ». In C. Depover & B. Noël (1999) *L'évaluation des compétences et des processus cognitifs*, De Boeck.

Laveault, D. (2012). Autorégulation et évaluation-soutien d'apprentissage. In L. Mottier Lopez & G. Figari, *Modélisations de l'évaluation en éducation* (p. 115-130). De Boeck Supérieur. https://doi.org/10.3917/dbu.lopez.2012.01.0115

Mottier-Lopez, L. (2015). *L'évaluation formative et certificative des apprentissages des élèves*. De Boeck Education.

Sarrazin, P., Tessier, D. & Trouilloud, D. (2006), Climat motivationnel instauré par l'enseignant et implication des élèves en classe : l'état des recherches. *Revue française de pédagogie*, n°157. (p.147-177).

Toczek, M.-C. & Souchal, C. (2017). Le pouvoir des contextes évaluatifs. *e-JIREF*, *3*(1 & 2), 21-35.

Quelles sont les perceptions par les parents et les élèves des pratiques évaluatives dans les classes ?

STÉPHANE BÉNIT
UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE (FRANCE)

Introduction

Pour répondre à la question complexe qui nous est posée, à savoir : « quelles sont les perceptions par les parents et les élèves des pratiques évaluatives dans les classes ? », nous présenterons une partie de nos travaux de thèse, ainsi qu'une autre recherche conduite avec des élèves de grande section de maternelle. L'objectif de notre propos est de montrer comment les enfants perçoivent les différentes pratiques évaluatives, au travers de différentes instances socialisatrices, famille, école, qu'elles soient explicites ou informelles (Genelot, 2023).

Une partie de ce texte sera consacrée au travail scolaire réalisé au domicile familial notamment pour préparer certaines évaluations formelles. Nous tenterons de montrer que certains parents ont des attentes spécifiques à l'endroit de l'école et à ce titre suivent de très près le travail de leurs enfants. Cette transmission culturelle stratégique (Draelants, 2016) a des conséquences sur les expériences des élèves parfois très jeunes. Une autre partie de ce texte s'attachera à montrer comment en salle de classe les élèves reconfigurent le statut de l'erreur en classe à l'aune de leur statut scolaire, façonné en partie par le statut social parental et par le travail effectué à domicile.

I. Accéder à l'expérience subjective des élèves

Notre objectif a été de documenter l'expérience des élèves face aux pratiques évaluatives ainsi que les injonctions parentales à propos du travail scolaire, à l'aide d'une démarche inductive permettant de révéler le sens donné par les acteurs à leurs activités quotidiennes. En effet, les élèves vivent certaines situations évaluatives de classe comme autant d'événements marquants de leur scolarité, évènements qu'ils peuvent choisir de relater à leurs parents ou tenter de masquer. Il faut tout d'abord clarifier conceptuellement la notion d'expérience (Dubet, 2004), et plus particulièrement celle d'expérience scolaire (Dubet & Martucelli, 2006) :

> *« L'expérience est ce qui m'arrive et ce que j'en fais. L'expérience, c'est la capacité, pour les acteurs, à mobiliser des matériaux qui sont distribués socialement et ne leur sont pas propres, mais de faire quelque chose de singulier et de propre avec ce matériau, comme un musicien contraint par un code harmonique mais qui invente malgré tout. »*
> (Dubet, Zeitler, & Guerin, 2012)

Ainsi chaque élève, en fonction des situations sociales dans lesquelles il est placé, et au travers des différentes instances socialisatrices auxquelles il est confronté en permanence (école, pairs, parents),

construit une façon d'agir, de réagir face aux évènements. Les élèves construisent leur expérience en fonction d'interrelations avec leur environnement social, l'environnement social familial dans lequel ils évoluent ne constituant qu'une partie de la contrainte du code harmonique du musicien. L'expérience doit être comprise comme la résultante des relations de l'enfant avec son environnement social et psychologique au travers de quatre dimensions : la relation avec les pairs, la relation avec les parents, la relation à soi et la relation à (ou aux) enseignant(s). Chaque élève construit sa propre expérience scolaire à l'aune du rapport qu'il entretient avec chacune des dimensions.

Pendant les temps d'observations en classe nous avons assisté à des situations évaluatives interactives qu'elles soient différées ou simultanées mais également à des situations évaluatives formalisées plus explicites auxquelles ont principalement accès les parents (Genelot, 2023). Toutes ces situations évaluatives ont un effet sur les élèves même si la littérature scientifique s'est largement focalisée sur les travaux relatifs aux évaluations formalisées sommatives pour en démontrer de façon indiscutable « une impossible évaluation non biaisée des compétences des élèves » quel que soit l'outil utilisé : notation ou évaluation par compétences (Merle, 2018, p. 159.).

II. Préalable méthodologique

Dans les deux recherches évoquées en introduction, des entretiens semi-directifs à visée compréhensive ont été conduits auprès d'enseignants du 1er et 2nd degrés, de parents d'élèves et d'élèves. Des observations de classe ont également été menées. Dans le**Erreur ! Source du renvoi introuvable.** Tableau 1 sont regroupées les informations relatives à l'enquête menée auprès des parents et enfants. L'ensemble des entretiens se sont déroulés au domicile familial en présence des parents et de l'enfant. Les modalités évaluatives des élèves enquêtés du 1er et 2nd degrés sont précisées dans le Tableau 1, ainsi que l'académie dans laquelle se sont déroulés les entretiens. Les élèves ont été répartis en 3 niveaux, en difficulté, moyen, en réussite sur la base de leur livret scolaire (LSU), des travaux de classe, et après échanges avec les enseignants et la direction de l'établissement.

Tableau 1: Corpus relatif à l'enquête sur le travail parental (45 heures d'observation au total)

Élèves en difficulté	Élèves de niveau moyen	Élèves en réussite
-Inès CP/CM2[51] (père employé SNCF) Créteil, évaluée avec notes -Maëlys (père conducteur de bus, mère hôtesse de caisse) Reims, évaluée sans note -Benjamin 6e (mère Assistante maternelle, père Ouvrier) Reims, évalué sans note	-Yoni CP (père : employé SNCF, mère : secrétaire) Créteil, évalué sans note -Sophie (Sœur jumelle de Cécile) 6e (mère enseignante du 1er degré, père cadre supérieur) Créteil, évaluée avec notes -Mathieu 6e (père enseignant contractuel, mère chargée de communication) Créteil, évalué par notes et pourcentages	-Thomas 5e (mère professeur de physique-chimie, père cadre), évalué avec notes -Mathilde CM2/6e (mère bibliothécaire) Créteil, évaluée avec notes -Basil CM2/6e (mère cadre supérieure, père enseignant du 1er degré) Créteil, évalué avec notes -Cécile 6e (mère enseignante du 1er degré, père cadre supérieur) Créteil, évaluée avec notes -Grégoire 6e (père et mère cadre hospitalier) Reims, évalué sans note -Maelle 6e (mère cadre infirmier, père cadre supérieur) Reims, évaluée sans note -Romuald 5e (mère : employée mairie, père policier municipal, parents en reconversion professionnelle) Reims, évalué avec note -Irine 5e (mère : ingénieure au foyer, père ingénieur) Reims, évaluée avec note

Le Tableau 2 regroupe le corpus des élèves de maternelle évalués par compétences[52]. Les entretiens menés avec les enfants se sont déroulés dans l'enceinte de l'école dans des temps périscolaires.

Tableau 2: Corpus relatif à l'enquête sur les élèves de grande section de maternelle (25 heures d'observation)

Élèves en difficulté	Élèves de niveau moyen	Élèves en réussite
-Nathan (mère au foyer, père en prison) Reims -Laura (parents employés) Reims -Julie (mère aide à domicile, père manœuvre) Reims	-Mathéo (parents employés) Reims -Louis (parents employés) Reims -Sandra (père travaille à l'Insee, mère au foyer) Reims	-Gladys (mère cuisinière, père - chômage) Reims -Clément (mère universitaire, père chef d'entreprise) Reims -Marie (mère universitaire, père chef d'entreprise) Reims -Anna (père et mère archéologues) Reims

Note : Le dispositif évaluatif sommatif est constitué de marques de tampon indiquant la réussite, de smileys (3 modalités, content, moyennent content, triste).

[51] CP/CM2, signifie que les enfants ont été suivis de la classe de CP à celle de CM2.

[52] Ce terme utilisé par les enseignantes et repris par nous, ne correspond en réalité que très partiellement à la logique de compétences, la référence des enseignantes est attachée à l'outil évaluatif plutôt qu'à l'objet compétence (Genelot *et al.*, 2016). Il s'agit d'évaluation sans note. Le dispositif évaluatif sommatif est constitué de marques de tampon indiquant la réussite, de smileys (3 modalités, content, moyennent content, triste).

Enquêter auprès d'enfants et accéder à leur expérience n'est pas chose aisée. Le statut intermédiaire « ni parent, ni maître » (Danic, Delalande & Rayou, 2006, p. 106) permet de placer l'enfant en position d'informateur, puisque c'est lui qui explique le sens qu'il confère à ses actions. Il est de surcroît en mesure de fournir les clefs de compréhension de ses agissements (Kaufmann, 2014). L'objectif pour l'enquêteur est de devenir « adulte qui vient apprendre auprès d'enfants ». (Delalande, 2010, p. 70).

Au retour de l'école, à l'intérieur des familles aux profils sociaux variés, les enfants ne bénéficient pas tous d'aides parentales équivalentes, ni d'obligations scolaires similaires. Certains parents fortement diplômés et rompus aux us et coutumes scolaires procèdent à des activités de suivi des devoirs, d'aide aux apprentissages de leçons, de préparation aux évaluations. Ce travail parental invisible interroge le caractère osmotique de la transmission du capital culturel[53] (Bourdieu & Passeron, 1964 ; Daenants & Ballatore, 2014), La transmission par osmose du capital culturel correspond à un simple bain culturel permettant à l'enfant de s'en imprégner durablement, sans qu'aucune intention éducative parentale ne soit requise (Draelants, 2016). Garcia (2018), quant à elle, considère que la réussite scolaire des enfants passe fréquemment par un véritable travail pédagogique parental qui s'apparente à une professionnalisation et reste finalement peu interrogé. C'est également ce que nous observons au travers de nos résultats d'enquête. De nombreux travaux sociologiques ont exploré la fabrication des inégalités à l'intérieur des salles de classe (Lahire, 1993 ; Bautier & Rochex, 2007 ; Bonnéry, 2007, 2009), mais les conditions et modalités de travail dans l'espace familial ont été moins questionnées, hormis dans quelques travaux (Lahire, 1995 ; Daverne, 2003, 2009 ; Groleau, 2017). De façon récente, Lahire *et al*. (2019) ou encore Garcia (2018) ont tenté de décrypter la complexité des processus favorisant la transmission du capital culturel à l'intérieur de familles aux profils sociaux variés. Interroger et tenter de comprendre comment le capital culturel est assimilé, voire refaçonné par certains enfants (Draelants, 2016) est un enjeu de premier plan dans la compréhension des inégalités scolaires tant ce dernier est considéré comme un très bon prédicteur de la carrière scolaire (Vermandele, Dupriez, Maroy *et al.*, 2012).

Notre focalisation sur la dimension évaluative des discours internes à la famille repose sur trois éléments. Premièrement, comme nombre de systèmes éducatifs, l'école française fonctionne avec des situations évaluatives formalisées explicites très formalisées, principalement accessibles aux parents car elles ont vocation à être externalisées en dehors de la salle de classe, ainsi que des situations évaluatives interactives, lesquelles restent le plus souvent confinées à la salle de classe et largement méconnues des parents (Genelot, 2023). Toutefois la particularité française en matière d'évaluation sommative repose d'une part sur le couplage d'une échelle numérique (notation sur 10, notation sur 20) et d'une échelle de réussite imposée par le livret scolaire unique (LSU) (non acquis, en cours d'acquisition, acquis, à dépasser), et d'autre part sur l'existence d'un nombre important d'éléments de l'échelle évaluative signifiant un travail non réussi. Par exemple un 2/10, impliquera d'obtenir un 8/10 au prochain travail pour revenir à la « moyenne » dans la discipline ; un non acquis, un point rouge[54] ne pouvant pas être systématiquement effacé au profit d'un travail plus réussi. De telles pratiques affectent durablement la mobilisation des élèves et l'estime de soi (Périer, 2008). Deuxièmement, les

[53] Le capital culturel peut être défini comme les biens culturels transmis par les différentes actions pédagogiques familiales (Bourdieu & Passeron, 1970). Il reste à comprendre par quels processus ce capital se transmet, quels sont les acteurs familiaux les plus impliqués, avec quelle efficacité sur le plan scolaire. En d'autres termes, il est question d'interroger la façon dont les héritiers héritent de divers capitaux.

[54] Cette situation peut être comparée avec des pays européens au sein desquels le travail jugé comme insuffisant est affublé d'un voire deux signes évaluatifs négatifs au maximum (Finlande, Allemagne) (Cnesco, 2014), la France possède des échelles plus étendues pour signaler l'insuffisance d'un travail.

résultats scolaires permettent à certains parents de jauger de l'efficacité de leur activité d'aide au travail de l'enfant. Dans l'espace domestique, les évaluations formalisées sont très fortement suivies par les parents, qu'il s'agisse d'évaluations avec ou sans notes, ces dernières étant souvent appelées « évaluations par compétences » par les enseignants (Genelot *et al*., 2016). Nombreux sont en effet les discours familiaux se rapportant à l'école qui traitent principalement des résultats scolaires. Pour autant, les questions relatives aux résultats scolaires sont souvent absentes des travaux sur la construction familiale des dispositions scolaires (Lahire, 1995), tout comme le croisement des discours subjectifs des parents et des enfants sur ces thématiques. Troisièmement, l'implication scolaire parentale est susceptible d'être jugée en retour par l'institution. Garcia (2018) montre par exemple dans son enquête que les enseignants n'apprécient guère que les parents se mêlent trop des questions pédagogiques en qualifiant ces interventions de « zones d'empiètement ».

Nous présentons dans ce texte des résumés de nos travaux illustrés par quelques exemples de verbatims. Dans la recherche réalisée avec parents et enfants, 12 familles aux profils sociaux variés, les élèves sont évalués par des modalités sommatives diversifiées (notes ou sans notes). Parents et enfants ont spontanément évoqué lors des entretiens les termes de notes ou de compétences pour décrire les évaluations sommatives, ces termes proviennent des discours et sont repris par les enfants. Les enfants ont principalement retenu le terme de « remarques » pour décrire les interventions professorales liées aux situations évaluatives interactives différées ou simultanées (Genelot, 2023). Nous avons donc repris cette terminologie au cours de nos entretiens. Nous avons en outre questionné les modalités d'aides reçues à la maison dans l'apprentissage des leçons et dans les moments spécifiques de préparation aux évaluations, leurs natures et formes, leurs fréquences, afin de cerner au mieux les dynamiques relationnelles familiales et enfantines à propos du travail scolaire. Nous prendrons également appui sur la typologie des situations évaluatives de Genelot (2023) pour contextualiser certains propos d'enquêtés.

III. L'intériorisation, dès un jeune âge, de l'importance de la performance à l'école et des attentes de résultats des adultes

Les discours enfantins montrent que certaines pratiques parentales incitent à la mobilisation scolaire des enfants (Merle, 2005). Pour certains d'entre eux, une intériorisation de l'impératif de réussite scolaire[55] progressive est à l'œuvre au travers de certaines contractualisations (Merle, 2005). L'obligation scolaire rappelée par les parents surpasse souvent l'envie de vaquer à des temps de loisirs.

> Enquêteur : « Alors tu n'as pas envie, mais tu fais quand même tes devoirs ? »
> Irine (5e en réussite, évaluée avec des notes) : « Bah, il faut quand même que je les fasse parce que sinon voilà… la plupart ça va, ce n'est pas une corvée, mais c'est vrai la plupart du temps, j'aimerais faire autre chose avant, mais maman me rappelle à l'ordre. »

[55] L'impératif de réussite scolaire est défini par Merle (2005, p. 97.) comme l'incitation parentale « à travailler pour être digne de ses parents, pour conserver leur affection et susciter leur fierté ». L'intériorisation quant à elle est susceptible de s'articuler avec une contractualisation, c'est-à-dire un système d'octroi d'avantages ou de sanctions en cas de moins bons résultats scolaires.

Les enfants en réussite scolaire ont ainsi souvent intériorisé[56] progressivement l'obligation scolaire, grâce aux suivis parentaux et professoraux quotidiens, à tel point que le travail scolaire s'est mué en disposition au fil du temps[57] (Darmon, 2019).

> Maelle (6e, en réussite, évaluée sans note) : « J'sais pas comment expliquer, c'est naturel pour moi de faire mes devoirs. »
> Enquêteur : « Et si tu les faisais pas, qu'est-ce qui se passerait ? »
> Maelle : « Bah j'pense que déjà mes parents me crieraient beaucoup. Enfin y a des professeurs qui mettent des croix et dès que t'as oublié deux fois de faire tes devoirs, t'as un mot dans le carnet, du coup, les parents ils le savent. Et du coup, ils te crient dessus, donc je pense que je me ferais crier par les professeurs et par mes parents. »

IV. Des configurations favorisant la réussite scolaire

La réussite scolaire apparaît dans certaines familles comme la résultante d'une alchimie complexe. La valorisation, la vision utilitariste de l'école pour obtenir un bon métier ont été évoquées par toutes les familles, quelles que soient leurs professions et catégories socioprofessionnelles (PCS). Les plus grosses différences entre familles résident dans les contractualisations implicites (Merle, 2005), c'est-à-dire dans le suivi et l'aide à l'apprentissage des leçons et devoirs, ainsi que dans l'exigence de résultats-seuils, quelle que soit la modalité évaluative formalisée de la classe (note ou sans note). Certaines familles effectuent un suivi régulier, font travailler les enfants, exigent des résultats, et de surcroît réalisent ce travail pédagogique tous les jours. L'ensemble de ces modalités permettent aux enfants d'intérioriser progressivement la nécessité du travail scolaire.

V. Des modalités de travail extrêmement diversifiées

Comme de très nombreuses études le montrent (Brugeilles & Sebille, 2009 ; Garcia, 2018), les parents s'attelant au suivi scolaire de l'enfant au sein de notre enquête sont très fréquemment les mères Ces parents experts sont fréquemment porteurs d'une culture et d'une connaissance des exercices scolaires. Ils décryptent les attentes professorales, aident les enfants à apprendre les leçons, fournissent des méthodologies de résolution d'exercices et anticipent les évaluations. De tels parents sont non seulement des initiés des parcours scolaires, mais de surcroît de fins connaisseurs des exercices scolaires et des méthodes de résolution, pour en avoir mis à l'épreuve un certain nombre eux-mêmes durant leur scolarité.

De nombreux parents aident régulièrement les enfants dans le suivi des devoirs et apprentissages, y compris les parents des familles défavorisées (Kakpo, 2012). Certaines situations scolaires obligent les parents à moduler leurs exigences et stopper l'aide aux devoirs. C'est le cas notamment de la famille de Maelys qui a stoppé l'aide aux devoirs. Maelys doit faire face à du harcèlement à l'école en plus d'une sorte de phobie scolaire déclenchée par la perspective des évaluations. Face à la complexité de

[56] Darmon (2019, p.113) montre par exemple que l'incorporation d'une disposition est « non seulement comme le résultat de pratiques répétées, mais comme ce qui est recherché et ce qui fait l'objet d'un travail spécifique visant à rendre le maintien de l'engagement » plus facile, moins coûteux, en s'habituant à le faire, ce qui est le cas du travail scolaire.

[57] L'intériorisation d'une disposition, si elle n'est jamais observable directement, est potentiellement repérable à l'aune de comparaisons avec d'autres élèves, mais également en fonction du conflit et des luttes internes générées : « je n'ai pas envie de ne pas faire mes devoirs mais je dois m'astreindre malgré tout à les faire (Darmon, 2019) ». Dans le cas de notre enquête, l'ensemble des élèves en réussite ont tenu ce genre de propos, faisant des devoirs une routine, une obligation.

la situation, les parents choisissent de limiter l'aide aux devoirs et se consacrent exclusivement au bien-être de leur fille.

Dans ce cadre, Maelys travaille seule, se contentant d'un apprentissage de surface, le plus souvent par cœur, sans qu'aucune question, qu'aucune relation avec les savoirs de la leçon ne soit envisagée.

> Enquêteur : « Comment fais-tu pour travailler ? »
> Maelys (élève de 6e en difficulté évaluée sans note) : « Ben je lis la leçon, une fois, si j'la sais pas je refais et après je récite. »
> Enquêteur : « Toute seule ? »
> Maelys : « Bah oui. »

VI. Les stratégies socialement distribuées d'anticipation et de préparation des évaluations

De nombreux parents aident régulièrement les enfants dans le suivi des devoirs et apprentissages. Cette pratique devient, au fil du temps, un véritable travail d'institution en mesure de transformer une obligation scolaire en une disposition au sens de Bourdieu. Ce travail parental ne peut pas toujours être amorcé le soir même, en fonction des horaires parentaux tardifs, il a alors lieu le weekend.

> Enquêteur : « Que mettez-vous en place pour suivre le travail scolaire ? »
> Madame Pa (PCS favorisée, mère de Maelle, scolarisée en 6e et en réussite, évaluée sans note) : « C'est un suivi des devoirs. Être sûre que les devoirs soient faits. Donc c'est juste pointer chaque soir si les devoirs sont faits. Et puis le week-end, on a un peu plus de temps pour approfondir, réviser les évaluations à venir. »

Le suivi des devoirs dans d'autres familles plus défavorisées est plus complexe,[58]. La maman de Benjamin regrette ainsi que les enseignants ne vérifient pas plus le travail scolaire en 6e, alors qu'un tel suivi était effectif en primaire.

> Madame N (PCS défavorisée, mère de Benjamin scolarisé en 6e et en difficulté, évalué avec des notes) : « Les devoirs ? Avec Benjamin ça se passe dans le bus, vite fait quoi ! Moi j'peux pas le suivre, les parents viennent récupérer les petits vers 18h30, après je prépare à manger. »

La situation de Benjamin est emblématique de la difficulté pour certains parents d'assurer le suivi de leurs enfants. Benjamin a été suivi par sa maman en primaire mais cette dernière n'effectue plus qu'épisodiquement le suivi, car Benjamin évite au maximum cette aide qu'il estime peu rentable (Kakpo, 2012), d'autant plus qu'elle avait lieu à des moments qui correspondaient à des temps de loisirs pour Benjamin.

> Benjamin (classe de 6e et en difficulté scolaire) : « Ben… Ben des exercices des évaluations, des fois les leçons je les comprends pas trop. »
> Enquêteur : « Même si tu relis plusieurs fois ? »
> Benjamin : « Ben je comprends un petit peu mais des fois non. »
> Enquêteur : « Personne ne t'aide ? »
> Benjamin : « Ben… J'veux pas trop non plus. »

[58] Avec ce type de dispositif d'enquête, il n'est pas possible de dégager des variables et le poids respectif de ces variables dans le type de soutien parental. En revanche, on peut identifier des contextes globaux favorables ou défavorables à certaines pratiques : le temps, la faible proximité au monde scolaire, l'expérience scolaire antérieure malheureuse, la faible estime de ses capacités à suivre l'enfant… sont autant d'éléments qui entrent en ligne de compte.

VII. Les tentatives de décodage des attentes professorales

Prolongation ultime du suivi des devoirs, la préparation des évaluations s'apparente dans certaines familles à un véritable entraînement au sein duquel rien n'est laissé au hasard. Certains parents préparent même des évaluations blanches pour les enfants, quel que soit le système évaluatif formalisé de la classe (Genelot, 2023).

> Madame Pa (PCSP favorisée, mère Maëlle, scolarisée en 6e et en réussite, évaluée par compétences) : « c'est vrai quand il y a une évaluation, il y a un regard un peu plus appuyé. Pour les devoirs, il n'y a pas un enjeu tel qu'une évaluation. »

De tels entraînements sont orchestrés bien souvent plusieurs jours à l'avance et sont répétés jusqu'à que le résultat soit à la hauteur de l'attente évaluative parentale.

> Madame Ga (PCS favorisée, mère d'Irine, scolarisée en 5e et en réussite, évaluée avec des notes) : « comme je lui dis tout le temps moi je trouve que c'est important de faire travailler la mémoire, donc quand il y a une évaluation, même déjà le weekend, je la fais déjà travailler pour que ça s'enregistre quoi. »

Des questions ciblées sont posées sur les leçons et sont envisagées des questions susceptibles d'être posées lors du contrôle. Les enfants intègrent progressivement cette méthodologie jusqu'à ce qu'elle fasse partie de leur routine de travail.

> Enquêteur : « comment fais-tu pour obtenir de bons résultats ? »
> Irine (scolarisée en 5e et en réussite, 5e évaluée avec des notes) : « Ben, je ne sais pas, écrire, comprendre, voir, ce qu'on pourrait avoir à l'évaluation aussi, de temps en temps. Des exercices où on a à réfléchir, donc si elle (l'enseignante) nous les fait avant (en classe) ce n'est pas pour rien, c'est que ça peut nous préparer pour l'éval. »

Toutefois lorsque l'intériorisation peine à être totalement effective, certains parents ont recours à une contractualisation explicite reposant sur l'octroi ou la suppression d'avantages divers (Merle, 2005).

VIII. Les notes et compétences seuils : exigence de résultats

Il s'agit pour l'enfant d'obtenir à chaque évaluation scolaire une note, un signe évaluatif attestant de la validation de la maîtrise d'une compétence au-dessus du seuil fixé par les parents. Ces seuils sont quelquefois imposés de façon assez autoritaire.

> Échange avec Julie et sa maman aide à domicile, à propos du travail de Julie, élève de GS en difficulté, évaluée sans note.
> Maman de Julie : « On lui a expliqué à la maison qu'elle allait passer à la grande école, qu'il fallait qu'elle parle bien qu'elle articule, qu'elle sache compter. Donc elle a des récompenses quand elle fait bien son travail. La dernière fois que tu as bien travaillé, on est allé au Mac Do. »
> Julie : « Oui. »
> Maman de Julie : « Les exercices on les fait souvent quand même, peut-être pas tous les jours mais un jour oui, un jour non, pour pas non plus que ça devienne heu, qu'elle travaille tout le temps. »

Niels, élève de CP en réussite, sans note, est invité par sa maman cantinière qui l'élève seule à obtenir uniquement des points bleus, elle lui demande une évaluation-seuil.

> Enquêteur : « Elle dit quoi maman quand tu ramènes du bleu ? »
> Niels : « Bah, que c'est très bien. »
> Enquêteur : « Et quand tu ramènes du orange ou du rouge, elle dit quoi maman ? »

Niels : « Elle se fâche. »
Enquêteur : « Elle se fâche ? Qu'est-ce qu'elle fait ? »
Niels : « Elle dit " t'as eu un point rouge ? ", après elle dit " va dans ta chambre ". »
Enquêteur : « Elle te dit quoi maman sur le travail ? »
Niels : « Beh, que si tu fais du travail beh tu vas avoir des bonnes notes et que tu seras un bon élève. »
Enquêteur : « Ah, d'accord. Et alors, elle te donne du travail à la maison. T'aimes bien quand elle te donne du travail toi ? »
Niels : « Non. »
Enquêteur : « Ah, d'accord. Et, tu le fais ou tu le fais pas le travail qu'elle te donne ? »
Niels : « Je le fais. »
Enquêteur : « Ah, bon. Qu'est-ce qui se passe si tu le fais pas ? »
Niels : « Je serai privé de télévision et je serai privé de sorties. »

Madame Ha, cadre supérieure, maman de Basil, élève en réussite de CM2 évalué sans et avec notes.
Enquêteur : « Est-ce qu'il y a une exigence particulière quant aux résultats ? »
Madame Ha : « Je l'ai formulé de manière plus ou moins abrupte devant eux, c'est un peu une source de conflit entre Jérôme (son mari) et moi, mais… compte tenu de l'environnement dans lequel ils évoluent, les mauvais résultats ne seraient que le reflet d'un manque de travail. Et donc les mauvais résultats ne sont pas admis. Je crois que j'ai toujours dit en dessous de seize, dix-sept c'est mauvais. »

Les enfants tendent dès lors à se conformer à l'exigence parentale, puisque planent sur eux les hypothétiques sanctions, en cas de non-respect de la note-seuil. De telles évaluations-seuils sont également documentées par Garcia (2018) chez de nombreuses de classes sociales moyennes et supérieures, dont l'enjeu est d'occuper une place satisfaisante au sein de la hiérarchie scolaire de la classe.

IX. La réception différenciée des messages portés par les enseignants lors des situations évaluatives interactives différées et simultanées en maternelle

En classe, lors des situations évaluatives interactives informelles, auxquelles les parents n'ont En classe, lors des situations évaluatives interactives informelles, auxquelles les parents n'ont pratiquement pas accès, il existe une réception différenciée des interactions évaluatives en fonction du statut scolaire des élèves. Ainsi, de nombreuses observations de classe montrent que certaines situations évaluatives différées et simultanées (Genelot, 2023) sont réinterprétés par les élèves. En premier lieu, il existe un discours professoral extrêmement récurrent en classe lié au statut de l'erreur dans l'apprentissage : « c'est pas grave de rater ». Toutefois, certains élèves font une lecture différente de la notion d'erreur en fonction de leur statut scolaire.

Enquêteur : « Est-ce que c'est grave quand on se trompe et qu'on a des difficultés ? »
Mathéo (élève moyen de GS, parents de PCS moyenne) fait non de la tête. »

Les élèves en difficulté reprennent à leur compte l'affirmation professorale comme un bouclier dans l'espoir de préserver leur estime de soi, sinon de tenter de s'en convaincre : « C'est pas grave de faire des erreurs » répètent-ils presqu'à l'unisson.

Enseignante de maternelle B : « En individuel (situation évaluative simultanée individuelle, (Genelot, 2023)), je leur demande d'essayer quoi qu'il arrive, mais aussi de persévérer. Il faut

souvent plusieurs essais avant d'aboutir. S'ils se trompent, ce n'est pas grave, ils sont là pour apprendre. »

Malheureusement, les élèves en difficulté sont souvent convaincus du bien-fondé de la non-gravité de l'erreur énoncée par l'enseignant. En règle générale, les enseignants dans leur classe tentent de remobiliser les élèves après erreur en leur signifiant par un feedback la non-gravité de l'essai infructueux. Il arrive toutefois que ces discours mobilisateurs ne soient pas prononcés par certains enseignants, lesquels préfèrent signaler leur mécontentement face à la tâche ratée par un élève pour la énième fois. Nathan, élève en difficulté, affirme que si « le travail n'est pas réussi le maître ne sera pas content et il va nous punir » (situation évaluative différée individuelle), de même que Laura, élève en difficulté, qui a peur que « la maîtresse ne soit pas contente ». Elle mime alors la colère de la maîtresse en montrant le « sourire pas content » du signe évaluatif apposé sur son travail. Julie, élève en difficulté, quant à elle énonce avec tout le paradoxe que cela suppose que « la maîtresse crie que c'est pas grave ! ». Quelques élèves en difficulté sont extrêmement soucieux de l'échec et vont solliciter l'aval de la maîtresse avant de s'engager dans la moindre situation évaluative même non formalisée.

Enquêteur : « Selon vous quels sont les effets de l'évaluation sur les élèves ? »
Enseignante de maternelle A : « Beaucoup ont le même comportement en situation d'évaluation que lors d'une activité habituelle. Certains sont fiers d'avoir réussi et quelques-uns se rendent compte qu'ils ont du mal. Pour ceux-là, certains préfèrent ne rien faire plutôt que de risquer l'échec, d'autres font n'importe quoi pour avoir vite fini, d'autres encore gribouillent leur travail, ou me demandent confirmation avant d'oser démarrer… »

X. De « c'est pas trop grave » à « c'est grave » chez les élèves moyens et en réussite

Les élèves aux résultats moyens sont déjà plus réservés à propos de la question de l'erreur. Sandra, élève moyenne, parents de PCS moyenne évoque à ce titre le fait « que ce n'est pas trop grave ». Lorsque l'on pose la question aux meilleurs élèves, ces derniers rétorquent assez rapidement que l'on peut se tromper certes, mais jusqu'à un certain point seulement, afin de ne pas ternir l'image que l'enseignant se fait d'eux et de trop modifier leur statut scolaire.

Enquêteur : « Tu penses que la maîtresse elle pense quoi quand on fait des erreurs ? »
Clément : « Elle dit c'est pas grave. »
Enquêteur : « Elle dit que ça n'est pas grave ?. »
Clément : « Mais si elle gomme à chaque fois, elle va plus gommer. (La maîtresse gomme une fois ou deux maximum pour refaire un tracé non réussi) »
Enquêteur : « Ah, pourquoi elle ne va plus gommer si elle gomme à chaque fois ? »
Clément : « Parce que c'est c'est c'est heu c'est heu sinon elle s'énerve. »
Enquêteur : « Ah, elle s'énerve ? Ça veut dire que l'on n'a pas le droit de faire beaucoup d'erreurs ? »
Clément acquiesce.

De telles expériences relatives aux situation évaluatives interactives différées montrent que d'importantes différences se créent relativement précocement au cours de la scolarité. Certains élèves en réussite utilisent des processus d'autorégulation de leurs apprentissages. Ils ont intégré les attentes professorales et sont en capacité d'analyser leurs erreurs et corriger leurs productions (Laveault, 2012).

Enquêteur : « Est-ce que des fois tu te corriges toute seule ? »

> Marie, élève en réussite, grande section, parents de CSP favorisée : « Parfois je regarde si j'ai pas fait d'erreurs et si c'est bon je lève le doigt parce j'ai pas envie de lever mon doigt quand c'est pas bon. Donc j'efface si c'est pas bon. »

Conclusion

L'expérience subjective des élèves face à l'évaluation dépend de différentes instances socialisatrices, l'école et les parents jouant les rôles les plus importants au début de la scolarité. Les parents, n'ayant que peu accès aux situations évaluatives interactives de classe, se concentrent davantage sur les situations évaluatives formalisées explicites. Ces évaluations leur servent de référence pour guider, modifier le cas échéant l'activité pédagogique déployée auprès de leurs enfants (Garcia, 2018). Certains parents préparent leurs enfants à un décryptage des attendus évaluatifs professoraux, quelle que soit la modalité formalisée, avec ou sans notes. Ce travail prolonge la lente intériorisation de l'impératif de réussite scolaire (Merle, 2005), puisque les enfants comprennent rapidement que ces méthodes leur permettent fréquemment d'obtenir de bons résultats. Ainsi certains enfants développent des pratiques d'autocontrôle et d'analyses de leurs erreurs et sont capables d'anticiper les évaluations à venir. L'expérience qu'ils construisent dès le plus jeune âge se différencie alors sensiblement de celles des autres enfants. Certains parents n'hésitent pas à exiger uniquement, dans le cas d'une évaluation par compétences/sans notes, les points de couleur indiquant la réussite (Garcia, 2018). Le rapport à l'erreur des enfants s'en trouve sensiblement modifié, car il faut répondre aux attentes professorales et aux attentes parentales. Au final, le travail pédagogique parental quotidien soulève la question de la création de potentielles inégalités : « le fait qu'une partie des élèves réussissent bien scolairement grâce à un travail invisible réalisé au sein de la famille conduit à valider une norme scolaire inaccessible sans travail parental » (Garcia, 2018, p. 15.).

Certaines situations évaluatives interactives sont diversement vécues par les élèves. Les situations évaluatives interactives simultanées censées permettre la régulation des apprentissages des élèves au travers de leur fonction de régulation-guidage (Mottier-Lopez, 2015) restent selon Genelot (2023) relativement peu fréquentes en classe. De plus, nos observations de classe montrent que ces interactions, lorsqu'elles ont lieu, n'ont pas forcément l'effet escompté sur les élèves. La nature du feedback, essentielle pour constituer une aide à l'apprentissage, et plus encore, la façon dont l'enfant interprète ce feedback, semblent orienter son activité dans les tâches qui lui sont proposées à l'école.

> Mathieu (6e, élève moyen, évalué avec des notes et des pourcentages) : « Des fois les remarques[59] de la maîtresse elles m'énervent, c'est-à-dire que par exemple on a appris un truc qu'on vient juste d'apprendre, et je fais une faute là-dessus c'est normal on vient juste de l'apprendre on est en train de l'apprendre et elle dit à ça « ce n'est pas normal », à chaque fois souvent elle dit ça. »

En effet, de nombreux élèves ont l'impression d'être évalués en permanence en classe, les situations évaluatives interactives simultanées, donc non formelles, étant souvent perçues par les élèves comme les prémices d'une future évaluation formelle. Dans ce cadre, tout feedback à propos du travail en cours un tant soit peu négatif n'est pas de bon augure dans la perspective d'une évaluation plus formelle.

[59] « Remarque » est le terme utilisé par les enfants pour désigner les feedbacks de l'enseignant lors des situations évaluatives interactives simultanées ou différées.

Références

Bénit, S. & Sarremejane, P. (2019). Les pratiques évaluatives à la maternelle : des expériences enfantines différenciées en fonction du statut scolaire. *Carrefours de l'éducation*, 49, pp.291-298.

Brugeilles, C. & Sebille, P. (2009). La participation des pères aux soins et à l'éducation des enfants. L'influence des rapports sociaux de ses entre les parents et entre les générations. *Politiques sociales et familiales*, n°95, pp.19-32.

Danic, I., Delalande, J. & Rayou, P. (2006). *Enquêter auprès d'enfants et de jeunes : objets, méthodes et terrains de recherche en sciences sociales*. PUR.

Delalande, J. (2010). Saisir les représentations et les expériences des enfants à l'école. L'exemple du passage au collège. *Agora débats/jeunesses* 2, (n° 55), pp. 67-82.

Dubet, F. (1994). *Sociologie de l'expérience*. Seuil, coll. « la couleur des idées ».

Dubet, F. & Martuccelli, D. (1996). À l'école. *Sociologie de l'expérience scolaire*. Seuil.

Dubet, F., Zeitler, A. & Guerin, J. (2012). « La construction de l'expérience ». *Recherche et formation*, 70, pp.119-120.

Bautier, É. & Rochex, J.-Y. (2007) « Ces malentendus qui font les différences ». In : Deauvieau, J &. Terrail, J.-P (dir.), *Les sociologues, l'école et la transmission des savoirs*. La Dispute, pp. 227-241.

Bonnéry, S. (2007). *Comprendre l'échec scolaire. Élèves en difficultés et dispositifs pédagogiques*. La Dispute.

Bourdieu, P. & Passeron, J.-C. (1964). *Les héritiers. Les étudiants et la culture*. Éditions de Minuit.

Darmon, M. (2019). Analyser empiriquement un inobservable : comment « attrape-t-on » une disposition ? Dans : Séverine Depoilly éd., *La différenciation sociale des enfants : Enquêter sur et dans les familles* (pp. 107-137). Presses universitaires de Vincennes.

Daverne, C. (2003). Des héritiers… déshérités ? *VEI Enjeux*-132, 2003, pp.90-105.

Daverne, C. (2009). Des trajectoires scolaires et professionnelles atypiques. Pourquoi 'être bien né ne suffit pas ? *L'orientation scolaire et professionnelle*, n°38, pp.307-323

Draelants, H & Ballatore, M. (2014). Capital culturel et reproduction scolaire. Un bilan critique, *Revue française de pédagogie*, pp.115-142.

Draelants, H. (2016). Formes et évolutions de la transmission culturelle. Le « modèle des héritiers » à l'épreuve des données PISA 2009 », *Revue française de pédagogie*, 194, pp. 5-28.

Garcia, S. (2018). *Le goût de l'effort. La construction familiale des dispositions scolaires*. PUF.

Groleau, A. (2017). Entre reproduction et individualisation. Regard sur les parcours scolaires dans l'enseignement supérieur québécois d'individus issus de familles scolarisées, *SociologieS* en ligne http://sociologies.revues.org/6501

Genelot, S., Gardes, D., Mansanti, J. & Pinsard, N. (2016). Evaluer par compétences ou évaluer sans noter ? Arcueil : *Éducation permanente*, pp.89-100.

Genelot, S. (2023). Observer les pratiques évaluatives des enseignants. Cnesco-Cnam.

Lahire, B. (1993). *Culture écrite et inégalités scolaires. Sociologie de l'» échec scolaire » à l'école primaire*. Presses universitaires de Lyon.

Lahire, B. (1995). *Tableaux de familles. Heurs et malheurs en milieux populaires*. vEHESS/Gallimard/ Seuil.

Lahire, B. (dir.). (2019). *Enfances de classe. De l'inégalité parmi les enfants*. Éditions du Seuil.

Laveault, D. (2007). De la régulation au réglage : étude des dispositifs d'évaluation favorisant l'autorégulation des apprentissages, in Allal (L.), Mottier (L.), *Régulation des apprentissages en situation scolaire et en formation.* De Boeck, pp. 207-234.

Kaufmann, J.-C. (2014). *L'entretien compréhensif, l'enquête et ses méthodes*, Armand Colin.

Merle, P. (2005). Les explications extrascolaires de la mobilisation des élèves. Contribution à une sociologie de l'expérience subjective, *Les sciences de l'éducation pour l'ère nouvelle*, n°38, 2, pp.89-114.

Merle, P. (2018). *Les pratiques d'évaluation scolaire. Historique, difficultés, perspectives*. PUF.

Périer, P. (2008). La scolarité inachevée. Sortie de collège et expérience subjective du pré-apprentissage, *L'orientation scolaire et professionnelle*, 37/2, pp. 241-265.

Vermandele, C., Dupriez, V., Maroy, C. & Van Camppenhoudt, M. (2012). « Réussir à l'université : l'influence persistante du capital culturel de la famille ». *Les Cahiers de recherche du Girsef*, n° 87.

Quelle est la relation entre évaluation et motivation dans le cadre des apprentissages scolaires ?

FABIEN FENOUILLET

UNIVERSITÉ PARIS-OUEST NANTERRE LA DÉFENSE (FRANCE)

Introduction

L'évaluation dans le monde scolaire évoque en premier lieu cette fameuse note que se voit attribuer un élève par un enseignant. À partir du moment où il est possible de qualifier cette note de « bonne » ou de « mauvaise », elle entre de plein pied dans l'univers de la motivation. Peu importe que cette note soit numérique, sous forme de lettres ou de codes colorés. L'objectif des quelques lignes qui vont suivre sera donc de comprendre comment se forge cette valeur de bonne ou mauvaise note dans l'esprit de l'élève afin de mieux cerner quelques clefs de compréhension des apprentissages scolaires.

I. Buts, performance et apprentissage

L'apprentissage et la motivation sont des phénomènes psychologiques non observables. Afin de les rendre visibles, il est nécessaire de passer par des indicateurs. Un même indicateur peut renvoyer à ces deux phénomènes en même temps. Cette intrication entre motivation et apprentissage a été mise en évidence de façon particulièrement saillante (ces résultats n'ont pas fait l'objet d'une réplique aussi claire chez l'homme) dès le début du XXe siècle par Tolman et Honzik (1930) à partir d'une expérience sur des animaux (voir Encadré 1).

Encadré 1 : L'expérience de Tolman et Honzik (1930) révélant les liens entre motivation et apprentissage

Dans la recherche menée par Tolman et Honzik (1930), le dispositif consiste à observer le nombre d'erreurs que commettent trois groupes de rats en parcourant un labyrinthe. Moins un groupe de rats fait d'erreurs, meilleur est son apprentissage, le nombre d'erreurs est donc à première vue un indicateur de l'apprentissage. Tous les groupes de rats sont affamés avant de parcourir le labyrinthe. Le premier groupe reçoit une boulette de nourriture à la sortie du labyrinthe, alors que le deuxième n'en reçoit pas. Les résultats, jour après jour, montrent que le groupe récompensé fait de moins en moins d'erreurs alors que le groupe qui n'est pas récompensé stagne assez rapidement. En première analyse, il est possible de dire que cette diminution du nombre d'erreurs indique que le groupe récompensé progresse dans son apprentissage du labyrinthe là où le deuxième groupe accumule moins de connaissances. Cette recherche comprend un troisième groupe qui n'est récompensé qu'au bout du onzième jour. Ce troisième groupe, sur les 11 premiers jours, stagne de manière identique au groupe non récompensé. Cependant, à partir du douzième jour, ses résultats rejoignent subitement ceux du groupe récompensé, sans avoir eu besoin de plusieurs jours. Ce rattrapage en un seul jour

indique que le troisième groupe n'a pas eu besoin d'apprendre pour parvenir au même niveau de performance que le groupe récompensé dès le premier jour. Il connaissait tout aussi bien le labyrinthe que ce groupe, simplement il n'avait pas pour but, avant le onzième jour, de le parcourir le plus rapidement possible afin de récupérer une boulette de nourriture. L'indicateur « nombre d'erreurs » mesure donc ici tout autant l'apprentissage du rat que sa motivation à atteindre un objectif.

Transposée au monde scolaire, cette recherche permet d'envisager la note comme porteuse de cette même ambiguïté qui rend poreuse la frontière entre apprentissage et motivation. D'un côté, la note peut être considérée comme un indicateur de l'apprentissage. D'un autre, l'enseignant part du principe que l'élève va chercher à obtenir la meilleure note possible et donc intègre une forte motivation pour cet objectif de recherche de la performance. Pour Dweck et Leggett (1988), les élèves ne visent pas nécessairement à avoir la meilleure note possible, ils n'ont pas tous le même niveau d'aspiration. Autrement dit, en fonction de leur objectif, les élèves ne vont pas considérer de la même façon une même note. Il y a ceux qui vont amplement se satisfaire d'un 10 sur 20 alors que d'autres vont considérer qu'il s'agit d'une performance médiocre. Pour Dweck et Leggett (1988), les premiers, qui ont principalement pour objectif d'éviter un désastre, poursuivent un **but d'évitement de la performance (ou but de performance-évitement)** alors que les seconds, qui visent l'excellence, s'inscrivent dans la cadre d'un **but d'approche de la performance (ou but de performance-approche)**. La méta analyse de Murayama et Elliot (2012) indique que les élèves qui ont pour objectif d'éviter d'avoir une mauvaise performance vont effectivement avoir des scores plus faibles par rapport à ceux qui recherchent la performance.

Cependant, tous les élèves ne sont pas motivés par la note. Dweck et Leggett (1988) mettent en évidence que d'autres élèves vont davantage s'intéresser à leur progression dans les apprentissages plutôt qu'à la note. Ces élèves vont également rechercher plus volontiers les activités difficiles et des feedbacks plus précis sur leurs échecs afin de mieux comprendre comment ajuster leurs stratégies. Les recherches indiquent que ces élèves qui poursuivent ce que Dweck et Leggett (1988) ont appelé un **but d'apprentissage (par opposition au but de performance)** vont beaucoup mieux réagir face à l'échec. Pour le mettre en évidence, Elliot et Dweck (1988) ont orienté les enfants d'une dizaine d'années soit vers un but de performance (en les focalisant sur leur score), soit vers un but d'apprentissage (en les focalisant sur les gains d'apprentissage que peut apporter l'activité) et ils leur ont ensuite demandé de résoudre des activités insolubles, sans leur dire qu'elles étaient insolubles. Avant ces activités sans solution, les enfants ont participé à une première tâche donnant lieu à la communication d'un score qui permettait d'induire soit que ce dernier était particulièrement élevé, soit l'inverse. Ce feedback était manipulé par l'expérimentateur ce qui induit un niveau de compétence artificiel soit faible soit élevé. Les auteurs ont mis en évidence que les enfants qui ont un **but d'évitement de la performance** (performance-évitement) et pour lesquels un niveau de compétence faible a artificiellement été attribué vont avoir tendance à utiliser des stratégies peu efficaces, à attribuer leur échec à un manque de capacité et à choisir les activités les plus faciles. À l'inverse, les individus qui poursuivent un but d'apprentissage, quel que soit le niveau de compétence artificiel induit par l'expérimentateur, persistent à croire qu'ils vont trouver une solution, continuent face à l'échec d'utiliser des stratégies efficaces et jamais ne s'estiment incompétents. Les élèves qui ont un but de performance-approche avec induction de compétence élevé se comportent de la même manière que ceux qui ont un but d'apprentissage quelque soir l'induction de compétence pour ces derniers.

L'orientation vers un but d'apprentissage ou de performance est liée à différents facteurs qu'il serait trop long d'expliciter dans ces quelques lignes. Parmi ceux-ci, Dweck et Leggett (1988) estiment cependant que **la conception que l'élève a de la compétence** joue un rôle prépondérant. En effet, l'enfant développe au cours de sa scolarité une conception naïve de ses compétences et plus globalement de l'intelligence. Ces conceptions expliquent en partie les buts que se fixent les élèves dans les apprentissages scolaires. La **conception ou théorie dite « fixiste » de l'intelligence** conduit l'élève à se considérer lui-même (le soi) comme une entité constituée de traits stables, fixes, innés, non contrôlables, que l'on peut difficilement changer. L'autre grande conception est celle de **la théorie dite « incrémentale »**, qui conduit l'élève à comprendre l'intelligence et donc à se considérer lui-même (le soi) comme constitué d'un ensemble de qualités malléables, contrôlables, qui peuvent s'améliorer avec le temps. Les élèves qui développent une conception fixiste de l'intelligence ont tendance à adopter des buts de performance (évitement ou approche) là où ceux qui ont une conception incrémentale vont plutôt s'orienter vers les buts d'apprentissage comme le montre Leggett (1985). Dweck et Leggett (1988) citent la recherche de Leggett (1985) qui montrent que les élèves qui adoptent une conception incrémentale de l'intelligence vont davantage adopter un but d'apprentissage et ceux qui ont une conception fixiste vont plutôt être enclin aller vers un but de performance.

En fonction de ces conceptions, les élèves ne vont pas expliquer les réussites et les échecs de la même façon. Lorsqu'ils sont confrontés à l'échec, les élèves qui ont une conception fixiste de l'intelligence vont estimer qu'ils n'ont pas les capacités d'apprentissage nécessaires et auront davantage tendance à se décourager que les autres. Cette problématique dans la prise en compte de l'échec en fonction de la conception de l'intelligence se retrouve jusque dans l'activité visible du cerveau. Moser *et al*. (2011) ont mis en place une étude au cours de laquelle ils ont questionné des étudiants d'une vingtaine d'années sur leurs conceptions de l'intelligence. Cette recherche reposait sur une tâche de discrimination visuelle portant sur des symboles qui à chaque essai donnait lieu à un feedback sur la justesse de la réponse du participant. Cette tâche est une adaptation de la *flanker task* (Eriksen & Eriksen, 1974) lors de laquelle les participants devaient identifier correctement la lettre centrale (cible) d'une chaîne de cinq lettres dans laquelle la cible était soit congruente (par exemple, « MMMMM ») soit incongruente (par exemple, « NNMNN »). Durant cette tâche, l'activité cérébrale des participants était enregistrée au moyen de l'électroencéphalographie. Moser *et al*. (2011) ont mis en évidence que lors des feedbacks négatifs, l'activité cérébrale des participants avec une conception fixiste était moins élevée que celle des participants avec une conception incrémentale, et ce plus particulièrement au niveau du cortex cingulaire antérieur, qui est une des aires impliquées dans la motivation mais aussi dans la prise en compte de l'erreur. De plus, les auteurs ont montré que cette activité cérébrale a une action significative sur la réussite des essais suivants. Ces résultats et bien d'autres ont conduit à la mise en place de programmes éducatifs visant à orienter les conceptions de l'intelligence vers la conception incrémentale, programmes qui ont montré un impact significatif sur la réussite scolaire des élèves les plus en difficulté (Yeager *et al*., 2019). Ces programmes ont été créés dans l'idée de changer les attributions inhérentes à la conception fixiste liées à l'échec, en montrant par exemple la flexibilité du cerveau mais aussi que l'échec n'est jamais certain (pour une revue sur ces programmes et leurs effets voir Sperduto, Boujon, Fenouillet, 2022).

II. Contrôlabilité du résultat de l'évaluation

Un autre aspect important de l'évaluation dans le système scolaire est sa répétition. Les élèves sont évalués pratiquement toutes les semaines voire plus, ce qui induit **un effet d'attente** qui peut devenir plus que problématique quand l'élève s'attend à échouer systématiquement. En effet, les recherches ont mis en évidence que cet effet d'attente peut induire, en fonction des paramètres de la situation, un état qui s'apparente à la dépression.

Les premières recherches dans ce domaine ont été conduites sur l'animal. En effet, ce phénomène étant lié aux spécificités de l'environnement, il se retrouve aussi bien chez l'homme que chez l'animal (voir Encadré 2).

Encadré 2 : L'expérience d'Overnier et Seligman (1967) sur la résignation apprise (ou impuissance acquise)

Dans leur recherche, Overmier et Seligman (1967) constituent trois groupes de chiens dans une expérience qui se déroule en deux temps. Lors du premier temps, les chiens du groupe dit « échappement » sont soumis à un stimulus aversif électrique auquel ils peuvent se soustraire en appuyant avec le museau sur un panneau en bois disposé devant eux à cet effet. Les animaux du groupe dit « inévitable » sont soumis au même dispositif à la différence que l'appui avec le museau ne provoque pas l'arrêt du stimulus aversif qui s'arrête malgré tout au bout de quelques secondes. Enfin, le groupe dit « contrôle » ne participe pas à cette première phase mais uniquement à la deuxième. Lors de cette deuxième phase, les trois groupes font l'objet d'un second apprentissage aversif où les animaux doivent apprendre à changer de compartiment dans une cage avec une séparation avec l'apparition d'une lumière sous peine de recevoir un choc électrique. Lors de cette deuxième phase, les résultats font apparaitre que les groupes « échappement » et « contrôle » connaissent un apprentissage rapide et subissent en moyenne moins de 2 chocs sur les 10 possibles. Les résultats du groupe « inévitable », à l'inverse, montrent que les animaux ont le plus grand mal à acquérir ce nouvel apprentissage et n'arrivent pas à échapper à 8 chocs sur 10 en moyenne. Cette contreperformance n'est pas liée au choc électrique de la première phase puisque les performances du groupe « échappement » sont équivalentes à celle du groupe « contrôle ». Ce dysfonctionnement de l'apprentissage est lié à **l'état de résignation apprise** (syndrome d'impuissance acquise ou encore impuissance apprise en fonction des traductions) qui est induit lorsque l'animal apprend lors de la première phase que son comportement ne peut contrôler l'environnement.

Les résultats d'Overnier et Seligman (1967, voir Encadré 2) ont été répliqués chez l'homme avec des sons désagréables (Hiroto, 1974). Cependant certaines recherches ont donné des résultats différents de ceux d'Hiroto. En effet, lorsqu'il est confronté à un environnement incontrôlable, l'être humain ne se résigne pas toujours. **La résignation chez l'homme dépend du raisonnement qu'il tient face à cette incontrôlabilité**. Abramson, Seligman et Teasdale (1978) ont mis en évidence que la résignation se produit quand les individus attribuent leur incapacité à contrôler l'environnement à des facteurs stables, comme le manque d'intelligence (ce qui renvoie à la conception fixiste présentée précédemment).

La résignation apprise est un modèle possible pour expliquer l'échec scolaire. En effet, les élèves en échec scolaire, incapables de trouver des solutions pour éviter les mauvaises notes qui vont invariablement sanctionner leurs « difficultés d'apprentissage », ont appris à se résigner (Lieury &

Fenouillet, 2013), et ce point a particulièrement été bien illustré dans la recherche de Diener et Dweck (1980). Dans cette recherche, les auteurs commencent par questionner des enfants d'une dizaine d'années sur les causes qu'ils estiment à l'origine de leurs résultats scolaires (bonnes ou mauvaises notes), ou à l'origine de leur réussite (victoire ou défaite) à des jeux. Les élèves qui indiquent un facteur stable pour une mauvaise note (par exemple, « j'ai eu une mauvaise note en mathématiques car je ne suis pas assez intelligent ») et un facteur instable pour des bonnes notes (par exemple, « j'ai eu de la chance lors du dernier contrôle ») sont classés parmi ceux qui s'orientent vers la résignation. Ils sont comparés à un groupe qui affiche le profil attributionnel inverse, autrement dit des élèves qui, lorsqu'ils rencontrent l'échec, s'orientent malgré tout vers la maîtrise de l'activité (ils poursuivent un but de maîtrise). Les auteurs vérifient que l'intelligence (mesurée par un test ici, à ne pas confondre avec la conception de l'intelligence) des élèves est identique entre ces deux groupes. Les élèves sont ensuite confrontés à une activité insoluble. Les résultats mettent en évidence que les élèves au profil résigné sont plus enclins à citer leur manque d'intelligence à la fin de l'activité pour justifier leur échec. Les auteurs constatent également que la qualité des stratégies mises en place s'est détériorée au fil des activités pour ces mêmes élèves, comme s'ils devenaient effectivement un peu moins intelligents. À l'inverse, certains élèves du groupe orienté vers la maîtrise augmentent leur niveau stratégique et la plupart maintiennent le même niveau d'efficience stratégique tout au long de l'activité. Ces résultats indiquent que les raisonnements attributionnels que tiennent les élèves sur leurs réussites et leurs échecs peuvent les conduire à « baisser les bras » dans leurs apprentissages, comme le postule la théorie de la résignation apprise.

III. Autonomie de l'individu dans ses apprentissages

L'évaluation scolaire peut induire une autre forme de contrôle qui concerne, cette fois, directement le comportement de l'élève. En effet, l'évaluation peut être utilisée afin de « motiver » l'élève à apprendre son cours en prévision d'une évaluation. Cette utilisation de l'évaluation revient à contrôler (c'est-à-dire vise à orienter) les comportements. Cette forme de contrôle du comportement par l'évaluation a été particulièrement étudiée dans le cadre de **la théorie de l'autodétermination** (SDT pour *Self determination theory*, Ryan & Deci, 2017). Pour comprendre la problématique que peuvent induire certaines pratiques d'évaluation, il est nécessaire dans un premier temps de donner quelques éléments de compréhension sur cette théorie.

Les premiers travaux sur la SDT ont permis de mettre en évidence qu'il existe au moins deux grandes formes de motivation : **intrinsèque** et **extrinsèque**. Cette distinction part du constat que la récompense n'a pas toujours pour effet de favoriser la répétition du comportement. Deci (1971) est le premier à avoir mis en évidence que la récompense diminue l'apparition du comportement dans les conditions de libre choix. Les conditions de libre choix sont celles où l'individu peut faire ce qu'il souhaite sans qu'aucune forme de pression ne lui soit imposée. Les activités humaines qui sont réalisées par libre choix sont nombreuses : on peut penser par exemple aux activités sportives, à la lecture de romans ou encore aux jeux. Ces activités, généralement réalisées pour le plaisir de les pratiquer en elles-mêmes, sont motivées intrinsèquement, alors que celles qui sont pratiquées pour atteindre un objectif autre sont motivées extrinsèquement. Ainsi, la mise en place d'une récompense tendrait à diminuer l'intérêt d'un individu pour une activité, si celle-ci était motivée intrinsèquement.

Cette séparation entre motivations intrinsèques et extrinsèques ne recouvre pas une autre distinction, encore plus importante, entre **motivations autonomes** et **motivations contrôlantes**. En effet, si les

motivations intrinsèques sont par nature autonomes, les motivations extrinsèques peuvent aussi bien être autonomes que contrôlantes. Il existe des motivations extrinsèques autonomes pour les activités qui ne sont pas réalisées pour le plaisir de les pratiquer en elles-mêmes mais qui restent importantes pour les individus car elles sont connectées à leurs valeurs. Les personnes qui accordent de l'importance à l'écologie vont par exemple trier leurs déchets même si elles ne trouvent pas cette activité particulièrement intéressante. Cette motivation à trier ses déchets est un exemple de motivation extrinsèque autonome. Cependant, si une personne trie ses déchets dans la perspective de recevoir de l'argent ou d'en économiser, elle le fait pour une motivation extrinsèque contrôlante car l'argent n'est pas une valeur à laquelle l'individu peut s'identifier personnellement. La SDT introduit donc la notion de contrôle du comportement comme étant un élément central dans l'analyse de toute forme de motivation. L'un des premiers champs d'application de la SDT est le monde scolaire où elle a permis de mettre en évidence les limites de certaines pratiques pédagogiques, notamment celles qui concernent l'évaluation.

Tableau 1 : Les différentes formes de motivations pour la théorie de l'autodétermination

Motivation autonome		**Motivation contrôlante**
Motivation intrinsèque autonome = plaisir de réaliser l'activité pour elle-même	Motivation extrinsèque autonome = importance pour le soi	Motivation extrinsèque contrôlante = pression ou utilité instrumentale
J'adore comprendre comment fonctionne l'écologie	En tant qu'écologiste, je ne me vois pas ne pas trier mes déchets	Je pense qu'il est important de payer ses déchets au poids afin d'en favoriser le triage

Dans le cadre scolaire, l'évaluation fait partie des facteurs que Ryan et Deci (2017) rangent sous l'appellation « **effet de sape** ». Ce dernier permet de regrouper sous un même terme les facteurs qui ont en commun un effet délétère sur les motivations autonomes, qu'elles soient intrinsèques ou extrinsèques. Cet effet délétère est lié au **contrôle** qu'induit l'évaluation, la mise en place d'une récompense ou encore la compétition (pour ne citer que quelques exemples d'effets de sape, dont la liste dépasse ces quelques lignes de présentation). La notion de contrôle renvoie ici à l'idée que l'individu évalue en permanence s'il estime être ou non à l'origine de son comportement. Par exemple, si on me donne de l'argent pour lire un livre, je ne serai pas à l'origine de la lecture de ce livre, c'est l'argent qui en est à l'origine. Dans cet exemple, l'individu évalue (il ne s'agit pas nécessairement d'un processus conscient) que cette activité de lecture est contrôlée, il s'agit donc d'une motivation extrinsèque contrôlante. Ce contrôle du comportement va impacter négativement l'intérêt de l'individu pour l'activité. Dans le cadre scolaire, Grolnick et Ryan (1987) ont ainsi mis en évidence que l'intérêt pour un texte est significativement diminué si on annonce avant la lecture de ce texte que celui-ci sera évalué. Même quand les évaluations sont positives (comme une bonne note), elles ont malgré tout pour effet de diminuer les motivations autonomes (Ryan, 1982). En fin de compte, récompenser une activité intéressante (c'est-à-dire pour laquelle les individus ont une motivation intrinsèque qui est une forme de motivation autonome) a pour effet de rendre cette activité moins intéressante.

Cependant, il est important de garder à l'esprit que ce n'est pas l'évaluation en tant que telle qui a un effet de sape, mais la façon dont on y recourt. Cette distinction vaut pour tous les effets de sape et a bien été mise en évidence dans le cadre de la récompense. Dans une étude sur l'école primaire, Marinak et Gambrell (2008) ont proposé à deux groupes d'élèves des récompenses pour lire un livre tandis qu'un troisième groupe dit « contrôle » n'en recevaient aucune. Les récompenses étaient

proposées avant une activité de lecture en libre choix proposée à tous les groupes. La récompense pour le groupe « livre » était la possibilité de choisir un livre. Pour le groupe « symbolique », il s'agissait de choisir un objet parmi plusieurs objets tel qu'un porte clef ou un bracelet d'amitié. Les résultats mettent en évidence que seule la récompense du groupe « symbolique » a un effet négatif sur le temps de lecture en libre choix. Ce résultat, comme d'autres, indique que ce n'est pas la récompense en tant que telle qui a un effet sur les motivations autodéterminées mais bien le contrôle qu'elle induit sur le comportement de l'individu. Même les évaluations négatives (comme une mauvaise note), quand elles sont données dans un contexte non contrôlant, n'ont pas pour effet de diminuer les motivations autonomes (Carpentier & Mageau, 2013).

Conclusion

L'évaluation telle qu'elle est pratiquée actuellement dans le système scolaire pose de nombreux défis pour lesquels les différentes approches de la motivation peuvent apporter quelques éclairages. Les approches motivationnelles de l'évaluation scolaire présentées dans cette note suggèrent que certaines pratiques évaluatives peuvent potentiellement éloigner l'apprenant de ce qui semble être le véritable intérêt de l'évaluation, à savoir le feedback sur les apprentissages, pour les améliorer. En effet, l'intérêt du feedback pour les apprentissages est montré par de nombreuses études et méta-analyses (Bangert-Drowns *et al*., 1991 ; Hattie & Timperley, 2007) et l'évaluation a pour effet de détendre ce lien quand on insiste sur la performance (Dweck & Leggett, 1988), voire de le casser quand l'individu n'est pas en mesure de comprendre son évaluation (ce qui qui est évalué, le pourquoi de l'évaluation et tout ce qui se rapporte au résultat de l'évaluation, Abramson, Seligman & Teasdale, 1978). Une des grandes problématiques liées à cette utilisation de l'évaluation dans le système scolaire réside sans doute dans le fait qu'elle sert avant tout à contrôler le comportement de l'élève (c'est-à-dire qu'elle a une visée de contrôle sur le comportement de l'élève), ce qui est une orientation qui s'avère dommageable pour les apprentissages les plus autonomes comme ceux qui sont réalisés par intérêt pour la matière scolaire en elle-même (Ryan & Deci, 2017). Bien qu'elle reste largement incomplète, cette courte introduction sur la relation entre motivation et évaluation se veut une invitation à des lectures plus étendues dans le domaine notamment de la motivation, champ théorique actuellement encore en pleine expansion (voir Lieury & Fenouillet, 2013 ; Fenouillet, 2016 ; Fenouillet, 2017 ; Carré & Fenouillet, 2019).

Références

Abramson, L. Y., Seligman M. E. P. & Teasdale, J., (1978). Learned helplessness in humans: critique and reformulation. *Journal of abnormal psychology*, 87, 49-74.

Bangert-Drowns, R.L., Kulik, C.-C, Kulik, J.A. & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213-238.

Carpentier, J. & Mageau, G. A. (2013). When change-oriented feedback enhances motivation, well-being and performance: A look at autonomy-supportive feedback in sport. *Psychology of Sport and Exercise, 14*(3), 423–435.

Carré, P. & Fenouillet, F. (2009, 2019). *Traité de psychologie de la motivation*. Dunod.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation*. Journal of Personality and Social Psychology*, 18, 105-115.

Dweck, C. & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, n°2, 256-273.

Diener, C. I., & Dweck, C. S. (1980). An analysis of learned helplessness: II. The processing of success. *Journal of Personality and Social Psychology*, 39,940-952.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task*. Perception & Psychophysics*, 16, 143–149.

Fenouillet, F. (2016). *Les théories de la motivation*. Dunod.

Fenouillet, F. (2017). *La motivation*. Les Topos, Dunod.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*. 77. 81-112. 10.3102/003465430298487.

Hiroto, D. S. (1974). Locus of control and learned helplessness. *Journal of experimental psychology*, 102, 187-193.

Grolnick, W. S. & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology*, 52, 890-898.

Leggett, E. L. (1985). Children's entity and incremental theories of intelligence: Relationships to achievement behavior. Paper presented at the annual meeting of the Eastern Psychological Association, Boston.

Lieury, A. & Fenouillet, F. (2013). *Motivation et Réussite scolaire*. Dunod

Marinak, B. A. & Gambrell, L. B. (2008). Intrinsic motivation and rewards: What sustains young children's engagement with text? *Literacy Research and Instruction*, 47(1), 9–26.

Moser, J. S., Schroder, H. S., Heeter, C., Moran, T. P. & Lee, Y. H. (2011). Mind Your Errors: Evidence for a Neural Mechanism Linking Growth Mind Set to Adaptive Post-Error Adjustments. *Psychological Science*, 22, 1484-1489

Murayama, K. & Elliot, A. J. (2012). The competition–performance relation: A meta-analytic review and test of the opposing processes model of competition and performance. *Psychological Bulletin*, 138(6), 1035–1070.

Overmier, J. B. & Seligman, M. E. P. (1967). Effects of inescapable shock upon subsequent escape and avoidance learning. *Journal of Comparative and Physiological Psychology*, 63, 23-33.

Tolman, E. C. & Honzik, C. H. (1930). Introduction and removal of reward, and maze performance in rats. *University of California Publications in Psychology*, 4, 257–275.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3), 450–461.

Ryan, R. M. & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press.

Sperduto, C., Boujon, C. & Fenouillet, F. (2022). Bilan des interventions sur la conception de l'intelligence pour favoriser la réussite scolaire et proposition à mettre en œuvre auprès de collégiens, lycéens et adolescents en situation de décrochage scolaire. Psychologie Française, article en révision.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., Hahn, P. R., … Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573(7774), 364–369.

Quels sont les liens entre l'évaluation, la perception que les élèves ont d'eux-mêmes et leurs progrès ?

PASCAL PANSU
UNIVERSITÉ GRENOBLE ALPES (FRANCE)

Introduction

Un problème souvent évoqué dans les enquêtes à large échelle (Timss, Pirls et Pisa) est la présence de croyances négatives de compétence chez certains élèves, qui peuvent être un frein à l'acquisition des apprentissages fondamentaux indispensables à leur avenir scolaire et à leur insertion sociale future. Ces croyances étant particulièrement marquées chez les élèves français, comprendre leur lien avec les apprentissages est une étape essentielle pour les enseignants, qui à l'instar d'autres personnes signifiantes dans l'entourage de l'enfant (par exemple parents, fratrie, amis), peuvent involontairement participer à les entretenir.

I. Le soi est embarqué dans les apprentissages

Au-delà de prérequis, réussir une tâche scolaire nécessite non seulement des habiletés pour l'accomplir, mais aussi d'avoir une bonne perception de ces dernières pour les mobiliser correctement. Cette perception correspond au jugement qu'un élève porte sur sa compétence. Elle se construit et évolue au fur et à mesure de ses expériences de réussite ou d'échec, des cibles de comparaison auxquelles l'élève s'identifie, et des feedbacks des personnes importantes pour lui : parents, enseignants, pairs. Une telle perception est conceptualisée dans la littérature à partir de notions différentes, mais proches, comme le concept de soi (Marsh, 1990), le sentiment d'efficacité personnelle (Bandura, 1977, 1986), ou encore le sentiment de compétence (Harter, 1982, 1999)[60]. Bien que certaines nuances soient avancées par les auteurs (par exemple l'orientation temporelle vers le passé ou le futur, la stabilité dans le temps), si on considère ces notions à un même niveau de généralité ou de spécificité, elles sont proches sur un plan empirique (Fleury-Roy & Bouffard, 2006 ;

[60] Dans cette note, nous parlons d'auto-évaluation de compétence scolaire et non d'estime de soi qui renvoie à la valeur que les gens s'accordent en tant que personne (Harter, 1999). Cette valeur générale (*global self-worth*) pouvant varier en fonction de ce que les gens valorisent dans leur groupe d'appartenance. Ainsi, un élève en décrochage scolaire ou en marge du système scolaire peut avoir une bonne estime de lui-même s'il n'accorde aucune importance à l'école et valorise des activités qui, à nos yeux, n'ont aucune importance (voire paraissent déviantes), mais sont importantes pour les membres de son groupe (par exemple un gang). Pour cette raison, le lien souvent supposé entre estime de soi et réussite scolaire n'est pas fiable et il serait vraiment imprudent de conclure à une relation de cause à effet. Il relève plus d'une croyance en une position normative – en référence aux valeurs dominantes d'une société que nous partageons – que de faits démontrés. Par conséquent, à la question, « faut-il augmenter l'estime de soi des individus pour leur redonner confiance, indépendamment de leurs résultats [et *a fortiori* lorsqu'ils échouent] ? La réponse scientifique est clairement non » (Martinot, 2008, p. 68).

Pajares, 1996 ; Joët, Nurra, Bressoux & Pansu, 2007). Par exemple, Pietsch, Walker et Chapman (2003) ont montré que les énoncés mesurant la dimension cognitive du concept de soi en mathématiques (« j'ai de bonnes notes en mathématiques ») et ceux mesurant le sentiment d'efficacité personnelle (« je suis capable de réussir les exercices qu'on attend de moi en mathématiques ») renvoient à un même facteur. Ce facteur correspond à une auto-évaluation de sa compétence. Et s'il est une conclusion qui fait consensus aujourd'hui, c'est bien celle selon laquelle se percevoir comme une personne ayant peu d'habiletés pour réaliser une tâche peut nuire à l'exécution des actions nécessaires à sa réalisation et, *in fine*, affecter sa réussite, alors que se percevoir comme compétent pour réaliser une tâche peut soutenir la mise en œuvre des comportements nécessaires à l'exécution de celle-ci et, *in fine*, mener à sa réussite (Bandura, 1977, 1986 ; Harter, 1999 ; Marsh & Craven, 2006).

II. L'évaluation qu'un élève fait de sa compétence scolaire est loin d'être toujours réaliste

L'évaluation qu'un élève fait de sa compétence à l'école n'est pas toujours réaliste. Elle peut être biaisée positivement ou négativement et présenter un décalage, plus ou moins marqué, entre sa performance effective et l'évaluation qu'il en fait. L'élève peut alors se percevoir meilleur (on parlera d'illusion de compétence), ou inversement, moins bon qu'il ne l'est (illusion d'incompétence). C'est ce qu'il est convenu d'appeler le biais d'auto-évaluation de compétence scolaire (Bouffard & Narciss, 2011 ; Bouffard, Pansu & Boissicat, 2013 ; Gramzow *et al*., 2003 ; Pansu *et al*., 2021). Parce qu'il traduit un décalage entre la performance effective de l'élève et l'évaluation qu'il en fait, le biais d'auto-évaluation de compétence scolaire a d'abord été envisagé comme un schéma d'interprétation de la réalité marqué par des distorsions cognitives dans le traitement des informations relatives aux expériences d'apprentissage (Beck, 1967 ; Markus, 1977). Il traduirait donc une tendance générale de la personne à traiter de manière biaisée les informations sur soi. Il serait le reflet de pensées automatiques, négatives ou positives, qui seraient à l'origine d'une déformation des informations sur soi qui conduit à sélectionner préférentiellement des informations erronées. Lorsque ce mode de pensée est négatif, il peut conduire l'élève à interpréter négativement les informations ambiguës et à favoriser le rappel de souvenirs d'échec plutôt que de réussite. Par exemple, Vaillancourt et Bouffard (2009) ont montré que la présence d'un biais négatif d'auto-évaluation était associée à un traitement négatif des situations positives, à une exagération de l'importance des événements négatifs mineurs et à un sentiment de responsabilité accru envers ces événements. Les résultats indiquent aussi que se combinent à ce schéma des croyances relevant de forces incontrôlables. Par exemple, les élèves peuvent croire que leurs résultats sont dus à la chance ou au hasard et qu'ils ne peuvent rien faire pour en changer l'issue. Ils peuvent alors présenter des manifestations proches de ce que Seligman et collaborateurs ont appelé l'impuissance acquise, connue pour générer des déficits aux plans cognitif, affectif et motivationnel (*cf*. Miller, & Seligman, 1975). Au plan cognitif, ce déficit se traduit par un sentiment de non-contrôle (échec perçu comme inévitable) et interfère avec les processus de résolution de problèmes, de réflexion critique, et l'attention que l'élève peut porter à la tâche. Sur le plan émotionnel, il s'exprime dans un état dépressif exprimé par de l'apathie, des sentiments envers soi négatifs, de la passivité, de la colère et/ou de la rébellion. Au plan motivationnel, ce sentiment d'impuissance paralyse l'initiation des efforts et la persévérance à la tâche dès que survient un obstacle au profit de ruminations sur son état. Au fil des années, enfermés dans une telle spirale, les élèves consolident leur système de réponses inappropriées et développent des attitudes peu favorables envers l'école, des comportements les poussant à s'écarter des apprentissages et pouvant induire chez eux un sentiment d'aliénation scolaire propice à un décrochage prématuré.

S'il est une conclusion bien établie ce jour, c'est bien celle concernant le caractère inadapté du biais négatif d'auto-évaluation de compétence scolaire à un niveau général (Bouffard & Narciss, 2011 ; Bouffard *et al*., 2013). Depuis les premières études de Phillips (1984, 1987) conduites auprès d'élèves à haut potentiel, les résultats des recherches menées auprès d'élèves ordinaires vont également dans ce sens : un biais d'auto-évaluation de compétence scolaire négative (illusion d'incompétence) est préjudiciable aux apprentissages à l'école et à l'adaptation sociale (Bédard, Bouffard & Pansu, 2014 ; Bouffard, Boisvert & Vezeau, 2003 ; Bouffard, Vezeau, Chouinard & Marcotte, 2006 ; Leduc & Bouffard, 2017 ; Miserandino, 1996 ; Phillips & Zimmerman, 1990). Tant aux plans affectif, cognitif que comportemental, avoir une vision biaisée négativement de sa compétence scolaire est nuisible à l'élève, *a fortiori* lorsque le domaine d'évaluation est important pour lui. Ces élèves sont fréquemment moins motivés, plus passifs dans leurs apprentissages, et ont souvent des attentes de résultats et des résultats effectivement plus faibles (Bouffard *et al*., 2003 ; Bouffard *et al*., 2013). Ils se présentent aussi souvent comme moins enclins que leurs pairs à participer en classe, moins capables, moins curieux, voire moins disposés à faire des efforts et à relever des défis que les autres. Leur satisfaction et leur image de soi sont plus faibles que celles des autres élèves (Bouffard *et al*., 2003). Ils ont souvent du mal à accepter la responsabilité d'un succès, qu'ils attribuent préférentiellement à des facteurs externes, comme la chance, la facilité de la tâche, ou encore à l'indulgence du professeur. En revanche, ils n'ont aucune difficulté, même lorsque cela relève de toute évidence de facteurs externes, à s'attribuer la pleine responsabilité de leur échec, la rejetant directement sur leur inaptitude ou faibles capacités à l'école (un facteur interne et stable). En agissant de la sorte, ils bloquent le processus motivationnel découlant (*a minima*) des perceptions positives de soi, et peuvent aller jusqu'à se convaincre qu'ils ne sont pas faits pour de tels apprentissages, ce qui affecte leur engagement et leur rendement scolaire. En somme, ces travaux dressent un tableau sombre des corrélats de la présence d'une illusion d'incompétence et soutiennent clairement qu'une telle vision est délétère sur le plan des apprentissages et du bien-être psychologique.

Concernant l'évaluation biaisée positivement de sa compétence, les conclusions sont moins consensuelles. Pour certains auteurs, surestimer sa performance peut être nuisible à l'élève au fil du temps, car commettre l'erreur de se croire plus compétent qu'il ne l'est en réalité, peut le conduire à consacrer moins de temps et d'efforts que ne le requiert la tâche à exécuter et à obtenir, *in fine*, une piètre performance (Metcalfe, 1998 ; Baumeister *et al*., 1996). Les tenants de cette position, plutôt défaitiste, focalisent sur l'importance pour l'élève de toujours savoir précisément où il se situe dans ses apprentissages pour mettre en œuvre des actions nécessaires à la réalisation de la tâche et de progresser (Butler & Winne, 1995 ; Butler, 2011). Des attentes trop élevées associées à des échecs répétés pourraient être menaçantes pour le soi, et conduire l'élève à préférer des problèmes posant peu de défis, voire à utiliser des stratégies d'évitement dans le seul but de protéger un sentiment de compétence fragilisé (Harter, 1982 ; Robins & Beer, 2001). Cependant, des résultats récents présentent une autre vision de la surévaluation de sa compétence, assurément plus optimiste. Ils soutiennent qu'avoir une vision biaisée positivement de sa compétence est lié à une meilleure adaptation scolaire qu'avoir une vision réaliste ou biaisée négativement de sa compétence (Bonneville-Roussy, Bouffard & Vezeau, 2017 ; Bouffard & Narciss, 2011 ; Jamain, Bouffard & Pansu, 2020). Ces résultats appuient la thèse d'un schéma de soi fonctionnel sur le plan des apprentissages scolaires et sociaux (Bouffard *et al*., 2013). Dans cette perspective, une vision optimiste est bénéfique pour au moins deux raisons : d'une part, elle protège l'élève des émotions négatives, de l'anxiété et de symptômes dépressifs, et d'autre part, elle le pousse à avoir des ambitions, mais aussi à être motivé

et à s'autoréguler pour les approcher (Bandura, 1997 ; Bouffard & Narciss, 2011 ; Shin *et al*., 2007). Les résultats de plusieurs études vont dans ce sens : les élèves qui surestiment leur compétence ont souvent de meilleurs résultats scolaires que leurs camarades et ont une bonne progression au fil de l'année (Bouffard, Vezeau, Roy, & Lengelé, 2011 ; Dupeyrat *et al*., 2011). Assidus, ils participent activement en classe et sont motivés (Chamorro-Premuzic & Furnham, 2006 ; Fleury-Roy & Bouffard, 2006). Plutôt satisfaits d'eux-mêmes, ils sont aussi mieux jugés que d'autres par leur enseignant qui tend à les voir plutôt à l'aise dans l'exercice scolaire, peu enclins au doute et comme étant plutôt prêts à relever des défis (Lévesque-Guillemette, Bouffard & Vezeau, 2015 ; Pansu, Brun, Bouffard & Joët, 2013). En somme, ces résultats vont parfaitement dans le sens des conclusions que tirait Kurman (2006), à savoir, que présenter une auto-évaluation biaisée positivement de sa compétence peut être salutaire puisqu'elle accroît la motivation de l'élève et le conduit à s'autoréguler pour s'améliorer suite à un échec.

Quoi qu'il soit, tous ces travaux ont abordé l'évaluation biaisée de compétence scolaire à un niveau global supposant implicitement une tendance chez certains élèves à être biaisé dans l'évaluation de leurs apprentissages scolaires (une caractéristique de leur personne), indépendamment des contenus disciplinaires.

III. Deux facettes du biais d'auto-évaluation de compétence scolaire : générale et spécifique

Si les résultats des travaux évoqués jusque-là permettent de conclure sans ambiguïté que les corrélats d'un biais négatif de compétence générale à l'école sont préjudiciables, ils ont été obtenus à partir d'une mesure globale du biais d'auto-évaluation de compétence scolaire. Cependant, la question de la transposition de ces conclusions aux travaux réalisés à partir d'une mesure spécifique ne va pas de soi – la différence entre auto-évaluation biaisée générale ou spécifique étant loin d'être anodine. En effet, cette distinction renvoie à deux facettes du soi emboîtées hiérarchiquement : le biais général d'auto-évaluation de « sa » compétence scolaire, en englobant le rapport à l'école dans son intégralité, est d'un ordre supérieur, alors que le biais spécifique qui se focalise sur un apprentissage particulier (français, mathématiques, géographie, etc.) est d'un ordre inférieur. Prendre en compte des facettes spécifiques permet d'éviter des généralisations systématiques qui conduiraient à considérer qu'un élève biaisé négativement (ou positivement) dans une matière est nécessairement biaisé négativement (ou positivement) dans d'autres matières. Dans le cadre de cette présentation, nous avancerons quelques résultats récents obtenus dans le programme de recherche *SchoolBias* soutenu par l'ANR 2006 (Pansu, de Place, Bouffard *et al*., 2021). Ce programme s'est focalisé sur l'examen du développement des biais en français et en mathématiques et leurs conséquences pour les élèves à deux niveaux d'analyse interreliés : intra-individuel (celui de l'élève) et inter-individuel (jugement de l'enseignant sur l'élève). 22 études ont été réalisées dans le cadre de ce programme et ont permis des avancées importantes dans la compréhension de ces deux volets.

Concernant le premier volet, centré sur le niveau intra-individuel, trois grands résultats ressortent. Premièrement, il semble acquis que l'évaluation biaisée de compétence scolaire peut exister tant à un niveau global que spécifique : les élèves peuvent présenter un biais d'auto-évaluation dans une matière, sans fatalement le présenter dans une autre. Par exemple, seuls 29 % des élèves présentant un biais négatif stable dans le temps, en lecture ou en mathématiques, le présentent aussi dans l'autre matière, alors que lorsque le biais est positif, la concordance entre les deux matières est de 50 %

(Jamain, 2019 ; Pansu *et al*, 2021). Ces résultats renforcent l'idée que les élèves ne développent pas automatiquement une même orientation biaisée de leur compétence dans toutes les disciplines scolaires et que le biais général n'est pas réductible à la somme des biais spécifiques dans les matières principales. Ce résultat est important, car il souligne que ce n'est peut-être pas tout le fonctionnement scolaire de l'élève qui est perturbé, mais celui propre à un domaine particulier. Autrement dit, l'évaluation biaisée de sa compétence ne traduit pas dans tous les cas une distorsion cognitive permanente relevant d'une caractéristique propre à la personnalité (un trait), mais peut être dépendante du contexte (par exemple, la discipline scolaire). Deuxièmement, les résultats soulignent, conformément à ceux obtenus avec une mesure de biais général, que cette distorsion n'est délétère que lorsque le biais d'auto-évaluation de compétence scolaire est négatif ; lorsque l'évaluation est biaisée positivement, elle est bénéfique à l'élève. Une analyse de la structure des schémas de pensée des élèves étaye de manière plus fine ces résultats. Cette structure est différente selon l'orientation du biais : pour les élèves qui se sous-estiment (illusion d'incompétence), elle semble s'organiser surtout autour du pôle évaluation qui apparaît central dans leurs préoccupations, alors qu'elle semble s'organiser autour de deux pôles distincts, l'un lié à l'évaluation et l'autre à l'action, chez les élèves qui se surestiment (illusion de compétence). Cela pourrait être rapproché de préoccupations centrées sur la performance (plutôt orientée évitement) pour les premiers, et de préoccupations centrées à la fois sur la performance (plutôt orientée approche) et la maîtrise pour les seconds. Troisièmement, ils mettent en avant que la comparaison avec les pairs de la classe fournit des informations importantes aux élèves pour s'auto-évaluer, mais aussi que le mécanisme par lequel opère cette comparaison est complexe. Ainsi, lorsque l'élève choisit de se comparer à des camarades qui réussissent effectivement mieux que lui (comparaison absolue), son biais d'auto-évaluation est d'autant plus positif qu'il s'identifie fortement à la cible de comparaison, alors que lorsque la comparaison est relative (perception des écarts de compétence entre les élèves), plus il perçoit sa cible de comparaison comme meilleure que lui, plus son biais d'auto-évaluation est négatif. Cela pointe une fois de plus le rôle des perceptions subjectives et de leur importance pour les élèves et leur construction du soi.

Concernant le second volet, centré sur le niveau inter-individuel, les résultats confirment l'importance de la perception de soi scolaire des élèves dans le jugement des enseignants : plus le biais d'auto-évaluation des élèves est positif, meilleur est le jugement de l'enseignant. Ils suggèrent que les enseignants accordent une plus-value aux élèves qui se surestiment, ces derniers apparaissant plus en conformité avec les attentes du système scolaire que ceux qui se sous-évaluent. Ils apparaissent plutôt à l'aise avec les valeurs des systèmes scolaires méritocratiques et sont perçus par les enseignants comme plus confiants et capables de bien s'autoréguler en classe. Ce résultat n'est pas propre au système éducatif français et se retrouve aussi dans d'autres systèmes culturels. Il est sans doute à rapprocher de ce qui se passe au niveau intra-individuel, ce jugement pouvant enfermer les élèves dans les attentes de leur enseignant (auto-réalisation des prophéties). Enfin un autre point important concerne le manque de discernement des enseignants quant aux élèves biaisés qu'ils ne sont généralement pas toujours capables d'identifier correctement. Un tel résultat peut apparaître surprenant, voire en décalage avec ceux présentés dans le point précédent, à savoir, une plus-value accordée aux élèves qui se surévaluent, d'une part, et une croyance quant à l'effet bénéfique de la surestimation pour leur bon fonctionnement scolaire, d'autre part. Bien qu'une majorité d'enseignants considère qu'il est préférable qu'un élève surévalue ses capacités plutôt que l'inverse, il semblerait que ce que peut dire un enseignant du biais est loin d'être toujours en rapport avec ce qu'il prend en compte pour émettre son jugement. Comme l'ont suggéré Fleury-Roy et Bouffard (2006), tout semble

se passer à un niveau implicite, l'enseignant pouvant repérer des comportements qui lui fournissent des indications sur ce qu'il pourrait attendre ou non de ces profils d'élèves. Cependant, il se peut aussi qu'il soit difficile pour les enseignants, au regard de leur position et des valeurs d'impartialité dont ils doivent faire preuve à l'école, de reconnaître ouvertement qu'ils favorisent un profil d'élève à un autre. Quoi qu'il en soit, ce constat souligne les efforts qui doivent être entrepris sur le plan de la communication et de la formation pour aider le personnel enseignant et éducatif à mieux saisir la complexité de ce processus et de son implication pour les apprentissages des élèves.

IV. Réflexions pour l'action

Sur un plan pragmatique, ces résultats invitent à réfléchir sur les « gestes professionnels » utiles à mettre en place pour prévenir une auto-évaluation biaisée négativement de « sa » compétence, mais aussi sur la manière de restaurer la confiance des élèves touchés par ce phénomène (Pansu *et al*., 2021). Parmi ces gestes, l'enseignant peut manifester des attentes claires sur l'organisation de la tâche et sur ce que les élèves doivent faire (Hattie, 2008 ; Jamain, Bouffard & Pansu, 2020). L'enseignant favorise ainsi la correspondance entre la définition qu'il donne d'une tâche et celle qu'en a l'élève, ce qui a pour effet de rassurer l'élève et de lui permettre d'avoir une bonne compréhension de ce qu'il faut faire. C'est, non seulement essentiel pour accroitre sa confiance, mais aussi un prérequis à la planification des actions à réaliser pour atteindre le but escompté, ce qui, *in fine*, facilite la mise en action de l'élève dans la tâche. Ces résultats appellent également à être vigilant quant à la manière dont les élèves expliquent leur réussite ou leur échec (Pansu, & Sarrazin, 2010) : font-ils des attributions internes (« je suis responsable de mes échecs ») ou externes (responsabilité externe : manque de chance, difficulté de l'exercice, imputation au professeur) ? Présentent-ils une vision fixiste (« je suis nul, je n'y arriverai jamais ») ou malléable de leurs capacités ? Finalement, les résultats des études présentés invitent à réfléchir à différents leviers pour favoriser l'apprentissage des élèves : l'outillage dont les enseignants ont besoin (1) pour être en capacité de repérer les élèves présentant des attributions dysfonctionnelles pour leurs apprentissages ; (2) pour centrer leurs feedbacks sur la tâche et non sur les capacités intrinsèques de l'élève (ce qui reviendrait à renforcer chez lui une vision stable et immuable de l'intelligence) ; (3) pour éviter un climat de trop grande compétitivité et référer l'élève à ses progrès plutôt qu'aux performances des autres (en faisant en sorte de ne pas lui imposer, même involontairement, de comparaison sociale forcée) ; (4) pour procéder par petites étapes, donner des buts proches en ayant des attentes explicites et claires, de manière à permettre aux élèves de mettre en place des stratégies d'autorégulation adaptées pour atteindre leurs objectifs immédiats et futurs.

Références

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs (États-Unis): Prentice-Hall.

Baumeister, R. F., Smart, L. & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: the dark side of high self-esteem. *Psychological Review*, *103*(1), 5-33. https://doi.org/10.1037/0033-295x.103.1.5

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects.* New York: Harper & Row.

Bédard, K., Bouffard, T. & Pansu, P. (2014). The risks for adolescents of negatively biased self-evaluations of social competence: The mediating role of social support. *Journal of Adolescence*, *37*(6), 787-798. https://doi.org/10.1016/j.adolescence.2014.05.004

Bonneville-Roussy, A., Bouffard, T. & Vezeau, C. (2017). Trajectories of self-evaluation bias in primary and secondary school: Parental antecedents and academic consequences. *Journal of School Psychology*, *63*, 1-12. https://doi.org/10.1016/j.jsp.2017.02.002

Bouffard, T., Boisvert, M. & Vezeau, C. (2003). The illusion of incompetence and its correlates among elementary school children and their parents. *Learning and individual differences*, *14*(1), 31-46. https://doi.org/10.1016/j.lindif.2003.07.001

Bouffard, T. & Narciss, S. (2011). Benefits and risks of positive biases in self-evaluation of academic competence: Introduction. *International Journal of Educational Research*, *4*(50), 205-208. https://doi.org/10.1016/j.ijer.2011.08.001

Bouffard, T., Pansu, P. & Boissicat, N. (2013). Quand se juger meilleur ou moins bon qu'il ne l'est s' avère profitable ou nuisible à l'élève. *Revue française de pédagogie. Recherches en éducation*, (182), 117-140. https://doi.org/10.4000/rfp.4020

Bouffard, T., Vezeau, C., Chouinard, R. & Marcotte, G. (2006). L'illusion d'incompétence et les facteurs associés chez l'élève du primaire. *Revue française de pédagogie. Recherches en éducation*, (155), 9-20. https://doi.org/10.4000/rfp.61

Bouffard, T., Vezeau, C., Roy, M. & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International journal of educational research*, *50*(4), 221-229. https://doi.org/10.1016/j.ijer.2011.08.003

Butler, R. (2011). Are positive illusions about academic competence always adaptive, under all circumstances: New results and future directions. *International Journal of Educational Research*, *50*(4), 251-256. https://doi.org/10.1016/j.ijer.2011.08.006

Butler, D. L. & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245-281. https://doi.org/10.3102/00346543065003245

Chamorro-Premuzic, T. & Furnham, A. (2006). Self-assessed intelligence and academic performance. *Educational Psychology, 26*, 769-779. https://doi.org/10.1080/01443410500390921

Dupeyrat, C., Escribe, C., Huet, N. & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research*, *50*(4), 241-250. https://doi.org/10.1016/j.ijer.2011.08.005

Fleury-Roy, M. H. & Bouffard, T. (2006). Teachers' recognition of children with an illusion of incompetence. *European Journal of Psychology of Education*, *21*(2), 149-161. https://doi.org/10.1007/bf03173574

Gramzow, R. H., Elliot, A. J., Asher, E. & McGregor, H. A. (2003). Self-evaluation bias and academic performance: Some ways and some reasons why. *Journal of Research in Personality*, *37*(2), 41-61. https://doi.org/10.1016/s0092-6566(02)00535-4

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53*(1), 87–97. https://doi.org/10.2307/1129640

Harter, S. (1999). *Distinguished contributions in psychology. The construction of the self: A developmental perspective*. New York, NY, US: Guilford Press.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Jamain, L. (2019). *Biais d'auto-évaluation de compétence en français et en mathématiques chez les élèves de primaire: évolution et implications pour l'adaptation et la réussite scolaire des élèves?* Thèse de doctorat. Université Grenoble Alpes.

Jamain, L., Bouffard, T. & Pansu, P. (2020). Le lien entre le biais d'auto-évaluation de compétence et la performance scolaire: rôle de l'autorégulation et de la compréhension des attentes de l'enseignant par l'élève. *Canadian Journal of Education/Revue canadienne de l'éducation*, *43*(1), 197-228.

Joët, G., Nurra, C., Bressoux, P. & Pansu, P. (2007). Le jugement scolaire: un déterminant des croyances sur soi des élèves. *Psychologie & Education (Dourdan)*, (3), 23-40.

Kurman, J. (2006). Self-enhancement, self-regulation and self-improvement following failures. British *Journal of Social Psychology, 45*(2), 339–356. https://doi.org/10.1348/014466605X42912

Leduc, C. & Bouffard, T. (2017). The impact of biased self-evaluations of school and social competence on academic and social functioning. *Learning and Individual Differences*, *55*, 193-201. https://doi.org/10.1016/j.lindif.2017.04.006

Lévesque-Guillemette, R., Bouffard, T. & Vezeau, C. (2015). Les liens entre le jugement de l'enseignant sur les biais d'autoévaluation de compétence de l'élève et la qualité de leur relation. *Revue des sciences de l'éducation*, *41*(2), 179-197. https://doi.org/10.7202/1034032ar

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*(2), 63-78. https://doi.org/10.1037/0022-3514.35.2.63

Marsh, H. W. (1990). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2*, 77-172. https://doi.org/10.1007/BF01322177

Marsh, H. W. & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Martinot, D. (2008). Le soi, les autres et la société. Grenoble : PuG.

Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality and Social Psychology Review, 2*, 100-110. https://doi.org/10.1207/s15327957pspr0202_3

Miller, W. R. & Seligman, M. E. (1975). Depression and learned helplessness in man. *Journal of Abnormal Psychology*, *84*(3), 228-238. https://doi.org/10.1037/h0076720

Miserandino, M. (1996). Children who do well in school: Individual differences in perceived competence and autonomy in above-average children. *Journal of Educational Psychology*, *88*(2), 203-214. https://doi.org/10.1037/0022-0663.88.2.203

OECD Publishing (2013). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know.* Organisation for Economic Co-operation and Development.

OECD Publishing (2020). *PISA 2018 Results (Volume I): What Students Know and Can Do*. Organisation for Economic Co-operation and Development.

OECD Publishing (2020). *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. Organisation for Economic Co-operation and Development.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of educational research*, *66*(4), 543-578. https://doi.org/10.3102/00346543066004543

Pansu, P., de Place, A. L., Bouffard, T., Blaise, F., Boissicat, N., ... & Vezeau, C (2021). *Le biais d'auto-évaluation de compétence scolaire* (SchoolBias). ANR-16-CE28-0011.

Pansu, P., Brun, L., Bouffard, T. & Joët, G. (2013). Les implications du biais d'évaluation de compétence scolaire des élèves dans le jugement des enseignants. *Communication au 35e congrès de la Société québécoise pour la recherche en psychologie, organisé à Saguenay (Canada).*

Pansu, P. & Sarrazin, M. (2010). Les théories de l'attribution appliquées à l'éducation : L'efficacité des programmes de reconversion attributionnelle à l'université. *La Psychologie Sociale: applicabilité et applications*, 237-266.

Phillips, D. (1984). The illusion of incompetence among academically competent children. Child Development, 55(6), 2000–2016. https://doi.org/10.2307/1129775

Phillips, D. A. (1987). Socialization of perceived academic competence among highly competent children. Child Development, 58(5), 1308–1320. https://doi.org/10.2307/1130623

Phillips, D. A. & Zimmerman, M. (1990). The developmental course of perceived competence and incompetence among competent children. In R. J. Sternberg & J.Kolligian (Ed.), *Competence considered* (pp. 41-66). New York: Yale University Press.

Pietsch, J., Walker, R. & Chapman, E. (2003). The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of Educational Psychology*, *95*(3), 589-603. https://doi.org/10.1037/0022-0663.95.3.589

Robins, R. W. & Beer, J. S. (2001). Positive illusions about the self: short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, *80*(2), 340-352. https://doi.org/10.1037/0022-3514.80.2.340

Shin, H., Bjorklund, D. F. & Beck, E. F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development*, *22*(2), 197-212. https://doi.org/10.1016/j.cogdev.2006.10.001

Vaillancourt, M. È. & Bouffard, T. (2009). Illusion d'incompétence, attitudes dysfonctionnelles et distorsions cognitives chez des élèves du primaire. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *41*(3), 151-160. https://doi.org/10.1037/a0014890

En classe, quels effets peuvent avoir les interactions à dimension évaluative sur les élèves en termes d'inégalités ?

SÉBASTIEN GOUDEAU
UNIVERSITÉ DE POITIERS (FRANCE)

Introduction

Il est aujourd'hui solidement établi que les situations dans lesquelles se trouvent les individus ont le pouvoir d'influencer leurs pensées, émotions et comportements (Ross & Nisbett, 1991). De nombreuses recherches mettent ainsi en évidence ce *pouvoir de la situation* sur les performances scolaires. Dans cette note, nous présentons des travaux en psychologie sociale (discipline qui étudie ce pouvoir de la situation) ayant montré comment les situations scolaires peuvent perturber les performances des élèves de classe populaire, et ainsi amplifier les inégalités scolaires.

I. La menace du stéréotype

Un premier élément qui caractérise les interactions scolaires à dimension évaluative concerne l'existence de stéréotypes sociaux, qui sont des croyances sociales qui associent des caractéristiques positives ou négatives aux membres d'un groupe social donné (Leyens *et al.*, 1996). Par exemple, les enfants de classes populaires sont souvent jugés comme moins « intelligents » et moins travailleurs que les enfants de classes sociales favorisées. Les filles quant à elles sont souvent perçues comme moins compétentes en mathématiques que les garçons. Ces stéréotypes, même s'ils ne renvoient à aucune réalité scientifique, ont le pouvoir d'affecter les performances des élèves qui en sont la cible, et ainsi de jouer un rôle dans le phénomène de construction des inégalités scolaires.

Des recherches montrent que lorsqu'ils sont placés en situation d'évaluation, les élèves de classes populaires peuvent craindre de confirmer le stéréotype négatif relatif à leur groupe social d'appartenance, ou d'être jugés conformément à ce stéréotype (Steele, 1997). Cette « menace du stéréotype » engendre du stress, mais également des émotions et des pensées négatives, ce qui peut perturber la performance lors de la réalisation d'une tâche (Schmader *et al.*, 2008). En effet, elle impose à l'élève de faire deux choses : la tâche à réaliser, mais aussi gérer le stress qu'entraîne la peur d'être perçu comme incompétent. Cela s'avère délétère pour les performances intellectuelles car une partie de l'attention n'est pas allouée à la tâche. Pour mettre en évidence ce phénomène, des chercheurs ont demandé à des étudiants de réaliser un test difficile qui était présenté soit comme une évaluation de leurs capacités intellectuelles, soit comme un test ayant pour but simplement d'étudier la manière dont on mémorise les choses (Croizet & Claire, 1998). Les résultats montrent que lorsque le test est présenté comme une évaluation des capacités intellectuelles, les étudiants de classes populaires réussissent moins bien que les étudiants de classes favorisées. En revanche, cet écart

disparait lorsque le test n'est pas présenté comme tel. Ces résultats ont également été répliqués chez des jeunes enfants de CP (Désert *et al*., 2009).

Par ailleurs, d'autres travaux ont montré que les stéréotypes de genre qui décrivent les filles comme moins compétentes en mathématiques et en sciences produisent exactement les mêmes effets sur les inégalités de genre. Dans une étude, des collégiens et collégiennes devaient mémoriser une figure géométrique complexe avant de la reproduire sans modèle (Huguet & Régner, 2007). Cette tâche leur était présentée soit comme de la géométrie, soit comme du dessin. Les résultats montrent que dans la condition « géométrie », les filles obtiennent une moins bonne performance que les garçons, alors que dans la condition « dessin », elles réussissent mieux que leurs pairs.

II. L'importance de la présentation des situations évaluatives

Ainsi, les travaux sur la menace du stéréotype suggèrent que les situations d'évaluation, loin d'être neutres, ont tendance à accroître les écarts liés à l'origine sociale et au genre. Il ne faudrait pas en conclure pour autant qu'il faut renoncer à toute forme d'évaluation des élèves. Ces travaux posent la question de savoir ce qui peut être menaçant pour des élèves lorsqu'ils sont évalués et comment modifier les formes de l'évaluation. Les évaluations peuvent être plus ou moins orientées vers la mesure des performances à l'aide d'un indicateur simple (une note sur 20 par exemple) ou vers des éléments qui mesurent la progression des élèves[61]. Les travaux sur la menace du stéréotype révèlent que le fait de penser qu'un test pourrait mesurer de façon fiable des capacités intellectuelles qui seraient fixes – et sur lesquelles les élèves auraient peu de contrôle – perturbe la performance des élèves. En revanche, des études ont montré que lorsque l'évaluation n'est plus perçue comme le révélateur de capacités intellectuelles fixes, mais comme un outil au service de la progression, alors ces effets de menace disparaissent. Par exemple, dans une de leurs études, Smeding et ses collaborateurs (2013) ont manipulé la signification associée à un examen de statistiques. L'examen était présenté à des étudiants en première année de licence à l'université, soit comme un outil destiné à améliorer leur apprentissage (c'est-à-dire avec une fonction de formation), soit comme un outil permettant de sélectionner les meilleurs étudiants (c'est-à-dire avec une fonction de sélection). Les résultats montrent que lorsque l'examen était présenté comme un outil de sélection, les étudiants de classe populaire réussissaient moins bien que les étudiants de classe favorisée. En revanche, cet écart disparaissait lorsque l'évaluation était présentée comme un outil au service de l'amélioration des compétences.

III. Les risques de la comparaison sociale

Un autre élément déterminant des contextes dans lesquels se déroulent les apprentissages scolaires est l'omniprésence de la comparaison sociale entre les élèves. La comparaison sociale est un processus psychologique qui consiste à mettre en relation avec soi des informations de nature sociale, comme par exemple, la performance d'un ou de plusieurs autres élèves (Wood, 1996). Dans la salle de classe, le simple fait que plusieurs élèves réalisent une tâche identique en même temps est suffisant pour déclencher ce processus de comparaison sociale (Muller & Butera, 2007). Une des raisons pour

[61] Les indicateurs simples, comme la note sur 20 par exemple, parce qu'ils sont extrêmement synthétiques et parce qu'ils autorisent la comparaison voire le classement entre les élèves, ont tendance à focaliser l'attention de l'élève sur la performance plutôt que sur les savoirs et compétences maitrisées et sur les stratégies à mettre en œuvre pour progresser (Nicholls, 1984 ; Butler, 1987). Dans ce sens, ils sont assez peu favorables à une évaluation formative même s'ils ne l'empêchent pas.

lesquelles les élèves deviennent alors très sensibles à la performance de leurs camarades, et qu'il leur est même bien difficile d'en faire abstraction, est que celle-ci permet de répondre à une question fondamentale pour les êtres humains : « Qu'est-ce que je vaux ? » (Festinger, 1954). En effet, savoir si l'on est « bon » ne s'obtient bien souvent qu'en référence aux autres. Des travaux en psychologie sociale ont ainsi mis en évidence que voir ou imaginer quelqu'un réussir mieux que soi peut être menaçant pour l'image de soi. En effet, la réussite d'autrui peut être en décalage avec le besoin de maintenir une bonne image de soi, ce qui entraîne une réaction de stress et des pensées négatives coûteuses pour les ressources attentionnelles[62]. Mais pourquoi la réussite des autres est-elle menaçante ? Pourquoi n'est-elle pas plutôt une source d'inspiration ou de motivation ? En effet, sous certaines conditions, la réussite des autres peut être une source d'inspiration bénéfique pour l'image de soi et les performances (Huguet *et al.*, 2001). Une des conditions pour que cet effet inspirant soit possible est que l'élève perçoive la réussite d'autrui comme atteignable (Lockwood & Kunda, 1997). Or, les écarts de réussite qui sont donnés à voir dans la classe sont bien souvent perçus comme le reflet de différences individuelles relativement stables d'intelligence ou de motivation (Croizet & Claire, 1998). Le fait de penser que l'on est moins intelligent que les autres, que ces différences sont stables et donc qu'il n'est pas possible de réduire l'écart entre soi et autrui, est menaçant pour l'image de soi (Normand & Croizet, 2013). Cette tendance à interpréter les écarts de réussite comme le produit de dispositions internes est favorisée par le fait que la classe est supposée être un environnement juste. Les élèves n'ont donc guère d'autre choix que d'interpréter les écarts de réussite comme le révélateur de différences individuelles de capacités ou d'efforts. Ce que les élèves ne réalisent pas en revanche, c'est que la situation dans laquelle ils ou elles se trouvent n'est pas aussi équitable qu'elle semble l'être. En effet, en raison de leur socialisation familiale, les élèves se retrouvent à une inégale distance de la culture scolaire. Les élèves socialisés dans les classes moyennes et favorisées arrivent à l'école avec une familiarité importante vis-à-vis des attendus et savoirs scolaires comparativement aux élèves de classe populaire.

Dans une série d'études, Goudeau et Croizet (2017) ont mis en évidence le fait que les situations de comparaison sociale imposées par le contexte de classe (par ex., un élève qui lève la main pour dire qu'il a terminé un exercice), jouaient un rôle dans l'amplification des inégalités scolaires. Ces comparaisons délétères résultent du fait que les situations scolaires mettent en scène les écarts de réussite entre élèves d'une manière qui ignore l'inégale familiarité vis-à-vis de l'arbitraire culturel scolaire, ce qui oblige les élèves à interpréter ces (dés)avantages comme le reflet de différences de capacités intellectuelles, interprétation qui peut interférer avec le fonctionnement cognitif des élèves désavantagés.

Une première étude réalisée auprès d'élèves de 6e révèle ainsi le fait que rendre visibles les différences de réussite entre élèves (c'est-à-dire, comparaison sociale) durant une tâche de lecture-compréhension (c'est-à-dire, lever la main lorsque l'on a répondu à une question) perturbe la performance des élèves de classe sociale populaire. Dans une deuxième étude, les auteurs ont

[62] Ce résultat peut paraître *a priori* contradictoire avec le fait que le regroupement d'élèves en difficulté par groupe de niveau ou classe peut aussi avoir des conséquences négatives sur leur estime de soi et leur réussite (Marsh, 1984 ; Ireson & Hallam, 2001). Cette contradiction pourrait s'expliquer – même si cela mériterait de faire l'objet de recherches plus approfondies - par le fait que les élèves en difficulté regroupés ensemble peuvent faire l'objets d'attentes moins élevées et d'un traitement différencié de la part des adultes (enseignants et parents), entrainant une moindre progression, et révélant ainsi un phénomène de prophétie auto-réalisatrice (effet Pygmalion, Rosenthal & Jacobson, 1968). Ce phénomène pourrait également être causé par le fait que les élèves interprètent leur position dans ce groupe ou cette classe comme un manque d'intelligence ou d'effort de leur part (Moradoglu, Goudeau, & Cimpian, 2022).

manipulé expérimentalement le niveau de capital culturel d'élèves de CM2 afin de montrer comment la non-reconnaissance des différences initiales de capital culturel pouvait entraver la réussite des élèves les moins familiarisés avec les normes scolaires. Dans une première étape, les élèves étaient plus ou moins familiarisés avec un nouveau système arbitraire d'écriture. Dans une deuxième étape, ce nouveau système d'écriture faisait l'objet d'une évaluation. Pour une moitié des élèves, cette évaluation se réalisait dans un contexte de comparaison sociale où ils devaient lever la main lorsqu'ils pensaient avoir décodé un mot, rendant ainsi les écarts de performance visibles dans la classe. Pour l'autre moitié des élèves, cette tâche se réalisait sans que ceux-ci aient la possibilité de situer leur performance par rapport à celle des autres élèves. Les résultats de cette étude indiquent que les élèves peu familiarisés avec le nouveau code d'écriture voyaient leur performance diminuer en situation de comparaison sociale comparativement à la condition contrôle. N'étant pas conscients des différences de familiarisation avec la tâche, ces élèves, désavantagés parce qu'ils avaient été moins entraînés, n'avaient pas d'autre choix que d'interpréter leur moindre réussite comme le signe d'une infériorité intellectuelle, ce qui faisait chuter leur performance. En revanche, une troisième étude a permis de montrer qu'il était possible de neutraliser ces effets menaçants de la comparaison sociale en amenant les élèves à interpréter les écarts de réussite comme la conséquence de différences d'entraînement plutôt que de différences de capacités. Ces résultats sont en accord avec d'autres travaux sur la comparaison sociale ayant montré que voir réussir autrui mieux que soi peut être inspirant, mais à la condition de penser qu'il est possible d'atteindre le niveau de réussite d'autrui (Lockwood & Kunda, 1997).

IV. Le cas d'interactions informelles en classe pouvant être perçues comme ayant une dimension évaluative par les élèves

Au-delà des évaluations « formelles » qui ont lieu régulièrement dans la classe, il existe des contextes qui peuvent revêtir une dimension évaluative pour les élèves alors qu'ils ne sont pas présentés comme tels par les enseignants. Par exemple, lors des moments d'échanges collectifs en classe, les questions que posent un enseignant à ses élèves, les feedbacks qu'il donne, la réaction des autres camarades de classe peuvent revêtir une dimension évaluative pour l'élève. En effet, malgré leur caractère informel, ces situations (donner une bonne ou mauvaise réponse, être interrogé ou non) peuvent donner lieu à des inférences (de la part de l'enseignant ou des élèves) sur les caractéristiques de l'élève. Par exemple, ne pas être interrogé ou ne pas trouver la réponse à une question posée à la classe peut avoir des conséquences sur l'image que les élèves (et les autres) se font d'eux-mêmes. Des recherches réalisées dans des classes d'écoles maternelles à partir d'observations vidéo pendant des moments de « regroupement » montrent ainsi que comparés aux enfants de classes moyennes et favorisées, les enfants de classe populaire ont tendance à moins parler lors des échanges collectifs (Goudeau *et al.*, en préparation). Plus spécifiquement, ils sont moins susceptibles de prendre la parole spontanément (par exemple, répondre à une question posée collectivement) et ils sont également moins susceptibles d'être interrogés par l'enseignant. Enfin, lorsqu'ils parlent, ils ont tendance à garder la parole moins longtemps. Globalement, les élèves de classe populaire parlent en moyenne pendant ces moments de regroupement de 50 % à 75 % moins que leurs pairs (prises de paroles sollicitées et non sollicitées). Par ailleurs, dès l'âge de 5 ans, les enfants perçoivent les élèves qui parlent le plus comme intrinsèquement plus intelligents que les autres. Ces situations d'échanges collectifs, qui ont généralement pour objectif de favoriser la participation de chacun, pourraient donc paradoxalement contribuer à amplifier les inégalités de départ. D'une part parce que tous les élèves n'ont pas la même

probabilité de prendre la parole, et d'autre part parce que ceux qui le font le plus (plutôt des élèves de milieux favorisés) vont être perçus comme plus intelligents que les autres.

Cet ensemble de recherches met en lumière la façon dont les interactions quotidiennes dans la salle de classe peuvent subtilement contribuer à construire au quotidien les inégalités sociales. Les effets des évaluations formelles (ou informelles) sont dus en partie à une perception négative de soi (ou la peur d'être perçu négativement par les autres). Ces perceptions négatives peuvent être liées à des stéréotypes ou des comparaisons que l'on juge défavorables et qui amènent à penser que l'on est moins intelligent que les autres. Ces pensées fonctionnent comme des interférences qui vont perturber l'apprentissage de façon momentanée en consommant une partie des ressources cognitives, ce qui peut s'apparenter à faire l'expérience d'une double tâche). En revanche, si ces situations perdurent dans le temps, ces situations peuvent amener les élèves à se dés-identifier du domaine en question (Steele, 1997). Par ailleurs, il ressort également de ces travaux que l'interprétation que font les élèves des différences de réussite joue un rôle déterminant dans la construction de ces inégalités. C'est parce que les élèves ont tendance à expliquer les écarts de réussite comme le reflet de caractéristiques internes (comme l'intelligence ou la motivation), plutôt que par des causes externes ou situationnelles (telles que les stéréotypes, les pratiques éducatives parentales) que les situations d'évaluation peuvent créer des écarts de performance. Cette tendance à expliquer les différences de réussite comme le produit de caractéristiques internes à l'élève est lié à la fois aux propriétés du système cognitif et à un ensemble de croyances culturelles. Ainsi, cette préférence pour les explications internes est favorisée par le fonctionnement du système attentionnel et de la mémoire. Par exemple, comparés à des facteurs tels que l'origine sociale, les facteurs intrinsèques sont plus saillants et observables – ils viennent à l'esprit plus facilement et ils sont plus simples à mémoriser (Horne, Muradoglu & Cimpian, 2019). Au-delà des propriétés du système cognitif, cette tendance est également renforcée par les caractéristiques du contexte culturel. Ainsi, en Europe et en Amérique du Nord, les individus sont plutôt perçus comme des entités autonomes, séparées, indépendantes les unes des autres. Par conséquent, les différences de comportements en général, et de réussite en particulier, y sont perçues comme le produit de caractéristiques individuelles, dégagées de l'influence du contexte social (Markus & Kitayama, 1991). Mais dans les contextes culturels où les individus sont considérés comme plus interdépendants et reliés, en Asie par exemple, les individus sont plus sensibles au rôle des facteurs externes (par exemple, la famille, l'enseignant) dans la réussite (Plaut & Markus, 2005). Les biais en faveur des explications internes sont également encouragés par les caractéristiques des salles de classe. En effet, dans une classe, les enfants ont généralement le même âge, le même enseignant, ils bénéficient généralement du même enseignement et réalisent les mêmes tâches. Cette apparente homogénéité de l'environnement de la classe, renforcée par la croyance dans l'égalité des chances, fait ressortir les différences de réussite, tout en détournant l'attention des contraintes externes qui peuvent agir sur les performances, comme les stéréotypes, les différences de temps et de ressources que les familles des enfants peuvent consacrer à leur aide (Croizet et al., 2017). Enfin, les comportements et le langage des adultes peuvent renforcer par inadvertance ce biais. Par exemple, il n'est pas rare de donner des feedbacks sur les performances ou le comportement des élèves d'une façon qui met en évidence explicitement ou implicitement leurs caractéristiques personnelles. Or, il est bien établi qu'utiliser des termes tels que « brillants » pour décrire les élèves très performants ou consoler les élèves en difficulté en disant, par exemple, que « tout le monde ne peut pas être bon en maths » peut accentuer cette tendance (Heyman, 2008 ; Rattan, Good, & Dweck, 2012).

Plusieurs recherches ont été menées pour tenter de réduire ces effets de « menace » de l'image de

soi. Une façon de réduire ces phénomènes est d'agir sur la métacognition, c'est-à-dire la manière dont les individus pensent et se représentent le fonctionnement cognitif. Ainsi, amener des étudiants à penser que le stress est bon pour la performance ou encore expliquer à des collégiens que l'intelligence est un attribut malléable qui se développe plutôt qu'une capacité fixe, permet de contrecarrer les effets de menace du stéréotype (Johns, Inzlicht & Schmader, 2008 ; Good, Aronson & Inzlicht, 2003). Il est également possible de neutraliser les effets menaçants de la comparaison sociale en amenant les élèves à interpréter les écarts de réussite comme la conséquence de différences d'entraînement plutôt que de différences de capacités (Goudeau & Croizet, 2017).

Conclusion

Ainsi, ce qui se passe concrètement tous les jours dans la classe peut favoriser ou défavoriser les élèves en fonction de leur origine sociale ou de leur genre. Au-delà de l'engagement des enseignants pour la réussite de tous les élèves, certaines croyances culturelles et la façon dont est organisée l'école influence les interactions dans la classe. Au cours de situations évaluatives formelles comme au cours d'interactions ordinaires en classe, interpréter sa difficulté ou sa moins bonne réussite comme le fait d'être moins intelligent amène les élèves à ressentir du stress et des pensées négatives, ce qui peut détériorer la motivation et la réussite. Étant donné que les élèves issus de milieux populaires sont davantage susceptibles que les autres d'éprouver des difficultés et de faire l'expérience de comparaisons sociales qui leur sont défavorables (en partie en raison du fait qu'ils possèdent un capital culturel plus éloigné des savoirs et attendus scolaires), le biais en faveur des explications internes creuse les inégalités dans la classe. De façon similaire, le fait de penser que les filles réussissent moins bien que les garçons en mathématiques et en sciences parce qu'elles seraient moins « douées » pour ces domaines peut diminuer leur intérêt à s'engager dans ces matières et dans les carrières scientifiques ultérieures. De plus, au-delà de l'amplification des inégalités, le fait d'expliquer les écarts de réussite dans la classe comme la conséquence de l'intelligence ou de la motivation des élèves tend à faire paraître ces différences entre les enfants comme légitimes et justes, ce qui peut être un frein à la volonté à les réduire. Bien sûr, il existe d'autres causes économiques et structurelles à ces inégalités, mais ce qui se passe pendant la classe n'est pas à négliger. S'il paraît difficile de modifier directement les contextes culturels et institutionnels, les enseignants ont le pouvoir en revanche, dans une certaine mesure, de modifier certaines croyances et pratiques à l'intérieur de la salle de classe. Conduire les élèves à changer la façon dont ils interprètent leurs réussites, difficultés et échecs, modifier le sens associé aux différences de performances dans la classe, limiter la compétition, ou encore prendre en compte le fait que les élèves arrivent en classe inégalement préparés aux apprentissages et que par conséquent ce qui a été acquis à travers l'expérience familiale pour certains doit faire l'objet d'un travail explicite en classe pour les autres, sont autant de pistes possibles pour réduire la construction des inégalités sociales.

Références

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, *79*, 474. doi:10.1037/0022-0663.79.4.474

Croizet, J.-C. & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, 24, 588-594. doi:10.1177/0146167298246003

Croizet, J. C., Goudeau, S., Marot, M. & Millet, M. (2017). How do educational contexts contribute to the social class achievement gap: Documenting symbolic violence from a social psychological point of view. *Current Opinion in Psychology*, *18*, 105-110.

Désert, M., Préaux, M. & Jund, R. (2009). So young and already victims of stereotype threat: Socio-economic status and performance of 6 to 9 years old children on Raven's progressive matrices. *European Journal of Psychology of Education,* 24, 207-218. doi:10.1007/BF03173012

Festinger, L. (1954). A theory of social comparison processes. Human Relations, 7, 117-140. doi:10.1177/001872675400700202

Good, C., Aronson, J. & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*, 645-662. doi:10.1016/j.appdev.2003.09.002

Goudeau, S. & Croizet, J. C. (2017). Hidden Advantages and Disadvantages of Social Class: How Classroom Settings Reproduce Social Inequality by Staging Unfair Comparison. *Psychological Science*, 28, 162-170. doi: 10.1177/0956797616676600

Goudeau, S., Autin, F., Markus, H. R., Stephens, N. M., Croizet, J.-C., & Cimpian, A. (en préparation). Cultural mismatch in preschool: How teachers' and children's interactions play a role in the amplification of inequalities.

Heyman, G. D. (2008). Talking about success: Implications for achievement motivation. *Journal of Applied Developmental Psychology, 29*(5), 361-370. https://doi.org/10.1016/j.appdev.2008.06.003

Horne, Z., Muradoglu, M. & Cimpian, A. (2019). Explanation as a cognitive process. *Trends in Cognitive Sciences*, *23*(3), 187-199.

Huguet, P., Dumas, F., Monteil, J.-M. & Genestoux, N. (2001). Social comparison choices in the classroom: Further evidence for students'upward comparison tendency and its beneficial impact on performance. *European Journal of Social Psychology*, 31, 557–578. doi:10.1002/ejsp.475

Huguet, P., & Régner, I. (2007) Stereotype threat among schoolgirls in quasiordinary classroom circumstances, *Journal of Educational Psychology*, 2007, 99, 545-560. https://doi.org/10.1037/0022-0663.99.3.545

Ireson, J., & Hallam, S. (2001) *Ability Grouping in Education*. Londres : Sage.

Johns, M., Inzlicht, M. & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General, 137,* 691-705. doi:10.1037/a0013834

Leyens, J. P., Yzerbyt, V. & Schadron, G. (1996). *Stéréotypes et cognition sociale*. Éditions Mardaga.

Lockwood, P. & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, 73, 91-103. doi:10.1037/0022-3514.73.1.91

Markus, H. R. & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. Psychological Review, 98(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Marsh, H. W. (1984). Self-concept, social comparison, and ability grouping: A reply to Kulik and Kulik. *American Educational Research Journal, 21*(4), 799-806.

Muller, D. & Butera, F. (2007). The distracting effect of self-evaluation threat in coaction and social comparison. *Journal of Personality and Social Psychology*, 93, 194-211. doi:10.1037/0022-3514.93.2.194

Muradoglu, M., Goudeau, S. & Cimpian, A. Children's Explanations for Ability Grouping (en préparation).

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328–346. doi:10.1037/0033-295X.91.3.328

Normand, A. & Croizet, J.-C. (2013). Upward Social Comparison Generates Attentional Focusing When the Dimension of Comparison is Self-Threatening. *Social Cognition*, 31, 336-348. doi:10.1521/soco.2013.31.3.336

Plaut, V. C. & Markus, H. R. (2005). The "inside" story: A cultural-historical analysis of being smart and motivated, American style. In A.J. Elliott, & C.S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 457-488). New York: Guilford Press.

Rattan, A., Good, C. & Dweck, C. S. (2012). "It's ok—Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology, 48*(3), 731–737. https://doi.org/10.1016/j.jesp.2011.12.012

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart& Winston.

Ross, L. & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. McGraw-Hill.

Schmader, T., Johns, M. & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336–356. doi:10.1037/0033- 295X.115.2.336

Smeding, A., Darnon, C., Souchal, C., Toczek-Capelle, M. C. & Butera, F. (2013). Reducing the socio-economic status achievement gap at university by promoting mastery-oriented assessment. *PLoS One*, *8*(8), e71678.

Steele, C. M. (1997). A threat in the air: How stereotypes shape the intellectual identities and performance of women and african-americans. *American Psychologist*, 52, 509-516. doi:10.1037/0003-066X.52.6.613

Wood, J. V. (1996). What is social comparison and how should we study it? *Personality and Social Psychology Bulletin*, 22, 520-537. doi:10.1177/0146167296225009

Comment l'évaluation peut-elle soutenir un suivi des progrès qui permette à tous les élèves d'apprendre ?

LIESJE COERTJENS
UCLOUVAIN (BELGIQUE)

Introduction

Le domaine de l'évaluation formative est vaste et de multiples options existent pour que le personnel enseignant puisse la mettre en œuvre. La plus connue consiste à fournir une rétroaction à l'élève ou à l'étudiant après l'évaluation (Shute, 2008 ; Wisniewski *et al.*, 2020). Il existe cependant une autre option, moins connue : l'utilisation de l'évaluation formative (et plus précisément d'une évaluation rigoureuse des progrès des élèves ou des étudiants) pour différencier l'enseignement.

Un modèle bien décrit pour cette combinaison du suivi des progrès et de différenciation est le modèle « *Multi-Tiered System of Support*[63] » (MTSS). Ce modèle a été développé pour identifier et aider les élèves et étudiants en difficulté (Fuchs *et al.*, 2003). Bien que la majorité des recherches menées jusqu'à présent sur ce modèle proviennent des États-Unis[64], quelques premières études ont été publiées en Europe (Suède, Wolff, 2011 ; Irlande, Fives *et al.*, 2013 ; United Kingdom, Hatcher *et al.*, 1994).

Dans ce qui suit, je vais d'abord détailler le modèle MTSS et le rôle clé des deux composantes de ce modèle, à savoir le suivi des progrès et la différenciation. Ensuite, les résultats de la recherche sur ce modèle MTSS et ses deux composantes seront détaillés.

Pour illustrer le fonctionnement du modèle et décrire les résultats de la recherche, nous nous concentrerons sur les compétences en lecture. Ce choix s'explique par trois raisons principales. Premièrement, les compétences en lecture sont essentielles et sont importantes pour l'apprentissage dans d'autres domaines également, tant dans l'enseignement primaire que secondaire. Deuxièmement, il semble que le modèle MTSS soit fréquemment utilisé et étudié dans le domaine des compétences en lecture. Troisièmement, l'apprentissage de la lecture est un exemple relativement simple, qui permettra, nous l'espérons, de saisir pleinement la logique du modèle MTSS et de ses résultats. Il convient de souligner que le modèle MTSS s'applique également à d'autres domaines (comme par exemple les mathématiques).

[63] Système de soutien à plusieurs niveaux.

[64] Le modèle MTSS est également utilisé dans les écoles qui ne participent pas à une étude de recherche, voir Berkeley et al. (2020).

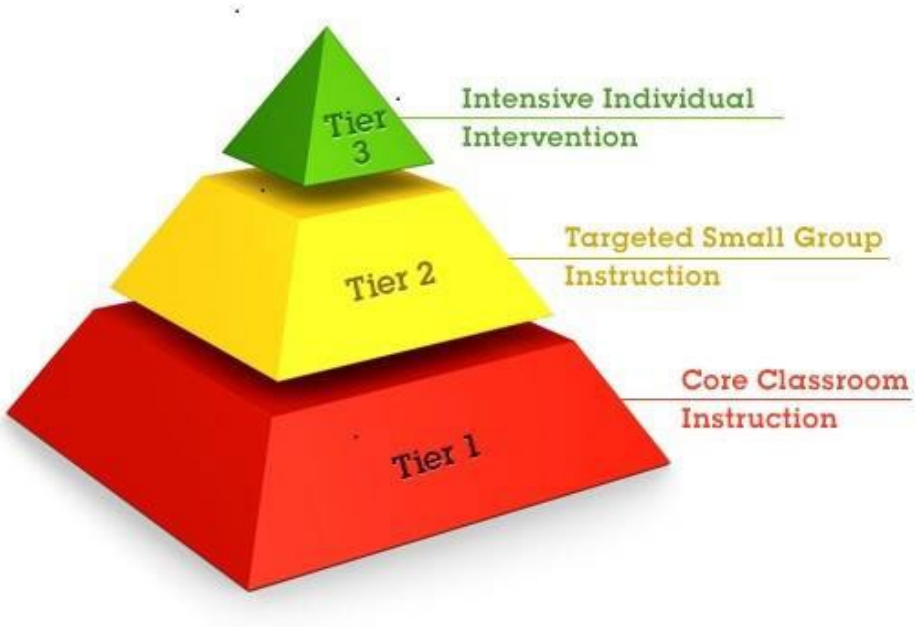
I. Le modèle « Multi-tiered system of support » (MTSS)

A. Le modèle MTSS « classique » et ses versions plus récentes

Dans sa version originale, le modèle MTSS comporte trois niveaux, qui sont présentés sous la forme d'une pyramide (voir Figure 1). Le premier niveau est constitué du programme d'enseignement élaboré par l'enseignant pour tous les élèves. Cela ne se limite pas à des explications de l'enseignant concernant un certain sujet (par exemple, dans l'apprentissage de la lecture, le son d'une nouvelle lettre), mais il s'agit de la combinaison de tous les moments de cours *ex cathedra* et de tous les exercices que l'enseignant prévoit pour que les élèves puissent maîtriser un certain sujet. Il est souligné dans le modèle que cette démarche doit être fondée sur des preuves (« *evidence-based* »), ce qui signifie qu'elle doit être conforme aux résultats de la recherche sur la manière d'enseigner un certain sujet.

Pour un grand nombre d'élèves (voir I. C.), cet enseignement de niveau 1 suffira pour maîtriser le nouveau contenu. Mais pour d'autres élèves, ce ne sera peut-être pas le cas. Le modèle MTSS suggère alors de fournir à ces élèves, un enseignement plus ciblé (comme la ré-explication de certains contenus, la réalisation d'exercices supplémentaires ou le *preteaching*[65]), pendant une courte période. Par exemple, plusieurs fois par semaine (3 à 5 fois), ces élèves bénéficient d'un enseignement supplémentaire de 30 minutes. Il s'agit du niveau 2 du modèle MTSS. Le modèle MTSS suggère que cet enseignement supplémentaire peut être dispensé en petits groupes (2 à 6 élèves maximum), mais dans la pratique, il est parfois dispensé individuellement (Gersten *et al.*, 2020 ; Neitzel *et al.*, 2021).

Figure 1 : Le modèle « *Multi-Tiered System of Support* » [66]



Tier 1 : niveau 1 – enseignement de base pour toute la classe
Tier 2 : niveau 2– enseignement ciblé en petits groupes
Tier 3 : niveau 3 – intervention individuelle intensive

[65] Enseignez aux élèves le contenu de la leçon à venir (en classe entière) à l'avance. De cette façon, ils ou elles travaillent deux fois la matière, ce qui peut être bénéfique pour leur apprentissage.

[66] Repris de https://www.fargo.k12.nd.us/Page/2059

De nouveau, pour un grand nombre d'élèves, l'enseignement de niveau 2 est suffisant pour l'acquisition des connaissances visées initialement. Cependant, pour quelques élèves, le niveau 2, ajouté au niveau 1, peut ne pas suffire (c'est-à-dire que ces élèves ne « répondent » pas de manière adéquate à l'intervention/au soutien, d'où le premier nom du modèle « réponse à l'intervention »). Pour eux, le modèle MTSS prévoit un niveau 3, à savoir un soutien individuel et intensif : c'est la principale caractéristique de ce niveau. L'enseignement est ici dispensé individuellement par un enseignant ou un membre de l'école à un élève, qui est ponctuellement retiré de la classe pour le suivre.

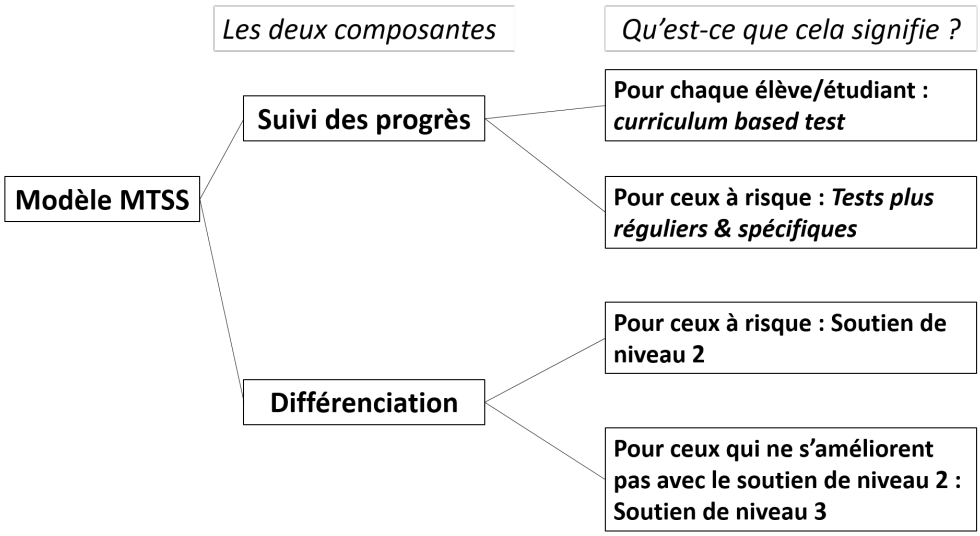
Certaines versions plus récentes du modèle MTSS divisent ce niveau 3 en deux : d'une part un soutien individuel et intensif tel que décrit ci-dessus, et d'autre part, une aide extérieure à l'école (par exemple le suivi par un logopède[67]) ou l'orientation vers la filière de l'éducation spécialisée.

Pour déterminer quels élèves ou étudiants ont besoin d'un soutien de niveau 2 et/ou 3, une évaluation rigoureuse est essentielle. La section suivante détaille cette évaluation de manière plus approfondie.

B. Quel est le rôle de l'évaluation dite « de suivi des progrès » (« *progress monitoring* ») dans le modèle MTSS ?

L'évaluation joue un rôle essentiel pour garantir que ni trop ni trop peu d'élèves et d'étudiants soient dirigés vers les niveaux 2 et 3, afin d'utiliser les ressources limitées de manière adéquate. **Le suivi des progrès comprend deux parties** (voir Figure 2).

Figure 2 : Carte mentale du modèle MTSS et de ses deux composantes



Premièrement, le niveau de compétence de tous les élèves est mesuré plusieurs fois par an en utilisant un « *curriculum-based test* ». Par exemple, au cours de la première année de l'enseignement primaire, lorsque les élèves apprennent à lire, ils lisent à haute voix des lettres et des mots pour discerner leur degré de maîtrise. Pour cela, un dispositif idéal consiste à faire effectuer un test court à l'élève, individuellement, en dehors de la classe, dans un endroit calme.

[67] Orthophoniste, en France.

Comme nous le verrons plus loin (voir I.D.1.), les *curriculum-based tests* sont élaborés par des experts (et non par les enseignants), pour un niveau d'âge précis, et dans le but de distinguer des différences fines de niveau, ce qui suppose que certaines questions ou tâches sont difficiles. Mais cela ne pénalise pas les élèves, car **ces tests ont une visée purement formative** : certes, ils permettent de faire un état des acquis de chaque élève (ce qui peut faire penser à une évaluation sommative), mais les résultats ne sont utilisés que pour déterminer le meilleur accompagnement possible pour chaque élève. Ils ne sont pas destinés à fournir une note pour le bulletin scolaire.

La fréquence à laquelle tous les élèves sont testés varie dans la littérature, mais tous les deux mois (Mathes *et al*., 2005), toutes les neuf semaines (Vernon-Feagans *et al*., 2010) ou trois fois par an (https://mtss4success.org/) semblent courants.

Un deuxième principe clé est que dès qu'un élève est identifié comme nécessitant un soutien de niveau 2, les tests sont plus réguliers et plus spécifiques. Cela a plusieurs objectifs : cela permet de voir si un mauvais résultat à un test basé n'est pas simplement le reflet d'un mauvais moment (par exemple, l'élève aurait très peu dormi). Il est certain que pour les élèves dont les performances sont très variables, il est difficile pour l'équipe éducative de discerner leur véritable niveau.

En outre, pour mieux comprendre les difficultés précises des élèves (par exemple, certaines lettres sont-elles plus difficiles pour les élèves que d'autres ?), le test est également plus spécifique. Il se concentre moins sur l'objectif final global (ce qui est le cas du *curriculum-based test*) et se concentre davantage sur des objectifs d'apprentissage spécifiques en relation avec l'enseignement dispensé au niveau 1 (par exemple l'association lettre-son). De cette manière, il fournit des pistes concrètes pour le soutien de niveau 2, l'idée sous-jacente étant que si un élève peut progresser sur ces objectifs plus spécifiques (association lettre-son), il progressera également sur l'objectif global (lire un texte).

En ce qui concerne la fréquence de ces tests plus spécifiques, les études diffèrent d'une fois toutes les quatre semaines (Little *et al*., 2012) à chaque semaine (McMaster *et al*., 2005 ; Vernon-Feagans *et al*., 2010).

C. La règle du 80 %-15 %-5 %, mais aussi du 60 %-30 %-10 % ou du 100 %-0 %-0 % est également possible !

Pour donner aux équipes scolaires et aux enseignants des idées sur le nombre probable d'élèves nécessitant du soutien de niveau 2 et 3, la littérature fournit des pourcentages. Bien qu'ils puissent varier d'une source à l'autre, il est fréquemment indiqué que pour 80 % des élèves, le niveau 1 suffira. Pour les autres élèves, le niveau 2 sera suffisant pour un grand nombre d'entre eux, soit 15 % de l'ensemble des élèves. Les 5 % restants auront besoin d'un enseignement supplémentaire de niveau 3.

Concrètement, pour une classe de 25 élèves, environ 20 progresseraient de manière satisfaisante dans leur apprentissage avec le niveau 1, environ 4 le feraient avec une combinaison des niveaux 1 et 2, tandis qu'un seul aurait besoin du soutien des trois niveaux.

Pourtant, chaque classe est différente. Cela n'aurait donc pas de sens d'appliquer la règle empirique du 80 %-15 %-5 % de manière systématique. En d'autres termes, si l'élève au rythme d'apprentissage le plus lent de la classe maîtrise le contenu avec seulement le niveau 1 (soit 100 %-0 %-0 %), il n'y a pas de raison pour mettre en place un enseignement de niveau 2 ! Dans d'autres classes au contraire, un plus grand nombre d'élèves aura besoin d'un enseignement de niveau 2 et 3 (par exemple 60 %-30 %-10 %).

D. La mise en place du modèle MTSS est un travail d'équipe

Le modèle MTSS suppose l'intervention conjointe de différents acteurs au service de l'apprentissage de tous les élèves. Dans cette sous-partie, nous donnons des précisions sur les personnes qui interviennent aux différentes étapes du dispositif.

1. La création de *curriculum-based tests*

Comme on l'a dit, les *curriculum-based tests* sont essentiels dans le modèle MTSS puisqu'ils permettent de repérer les élèves qui ont besoin des niveaux 2 et/ou 3 d'intervention. Ces tests ne sont pas à la charge des enseignants, mais élaborés par des chercheurs (par exemple, les tests DIBELS, https://dibels.uoregon.edu/, utilisés dans Wanzek & Vaugh, 2008 ; Denton *et al.*, 2013) ou par des agences gouvernementales (Vernon-Feagans *et al.*, 2010).

Pourquoi les *curriculum-based tests* ne sont pas élaborés par les enseignants ? Prenons à nouveau l'exemple de l'apprentissage de la lecture : supposons qu'un ensemble de textes *a priori* de plus en plus difficiles à lire par les élèves soit disponible tout au long de leur scolarité. Le travail de création ou de sélection de ces textes ne suffit pas. Il faudrait encore les valider en les testant sur un grand nombre d'élèves. Cela permettrait d'éviter les textes inappropriés (c'est-à-dire censés être d'un certain niveau, mais qui sont en réalité trop difficiles ou trop faciles). En outre, cette validation à grande échelle est nécessaire pour discerner les seuils d'un niveau de lecture « insuffisant ». La fixation de seuils trop ambitieux (c'est-à-dire un niveau de lecture que très peu d'élèves atteignent à un moment donné) ou trop bas (c'est-à-dire un niveau de lecture que tous les élèves atteignent, même ceux qui développeront des problèmes en lecture pour la suite) entraverait l'utilité du test. Ce travail de validation ne peut donc pas être concrètement réalisé par les enseignants eux-mêmes dans leur classe.

Pour permettre au modèle MTSS d'être largement utilisé dans les écoles, cela supposerait donc un panel de *curriculum-based tests* validés, permettant de suivre les progrès des élèves ou des étudiants sur plusieurs années. Ce panel de tests serait accompagné d'un manuel destiné au personnel enseignant, donnant des indications sur l'administration des tests et définissant les seuils permettant de déterminer quels élèves ou étudiants ont besoin d'un soutien de niveau 2.

2. Qui administre le *curriculum-based test* dans les écoles ?

Une fois un *curriculum-based test* créé et validé, il peut être administré dans les écoles. Le degré de collaboration nécessaire au sein de l'école varie en fonction du type de test. Reprenons notre exemple de l'apprentissage de la lecture. Un élève devra quitter la salle de classe pour lire quelques mots ou un texte à haute voix dans un endroit calme. Cela peut être devant l'enseignant, mais un membre de l'équipe (autre enseignant ou assistant pédagogique) devra alors prendre en charge la classe. Une autre possibilité est qu'un membre de l'équipe administre le test pour chacun des élèves. Peut-être que pour d'autres domaines, comme les mathématiques, il est possible de regrouper quelques élèves car le test peut être un test écrit. Ces éléments doivent être précisés dans le manuel accompagnant le test, mais la considération essentielle pour l'équipe d'enseignants est toujours d'obtenir une appréciation la plus précise possible du niveau des élèves ou des étudiants. Par conséquent, les conditions d'administration du test doivent être optimales pour permettre une vision correcte du niveau de maitrise des élèves ou des étudiants.

3. Qui réalise le soutien de niveau 2 ?

Lorsque les résultats d'un *curriculum-based test* sont connus, les élèves qui ont besoin d'un soutien de niveau 2 sont identifiés. Pour assurer ce soutien de niveau 2, deux possibilités existent. La première possibilité est que l'enseignant propose des exercices pour que tous les élèves puissent travailler de manière indépendante et qu'il regroupe ensuite les élèves qui ont besoin d'un soutien de niveau 2 dans la classe pour travailler avec eux de manière plus intensive. À ma connaissance, aucune étude empirique n'a utilisé cette première option, peut-être en raison de la difficulté à l'organiser.

Une deuxième possibilité est que l'élève ou les élèves nécessitant un soutien de niveau 2 quittent la classe individuellement ou par groupes de 2 à 6. De nouveau, l'enseignant peut effectuer ce soutien de niveau 2, mais il faudrait alors qu'un autre enseignant ou un assistant pédagogique prenne en charge la classe. Ou encore, un assistant pédagogique ou un bénévole pourrait assurer le soutien de niveau 2 (Gersten *et al.*, 2020 ; Neitzel *et al.*, 2021). Dans ce cas, cette personne est informée par l'enseignant sur le contenu et les méthodes utilisées en classe et, l'enseignant lui fournit des exercices.

À première vue, il peut sembler étrange de laisser une personne moins qualifiée assurer le soutien de niveau 2, mais les recherches sur l'apprentissage de la lecture au début de l'école élémentaire aux États-Unis suggèrent des avantages évidents à travailler de manière plus intensive avec un petit groupe d'élèves, même lorsque cette personne n'est pas un enseignant certifié (Gersten *et al.*, 2020). À partir de ces résultats de recherche, et même si le soutien apporté par un enseignant est légèrement (mais pas de manière significative) meilleur, il apparaît qu'un soutien aux élèves en difficulté par une personne moins qualifiée qu'un enseignant est préférable à une absence de soutien (Gersten *et al.*, 2020 ; Neitzel *et al.*, 2021).

4. Qui réalise le soutien de niveau 3 ?

Si un élève ne progresse pas après avoir mis en place un soutien de niveau 2, un soutien de niveau 3 est nécessaire. Il s'agit d'un soutien individuel, c'est-à-dire que l'élève quitte la classe pour une durée déterminée. En ce qui concerne la personne qui fournit ce soutien de niveau 3, les mêmes options que pour le niveau 2 existent : enseignant, autre enseignant de l'école, assistant pédagogique ou bénévole. En outre, le niveau 3 est parfois assuré par une personne plus spécialisée, comme un orthophoniste dans le cas de l'apprentissage de la lecture. Ce dernier soutien peut se faire soit à l'école, soit en dehors des heures de classe.

Pour le niveau 3, la même logique que pour le niveau 2 s'applique : d'après une étude récente, mieux vaut un soutien de niveau 3 réalisé par une personne moins qualifiée qu'un enseignant qu'aucun soutien de niveau 3 (Neitzel *et al.*, 2021). Les chercheurs observent toutefois que les bénévoles non rémunérés entraînent un gain moindre en matière de compétences en lecture, comparativement à un enseignant certifié (Neitzel *et al.*, 2021).

5. Qui crée et administre les évaluations supplémentaires pour les élèves ou étudiants bénéficiant d'un soutien de niveau 2 et 3 ?

Pour les élèves et étudiants identifiés comme étant « à risque » sur la base du *curriculum-based test*, bénéficiant d'un soutien de niveau 2 (ou 3), des tests plus fréquents et plus détaillés sont recommandés. Encore une fois, ce n'est pas aux enseignants de créer ces tests : ils sont créés soit par une équipe de recherche, soit par une équipe de conseillers pédagogiques. L'équipe pédagogique

sélectionne ensuite les tests les plus appropriés en fonction de la compétence étudiée (par exemple, la lecture) et de l'âge de l'apprenant.

En ce qui concerne l'apprentissage de la lecture, si les résultats d'un *curriculum-based test* indiquent qu'un élève peut avoir des difficultés, des tests supplémentaires (par exemple, sur la conscience phonologique, la fluidité, le décodage...) peuvent révéler quel aspect de l'apprentissage de la lecture est le plus problématique pour lui. Pour s'assurer d'une vision adéquate, il semble judicieux de réaliser ces tests dans une pièce calme et individuellement. Une fois cet ou ces aspects identifiés, le soutien de niveau 2 (ou 3) peut être adapté en conséquence.

Lorsque les sessions de soutien de niveau 2 ou 3 sont en cours, des tests supplémentaires réguliers sur ces compétences plus spécifiques sont recommandés. La littérature sur le modèle MTSS n'indique pas clairement qui administre ces tests, mais il semble que cela fasse souvent partie des sessions de soutien de niveau 2 ou 3 mises en place. Autrement dit, la personne qui assure ces sessions administre également ces tests.

II. Que savons-nous de l'impact du modèle MTSS et ses deux composantes sur la lecture ?

Dans les sections suivantes, nous détaillerons les résultats de la recherche concernant le modèle MTSS dans son ensemble et les résultats pour les deux composantes clés, à savoir le suivi des progrès et la différenciation.

Cinq méta-analyses publiées récemment semblent particulièrement pertinentes pour évaluer l'impact du modèle MTSS. Le Tableau 1 fournit plus d'informations sur les études incluses dans chacune de ces méta-analyses et sur la composante du modèle MTSS examinée.

A. L'impact du modèle MTSS

Récemment, Neitzel *et al*. (2021) ont publié une méta-analyse sur les programmes de soutien pour les lecteurs en difficulté. Quatre études empiriques rigoureuses ont examiné l'effet du modèle MTSS, soit appliquées à l'ensemble de l'école, soit au niveau de la classe.

Les résultats montrent un effet positif sur la lecture (0,27), mais en raison du petit nombre d'études (seulement quatre), cet effet n'était pas significatif (p=0,098). Il existe donc une première piste qui suggère que le modèle MTSS peut avoir un impact positif sur les résultats des élèves en lecture.

L'étude de Smith *et al.* (2016) est l'une des quatre études incluses dans la méta-analyse. Afin de donner un meilleur aperçu de la manière dont le modèle MTSS peut être mis en œuvre, nous la décrivons brièvement ci-dessous.

Les 44 écoles participantes de l'Oregon (États-Unis) ont été assignées aléatoirement à la condition de traitement (groupe-test, c'est-à-dire élèves bénéficiant de la mise en place d'un soutien selon le modèle MTSS) ou de comparaison (groupe-témoin, c'est-à-dire élèves ne bénéficiant pas du dispositif). Environ 3 000 élèves de première année ont participé, dont la moitié environ avait un accès à la restauration scolaire gratuit ou à prix réduit, ce qui est un indicateur du faible statut socio-économique de sa famille. L'étude s'est déroulée pendant toute la première année en primaire[68].

[68] Cours préparatoire en France

Un état des acquis de chaque élève a été effectué à l'automne de l'année scolaire. Tous les élèves de première année ont passé le *Stanford Achievement Test*, un *curriculum-based test* qui a été validé et normalisé par d'autres chercheurs au préalable. Plus précisément, les élèves ont passé les parties de ce test consacrées à la lecture de mots et de phrases, qui sont administrées à l'ensemble de la classe en même temps. Si un élève obtenait un score inférieur au 30e percentile, il était identifié comme « à risque » et recevait un soutien supplémentaire.

Tous les élèves ont reçu un enseignement de la lecture de 90 minutes par jour (niveau 1). Les enseignants ont participé à 5 journées de formation au cours de l'année scolaire, au cours desquelles les points suivants ont été soulignés : « (a) routines d'enseignement ; (b) aperçu de la recherche sur le contenu et les compétences de la lecture débutante, notamment la conscience phonémique, la phonétique, le vocabulaire, la compréhension et la fluidité de la lecture de textes connectés ; et (c) stratégies pour accroître l'engagement des élèves » (p. 556).

Les élèves ayant droit à un soutien de niveau 2 ont reçu un enseignement supplémentaire de 30 minutes en petits groupes (3 à 5 élèves) sur la lecture. Au cours de ces séances quotidiennes, qui ont duré toute l'année scolaire, un instructeur a travaillé sur « les compétences fondamentales en lecture, y compris la conscience phonémique, la lecture de mots et la fluidité de la lecture de textes reliés » (p. 555). Les instructeurs étaient principalement des assistants pédagogiques ; seuls quelques-uns étaient des enseignants certifiés. Les instructeurs ont participé à trois journées de formation et ont reçu un accompagnement (*coaching*) sur leur lieu de travail.

Les résultats indiquent que les élèves de la condition de traitement ont significativement plus progressé en lecture de mots que les élèves de la condition de comparaison (taille de l'effet de 0,24). Il existe également un effet sur la lecture de phrases (taille de l'effet de 0,18), mais il n'est pas significatif. La présence d'un effet est remarquable dans ce cas, car les écoles de la condition de comparaison ont également fourni un soutien supplémentaire aux lecteurs en difficulté en petits groupes. Cela signifie que l'effet positif était dû à la formation des instructeurs et à l'alignement explicite entre les niveaux 1 et 2, facilité par le matériel fourni aux enseignants et aux instructeurs pour cet enseignement de niveau 1 et 2.

Tableau 1 : Informations sur les différentes méta-analyses

Article	Études incluses dans la méta-analyse	Impact sur quel aspect ?
Dietrichson *et al.* 2017	Méta-analyse de 101 études (95 % provenant des États-Unis), réalisées à partir de 2000, qui portent sur des interventions en matière de mathématiques ou de lecture ciblant des élèves de faible statut socioéconomique. Pour être incluses dans la méta-analyse, les études devaient inclure un groupe de contrôle (soit le traitement habituel, soit un traitement placebo) et des tests standardisés[69]. En raison du petit nombre d'études et comme les résultats statistiques n'indiquent pas de grandes différences, l'effet des interventions en mathématiques (21 % des études) et en lecture est regroupé. L'étude s'est concentrée sur l'enseignement primaire et le début de l'enseignement secondaire.	5 études comprenaient une composante supplémentaire de retour d'information ou de suivi des progrès, afin d'optimiser l'enseignement futur (appelée « *feedback and progress monitoring* »). Les auteurs ajoutent que la plupart de ces études comprenaient également d'autres interventions, comme un accompagnement des enseignants (*coaching*). Du soutien de niveau 2 (appelées « *tutoring* » dans cette étude, définies comme un enseignement supplémentaire pour 1 à 5 élèves) ont été trouvées dans 36 études.
Dietrichson *et al.* 2020	Méta-analyse comprenant 71 études dans l'enseignement secondaire (83 % provenant des États-Unis), réalisées à partir de 1980, qui portent sur des interventions en matière de mathématiques ou de lecture ciblant des élèves de faible statut socio-économique. Les études devaient inclure un groupe de contrôle (soit le traitement habituel, soit un traitement placebo) et des tests standardisés. Comme certaines études combinent des interventions en lecture et en mathématiques, l'effet des interventions n'a pas été séparé. Pourtant, les interventions étaient plus souvent examinées à l'aide de tests de lecture (32 % utilisaient un test de mathématiques).	Dans 19 études, les enseignants ont reçu les résultats de leurs élèves au *curriculum-based test*, qualifiées de suivi des progrès. Cependant, ces méthodes étaient toujours combinées à d'autres méthodes, comme un accompagnement des enseignants (*coaching*) ou le soutien de niveau 2. Dans 22 études, l'enseignement en petits groupes a été utilisé, avec jusqu'à 5 élèves.
Dietrichson *et al.* 2021	Méta-analyse comprenant 205 études dans l'enseignement primaire (86 % provenant des États-Unis), réalisées à partir de 1980, qui portent sur des interventions en matière de mathématiques ou de lecture ciblant des élèves de faible statut socioéconomique. Les	Dans 19 études (dont 10 portant uniquement sur la lecture), les enseignants ont reçu des informations sur le progrès des élèves (par exemple à l'aide de *curriculum-based tests*). Cependant, ces méthodes étaient presque toujours combinées à d'autres méthodes : les deux études dans lesquelles ce n'était

[69] Un test standardisé est un test créé et validé par des chercheurs ou par des agences gouvernementales. Il diffère d'un *curriculum-based test* car, dans ce dernier cas, des seuils sont spécifiées pour interpréter les résultats (voir I. D. 1.).

	études devaient inclure un groupe de contrôle (soit le traitement habituel, soit un traitement placebo) et des tests standardisés. Comme certaines études combinent des interventions en lecture et en mathématiques, l'effet des interventions n'a pas été séparé. Pourtant, les interventions se concentrent plus souvent sur la lecture (18 % ont utilisé un test de mathématiques).	pas le cas portaient sur les mathématiques et étaient relativement anciennes (Fuchs *et al*., 1991 et Fuchs *et al*., 1996). Dans 118 études, un enseignement en petits groupes (semblable à un soutien de niveau 2) a été utilisé, avec un maximum de 5 élèves. 67 séances de soutien étaient individuelles, 48 en groupes de 2 ou 3, et 35 en groupes de 4 ou 5.
Gersten *et al.* 2020	Méta-analyse de 33 études provenant des États-Unis (publiées entre 2002 et 2017), qui portent sur des interventions en matière de lecture ciblant des élèves en 1e, 2e ou 3e primaire[70]. Les études devaient porter sur du soutien de niveau 2 (soit en petits groupes, soit individuellement) et durer au moins 8 heures. Les études devaient inclure un groupe de contrôle et des tests standardisés pour identifier les élèves à risque.	Du soutien de niveau 2 (défini comme un enseignement supplémentaire pour 1 à 5 élèves, en général 30 minutes par jour) ont été trouvées dans 33 études. 21 de ces études comprenaient des petits groupes (2 à 5 élèves) tandis que dans 12 études, un tutorat individuel était proposé. La durée médiane de la séance de tutorat était d'environ 25 minutes par jour, mais variait de 9 minutes à 52 minutes par jour. Ce tutorat était assuré par une personne sans diplôme d'enseignant (17 études), un chercheur titulaire d'un bachelier (9 études) ou des enseignants (7 études).
Neitzel *et al.* 2021	Méta-analyse de 65 études (95 % provenant des États-Unis), qui portent sur des interventions en matière de lecture ciblant des élèves en maternelle ou en 1e primaire, publiées entre 1990 et 2020. Les études devaient se dérouler sur une période d'au moins 12 semaines et être réalisées par des enseignants ou des assistants pédagogiques. Les études devaient inclure un groupe de contrôle, qui recevrait le traitement habituel, et des tests standardisés.	Quatre études ont examiné l'impact du modèle MTSS appliqué au niveau de la classe ou de l'école. Cela signifie que ces études incluaient à la fois le suivi des progrès et la différenciation (soutien de niveau 2 et, parfois, de niveau 3). Du soutien de niveau 2 (appelé « *One-to-small group tutoring* » dans cette étude, correspondant à un enseignement supplémentaire pour 1 à 6 élèves, en général 30 minutes par jour) ont été trouvées dans 23 études. 19 de ces études comprenaient des petits groupes (2 à 6 élèves), tandis que dans 4 études, des assistants pédagogiques encadraient ces petits groupes. Du soutien de niveau 3 (appelé « *One-to-one tutoring* » dans cette étude, correspondant à un enseignement supplémentaire pour 1 élève) ont été trouvées dans 25 études. Ce tutorat était assuré par des bénévoles non rémunérés (4 études), des bénévoles rémunérés (3 études), des assistants pédagogiques (4 études) ou des enseignants (14 études).

[70] Cela correspond à cours préparatoire, cours élémentaire niveau 1 et cours élémentaire niveau 2 respectivement

B. L'impact de la composante « suivi des progrès »

Pour discerner quels élèves et étudiants sont à risque et devraient bénéficier d'un soutien de niveau 2 (et plus tard de niveau 3), le suivi des progrès des élèves ou des étudiants est important. Trois méta-analyses ont examiné l'effet de différentes interventions ciblant les élèves à risque (Dietrichson *et al*., 2017 ; Dietrichson *et al*., 2021; Dietrichson *et al*.,2020). Toutes trois comprenaient des études à la fois sur la lecture et les mathématiques, mais les études sur la lecture aient été prédominantes pour l'ensemble des interventions (seulement 21 %, 18 % et 32 % des études étaient axées sur les mathématiques dans chacune des trois méta-analyses).

Étant donné le faible nombre d'études portant uniquement sur l'impact de la composante « le suivi des progrès » (voir Tableau 1), les résultats doivent être interprétés avec prudence. De plus, toutes les études sur le suivi des progrès des élèves concernant la lecture dans l'enseignement primaire comprenaient également d'autres composantes d'intervention (par exemple, formation du personnel) (Dietrichson *et al*., 2021). Il est donc difficile de démêler les effets et d'obtenir une image précise de l'effet du suivi des progrès en tant que tel.

Malgré ces limites, il semble que le suivi des progrès améliore les performances des élèves en lecture et en mathématiques : selon les différentes méta-analyses, les tailles d'effet moyennes étaient de 0,32 (Dietrichson *et al*., 2017, 5 études en primaire ou au début du secondaire), 0,173 (Dietrichson *et al*., 2021, 19 études en primaire) et 0,188 (Dietrichson *et al*.,2020, 19 études en secondaire).

C. L'impact de la composante « différenciation »

Pour les élèves ou étudiants qui ne progressent pas suffisamment en lecture avec l'enseignement de niveau 1 en classe, un soutien de niveau 2 est ajouté. De multiples méta-analyses ont examiné l'efficacité de ce soutien de niveau 2 (voir Tableau 1).

Différentes méta-analyses concluent que le soutien de niveau 2 améliore les résultats en lecture dans l'enseignement primaire (Neitzel *et al*., 2021 : 23 études avec une taille d'effet moyenne de 0,24, p<0,001 ; Dietrichson *et al*., 2017 : 36 études avec une taille d'effet moyenne de 0,36, intervalle de confiance de 95 % : 0,26-0,45 ; Dietrichson *et al*., 2021 : 118 études avec une taille d'effet moyenne de 0,376, intervalle de confiance de 95 % : 0,314-0,438). Ces résultats sont en accord avec l'affirmation de Baker *et al.* (2010) selon laquelle « environ 30 minutes d'enseignement en petits groupes chaque jour (3 à 5 jours par semaine) est très efficace pour les élèves qui ont des difficultés à apprendre à lire » (p. 3).

Comme le recommande le modèle MTSS, ce soutien de niveau 2 se fait en petits groupes (généralement 2 à 6 élèves). Pourtant, dans de nombreuses études, ce soutien est individuel. Par conséquent, les méta-analyses ont également comparé si ce choix (séance en petit groupe ou tutorat individuel) avait un impact sur l'efficacité du dispositif. Les résultats de différentes méta-analyses convergent à nouveau : le tutorat individuel est légèrement plus efficace, mais les différences avec un soutien en petit groupe sont faibles et non significatives (Dietrichson *et al*., 2017 ; 2021). Gersten *et al.* (2020) par exemple ont trouvé une taille d'effet de 0,31 pour les petits groupes et de 0,46 pour les tutorats individuels. Dans l'étude de Neitzel *et al*. (2021), la taille de l'effet était de 0,24 pour un petit groupe (2 à 6 élèves), contre 0,41 pour le tutorat individuel. Les chercheurs ont également testé si cette différence était significative, et l'ont fait par niveau scolaire. Les résultats indiquent qu'en première année (environ à 6 ans), le tutorat individuel est plus efficace que le travail en petit groupe,

mais ce n'est pas le cas en deuxième et troisième année (7-8 ans) (Neitzel *et al*., 2021). En résumé, les résultats suggèrent que le soutien de niveau 2 peut être organisé efficacement en petits groupes, mais qu'en première année primaire, le soutien individuel gagnerait à être privilégié.

Le niveau 3 est ajouté si le soutien de niveau 2 n'est pas suffisant pour un élève spécifique. Ce soutien de niveau 3 est toujours individuel. Une seule méta-analyse a examiné l'effet de ce soutien de niveau 3 : Neitzel *et al*. (2021) ont identifié 25 études, avec un effet moyen de 0,41. Il apparaît donc que le soutien de niveau 3 est efficace pour aider à apprendre à lire aux élèves en primaire pour lesquels les niveaux 1 et 2 n'ont pas suffi.

En guise de conclusion

Cette note avait pour but de détailler une possibilité moins connue d'évaluation formative, à savoir le suivi des progrès en vue de la différenciation. Cette articulation du suivi des progrès et de la différenciation est évidente dans le modèle MTSS. Ce modèle a été décrit en détail ainsi que les différents acteurs intervenant dans sa mise en place dans une école.

Ensuite, les résultats de la recherche empirique ont été présentés, axés sur l'application du modèle MTSS ou de l'une de ses deux composantes pour l'apprentissage de la lecture au début de l'enseignement primaire. Dans l'ensemble, il existe des premières preuves de l'efficacité du modèle MTSS, même s'il est évident que des recherches supplémentaires sont nécessaires. En ce qui concerne la composante « suivi des progrès », trop peu d'études l'ont examinée séparément pour pouvoir conclure sur son impact. En ce qui concerne la composante « différenciation » (c'est-à-dire le soutien de niveaux 2 et 3), de nombreuses méta-analyses rigoureuses montrent qu'elle est efficace pour améliorer les compétences en lecture des élèves.

Malgré ces (premiers) résultats prometteurs, deux remarques critiques sur la base de la littérature actuelle sont à noter. Premièrement, les études sont principalement dirigées par des chercheurs : les chercheurs ont développé des matériels, fourni des formations, évalué les progrès des enfants et vérifié la mise en œuvre correcte du soutien de niveau 2 et 3. Ces conditions sont souvent favorables à l'obtention de tailles d'effet (plus) positives et significatives. Il reste à déterminer si des effets similaires peuvent être obtenus lorsque le modèle MTSS est mis en œuvre dans un cadre scolaire plus ordinaire (hors expérimentation), par exemple lorsque le personnel enseignant évalue les progrès des élèves ou étudiants et utilise le matériel pour le soutien de niveau 2 sans formation supplémentaire.

Deuxièmement, la littérature se concentre principalement sur (le début de) l'enseignement primaire. Bien qu'il existe une première méta-analyse des études menées dans l'enseignement secondaire (Dietrichson *et al.*, 2020), le modèle MTSS reste relativement peu étudié dans l'enseignement secondaire.

Références

Baker, S. K., Fien, H., & Baker, D. L. (2010). Robust Reading Instruction in the Early Grades: Conceptual and Practical Issues in the Integration and Evaluation of Tier 1 and tier 2 Instructional Supports. *Focus on Exceptional Children*, 42(9). doi:10.17161/foec.v42i9.6693

Berkeley, S., Scanlon, D., Bailey, T. R., Sutton, J. C., & Sacco, D. M. (2020). A Snapshot of RTI Implementation a Decade Later: New Picture, Same Story. *Journal of Learning Disabilities, 53*(5), 332-342. doi:10.1177/0022219420915867

Denton, C. A., Tolar, T. D., Fletcher, J. M., Barth, A. E., Vaughn, S., & Francis, D. J. (2013). Effects of tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. *Journal of Educational Psychology*, 105, 633-648. doi:10.1037/a0032581

Dietrichson, J., Bøg, M., Filges, T. & Jørgensen, A.-M. K. (2017). Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis. *Review of Educational Research*, 87(2), 243-282. https://doi.org/10.3102/0034654316687036

Dietrichson, J., Filges, T., Klokker, R. H., Viinholt, B. C. A., Bøg, M., & Jensen, U. H. (2020). Targeted school-based interventions for improving reading and mathematics for students with, or at risk of, academic difficulties in Grades 7–12: A systematic review. *Campbell Systematic Reviews*, 16(2), e1081. doi:https://doi.org/10.1002/cl2.1081

Dietrichson, J., Filges, T., Seerup, J. K., Klokker, R. H., Viinholt, B. C. A., Bøg, M., & Eiberg, M. (2021). Targeted school-based interventions for improving reading and mathematics for students with or at risk of academic difficulties in Grades K-6: A systematic review. *Campbell Systematic Reviews*, 17(2), e1152. doi:https://doi.org/10.1002/cl2.1152

Fives, A., Kearns, N., Devaney, C., Canavan, J., Russell, D., Lyons, R., …O'Brien, A. (2013). A one-to-one programme for at-risk readers delivered by older adult volunteers. *Review of Education*, 1(3), 254–280. https://doi.org/10.1002/rev3.3016

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-Intervention: Definitions, Evidence, and Implications for the Learning Disabilities Construct. *Learning Disabilities Research & Practice*, 18(3), 157-171. doi:https://doi.org/10.1111/1540-5826.00072

Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020, 2020/04/02). Meta-Analysis of the Impact of Reading Interventions for Students in the Primary Grades. *Journal of Research on Educational Effectiveness*, 13(2), 401-427. https://doi.org/10.1080/19345747.2019.1689591

Hatcher, P.J., Hulme, C., & Ellis, A.W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, 65(1), 41–57. https://doi.org/10.2307/1131364

Little, M. E., Rawlinson, D. A., Simmons, D. C., Kim, M., Kwok, O.-m., Hagan-Burke, S., . . . Coyne, M. D. (2012). A Comparison of Responsive Interventions on Kindergarteners' Early Reading Achievement. *Learning Disabilities Research & Practice, 27*(4), 189-202. doi:https://doi.org/10.1111/j.1540-5826.2012.00366.x

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40*(2), 148-182. doi:https://doi.org/10.1598/RRQ.40.2.2

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to Nonresponders: An Experimental Field Trial of Identification and Intervention Methods. *Exceptional Children*, 71, 445-463.

Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2022). A Synthesis of Quantitative Research on Programs for Struggling Readers in Elementary Schools. *Reading Research Quarterly*, 57(1), 149-179. doi:https://doi.org/10.1002/rrq.379

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153-189. doi:10.3102/0034654307313795

Smith, J. L. M., Nelson, N. J., Smolkowski, K., Baker, S. K., Fien, H., & Kosty, D. (2016). Examining the Efficacy of a Multitiered Intervention for At-Risk Readers in Grade 1. *The Elementary School Journal*, 116(4), 549-573. doi:10.1086/686249

Vernon-Feagans, L., Gallagher, K., Ginsberg, M. C., Amendum, S., Kainz, K., Rose, J., & Burchinal, M. (2010). A Diagnostic Teaching Intervention for Classroom Teachers: Helping Struggling Readers in Early Elementary School. *Learning Disabilities Research & Practice*, 25(4), 183-193. doi:https://doi.org/10.1111/j.1540-5826.2010.00316.x

Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *J Learn Disabil*, 41(2), 126-142. doi:10.1177/0022219407313426

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology*, 10(3087). doi:10.3389/fpsyg.2019.03087

Wolff, U. (2011). Effects of a randomised reading intervention study: An application of structural equation modelling. *Dyslexia*, 17(4), 295–311. https://doi.org/10.1002/dys.438

Peut-on évaluer différemment des enfants différents ?

MICKAËL JURY
UNIVERSITÉ CLERMONT AUVERGNE (FRANCE)

Introduction

La loi du 8 juillet 2013 d'orientation et de programmation pour la refondation de l'école a consacré le principe de la scolarisation inclusive dans le code de l'éducation. Il est ainsi rappelé à l'article premier de celui-ci (art. L111-1) que le système éducatif et l'ensemble des acteurs et actrices qui œuvrent en son sein doivent s'assurer que tous les élèves, notamment celles et ceux en situation de handicap, puissent bénéficier d'une scolarisation dans le milieu ordinaire. Pour ce faire, l'article L111-2 du même code précise que « des dispositions appropriées [doivent rendre] possible l'accès de chacun, en fonction de ses aptitudes et de ses besoins particuliers, aux différents types ou niveaux de la formation scolaire. » En d'autres mots, les enseignantes et enseignants doivent s'assurer que l'ensemble des élèves puissent accéder au contenu des leçons dispensées, mais aussi que les évaluations proposées leur permettent de progresser et traduire réellement leurs compétences. Pourtant, touchant ce dernier point, il semble qu'il y ait une tension particulière au sein de la communauté éducative (Dubois, 2016). En effet, modifier les conditions dans lesquelles les évaluations se déroulent (par ex., *via* le recours à un ou une accompagnante d'élèves en situation de handicap – AESH – ou à une modalité différente comme l'oral au lieu de l'écrit pour compenser les difficultés d'un élève ayant des difficultés graphomotrices) peut laisser penser que l'évaluation ne permettrait plus de bien appréhender le niveau réel de compétence de l'élève ou même que celle-ci serait injuste puisque certains élèves seraient aidés par rapport à d'autres.

L'objectif de cette note de synthèse est de présenter des éléments issus de la recherche pour discuter de cette tension et essayer de la réduire.

Ainsi, dans un premier temps, des travaux issus de la sociologie et de la psychologie sociale de l'éducation seront évoqués afin d'essayer de comprendre pourquoi cette question de l'évaluation des élèves en situation de handicap est particulièrement sensible. Puis, des éléments sur les adaptations aux évaluations seront présentés et discutés à l'aide de travaux scientifiques afin de mettre en évidence comment les enseignants peuvent envisager des évaluations ne dressant pas de barrières supplémentaires pour les élèves en situation de handicap. Enfin, dans une dernière partie, une discussion questionnera le bien-fondé de cette tension lorsque l'on considère l'évaluation comme étant au service des apprentissages des élèves.

I. Sélection et inclusion : une incompatibilité particulièrement forte lorsqu'il s'agit d'évaluer les élèves

Pour comprendre pourquoi la question de l'évaluation des compétences des élèves en situation de handicap est source de tension pour la communauté éducative, il est important de revenir sur le fonctionnement de notre système éducatif.

D'après la perspective fonctionnaliste en éducation (Dornbusch *et al*., 1996), les systèmes éducatifs occidentaux auraient à la fois une fonction de formation mais aussi une fonction de sélection. Ainsi, si l'École doit « transmettre » des savoirs afin que chacun puisse « développer un socle commun de connaissances, de compétences et de culture devant permettre la poursuite d'études, la construction d'un avenir personnel et professionnel et préparer à l'exercice de la citoyenneté » (art. L122-1-1 du Code de l'éducation), il apparait qu'elle doit aussi permettre l'identification, parmi l'ensemble des élèves, de celles et ceux qui sont les plus aptes à obtenir les diplômes, celles et ceux qui seraient les plus méritants (Mijs, 2016). Dit autrement, le système éducatif devrait donc, grâce à une démocratisation de l'accès aux savoirs, offrir à l'ensemble des élèves les mêmes opportunités de développer leur potentiel, tout en ayant pour objectif de « faire le tri » afin d'assurer aux élèves sélectionnés une place correspondant à leur mérite individuel (Autin *et al*., 2015 ; Darnon *et al*., 2012).

Dans le système éducatif français, cette double fonction semble solidement ancrée. En effet, comme le rappelle Duru-Bellat (2019), le système actuel s'est construit sur l'École de Jules Ferry, à la fin du XIXe siècle, au sein de laquelle un ordre duel existe entre l'enseignement primaire, centré sur le fait fournir une instruction de base, et l'enseignement secondaire (les collèges et lycées) destinés à former les élites. Si l'enseignement primaire accueillait l'ensemble de la population, notamment à travers l'instauration de l'instruction obligatoire à l'âge de 6 ans, les collèges et lycées étaient réservés à une faible partie de celle-ci, les meilleurs, qui partageaient par ailleurs la particularité de venir quasi unanimement de milieux sociaux favorisés. Dans la première partie du XXe siècle, cet ordre social, de par les inégalités sociales qu'il maintient, sera rediscuté laissant émerger alors l'idée que l'accès aux plus hautes positions de l'État devait se faire sur la base de l'identification des enfants les plus méritants, indépendamment de leur milieu social d'origine. On parle alors d'« élitisme républicain ». Une nouvelle fois, consciente des inégalités importantes entre les élèves que génère ce principe, l'École tentera de se renouveler à travers la massification et « l'élévation continue du niveau culturel de la nation » (plan Langevin-Wallon, 1946). L'objectif est alors d'ouvrir la voie à une compétition plus juste pour les places les plus prestigieuses dans la société. Ainsi, si cette idée de sélection semble s'être peu à peu effacée dans les textes officiels, elle continue toutefois d'infuser régulièrement les discours des responsables de notre pays laissant planer une certaine ambiguïté. Par exemple, en 1985, Jean-Pierre Chevènement alors ministre de l'Éducation en France, rappellera son attachement à l'élitisme républicain, tout en reconnaissant que le premier objectif de l'École n'était pas « de sélectionner les meilleurs » mais bien de « donner à tous les enfants une bonne formation de base car tous les enfants peuvent apprendre ». Beaucoup plus récemment, Pap Ndiaye, lors de son intronisation en tant que Ministre de l'Éducation nationale, a rappelé qu'il était « un pur produit de la méritocratie républicaine ». Selon Duru-Bellat (2019), le maintien de l'ambivalence autour de l'existence de la fonction de sélection tiendrait notamment au risque de voir s'effondrer « la valeur marchande des diplômes et l'appariement efficace entre formations et emplois » (p. 52). Ce rapport à la sélection contribuerait ainsi à expliquer la valeur très forte que la société accorde à la notion de mérite scolaire.

La survalorisation du mérite dans notre système scolaire interroge la compatibilité avec le paradigme de l'école inclusive. Si vouloir identifier les élèves les plus méritants ne semble pas, *a priori*, en contradiction avec le fait de proposer des réponses pédagogiques adaptées aux besoins de tous les élèves, il apparait que la réalité puisse être un peu différente. En effet, tout se passe comme si le fait de lever les barrières à la participation de certains par la mise en place d'adaptations allait à l'encontre du principe d'une comparaison juste entre les élèves. Dit autrement, par la mise en place de ces adaptations lors des évaluations, les enseignants et enseignantes pourraient avoir la sensation que les élèves ne sont plus sur la même ligne de départ et que les résultats de la compétition méritocratique seraient alors faussés (pour une proposition similaire voir Barton & Slee, 1999 ; Corbett, 1999 ; Wong *et al*., 2004). Les enseignants feraient alors face à un dilemme entre égalité et équité (Thomazet *et al*., 2014). Khamzina *et al.* (2021) ont récemment testé cette hypothèse dans une série d'études interrogeant près de 900 enseignantes et enseignants français et leurs résultats appuient l'idée que la fonction de sélection du système éducatif ne serait pas compatible avec l'école inclusive puisque plus les enseignantes et enseignants pensent que le système éducatif participe à la sélection des élèves, plus ils croient en la méritocratie scolaire et moins ils soutiennent la mise en œuvre de l'école inclusive. Autrement dit, pour certains enseignants interrogés, vouloir ouvrir les portes de l'école à toutes et tous ne semble pas cohérent avec l'objectif d'identifier les plus méritants.

Pour réduire cette tension, il convient de s'interroger sur les bénéfices réels de la mise en place d'adaptations lors des évaluations. Qu'en dit la littérature scientifique ? Est-ce que proposer un autre format d'évaluation crée une rupture dans le « jeu » méritocratique ? Dans la suite de cette note, nous nous intéresserons à différentes adaptations disponibles pour les évaluations et examinerons ce que la littérature peut dire de leur efficacité.

II. Quelles pratiques pédagogiques pour l'évaluation des élèves en situation de handicap ?

Si les éléments présentés jusque-là mettent en exergue une potentielle tension pour les enseignantes et enseignants lorsqu'ils doivent évaluer les élèves en situation de handicap, il existe pourtant certains repères, notamment identifiés dans la littérature anglophone en éducation[71], afin de les accompagner dans cette pratique pédagogique si particulière.

Comme nous venons de l'évoquer, quel que soit l'objectif d'une évaluation, ou le type d'adaptations proposé, la préoccupation des enseignants doit être d'obtenir une mesure fiable des compétences des élèves sans compromettre l'équité entre les élèves. Pour ce faire, Bolt et Roach (2009), sur la base d'une revue de la littérature, ont proposé une liste de bonnes pratiques touchant l'aménagement des évaluations qui visent à lever les barrières à la participation des élèves en situation de handicap sans pour autant leur fournir un avantage.

[71] L'importance de la pratique des tests standardisés aux États-Unis d'Amérique a fortement coloré les travaux réalisés dans ce champ disciplinaire. En effet, afin de s'assurer de l'équité entre les élèves et étudiants, de nombreuses recherches expérimentales, quasi-expérimentales ou même qualitatives ont été conduites par des équipes de chercheurs et chercheuses en éducation. Il est à noter que cette thématique est beaucoup moins présente dans la littérature francophone. Ainsi, les données reportées ici ont principalement été recueillies auprès d'élèves et d'étudiants nord-américains. Nous invitons les lectrices et lecteurs à garder cet élément de contexte en tête à la lecture de cette note.

Comme le montre le Tableau 1, ci-dessous, il est ainsi possible de travailler sur la présentation de l'évaluation, le type de réponses à fournir, l'organisation ou encore les conditions de déroulement de l'épreuve.

Tableau 1 : Types d'adaptations pour l'évaluation des élèves en situation de handicap (extraits et adaptés depuis Bolt & Roach, 2009)

Catégorie	Adaptations	Définition
Présentation	Police : type et taille	L'évaluation peut utiliser une police adaptée en termes de type (sans serif), de taille (plus grande) ou d'espacement.
	Lecture à voix haute	L'AESH, un pair ou l'enseignant peut lire à voix haute les instructions, les questions et les réponses de l'évaluation à l'élève. Les enseignants peuvent également s'enregistrer pendant qu'ils lisent les instructions et transmettre le fichier audio à l'élève.
	Consignes	Les enseignants doivent s'assurer que leurs instructions sont claires et sans ambiguïté. Ainsi, ils peuvent expliciter leurs attentes et séquencer les consignes le plus possible. En outre, l'utilisation de pictogrammes au lieu des mots peut se révéler utile (pour un logiciel libre, voir https://www.pictoselector.eu).
Réponses	Type de réponses	Une évaluation n'est pas nécessairement écrite, la possibilité de répondre *via* d'autres modalités (par ex., l'oral, des QCM) peut être offerte.
	Scripteur	Si l'écriture est indispensable pour l'évaluation, il est possible de permettre à une autre personne d'écrire la réponse fournie pour l'élève. Si des doutes se posent quant à l'investissement d'un secrétaire, il est du devoir de l'enseignant de discuter et clarifier le rôle de celui-ci.
Organisation	Temps supplémentaire	Offrir un délai supplémentaire pour les élèves qui pourraient en avoir besoin est une solution déjà largement utilisée par les enseignants ou même dans les examens nationaux (voir plus bas). Pour autant, il nous semble important ici de la rappeler.
	Pauses dans l'évaluation	Contrairement à la simple prolongation de la durée des évaluations, certains élèves peuvent bénéficier de pauses pendant les examens. Par conséquent, les évaluations doivent être conçues *a priori* de manière à permettre ces pauses sans nuire à la validité de la mesure des compétences envisagée.

Dans la partie suivante, nous proposons de regarder un peu plus précisément des travaux ayant testé empiriquement les effets de certaines de ces propositions.

III. Effet des adaptations aux évaluations : quelques éléments de la littérature

Lorsque les enseignants et enseignantes proposent une évaluation adaptée à un élève en situation de handicap, ils peuvent se demander si la note que celui-ci obtient à l'issue de celle-ci a la même valeur que celle attribuée à un camarade ayant fait l'évaluation classique. En d'autres mots, ils peuvent se demander si l'évaluation adaptée permet de mesurer fidèlement le niveau de maîtrise de la compétence visée, si celle-ci est efficace.

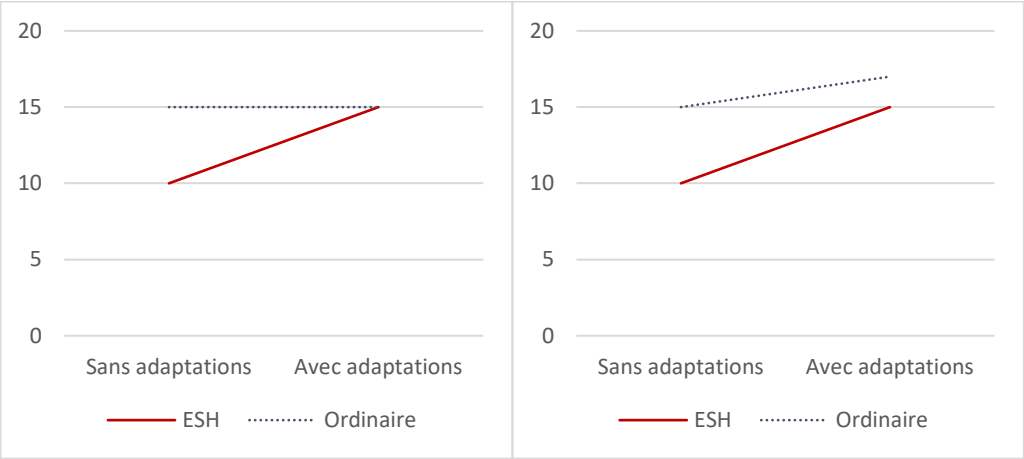
La question de l'efficacité des adaptations a fait l'objet d'une attention soutenue et des recherches ressort qu'une adaptation peut être considérée comme efficace :

- si elle permet aux élèves en situation de handicap qui en bénéficient de mieux réussir que s'ils ne l'avaient pas mais aussi qu'elle n'améliore pas la performance des élèves qui ne sont pas en situation de handicap si on la leur propose (hypothèse d'interaction, voir Figure 1, partie gauche) ;

Ou :

- si elle améliore la performance des élèves en situation de handicap qui en bénéficient dans une plus large mesure que celle des élèves qui ne sont pas en situation de handicap si on la leur propose (hypothèse du boost différencié, voir Figure 1, partie droite).

Figure 1 : Représentation des hypothèses d'efficacité des aménagements (hypothèse d'interaction à gauche, hypothèse du boost différencié à droite)



Parmi les propositions formulées dans le Tableau 1, ci-dessus, nous nous arrêtons spécifiquement sur deux qui pourraient être mises en place pour des élèves présentant des difficultés de lecture (par ex., des élèves dyslexiques). Pour ces derniers, la réalisation d'une évaluation dans le même temps que celui accordé aux autres élèves de la classe peut en effet poser des difficultés supplémentaires. Le temps nécessaire pour décoder les consignes pourrait notamment les empêcher de mobiliser et donc de témoigner de leurs compétences dans le temps imparti. Pour lever cette barrière, il est alors possible de leur proposer, par exemple, du temps supplémentaire pour réaliser l'évaluation ou la lecture à voix haute des consignes. Ci-après sont présentés des résultats spécifiques associés à ces deux adaptations.

A. Temps supplémentaire pour réaliser une évaluation

Le temps supplémentaire pour réaliser une évaluation est une adaptation très souvent utilisée pour les examens (par ex., en 2014-2015, cela concernait près de 90 % des étudiants en situation de handicap en France[72]). Comme son nom l'indique explicitement, il s'agit de permettre à l'élève en situation de handicap de bénéficier d'un temps supplémentaire (généralement un tiers supplémentaire du temps alloué) pour réaliser l'évaluation.

Touchant les effets de celle-ci, Sireci *et al*. (2005) disaient notamment, à l'issue d'une revue de la littérature, identifier un soutien modéré à l'hypothèse d'interaction évoquée précédemment. Plus précisément, parmi les 8 études examinées et conduites auprès de collégiens, lycéens et étudiants d'universités, 5 identifiaient des bénéfices de cette pratique pour les étudiants en situation de handicap (2 mettaient en évidence un bénéfice pour tous les élèves, 1 ne traduisait de bénéfice pour personne). Lovett (2010) concluait, *via* l'intégration de nouveaux travaux, que tous les élèves bénéficient de l'octroi d'un temps supplémentaire, même si le gain semble un peu plus grand pour ceux en situation de handicap.

Très récemment, Furlano *et al.* (2021) ont réalisé une étude expérimentale mettant à l'épreuve cette adaptation. Dans cette étude, ils proposaient à des étudiants d'université au Canada de compléter un test de mesure des compétences cognitives, soit dans le temps classiquement imparti, soit avec 50 % de temps en plus. Parmi les 234 étudiants ayant participé (moyenne d'âge, 18,45 ans), 27 indiquaient avoir un diagnostic de trouble des apprentissages. Lorsque les performances des deux groupes d'étudiants étaient comparées (sur le nombre d'items tentés et le score global), seul un effet des conditions de réalisation apparaissait clairement, dans le sens où les étudiants ayant passé le test avec plus de temps réussissaient mieux que leurs pairs ayant passé le test dans le temps classiquement imparti. Il n'y avait donc pas d'effet différencié de ces conditions selon le profil de l'étudiant. Dit autrement, les étudiants présentant un trouble des apprentissages n'avaient pas davantage bénéficié du temps supplémentaire.

Ainsi, au-delà de ces quelques exemples, il semble que les résultats des différents travaux dans cette littérature ne permettent donc pas d'être unanime quant à l'efficacité de cette adaptation pour l'évaluation (Lovett & Nelson, 2021). En effet, si elle semble bien bénéfique pour les élèves et étudiants en situation de handicap, elle ne peut pas être considérée comme une adaptation efficace au regard des conditions présentées précédemment (hypothèses d'interaction ou de boost différencié). Néanmoins, de manière pragmatique, considérer cette adaptation pour compenser certaines difficultés apparait comme pertinent. En outre, lorsque l'on interroge les étudiants d'université sur cette modalité, ils indiquent percevoir des bénéfices associés à celle-ci tels que la réduction de l'anxiété ou encore le sentiment de pouvoir démontrer réellement leurs compétences (Slaughter *et al*., 2020).

Il faut toutefois noter que sur le plan organisationnel, il n'est pas toujours possible de laisser 20 minutes en plus sur une évaluation d'une heure à un élève sans risquer de désorganiser sa journée d'école ou celle des autres élèves et professeurs. Une alternative, pour dépasser ce point précis, serait de mettre en œuvre une adaptation visant à soulager l'activité de lecture des consignes.

[72] Rapport n°2018-35 de l'Inspection Générale de l'Administration, l'Éducation nationale, et de la Recherche : « Les aménagements d'épreuves d'examens pour les élèves et étudiants en situation de handicap »

B. Lecture des consignes lors d'une évaluation

Pour lever les barrières rencontrées par les élèves présentant des difficultés de décodage, il est également possible de leur proposer la lecture des consignes ou des textes à étudier. Soulager cette partie de l'évaluation pourrait, par exemple, leur permettre plus facilement de montrer leur compétence en compréhension de texte ou encore en calcul (Andreou *et al*., 2019).

Sireci *et al.* (2005), dans la même revue de la littérature que celle citée précédemment, faisaient une synthèse de 10 études ayant examiné la mise en place de cette adaptation auprès d'élèves en école primaire, collège et lycée. Les résultats de 5 d'entre elles apportaient un soutien à l'idée que cela serait une réponse pédagogique satisfaisante, dans le sens où elle permettrait aux élèves en situation de handicap qui en bénéficient de mieux réussir sans pour autant améliorer les performances des élèves qui ne sont pas en situation de handicap (*cf*. hypothèse d'interaction).

Dans des travaux récents, Giusto et Ehri (2019) ont proposé à des élèves américains de niveau CE2 de réaliser un test de compréhension. L'échantillon incluait des élèves considérés comme des décodeurs moyens ou faibles. Les participants réalisaient le test selon 3 conditions expérimentales : seul face au livret (consigne classique), avec un accompagnement pour la réalisation du test et la lecture des consignes et des questions par un adulte, avec un accompagnement de la réalisation du test sans lecture des consignes et des questions par un adulte. Les résultats de l'analyse, sur un petit échantillon de 82 élèves (dont seulement 28 faibles lecteurs), tendent à appuyer l'hypothèse d'interaction puisque les élèves faibles décodeurs ont obtenu de meilleures performances seulement lorsque les consignes et questions étaient lues par l'adulte et que les performances des décodeurs moyens étaient inchangées selon les conditions de passation du test (on peut toutefois regretter qu'il n'y ait pas eu de bons décodeurs dans cette étude afin de comparer la pratique sur l'ensemble des élèves).

Ainsi, par rapport au temps supplémentaire pour réaliser l'évaluation, les différents travaux disponibles sur la lecture des consignes semblent soutenir davantage l'idée d'une adaptation « efficace » (soutien à l'hypothèse d'interaction). Si, comme précédemment, la littérature n'est pas unanime quant aux effets observés et qu'il convient d'être particulièrement prudent en termes d'implication, cet ajustement dans les pratiques évaluatives des enseignants apparait toutefois comme pertinent pour accompagner les élèves qui ont des difficultés de décodage, particulièrement ceux et celles à l'école élémentaire et au collège (Giusto & Ehri, 2019 ; Lovett & Nelson, 2021 ; Sireci *et al*., 2005).

L'ensemble des éléments présentés jusqu'à maintenant visait à contribuer au débat suggérant que la mise en place d'adaptations pour les examens pourrait créer une rupture d'égalité dans l'évaluation des élèves. Si cela ne semble pas être le cas, nous proposons, dans la partie suivante, une discussion autour de l'intérêt même de ce débat lorsque l'évaluation contribue avant tout à « aider l'élève à progresser et [à] rendre compte de ses acquis » (art. D 321-23 du code de l'éducation).

IV. Un débat, quel débat ?

Butera (2011, voir aussi Butera et al., 2021) rappelle que les élèves et enseignants évoluent dans des environnements compétitifs où les notes que les seconds adressent aux premiers (au collège, lycée, voire même encore parfois à l'école élémentaire) sont toujours aujourd'hui perçues comme une mesure fiable de l'apprentissage et nécessaires pour mesurer la « valeur » des élèves. Cette ambiguïté persistante sur le statut des notes contribue, très probablement, à expliquer pourquoi l'évaluation et

les adaptations qui peuvent être mises en place pour des élèves en situation de handicap sont sources de craintes, notamment quant à l'égalité entre les élèves (dilemme égalité-équité, voir Thomazet *et al.*, 2014). Pourtant, il est plus que nécessaire de rappeler comme Mottier-Lopez (2015) et d'autres avant elle, que l'évaluation n'a pas pour objectif prioritaire de sélectionner ou comparer les élèves mais bien, comme l'indique également le code de l'éducation, de les aider dans leurs apprentissages.

Dans cet objectif-là, la question qui traverse en partie cette note de synthèse (c'est-à-dire, est-ce juste d'évaluer différemment des élèves différents ?) pourrait finalement sembler inopportune. En effet, si l'enjeu d'une mesure fiable du niveau de maîtrise d'une compétence par l'évaluation adaptée reste fondamental dans une perspective formative de l'évaluation, la tension quant à l'égalité entre les élèves serait beaucoup moins saillante si le standard de définition de la compétence est l'individu plutôt qu'autrui (Elliot *et al*., 2011). Ainsi, pour dépasser ce débat, il semble urgent de continuer à faire évoluer les pratiques d'évaluation des enseignantes et enseignants.

Pour atteindre ce but et mettre en œuvre une évaluation plus inclusive, Wakeman *et al*. (2021), dans une revue de la trop faible littérature sur le sujet, proposent d'inviter les enseignants à penser en dehors des usages habituels de la note et de l'évaluation. Les autrices estiment notamment qu'il faudrait retravailler la manière dont sont construites ces tâches (quelles compétences sont visées ? construites par qui ? avec qui ?) et la manière dont les résultats à celles-ci sont communiqués aux élèves et responsables légaux. Elles insistent particulièrement sur l'importance de bien identifier les compétences et sous-compétences à évaluer afin que l'évaluation ne soit pas un agrégat difficilement interprétable. Dit autrement, celle-ci ne doit pas confondre des indicateurs renvoyant à la maîtrise d'une compétence, à la réalisation d'une tâche, à la participation de l'élève ou à ses progrès. Le fait de décomposer les compétences (comme dans le livret de scolarisation unique par exemple) pourrait ainsi permettre de jauger spécifiquement les réussites et difficultés des élèves sur des points très précis. Enfin, les autrices insistent sur l'intérêt de la co-construction des évaluations. Pour évaluer la compétence d'un élève bénéficiant d'un dispositif ULIS[73], il est donc fondamental que l'enseignant spécialisé, coordonnateur du dispositif, soit partie prenante.

En améliorant la clarté, la précision, et le sens d'une évaluation, qu'elle soit à destination d'un ou une élève en situation de handicap ou de toute la classe, cela rend beaucoup plus facile l'interprétation du jugement réalisé (par une note ou une appréciation) et des écarts entre les élèves, réduisant très probablement ainsi le risque que des inégalités de traitement puissent être perçues.

Conclusion

Permettre à tous les élèves, indépendamment de leurs particularités, d'avoir les mêmes opportunités de réussite est un objectif central de notre École. Si de nombreux défis s'imposent aux enseignants et enseignantes pour atteindre celui-ci, l'un des plus importants est probablement celui de l'évaluation des compétences des élèves en situation de handicap (Dubois, 2016).

Cette note de synthèse a présenté une des raisons expliquant potentiellement pourquoi cette activité est encore aujourd'hui difficile. Le fonctionnement du système, à travers sa mission de sélection et la promotion de la méritocratie, peut venir implicitement créer des barrières dans la mise en place d'adaptations pédagogiques pour l'évaluation. En effet, cela pourrait laisser à penser que l'on combat une injustice (la situation de handicap) en en installant une autre (vis-à-vis des pairs ordinaires). Aussi

[73] Unité Localisée pour l'Inclusion Scolaire

s'agit-il, pour les enseignants, de fournir des adaptations aux élèves en situation de handicap afin de leur permettre d'exprimer pleinement leurs compétences tout en ne créant pas de rupture au principe d'égalité avec les autres élèves.

Pour ce faire, des adaptations ont reçu une attention soutenue dans la littérature scientifique. Parmi celles-ci, nous avons évoqué le fait que le temps supplémentaire ou la lecture à voix haute des consignes semblent des pistes intéressantes pour lever des barrières que certains élèves pourraient rencontrer lors des évaluations. Si cela est prometteur et pourrait aider à la mise en œuvre de l'école inclusive, nous souhaitons néanmoins alerter les lecteurs sur plusieurs points de vigilance à l'issue de cette note de synthèse.

Premièrement, et comme évoqué dans la partie 0, l'incompatibilité entre sélection et évaluations adaptées nous semble reposer principalement sur une conception normative des évaluations (c'est-à-dire, utile pour comparer les niveaux des élèves). Une conception beaucoup plus formative de celles-ci, centrée notamment sur les progrès des élèves, rendrait la tension évoquée ici beaucoup plus faible.

Deuxièmement, les conditions de réalisation des travaux sur l'efficacité des adaptations pédagogiques présentées dans la partie 0 doivent nous amener à beaucoup de prudence. En effet, de petits échantillons, des participants pour lesquels l'adaptation proposée ne répond pas nécessairement aux besoins, ou encore des contextes parfois éloignés des situations de classe du quotidien nous paraissent autant de points à considérer au moment de lire ces résultats.

Finalement, le recours systématique à des adaptations individuelles lors des évaluations peut être stigmatisant pour les élèves en bénéficiant et ainsi éloigner de l'idéal inclusif voulu par le changement de paradigme (Thomazet, 2008). Une réflexion plus en amont sur les barrières potentielles lors d'une évaluation et une réponse plus collective que singulière (par ex., si je lis les consignes à haute voix pour un élève, pourquoi ne pas les lire pour toutes et tous ?) s'accorderaient probablement davantage avec l'idée d'une école dans laquelle chacun, quelle que soient la situation de handicap qu'il rencontre, ses difficultés passagères ou persistantes ou son environnement de vie, a, de droit, sa place.

Références

Andreou, G., Athanasiadou, P. & Tzivinikou, S. (2019). Accommodations on reading comprehension assessment for students with learning disabilities: A review study. *Psychology, 10*(4), 521-538.

Autin, F., Batruch, A. & Butera, F. (2015). Social justice in education: How the function of selection in educational institutions predicts support for (non)egalitarian assessment practices. *Frontiers in Psychology, 6*, 1–13.

Barton, L. E. N. & Slee, R. (1999). Competition, selection and inclusive education: some observations. *International journal of inclusive education, 3*(1), 3-12.

Bolt, S. & Roach, A. T. (2009). *Inclusive assessment and accountability: A guide to accommodations for students with diverse needs.* Guilford Press.

Butera, F. (2011). La menace des notes. In F. Butera, C. Buchs, & C. Darnon (Eds.), *L'évaluation, une menace ?* (pp. 45–53). Presses Universitaires de France.

Butera, F., Swiatkowski, W. & Dompnier, B. (2021). Competition in education. In S. Garcia, A. Tor, & A. Elliot(Eds.), *The Oxford Handbook on the Psychology of Competition*. Oxford University Press.

Corbett, J. (1999). Inclusive education and school culture. *International Journal of Inclusive Education, 3*(1), 53–61.

Darnon, C., Dompnier, B. & Marijn Poortvliet, P. (2012). Achievement goals in educational contexts: A social psychology perspective. *Social and Personality Psychology Compass, 6*(10), 760–771.

Dornbusch, S. M., Glasgow, K. L. & Lin, I. C. (1996). The social structure of schooling. *Annual Review of Psychology, 47*, 401–429.

Dubois, B. (2016). La flexibilité d'une évaluation : une réponse à la prise en compte de la diversité des élèves ? *La Nouvelle Revue de l'adaptation et de La Scolarisation, 2*, 79– 87.

Duru-Bellat, M. (2019). *Le mérite contre la justice*. Presses de Sciences Po.

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3× 2 achievement goal model. *Journal of Educational Psychology, 103*(3), 632-648.

Furlano, R., Goodrich, S., Darr, W., Gibbard, K., Ho, J. L., Wang, Z., ... & Hausdorf, P. A. (2021). Determining the appropriateness of extended time accommodations in standardized cognitive ability testing. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 53*(2), 152-163.

Giusto, M. & Ehri, L. C. (2019). Effectiveness of a partial read-aloud test accommodation to assess reading comprehension in students with a reading disability. *Journal of learning disabilities, 52*(3), 259-270

Khamzina, K., Jury, M., Ducreux, E. & Desombre, C. (2021). The conflict between inclusive education and the selection function of schools in the minds of French teachers. *Teaching and Teacher Education, 106*, 103454.

Lovett, B. J. (2010). Extended time testing accommodations for students with disabilities: Answers to five fundamental questions. *Review of Educational Research, 80*(4), 611-638.

Lovett, B. J. & Nelson, J. M. (2021). Systematic review: Educational accommodations for children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 60*(4), 448-457.

Mijs, J. J. B. (2016). The unfulfillable promise of meritocracy: Three lessons and their implications for justice in education. *Social Justice Research, 29*(1), 14–34.

Mottier Lopez, L. (2015). *Évaluation formative et certificative des apprentissages*. De Boeck

Sireci, S. G., Scarpati, S. E. & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of educational research, 75*(4), 457-490.

Slaughter, M. H., Lindstrom, J. H. & Anderson, R. (2020). Perceptions of extended time accommodations among postsecondary students with disabilities. *Exceptionality*, 1-15.

Thomazet, S. (2008). L'intégration a des limites, pas l'école inclusive ! *Revue des Sciences de l'Éducation, 34*(1), 123-139.

Thomazet, S., Merini, C. & Gaime, E. (2014). Travailler ensemble au service de tous les élèves. *La Nouvelle Revue de l'Adaptation et de la Scolarisation, 65*(1), 69-80.

Wakeman, S. Y., Thurlow, M., Reyes, E. & Kearns, J. (2021). Fair and equitable grading for ALL students in inclusive settings. *Inclusive Practices*, 2732474521105539

Wong, D. K., Pearson, V. & Lo, E. M. (2004). Competing philosophies in the classroom: a challenge to Hong Kong teachers. *International Journal of Inclusive Education, 8*(3), 261–279.

Comment la planification peut-elle permettre de passer d'une évaluation de l'apprentissage à une évaluation pour l'apprentissage ?

GONZAGUE YERLY

HAUTE ÉCOLE PÉDAGOGIQUE FRIBOURG (SUISSE)

Introduction

Cette note discute de l'apport de la planification des démarches d'évaluation comme levier pour aider les enseignants[74] à mettre en œuvre une évaluation *pour* l'apprentissage (*assessment for learning)* (De Ketele, 2010 ; Wiliam, 2011), nommée également « évaluation-soutien d'apprentissage » (Allal & Laveault, 2009 ; Laveault & Allal, 2016). Elle amène tout d'abord des résultats empiriques quant aux apports de la planification pour l'enseignement et l'apprentissage de manière générale, puis de manière spécifique pour l'évaluation des apprentissages. Elle amène ensuite des éléments théoriques et pratiques sur les apports d'une planification « flexible » des démarches d'évaluation des apprentissages, telle que proposée initialement par l'*Assessment Reform Group* (Broadfoot *et al*., 2002).

I. Les apports de la planification pour l'enseignement et l'apprentissage. Et l'évaluation des apprentissages ?

La planification est une pratique complexe pour laquelle les enseignants consacrent beaucoup d'énergie dans les phases pré-, inter- et postactives de leur enseignement (Dessous, 2002 ; Wanlin, 2009). Elle comprend la réflexion, la prise de décisions, l'opérationnalisation et le jugement des choix opérés (Clark & Yinger, 1979). Elle est une compétence cruciale pour un enseignement de qualité, c'est-à-dire qui permet un véritable apprentissage. En effet, différentes synthèses de travaux empiriques (Dessous, 2002 ; Hattie, 2009 ; Wanlin, 2009) démontrent, aussi bien au primaire qu'au secondaire, que les enseignants « efficaces » sont ceux qui savent planifier leur enseignement et l'apprentissage des élèves. Par exemple, les études expérimentales de Zahorik (1970) ou de Byra et Coulon (1994) (échantillon d'enseignants qui planifient une leçon *vs*. qui ne la planifient pas) montrent que les enseignants qui planifient sont « plus centrés sur les élèves et produisent des leçons de meilleure qualité » (Dessous, 2002). Toutefois, ces travaux soulèvent également le manque de réussite d'une planification trop rigide, qui ne s'adapte pas aux réactions des élèves.

Les principales données de recherche sur la planification sont fournies par le champ de l'étude de la pensée des enseignants. Elles sont recueillies par différentes méthodes (journal de bord, questionnaire, entretiens d'explicitation, etc.). Selon Wanlin (2016), reprenant différentes synthèses

[74] Par mesure de lisibilité, seul le masculin est utilisé comme générique.

de ces travaux scientifiques (Dessus, 2002 ; Tochon, 1993 ; Wanlin, 2009), les enseignants disent planifier pour trois fonctions, qui ne sont pas hiérarchisées : 1) pour des raisons personnelles et psychologiques : la planification permet de diminuer l'anxiété de l'enseignant, permet un meilleur contrôle de l'imprévu, renforce la maîtrise personnelle des savoirs enseignés, permet de développer des routines positives ; 2) pour assurer une meilleur cohérence pédagogique : la planification permet de définir et d'organiser plus précisément les buts visés, les activités d'apprentissage et leur cohérence ; 3) pour des arguments administratifs et organisationnels : la planification permet de s'assurer de couvrir l'entier du programme scolaire.

La synthèse de Wanlin (2009), qui reprend les principaux travaux de recherche francophones et anglophones de 1970 à 2009, montre que si l'enseignement et l'apprentissage sont – à des degrés variables et par des méthodes diverses – planifiés par les enseignants, les éléments liés à l'évaluation des apprentissages des élèves ne sont, très souvent, pas pris en compte dans leurs planifications. Aussi selon cette synthèse, les enseignants considèrent les objectifs et l'évaluation des apprentissages comme des éléments formels et consacrent davantage de temps à planifier des éléments plus pratiques, telles que les activités ou tâches des élèves. Ce constat est par exemple corroboré plus récemment par la recherche de Sallin (2020) dans l'enseignement secondaire en Suisse. Cette recherche qualitative (entretiens et analyse de documents de planifications auprès de 10 enseignants responsables chacun deux à trois disciplines scolaires) démontre que pour les enseignants interrogés, planifier l'évaluation des apprentissages se limite surtout à fixer une date pour un test sommatif. Pourtant, comme nous l'avons exposé, les apports de la planification sont bien documentés. Pourquoi n'en serait-il pas de même pour l'évaluation des apprentissages ? L'évaluation des apprentissages est un « processus » qui comprend différentes étapes (définir les objets à évaluer, collecter des informations, interpréter les informations, construire un jugement et prendre des décisions, communiquer et agir) (Allal, 2008 ; Durand & Chouinard, 2012). Toutefois, sans planification, l'évaluation des apprentissages risque de rester uniquement un « après-coup » (Yerly *et al*., 2019), un test mis en place dans l'urgence des calendriers scolaires. Dans un pareil cas, l'évaluation des apprentissages s'apparente davantage à un acte administratif et s'éloigne du paradigme contemporain de l'évaluation *pour* apprendre (Fagnant & Goffin, 2017).

Des travaux de recherche, notamment dans le champ de l'évaluation des apprentissages (Mottier Lopez et Laveault, 2008), démontrent depuis les années 1970, l'importance d'identifier et de communiquer clairement les objectifs d'apprentissage aux élèves afin de favoriser leurs apprentissages. Tourneur (1989) recense une vingtaine d'études expérimentales qui démontrent les effets positifs de la communication des objectifs aux élèves. Plus récemment, les travaux de synthèse de Hattie (2009), montrent un effet positif très important des intentions (*goal*) sur l'apprentissage des élèves (d = 0,56 ; 34e rang ; 11 méta-analyses). Un « apprentissage ciblé » (Hattie, 2017) comprend les intentions (objectifs) mais également les critères d'évaluation. La synthèse de Hattie (2009) montre que la planification et la communication des « intentions » favorisent également un autre facteur parmi les plus puissants pour améliorer l'apprentissages des élèves : la clarté de l'enseignant (*teacher clarity*) (d = 0,75 ; 8e rang ; 1 méta-analyse).

Sur la base des résultats de ses synthèses, une des stratégies proposées à destination des enseignants par Hattie (2012 ; 2017) afin de toucher les deux aspects précités (apprentissage ciblé ; clarté) est la « conception à rebours » (*backward design*) conceptualisée par Wiggins et McTighe (2005). Pour Hattie (2017), la « conception à rebours » est « l'une des meilleures façons de maximiser l'apprentissage »

(p. 152). La « conception à rebours » (nommé également « planification à rebours ») souligne l'importance d'une démarche de planification en trois étapes successives (Tomlinson & McTighe, 2010, p.26) :

1. Déterminer les résultats escomptés. Commencer par identifier clairement les cibles à atteindre (compétences et/ou objectifs) dans les programmes et fixer des priorités pour viser un apprentissage durable ;
2. Déterminer des preuves valables. Identifier plus précisément à quoi ressemble la réussite (par exemple des situations prototypiques, les critères de réussite) et prévoir des démarches d'évaluation à mettre en œuvre en cours d'apprentissage qui correspondent aux résultats escomptés ;
3. Prévoir les expériences d'apprentissage et la pédagogie à privilégier. Prévoir les moyens principaux pour l'atteindre les résultats escomptés : les pré-requis nécessaires, les principales activités, les différentes stratégies d'enseignement.

La « conception à rebours » place l'évaluation au cœur de la planification. Elle va plus loin que la simple identification des objectifs d'apprentissages et demande aux enseignants, lors de la planification, de « réfléchir comme des évaluateurs » (Tomlinson et McTighe, 2010, p. 26). Selon Tomlinson et McTighe (2010), elle permet aux enseignants d'éviter certains écueils : éviter de planifier essentiellement des activités, certes intéressantes mais parfois éloignées des buts d'apprentissage (surtout au primaire) ; éviter de planifier la réalisation de l'ensemble des contenus issus des manuels scolaires et non l'atteinte des objectifs et/ou compétences visées (surtout au secondaire).

II. Une planification « flexible » comme levier vers une évaluation-soutien d'apprentissage

Les travaux de recherche, mais également les textes prescriptifs au sujet de l'évaluation des apprentissages de nombreux systèmes éducatifs, s'orientent actuellement vers le paradigme de l'évaluation *pour* les apprentissages (De Ketele, 2010) ou *assessment for learning* (Wiliam, 2011). Cette évolution est notamment liée au développement des théories de l'apprentissage (Allal, 2019 ; Black & Wiliam, 2018), mais aussi à l'évolution des systèmes éducatifs, notamment avec l'approche par compétences (Legendre & Morrissette, 2014). Ancré dans ce paradigme, le concept d'« évaluation-soutien d'apprentissage » (Allal & Laveault, 2009 ; Laveault & Allal, 2016) amène une dynamisation du concept d'évaluation des apprentissages, sous ses différentes formes (formative et sommative), avec comme intention principales qu'elles servent à soutenir les élèves dans leurs apprentissages. Y sont notamment promus l'évaluation en continu, l'engagement actif de l'élève dans le processus évaluatif et les synergies entre les différentes formes d'évaluation (formatives et sommatives).

Parmi les dix principes édictés pour guider les enseignants vers une évaluation *pour* apprendre (*assessment for learning*), l'*Assessment Reform Group* (Broadfoot *et al*., 2002) défend l'idée que « l'évaluation devrait faire partie d'une planification efficace de l'enseignement et de l'apprentissage » (traduction libre, p. 2). Selon ce collectif de chercheurs anglophones reconnus, et à l'image de la « conception à rebours » (Wiggins & McTighe, 2005), la planification doit permettre à l'enseignant de bien identifier les cibles et les critères de réussite, de prévoir le recueil de l'information et d'organiser des opportunités de feedbacks. Néanmoins, la planification doit rester « flexible », c'est-à-dire adaptable selon les événements en classe. Aussi, selon ce principe, le rôle et l'implication des élèves dans le processus devraient également être planifiés par l'enseignant. Ainsi, une planification

« flexible » permet à l'enseignant de préparer un projet initial clair et précis mais laisse également une place importante à la réaction aux événements en classe et à l'interaction avec les élèves. Elle évite les effets néfastes d'une planification trop rigide. En effet, il est reconnu que les enseignants « experts » sont insatisfaits par des plans rigides, les jugent inadéquats (Tochon, 1993, 2013). Les enseignants qui sont des « experts adaptatifs » (Hattie, 2017) sont ceux qui sont capables de voir l'apprentissage du point de vue des élèves, de les situer par rapport aux cibles d'apprentissage et de flexibiliser leurs stratégies et leurs plans initiaux. Cette flexibilité est rendue possible grâce à une connaissance préalable des attentes et à la maîtrise de différentes stratégies.

Aussi, pour Black & Wiliam (2009), préciser et communiquer les intentions d'apprentissage et les critères de réussite est la première des cinq stratégies-clés afin de mettre en place une évaluation *pour* apprendre. Toutefois, ils soulignent également l'importance du feedback régulier et l'implication des élèves dans le processus évaluatif et que ces deux éléments sont rendus possibles et efficaces grâce à des cibles claires. Il n'est pas inutile de rappeler ici la puissance du feedback sur l'apprentissage des élèves selon la synthèse de Hattie (2009) (d = 0,73 ; 10e rang ; 23 méta-analyses). Toutefois, pour un effet plus marqué, Black et Wiliam (2009) relèvent que le soutien aux apprentissages des élèves passe par une pratique fréquente de l'évaluation formative, également spontanée et informelle dans l'action. En accompagnant des enseignants de mathématiques et de sciences naturelles à utiliser l'évaluation formative dans l'action, ces auteurs ont pu démontrer une augmentation du rendement des élèves d'environ 80 %, mesuré à l'aide de tests standardisés. Encore une fois, pour atteindre de tels résultats, Black et Wiliam (2009) soulignent que les cibles doivent être claires pour l'enseignant et comprises par les élèves, mais également que les élèves doivent être formés et impliqués dans le processus.

Les différents résultats de recherche présentés permettent de spécifier les qualités nécessaires d'une planification « flexible ». Une planification « flexible » rassemble les différents critères de qualité d'une planification proposés par Maulini (2004) : elle est à la fois hiérarchisée, dynamique et interactive.

- **Hiérarchisée**. À l'image du concept de « conception à rebours » (Wiggins & McTighe, 2005), une planification flexible oriente l'enseignant, dans la phase préactive, dans la construction d'un projet clair et précis. Elle permet d'en identifier les priorités qui seront les grandes balises du projet (objectifs/compétences et critères de réussite ; traces de l'apprentissage ; activités significatives). L'enseignant sait où il doit emmener ses élèves sans toutefois préparer l'intégralité de la séquence dans les détails au risque d'une trop grande rigidité.

- **Dynamique**. Par une planification dynamique, l'enseignant se prépare, dans la phase interactive, à adapter son plan initial selon les événements de la classe, par exemple en prenant davantage de temps ou en accélérant selon ses observations. Des retours en arrière ou des sauts en avant sont également favorisés. Des « espaces d'improvisation » (Tochon, 2013) sont prévus et sont rendus possibles grâce à des éléments bien hiérarchisés.

- **Interactive**. Une planification flexible prévoit de laisser un rôle important aux élèves, des éléments qui permettent la prise en compte de leurs réactions et de leurs besoins. Des moments sont prévus dans la planification de la séquence afin de les impliquer dans certains choix, de modifier le projet initial ou d'ouvrir des alternatives dans l'interaction.

III. Quels sont les apports d'une planification flexible des démarches d'évaluation des apprentissages ?

Une planification « flexible » des démarches d'évaluation vise la mise en œuvre d'une évaluation qui soutient les apprentissages (Laveault & Allal, 2016) et donc, *in fine*, vise un apprentissage des élèves en profondeur (Tochon, 2013 ; Fullan *et al*., 2018). En quoi une telle planification « flexible », comme définie auparavant, peut-elle favoriser la mise en œuvre d'une telle évaluation ? Pour répondre à cette question, nous discutons de ses apports quant à trois éléments-clés du concept « évaluation-soutien d'apprentissage » (Laveault & Allal, 2016) :

- *Une planification flexible de l'évaluation pour mieux intégrer l'évaluation en continu dans la séquence*
  Une évaluation en continu vise la régulation de l'apprentissage tout au long de la séquence, c'est-à-dire une régulation « interactive » (Allal, 2019). Selon Allal (2007) et Laveault (2007), le processus de régulation nécessite 1) des cibles claires, 2) un contrôle de la progression des élèves par rapport à ces cibles, 3) des feedbacks et 4) des actions de l'enseignant et/ou de l'élève (régulation ; autorégulation ; corégulation) pour orienter et différencier l'apprentissage vers ces cibles. La planification amène l'enseignant à identifier de manière précise les cibles et les critères avant de débuter la séquence d'enseignement/apprentissage, mais également à s'y référer tout au long de celle-ci pour récolter des traces variées et en apporter des feedbacks réguliers, situés dans l'action et en fort lien avec les cibles définies et comprise par les élèves. La planification des démarches d'évaluation favorise ainsi le « dialogue interactif » (Black & Wiliam, 2018) avec les élèves tout au long d'une séquence, notamment de manière informelle, tout en identifiant les moments cruciaux où leur formalisation est nécessaire. Ces prises de données régulières permettent l'analyse et la prise de décisions dans l'interaction dans le but de modifier les choix initiaux ou de s'assurer de leur bonne conduite. L'évaluation fait ainsi partie des pratiques quotidiennes, de manière informelle et non chronophage.

- *Une planification flexible de l'évaluation pour renforcer l'articulation entre formatif et sommatif*
  L'idée des possibles synergies entre évaluation formative et sommative est déjà ancienne (Black & Wiliam, 2018 ; Harlen, 2005). Selon Laveault et Allal (2016), les synergies peuvent se faire de trois manières, toutes avec l'intention de soutenir les apprentissages : 1) les données récoltées durant l'apprentissage (évaluation formative) peuvent contribuer à ajuster l'évaluation sommative ; 2) l'évaluation sommative peut être une opportunité d'apprendre, si son *design* permet un véritable feedback et des actions de régulation ; 3) le processus peut être balisé en étapes formatives et sommatives pour une même production travaillée durant la séquence. Ces différentes voies permettent, mais aussi exigent, l'« alignement curriculaire » (Anderson, 2002 ; Pasquini, 2019) tout au long d'une séquence et la triangulation de données d'évaluation significatives afin de construire un « jugement professionnel » (Lafortune & Allal, 2007) tout au long de la séquence. Il est donc nécessaire de récolter des traces (ou preuves) d'apprentissage tout au long du processus. Ceci ne peut pas se créer sur le moment, mais demande une véritable planification. Celle-ci pourra ensuite être régulée tout au long de la séquence tout en restant proche des cibles identifiées.

- *Une planification flexible de l'évaluation pour renforcer l'implication des élèves ?*
  Une planification flexible, par son caractère interactif, prévoit des moments qui impliquent les élèves dans l'analyse du déroulement de l'apprentissage et les décisions à prendre. Elle leur donne un véritable rôle d'acteur. L'évaluation en elle-même est considérée comme un apprentissage (*assessment as learning*) (Earl, 2013). L'évaluation des apprentissages, sous ses différentes formes (formative et sommative), vise à développer les compétences d'autorégulation des élèves (Andrade et Brookhart, 2020). Pour ce faire, la planification initiale hiérarchisée permet plus aisément de communiquer les intentions, pour que les apprenants se les approprient. L'enseignant prévoit des moments afin d'expliciter aux élèves quels sont les cibles et les critères et de proposer des situations pour qu'ils se les approprient. L'étude par questionnaire de Yerly & Berger (2022) montre par exemple, dans la formation universitaire, que les perceptions que se font les étudiants des intentions du professeur pour un examen peuvent influencer positivement leurs pratiques d'apprentissage autorégulés. Dans une planification flexible, l'enseignant peut prévoir également des moments et actions afin de co-construire ou d'adapter le « référentiel » d'évaluation avec les élèves (Dechamboux & Mottier Lopez, 2019). Il peut également impliquer les élèves dans le choix (différencié) des modalités des différentes situations évaluatives. Dans sa planification, l'enseignant peut également prévoir des moments ou étapes propices à l'autoévaluation, l'évaluation par les pairs ou la co-évaluation qui seront rendus plus efficaces grâce à une meilleure connaissance des cibles d'apprentissage. Le but ultime consiste à ce que les élèves deviennent de meilleurs apprenants, capable de se donner du feedback à eux-mêmes (Wyatt-Smith, 2021).

IV. Comment pratiquer une évaluation flexible des démarches d'évaluation ?

La planification des démarches d'évaluation se fait conjointement à la planification de l'enseignement et de l'apprentissage, afin de créer une séquence d'« enseignement-apprentissage-évaluation » (Gérard, 2013). Plus concrètement, quels peuvent être les éléments pris en compte dans une planification « flexible » des démarches d'évaluation des apprentissages ? Le Tableau 1, ci-dessous propose des pistes, non exhaustives mais inspirées des différents éléments des sections précédentes, afin de renforcer les différentes caractéristiques d'une planification flexible. Elles sont catégorisées selon les trois critères proposés initialement par Maulini (2004).

Tableau 1 : Éléments pouvant être pris en compte dans la planification flexible des démarches d'évaluation des apprentissages

Hiérarchisée	Définir les cibles (compétences ou objectifs) et les critères de réussite avant de planifier tout autre élément, en lien avec l'analyse du plan d'études et le choix des approches didactiques liées aux disciplines scolaires. Identifier et articuler les étapes de la séquence qui visent l'acquisition de ressources et celles qui visent leur mobilisation (situation de compétences). Planifier les différentes situations/moments d'évaluation (formatives et sommatives) incontournables afin de les agencer et de les articuler, en lien avec les différentes phases d'une séquence (découverte, construction, entraînement, transfert).
Dynamique	Planifier la séquence dans ses grandes lignes (hiérarchisée), afin de réserver des espaces d'ouverture qui permettent des approfondissements, des retours en arrière. Planifier des moyens d'observation (formels et informels) en continu de la progression des élèves par rapport aux cibles visées qui permettent de réguler le projet initial. Planifier des moments privilégiés pour différencier son enseignement en lien avec les prises de données.
Interactive	Planifier de brefs moments (discussions métacognitives) pour expliciter aux élèves les intentions et les démarches d'évaluation et leur agencement dans la séquence. Planifier des moments, au début de séquence, pour co-construire / formuler les critères de réussite avec les élèves. Planifier des moments d'autoévaluation ; d'évaluation par les pairs ; de co-évaluation par rapport aux cibles visées.

Conclusion

L'objectif de cette note était de discuter des apports de la planification dans la mise en œuvre effective d'une évaluation qui soutient les élèves dans leurs apprentissages, afin de dépasser les intentions, souvent acceptées mais peu réalisées. Notre contribution aura été de mettre en lien les différentes connaissances scientifiques quant aux apports de la planification de manière générale pour l'enseignement et l'apprentissage et ceux spécifiques au champ de l'évaluation des apprentissages. Les connaissances scientifiques actuelles plaident pour une planification « flexible » pour atteindre une évaluation *pour* apprendre, c'est-à-dire une planification approfondie et structurée (hiérarchisée) mais qui laisse aussi une place importante aux ajustements dans l'interaction (dynamique) et qui implique les apprenants (interactive). Toutefois, bien que les apports de la planification pour l'enseignement et l'apprentissage soient fortement étayés par la littérature, ils restent encore peu éprouvés sur le terrain en ce qui concerne l'évaluation des apprentissages, d'autant plus pour une planification « flexible ». Aussi, le développement de telles pratiques risque de se confronter à d'importants obstacles. Comme nous l'avons mentionné, une pratique flexible est plutôt le fait d'enseignants « experts ». Une telle planification nécessite donc des connaissances théoriques et expérientielles importantes de la part des enseignants (contenus des programmes, fonctionnement des élèves, stratégies didactiques, etc.) acquises en formation (initiale et continue) et en exercice sur le terrain, mais également des ressources et des conditions contextuelles favorables (travail en équipe, soutien hiérarchique, etc.).

Références

Allal, L. (2007). Régulation des apprentissages : Orientations conceptuelles pour la recherche et la pratique en éducation. In L. Allal & L. Mottier Lopez (Eds.), *Régulation des apprentissages en situation scolaire et en formation* (pp. 7–23). De Boeck.

Allal, L. (2008). Évaluation des apprentissages. In A. van Zanten (dir.), *Dictionnaire de l'éducation* (pp.312-314). Presses Universitaires de France.

Allal, L. (2019). Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice, 27*(4), 332-349

Anderson, L. W. (2002). Curricular Alignment: A Re-Examination, *Theory into Practice, 41*(4), 255-260.

Black, P. J. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.

Black, P. et Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551-575.

Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M. & Stobart, G. (2002). *Assessment for learning: 10 principles.* University of Cambridge School of Education

Byra, M. & Coulon, S. C. (1994). The effect of planning on the instructional behaviors of preservice teachers. *Journal of Teaching in Physical Education*, 13(*3*), 123-139.

Clark, C. M. & Yinger, R. J. (1987). Teacher planning. In J. Calderhead (Éd.), *Exploring teachers' thinking* (pp. 84-103). Cassell.

Dechamboux, L. et Mottier Lopez, L. M. (2019). Co-construire le référentiel de l'évaluation formative pour soutenir un processus de co-régulation dans la microculture de classe. *e-JIREF, 5*(2), 86 −111.

De Ketele, J.-M. (2010). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée*, *15*(1), 25-37.

Dessus, P. (2002). Les effets de la planification sur l'activité de l'enseignant en classe. In P. Bressoux (Éd.), *Les stratégies de l'enseignant en situation d'interaction* (pp. 17-28).

Durand, M.-J., & Chouinard, R. (2012). *L'évaluation des apprentissages. De la planification de la démarche à la communication des résultats.* Heurtebise.

Earl, L. (2013). *Assessment as learning.* Corwin Press.

Fagnant, A. & Goffin, C. (2017). Les conceptions des futurs enseignants du secondaire en matière d'évaluation : Entre un accord de principe et une vision limitée de l'évaluation formative. *Mesure et évaluation en éducation*, 40(1), 1-32.

Fullan, M., Quinn, J., & McEachen, J. (2018). *L'apprentissage en profondeur : s'ouvrir au monde, changer le monde*. Presses de l'Université du Québec.

Gérard, F.-M. (2013). L'évaluation au service de la régulation des apprentissages : Enjeux, nécessités et difficultés. *Revue française de linguistique appliquée*, *18*(1), 75-92.

Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies. *Curriculum Journal, 16*, 207–223.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning.* Routledge.

Hattie, J. (2017). *L'apprentissage visible pour les enseignants. Connaître son impact pour maximiser le rendement des élèves.* Québec : Presses de l'Université du Québec.

Lafortune, L. & Allal, L. (2007). *Jugement professionnel en évaluation. Pratiques enseignants au Québec et à Genève*. Presses de l'Université du Québec.

Laveault, D. & Allal, L. (2016) *Assessment for Learning: Meeting the Challenge of Implementation.* Springer.

Laveault, D. (2007). De la « régulation » au « réglage » : élaboration d'un modèle d'autoévaluation des apprentissages. In L. Allal & L. Mottier Lopez (éd.), *Régulation des apprentissages en situation scolaire et en formation* (pp. 207-234). De Boeck.

Legendre, M.-F. et Morissette, J. (2014). Défis et enjeux de l'approche par compétences dans un contexte d'obligation de résultats. Dans J. Morissette et M.-F. Legendre (dir.), *Enseigner et évaluer. Regards sur les enjeux éthiques et sociopolitiques* (p. 211 −245). Presses de l'Université Laval.

Maulini, O. (2004). Faire le programme ou faire son métier ? Planifier le travail : une compétence stratégique pour les (jeunes) enseignants. *Résonances*, 1, 6-8.

Mottier Lopez, L. & Laveault, D. (2008). L'évaluation des apprentissages en contexte scolaire : développements, enjeux et controverses. *Mesure et évaluation en éducation*, 31(*3*), 5–34

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.

Tochon, F. V. (2013). Planification ouverte de l'enseignement dans une approche profonde de l'apprentissage. In M. De Kesel, M. Bouhon, J.-L. Dufays & J.Plumat (Dir.) *La planification des apprentissages. Comment les enseignants préparent-ils leurs cours ?* (pp. 31-62). Presses universitaires de Louvain.

Tochon, F. V. (2013). Planification ouverte de l'enseignement dans une approche profonde de l'apprentissage. In M. De Kesel, M. Bouhon, J.-L. Dufays & J.Plumat (Dir.) *La planification des apprentissages. Comment les enseignants préparent-ils leurs cours ?* (pp. 31-62). Louvain : Presses universitaires de Louvain.

Tomlinson, C. A. et McTighe, J. (2010). *Intégrer la différenciation pédagogique et la planification à rebours.* Chenelière éducation.

Sallin, L. (2020). *L'évaluation des apprentissages, une pratique qui se planifie ? Regard sur les pratiques des enseignants fribourgeois de français du cycle 3.* Mémoire de Master (non publié), Université de Fribourg

Stiggins, R. (2005). From formative assessment to assessment for learning A path to success in standards-based schools. *Phi Delta Kappan*, *87*, 324-328.

Wanlin, P. (2009). La pensée des enseignants lors de la planification de leur enseignement. *Revue française de pédagogie.* (166), 89-128.

Wiggins, G. & McThighe, J. (2005). *Understanding by design*. Alexandria (USA): Association for Supervision and Curriculum Development.

Wiliam, D. (2011). What Is Assessment for Learning? *Studies in Educational Evaluation, 37*, 3-14.

Yerly, G., Coen, P.-F & Monnard, I. (2019). Evaluer en dépassant l'arithmétique des scores chiffrés. *Educateur*, *Les Dossiers, 16*, 14-16.

Yerly, G. & Berger, J.-L. (2022). Des pratiques d'évaluation constructives à l'Université ? Sous quelles conditions un examen peut-il soutenir les processus d'autorégulation de l'apprentissage. *La Revue LEeE*, *6*.

Zahorik, J. A. (1970). *The effect of planning on teaching*. The Elementary School Journal, 71, 143-151.

À quelles conditions développer des pratiques de notation au service des décisions des enseignants et des apprentissages des élèves ?

RAPHAËL PASQUINI

HAUTE ÉCOLE PÉDAGOGIQUE DU CANTON DE VAUD (SUISSE)

Introduction

L'étude « *Teaching and Learning International Survey* » (TALIS) de 2018 a montré qu'une immense majorité des élèves qui fréquentent les écoles des pays dits industrialisés voient leurs apprentissages essentiellement évalués au moyen d'épreuves sommatives[75] notées, élaborées par l'enseignant, passées en classe ou données en devoirs à la maison, sous une forme traditionnelle « papier-crayon » (Quittre *et al*., 2019). En fonction du système et des degrés scolaires, le moment de l'introduction de la note varie et ses modes de structuration diffèrent : les échelles utilisées, la fixation des seuils de réussite ou le nombre de notes qualifiant l'échec peuvent changer. Toutefois, malgré ces spécificités contextuelles, la note est généralisée dans tous les systèmes scolaires (*Organisation for Economic Cooperation and Development* [OECD], 2012). Il devient dès lors raisonnable d'admettre, qu'à certaines conditions, elle représente l'information la plus concrète que les élèves peuvent obtenir sur les bilans de leurs apprentissages (Guskey & Link, 2019).

La question séculaire visant à se positionner « pour » ou « contre » la note est donc pour nous surannée et peu en lien avec la réalité des systèmes éducatifs. À ce titre, rappelons l'inefficacité de toutes les réformes mises en œuvre depuis vingt ans en francophonie à large échelle pour remplacer la note par d'autres codes[76]. Des recherches ont en effet mis en évidence qu'avec des appréciations lettrées (par exemple, « atteint »), avec des codes de couleurs ou d'autres symboles, les enseignants fonctionnent comme avec des pourcentages ou des notes chiffrées (Kohn, 2011). L'universalité de la note tiendrait notamment au fait que le corps enseignant entretient avec cette dernière une relation ambigüe : d'un côté, imposée par l'institution-École, elle est le principal vecteur par lequel il lui est demandé de « rendre compte » de son travail, de documenter l'avancée dans les programmes et de pronostiquer la suite du parcours de chaque élève. De l'autre, la note, au quotidien, prend une place conséquente dans le temps scolaire au détriment parfois de l'apprentissage : nombre d'enseignants disent ne faire que noter et s'en plaignent (Brookhart, 2013 ; Dubus, 2006 ; Kitabgi, 2009). Cette « course d'obstacles » que la note peut parfois provoquer a des effets sur le stress des élèves, sur leur

[75] L'évaluation sommative intervient à la fin de l'étude d'un sujet ou d'une séquence d'enseignement en dressant un bilan des apprentissages réalisés, sa fonction première étant de les certifier ou de les attester. Elle fait l'objet de l'attribution d'une note le plus souvent chiffrée. Elle est parfois synonyme d'évaluation certificative.

[76] Nous pensons par exemple à la réforme ministérielle au Québec en 2011, à la réforme École Vaudoise en Mutation dans le canton de Vaud à la fin des années 1990 ou à la décision du canton de Zürich de réintroduire les notes à la rentrée 2022.

motivation et sur leur rapport au savoir, et elle impacte également les parents qui n'ont comme balises de la réussite de leur enfant que ses notes et ses moyennes.

Tout en considérant les effets néfastes que la note peut provoquer et la logique de sélection qu'elle induit, et prenant acte que les pratiques de notation relèvent souvent d'habitudes peu questionnées (Brookhart *et al*., 2016 ; Moss, 2013), nos travaux étudient les conditions nécessaires pour que la note se réfère aux apprentissages enseignés, réalisés et évalués, afin qu'elle puisse informer l'enseignant et l'élève de ce qui a été réussi et de ce qui doit encore être travaillé. L'enjeu est de pouvoir offrir des possibilités de rétroactions aux élèves afin de soutenir leurs apprentissages à l'aide de notes à haute valeur informative. En cela, nous parlons de note constructive (Pasquini, 2021).

I. Notre démarche de formation pour favoriser le développement de pratiques de notation au service des apprentissages des élèves

Nos travaux et nos actions de formation s'appuient sur un nombre conséquent d'études réalisées dans la recherche anglophone, qui cherche depuis plusieurs décennies à mieux comprendre les conditions nécessaires au développement de pratiques de notation constructives (Brookhart, 2022 ; Guskey & Brookhart, 2019). Nous en déclinons ici les principaux apports.

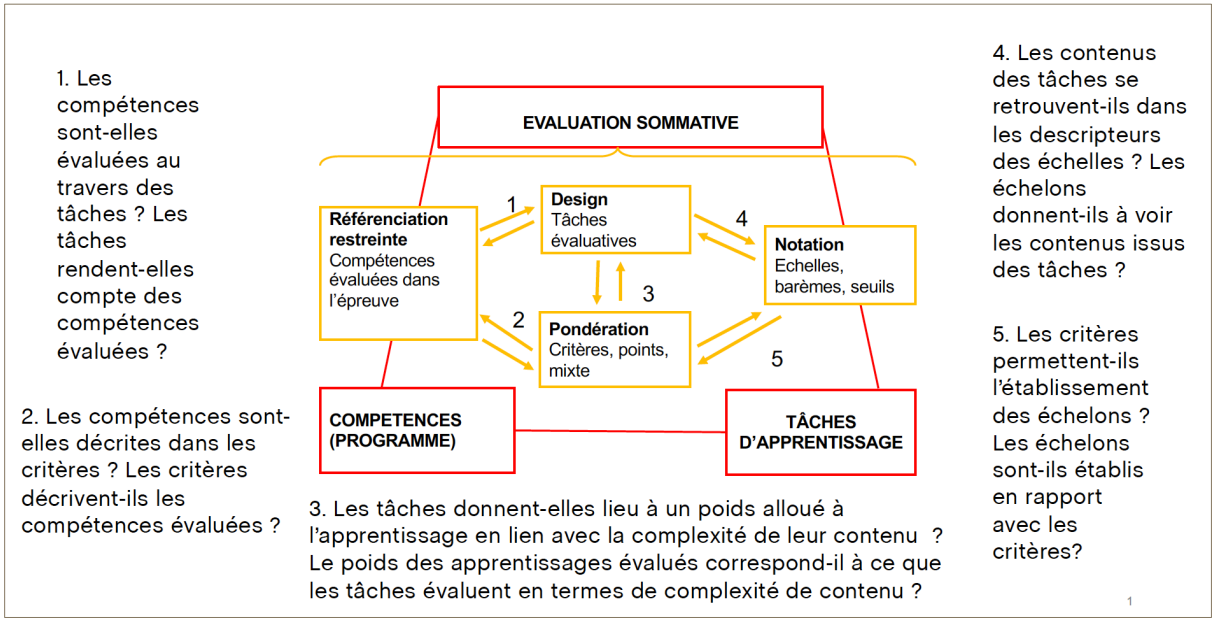
A. L'alignement curriculaire et son acception élargie

Anderson (2002) définit **l'alignement curriculaire comme étant « le niveau de cohérence entre les compétences du programme et l'évaluation, entre les compétences du programme et les activités d'apprentissage, et enfin entre l'évaluation et les activités d'apprentissage »** (p. 257, traduction libre). Les modalités d'enseignement ainsi que les opportunités d'apprentissage offertes aux élèves au travers de tâches d'apprentissage font également partie du modèle. Lorsque c'est le cas, nous parlons **d'alignement constructif** (Biggs, 2003). La notion d'alignement relève d'une conceptualisation de la cohérence à l'aune d'une taxonomie d'habiletés cognitives appariées à des contenus (Anderson & Krathwohl, 2001)[77]. La dimension curriculaire renvoie au fait que cette cohérence est pensée depuis un programme. L'élaboration d'épreuves évaluatives sommatives notées est ici conçue comme un processus complexe, dynamique, contextuel, centré sur les apprentissages (par ex., Biggs, 1999, 2003 ; Gauthier *et al*., 2005 ; Martone & Sireci, 2009).

Face à certaines incomplétudes du modèle de départ qui ne conceptualise pas suffisamment les composantes spécifiques à l'évaluation sommative que sont la pondération et la notation, nous avons élaboré le modèle de **l'alignement curriculaire élargi** en nous basant sur la littérature et en observant les pratiques effectives (Figure 1- Pasquini, 2018, 2019a, 2020, 2021). C'est à l'aide de ce modèle que nous travaillons avec les enseignants depuis quatre ans, en formation initiale et continue, au primaire, au secondaire et au post-obligatoire, dans toutes les disciplines.

[77] Cette taxonomie permet de formuler – ou d'analyser – des objectifs d'apprentissage (ou compétences) en croisant six niveaux d'habiletés cognitives croissant en complexité (se souvenir, comprendre, appliquer, analyser, évaluer et créer) avec quatre contenus de nature différente (les faits, les concepts, les procédures et les démarches métacognitives).

Figure 1 : L'alignement curriculaire élargi et les questions qu'il suscite (Pasquini, 2018, 2019a)



Le modèle ainsi conceptualisé s'inscrit au cœur du modèle premier où interagissent en cohérence les tâches d'apprentissage, les compétences issues des programmes et les pratiques d'évaluation, que nous centrons volontairement sur l'évaluation sommative (traits rouges dans la Figure 1). Cette dernière déploie ensuite ses quatre composantes (accolade en jaune) en les mettant en liens réciproques (doubles flèches en jaune). Ainsi, la référenciation restreinte décrit les choix réalisés dans les compétences à évaluer d'un point de vue sommatif. Le *design* (Wiggins & McTighe, 2005) explicite les processus de construction des tâches évaluatives, du point de vue des contenus sur lesquels elles portent et des habiletés cognitives que les élèves doivent mobiliser. La pondération renvoie au poids alloué aux apprentissages et aux outils exploités pour ce faire : des critères, des points, ou des critères appariés à des points comme modalité mixte (Andrade, 2005 ; Jönsson, 2014). Enfin, la notation cible les démarches de construction de la note chiffrée en lien avec l'apprentissage à l'aide d'outils comme les échelles (Brookhart, 2017 ; Marzano & Heflebower, 2011). Les questions proposées, que nous abordons avec les enseignants, concrétisent les liens de cohérence d'un point de vue taxonomique et de contenu entre chaque composante. Le modèle ainsi élargi – qui désigne le fort lien de cohérence systémique existant entre les compétences évaluées (référenciation restreinte), les tâches évaluatives (*design*), les points et/ou les critères attribués (pondération) et l'échelle (notation) (Pasquini, 2018, 2019a) - permet d'appréhender les conditions concrètes nécessaires à l'élaboration d'une note constructive.

B. La planification à rebours

Pragmatiquement, nous favorisons l'appropriation de ce modèle en travaillant avec les enseignants sur la planification à rebours (Gauthier *et al*., 2005 ; Tomlinson & McTighe, 2010) dont les étapes sont les suivantes :

- déterminer les compétences du programme à atteindre ;
- identifier les critères clés caractérisant l'atteinte de ces compétences ;

- construire des situations et des tâches d'apprentissage alignées avec les compétences et les critères ;
- collecter des informations sur les apprentissages des élèves tout au long de la séquence en lien avec les attentes formulées en pensant aux attentes relatives à la certification.

Ce travail demande de la flexibilité, s'adapte à la discipline et aux contextes ainsi qu'aux publics d'élèves. Il se réalise idéalement en équipe, jusqu'à l'élaboration détaillée de l'épreuve critériée finale qui vient à la fin du processus, en se basant sur la réflexion menée sur l'apprentissage en amont (Pasquini, 2019b).

C. L'approche critériée

Dans un récent article théorique recensant un grand nombre d'études anglophones, Brookhart (2022) montre que les pratiques de notation constructives doivent s'appuyer sur des critères. Étant qualitatifs, les critères représentent l'unique outil permettant de développer des pratiques de notation à haute valeur informative. Dans le cadre des formations que nous menons, nous discutons quatre de leurs caractéristiques incontournables avec les enseignants (Balan et Jönsson, 2018) :

- En se centrant sur les apprentissages des élèves, l'évaluation critériée permet de formuler des rétroactions (ou *feedback*) soutenant l'apprentissage. Il est de fait plus constructif de dire à un élève sur quel critère il a par exemple échoué, plutôt que de lui décompter le nombre de points qui lui ont manqué pour obtenir une note satisfaisante ou supérieure.

- Référés à des compétences, les critères situent chaque élève par rapport à la maîtrise de ces derniers. La comparaison entre élèves est inexistante, ce qui évite de faire jouer à l'évaluation sommative un rôle de comparaison. La certification des apprentissages reprend donc ici le premier rôle, donnant à la note bilan une valeur pédagogique.

- Les critères étant communiqués dès que possible, les élèves sont rapidement et fréquemment confrontés aux contenus et aux apprentissages travaillés et évalués. Ils sont alors petit à petit en mesure de s'approprier une partie de la démarche évaluative. En cela, les critères deviennent des outils pour soutenir les apprentissages des élèves déjà pendant l'enseignement.

- La rétroaction fondée sur des critères permet à l'enseignant de mettre en évidence, à côté de la note chiffrée, ce que les élèves ont réussi, mais surtout ce qu'ils vont devoir encore travailler sous son guidage. La rétroaction est donc ici aussi constructive et qualitative, dans la mesure où elle se centre sur l'apprentissage (« nous allons reprendre la notion de schéma narratif », par exemple en français) et non sur la personne de l'élève (« tu aurais pu te donner un peu plus de peine »), ni sur un total de points (« avec deux points de plus, tu aurais eu 12 »), ces deux derniers commentaires n'ayant aucun impact sur les apprentissages des élèves (De Ketele & Freres, 2009).

En s'appuyant sur des critères, nous observons que l'enseignant, pour établir des seuils de notes, change son questionnement, qui passe de « à quel pourcentage de points met-on le seuil de suffisance ? » à « quelles preuves d'apprentissage l'élève doit-il fournir dans cette situation pour obtenir telle note, synonyme d'apprentissages en construction ou réalisés ? ». Dans une telle

perspective, les points peuvent toujours être exploités, mais ils s'apparient à des critères qui renvoient aux apprentissages évalués (Jönsson, 2014).

D. Une critique des échelles de notation arithmétiques et des procédures de mesure

Comme le dit McMillan (2019), « les pratiques de notation font référence à la manière dont les enseignants utilisent les informations provenant des épreuves sommatives pour déterminer et communiquer les notes des élèves » (traduction libre, p. 85). En cela, toute pratique de notation constructive est qualitative puisqu'elle cherche à identifier des indices de réussite. Dans cette perspective, nous amenons les enseignants à construire des barèmes de notation en référence à des critères, adaptés aux situations d'enseignement-apprentissage-évaluation, et qui tiennent compte des spécificités de la discipline, du temps d'enseignement et d'apprentissage effectif, du public d'élèves, de la complexité du contenu et de l'épreuve. En parallèle, le fait de partager les travaux de DeLandsheere (1992) avec les enseignants leur permet de devenir progressivement critiques face aux échelles numériques basées uniquement sur des points, car elles ont quatre faiblesses qui empêchent l'élaboration de notes constructives :

1. *L'échelle n'a ni zéro ni maximum naturels ou absolus.* Elle commence et finit n'importe où, au choix de celui ou celle qui l'utilise.
2. *Les degrés, ou échelons, ne sont pas de la même grandeur à l'intérieur d'une même échelle.* D'un point de vue d'apprentissage, il est effectivement impossible de démontrer que l'écart séparant la bonne de la très bonne connaissance d'un sujet est le même que l'écart séparant la connaissance moyenne de la bonne.
3. *Des degrés ou échelons correspondant d'une échelle à l'autre pour une même discipline ne sont pas de même grandeur.* Par exemple, malgré l'usage d'échelles arithmétiques, « bon » en expression orale en français pour une enseignante ou un enseignant n'est pas quantitativement égal à « bon » en expression orale en français pour un autre.
4. Des échelles portant sur des disciplines différentes ne sont pas comparables.

Guskey (2013) met également en évidence quelques problèmes des échelles élaborées uniquement sur la base de points, que nous mettons également en discussion (p. 70-72) :

- Par convention, les enseignants qui utilisent de telles échelles posent le plus souvent le seuil de suffisance autour de 60 % à 65 % du total des points. Le résultat est une échelle qui identifie potentiellement deux tiers des points pour décrire l'échec et seulement un tiers pour la réussite. Quel message donne-t-on aux élèves avec de tels outils ? Est-ce à dire qu'il est plus facile de différencier l'échec que la réussite ?
- Dans cette perspective, plus le nombre de points est important dans l'épreuve, plus le phénomène est ample. Comme DeLandsheere, l'auteur observe que de nombreux enseignants n'utilisent pas toutes les notes insuffisantes, ni celles qui décrivent l'excellence. Pourquoi alors en avoir autant ?
- Une croyance très répandue est que plus il y a de points à disposition pour calculer la note, plus cette dernière est précise et objective. Pourtant, dans une perspective qualitative, il est

impossible de différencier l'accumulation de 45 et de 50 points. Augmenter la gradation d'une échelle en pourcentages est donc un leurre de précision.

- Les échelles en pourcentages ne fonctionnent que sur l'exactitude des calculs des seuils des notes, sans rapport avec la maîtrise avérée des apprentissages des élèves dans les tâches évaluatives. Établir des seuils de manière appropriée doit plutôt combiner le jugement professionnel des enseignants en référence aux contenus abordés et aux habiletés cognitives requises par les tâches évaluatives, qui devraient être alignées avec les compétences *via* des critères (Tierney *et al*., 2011).

Tous ces éléments sont abordés progressivement en formation, et participent à l'appropriation de la logique d'une évaluation soutien d'apprentissage et d'une note constructive.

II. Les enjeux que nous relevons avec les enseignants

Lorsque nous accompagnons des enseignants, nous abordons trois enjeux majeurs relatifs au développement de pratiques de notation constructives.

A. L'évaluation sommative et le soutien des apprentissages des élèves

Nous observons que l'évaluation notée est parfois exploitée à des fins non pédagogiques (Hadji, 2012) : maintenir les élèves dans le travail, les motiver à apprendre, gérer la discipline, asseoir le pouvoir de l'enseignant, ou encore récompenser (un effort par exemple). Elle est dès lors souvent au service d'une entreprise qui la dépasse. Or, en l'intégrant dans une réflexion systémique centrée sur les processus d'enseignement-apprentissage, nous soulignons que l'évaluation sommative notée participe à faire progresser chaque élève (Harlen 2005, 2012 ; Laveault et Allal, 2016 ; Mottier Lopez, 2015).

B. Des pratiques de notation en liens avec les objectifs pédagogiques

De nombreux travaux montrent que plusieurs éléments interviennent dans les pratiques de notation : la culture de l'établissement (Duru-Bellat, 2010 ; Kelly, 2008), le niveau des élèves (Guskey, 2009), leur milieu socio-économique (Hattie, 2008 ; Klapp Lekholm, 2011), le degré de sévérité de l'évaluateur (Kitabgi, 2009), ses valeurs (Sun & Cheng, 2014), ses émotions (Brackett *et al*., 2013 ; Tierney, 2015), son niveau de connaissances didactiques, pédagogiques et évaluatives (Xu & Brown, 2016) ou encore sa capacité de comprendre et d'interpréter la logique des prescriptions auxquelles il doit se soumettre (Tierney *et al.*, 2011) sont autant de freins potentiels à une construction de notes constructives. Par ailleurs, nous avons mis en évidence que les prises de décisions des enseignants lorsqu'ils notent font l'objet de multiples négociations entre différents facteurs aboutissant régulièrement à des dilemmes (Pasquini et DeLuca, 2021). Les enseignants reconnaissant ces biais, nous soulignons toujours avec eux que les notes ne devraient poursuivre que des objectifs pédagogiques : (1) certifier les apprentissages considérés comme importants dans les programmes et les établissements, (2) offrir des possibilités des rétroactions formatives aux élèves, et (3) donner des informations aux enseignants pour leur permettre de planifier les apprentissages à poursuivre (Carey & Carifio, 2012).

C. Les barèmes et les échelles : des outils à prendre au sérieux

À ce niveau, nous engageons toujours avec les enseignants une démarche compréhensive sur les questions de l'élaboration des notes et des différentes techniques de manipulation, agrégation, conservation qu'ils leur font subir, ainsi que des genres de décisions auxquelles ces techniques conduisent (Dubus, 2006). Plus spécifiquement, nous accompagnons les enseignants pour qu'ils comprennent mieux sur quelles bases ils décident de la valeur d'une note – et plus particulièrement celle de la suffisance, le 10, qui rend compte du seuil de réussite, le plus important (Dwyer, 1996) – et comment se déploie l'ordonnancement des notes sur les échelles.

Ces réflexions nous permettent également de remettre en question la construction de notes normatives, où l'élève est classé à partir de ses résultats par comparaison avec ceux de ses camarades. Nous soulignons le caractère discriminant de telles pratiques qui aboutissent à la mise en évidence des différences individuelles (Shepard *et al.,* 2018), poursuivent une visée de sélection et de compétition (par. ex., Brookhart, 2005 ; Mottier Lopez, 2015) et produisent toujours de l'échec (Hadji, 2016). Concrètement, l'évaluation normative se traduit par la création d'une courbe de Gauss ou « en cloche ». Elle est très présente en France selon Antibi (2003), qui l'a thématisée sous le nom de « constante macabre ».

III. Que nous apprennent les réflexions du corps enseignant aux prises avec la mise en œuvre d'une note constructive ?

A. Nuancer la qualité est parfois délicat

Tous degrés et disciplines confondus, de nombreux enseignants nous partagent leur difficulté à qualifier graduellement la certification d'une compétence, ne sachant pas par exemple comment expliciter la différence entre une excellente et une bonne performance. En effet, plus l'échelle de notation est ample, plus la définition de chaque note en lien avec des critères est délicate (Guskey & Brookhart, 2019). Pour cela, nous les encourageons à définir les attentes au travers de critères pour signifier au moins la valeur de la note suffisante symbolisant le seuil entre réussite et échec.

B. Élaborer des critères ne va pas de soi

Dans de multiples situations d'évaluation, les enseignants que nous rencontrons travaillent et évaluent sans critère, ce qui tendrait à confirmer plusieurs résultats de recherche (Nordrum *et al*., 2013). Cela est moins dû au fait que les enseignants ne les connaissent pas qu'à leur difficulté à expliciter concrètement leurs attentes. Ce travail de formulation explicite est souvent perçu comme difficile (Wyatt-Smith & Adie, 2021) et chronophage. Mais la démarche induite par la planification à rebours est ici aidante, même si elle est souvent jugée complexe car elle oblige à penser aux apprentissages puis aux tâches en rapport, alors que les habitudes vont souvent dans l'autre sens.

C. Tenir compte des spécificités de sa discipline est fondamental

Nous observons que le contenu disciplinaire joue un rôle majeur dans la possibilité, ou non, de construire des notes à haute valeur informative basées sur des critères (Duncan & Noonan, 2007 ; Meier *et al*., 2006 ; Prøitz, 2013). Par exemple, au primaire et au secondaire, les disciplines des langues semblent mieux se prêter à l'exercice, notamment en compréhension et en production de l'écrit. Cela est probablement dû au fait qu'il est plus aisé de faire la différence entre une performance juste

suffisante ou bonne d'une compréhension d'œuvre littéraire en italien par exemple, mais que le même exercice semble plus délicat lorsqu'il s'agit de qualifier la bonne ou très bonne utilisation du cercle chromatique en arts visuels. Cela étant, nos dernières recherches montrent que des enseignants parviennent à critérier tout de même des notes dans des disciplines comme les mathématiques (Mrazek & Morales Villabona, 2022).

Conclusion : Deux éléments pour penser la formation

A. La réception de la formation par les enseignants

Les formations continues que nous dispensons sont toutes évaluées selon une démarche spécifique : les enseignants sont questionnés en ligne sur les apports et limites de la formation à la fin du dispositif, puis une seconde fois quelques mois plus tard, afin de voir dans quelle mesure les compétences travaillées sont mises en œuvre. Aussi, tous degrés confondus, nous observons que les enseignants avec lesquels nous collaborons soulignent la pertinence des apports proposés. Ils relèvent toutefois que la réflexion exigée est complexe, demande un temps d'appropriation et d'expérimentation conséquent et nécessite une coordination en équipe. C'est la raison pour laquelle nous privilégions des dispositifs collaboratifs en écoles, centrés sur les questions et problèmes relevés par les enseignants en fonction de leurs contextes. Nous travaillons depuis les pratiques réelles des participants, de manière à tenir compte de leurs élèves, de leurs contraintes et de leurs prescriptions. Cette méthode de travail nous permet d'accompagner chaque année la mise en œuvre de démarches dans plusieurs classes du primaire et du secondaire, qui montrent toutes que l'on peut conjuguer cohérence, transparence et exigence (par exemple, voir Pasquini, 2020).

B. Le changement de posture des enseignants

Notre approche permet le plus souvent aux enseignants de modifier leurs conceptions de l'évaluation et de la notation, de prendre conscience des phénomènes de cohérence à ce niveau. Globalement, ils s'approprient plus profondément le programme, hiérarchisent les compétences à enseigner et à évaluer, élaborent des tâches évaluatives complexes, critérient en référence aux compétences évaluées et concentrent leur enseignement sur les contenus qu'ils jugent importants pour les progrès des élèves (Pasquini, 2018 ; Pasquini & Morales Villabona, 2022a). À ce titre, nous pensons que leur identité d'évaluateur (Looney *et al*. 2017) ainsi que leurs connaissances et compétences en évaluation (Xu & Brown, 2016) évoluent positivement (Pasquini & Morales Villabona, 2022b). Des tensions restent toutefois toujours présentes dans leurs pratiques de notation, notamment dans leur difficulté à mettre en œuvre des démarches qualitatives dans des environnements et des équipes aux pratiques traditionnelles, centrées sur les notes strictes et les moyennes (Mottier Lopez *et al*., 2012).

Références

Anderson, L. W. (2002). Curricular alignment : A re-examination. *Theory into Practice*, 41(4), 255-260. https://doi.org/10.1207/s15430421tip4104_9

Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53*(1), 27-31.

Antibi, A. (2003). *La constante macabre ou comment a-t-on découragé des générations d'élèves.* Math'Adore.

Balan, A. et Jönsson, A. (2018). Increased explicitness of assessment criteria: effects on student motivation and performance. *Frontiers in Education, 3*(81), 56-68.

Brackett, M. A., Floman, J. L., Ashton-James, C., De, L. & Salovey, P. (2013). The influence of teacher emotion on grading practices: a preliminary look at the evaluation of student writing. *Teachers and Teaching*, 19(6), 634-646. https ://doi.org/10.1080/13540602.2013.827453

Brookhart, S. M. (2013). Grading. Dans J. McMillan (dir.), *SAGE handbook of research on classroom assessment* (p. 257-271). SAGE Publications.

Brookhart, S. M. (2017). *How to use grading to improve learning*. ASCD.

Brookhart, S. M. (2022). Connecting grading with other assessment practices. *La Revue LEeE (Assessment in Education Journal)*, (6). https://doi.org/10.48325/rleee.006.01

Butera, F., Buchs, C. et Darnon, C. (2011). *L'évaluation, une menace ?* PUF.

Carey, T. et Carifio, J. (2012). The minimum grading controversy: Results of a quantitative study of seven years of grading from an urban high school. *Educational Researcher*, 41(6), 201-208. https://doi.org/10.3102/0013189X12453309

Crahay, M. (2007). *Peut-on lutter contre l'échec scolaire* ? De Boeck.

De Ketele, J.-M. et Freres, G. (2009). Les commentaires des enseignants et des élèves : simples jugements ou processus évaluatifs ? *Les Cahiers de Recherche en Éducation et Formation*, 74, 4-32.

Dubus, A. (2006). *La notation des élèves. Comment utiliser la docimologie pour une évaluation raisonnée*. Armand Colin.

Duncan, R. C. et Noonan, B. (2007). Factors affecting teachers'grading and assessment practices. *The Alberta Journal of Educational Research*, 53(1), 1-21.

Duru-Bellat, M. (2010). Ce que la mixité fait aux élèves. *Revue de l'OFCE*, 3 (114), 197-212. https://doi.org/10.3917/reof.114.0197

Dwyer, C. A. (1996). Cut scores and testing: statistics, judgment, truth, and error. *Psychological Assessment*, 8(4), 360-362. https://doi.org/10.1037/1040-3590.8.4.360

Gauthier, C., Mellouki, M., Bissonnette, S. et Richard, M. (2005). *Écoles efficaces et réussite scolaire des élèves à risque. Un état de la recherche.* (Rapport de recherche du Centre de recherche interuniversitaire sur la formation et la profession enseignante). CRIFPE, Université Laval.

Guskey, T. R. (2009, avril). Bound by tradition: Teachers' views of crucial grading and reporting issues [communication orale]. *American Educational Research Association*, San Francisco, Etats-Unis.

Guskey, T. R. (2013). The case against percentage grades. *Educational, School, and Counseling Psychology Faculty Publications*, 22. https ://uknowledge.uky.edu/edp_facpub/22

Guskey, T. R. et Brookhart, S. M. (Ed.) (2019). *What we know about grading*. ASCD.

Guskey, T. R. et Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice*, 26(3), 303-320. https://doi.org/10.1080/0969594X.2018.1555515

Hadji, C. (2012). *Faut-il avoir peur de l'évaluation* ? De Boeck.

Hadji, C. (2016). Le système éducatif peut-il passer de l'évaluation normative à une évaluation constructive ? Dans B. Bocquet (dir.), *La fièvre de l'évaluation quels symptômes, quels traitements* ? (p. 113-135). Presses Universitaires du Septentrion.

Harlen, W. (2005). Teacher's summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal, 16*(2), 207-223. https://doi. org/10.1080/09585170500136093

Harlen, W. (2012). On the relationship between assessment or formative and summative purposes. Dans J. Gardner (dir.), *Assessment and learning* (p. 87–102). Sage.

Hattie, J. (2008). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Jönsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment et Evaluation in Higher Education*, 39(7), 840-852. https://doi.org/10.1080/02602938.2013.875117

Kelly, S. (2008). What types of students' efforts are rewarded with high marks? *Sociology of Education*, 81(1), 32-52. https://doi.org/10.1177/003804070808100102

Kitabgi, S. (2009). *L'évaluation scolaire est-elle au service de l'orientation ?* Centre d'orientation de CCIP. https://jmb21470.pagesperso-orange.fr/mysegpa21/documents/Etude-evaluation-scolaire-orientation-2009.pdf

Klapp Lekholm, A. (2011). Effects of school characteristics on grades in compulsory school. *Scandinavian Journal of Educational Research*, 55 (6), 587–608. https://doi.org/10.1080/00313831.2011.555923

Kohn, A. (2011). The case against grades. *Effective Grading Practices*, 69(3), 28-33.

Laveault, D. et Allal, L. (Eds.). (2016). *Assessment for learning: Meeting the challenge of implementation*. Springer International.

Looney, A., Cumming, J., van Der Kleij, F. et Harris, K. (2017): Reconceptualising the role of teachers as assessors : teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442-467. https://doi.org/10.1080/0969594X.2016.1268090

Martone, A. et Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research, 79*(4), 1332-1361. https://doi.org/10.3102/0034654309341375

Marzano, R. J. et Heflebower, T. (2011). Grades that show what students know. *Effective Grading Practices, 69*(3), 34-39.

McMillan, J. H. (2019). Surveys of teachers' grading practices and perceptions. Dans T. R. Guskey et S. M. Brookhart (dir.), *What we know about grading* (p. 84-112). ASCD.

Meier, S. L., Rich, B. S., Cady, J. (2006). Teachers' use of rubrics to score non-traditional tasks: factors related to discrepancies in scoring. *Assessment in Education: Principles, Policy & Practice*, 13(1), 69-95. https://doi.org/10.1080/09695940600563512

Merle, P. (2012). L'évaluation par les notes : quelle fiabilité et quelles réformes ? *Regards croisés sur l'économie*, 2(12), 218-230. https://doi.org/10.3917/rce.012.0218

Moss, C. (2013). Research on classroom summative assessment. Dans J. McMillan (dir.), *SAGE handbook of research on classroom assessment* (p. 235-256). SAGE Publications.

Mottier Lopez, L., Tessaro, W., Dechamboux, L. et Morales Villabona, F. (2012). La modération sociale : un dispositif soutenant l'émergence de savoirs négociés sur l'évaluation certificative des apprentissages des élèves. *Questions Vives*, 6(18). https://doi.org/10.4000/questionsvives.1235

Mottier Lopez, L. (2015). *Évaluations formative et certificative des apprentissages : enjeux pour l'enseignement*. De Boeck.

Mrazek, J. et Morales Villabona, F. (2022, prépublication). Évaluation critériée et résolution de problèmes numériques. Un récit de pratique accompagné de pistes de réflexion sur l'évaluation constructive. *La Revue LEeE,6.* http://revue.leee.online/index.php/info/article/view/137

Nordrum, L., Evans, K. et Gustafsson, M. (2013). Comparing student learning experiences of in-text commentary and rubric-articulated feedback: strategies for formative assessment. *Assessment and Evaluation in Higher Education*, 38, 919-940. https ://doi.org/10.1080/02602938.2012.758229

Organisation for Economic Cooperation and Development (OECD) (2012). *Grade Expectations: How marks and education policies shape student's ambitions*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264187528-en

Pasquini, R. (2019a). Élargir conceptuellement le mode le 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.

Pasquini, R. (2019b). Construire des évaluations sommatives communes : une mauvaise « bonne idée » ? *Éducateur*, 6, 4-5.

Pasquini, R. (2021). *Quand la note devient constructive. Évaluer pour certifier et soutenir les apprentissages*. Presses de l'Université Laval

Pasquini, R. et Deluca, C. (2021). Grading in a dilemmatic space: An exploratory cross-cultural analysis of Mathematics and Language Secondary teachers. *Comparative and International Education/Éducation comparée et internationale*. 49(2), 51-70. https://doi.org/10.5206/cieeci.v50i1.14133

Pasquini, R. et Morales Villabona, F. (2022a). Comment appréhender une littératie en évaluation pour documenter et comprendre le développement des compétences des enseignants ? *Revue canadienne de l'éducation*, *45*(2). 512-544. https://cje-rce.ca/wp-content/uploads/sites/2/2021/03/45-2-9-5171-Pasquini-512-554.pdf

Pasquini, R. et Morales Villabona, F. (2022b, 22 avril). *When Assessment for Learning Meets Teachers' Assessment Culture: Issues and Prospects* [communication orale dans un symposium]. Colloque de l'AERA 2022, San Diego, Etats-Unis.

Prøitz, T. S. (2013). Variations in grading practice – subjects matter. *Education Inquiry*, 4(3), 555-575. https://doi.org/10.3402/edui.v4i3.22629

Quittre, V., Dupont, V., et Lafontaine, D. (2019). Enseigner au quotidien. Talis 2018. http://www.enseignement.be/index.php?page=28161

Shepard, L. A., Penuel, W. R. et Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21-34. https ://doi.org/10.1111/emip.12189

Sun, Y. et Cheng, L. (2014). Teachers' grading practices: meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326-343. https://doi.org/10.1080/0969594X.2013.768207

Tierney, R. D. (2015). Altered grades: A grey zone in the ethics of classroom assessment. *Assessment Matters*, 8, 5-30. http://dx.doi.org/10.18296/am.0002

Tierney, R. D., Simon, M. et Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards-based grading. *The Educational Forum*, 75(3), 210-227. https://doi.org/10.1080/00131725.2011.577669

Tomlinson, C. A. et McTighe, J. (2010). *Intégrer la différenciation pédagogique et la planification à rebours*. Chenelière Éducation.

Wyatt-Smith, C. et Adie, L. (2021). The development of students' evaluative expertise: enabling conditions for integrating criteria into pedagogic practice. *Journal of Curriculum Studies*, 53(4), 399-419. https://doi.org/10.1080/00220272.2019.1624831

Xu, Y., et Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.

Quels sont les apports de la psychologie cognitive pour articuler évaluation et apprentissage ?

DANIEL GAONAC'H
UNIVERSITÉ DE POITIERS (FRANCE)

Introduction

L'accent mis sur l'intérêt des évaluations formatives repose sur l'idée que le fait d'évaluer doit permettre (et peut permettre) d'améliorer la qualité des apprentissages réalisés par les élèves. Les écrits qui visent à théoriser ce concept (par exemple Allal, 1991) mettent souvent en avant l'idée qu'un tel mode d'évaluation vise à une *régulation* efficiente de l'enseignement et/ou des apprentissages qui en découlent. Ce qui caractérise l'évaluation formative n'est donc pas tant l'absence de notes, de classement, de pénalisation, de publicité institutionnelle (pairs, parents). Ce mode d'évaluation implique surtout que soient fournies à l'enseignant et/ou à l'élève des données sur l'état d'un processus d'apprentissage en cours, et que ces données permettent une prise de décision sur la conduite à tenir, c'est-à-dire sur la mise en œuvre des activités supposées rendre plus efficace l'apprentissage, au regard des objectifs qui ont été fixés.

On peut ainsi retenir deux caractéristiques de cette conception :

- Même si le processus d'apprentissage inclut une séquence d'évaluation, cette séquence ne constitue pas un élément intrinsèque de ce processus ; elle en reste un élément extérieur, dont la pertinence repose sur sa capacité à modifier le déroulé du processus d'apprentissage lui-même.
- Elle semble se référer toujours à une action volontaire sur cette nouvelle séquence, à travers la prise de conscience et l'analyse du résultat de la séquence antérieure.

C'est à propos de ces deux dimensions que nous souhaitons illustrer une vision différente de l'articulation entre évaluation et apprentissage. Nous défendons ici l'idée que l'importance de cette articulation peut être mise en évidence si l'on prend en compte les processus cognitifs qui sont en jeu lorsque les situations d'apprentissage impliquent à la fois des processus de mémorisation et des processus d'évaluation. Il s'agit donc, en s'appuyant sur la connaissance qu'on peut avoir des processus d'apprentissage, d'examiner de quelle manière ces processus généraux peuvent intégrer des processus d'évaluation, et à quelles conditions cette intégration peut s'avérer particulièrement efficace. Pour étayer ce point de vue, on peut se référer notamment à deux exemples :

- **L'effet-test**, qui montre que l'évaluation d'un apprentissage, indépendamment de toute « notation », constitue par elle-même un élément favorisant celui-ci. On examinera quelles hypothèses on peut mettre en avant pour expliquer cet effet, notamment en analysant certaines des conditions de son apparition.

- **Les recherches dans le domaine de la métacognition** qui conduisent à contester l'idée que la prise de conscience des processus en jeu dans les apprentissages soit réellement un facteur favorable à la réalisation de ceux-ci. On peut notamment mettre en avant l'existence et l'efficacité de processus de régulation implicites, à travers lesquels « l'évaluation » de son apprentissage par l'apprenant ne tient plus à la présence d'une séquence explicite le conduisant à en évaluer l'efficacité.

I. L'effet-test : quels sont les processus impliqués ?

Lorsqu'il y a injonction de « revoir » un cours, la stratégie spontanée est souvent de relire. Mais de nombreuses recherches montrent que la simple révision des informations à mémoriser conduit toujours à des performances inférieures, et souvent largement inférieures, à temps égal, à celles qui sont issues d'une pratique plus élaborée : il s'agit de toutes les situations où l'apprenant est conduit à retrouver par lui-même et à produire, sous différentes formes, les informations mémorisées. On désigne cet effet favorable par l'expression « effet-test ».

Cette pratique constitue un grand classique dans des domaines bien déterminés : on pense notamment à la « récitation » de textes, de poésies, souvent demandée par l'enseignant dans la classe, mais aussi pratiquée spontanément par les apprenants. Mais de nombreuses données montrent la pertinence de cette pratique dans de nombreux domaines (pour une revue, voir Roediger et Butler, 2011, et en français dans Brown, Roediger et McDaniel, 2016) : par exemple l'apprentissage d'un vocabulaire en langue étrangère (Karpicke & Roediger, 2008), l'étude d'un cours d'histoire (Carpenter, Pashler & Cepeda, 2009), ou des leçons de sciences (McDaniel *et al*., 2011), et ce à tous les niveaux de la scolarité ou chez l'adulte.

Il est important de relever que cet effet joue y compris lorsque l'apprenant *n'a pas la possibilité de vérifier si les réponses qu'il donne sont correctes ou pas*. Certes l'effet est alors amoindri par rapport à la situation dans laquelle il y a vérification de cette exactitude (69 % de réponses correctes au test final, contre 51 % : Butler, Karpicke & Roediger, 2007), mais il existe bien par rapport à la situation contrôle sans révision sous forme de test (31 %). Si donc il existe l'effet d'une évaluation explicite, qui peut être à l'origine d'une régulation relevant d'un processus métacognitif, l'effet-test ne peut se réduire à une telle régulation. De notre point de vue, il s'agit d'un effet direct sur les processus d'apprentissage mis en œuvre dans ces situations de test. Pour comprendre la nature de ces processus, il faut s'intéresser de manière plus précise à certains aspects des conditions de réalisation de ces tests : il s'agit de **l'optimisation du délai entre apprentissage et test**, et de **la définition des contenus d'apprentissage sur lesquels porte le test**.

On relève notamment que l'effet est plus marqué si on introduit un délai entre l'étude et le test (Butler, Karpicke & Roediger, 2007) : un test immédiat n'a que très peu d'effet, et la nécessité d'un délai suffisamment long apparaît surtout lorsque la mémorisation finale est évaluée à long terme (rappel final plusieurs semaines ou plusieurs mois après les exercices). **Le délai optimal entre apprentissage et test peut dépendre de l'objet d'apprentissage**. Dans le cas d'un vocabulaire en langue étrangère, Pyc et Rawson (2010) montrent qu'un délai de 6 minutes induit des performances meilleures qu'un délai d'une minute. Mais sur des connaissances plus complexes, on montre qu'un délai en jours voire en semaines est plus efficace. Dans le cas du cours d'histoire évoqué plus haut, les performances finales sont meilleures lorsque la révision sous forme de test est réalisée 16 semaines après la leçon initiale. Dans le cas des leçons de sciences évoquées plus haut, l'amélioration liée à l'effet-test est la

plus forte si on ménage un délai d'une vingtaine de jours entre la leçon et le test, et ce même s'il s'agit d'un test unique, qui joue tout autant que 3 tests menés durant cette période. De manière générale, la répétition d'un test n'a que peu d'effet : **un seul test peut suffire à améliorer les performances, à condition qu'il soit réalisé au bon moment** (Landauer & Bjork, 1978). De plus, le caractère favorable d'un tel délai vaut y compris lorsque l'apprenant dispose d'un feedback : contrairement à ce qu'on peut penser intuitivement, le feedback le plus efficace n'est pas celui qui intervient aussitôt après que l'élève a donné sa réponse ; **il est préférable de faire un retour en fin de test**, c'est-à-dire sur l'ensemble des connaissances qui font l'objet d'un apprentissage, et non pas de manière ponctuelle sur chaque élément de ces connaissances (Mullet *et al.*, 2014).

C'est en prenant en compte ces processus différés qu'on peut comprendre en quoi des facteurs comme l'introduction d'un délai ou la prise en compte d'un ensemble d'informations reliées, peuvent constituer des conditions favorables à l'établissement de l'effet-test. **Celui-ci ne peut être alors considéré comme un simple « état des lieux » d'une séquence d'apprentissage, mais constitue en soi un processus qui s'intègre dans les processus d'apprentissage**.

II. L'explicite et l'implicite dans la régulation des apprentissages

Nous sommes partis du constat que les bénéfices d'une évaluation formative sont souvent de fait attribués à une composante métacognitive, c'est-à-dire à la conscience que l'apprenant lui-même peut avoir, spontanément ou à travers des indications qui lui sont fournies, du fonctionnement de son propre système cognitif, et dans le cas d'un apprentissage de la manière dont il met en œuvre cet apprentissage. En effet, l'opportunité de contrôler et de réguler l'avancée d'un apprentissage est supposée être liée à la connaissance par l'élève de sa progression dans cet apprentissage et des difficultés rencontrées, à une meilleure représentation par l'élève de la tâche mise en œuvre, du but des exercices proposés, et donc à la possibilité de corriger son action sur la base de cette connaissance et cette représentation.

Si de nombreuses données permettent d'affirmer que les compétences métacognitives d'un individu peuvent avoir des effets bénéfiques sur la réussite d'un apprentissage (par exemple de Boer *et al*., 2018, pour une synthèse), il faut cependant d'abord admettre que ce n'est pas toujours le cas, et la divergence entre croyances affichées et stratégies mises en œuvre constitue un des thèmes de prédilection des recherches sur la métacognition. Pour illustrer cette divergence, on peut revenir un moment sur le dossier de l'effet-test. Alors que les données empiriques montrent de manière claire la solidité de cet effet, il est caractéristique que **les apprenants eux-mêmes sont loin d'être conscients de la supériorité de ce mode de révision.** Quand on interroge des étudiants sur la manière dont il convient d'apprendre un cours, pour une majorité d'entre eux la lecture répétée d'un chapitre est préférable à une alternance lecture-rappel (Karpicke, Butler & Roediger, 2009). Il est frappant de constater que l'absence de conscience du bénéfice tiré d'une phase de test est avérée, y compris si l'on interroge l'apprenant juste après qu'une telle phase ait été mise en œuvre. Dans leur expérience sur l'apprentissage d'un vocabulaire en langue étrangère, Karpicke & Roediger (2008) avaient montré qu'aussitôt après la phase d'apprentissage, une phase de test sur ce qui venait d'être appris était nettement plus efficace qu'une nouvelle présentation du vocabulaire : dans un test effectué une semaine plus tard (rappel final), ils observaient 80 % de rappel correct chez les élèves qui avaient effectué un test contre 35 % chez ceux qui avait bénéficié d'une seconde présentation. Ils avaient demandé aux participants, aussitôt après la phase de test, d'estimer le nombre de mots dont ils pourraient se souvenir par la suite : ils n'avaient alors constaté pour cette estimation aucune

différence entre les deux groupes d'apprenants. Des interviews menées ensuite auprès des participants montraient que ceux qui admettaient l'utilité d'un test de ce qu'ils avaient étudié n'admettaient cette utilité qu'au titre d'un contrôle des effets de leur apprentissage, et non d'une façon de le consolider.

L'effet-test n'est pas le seul effet à propos duquel on peut évoquer une telle cécité. Un autre exemple caractéristique porte sur l'opposition apprentissage massé / apprentissage distribué : la répétition immédiate d'un matériau objet d'apprentissage, plutôt qu'un exercice différé, est le plus souvent considérée comme le moyen le plus efficace d'apprendre (Bjork, Dunlosky & Kornell, 2013). L'explication proposée est que la simple présentation (ou la lecture) répétée d'un même contenu produit chez l'apprenant une « illusion de savoir », car cette répétition accroît la facilité avec laquelle ce contenu est traité au moment de la présentation, quand bien même cette facilité ne constitue en rien un élément favorable à la mémorisation. Il s'agit là d'un phénomène très général : les prédictions faites par l'apprenant lui-même sur l'établissement d'un souvenir en mémoire sont fondées principalement sur la facilité du traitement des contenus à mémoriser. De manière plus générale encore, les apprenants ont souvent des représentations très fausses de ce qui fait qu'un enseignement est efficace pour eux (Uttl, White & Gonzalez, 2017).

Faut-il pour autant considérer que, y compris dans le cas d'une évaluation formative, la mauvaise représentation qu'on peut avoir des processus d'apprentissage réellement en jeu rende peu efficace toute prise de conscience de l'état d'un apprentissage ? On ne peut aborder ce dossier sans prendre en compte une distinction, devenue centrale dans le domaine de la métacognition, entre des **processus explicites** et des **processus implicites**. Veenman, van Hout-Wolters et Afflerbach (2006), par exemple, distinguent :

- Des processus liés à des connaissances métacognitives conscientes. Ces connaissances peuvent jouer un rôle, mais ne garantissent nullement l'efficacité d'un apprentissage, car elles peuvent être correctes ou incorrectes.
- Des processus de régulation, fondés sur des mécanismes de feedback. Ces mécanismes ne sont pas nécessairement conscients, même s'ils peuvent néanmoins déboucher, tout au moins à long terme, sur l'élaboration de connaissances conscientes.

Pour reprendre une distinction fondamentale dans les recherches actuelles sur la structure de la mémoire humaine (voir par exemple dans Gaonac'h, 2022, chapitre 1), on peut considérer que les premiers processus relèvent de connaissances déclaratives (explicites), et les seconds de connaissances procédurales (implicites, mais éventuellement explicitables). Tsalas, Sodian & Paulus (2017) ont opérationnalisé cette dimension explicite/implicite de la métacognition, chez des enfants de 10 ans et chez des adultes, à travers une épreuve d'apprentissage de paires de dessins associés. Certaines de ces paires correspondaient à des associations fréquentes (*chien* – *chat*), d'autres à des associations rares (*chien* – *tomate*). La composante explicite était évaluée à travers un questionnaire standardisé relatif aux stratégies de mémorisation. La composante implicite était évaluée à travers deux indices : une estimation faite par chaque participant et, pour chaque paire, de sa capacité à rappeler à partir du premier dessin celui qui lui était associé ; une mesure de la durée d'étude par chaque participant de chacune des paires de dessins. On testait ensuite, après la réalisation d'une tâche distractrice, le rappel de ces dessins associés. Les données recueillies conduisent à plusieurs constats :

- Il n'y a pas de lien entre la composante explicite et la composante implicite : faire preuve de davantage de connaissances métacognitives n'induit nécessairement ni des estimations ni des durées d'étude adaptées.
- Il n'y a pas de lien entre connaissances explicites et performances de rappel.
- **Le lien est en revanche avéré entre les indices de la composante implicite et ces performances**. Il est plus important chez les adultes que chez les enfants.

Ce profil de données est par ailleurs éclairé par ce qu'on observe à partir de l'évaluation des capacités exécutives des participants, celles qui permettent à l'individu la *gestion* de ses activités cognitives : attention aux éléments importants, maintien des informations pertinentes, effacement de celles qui ne le sont plus, etc. Il s'agit pour ce qui nous concerne ici de compétences de contrôle cognitif qui peuvent être en jeu notamment dans le contrôle volontaire de stratégies de mémorisation (répétition mentale de certaines informations, établissement de liens entre différentes informations, etc.). De manière apparemment paradoxale, c'est chez les enfants, et non chez les adultes, qu'on observe un lien entre capacités de contrôle et performances de rappel, ce qui peut correspondre à une utilisation plus contrôlée des stratégies de mémorisation chez les enfants. Cela ne signifie pas pour autant que ces stratégies contrôlées soient particulièrement efficaces : cela montre surtout que l'enjeu du développement cognitif, plus que sur la connaissance explicite des mécanismes mémoriels, porte sur la mise en œuvre automatique (sans contrôle) de mécanismes de régulation, ce qui relève d'un apprentissage et n'est attesté qu'assez tard chez les adolescents. Le poids plus important du contrôle cognitif chez les enfants ne tient pas à de meilleures capacités de contrôle, mais à la nécessité de compenser la plus faible automatisation de la régulation des processus d'apprentissage.

On sait depuis longtemps que l'acquisition des compétences métacognitives peut être tardive : même chez l'adulte, la mise en œuvre de stratégies efficaces n'est ni spontanée ni générale. Dès ses premières publications sur la métacognition, Flavell (1971) mettait d'ailleurs en valeur, pour rendre compte de l'acquisition de capacités métacognitives, la notion « **d'expériences métacognitives** », c'est-à-dire la confrontation répétée à des situations variées, qui constituent des occasions de prendre en compte de manière fine les situations d'apprentissage et leurs caractéristiques. Les processus de régulation implicites que nous évoquons ici tiendraient notamment à la multiplicité et à la diversité des « expériences » cognitives relatives à un contenu d'apprentissage. L'évaluation de son apprentissage par l'apprenant ne tient plus alors à la présence d'une séquence explicite lors de laquelle il en évaluerait l'efficacité, mais à la confrontation à des situations suffisamment variées dans lesquelles il est amené à mettre en œuvre des contenus d'apprentissage qui peuvent être proches mais utilisés dans des conditions différentes.

Cela revient à dire que l'acquisition de compétences métacognitives prend du temps car elle se réalise au fur et à mesure qu'un individu est confronté à une variété de situations dans lesquelles il est amené à (re)mobiliser des apprentissages. De ce point de vue, l'adulte a plus de compétences métacognitives parce qu'il a été confronté à davantage de situations, alors même que la mise en œuvre de ces compétences n'apparait pas nécessairement de manière explicite lors de la réalisation d'un apprentissage.

On peut ainsi aboutir à la conclusion que **la fréquence, et surtout la variété, des situations d'évaluation, constituent en soi des conditions favorables au développement de compétences de régulation qui ne relèvent pas nécessairement d'une régulation explicite**, liée à la prise de conscience

de l'état d'un apprentissage. Les situations d'évaluation doivent aussi permettre à l'apprenant d'être confronté, à travers différents contenus d'apprentissage, à des contraintes d'apprentissage suffisamment variées pour que les mécanismes implicites, dont on connaît maintenant la puissance, puissent devenir opérationnels. On pourra rapprocher ce point de vue de celui qui a été largement développé par tout un pan de la psychologie cognitive consacré à « l'expertise », dont on a montré, notamment à l'origine à propos du jeu d'échecs, qu'elle pouvait relever fortement de la référence implicite et automatisée à des situations prototypiques surentraînées (Gaonac'h, 2019, chapitre 8).

Cette référence aux dimensions implicites du fonctionnement cognitif conduit à insister sur le fait que la mise en œuvre, dans les apprentissages, de stratégies explicites, y compris sans doute dans le cadre d'une évaluation formative, est loin d'être spontanée et efficace. Si l'on suit le raisonnement qui prévaut dans la mise en exergue de processus implicites de régulation, c'est sans doute à travers la confrontation répétée à des contextes variés d'évaluation que peuvent être construites des compétences de régulation qui peuvent rester implicites sans perdre leur efficacité : c'est alors, **plus que leur explicitation, la multiplicité et la diversité des expériences cognitives relatives à un contenu d'apprentissage qui contribuent à cette efficacité**.

On peut rapprocher ce point de vue de celui défendu notamment par Lieury (1991 ; voir Gaonach, 2022, Chapitre 1, pour une synthèse), qui argumente que la construction des connaissances doit s'appuyer sur la présentation d'« épisodes » variés. Ce point de vue peut notamment être illustré simplement à travers les recherches sur le rôle des exemples dans les apprentissages. Les recherches sur l'utilisation des exemples démontrent bien leur pertinence pédagogique, mais concluent souvent à la nécessité d'introduire, pour un même exposé, plusieurs exemples présentant des caractéristiques différentes. Cet effet est particulièrement marqué pour les débutants ou les élèves en difficulté. Chi *et al*. (1989) ont ainsi analysé la façon dont des étudiants résolvent des problèmes de mécanique lorsqu'ils doivent transférer un entraînement initial dans des situations variées. Ils constatent que les étudiants ayant un faible niveau dans la discipline doivent retourner aux exemples fournis lors de l'entraînement, alors que pour les étudiants de bon niveau, la référence aux exemples pertinents étudiés auparavant est instantanée. Ainsi, la différence entre les étudiants de différents niveaux ne tient pas tant à des capacités de raisonnement qu'à une exploitation automatique d'exemples pertinents, rendue possible par la confrontation répétée à des exemples suffisamment variés.

Conclusion

Nous avons à travers cet exposé cherché à illustrer deux principes qui nous paraissent éclairer en quoi une évaluation formative peut constituer un déterminant important des apprentissages, non pas comme un « contrôle » extérieur au processus d'apprentissage, mais comme une composante inhérente à ce processus. **L'évaluation doit être conçue comme un élément intégré dans le processus d'apprentissage** en conformité avec ce qu'on connaît des processus de consolidation et de récupération des informations en mémoire, et de la manière dont les apprenants peuvent exercer un contrôle sur ces processus.

Références

Allal, L. (1991). Vers une pratique de l'évaluation formative : matériel de formation continue des enseignants. De Boeck.

Bjork, R.A., Dunlosky, J. & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417-444. https://doi.org/10.1146/annurev-psych-113011-143823

Bjork, E.L., Little, J.L. & Storm, B.C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition*, 3(3), 165-170. https://doi.org/10.1016/j.jarmac.2014.03.002

Brown, P.C., Roediger, H.L. & McDaniel, M.A. (2016). Mets-toi ça dans la tête : Les stratégies d'apprentissage à la lumière des sciences cognitives. Markus Haller.

Butler, A.C., Karpicke, J.D., & Roediger III, H.L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13(4), 273-281. https://doi.org/10.1037/1076-898x.13.4.273

Carpenter, S.K., Pashler, H. & Cepeda, N.J. (2009). Using tests to enhance 8th grade students' retention of U.S. History facts. *Applied Cognitive Psychology*, 23(6), 760-771. https://doi.org/10.1002/acp.1507

de Boer, H., Donker, A. S., Kostons, D. D. & van der Werf, G. P. (2018). Long-term effects of metacognitive strategy instruction on student academic performance: A meta-analysis. *Educational Research Review*, *24*, 98-115. Flavell, J.H. (1971). What is memory development the development of? *Human Development*, 14(4), 272-278. https://doi.org/10.1159/000271221

Gaonac'h, D. (2019). *Quand le cerveau se cultive : Psychologie cognitive des apprentissages*. Hachette.

Gaonac'h, D. (2022). *Les élèves et la mémoire*. Retz (Mythes et Réalités).

Karpicke, J.D., Butler, A.C. & Roediger, H.L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, 17(4), 471-479. https://doi.org/10.1080/09658210802647009

Karpicke, J.D. & Roediger, H.L. (2008). The critical importance of retrieval for learning. Science, 319(5865), 966-968. https://doi.org/10.1126/science.1152408

Landauer, T.K. & Bjork, R.A. (1978). Optimum rehearsal patterns and name learning. *In* M.M. Gruneberg et al. (Eds.), *Practical aspects of memory* (pp. 625-632). Academic Press.

McDaniel, M.A. *et al.* (2011). Test-enhanced in a Middle School Science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, 103(2), 399-414. https://doi.org/10.1037/a0021782

Mullet, H.G. *et al.* (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition*, 3(3), 222-229. https://doi.org/10.1016/j.jarmac.2014.05.001

Pyc, M.A. & Rawson, K.A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330(6002), 335. https://doi.org/10.1126/science.1191465

Roediger, H.L. & Butler, A.C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20-27. https://doi.org/10.1016/j.tics.2010.09.003

Rowland, C.A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432-1463. https://doi.org/10.1037/a0037559

Tsalas, N., Sodian, B. & Paulus, M. (2017). Correlates of metacognitive control in 10-year old children and adults. *Metacognition Learning*, 12(3), 297-314. https://doi.org/10.1007/s11409-016-9168-4

Uttl, B., White, C. A. & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22-42.

Veenman, M.V.J., van Hout-Wolters, B.H.A.M. & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, 1, 3-14. https://doi.org/10.1007/s11409-006-6893-0

Zaromb, F.M. & Roediger, H.L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8), 995-1008. https://doi.org/10.3758/mc.38.8.995

Comment le numérique peut-il aider les enseignants à mettre en œuvre des évaluations formatives ?

FRANCK SILVESTRE
UNIVERSITÉ DE TOULOUSE (FRANCE)

Introduction

Le champ de recherche des Environnements Informatiques pour l'Apprentissage Humain (EIAH) s'intéresse à la production de connaissances en ingénierie des systèmes technologiques pouvant soutenir l'apprentissage et/ou l'enseignement humain. Ce champ de recherche s'appuie en général sur des cadres théoriques et des connaissances issues de recherches en sciences de l'éducation, en psychologie cognitive, ou de toute autre discipline des sciences humaines et sociales pouvant éclairer les situations d'apprentissage ou d'enseignement. L'évaluation formative et les notions qui lui sont connexes comme le feedback ont fait l'objet de nombreux travaux de recherche. En tant que telle, l'évaluation formative est un bon candidat à la recherche en EIAH qui s'intéresse à son sujet à la manière de soutenir la mise en œuvre de séquences d'évaluations formatives à l'aide de la technologie.

Dans cette note, après avoir rappelé les définitions et notions clés relatives à l'évaluation formative, nous décrivons comment, dans le contexte d'un enseignement de masse, la technologie est apparue pour soutenir sa mise en œuvre. Nous présentons ensuite, comment ces technologies s'adaptent aujourd'hui, notamment en mobilisant l'analyse des données d'apprentissage (*learning analytics*), pour assister l'exécution d'évaluations formatives promouvant l'argumentation écrite et l'instruction entre pairs.

I. Les notions clés

En 2007, Wiliam introduit les expressions « évaluation de l'apprentissage » et « évaluation pour l'apprentissage » pour distinguer respectivement évaluation sommative et évaluation formative (Wiliam, 2007). Les évaluations sommatives ont pour objectif la mesure du niveau d'atteinte par les étudiants des objectifs d'un programme scolaire ; elles permettent, à la fin d'une unité d'enseignement, de comparer le niveau ou la performance de l'étudiant par rapport à un niveau de référence attendu en vue de l'éventuelle délivrance d'une certification (Sadler, 1989). L'évaluation formative, quant à elle, s'inscrit dans une démarche de régulation des apprentissages (Perrenoud, 1998 ; Wiliam, 2007) poursuivant ainsi un autre objectif décrit dans (Carnegie Mellon University, s. d.) :

> *« L'objectif des évaluations formatives est de mesurer le niveau d'apprentissage des étudiants pour fournir un* feedback *qui peut être utilisé par les enseignants pour améliorer leur enseignement et par les étudiants pour améliorer leur apprentissage. De manière plus spécifique, les évaluations formatives :*

> *- aident les étudiants à identifier leur force et leur faiblesse et à cibler les sujets qui nécessitent plus de travail ;*
>
> *- aident les enseignants à reconnaître les difficultés des étudiants et à traiter les problèmes immédiatement. »*

Cette définition explicite le rôle clé joué par le concept de feedback au cœur de l'évaluation formative, et corrobore ainsi l'idée que les concepts de feedback et d'évaluation formative sont étroitement liés (Black & Wiliam, 1998).

Selon Sadler, un feedback est une information portant sur l'écart entre un niveau actuel et un niveau de référence. Dans le contexte d'une évaluation formative, le niveau observé est le niveau d'atteinte par un apprenant d'un ou de plusieurs objectifs d'un enseignement ; cette information est utilisée pour modifier cet écart d'une quelconque manière (Sadler, 1989).

Sadler identifie trois conditions nécessaires pour que les apprenants tirent un bénéfice d'un feedback dans le contexte d'une tâche d'apprentissage ; il déclare qu'un apprenant doit connaître :

- ce qu'est la performance attendue ;
- comment se positionne sa performance actuelle par rapport à la performance attendue ;
- comment agir pour réduire l'écart entre sa performance actuelle et la performance attendue.

De plus, les études sur l'engagement cognitif des apprenants montrent que l'argumentation entre pairs, plus généralement la possibilité de s'appuyer sur les connaissances des autres apprenants pour construire sa propre connaissance conduit à un engagement cognitif important (Chi *et al*., 2018) et donc à de meilleurs résultats d'apprentissage. En plus du feedback « traditionnel », l'introduction de phase d'interactions entre pairs au sein des évaluations formatives, quelle que soit la forme que prennent ces interactions, est une clé de succès dans leur mise en œuvre.

L'évaluation formative et la fourniture de feedback *ad hoc* sont des leviers pour améliorer les résultats d'apprentissage, quels que soient le niveau et les disciplines dans lesquels elles opèrent (Black & Wiliam, 2009 ; Hattie & Timperley, 2007). L'évaluation formative est donc idéalement mise en œuvre de manière fréquente et systématique.

II. Les évaluations formatives et leur mise en œuvre dans la classe

Black et Wiliam, dans le cadre de leur théorie de l'évaluation formative (Black & Wiliam, 2009), ont identifié 5 activités de référence dans la mise en œuvre d'évaluations formatives :

- le partage des critères de succès avec les apprenants ;
- le questionnement à la classe ;
- l'évaluation sous forme de commentaires plutôt qu'à l'aide de notes chiffrées uniquement ;
- l'auto-évaluation et l'évaluation par les pairs ;
- l'utilisation de tests sommatifs en vue d'usage formatif.

Ces 5 activités s'inscrivent plus largement dans 5 stratégies de mise en œuvre que Black et Wiliam formulent comme suit :

1. clarifier et partager les objectifs d'apprentissage et les critères de succès ;
2. concevoir des situations efficaces de discussions en classe et d'autres tâches d'apprentissage qui élicitent les preuves du niveau d'avancement des apprenants dans leur apprentissage ;
3. fournir des feedbacks qui permettent aux apprenants de progresser dans leur apprentissage ;
4. faire en sorte que les apprenants deviennent des ressources d'instruction pour les autres apprenants ;
5. faire en sorte que les apprenants deviennent responsables de leur propre apprentissage.

Concrètement, l'évaluation formative peut se décliner à l'oral : en face à face, c'est ce qui se produit quand un enseignant interroge un ou plusieurs élèves en classe. La réponse fournie, censée refléter le niveau de l'élève sur l'objectif d'apprentissage ciblé, est analysée et commentée par les pairs ou directement par l'enseignant. Ces échanges et l'intervention de l'enseignant constituent le feedback adressé aux élèves. Dans ce contexte, plus le nombre d'élèves est important, plus il est difficile de collecter et de traiter l'ensemble des réponses apportées par les élèves. Pour une question donnée, il est alors très probable que seul un petit lot d'élèves soit engagé fortement dans la séquence d'évaluation formative.

Une autre forme de pratique consiste à donner un travail écrit à réaliser en classe ou à la maison qui sera relevé et corrigé par l'enseignant afin de fournir un feedback personnalisé. Cette forme d'activité peut être dispensée entièrement à distance ou dans un contexte hybride : les élèves travaillent en dehors de la classe et tout ou partie de la restitution s'effectue en face à face.

Cette démarche vise l'engagement de tous les élèves dans des activités d'apprentissage, mais pose au moins deux problèmes :

- le feedback est fourni loin du temps de l'activité, ce qui peut nuire à son efficacité (Hattie & Timperley, 2007) ;
- le travail de l'enseignant est difficilement soutenable si l'objectif est une mise en œuvre fréquente et systématique de l'évaluation formative.

En synthèse, la mise en œuvre de l'évaluation formative augmente en complexité dès lors que le nombre d'élèves augmente dans la classe. Des systèmes technologiques sont apparus pour collecter et analyser un grand nombre de réponses obtenues à une question posée à la classe afin de fournir, au bon moment, un feedback adapté, c'est-à-dire aidant, *in fine,* l'apprenant à apprendre.

III. Les évaluations formatives supportées par la technologie

A. Les tests autocorrectifs

Dès 1993, plusieurs travaux (Clariana, 1993 ; Zakrzewski & Bull, 1998) montrent l'efficacité de dispositifs d'auto-évaluation informatisés pour la mise en œuvre d'évaluations formatives dans l'enseignement supérieur.

Différents travaux (Miller, 2009 ; Ricketts & Wilks, 2002) montrent que les évaluations informatisées facilitent la mise en œuvre d'évaluations formatives fréquentes pour des cohortes comprenant un grand nombre d'étudiants. La fourniture d'un feedback aux étudiants relatif à leurs travaux ou à leurs résultats d'évaluation est une clé de succès pour l'amélioration de leur apprentissage (Hattie &

Timperley, 2007 ; Higgins, 2000). Nicol et Macfarlane-Dick préconisent ainsi l'utilisation de tests en ligne pour que le feedback puisse être mis à disposition des apprenants à n'importe quel instant, depuis n'importe quel lieu, aussi souvent que l'apprenant le souhaite (Nicol & Macfarlane-Dick, 2006).

La mise en place des tests autocorrectifs est triviale pour évaluer la maîtrise de connaissances (utilisation de questionnaires mobilisant des questions à choix multiples, des questions d'appariement, des textes à trous, etc.). La compréhension de concepts peut aussi être évaluée avec des questions fermées : les travaux sur les c*onceptests,* correspondant à un lot de questions permettant de tester la compréhension d'une notion plutôt que la mémorisation superficielle d'une formule ou d'un énoncé, rendent possible la correction automatique de questions fermées évaluant la compréhension des apprenants sur des concepts dans des domaines variés (Falconer, 2004 ; McConnell *et al.*, 2006 ; Schlatter, 2002). L'accès aux questionnaires interactifs des études PISA permet de consulter un grand nombre de questions fermées, qui peuvent donc être corrigées de manière automatique, visant à évaluer des compétences de résolution de problèmes en mathématiques et en sciences (*PISA 2015 - PISA*, s. d.).

Certaines activités complexes restent difficiles à évaluer de manière automatique. Par exemple, les études menées sur des technologies proposées pour évaluer automatiquement les compétences de rédaction d'essais peuvent mentionner des réserves sur la qualité de l'évaluation ou sur celle du feedback (Hoang & Kunnan, 2016) ou, quand les résultats sont positifs (amélioration de la compétence d'écriture constatée) pointer la nécessité pour l'enseignant de bien comprendre comment tirer parti de l'outil pour l'obtention d'effets positifs (Moore & MacArthur, 2016) .

B. Les systèmes de votes interactifs

Les systèmes de votes interactifs permettent à une classe entière ou un amphithéâtre de répondre à une question à choix multiples affichée sur un écran (Kay & LeSage, 2009). Chaque étudiant utilise un dispositif (boîtier de vote, plus récemment smartphone, tablette ou ordinateur) lui permettant de saisir la réponse à la question. Les réponses sont collectées en temps réel et les résultats affichés après que l'ensemble des étudiants ont répondu à la question. Les résultats sont en général présentés sous forme d'histogrammes et préservent l'anonymat des participants. Diverses études ont montré les bénéfices apportés par l'utilisation des systèmes de votes interactifs. Ceux-ci s'avèrent très efficaces pour motiver et engager les étudiants (Gauci *et al*., 2009 ; Uhari *et al.*, 2003) ainsi que pour améliorer les résultats des étudiants participant aux votes interactifs (Shaffer & Collura, 2009). L'étude de Gauci *et al*. (Gauci *et al*., 2009) insiste sur le fait que les résultats obtenus permettent d'orienter les discussions pendant le cours. Ces résultats donnent l'occasion à tous les participants (enseignants et apprenants) de donner et de recevoir plus d'explications ou de résoudre des conflits sociocognitifs.

C. L'instruction par les pairs

L'approche baptisée « instruction par les pairs » (IP – « *peer-Instruction »* en anglais) a été introduite par Mazur pour la première fois en 1991 durant des cours de physique pour des étudiants de licence. L'approche consiste, pour une même question de compréhension (issue d'un *conceptest*), à sonder les étudiants deux fois : une première fois après une phase de réflexion individuelle, puis une deuxième fois après que les étudiants ont été invités à convaincre leurs voisins du bien-fondé de la réponse qu'ils ont fournie à l'issue du premier sondage. À suite du deuxième vote, un feedback collectif est donné par l'enseignant sur la réponse correcte et le raisonnement attendu. Les résultats du deuxième

sondage révèlent d'une part un plus grand nombre de bonnes réponses, mais surtout que les bénéfices en termes d'apprentissage se trouvent renforcés par rapport à une approche classique (Mazur, 1997).

Devant les résultats probants apportés par l'approche IP (Crouch & Mazur, 2001), beaucoup d'universités se sont lancées dans l'utilisation des systèmes de votes interactifs en appliquant la méthode de Mazur (Burnstein & Lederman, 2001 ; Cline, 2006 ; Parmentier *et al.*, 2015), y compris en France (Rudolph *et al*., 2014 ; J. F. Parmentier *et al*., 2015).

En forçant la discussion entre pairs entre les deux phases d'interrogation de l'assemblée, l'approche IP renforce d'une part la mise en place de situations de discussions en classe, et d'autre part le fait que les étudiants deviennent des ressources d'instructions pour les autres.

D. Le développement des services Web

Le développement des technologies permet la mise en œuvre d'évaluations formatives à grande échelle. Le développement d'Internet a permis de porter les dispositifs technologiques initiaux sur le Web rendant accessibles les services offerts *via* tout dispositif connecté au réseau des réseaux : ordinateur, tablette, smartphone.

Divers outils permettent la mise en place de tests autocorrectifs. Les principales plateformes de cours en ligne, telles que Moodle, mettent à disposition des services de création et de passage de tests autocorrectifs. La plupart des environnements numériques de travail fournissent aussi de tels dispositifs. Enfin, le réseau Canopé propose le service en ligne Quizinière[78].

Les services Kahoot[79], Socrative[80], Wooclap[81] et Plickers[82] sont des exemples (liste non exhaustive) de services Web soutenant la mise en œuvre d'évaluations formatives sur le modèle des systèmes de votes interactifs. Les études portant sur ces outils pointent des bénéfices similaires à l'utilisation des dispositifs antérieurs : engagement et motivation renforcés pour les étudiants (Chng & Gurvitch, 2018 ; Wang & Tahir, 2020 ; Wash, 2014).

Cette transition vers les services Web a permis d'une part un enrichissement des interactions et d'autre part une mobilisation de ces technologies dans des contextes d'apprentissage à distance ou hybride.

IV. L'évaluation formative promouvant l'argumentation écrite et l'interaction entre pairs : le cas du processus à deux votes

L'instruction par les pairs est une approche d'évaluation formative soutenue par la technologie et promouvant l'interaction entre pairs entre deux phases de questionnement. Dans sa forme originale, les interactions d'effectuent à l'oral. Deux plateformes, Elaastic[83] (J.-F. Parmentier & Silvestre, 2019) et myDalite[84] (Charles *et al*., 2019), se sont intéressées à la transposition de cette approche afin de promouvoir l'argumentation écrite tout en maintenant les interactions entre pairs. Toutes deux permettent la pratique d'évaluations formatives en face à face, à distance ou dans un contexte

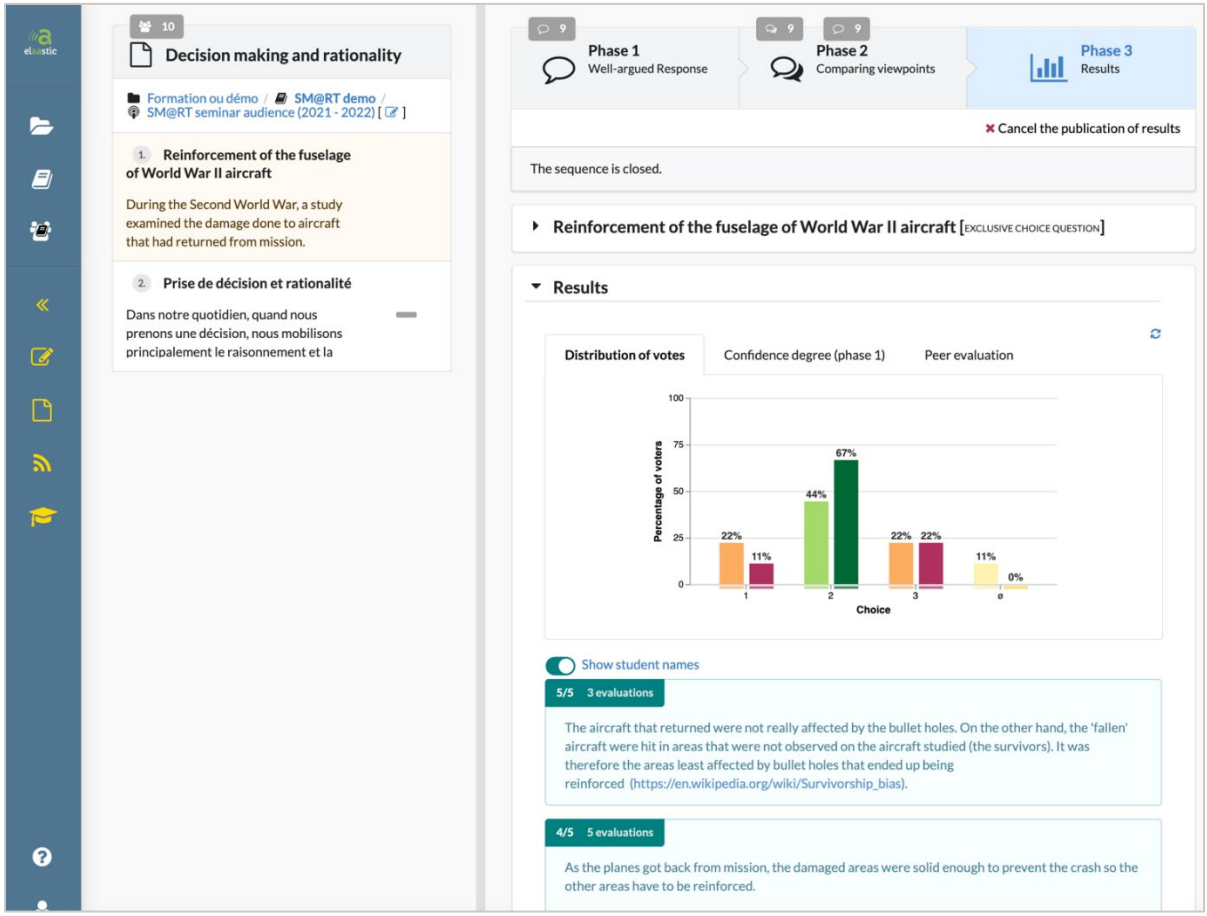
[78] https://www.quiziniere.com/
[79] https://kahoot.it/
[80] https://www.socrative.com/
[81] https://www.wooclap.com/fr/
[82] https://get.plickers.com/
[83] https://elaastic.irit.fr/
[84] https://mydalite.org/fr/

d’enseignement hybride. Les séquences d’évaluations formatives menées en mobilisant l’instruction par les pairs, Elaastic ou myDalite s’inscrivent donc toutes dans un processus dans lequel les apprenants répondent une première fois à une question à choix, « subissent un traitement pédagogique » correspondant à une confrontation de point de vue, puis répondent de nouveau à la même question. La confrontation de point de vue prend alors différentes formes dépendant de l’outil mobilisé pour orchestrer la séquence. Nous avons baptisé cette séquence générique d’évaluation formative « processus d’évaluation formative à 2 votes » (*2-votes based process* en anglais) (Andriamiseza *et al*., 2021a, 2021b, 2021c), elle est illustrée dans la Figure 1.

Figure 1 : Processus d’évaluation formative à 2 votes

Dans Elaastic et myDalite, lors du premier vote, il est demandé aux apprenants de fournir une explication écrite pour justifier leurs choix. Dans myDalite, le deuxième vote prend la forme d'une sélection d'une explication proposée par les autres apprenants. Ensuite, la plateforme fournit aux enseignants un feedback détaillant le nombre d'apprenants qui sont passés de la mauvaise à la bonne réponse, de la bonne à la mauvaise réponse, et le nombre d'apprenants étant restés sur une bonne ou mauvaise réponse. Dans Elaastic, le deuxième vote est proposé après avoir engagé les apprenants dans une phase d'évaluation par les pairs anonymisée : il leur est demandé d'évaluer plusieurs justifications de leurs pairs en utilisant une échelle de Likert à 5 niveaux leur permettant d'indiquer leur degré d'accord avec l'explication évaluée. À tout moment de la séquence, Elaastic permet à l'enseignant de consulter la répartition des votes des apprenants, les explications fournies et les résultats de l'évaluation par les pairs (Figure 2).

Figure 2 : Restitution des résultats d'une séquence orchestrée avec Elaastic



V. Pratiques et recherches autour de la plateforme Elaastic

Elaastic a d'abord été déployé dans l'enseignement supérieur, principalement pour des enseignements scientifiques (informatique, management de projet, physique) du niveau licence au niveau master. L'étude empirique réalisée à partir des données d'apprentissage collectées dans le supérieur depuis 2015 a permis d'enrichir le processus en mode synchrone (face à face ou à distance) avec la fourniture de recommandations à destination de l'enseignant en cours de séquence : celui-ci peut être invité à raccourcir une séquence si le système évalue que la phase d'évaluation par les pairs a peu de chance d'être efficace, ou à porter attention à certaines explications écrites lors de la restitution en fonction des résultats obtenus sur les différentes activités composant la séquence (Andriamiseza *et al*., 2021a, 2021b, 2021c).

En 2019, dans le contexte du projet B4MATIVE! mené en partenariat avec l'académie de Nancy-Metz et financé par la direction du numérique pour l'éducation, Elaastic a été déployé dans secondaire, du collège au lycée, dans des champs disciplinaires plus variés : sciences, langues, musique, etc. Peu d'usages ont été constatés pendant la période de la pandémie : il a été en effet compliqué d'accompagner les enseignants à distance sur un tel outil dans un contexte d'urgence à s'adapter aux contraintes des confinements successifs.

Plus récemment, en 2021, a débuté une recherche collaborative avec le collège Jean-Paul Laurens en Haute-Garonne, dans lequel des enseignants de SVT, français, technologie, histoire-géographie et mathématiques se sont engagés fortement dans l'utilisation d'Elaastic. Il a été collecté en une année

scolaire un volume de données permettant de répliquer les tests statistiques que nous avions appliqués sur les données du supérieur, dont les résultats sont cohérents avec les précédents[85]. De manière qualitative, les enseignants du collège mentionnent dans leurs rapports d'usages des bénéfices sur différents aspects de l'apprentissage, comme l'engagement et la motivation des élèves, ou encore la progression sur l'argumentation écrite, comme le suggèrent ci-dessous les extraits d'un rapport d'usage en classe de 6e de l'enseignante d'histoire-géographie engagée dans la recherche collaborative :

> « …le recours au numérique suscite une motivation accrue de tous les élèves, même ceux en difficultés ou ceux à profil particulier qui bénéficient de dispositifs spécifiques (PAP, Gevasco). Les élèves qui ne sont pas forcément à l'aise avec l'informatique ont trouvé l'utilisation de la plateforme tout à fait accessible, et ils ont rencontré moins de difficultés techniques que lors de l'utilisation d'autres supports numériques. »

> « L'ouverture et la fermeture de chaque phase permet au groupe de se canaliser : les élèves doivent en effet être attentifs aux consignes données à l'oral par l'enseignante (exemple : « tout le monde a bien répondu à la question 1 ? je ferme la phase 1 dans une minute ; attention je ferme la phase 1 ; j'ouvre à présent la phase 2 ») comme aux consignes écrites de l'outil. Ces phases permettent de rythmer le travail tout en favorisant la concentration des élèves, même de ceux qui ont plus de mal à se centrer sur une tâche, puis à se recentrer sur la tâche suivante. »

> « La qualité de la réponse justifiée a été bien meilleure que lors des autres travaux d'argumentation menés en classe (qu'ils aient été faits collectivement, en binômes ou en équipes de quatre), même si l'effet d'entraînement a incontestablement joué avec ce quatrième retour sur la démarche critique. La plus-value est indéniable en ce qui concerne l'objectif cognitif et la manière de penser et de réfléchir des élèves : ils ont pris davantage le temps de réfléchir individuellement à leur justification et au choix de leurs arguments (volontairement nombreux dans cette leçon). Ils se sont appliqués à taper leur réponse et se sont davantage relus, soignant la maîtrise de la langue et l'expression écrite [l'enseignante les avait avertis que leur réponse serait visible par toute la classe]. Ils ont mieux réinvesti ce qui avait été fait en classe (certains ont même eu le réflexe de prendre leur manuel et leur cahier pour utiliser les outils déjà travaillés lors des cours précédents). »

Conclusion

La mise en œuvre de séquences d'évaluation formative en classe ou hors la classe est un levier pour améliorer les apprentissages. Cependant, dans un contexte d'enseignement de masse, sa mise en œuvre est complexe et pose des problèmes de soutenabilité pour les enseignants. Les technologies numériques apportent depuis quelques décennies des réponses de plus en plus sophistiquées pour répondre à ces défis. L'avènement et la démocratisation d'Internet et des dispositifs connectés permettent de proposer aujourd'hui des outils promouvant des interactions riches et collaboratives. Les travaux menés autour de la plateforme Elaastic s'inscrivent dans une démarche visant à analyser et traiter les données d'usages en temps réel ou en différé pour exploiter des variables que l'enseignant n'est pas en capacité de traiter sans l'aide de la machine : dans cette approche, le numérique peut aider à rendre les processus d'apprentissage et leurs résultats « visibles ».

[85] Les résultats de cette dernière étude ne sont pas encore publiés à ce jour.

Références

Andriamiseza, R., Silvestre, F., Parmentier, J.-F., & Broisin, J. (2021a). Recommendations for Orchestration of Formative Assessment Sequences : A Data-Driven Approach. In T. De Laet, R. Klemke, C. Alario-Hoyos, I. Hilliger, & A. Ortega-Arranz (Éds.), *Technology-Enhanced Learning for a Free, Safe, and Sustainable World* (Vol. 12884, p. 245-259). Springer International Publishing. https://doi.org/10.1007/978-3-030-86436-1_19

Andriamiseza, R., Silvestre, F., Parmentier, J.-F., & Broisin, J. (2021b). Vers la conception de feedback pour enseignants dans un contexte d'évaluation formative à grande échelle : Une approche analytique. In M. Lefevre, C. Michel, T. Geoffre, M. Rodi, L. Alvarez, & A. Karoui (Éds.), *10ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2021)* (p. 46-57). ATIEF : Association des Technologies de l'Informationpour l'Éducation et la Formation. https://hal.archives-ouvertes.fr/hal-03292736

Andriamiseza, R., Silvestre, F., Parmentier, J.-F., & Broisin, J. (2021c). Data-informed Decision-making in TEFA Processes : An Empirical Study of a Process Derived from Peer-Instruction. *Proceedings of the Eighth ACM Conference on Learning @ Scale*, 259-262. https://doi.org/10.1145/3430895.3460153

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

Burnstein, R. A., & Lederman, L. M. (2001). Using wireless keypads in lecture classes. *The Physics Teacher*, *39*(1), 8-11. https://doi.org/10.1119/1.1343420

Carnegie Mellon University. (s. d.). *Formative vs Summative Assessment*. Consulté 22 juillet 2022, à l'adresse https://www.cmu.edu/teaching/assessment/basics/formative-summative.html

Charles, E. S., Lasry, N., Bhatnagar, S., Adams, R., Lenton, K., Brouillette, Y., Dugdale, M., Whittaker, C., & Jackson, P. (2019). Harnessing peer instruction in- and out- of class with myDALITE. *Fifteenth Conference on Education and Training in Optics and Photonics: ETOP 2019 (2019), Paper 11143_89*, 11143_89. https://www.osapublishing.org/abstract.cfm?uri=ETOP-2019-11143_89

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP Theory of Cognitive Engagement Into Practice. *Cognitive Science*, *42*(6), 1777-1832. https://doi.org/10.1111/cogs.12626

Chng, L., & Gurvitch, R. (2018). Using Plickers as an Assessment Tool in Health and Physical Education Settings. *Journal of Physical Education, Recreation & Dance*, *89*(2), 19-25. https://doi.org/10.1080/07303084.2017.1404510

Clariana, R. B. (1993). A Review of Multiple-Try Feedback in Traditional and Computer-Based Instruction. *Journal of Computer-Based Instruction*, *20*(3), 67-74.

Cline, K. S. (2006). Sharing Teaching Ideas : Classroom Voting in Mathematics. *The Mathematics Teacher*, *100*(2), 100-104. https://doi.org/10.5951/MT.100.2.0100

Crouch, C. H., & Mazur, E. (2001). Peer Instruction : Ten years of experience and results. *American Journal of Physics*, *69*(9), 970-977. https://doi.org/10.1119/1.1374249

Falconer, J. L. (2004). Use of ConcepTests and instant Feedback in Thermodynamics. *Chemical Engineering Education*, *38*(1), 64-67.

Gauci, S. A., Dantas, A. M., Williams, D. A., & Kemm, R. E. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, *33*(1), 60-71. https://doi.org/10.1152/advan.00109.2007

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81-112. https://doi.org/10.3102/003465430298487

Higgins, R. (2000, octobre 6). *« Be more critical! » : Rethinking assessment feedback*. Education-line. http://www.leeds.ac.uk/educol/documents/00001548.htm

Hoang, G. T. L., & Kunnan, A. J. (2016). Automated Essay Evaluation for English Language Learners:A Case Study of MY Access. *Language Assessment Quarterly*, *13*(4), 359-376. https://doi.org/10.1080/15434303.2016.1230121

Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems : A review of the literature. *Computers & Education*, *53*(3), 819-827. https://doi.org/10.1016/j.compedu.2009.05.001

Mazur, E. (1997). Peer instruction : Getting students to think in class. *AIP Conference Proceedings*, *399*(1), 981-988. https://doi.org/10.1063/1.53199

McConnell, D. A., Steer, D. N., Owens, K. D., Knott, J. R., Van Horn, S., Borowski, W., Dick, J., Foos, A., Malone, M., & McGrew, H. (2006). Using conceptests to assess and improve student conceptual understanding in introductory geoscience courses. *Journal of Geoscience Education*, *54*(1), 61-68.

Miller, T. (2009). Formative computer-based assessment in higher education : The effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education*, *34*(2), 181-192. https://doi.org/10.1080/02602930801956075

Moore, N. S., & MacArthur, C. A. (2016). Student use of automated essay evaluation technology during revision. *Journal of Writing Research*, *8*(1), 149-175. https://doi.org/10.17239/jowr-2016.08.01.05

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning : A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Parmentier, J. F., Lamine, B., & Bonnafé, S. (2015). *Changer les conceptions en mécanique des étudiants en L1 à l'Université*. 4.

Parmentier, J.-F., & Silvestre, F. (2019). La (dé-)synchronisation des transitions dans un processus d'évaluation formative exécuté à distance : Impact sur l'engagement des étudiants. In J. Broisin, E. Sanchez, A. Yessad, & F. Chenevotot (Éds.), *9ème Conférence sur les EnvironnementsInformatiques pour l'Apprentissage Humain—EIAH 2019*. ATIEF : Association des Technologies de l'Information pour l' Education et la Formation. https://hal.archives-ouvertes.fr/hal-02904155

Perrenoud, P. (1998). From Formative Evaluation to a Controlled Regulation of Learning Processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 85-102. https://doi.org/10.1080/0969595980050105

PISA 2015—PISA. (s. d.). Consulté 24 août 2022, à l'adresse https://www.oecd.org/pisa/test/pisa2015/#d.en.537240

Ricketts, C., & Wilks, S. J. (2002). Improving Student Performance Through Computer-based Assessment : Insights from recent research. *Assessment & Evaluation in Higher Education*, *27*(5), 475-479. https://doi.org/10.1080/0260293022000009348

Rudolph, A. L., Lamine, B., Joyce, M., Vignolles, H., & Consiglio, D. (2014). Introduction of interactive learning into French university physics classrooms. *Physical Review Special Topics - Physics Education Research*, *10*(1), 010103. https://doi.org/10.1103/PhysRevSTPER.10.010103

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119-144. https://doi.org/10.1007/BF00117714

Schlatter, M. D. (2002). Writing ConcepTests for a multivariable calculus class. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *12*(4), 305-314.

Shaffer, D. M., & Collura, M. J. (2009). Evaluating the Effectiveness of a Personal Response System in the Classroom. *Teaching of Psychology*, *36*(4), 273-277. https://doi.org/10.1080/00986280903175749

Uhari, M., Renko, M., & Soini, H. (2003). Experiences of using an interactive audience response system in lectures. *BMC Medical Education*, *3*(1), 12. https://doi.org/10.1186/1472-6920-3-12

Wang, A. I., & Tahir, R. (2020). The effect of using Kahoot ! For learning – A literature review. *Computers & Education*, *149*, 103818. https://doi.org/10.1016/j.compedu.2020.103818

Wash, P. D. (2014). Taking advantage of mobile devices : Using Socrative in the classroom. *Journal of Teaching and Learning with Technology*, 99-101. https://doi.org/10.14434/jotlt.v3n1.5016

Wiliam, D. (2007). *Keeping learning on track : Formative assessment and the regulation of learning*.

Zakrzewski, S., & Bull, J. (1998). The Mass Implementation and Evaluation of Computer-based Assessments. *Assessment & Evaluation in Higher Education*, *23*(2), 141-152. https://doi.org/10.1080/0260293980230203

Comment peut-on rendre compte des progrès des élèves aux familles ? L'exemple de l'école maternelle.

CHRISTOPHE JOIGNEAUX
UNIVERSITÉ PARIS-EST CRÉTEIL (FRANCE)

Introduction

Dans les années 1960 et au début des années 1970, les textes officiels et professionnels relatifs à l'école maternelle française mettaient surtout en avant son modèle pédagogique unique, par son souci de respecter le développement et l'épanouissement du jeune enfant, ses capacités de création, ses besoins affectifs, sa faculté à faire ses propres découvertes… À partir du milieu des années 1970, avec les grandes réformes visant à unifier le système éducatif et à allonger la scolarité obligatoire, on commence à observer de petites inflexions sur ce plan jusqu'au grand tournant des années 1980, à partir duquel ils mettent beaucoup plus l'accent sur les apprentissages à réaliser à l'école maternelle pour préparer ceux qui sont programmés à l'école élémentaire. Cette mission propédeutique de l'école maternelle va devenir progressivement prépondérante, comme le relève l'usage croissant du vocabulaire scolaire jusque-là beaucoup moins utilisé que dans les segments ultérieurs de la scolarité. Ainsi sont de plus en plus présents dans les programmes les termes d'élèves, de « première école », d'apprentissages, d'enseignement, de programmes, de compétences et… d'évaluations (Joigneaux, 2019).

Pour autant, l'évaluation à l'école maternelle continue à faire débat, parce que ne vont pas de soi les réponses qui sont aujourd'hui données aux questions suivantes : faut-il commencer à évaluer à l'école maternelle ? Et si oui, comment évaluer les apprentissages de si jeunes élèves et comment en rendre compte aux parents ?

Cette évolution des textes règlementaires relatifs aux missions de l'école maternelle semble donner une réponse toute trouvée : comme elle doit être la première école, il n'y a aucune raison que l'évaluation des apprentissages que peuvent y faire les élèves ne soit pas présente. Pourtant, quand on étudie de plus près à la fois les textes prescriptifs relatifs à l'évaluation à l'école maternelle et les pratiques de cette dernière dans les classes, on se rend vite compte que cette réponse ne va pas de soi, parce qu'elle ne rend pas compte de la complexité des enjeux de la publicisation de l'évaluation des apprentissages présents non seulement à l'école à ce premier niveau de la scolarité, mais aussi sans doute tout au long de celle-ci. C'est ce que je vais tenter de montrer dans cette note[86].

[86] Le lecteur intéressé pourra approfondir la lecture de cette note en consultant la référence suivante : C. Joigneaux, 2016, « L'évaluation dans les écoles maternelles françaises. Quelles cohérences pour quels effets ? » *in* C. Veuthey, G. Marcoux, T. Grange (dir.), *L'école première en question. Analyses et réflexions à partir des pratiques d'évaluation,* EME éditions, pp. 39-54.

I. Une évolution des pratiques d'évaluation promue par les programmes 2015 et 2021 de l'école maternelle

Les programmes de l'école maternelle de 2015 et de 2021 (le dernier en date) emploient une expression inédite (on ne la trouve pas dans les programmes antérieurs) : « l'évaluation positive ». À ce point inédite d'ailleurs qu'elle a heurté, à la suite de la parution du programme de 2015, beaucoup d'enseignants et d'enseignantes de maternelle, comme ils nous l'ont confié, à l'occasion d'échanges qui ont suivi des conférences : ils affirmaient être meurtris qu'on les soupçonne de faire de « l'évaluation négative ». Et quand on lit attentivement le paragraphe que consacrent ces programmes à l'évaluation positive, on peut effectivement y trouver une dénonciation implicite de pratiques d'évaluation qui ne correspondraient pas à ce modèle, et qui sont décrites en creux comme ceci : des évaluations qui ne mettent en valeur que « les résultats obtenus » des élèves et/ou qui ne constitueraient que des « instruments de prédiction et de sélection ». L'évaluation positive qui doit remplacer ces anciennes pratiques, disent ces textes, ne peut se réduire à cela car elle « s'attache à mettre en valeur, au-delà du résultat obtenu, le cheminement de l'enfant et les progrès qu'il fait par rapport à lui-même ».

Deux modèles de pratiques bien différentes de l'évaluation des apprentissages des élèves de maternelle sont donc opposés dans ces programmes. D'un côté, un modèle qui ne se préoccuperait que des résultats des élèves parce qu'ils seraient prédictifs de leurs apprentissages futurs et qu'ils permettraient pour cette raison de sélectionner les plus aptes à les faire. Et de l'autre côté, un modèle plus novateur, qui devrait remplacer le premier parce qu'il cherche à mesurer non seulement ce que les élèves auraient dû avoir appris, au moment d'une évaluation donnée, mais aussi les « progrès » qu'ils ont pu faire entre plusieurs évaluations. En privilégiant pour évaluer les apprentissages et les progrès des élèves non pas le support de fiches, mais au moyen de l'observation de l'activité des élèves et des interprétations que peuvent en faire les enseignants : « elle [l'évaluation positive] repose sur une observation attentive et une interprétation de ce que chaque enfant dit ou fait ».

Ces deux modèles correspondent donc à deux philosophies de l'évaluation des apprentissages très différentes. L'un considère en effet que le plus important est de mesurer les apprentissages que les élèves ont réalisé à un moment donné, et donc, avant tout, le produit de ces apprentissages. L'autre considère qu'il est tout aussi important de mesurer le processus d'apprentissage lui-même, dans toute sa profondeur temporelle. Et que par conséquent, l'étalon pour cette mesure ne peut pas être seulement constitué des progressions d'acquisition des savoirs et des compétences constituant les programmes scolaires, mais qu'il doit aussi et avant tout tenir compte de la temporalité des apprentissages, propre à chaque élève.

C'est pour mieux respecter cette temporalité des apprentissages de chaque élève que ces programmes listent des apprentissages à réaliser – qualifiés d'« attendus » – durant toute la scolarité à l'école maternelle, et non pas durant chacune des années de cette scolarité, contrairement aux programmes précédents. Cette évolution des programmes n'a pas manqué de désarçonner de nombreux enseignants de maternelle qui regrettaient de ne plus avoir de repères annuels pour programmer et évaluer les apprentissages de leurs élèves, ainsi que pour élaborer les livrets ou carnets d'évaluation. En effet, pour beaucoup d'entre eux, ces nouvelles prescriptions étaient en décalage avec les mises en forme de leurs évaluations, comme le sous-entendait, on l'a vu, le programme de 2015 et comme le donnent à penser les enquêtes que nous avons pu mener à propos de la façon dont les évaluations

sont présentées aux parents, notamment celles qui ont eu lieu juste avant la publication de ce programme.

II. La diversité des comptes rendus aux parents d'élève des évaluations faites en classe

Durant les années scolaires 2013-2014 et 2014-2015, nous avons réalisé deux enquêtes sur les livrets ou carnets d'évaluation (aussi qualifiés de carnet de suivis, de réussite…)[87] dans deux académies (Nord-Pas-de-Calais puis Créteil), couplées à des entretiens et (plus rarement) des observations avec certains des enseignants utilisant ces supports, et prolongées depuis par des recueils de données dépendantes de nos accès dans les écoles en tant que formateur d'enseignants du primaire à Créteil (Joigneaux, 2016).

Les principaux résultats qu'on peut dégager de toutes ces enquêtes est que dans 90 % des livrets analysés, est essentiellement évaluée la maîtrise de compétences ou de savoirs très précis, très liés à certains types de tâches et aux contextes de leur réalisation. Exemple en petite section (première année de maternelle) : « coller des gommettes dans une forme dessinée ». Ces compétences correspondent donc à des consignes très détaillées ainsi qu'à des supports et contextes d'évaluation très précis. Elles semblent de ce fait très faciles à évaluer car leur « niveau » d'acquisition peut être facilement quantifié. Pour reprendre le même exemple : la présence ou non de gommettes dans une forme dessinée, puis leur nombre dans et en dehors de cette forme. Mais la structure des livrets suggère que ces évaluations ne doivent être réalisées qu'assez ponctuellement dans la mesure où elles ne figurent qu'au sein de plages temporelles très réduites par rapport à l'étendue de l'année scolaire : le plus souvent une période (l'année scolaire de l'école primaire est décomposée en « périodes » qui sont délimitées par les vacances scolaires) ou un trimestre, jamais au-delà d'une année scolaire.

A contrario, dans 10 % des livrets analysés, est évaluée la maîtrise de compétences ou de savoirs beaucoup plus transversaux, à plusieurs reprises, durant plusieurs périodes de l'année, voire durant toutes les années de la scolarité en maternelle. Exemple : « se repérer dans un livre et l'espace d'une page ». Contrairement à celles qu'on peut trouver dans le premier type de livrets, ces compétences plus transversales peuvent être évaluées dans des contextes (situations, supports, exercices…) très différents et peuvent être plus difficilement quantifiables. Pour reprendre l'exemple de l'évaluation de la compétence relative au repérage dans les livres, elle peut être faite toutes les fois où les élèves manipulent un des livres présents dans la classe, y compris à l'occasion de moments relativement informels comme quand les élèves échangent librement (sans que l'enseignant dirige les échanges) dans les « coins lecture » (espaces où sont rassemblés des supports écrits mis à la disposition des élèves). Cela suppose le recours à des observations et des situations d'évaluation plus informelles et plus continues que celles dans lesquelles sont évaluées les compétences plus ciblées contenues dans le premier type de livrets.

C'est la même proportion (90%-10%) entre ces deux types de supports de compte rendu d'évaluation que nous avons retrouvé de façon moins systématique depuis dix ans dans l'académie de Créteil.

Ces résultats suggèrent qu'il existe bien deux modèles assez contrastés sinon d'évaluations, du moins de comptes rendus donnés aux parents d'élèves de l'évaluation des apprentissages en maternelle. On trouve d'un côté un modèle de comptes rendus qui cherche avant tout à mesurer les écarts possibles entre les tâches très ciblées qu'un enfant est capable de faire à un instant *t* et ce qui est attendu de

[87] Pour faire simple, nous parlerons le plus souvent, dans la suite de ce texte, de livrets d'évaluation.

lui, durant cette année scolaire ou même durant une période de l'année plus limitée. Et d'un autre côté, un modèle de comptes rendus d'évaluation plus « positif » visant plutôt à observer les progrès d'un même élève dans l'acquisition d'une même compétence ou d'un même savoir, plusieurs fois durant une même année ou au cours de plusieurs années successives. Le second cherche donc plutôt à présenter le processus d'apprentissage à partir de l'évaluation de compétences ou de savoirs relativement transversaux, et non seulement un de ces produits temporaires et réduits à des compétences ou savoirs très ciblés, tel qu'il peut se manifester à un l'instant où il est évalué, comme c'est le cas dans le cadre du premier modèle de comptes rendus.

III. Les enjeux et les obstacles de la formalisation d'une évaluation plus « positive »

De la coexistence à la maternelle des deux types de comptes rendus d'évaluations exposés ci-dessus, on peut tirer deux enseignements concernant les enjeux et les obstacles à la formalisation d'une évaluation davantage centrée sur les processus d'apprentissage, et pas seulement sur leurs produits temporaires.

Le premier enseignement concerne les enjeux d'une évaluation plus positive. On a vu que ce qui différencie le plus les deux grands types de comptes rendus d'évaluation que nous avons distingués à partir de l'analyse de notre échantillon de livrets est leur propension à rendre compte exclusivement soit des processus, soit des produits d'apprentissage. On pourrait penser que cette différence est assez secondaire et c'est d'ailleurs sans doute pour cette raison que son enjeu pour les apprentissages a pu longtemps être sous-estimé, y compris par les chercheurs. Mais ne donner à voir aux familles que des apprentissages très ponctuels réalisés dans des domaines de savoirs ou de compétences très restreints semble conduire les enseignants à exclure des évaluations présentées dans les supports transmis aux familles les compétences les plus transversales, comme « se repérer dans un livre et dans l'espace d'une page » et plus généralement toutes celles qui concernent l'autonomie cognitive ou le rapport au langage (oral et écrit) des élèves. Or un grand nombre de travaux ont montré que l'enseignement et l'évaluation de ce type de compétences peuvent être décisives pour rendre plus explicites les pédagogies et ainsi ne pas accentuer les inégalités existantes entre les élèves dès l'école maternelle (voir par exemple Bourdieu & Passeron, 1970 ; Bernstein, 1975 ; Charlot *et al.*, 1992 ; Lahire, 1993 ; Joigneaux, 2009). Et plus précisément s'agissant des livrets d'évaluation, ne pas rendre compte des progrès réalisés dans l'acquisition de telles compétences, c'est faire courir le risque de rendre implicite aux familles les plus éloignées de la culture scolaire ce qui fonde cette dernière et, en regard, le degré d'acculturation scolaire de leur enfant.

Pourquoi une évaluation plus « positive » et plus explicite n'est-elle pas donnée plus à voir aux parents au moyen de supports rendant davantage compte des progrès des élèves en évaluant des compétences et des savoirs plus transversaux et plus liés à la réussite scolaire à long terme ? Il est difficile de répondre à cette question parce qu'à notre connaissance, elle n'a pas été posée en ces termes. Mais certaines recherches ou nos propres enquêtes peuvent indirectement apporter quelques éléments de réponse à cette question.

Première réponse possible, explorée en particulier par des chercheurs appartenant au réseau Reseida[88], ces compétences et ces savoirs sont peu évalués parce que « transparents » à une grande

[88] Reseida, acronyme de Recherches sur la Socialisation, l'Enseignement, les Inégalités et les Différenciations dans les Apprentissages. Ces recherches sont rassemblées dans l'ouvrage de Rochex & Crinon (2011).

partie des enseignants. Ils le seraient en raison des insuffisances actuelles de la formation et/ou d'un déficit de diffusion des recherches sur les inégalités scolaires visant à cerner les dimensions les plus transversales de culture scolaire qui pourraient expliquer la reproduction des inégalités à l'école.

Autre réponse possible, plutôt alternative, que nous avons commencé à suggérer plus haut : ces types de savoirs et de compétences ne sont pas évalués ou, s'ils sont évalués, ne sont pas souvent présentés aux parents parce qu'ils sont difficilement formalisables et décomposables en items dans des livrets. Nous en avons fait indirectement l'expérience en présentant à des enseignants les résultats de nos recherches sur les difficultés rencontrées par un grand nombre d'élèves de milieux populaire à exploiter les pouvoirs réflexifs de l'écrit, c'est-à-dire les possibilités qu'offre la permanence de l'écrit de reconsidérer ce qu'on a déjà lu (ou parcouru des yeux) ou écrit (ou tracé) quand par exemple on les fait travailler sur des fiches (feuilles préimprimées) à l'école maternelle (Joigneaux, 2009). En effet, sans formaliser cette compétence transversale comme cela, les enseignants se montrent souvent d'accord avec cette piste d'analyse des ressorts langagiers des inégalités scolaires, mais après avoir désigné ce type de difficultés de façon plus informelle par des formulations du type : « ces élèves sont moins bien organisés que les autres ». Autrement dit, ils ont une connaissance relativement informelle de cette compétence, ils l'ont bien souvent repérée mais ils ont des difficultés à la formaliser, que ce soit à l'oral ou dans des livrets destinés aux parents d'élève.

Une autre explication encore peut être avancée, complémentaire à la précédente mais plus difficilement objectivable : pour un faisceau de raisons différentes, les enseignants évaluent surtout en pratique, c'est-à-dire dans des situations relativement informelles, ce type de compétences et de savoirs. D'abord parce que comme le suggèrent nos enquêtes présentées plus haut, évaluer des compétences relativement transversales est moins aisé que de mesurer les acquisitions relatives à des compétences ou savoirs à spectre moins large, qu'on peut par conséquent plus facilement associer à des tâches et des supports faciles à mobiliser dans le quotidien ordinaire de la classe, comme le fait de coller des gommettes dans des formes données, on l'a vu. Ensuite parce que, comme on l'a également déjà écrit, il est aussi plus facile de mesurer des degrés différents d'acquisition relatifs à ces savoirs ou compétences parce qu'ils sont souvent quantifiables : le comptage du nombre de gommettes présentes dans une forme donnée est rapide et aisé. En outre, cela peut se faire sur des supports papier, et donc l'évaluation proprement dite peut se faire en dehors du temps de classe, contrairement aux observations de compétences plus transversales.

Alors que la mise en œuvre systématique de ces observations est beaucoup plus coûteuse en temps (cela requiert d'observer chaque élève, individuellement) et nécessite la mise en œuvre de stratégies pédagogiques beaucoup plus complexes parce que pendant que l'enseignant observe un (groupe d') élève(s), il faut que les autres soient suffisamment autonomes pour ne pas interférer dans ce processus. Or cette pédagogie de l'autonomie (Lahire, 2005) n'est pas facile à mettre en place, comme nous l'ont rapporté la grande majorité des enseignants sur lesquels ont porté nos enquêtes depuis 25 ans, en maternelle mais aussi au collège (Joigneaux & Rockwell, 2020 ; Netter & Joigneaux, 2022 ;).

Au final, cette analyse des enjeux et des obstacles de la formalisation d'évaluations davantage centrées sur les progrès d'acquisition des élèves de savoirs et compétences suffisamment transversaux pour donner à voir les principaux réquisits de la culture scolaire, donne à voir toute la complexité de mettre en œuvre ce que certains chercheurs ont qualifié de pédagogie de l'autonomie (Lahire, 2005 ; Netter & Joigneaux, 2022), mais qui pourrait aussi être qualifiée de pédagogie du développement. Que ce soit en effet dans sa conception par la formalisation de ce type de savoirs et de compétences et de sa

mesure, sa mise en œuvre ou sa communication notamment aux parents, ce type d'évaluation implique en effet de mieux articuler le temps standardisé des apprentissages scolairement programmés et les temporalités nécessairement singulières du développement de chaque élève. Or cela nécessite beaucoup plus de temps d'observations consacrés à ce que font, disent ou même pensent les élèves, de formalisations pour en restituer les résultats, ce qu'un enseignant, seul (sans accompagnement institutionnel) et sans coopération au sein de son école ; dans une classe surchargée, peut difficilement mettre en œuvre[89].

[89] Ce que nous avons cherché à montrer indirectement à travers l'analyse de la grande solitude institutionnelle des Professeurs des Écoles débutants, actuellement (voir Broccolichi *et al.*, 2018).

Références

Bourdieu, P. & Passeron, J.C. (1970). *La Reproduction*. Les Éditions de Minuit.

Bernstein, B. (1975). *Langage et classes sociales*. Les Éditions de Minuit.

Broccolichi, S., Joigneaux, C. & Mierzejewski, S. (2018). *Le parcours du débutant. Enquête sur les premières années d'enseignement à l'école primaire*. Presses universitaires de l'Artois.

Charlot, B., Bautier, E. & Rochex, J.Y. (1992). *École et savoir dans les banlieues… et ailleurs*. A. Colin.

Joigneaux, C. (2009). « La construction des inégalités scolaires dès l'école maternelle ». *Revue Française de Pédagogie*, 169, pp. 17-28

Joigneaux, C. (2016). L'évaluation dans les écoles maternelles françaises. Quelles cohérences pour quels effets ? *In* C. Veuthey, G. Marcoux, T. Grange (dir.), *L'école première en question. Analyses et réflexions à partir des pratiques d'évaluation,* EME éditions, pp. 39-54.

Joigneaux, C. (2019). *Culture et inégalités scolaires dès l'école maternelle. Une approche ethno-historique*. Mémoire d'habilitation à diriger des recherches. Université de Paris 8.

Joigneaux, C. & Rockwell, E. (2020). Cause or consequence? Framing and keying interaction in a French secondary classroom. *Learning, Culture and Social Interaction*, 24

Lahire, B. (1993) *Culture écrite et inégalités scolaires*. Presses universitaires de Lyon.

Lahire, B. (2005). Fabriquer un type d'homme « autonome » : analyse de dispositifs scolaires. *In L'esprit sociologique*, La Découverte, pp. 322-347.

Netter, J. & Joigneaux, C. (2022). Les pédagogies de l'autonomie, entre dispositifs et pratiques. L'exemple d'une classe de maternelle française. *Carrefours de l'éducation*, à paraitre.

Rochex, J.Y. & Crinon, J. (dir.) (2011). *La construction des inégalités scolaires. Au cœur des pratiques et des dispositifs d'enseignement*. Presses universitaires de Rennes.

Comment peut-on définir les attentes évaluatives en classes ? Une synthèse des réflexions en EPS

LÉA GOTTSMANN

ÉCOLE NORMALE SUPÉRIEURE DE RENNES (FRANCE)

Introduction

La question de l'évaluation à l'école est depuis longtemps un sujet de débats et de tensions, évoluant au regard des finalités de l'école et de sa contribution aux enjeux éducatifs. Longtemps perçue comme un moyen de sélection ou de sanction des apprentissages réalisés, de nombreuses propositions se développent aujourd'hui pour chercher à rendre l'évaluation utile pour les apprentissages des élèves. Ces tentatives cherchent à dépasser l'opposition traditionnelle entre une évaluation formative et une évaluation certificative, mais aussi à réduire la place des notes pour impulser de nouveaux paradigmes d'évaluation (Merle, 2015). En effet, selon cet auteur, l'évaluation par notes est issue d'une tradition des concours des grandes écoles, créée dans un objectif de forte sélection, ne correspondant pas toujours aux enjeux scolaires : « la logique du concours, hiérarchiser et classer, n'entretient pas de rapport avec les apprentissages scolaires » (*op. cit.*, p. 86).

En éducation physique et sportive (EPS), les formes d'évaluation ont aussi évolué, passant d'une évaluation unique de la performance vers des évaluations qui cherchent à prendre en compte l'élève de façon plus qualitative et personnalisée (Attali & St Martin, 2010). C'est notamment le développement de référentiels qui évaluent les processus d'apprentissage, les progrès et pas uniquement une performance lors d'une production finale. L'objectif est ainsi de pouvoir élargir le spectre d'évaluation de l'élève de telle sorte que « le sens de l'initiative, de l'autonomie, la capacité à prendre des responsabilités, la qualité du travail fourni font désormais partie des critères officiels de réussite » (Attali & St Martin, 2010, p. 62). Cette évolution va dans le sens de valoriser et d'évaluer des compétences méthodologiques et sociales qui doivent contribuer à la formation d'un citoyen qui sera physiquement actif dans sa vie future. En effet, « au regard des enjeux de société, l'EPS contribue à développer une culture de l'activité physique régulière et durable, levier indispensable de l'amélioration de la santé publique » (BO spécial n°1 du 22 janvier 2019).

L'introduction d'une approche par compétences s'inscrit particulièrement dans cette démarche de valoriser les progrès, l'engagement, le processus d'apprentissage ; et pas uniquement la performance individuelle. Mais plusieurs freins et tensions apparaissent au sein de la communauté éducative, certains pouvant être qualifiés d'obstacles épistémologiques et idéologiques à l'introduction du concept de compétence dans les pratiques (Gottsmann & Delignières, 2016). C'est notamment la difficulté à sortir d'une conception analytique des apprentissages et des « tâches » d'évaluation, sans prendre en compte la complexité de l'activité d'un élève et des situations plus proches de la vie réelle. En effet, une tâche est caractérisée par un but précis à atteindre, au cours de laquelle l'enseignant

attend une réponse fermée ou un comportement unique. Une telle conception amène à évaluer de façon segmentée ou artificielle les différentes composantes d'une compétence que sont les connaissances, les attitudes et les capacités dans des situations fermées ou séparées. Issue d'une perspective cognitiviste, cette conception amène l'enseignant à isoler les apprentissages les uns des autres mais ne « garantit pas que l'élève saura les utiliser à bon escient dans une situation réelle et complexe » (Terré, Sève & Saury, 2016, p. 72), ce qui est pourtant attendu dans une évaluation des compétences. Cette vision analytique des apprentissages peut être dépassée par les conceptions enactives de la compétence qui évoquent l'idée de définir un « agir compétent » par rapport aux situations dans lesquelles il agit (Masciotra & Medzo, 2009). En s'appuyant sur le cadre théorique de l'enaction qui pose l'hypothèse d'un couplage indissociable entre l'acteur et son environnement, l'élève et sa situation ne doivent pas être pensés de façon isolée ou séparée. Un élève compétent est donc un élève « qui parvient à reconnaître le potentiel offert par sa situation et à l'exploiter », et non pas un élève qui résout un problème de façon isolée et déconnectée (Terré *et al*, 2016). Partant de l'idée qu'une compétence ne peut être définie en dehors d'une famille de situations, l'évaluation des compétences doit donc se faire dans ce cadre, au sein de situations qui sont familières pour les élèves afin de favoriser la continuité des apprentissages et des expériences. La relation entre l'élève et son environnement est remise au cœur de l'activité : l'un ne peut être pensé sans l'autre et l'évaluation de l'élève doit se penser de façon contextualisée, et non plus de façon séparée entre les apprentissages et l'évaluation (Paintendre, Terré & Gottsmann, 2021).

Au regard de ces éléments introductifs, cette note cherche à appréhender certains points de tension en EPS, en lien notamment avec la spécificité de l'évaluation dans cette discipline : a) la relation forte aux pratiques sociales de références qui impacte les procédures d'évaluation ainsi que le sens donné à l'activité des élèves ; b) la tendance à séparer l'évaluation des dimensions motrices, méthodologiques et sociales dans l'activité des élèves ; c) la question du rapport entre ce qui est enseigné en EPS, et donc ce qui est censé être évalué, en lien avec la perception par les élèves de ces situations d'apprentissage et d'évaluation. Des résultats issus de la littérature scientifique et des propositions professionnelles seront mobilisés afin de mieux comprendre les enjeux relatifs aux procédures d'évaluation en EPS, et plus globalement dans la sphère scolaire.

I. Relation aux pratiques sociales de référence

Une des principales spécificités de l'EPS en tant que discipline scolaire renvoie à son rapport aux pratiques sociales de référence, c'est-à-dire les activités physiques, sportives et artistiques sur lesquelles elle s'appuie dans son enseignement et dans son évaluation (par exemple, les activités athlétiques, les danses, les activités de pleine nature, les sports collectifs). Le rapport à ces pratiques fonde l'identité de la discipline, mais engendre aussi des confusions entre la pratique du « sport » en clubs ou en associations, et les pratiques en EPS à l'école. L'image du « sportif » reste encore centrale et peut parfois être considérée comme un frein à l'évolution des formats d'évaluation qui rendraient compte d'une activité plus scolaire des élèves, en dehors d'une unique performance motrice. Plusieurs études mettent d'ailleurs en évidence que les évaluations restent majoritairement centrées sur des aspects de performances ou de connaissances puisqu'elles sont perçues par les enseignants comme laissant moins de place à la subjectivité et qu'elles sont plus faciles à mettre en place (Lopez-Pastor *et al*., 2013 ; Mougenot, 2013). C'est d'autant plus vrai pour des activités d'opposition ou des sports collectifs où les enseignants se sentent plus en difficulté pour mettre en place des critères d'évaluation objectifs. En effet, un ensemble de paramètres entrent en compte dans l'évaluation des élèves : les

choix opérés par les élèves de façon individuelle ou collective dans le jeu, les habiletés motrices et techniques de jeu, l'investissement au sein d'une équipe ou dans des rôles sociaux, etc. Plus encore, la possibilité d'évaluer d'autres compétences, l'investissement ou le progrès des élèves fait partie des préoccupations majeures dans l'activité d'évaluation des enseignants (Brau-Antony & Hubert, 2014), mais implique pour eux un ensemble de difficultés. Prendre en compte d'autres dimensions de l'activité d'un élève, par rapport à la seule dimension motrice et performative, peut être perçu comme plus subjectif, plus difficile à évaluer et moins équitable pour les élèves, ou relevant d'autres dimensions que des apprentissages réalisés dans le cadre de l'EPS.

Une deuxième tension centrale pour les enseignants à propos de l'évaluation concerne l'adaptation des référentiels nationaux, qui impulsent un ensemble de directives et de prescriptions, par rapport aux spécificités locales d'un établissement, d'une équipe éducative et des besoins des élèves. C'est l'une des préoccupations majeures des enseignants dans l'évaluation en EPS, particulièrement dans le cadre des référentiels d'évaluation certificatifs où les enseignants se dotent « d'une certaine autonomie qui leur permet de prendre des distances par rapport aux prescriptions officielles » (Brau-Antony & Hubert, 2014, p. 14). Il existe donc une certaine hétérogénéité dans les pratiques d'évaluation, dépendant d'un ensemble de critères et de spécificités définis au niveau local, selon les conceptions des enseignants, la dynamique éducative de l'établissement et les profils des élèves. Cette hétérogénéité est accentuée par une évolution récente dans les programmes d'EPS : les modalités d'évaluation sont désormais définies au niveau national par des « attendus de fin de cycle » qui sont ensuite déclinés et adaptés au niveau local par les équipes éducatives. C'est dans ce cadre que le développement de formes scolaires de pratiques tente de répondre à un équilibre entre la référence aux pratiques scolaires de référence et la réponse aux enjeux scolaires. Les équipes éducatives sont alors amenées à définir des formes scolaires de pratique, qui représentent la définition au niveau local du cadrage national et académique, en fonction des spécificités du contexte d'enseignement. Ces formes scolaires de pratique doivent être « porteuses du fond culturel des activités physiques sportives et artistiques » tout en étant porteuses « de contenus d'enseignement prioritaires définis par l'enseignant » (BO spécial n°1 du 22 janvier 2019). Les programmes disciplinaires encouragent d'ailleurs les équipes pédagogiques à définir des formes scolaires de pratique communes à l'établissement dans le projet pédagogique d'EPS, afin de permettre une continuité dans les apprentissages et dans les modalités d'évaluation. Certains y voient une vraie plus-value dans une adaptation plus importante à l'hétérogénéité des élèves dans les établissements, afin de leur proposer des formes de pratique et des évaluations correspondant mieux à leurs profils, sans pour autant perdre le fond culturel de l'activité support et les exigences institutionnelles. C'est aussi une volonté institutionnelle de favoriser une dynamique de travail au sein de l'équipe pédagogique dans la conception de référentiels communs entre les enseignants. Pour d'autres, cette adaptation locale peut à l'inverse contribuer aux inégalités de réussite sur le territoire national, tout autant que la généralisation du contrôle continu aux épreuves nationales par exemple.

II. La tendance à séparer l'évaluation des dimensions motrices, méthodologiques et sociales des compétences

La tendance à segmenter l'évaluation des différentes dimensions de la compétence vient d'une conception réductionniste et dualiste de l'apprentissage, considérant le corps séparé de l'esprit (Paintendre *et al*, 2021). Cette tendance se caractérise en EPS par une distinction entre les composantes motrices (ce qui relève de la technique, des habiletés motrices), méthodologiques

(dimensions réflexives sur l'activité par exemple) et sociales (relationnelles) (Saury, Adé, Gal-Petitfaux, Huet, Sève & Trohel, 2013). Ce sont notamment les programmes qui continuent de définir les apprentissages et leur évaluation de cette façon, définissant les compétences par exemple par le triptyque connaissances, capacités, attitudes. Plus encore, l'évaluation au baccalauréat continue dans cette voie avec une répartition de la note sur des attendus de fin de lycée (AFL) « moteurs » et des attendus plus méthodologiques et sociaux. Ainsi, la note en EPS est répartie de la façon suivante : 12 points concernant l'AFL1 ; et 8 points concernant les AFL2 et AFL3. Pour ces 8 points, les élèves peuvent choisir la répartition qu'ils souhaitent, à condition que chaque AFL soit évalué au minimum sur 2 points. Par exemple dans le cadre de l'activité vitesse-relais[90] l'AFL1 renvoie à « s'engager pour produire une performance maximale à l'aide de techniques efficaces » ; l'AFL2 correspond à « s'entraîner, individuellement et collectivement pour réaliser une performance » ; et enfin l'AFL3 qui correspond à « choisir et assumer les rôles qui permettent un fonctionnement collectif solidaire ». Les élèves sont donc amenés à réaliser plusieurs choix, dans la répartition chiffrée de leur évaluation et dans les rôles qu'ils souhaitent mener (starter, chronométreur, partenaire ou coach par exemple). Même si la question du choix par les élèves peut être intéressante pour mieux prendre en compte leurs spécificités et leurs besoins, cette segmentation de l'activité des élèves continue, en évaluant de façon différenciée ce qui est de l'ordre du moteur, du méthodologique et du social, alors qu'ils sont imbriqués dans la pratique d'un élève (Paintendre *& al*, 2021).

Plus encore, on peut observer de nombreuses propositions professionnelles qui cherchent à développer des situations « complexes » pour l'apprentissage et l'évaluation des compétences. Ces situations complexes amènent les élèves à faire des choix, à manipuler des variables en respectant quatre principes (Roegiers, 2011) : a) la mobilisation d'un ensemble de ressources de façon combinée ; b) la finalisation de la situation par une dimension sociale et culturelle ; c) la confrontation à un problème global ; d) une part d'inédit. Certaines études en EPS montrent l'intérêt d'un dispositif d'apprentissage par situations complexes en comparaison avec un enseignement plus traditionnel (Le Briquer, 2017). En effet, dans les situations complexes, les progrès en natation de vitesse sont plus importants pour les élèves, ainsi qu'un développement des interactions sociales. Pour d'autres auteurs, le risque d'aller vers une complexité à l'extrême des dispositifs d'apprentissage et d'évaluation des compétences est réel (Gottsmann, 2019). Certains alertent notamment sur la place importante donnée à l'activité réflexive dans ces situations, pouvant amener à une possible augmentation des inégalités des élèves, encore plus que dans les approches plus traditionnelles des apprentissages (Chenu, Crahay & Lafontaine, 2014). Plus encore, la question de présenter des situations « inédites » lors de l'évaluation pour favoriser le réinvestissement des apprentissages interroge par rapport à la dimension familière de la compétence et des situations censées la révéler.

Certaines propositions professionnelles cherchent donc à donner du sens à ces différentes dimensions de la compétence en développant des outils d'évaluation qui viennent à la fois rendre compte de la compétence attendue par l'enseignant, et aider l'élève dans son processus d'apprentissage. L'objectif est de faire plus de liens entre les apprentissages et les modalités d'évaluation, mais aussi de permettre aux élèves de mieux se connaître pour une pratique à plus long terme en tant que futur citoyen. Ce sont par exemple les propositions du groupe « ÉPIC » de l'AEEPS[91] qui cherchent à développer des outils d'évaluation de la compétence que les élèves s'approprient au fil de leur activité. Une

[90] Fiche d'illustration en vitesse-relais : https://eduscol.education.fr/1748/programmes-et-ressources-en-eps-voie-gt

[91] ÉPIC : Évaluation Par Indicateurs de Compétences, Groupe ressource de l'Association pour l'Enseignement de l'Éducation Physique et Sportive (www.aeeps.org)

proposition en demi-fond cherche par exemple à rééquilibrer l'évaluation unique de la performance relative, c'est-à-dire par rapport à soi et l'évaluation de la performance brute et normée, qui ne prend pas en compte l'hétérogénéité des élèves (Testud & Rossi, 2017). Les auteurs cherchent ainsi à développer une « intentionnalité de performance » chez tous les élèves, malgré leurs différences en termes de capacité à produire et à supporter des efforts. La forme de pratique se caractérise par des courses en peloton autour d'un travail technique sur les allures et sur la position de « lièvre », dans des groupes de courses homogènes. Ainsi, selon le niveau de course des élèves, des distances différentes sont proposées, ainsi que deux modalités de courses (3 et 9 minutes) pour sensibiliser aux changements de rythme et au fait de vivre des efforts intenses. Dans le cadre de l'évaluation, plusieurs scores sont envisagés pour donner des indicateurs sur le développement de la compétence attendue : un score de « performance » renvoyant à l'écart entre le lièvre et les autres coureurs ; un score individuel renvoyant à l'écart entre les vitesses sur les différentes courses. Les auteurs défendent cet outil d'évaluation comme une « feuille de route » tout au long de la séquence pour une évaluation cohérente avec l'expérience souhaitée pour les élèves, sans segmenter l'activité motrice, méthodologique et sociale. De plus, cet exemple permet de mettre en évidence des formats d'évaluation qui permettent de prendre en compte la spécificité des élèves, par exemple entre les filles et les garçons, mais aussi par rapport aux élèves à besoins particuliers. Cette question nécessite d'être réellement prise en charge par l'EPS dans le cadre scolaire : c'est la seule discipline pour laquelle certains élèves sont déclarés inaptes toute ou une partie de leur scolarité, sans avoir accès aux apprentissages visés et aux évaluations.

III. Ce qui est enseigné en EPS et donc ce qui est censé être évalué

La question de l'évaluation porte également sur ce qui est évalué par rapport à ce qui est enseigné ou visé en termes d'apprentissage. Comme d'autres disciplines, certains interrogent le fait que l'EPS évalue majoritairement des dispositions acquises en dehors de l'école, plutôt que des apprentissages réalisés dans le cadre de l'EPS (Brau-Antony & Hubert, 2014). La complémentarité entre ce qui est appris et ce qui est évalué est important à envisager, notamment lorsque l'on s'intéresse au sens que les élèves donnent aux situations d'évaluation. En effet, les enjeux dans les situations d'évaluation « sont perçus comme significatifs et importants dans l'expérience des élèves » (Saury *et al*, 2013, p. 92) et cette perception impacte fortement leur mode d'engagement et leurs apprentissages. L'analyse de l'expérience des élèves permet de mettre en évidence des modalités d'engagement antagonistes selon si la situation est perçue comme évaluée ou non évaluée. Plus précisément, Saury & Rossard (2009) ont analysé les préoccupations des élèves au cours de différentes situations, coopératives ou compétitives en badminton ainsi que la place de la perception des enjeux évaluatifs en leurs seins. Les résultats révèlent que les enjeux d'évaluation perçus par les élèves sont très prégnants dans trois types de situations : a) lorsque les situations sont explicitement présentées par l'enseignant comme évaluatives (attribution d'une note) ; b) lorsque les situations amènent à un classement entre les élèves, sans forcément être évaluées ; c) lorsque les situations intègrent des dispositifs de co-observation entre eux. Cette perception évaluative par les élèves va même parfois à l'encontre de préoccupations plutôt tournées la recherche de plaisir ou d'intérêt dans la situation, pourtant favorables à un engagement plus durable dans la pratique. La question de la perception par les élèves des enjeux évaluatifs et du sens qu'ils y donnent doit donc être appréhendée pour mieux envisager les relations entre les apprentissages, leur évaluation et ce qui sera utile pour les élèves dans leur future vie de citoyen.

Au-delà de la perception des enjeux évaluatifs par les élèves, d'autres études sur l'analyse de l'expérience des élèves mettent en évidence des décalages entre l'exploitation par les élèves des artefacts dans des situations d'apprentissage prévues par l'enseignant lorsqu'ils cherchent à développer et évaluer des compétences (Gottsmann, Trohel & Gal-Petitfaux, 2021). C'est par exemple le fait de mettre en place une règle de 10 points bonus lorsqu'un binôme d'élèves remporte le point en tennis de table après avoir pris un temps mort : l'objectif de l'enseignant est d'encourager chez les élèves des interactions coopératives au profit du gain du match, qui seront évaluées comme composantes de la compétence visée. Pourtant, lorsque l'on s'intéresse à l'expérience vécue des élèves dans cette situation, leurs préoccupations sont tournées de façon majoritaire vers le fait de marquer 10 points, peu importe la façon de faire. Plus encore, lorsque l'enseignant passe au sein des groupes dans la classe, les élèves « masquent » cette activité pour montrer qu'ils répondent aux exigences scolaires de l'enseignant (ils font croire à l'enseignant qu'ils se sont donnés une consigne utile), sans pour autant y donner du sens ou construire ce que l'enseignant vise en termes d'apprentissage. Cet exemple révèle donc aussi tous les moments d'évaluation plus implicites au cours des leçons que les élèves perçoivent et qui impactent leur engagement et ce qu'ils apprennent. Il semble donc important d'interroger le sens que les élèves donnent aux dispositifs mis en place par l'enseignant par rapport à ce qui est évalué et par rapport à ce que l'on cherche à développer chez eux. C'est notamment la question sous-jacente d'évaluer des compétences « scolaires », c'est-à-dire que les élèves soient capables de répondre à des exigences fixées par l'enseignant, ou bien d'évaluer des compétences qui sont signifiantes pour les élèves dans la formation d'un citoyen et dans leur vie future.

Conclusion

Cette note permet de proposer un aperçu rapide des évolutions qui ont eu lieu dans les pratiques d'évaluation en EPS, de la relation aux pratiques sociales de référence, vers l'évaluation des différentes composantes d'une activité en lien avec son environnement et enfin de la question de ce qui est évalué par rapport à ce qui est enseigné. Même si les propositions sont nombreuses aujourd'hui pour chercher à faire plus de liens entre des évaluations formatives et certificatives et pour que les évaluations prennent plus de sens pour les élèves et soient plus lisibles, il semble qu'il reste encore des points de tensions à un réel changement paradigmatique en matière d'évaluation. Certains de ces obstacles se posent particulièrement en EPS, par rapport à d'autres disciplines, en raison des spécificités qui ont été présentées. D'autres font écho à d'autres disciplines, par exemple la spécificité scolaire par rapport aux pratiques sociales de référence, ou encore par rapport aux apprentissages réalisés en dehors de l'école. Au-delà de ces points de tension, il semble nécessaire aussi de s'interroger sur la pression évaluative que les élèves subissent tout au long de leur parcours, et au quotidien dans les classes, qui peut nuire – comme nous l'avons vu – à leur expérience scolaire ou aux apprentissages visés par les enseignants. S'interroger sur le sens et la place que l'évaluation prend dans l'expérience des élèves, en lien avec ce qu'ils perçoivent, ce qu'ils ont appris et ce qu'ils retiendront pour leur vie future, semble essentiel au regard de la littérature à ce sujet.

Références

Attali, M. & St Martin, J. (2010). L'évaluation en EPS : entre légitimité disciplinaire et défis culturels (1959-2009), *Les sciences de l'éducation – Pour l'Ere nouvelle*, 43, 55-81.

Brau-Antony, S. & Hubert, B. (2014). Curriculum en EPS et évaluation certificative au baccalauréat, *Questions vives*, 22.

Chenu, F., Crahay, M. & Lafontaine, D. (2014). Par-delà l'approche par compétences : quelle place réserver aux savoirs, à leur enseignement et à leur évaluation ?, *Education & Formation*, e-302, 17–29.

Gottsmann, L. (2019). *La construction d'un agir compétent en Éducation Physique: analyse de l'expérience des élèves et du processus de typification des connaissances.* Université Clermont Auvergne, Thèse de doctorat non publiée.

Gottsmann, L., & Delignières, D. (2016). A propos des obstacles épistémologiques à l'émergence du concept de compétence, *Movement & Sport Sciences*, 4, 71-81.

Gottsmann, L., Trohel, J. & Gal-Petitfaux, N. (2021). The role of artefacts in developing competences in physical education: exploring students' experience and knowledge built in action, *Physical Education and Sport Pedagogy*, 1-15.

Le Briquer, Y. (2017). *L'apprentissage par situation complexe en EPS au collège, rôles de la coopération et de la compétition en natation de vitesse chez des élèves de 5ème.* Université Bordeaux, Thèse de doctorat non publiée.

Lopez-Pastor, V.M., Kirk, D., Lorente-Catalan, E., MacPhail, A. & Macdonald, D. (2013). Alternative assessment in physical education: a review of international literature, *Sport, Education and Society*, 18, 1, 57-76.

Masciotra, D., & Medzo, F. (2009). *Développer un agir compétent : vers un curriculum pour la vie*. De Boeck.

Merle, P. (2015). L'école française et l'invention de la note. Un éclairage historique sur les polémiques contemporaines, *Revue française de pédagogie*, 193, 77–88

Mougenot, L. (2013). *L'évaluation scolaire en question : de l'activité des enseignants aux conduites des élèves : impact des évaluations proposées et des modalités de groupement sur les conduites des élèves du secondaire en éducation physique et sportive.* Université René Descartes, Paris V, Thèse de doctorat non publiée.

Paintendre, A., Terré, N. et Gottsmann, L. (2021). Vers une conception holiste de l'activité de l'élève et de ses apprentissages : repenser la relation à son corps et à son environnement ? In T. Froissart, A. Paintendre & J. Saint-Martin (Eds.), *L'EPS du XXIe siècle ou les enjeux d'une EP de qualité (1981-2021)*. éPUre

Roegiers, X. (2011) *Des situations pour intégrer les acquis scolaires*. De Boeck Supérieur.

Terré, N., Sève, C., & Saury, J. (2016). Une approche énactive du développement des compétences en milieu scolaire. *Education et Francophonie, 44(2)*, 68-85.

Testud, E., & Rossi, D. (2017). Courir au record ou maîtriser un affrontement ? Une modalité de pratique du demi-fond et un projet de formation des élèves. *Enseigner l'EPS, 271*, 19-25.

Saury, J., Adé, D., Gal-Petitfaux, N., Huet, B., Sève, C. & Trohel, J. (2013). *Actions, significations et apprentissages en EPS. Une approche centrée sur les cours d'expériences des élèves et des enseignants*. Revue EP&S.

Saury, J., & Rossard, C. (2009). Les préoccupations des élèves durant des tâches d'apprentissage coopératives et compétitives en badminton : une étude de cas, *Revue des Sciences de l'Education*, *35*(3), 195-216.

Quels apports de la didactique pour penser l'évaluation ? L'exemple des mathématiques

JULIE HOROKS ET JULIA PILET

UNIVERSITÉ PARIS-EST CRÉTEIL (FRANCE)

Introduction

Le document d'accompagnement relatif à l'évaluation des mathématiques au cycle 3 (CM1-CM2-6e) (*Ressources pour l'évaluation en mathématiques*, 2016[92]) rappelle que la dimension sommative de l'évaluation (au terme d'un processus d'enseignement) doit s'articuler à une stratégie d'évaluation plus large, comportant d'autres volets, tels que l'évaluation diagnostique notamment (au début d'une séquence d'enseignement) et l'évaluation formative (en cours d'activité). Concernant ces deux derniers volets, le document d'accompagnement intitulé *Différenciation pédagogique*[93] (2016) précise que « le repérage, l'identification et le traitement des blocages et des erreurs constituent à la fois pour l'élève un levier pour progresser dans ses apprentissages et pour le professeur un appui pour réguler son enseignement » (page 5). Au-delà de ces injonctions, comment accompagner les enseignants à pratiquer une évaluation au service des apprentissages ? Nous proposons ici des outils théoriques pour décrire les pratiques d'évaluation que nous opérationnalisons pour, d'une part, dresser un panorama des pratiques d'évaluation des enseignants du secondaire français en mathématiques et, d'autre part, pour proposer des leviers de formation à travers la question de l'évaluation.

I. Des recherches sur l'évaluation

Depuis les vingt dernières années, l'évaluation est une thématique que les recherches en didactique des mathématiques investissent dans diverses entrées : évaluations nationales et internationales et leurs comparaisons (Bodin, 2006 ; Artigue & Winslow, 2010 ; Roditi & Salles, 2015, Sayac & Grapin, 2016), évaluation diagnostique et parcours d'enseignement différencié (Grugeon, 1997 ; Grugeon-Allys, Pilet, Chenevotot, Delozanne, 2012 ; Pilet, 2015 ; Grugeon-Allys, 2016, 2018), évaluation formative (Coppé & Moulin, 2017, Chanudet, Coppé, Gandit, Moulin, 2018), évaluation des compétences (Winslow, 2005), pratiques évaluatives (Sayac, 2017 ; Horoks, Kiwan, Pilet, Roditi, Haspekian, 2018 ; Chanudet, 2019, Horoks, 2022). Deux projets de grande ampleur ont permis des

[92] https://eduscol.education.fr/document/16528/download

[93] https://eduscol.education.fr/document/17197/download

avancées significatives : un projet européen *ASSIST-ME*[94] (2013-2017) et un projet financé par l'Agence nationale de la recherche *Néopraéval*[95] (2014-2018) dans lequel l'évaluation en mathématiques était abordée à travers la conception de dispositifs d'évaluation et leurs usages dans les pratiques enseignantes. À la lumière de ces avancées, cette contribution s'inscrit dans une réflexion sur les pratiques enseignantes, principalement du secondaire, autour des apports potentiels de l'évaluation pour faire apprendre les mathématiques.

A. Définitions de l'évaluation : quels liens avec les apprentissages ?

Quelle que soient les fonctions données à l'évaluation (Rey & Feyfant, 2014), « évaluer » peut être défini par : « recueillir un ensemble d'informations suffisamment pertinentes, valides et fiables, examiner le degré d'adéquation entre cet ensemble d'informations et un ensemble de critères adéquats aux objectifs fixés au départ ou ajustés en cours de route, en vue de prendre une décision. » (De Ketele, Gerard, Roegiers, 1997, p. 34).

L'évaluation formative, qui nous intéresse plus particulièrement ici, est définie dans de nombreux travaux comme une évaluation *pour* l'apprentissage, en opposition à une évaluation *de* l'apprentissage, souvent dominante dans les systèmes éducatifs. Sa définition a été élargie au cours du temps (Allal & Mottier Lopez, 2005). Elle a d'abord été considérée comme un moment d'enseignement spécifique qui a lieu après une phase d'enseignement, on y a inclus par la suite également les interactions menées de manière plus informelle et pouvant être intégrées dans chaque activité d'enseignement et d'apprentissage. Pour Black et Wiliam (1998), l'évaluation est formative lorsque les informations recueillies par l'enseignant sont utilisées pour répondre aux besoins des élèves et lorsqu'en retour l'élève s'engage dans la tâche[96] et peut s'auto-évaluer. Quant à Ash et Levitt (2003), ils soutiennent que l'évaluation formative est une interaction entre l'enseignante et l'élève, qui s'appuie sur ce que l'élève sait déjà faire. L'enseignante recueille des indices de l'activité de l'élève, dans le but de les analyser et de prévoir l'étape suivante pour aider l'élève à évoluer. Pour Shavelson, Young, Ayala, Brandon, Furtak, Ruiz-Primo et Yin (2008) enfin, l'évaluation formative peut prendre plusieurs formes selon qu'elle s'intègre dans des évaluations formelles et planifiées, au niveau global, ou dans d'autres, moins formelles et dans l'interaction, ou encore dans celles qui peuvent arriver « *on the fly* » de façon spontanée, au niveau micro ou local, à l'initiative du professeur ou d'un élève.

B. La place des savoirs dans l'évaluation

Dans notre approche de l'évaluation par prise d'informations et de décisions, nous accordons une place centrale aux contenus mathématiques. Selon Bain (1988) pour l'enseignement du français ou Bodin (1997) pour les mathématiques, pour assurer sa fonction d'amélioration de l'apprentissage, l'évaluation formative ne peut être pensée indépendamment des savoirs enseignés. Outiller les enseignants pour qu'ils dépassent des catégories d'élèves « bons, moyens, faibles », souvent

---

[94] . Le projet européen « *Assess Inquiry in Science, Technology and Mathematics Education* » (ASSIST-ME,2013-2017) étudie la mise en œuvre combinée de l'évaluation sommative et formative dans un enseignement de sciences expérimentales, de technologie et de mathématiques fondé sur les démarches d'investigation. https://cordis.europa.eu/project/id/321428

[95] Projet NéoPraéval « Nouveaux Outils pour de nouvelles pratiques d'évaluation et d'enseignement des mathématiques » https://www.ldar.website/neopraeval

[96] Dans le champ scolaire, pour un élève, une tâche est ce qu'il a à faire, souvent dans des énoncés proposés par l'enseignant, et avec des consignes de travail mathématique associées, alors que l'activité est ce qu'il va mettre en œuvre, de façon visible et invisible, pour effectuer cette tâche, ce qu'il fait ou ne fait pas, ce qu'il dit ou ne dit pas. Cette distinction va évidemment à l'encontre de formulations courantes qu'on trouve dans les manuels scolaires telles que « activités d'introduction » ou « activités d'entraînement », ou de celles utilisées dans les programmes.

indépendantes de ce qui est enseigné, nécessite de les accompagner dans le repérage et surtout la caractérisation des besoins d'apprentissage des élèves dans la discipline ou le domaine qu'ils enseignent.

L'évaluation formative dépend de la façon dont l'enseignant s'appuie sur les activités mathématiques effectives visibles des élèves (tout ce que les élèves font en lien avec la résolution de l'exercice) sur des tâches choisies pour faire apparaître certains savoirs et savoir-faire des élèves. Un tel appui implique de recueillir des informations fiables sur ces activités et d'en faire une exploitation pertinente pour les apprentissages. Cela suppose non seulement une anticipation des procédures susceptibles d'être mises en œuvre, mais aussi, une gestion de la classe prenant en compte, dans les connaissances ou les procédures repérées chez les élèves, les spécificités du savoir en jeu. Cette gestion passe également par les échanges, collectifs et individuels, entre l'enseignant et les élèves, ce que les travaux sur l'évaluation formative associent souvent à la notion de feedback. Là encore, les spécificités du savoir jouent un rôle central.

C. Les feedbacks et l'évaluation formative

Dans leur revue de littérature anglo-saxonne, Georges et Pansu (2011) s'intéressent au rôle des feedbacks scolaires pour renforcer la motivation et soutenir la réussite des élèves, mais leur efficacité du point de vue de la régulation des processus d'apprentissage peut être questionnée (Perrenoud, 1998). Un feedback de type « réponse incorrecte » ne semble pas suffisant pour permettre à l'élève de comprendre la distance entre ce qu'il a produit et ce qui est attendu. Ou encore, une formulation d'ordre légal, du type « on a le droit de », faible mathématiquement, semble être un feedback de portée limitée pour amener les élèves à mettre du sens et à ne pas réduire les mathématiques à un ensemble de règles à appliquer (en algèbre, « on n'a pas le droit d'additionner des nombres et des lettres » pour l'erreur « 3+x est égal à 3x »). Des feedbacks avec une portée générale, appuyés par les propriétés mathématiques, sembleraient plus formatifs.

D. Définir des pratiques d'évaluation en mathématiques

Compte tenu de l'approche choisie, les « pratiques d'évaluation » des apprentissages des élèves en mathématiques peuvent être définies par tout ce qui dans la pratique de l'enseignant participe à :

- une prise d'informations sur les activités mathématiques des élèves et/ou les connaissances mathématiques que ces activités mobilisent, qui peuvent être repérées à différents niveaux : dans les résultats produits par les élèves ou les procédures qui les ont menés à ces résultats, à travers la résolution d'un ensemble de tâches ;

- une interprétation de ces informations, en référence à des éléments qui peuvent être variés et plus ou moins locaux (les réponses attendues par l'institution à un niveau donné, ou par l'enseignant dans un établissement donné, voire pour une classe ou un élève particulier), mais aussi en référence aux savoirs mathématiques visés et à la façon de les mobiliser avec plus ou moins d'expertise dans ces tâches ;

- une exploitation enfin de ces informations, à plus ou moins court terme (selon que cela infléchit ou non le projet d'enseignement, immédiatement ou lors de futures séances), donc à la fois à travers l'organisation des interactions avec les élèves, l'élaboration du discours mathématique porté sur leurs productions pour les valider ou les invalider, les liens tissés entre ces productions et les savoirs visés pour les situer par rapport à ce qui est attendu, mais aussi plus globalement la

régulation des apprentissages (Allal & Mottier Lopez, 2005 ; Mottier Lopez, 2012) pour l'ensemble de la classe ou pour certains élèves en particulier. La prise de décision menant à cette exploitation des informations recueillies peut être différée de la prise d'informations. Elle peut aussi ne pas intervenir.

II. Les pratiques d'évaluation dans l'enquête nationale à grande échelle Praesco

En France, l'enquête Praesco (Enquête sur les pratiques enseignantes spécifiques aux contenus), menée la Depp (Direction de l'évaluation, de la prospective et de la performance) en collaboration avec une équipe de chercheurs et d'enseignants, a permis de dresser un panorama[97] des pratiques déclarées d'enseignement de collège en mathématiques[98]. L'enquête repose sur un questionnaire qui a été soumis à un échantillon représentatif d'enseignants de mathématiques de troisième. La partie de l'enquête qui porte sur l'évaluation repose surtout sur des items portant spécifiquement sur les pratiques d'évaluation diagnostique (17 items) ou sommative (36 items). Les questions relatives à l'évaluation formative sont quant à elles réparties dans l'ensemble du questionnaire, et moins identifiables par les enseignants comme des questions relevant de l'évaluation. Ces questions concernent la gestion de certaines erreurs, le caractère mathématique des feedbacks proposés aux élèves, ou encore l'implication plus ou moins grande des élèves dans certaines interactions relevant de l'évaluation formative.

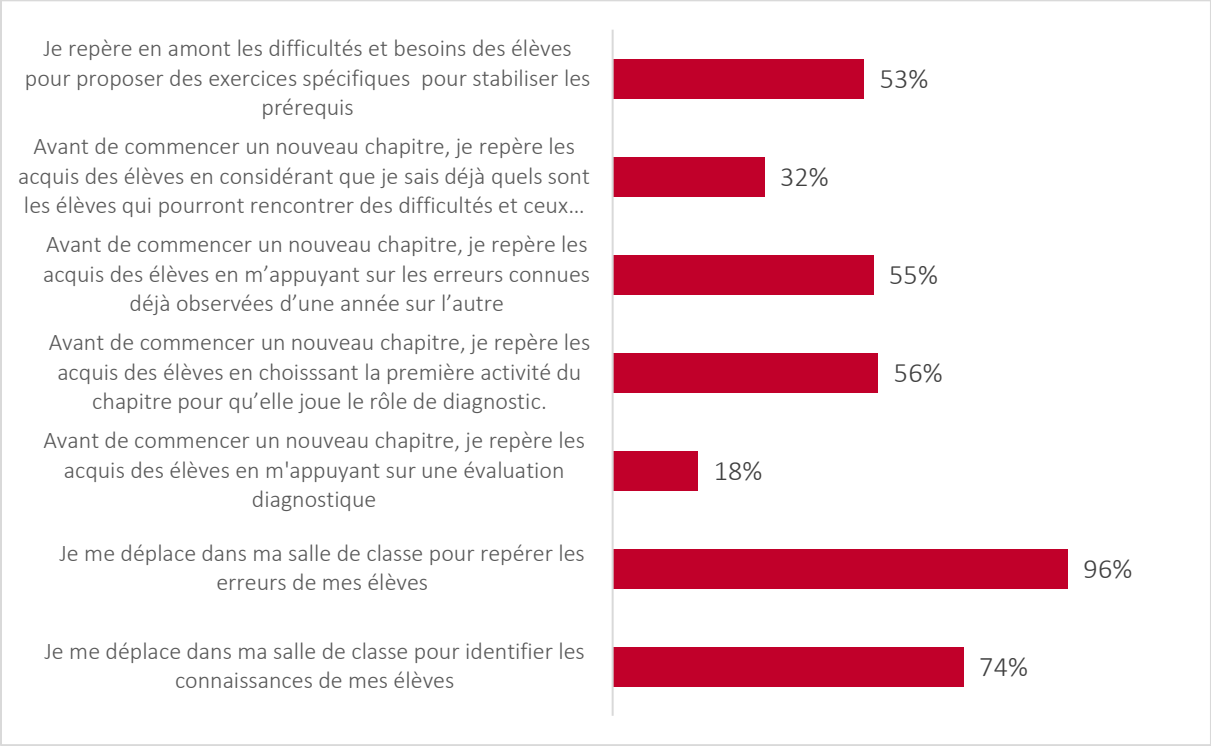
A. L'évaluation diagnostique

L'évaluation diagnostique visant à repérer des prérequis, des conceptions et des erreurs des élèves occupe une faible place au sein des pratiques déclarées. Un grand nombre d'enseignants déclarent fréquemment se déplacer dans la salle de classe pour repérer les erreurs de leurs élèves (96 %) et identifier leurs connaissances (74 %). Mais seulement 18 % déclarent s'appuyer fréquemment sur une évaluation diagnostique avant un nouveau chapitre. 56 % des enseignants déclarent choisir régulièrement la première activité d'un chapitre pour lui faire jouer le rôle de diagnostic, et 53 % déclarent repérer en amont les besoins des élèves pour leur proposer des exercices spécifiques en vue de stabiliser les prérequis. Plus de la moitié des enseignants (55 %) indiquent également s'appuyer « souvent » ou « très souvent » sur les erreurs connues déjà observées les années précédentes. Près d'un tiers des enseignants disent estimer régulièrement les acquis des élèves grâce à ce qu'ils savent ou pensent savoir *a priori* des difficultés de leurs élèves (voir Figure 1).

[97] https://www.education.gouv.fr/premiers-resultats-de-l-enquetesur-les-pratiques-d-enseignementdes-mathematiques-praesco-en-classe-309566

[98] Une enquête a été réalisée en parallèle sur les pratiques d'enseignement de contenus spécifiques à l'école primaire https://www.education.gouv.fr/premiers-resultats-de-l-enquete-sur-les-pratiques-d-enseignement-des-mathematiques-praesco-en-classe-309564

Figure 1 : Pratiques déclarées d'évaluation diagnostique par des enseignants de 3e en mathématiques

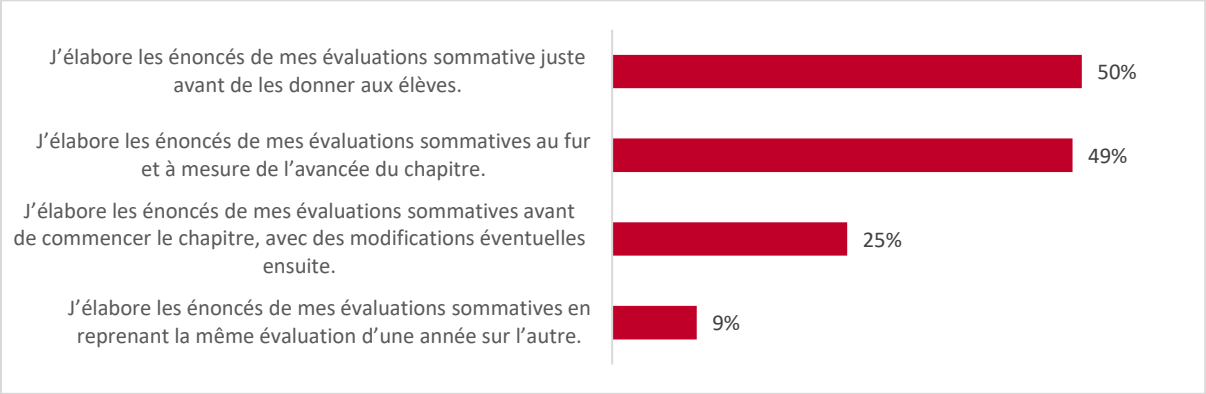


Source : Enquête Praesco (2019). Depp-MENJ.

B. L'évaluation sommative

L'évaluation sommative est peu anticipée dans la préparation des enseignants et présente peu les objectifs d'évaluation aux élèves. Un quart des enseignants déclarent élaborer régulièrement les énoncés d'évaluation sommative avant de commencer un chapitre. La moitié disent les concevoir au fur et à mesure de l'avancée du chapitre ou juste avant de les donner aux élèves. Une minorité (9 %) déclare s'appuyer « souvent » ou « très souvent » sur les évaluations proposées les années précédentes. Les élèves sont préparés aux tâches d'évaluation formelle principalement par des conseils d'ordres généraux et peu par des conseils spécifiques aux contenus évalués : 28 % des enseignants indiquent proposer une grille explicitant les compétences et les différents types d'exercices associés au thème « mathématique de l'évaluation ».

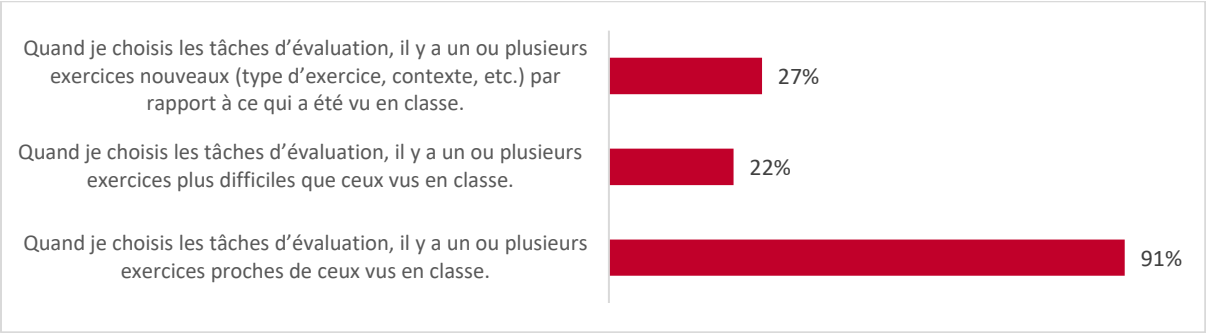
Figure 2 : Pratiques déclarées d'évaluation sommative par des enseignants de 3e en mathématiques



Source : Enquête Praesco (2019). Depp-MENJ.

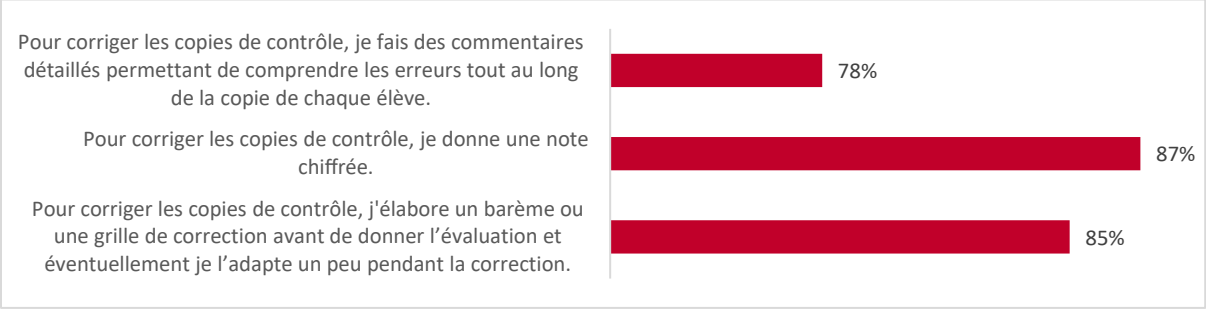
Ainsi le « contrat » d'évaluation sommative, quand il existe, apparaît expliciter peu aux élèves les objectifs d'apprentissage et le travail personnel à réaliser. La grande majorité des enseignants (91 %) déclarent concevoir des évaluations sommatives avec des exercices proches de ceux vus en classe. Ils sont moins nombreux à inclure fréquemment des exercices nouveaux (27 %) ou plus difficiles (22 %) par rapport à ce qui a été vu auparavant avec les élèves. 85 % des enseignants déclarent élaborer fréquemment un barème ou une grille de correction avant de donner l'évaluation et 87 % attribuer fréquemment une note chiffrée. Enfin, 78 % des enseignants indiquent annoter les copies.

Figure 3 : Pratiques déclarées de choix des tâches d'évaluation par des enseignants de 3e en mathématiques



Source : Enquête Praesco (2019). Depp-MENJ.

Figure 4 : Pratiques déclarées de correction des copies par des enseignants de 3e en mathématiques

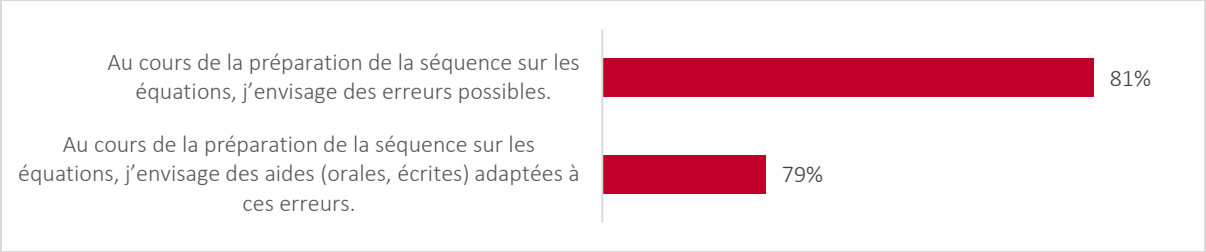


Source : Enquête Praesco (2019). Depp-MENJ.

C. L'évaluation formative

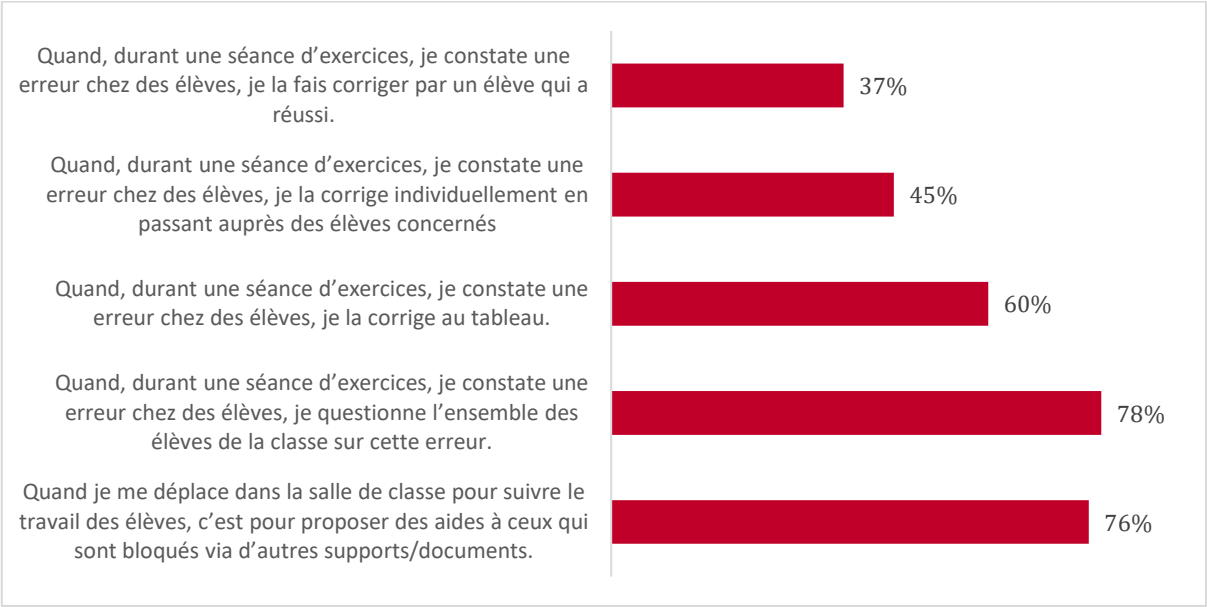
L'évaluation formative est marquée par une volonté de repérer les procédures et les erreurs des élèves mais des moyens limités pour les traiter. Environ 80 % des enseignants déclarent fréquemment anticiper dès la préparation des erreurs possibles et des aides (orales ou écrites) adaptées pour la résolution des équations. Les enseignants sont très nombreux (76 %) à indiquer se déplacer dans la classe pour proposer des aides aux élèves bloqués. Les erreurs constatées au cours d'une séance d'exercices sont traitées selon des modalités variées : 78 % des enseignants disent questionner régulièrement l'ensemble des élèves de la classe à propos de ces erreurs et 60 % les corriger eux-mêmes au tableau « souvent » ou « très souvent ». Seuls 45 % d'entre eux interviennent fréquemment envers les élèves concernés par les erreurs en passant individuellement auprès d'eux. 37 % des enseignants font corriger régulièrement une erreur constatée dans un exercice par un élève qui l'a réussi.

Figure 5 : Pratiques déclarées dans la préparation de séquences par des enseignants de 3e en mathématiques



Source : Enquête Praesco (2019). Depp-MENJ.

Figure 6 : Pratiques déclarées durant des séances d'exercices par des enseignants de 3e en mathématiques

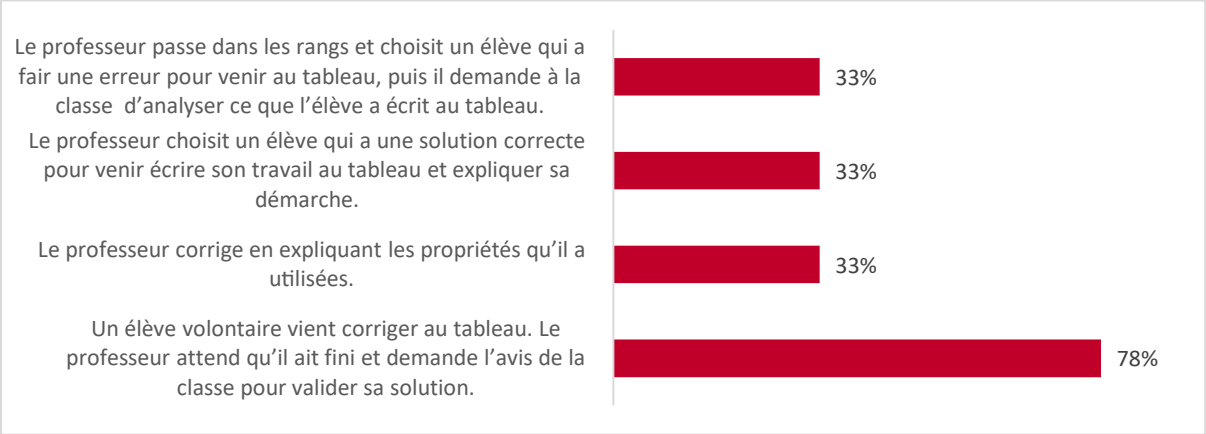


Source : Enquête Praesco (2019). Depp-MENJ.

Par ailleurs, il était demandé dans l'enquête aux enseignants d'indiquer la fréquence avec laquelle ils mettent en œuvre des corrections similaires à quatre déroulements proposés. Le déroulement qu'ils mettent en œuvre le plus fréquemment correspond, pour 78 % d'entre eux, au cas où un élève

volontaire vient corriger au tableau et où la classe est sollicitée pour valider la solution proposée par cet élève. Les trois autres déroulements de correction proposés, dans lesquels la correction est prise en charge par l'enseignant, par un élève qui a réussi, ou par un élève qui a fait une erreur de calcul littéral, recueillent des taux de réponse proches d'un tiers des enseignants.

Figure 7 : Pratiques déclarées pour des corrections d'exercices par des enseignants de 3e en mathématiques



Source : Enquête Praesco (2019). Depp-MENJ.

Ces résultats suggèrent que certaines pratiques s'appuient peu sur la fonction formative de l'évaluation. L'évaluation pour faire apprendre les mathématiques aux élèves ne semble pas suffisamment présente dans la classe pour favoriser la prise d'initiatives chez les élèves et leur donner des moyens de contrôle de leur action. Cet état des lieux sur les pratiques déclarées des enseignants sur l'évaluation conduit à interroger les leviers possibles pour les faire évoluer.

III. Un exemple de recherche sur les pratiques d'évaluation pour identifier des leviers pour les faire évoluer

Le lieu d'éducation associé (LéA) du réseau de collèges Martin-du-Gard (RMG)[99] de l'académie de Créteil, rassemblant une douzaine d'enseignants de collège et des chercheurs en didactiques des mathématiques, a visé à développer de nouvelles pratiques d'enseignement et d'évaluation en calcul littéral. Cette collaboration dure depuis plus de 6 années[100] et ses effets sur les pratiques d'évaluation des enseignants ont été analysés (cf. Horoks & Pilet, 2015 ; Pilet & Horoks, 2019). Les principales évolutions des pratiques constatées sont résumées ci-dessous.

Les choix des enseignants pour leurs évaluations sommatives sont relativement stables. L'écart entre les exercices proposés pendant une séquence d'enseignement et ceux sur lesquels les élèves sont évalués reste globalement le même, en termes de nature et de complexité, après plusieurs années de travail collectif. Ceci est à relier avec les fonctions de l'évaluation privilégiées par les enseignants, qui évoluent peu. L'évaluation diagnostique reste encore peu présente dans les pratiques des enseignants, malgré un travail collaboratif largement appuyé sur des tests diagnostiques des compétences algébriques des élèves. Cependant, les enseignants semblent identifier davantage les connaissances mathématiques mobilisées par les élèves lorsqu'ils résolvent les exercices. Mais ils exploitent peu les

[99] http://ife.ens-lyon.fr/lea/le-reseau/anciens-lea/lea-reseau-de-colleges-martin-du-gard

[100] Se référer à Allard, Horoks et Pilet (2022) pour une présentation des principes de travail du LéA.

procédures mathématiques ainsi repérées chez les élèves pour les hiérarchiser et les relier aux notions mathématiques visées.

Ces évolutions sont probablement liées au travail collaboratif, notamment celui mené sur l'analyse de productions d'élèves. Certaines contraintes institutionnelles peuvent par ailleurs influencer les pratiques des enseignants : des grilles d'évaluation sommative ont été conçues par les enseignants lors de la troisième année du travail du LéA RMG, à leur propre initiative, pour répondre à un besoin face à des demandes institutionnelles prescrivant l'usage des compétences. Ces grilles, inspirées du travail sur les tâches réalisées dans le LéA RMG, ont permis aux enseignants une analyse plus fine des activités mathématiques déployées par les élèves dans les évaluations formelles, avec un retour possible et des modalités de travail permettant un partage des critères d'évaluation et de réussite en amont et en aval des évaluations avec les élèves.

Mais un tel dispositif est coûteux en temps, bénéficie d'un soutien institutionnel et ne concerne que des enseignants volontaires, ce qui questionne sa reproductibilité dans d'autres contextes et pour toucher davantage d'enseignants.

En conclusion : les questions posées par la conférence de consensus du Cnesco

Nous avons tenté d'apporter dans cette note des éléments de réponse aux deux questions suivantes posées par la conférence de consensus du Cnesco :

1. Quelle est la place de l'évaluation pour faire apprendre les élèves, dans les pratiques des enseignants en mathématiques en France ?
2. Comment aider les enseignants à mettre en place une évaluation au service des apprentissages en mathématiques ?

Nos analyses à grande échelle et sur un temps long nous indiquent que malgré la volonté les enseignants d'accompagner au mieux les élèves en vue de leur réussite, les moyens qu'ils mettent en œuvre pour y parvenir, en particulier à travers l'évaluation des apprentissages, semblent encore relativement limités. Une piste possible serait selon nous d'outiller les enseignants à analyser plus finement les contenus mathématiques à enseigner et les procédures mathématiques développées par les élèves. Ainsi la question de l'évaluation, qui concerne à la fois les contenus enseignés et leur mise en œuvre en classe, nous paraît être un levier pertinent pour le développement professionnel des enseignants. Cet enjeu de formation est probablement partagé avec d'autres disciplines, mais la façon d'y répondre dépendra des savoirs enseignés.

Références

Allal, L. & Mottier-Lopez, L. (2005). Formative assessment of learning: A review of publications in French. In *Formative Assessment - Improving Learning in Secondary Classrooms* (p. 241-264). Paris: OECD Publication.

Allard, C., Horoks, J. & Pilet, J. (2022). Principes de travail collaboratif entre chercheur.e.s et enseignant.e.s : le cas du LéA RMG, *Éducation et Didactique*, 16(1), 49-66.

Artigue, M. & Winslow, C. (2010). International comparative studies in mathematics education: a view from the anthropological theory of didactics. *Recherches en didactique des mathématiques*, 30(1), 47-82.

Ash, D. & Levitt, K. (2003). Working within the Zone of Proximal Development: Formative Assessment as Professional Development. *Journal of Science Teacher Education, 14(1)*, 1-313.

Bain, D. (1988). L'évaluation fait fausse route : de là, la nécessité de changer de CAP. *Mesure et évaluation, 10(4)*, 23-32.

Bodin, A. (1997). L'évaluation du savoir mathématique. Questions et méthodes. *Recherches en didactique des mathématiques*, 17(1), 49-96.

Bodin, A. (2006). Les mathématiques face aux évaluations nationales et internationales. *Repères IREM*, 65, 55-89.

Black, P. & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. Granada Learning.

Chanudet, M., Coppé, S., Gandit, M. & Moulin, M. (2018). Analyse des interactions didactiques dans une perspective d'évaluation formative. In S. Coppé & E. Roditi (dir.), *Actes de la 19e école d'été de didactique des mathématiques* (p. 453-473). La pensée sauvage.

Chanudet, M. (2019). *Étude des pratiques évaluatives des enseignants dans le cadre d'un enseignement centré sur la résolution de problèmes en mathématiques*. Université de Genève. Thèse, 2019. doi: 10.13097/archive-ouverte/unige:125833

Coppé, S. & Moulin, M. (2017). Évaluation entre pairs et débat argumenté dans le cadre d'un problème complexe en mathématiques. *Canadian Journal of Science, Mathematics and Technology Education, 17(4)*, 308-327.

De Ketele, J. M., Gerard, F. M. & Roegiers, X. (1997). L'évaluation et l'observation scolaire : deux démarches complémentaires. *Éducations - Revue de diffusion des savoirs en éducation*, 12, 33-37.

Georges, F. & Pansu, E. (2011). Les feedbacks à l'école : un gage de régulation des comportements scolaires. *Revue française de pédagogie*, n°176, pp.101-146. Grugeon, B. (1997). Conception et exploitation d'une structure d'analyse multidimensionnelle en algèbre élémentaire. *Recherches en didactique des mathématiques*, *17(2)*, 49-96.

Grugeon-Allys, B. (2016). Évaluer en mathématiques : une approche didactique et épistémologique, In P. Detroz, M. Crahay, & A. Fagnant (ed), *L'évaluation à la lumière des contextes et des disciplines* (pp. 63-90). De Boeck.

Grugeon-Allys, B., Pilet, J., Chenevotot, F. & Delozanne, E. (2012). Diagnostic et parcours différenciés d'enseignement en algèbre élémentaire. In L. Coulange, J.-P. Drouhard, J.-L. Dorier. & A. Robert (ed),

Recherche en Didactique des Mathématiques, Enseignement de l'algèbre élémentaire, Bilan et perspectives, Hors-série, 137-162. La pensée sauvage.

Grugeon-Allys, B. (2018). Évaluer en mathématiques : des pistes pour un enrichissement réciproque de la recherche en didactique des mathématiques et en évaluation. In S. Coppé & E. Roditi (dir.), *Actes de la 19e école d'été de didactique des mathématiques* (p. 165-194). Grenoble : La pensée sauvage.

Horoks, J. (2022). Des pratiques aux apprentissages mathématiques, en passant par la formation. : *Circulation des savoirs issus des recherches en Didactique des Mathématiques.* [Note de synthèse pour l'habilitation à diriger des recherches, CY Cergy Paris Université, France], ⟨tel-03613699⟩.

Horoks, J. & Pilet, J. (2015) Etudier et faire évoluer les pratiques d'évaluation des enseignants de mathématiques en algèbre au collège dans le cadre d'un Léa, In L. Theis, Actes EMF2015, *Pluralités culturelles et universalité des mathématiques : enjeux et perspectives pour leur enseignement et leur apprentissage*, Alger, Algerie, 10 au 14 octobre 2015, GT9, 491-501.

Horoks, J., Kiwan, M., Pilet, J., Roditi, E. & Haspekian, M. (2019). Régulation des apprentissages et évaluation formative : quels regards didactiques ? In S. Coppé & E. Roditi (Eds.), *Actes de la XIXème Ecole d'été de didactique des mathématiques, Paris, août 2017* (p. 429-452). Grenoble : La Pensée Sauvage.

Mottier Lopez, L. (2012). *La régulation des apprentissages en classe*. De Boeck.

Perrenoud, P. (1998). From Formative Evaluation to a Controlled Regulation of Learning Processes. Towards a wider conceptual field, *Assessment in Education*, Vol. 5, N°1, pp.85-102.

Pilet, J. & Horoks, J. (2019). Effets potentiels d'une évolution des pratiques enseignantes d'évaluation sur les apprentissages algébriques des élèves au collège. In M. Abboud (Ed.), *Actes du Colloque EMF 2018* (p. 1030-1038). IREM de Paris.

Pilet, J. (2015). Réguler l'enseignement en algèbre élémentaire par des parcours d'enseignement différencié. *Recherches en didactique des mathématiques, 35(3),* 273-312.

Rey, O. & Feyfant, A. (2014). *Evaluer pour (mieux) faire apprendre*. Dossier d'actualité Veille et Analyses, 94, 44.

Roditi, E. & Salles, F. (2015). Nouvelles analyses de l'enquête PISA 2012 en mathématiques. *Éducation et formations*, 86-87, 236-267.

Sayac, N. (2017). *Approche didactique de l'évaluation et de ses pratiques en mathématiques : enjeux d'apprentissages et de formation*. Habilitation à Diriger des Recherches. Université Paris-Diderot, Paris.

Sayac, N. & Grapin, N (2016). Stratégies et degrés de certitude des filles et des garçons en mathématiques : quelles différences pour quels résultats ? *Repères IREM*, 104, 43-57.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A. & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21(4), 295-314.

Winslow, C. (2005). Définir les objectifs de l'enseignement mathématique : La dialectique matières-compétences. *Annales de didactique et des sciences cognitives, 10,* 131–156.

Comment l'évaluation peut-elle favoriser la compréhension de l'oral ? L'exemple des langues vivantes étrangères.

STÉPHANIE ROUSSEL
UNIVERSITÉ DE BORDEAUX (FRANCE)

Introduction

Dans le cadre scolaire, un certain nombre de directives institutionnelles, comme le socle commun de connaissances, de compétences et de culture (2015) et le cadre européen commun de références pour les langues (CECRL, 2001, 2018) régissent l'évaluation de la compréhension de l'oral (désormais CO) en langue seconde (L2). Ces textes officiels poursuivent deux objectifs principaux : fixer pour les élèves des objectifs de maîtrise, et harmoniser les pratiques d'évaluation de cette compétence, à la fois aux niveaux national et européen.

Si les descripteurs et grilles d'items qu'ils contiennent permettent aux enseignants et aux élèves d'évoluer dans un cadre et de satisfaire aux standards scolaires, ils ne donnent en revanche aucune indication sur la manière dont les élèves peuvent apprendre à comprendre, et donc sur la manière dont il est possible, pour les enseignants, de concevoir des situations d'entraînement de cette compétence fondées sur des données issues d'une évaluation (que celle-ci soit formelle ou non). Autrement dit, le cadre ne dit pas comment mettre l'évaluation au service de l'apprentissage et des progrès des élèves. Ce n'est d'ailleurs pas son objectif. Entre l'évaluation et l'apprentissage, le lien parait alors ténu et l'articulation fragile.

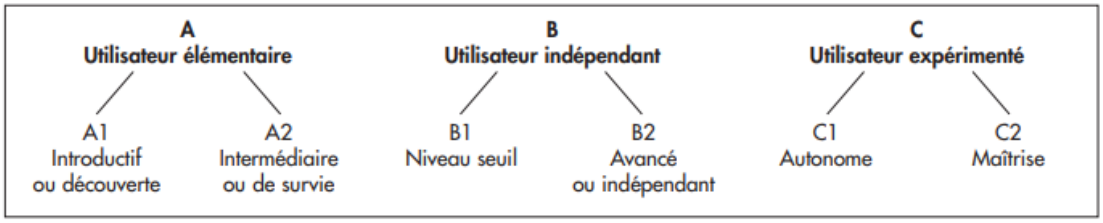
Dans cette note, après avoir décrit les outils institutionnels à la disposition des enseignants pour évaluer la compréhension orale (CO), nous allons explorer le lien entre l'évaluation et l'apprentissage dans le domaine de la CO en nous appuyant sur les recherches qui se sont intéressées à cette compétence. Nous verrons que l'évaluation de la CO pose de nombreux défis non seulement aux enseignants, mais aussi aux chercheurs. Cette activité repose, en effet, sur des processus internes, dont seuls des indices comportementaux externes (production verbale ou non verbale) fournissent des traces. L'étude de ces traces et leur prise en compte permettent cependant de dégager des pistes pour comprendre les ressorts de la CO. Dès lors, il est possible de concevoir une articulation entre l'évaluation (formelle ou informelle) de ce que comprennent les élèves et la mise en œuvre d'une séquence d'enseignement-apprentissage (situations d'entraînement) pertinente pour améliorer leur CO.

I. Des outils d'évaluation largement imposés par l'institution

A. Le socle commun de connaissances, de compétences et de culture et le CECRL

En langues vivantes étrangères, les programmes ainsi que les compétences sont déclinés à partir du socle commun de connaissances, de compétences et de culture (Bulletin officiel n°17 du 23 avril 2015). L'évaluation des langues vivantes étrangères fait partie de l'évaluation de la composante 2 (comprendre, s'exprimer en utilisant une langue étrangère et, le cas échéant, une langue régionale) du domaine 1 (les langages pour penser et communiquer) du socle. Le socle commun est lui-même adossé au CECRL, qui permet de positionner l'élève à un certain niveau (de A1 à C2) dans les cinq activités langagières (CO, compréhension de l'écrit, expression orale en interaction, expression orale en continu, expression écrite), et reconnait trois grands types d'utilisateurs des langues secondes : l'utilisateur élémentaire (A), indépendant (B) et expérimenté (C).

Figure 1 : niveaux du CECRL (p. 25)



B. La compréhension de l'oral

Dans le volume complémentaire du CECRL (2018), la grille de descripteurs proposée pour la compréhension générale de l'oral (Figure 2) est complétée par d'autres qui distinguent différentes situations de CO : comprendre en tant qu'auditeur, comprendre des annonces et des instructions, comprendre des émissions de radio et des enregistrements. Il existe également une échelle à part pour « Comprendre des émissions de télévision et des films » dans la section « Réception audiovisuelle ».

Figure 2: Descripteurs du CECRL pour la compréhension de l'oral - Volume Complémentaire avec de Nouveaux Descripteurs (CECR, 2018, p. 58)

RÉCEPTION ORALE	
COMPRÉHENSION GÉNÉRALE DE L'ORAL	
C2	Peut comprendre sans effort pratiquement toute langue orale, en direct ou enregistrée, si le débit est naturel.
C1	Peut suivre une intervention d'une certaine longueur sur des sujets abstraits ou complexes même hors de son domaine mais peut avoir besoin de faire confirmer quelques détails, notamment si l'accent n'est pas familier. Peut reconnaître une gamme étendue d'expressions idiomatiques et de tournures courantes ainsi que les changements de registre. Peut suivre une intervention d'une certaine longueur même si elle n'est pas clairement structurée et même si les relations entre les idées sont seulement implicites et non explicitement indiquées.
B2	Peut comprendre une langue orale standard, en direct ou à la radio, sur des sujets familiers ou non familiers se rencontrant normalement dans la vie personnelle, sociale, académique ou professionnelle. Seul un très fort bruit de fond, une structure inadaptée du discours ou l'utilisation d'expressions idiomatiques peuvent gêner sa compréhension.
	Peut comprendre les idées principales d'interventions complexes du point de vue du fond et de la forme sur un sujet concret ou abstrait et dans une langue standard, y compris des discussions techniques dans son domaine de spécialisation. Peut suivre une intervention d'une certaine longueur comportant une argumentation complexe à condition que le sujet soit assez familier et que le plan général de l'exposé soit indiqué par des marqueurs explicites.
B1	Peut comprendre une information factuelle directe sur des sujets de la vie quotidienne ou relatifs au travail en distinguant les messages généraux et les points de détail, à condition que l'articulation soit claire et l'accent courant.
	Peut comprendre les points principaux d'une intervention dans une langue claire et standard sur des sujets familiers rencontrés régulièrement au travail, à l'école, pendant les loisirs, y compris des récits courts.
A2	Peut comprendre assez pour pouvoir répondre à des besoins concrets à condition que la diction soit claire et le débit lent.
	Peut comprendre des expressions et des mots relatifs à des domaines de priorité immédiate (par exemple, information personnelle et familiale de base, achats, géographie locale, emploi), à condition que la diction soit claire, bien articulée et lente.
A1	Peut comprendre une intervention si elle est lente, soigneusement articulée et comprend de longues pauses qui permettent d'en assimiler le sens. **Peut reconnaître une information concrète à propos d'un sujet familier et quotidien, à condition que le débit soit lent et que l'information soit claire (par exemple sur un lieu ou un horaire).**
Pré-A1	**Peut comprendre des questions et des affirmations courtes et très simples, à condition qu'elles soient prononcées clairement et lentement, illustrées par des gestes ou des images facilitant la compréhension et qu'elles soient éventuellement répétées.** **Peut reconnaître des mots familiers, à condition qu'ils soient prononcés clairement et lentement dans un contexte clairement défini, quotidien et familier.** **Peut reconnaître des nombres, des prix, des dates et les jours de la semaine à condition qu'ils soient prononcés clairement et lentement dans un contexte clairement défini, quotidien et familier.**

Le CECRL présente l'avantage - et peut-être à l'époque la nouveauté - de considérer officiellement la classe comme un environnement qui prépare les élèves à s'adapter à leur environnement social et multiculturel, et se réclame ainsi d'une perspective actionnelle[101], même si aucune référence théorique n'y est mentionnée pour justifier cette approche (Maurer & Puren, 2019) et que l'on a plutôt l'impression d'un prolongement de l'approche communicative, à tout le moins, comme nous le verrons, du point de vue des objectifs affichés.

Par ailleurs, le CECRL a pour ambition d'harmoniser l'évaluation des langues en Europe, ce qui constitue un autre avantage incontestable pour les étudiants qui circulent en Europe, d'une université à l'autre par exemple. Leur niveau de langue devient lisible d'un pays à l'autre. Cet avantage est minoré par une mise en œuvre irrationnelle du CECRL dans certaines universités, qui considèrent que chaque semestre, chaque étudiant devrait valider un niveau supplémentaire… Dans les programmes scolaires, le CECRL est aussi décliné au plan national.

C. Exemples de grilles d'évaluation fournies aux enseignants

Sur le site Éduscol, plusieurs exemples de situations d'évaluation de la CO sont mis à la disposition des enseignants. Pour les élèves de collège, l'un d'entre eux propose, en allemand, la situation suivante : « Tu es membre du ciné-club langues vivantes de ton collège et tu participes à la sélection du film qui sera prochainement projeté. Tu vas regarder un document vidéo pour en savoir plus sur le film *Das Wunder von Bern* ». Le document mis à la disposition des enseignants précise que les élèves peuvent soit rendre compte de ce qu'ils ont compris, en langue cible ou en français, soit « être évalués à partir d'un protocole gradué et progressif, construit à partir d'exercices permettant d'apprécier et d'évaluer le niveau de l'élève. » Nous avons donc ici deux tâches verbales (restitution en langue cible et réponses à des questions ciblées) qui permettent à l'enseignant de collecter des indices de compréhension. Plusieurs tableaux adossés au CECRL proposent ensuite des exemples d'éléments linguistiques et culturels que l'élève pourra relever dans les supports vidéo (voir Figure 3).

[101] Dans la perspective actionnelle, les élèves écrivent, parlent, lisent dans un but social précis. La perspective actionnelle va donc plus loin de la perspective communicative. Comme l'expriment bien Baily et Cohen (2009), dans l'approche communicative la priorité est donnée au sens et au contexte de l'énoncé dans une situation de communication, par opposition aux approches précédentes (grammaire-traduction, audio-visuelle, audio-orale) qui mettaient davantage l'accent sur les formes et les structures des langues.

Figure 3: Grilles d'évaluation fournies aux enseignants pour l'évaluation de la compréhension au niveau A1 et B1

Niveau A1 du CECRL

Attendus de fin de cycle
Peut comprendre des mots familiers et des expressions courantes sur lui-même, sa famille et son environnement.

REPÈRES DE PROGRESSIVITÉ (LINGUISTIQUES ET CULTURELS)	EXEMPLES D'ÉLÉMENTS LINGUISTIQUES ET CULTURELS À REPÉRER
Repérer des indices sonores simples et des expressions courantes pour construire une amorce de compréhension du document.	L'élève a repéré les indices visuels pour identifier la nature du document, et les interlocuteurs : • il s'agit des extraits du film. On peut également entendre des spectateurs qui s'expriment sur le film en sortant du cinéma.
Isoler des informations simples dans un message.	L'élève a relevé les informations isolées suivantes (mots transparents, ou répétés un nombre de fois suffisant) : • *Kind, Vater, Tor, super* Amorces de compréhension possibles : • la vidéo montre qu'il y a des émotions dans le film ; • les spectateurs sont contents d'avoir vu ce film…
Percevoir quelques références culturelles habituelles, simples et reconnaissables appartenant à la langue du pays concerné.	L'élève a su repérer des références culturelles visibles : • l'histoire d'un match de football ; les Allemands jouent en finale.

Niveau B1 du CECRL

Attendus de fin de cycle
Peut comprendre une information factuelle sur des sujets simples en distinguant l'idée générale et les points de détail, à condition que l'articulation soit claire et l'accent courant.

REPÈRES DE PROGRESSIVITÉ (LINGUISTIQUES ET CULTURELS)	EXEMPLES D'ÉLÉMENTS LINGUISTIQUES ET CULTURELS À REPÉRER
Comprendre un message oral en continu sur un point d'intérêt personnel ou sur un sujet déjà connu (en lien avec les thèmes culturels du programme).	L'élève est capable de mettre en relation les informations repérées pour retranscrire le sujet du document : • les spectateurs trouvent le film superbe ; • une jeune fille pense qu'il est adapté à tous les âges.
Suivre les points principaux d'une discussion d'une certaine longueur sur un sujet familier ou d'actualité. Suivre le plan général d'exposés courts sur des sujets familiers ou déjà connus. Relever et comprendre les points de détail.	L'élève a repéré les informations essentielles du document, ainsi que les points de détail, et les met en relation pour en reconstruire le sens : • chaque enfant a besoin d'un père ; • le père revient de sa captivité en tant que prisonnier de guerre ; • chacun a besoin d'un rêve ; • il y a une grande compétition de football ; • chaque pays a besoin d'une légende : les Allemands gagnent la finale de la coupe du monde 1954.
Mobiliser des références culturelles pour interpréter les éléments d'un document sonore.	L'élève mobilise des références dans le thème donné : • Allemagne, Suisse, après-guerre…

Les indications fournies aux enseignants sur l'évaluation de la CO prennent donc majoritairement la forme de grilles d'évaluation, de descripteurs, de tableaux et d'exemples de situations d'évaluation « clef-en main ».

D. Le CECRL, un outil d'évaluation critiquable

L'une des critiques adressées au CECRL, en tant qu'outil d'évaluation, concerne le caractère approximatif des items, et apparait clairement lorsqu'on consulte par exemple la grille de descripteurs présentée Figure 2. On compte alors un nombre important d'adverbes qui laissent place à la subjectivité et à l'interprétation individuelle : « pratiquement » dans l'expression « pratiquement toute langue orale », ou « assez » dans l'expression « comprendre assez », et d'adjectifs imprécis comme « certain » dans « une certaine longueur » ou « complexes » dans « des interventions complexes ». Selon Puren (2019), ce flou serait intentionnel et bénéficierait à des opérateurs privés (*Cambridge University, Goethe Institut, Instituto Cervantes*), dont les certifications seraient plus fiables car ils utilisent des descripteurs plus précis et plus calibrés. La polémique prend alors une dimension politique et même financière. Il reste qu'il est difficile pour les enseignants - malgré la liberté et la possibilité de créativité laissée par le cadre - dans leurs classes, de juger si le document choisi a bien « une certaine longueur » et si l'élève a « compris assez », ce qui fait du CECRL un outil non seulement complexe à utiliser, mais aussi peu susceptible de servir précisément les objectifs d'harmonisation qu'il s'est fixé. Tardieu (2010) intitule d'ailleurs l'un de ses articles « Votre B1 est-il mon B1 ? ».

Pour trouver d'autres outils mobilisables en classe pour évaluer la CO, on peut interroger les travaux de recherche qui se sont intéressés à cette compétence et aux comportements des élèves pendant une tâche de CO.

II. La recherche sur les compétences de compréhension de l'oral : des leviers pour la conception d'entraînements

A. Les protocoles verbaux

De nombreuses recherches (Goh, 1998, 2000 ; Vandergrift, 2003, 2006 ; Vandergrift & Tafaghodtari, 2010 ; Vandergrift & Goh, 2012) ont étudié le comportement spontané des apprenants lors d'une tâche de rappel en langue seconde, c'est-à-dire en leur demandant de raconter à l'oral, au fur et à mesure de l'écoute ou après l'écoute entière, en langue première (L1), ce qu'ils comprenaient, puis en étudiant finement ces protocoles verbaux. L'étude de Vandergrift (2003) compare les différentes stratégies de CO utilisées par de jeunes anglophones qui apprennent le français langue seconde. Dans cette étude, les participants écoutent un dialogue en français dans lequel un animateur radio annonce à une auditrice qu'elle a gagné un séjour au ski. Vandergrift propose à tous les participants une écoute fractionnée du document sonore. Après l'écoute de chaque extrait, les participants sont invités à dire ce qui leur vient à l'esprit. Ci-dessous (Tableau 1) un extrait (Vandergrift, 2003, p. 478) du rappel en anglais de deux apprenantes (Rose et Nina), qui n'ont pas le même niveau initial en français, langue seconde.

Tableau 1 : Extrait du rappel en français de deux élèves anglophones (Vandergrift, 2003 p. 478)

« Allô, est-ce que je peux parler à Mademoiselle Hélène Petit, s'il vous plaît. *- C'est moi-même, monsieur. »*	
***Rose**: ''Hello, can I talk to Mademoiselle ... '' I don't know the name. And then she said, ''That's me.''*	***Nina**: This is a conversation on the phone. Yeah, and a male asked to talk to a female about something. Yeah.*
***Int**.: O.K. Anything that you're thinking about?*	***Int**.: What's going on in your mind?*
***Rose**: No.*	***Nina**: I think it has something to do with, like something, if they're advertising something because if they were going to have a long conversation like, between, like a friend, they would say, they won't call it ''Mrs.'' whatever her name is.*

« Ah, bonjour et joyeuse Saint-Valentin, Hélène. Ici Robert Bélair de CKAC. J'ai une très bonne nouvelle à vous annoncer aujourd'hui. *- Oui ? »*	
***Rose**:* ''Hi. I got something new.'' ''Nouvelle'' I don't know. That's all I got.	***Nina**: This is a conversation on the phone. Yeah, and a male asked to talk to a female about something. Yeah.*
***Int**.:* O.K. Anything you're thinking about now?	***Int**.: What's going on in your mind?*
***Rose**: Rose: No, just nervous.*	***Nina**: Nina: Something new. Because he said, ''nouvelle,'' I heard some- thing, or something new today, he said ''aujourd'hui.'' Maybe he is advertising something like, today, I have something new, or...*
	***Int.:** Anything else going on?*
	***Nina:** Nothing now.*

On peut observer, à travers la comparaison entre les deux apprenantes, que la compétence linguistique initiale correspond à d'importantes différences dans la manière dont elles rapportent ce qu'elles ont compris. Alors que Rose traduit brièvement dans sa langue première (l'anglais) ce qu'elle entend en français sans relier ces éléments, ni au cotexte (à l'ensemble des éléments du discours oral), ni au contexte, Nina s'engage spontanément dans l'élaboration d'hypothèses sur la suite du dialogue. Elle s'appuie sur la manière dont les locuteurs s'adressent l'un à l'autre et en infère le type de relation qu'ils entretiennent. Elle comprend assez rapidement qu'il s'agit d'une émission de radio. De l'intonation de la locutrice, Nina déduit la surprise, qu'elle combine avec les éléments lexicaux, pour établir que celle-ci a sans doute gagné quelque chose et qu'elle ne s'y attendait pas. Vandergrift (2003) observe ainsi que les auditeurs les plus compétents en L2 utilisent un plus grand nombre de stratégies de contrôle ou de régulation de la tâche de compréhension. Pendant l'écoute, ils ne cessent d'interroger ce qu'ils entendent et de confronter ce qu'ils perçoivent avec leurs connaissances du monde (concernant l'intonation ou les relations entre les protagonistes), faisant ainsi preuve d'ouverture et de souplesse dans leur approche. L'auteur de l'étude insiste sur la capacité de Nina à prédire ce qui va suivre, à planifier son écoute, à ajuster son attention, à repérer les problèmes et à les résoudre. Toutes ces compétences et stratégies sont visibles et pourraient donc parfaitement faire partie de l'évaluation de la CO en classe : cela permettrait de rendre compte, non pas uniquement de ce qui est compris, mais aussi du processus qui conduit à la compréhension, c'est-à-dire des étapes de

l'apprentissage. En d'autres termes, tandis que les descripteurs attestent d'un niveau, l'analyse des rappels permet aux enseignant de formuler des feedbacks aux élèves et d'ajuster leur enseignement.

Pour trouver des explications théoriques au comportement d'écoute de Rose et de Nina, il faut se tourner vers les mécanismes cognitifs de la CO. Ces explications théoriques sont utiles pour avoir une vision fine de ce qui se joue en situation de CO en L2 et ainsi prendre en compte d'autres indices d'apprentissage, qu'il est possible d'évaluer pour mieux accompagner les progrès des élèves.

B. Les mécanismes cognitifs de la compréhension de l'oral

Du point de vue du traitement de l'information orale, comprendre, c'est-à-dire construire le sens d'un discours, implique un certain nombre d'opérations dites de « bas niveau » parce qu'elles concernent la manipulation directe du langage (traitements phonologiques, décodage, segmentation, traitements lexicaux et syntaxiques, etc.) et de « haut niveau » (régulation de la tâche, anticipation, maintien de la cohérence du discours, activation des connaissances préalables en mémoire à long terme), qui interagissent entre elles. Ces opérations interactives sollicitent fortement la mémoire de travail, dont on sait qu'elle a une capacité limitée.

En langue seconde, contrairement à ce qui se passe en L1, la faible automatisation des processus de « bas niveau » rend ces opérations très coûteuses sur le plan cognitif. Ce coût cognitif est d'autant plus important, lorsqu'il s'agit de la CO, que cette activité se déroule le plus souvent « en temps réel », avec des contraintes temporelles fortes (Gaonac'h & Fayol, 2003). Outre le risque d'échec dans la réalisation de ces opérations, ce coût supplémentaire, la mobilisation des ressources cognitives nécessaires peut se faire au détriment de la mise en œuvre des processus de haut niveau. C'est, dans l'exemple ci-dessus, ce qui explique le comportement de Rose, dont l'attention est accaparée par les processus de décodage : le coût cognitif et la saturation de la mémoire de travail, provoquée par une automatisation insuffisante des processus de bas niveau en L2, rendent plus difficile l'interaction entre les différents processus, eux-mêmes situés à différents niveaux. Chez Nina, en revanche, l'automatisation plus importante des processus de bas niveau lui permet de mobiliser des processus de construction du sens. L'ensemble de ces contraintes explique d'ailleurs que les élèves ressentent souvent une anxiété importante dans les tâches de compréhension orale.

C. La compréhension de l'oral, une activité angoissante pour les apprenants

Les travaux de Goh (2000), d'Elkhafaifi (2005), mais aussi de Vandergrift (Vandergrift & Goh, 2012) par exemple, sont ceux qui sont les plus largement cités lorsqu'il s'agit d'examiner le ressenti des élèves face à une tâche de CO en langue seconde. Goh (2000) explique par exemple que lors de la phase de perception de l'*input* sonore certains apprenants disent ne pas reconnaître des mots qu'ils connaissent pourtant, négliger la partie suivante du discours lorsqu'ils pensent au sens, ne pas parvenir à découper le flux du discours, manquer le début du discours. Lors de la phase d'analyse syntaxique, pendant laquelle les mots sont combinés pour former une représentation mentale cohérente à la fois avec les éléments syntaxiques repérés et avec les connaissances stockées en mémoire à long terme, les apprenants déclarent par exemple oublier rapidement ce qu'ils ont entendu, être incapables de former une représentation mentale de ce qu'ils entendent. Lors de la phase d'utilisation (dernière phase du modèle de la compréhension, élaboré par Anderson (1995), voir aussi Fayol et Gaonac'h, 2003), les apprenants devraient pouvoir inférer et déduire le sens à partir de ce qu'ils ont compris, or certains expliquent qu'ils comprennent les mots, mais pas le message ni les idées principales qu'ils véhiculent.

En exposant les mécanismes cognitifs de la compréhension, le rôle de la mémoire de travail et des connaissances du monde, l'objectif est non seulement de fournir une explication aux difficultés ressenties par les élèves, mais aussi de les situer aux différents niveaux du processus de compréhension pour proposer des leviers et des pistes d'entraînement aux enseignants. Il faut retenir de ces travaux la grande variété interindividuelle : les comportements des élèves face à une tâche de CO sont très différents. Leur évaluation peut donc difficilement se borner à la question de savoir s'ils ont compris ou pas, ni même d'ailleurs à ce qu'ils ont compris.

III. Évaluer pour faire apprendre/faire progresser (entrainer)

A. L'absence d'articulation entre l'évaluation et l'apprentissage dans le CECRL

Le CECRL, qui pourtant s'intitule « apprendre, enseigner, évaluer », ignore complètement la méthodologie d'enseignement-apprentissage (Maurer & Puren, 2019). Le cadre ne dit pas comment les élèves peuvent atteindre tel ou tel niveau, les descripteurs ne donnent aucune piste didactique pour entrainer les élèves à comprendre, ni d'ailleurs pour interpréter les difficultés des élèves et donc y apporter une réponse en proposant des entraînements adaptés. Il n'y est donc pas question d'apprendre ou d'enseigner, contrairement à ce qui est annoncé dans le titre, mais uniquement d'évaluer, et c'est sans doute l'effet le plus délétère de ces descripteurs sur les pratiques enseignantes.

Pourtant, la question corollaire à celle de l'évaluation de la CO est bien celle de l'entraînement de cette compétence. En s'appuyant sur les recherches que nous avons présentées plus haut, on peut concevoir des situations d'évaluation et des entraînements pour apprendre aux élèves à comprendre. Une conclusion importante ressort notamment des données théoriques et empiriques que nous venons de présenter : tous les apprenants ne procèdent pas de la même façon pour écouter et comprendre un document sonore en langue étrangère. En CO, l'évaluation pourrait donc servir à repérer finement les processus cognitifs en jeu, en tenant compte de différences individuelles majeures. Pour atteindre cet objectif, plusieurs outils sont à la disposition des enseignants.

B. Questionnaire métacognitif

Le *Metacognitive Awareness Listening Questionnaire* (MALQ) a été élaboré pour évaluer la conscience métacognitive des apprenants de L2 concernant l'utilisation de stratégies d'écoute-compréhension. Il permet également aux auditeurs de s'autoévaluer (Vandergrift, Goh, Mareschal & Tafaghodtari, 2006) et ainsi d'acquérir de l'autonomie face à la tâche de compréhension. Le questionnaire comprend 21 items, qui représentent cinq domaines de « conscience métacognitive » concernant la CO en langue seconde : résolution de problèmes, planification et évaluation, traduction mentale, attention dirigée et connaissance de soi-même. Le questionnaire est administré après une tâche d'écoute, et les répondants sont invités à déclarer les stratégies qu'ils ont utilisées pour accomplir la tâche. Les recherches ont montré que la conscience métacognitive est positivement liée à la performance en CO. Le MALQ, ou ses adaptations, peut être utilisé en classe pour évaluer le développement métacognitif des apprenants de L2 tout au long d'une année scolaire et les inciter par exemple à utiliser les mots qu'ils comprennent pour deviner le sens de mots qu'ils ne comprennent pas (item 5). L'auto évaluation peut donc aider les élèves à progresser, en les conduisant à prendre conscience des étapes intermédiaires de la CO et des processus mentaux en jeu.

Tableau 2 : Premiers items du Metacognitive Awareness Listening Questionnaire (MALQ), (Vandergrift, Goh, Mareschal & Tafaghodtari, 2006, p. 462, traduction française des auteurs)

Planification et évaluation	1. Avant d'écouter un extrait audio, j'ai en tête un plan sur la façon dont je vais écouter l'extrait.	1 2 3 4 5 6
Attention dirigée	2. Je me concentre davantage sur le texte quand j'ai de la difficulté à comprendre.	1 2 3 4 5 6
Connaissance sur soi-même	3. Je trouve que comprendre un extrait sonore en français est plus difficile que lire, parler, ou écrire en français.	1 2 3 4 5 6
Traduction mentale	4. Je traduis dans ma tête au fur et à mesure que j'écoute.	1 2 3 4 5 6
Résolution de problèmes	5. J'utilise les mots que je comprends pour deviner le sens des mots incompris.	1 2 3 4 5 6
Attention dirigée	6. Lorsque je sens que mon esprit s'égare, je récupère ma concentration immédiatement.	1 2 3 4 5 6
Résolution de problèmes	7. Tout au long de l'exercice d'écoute, je compare ce que je comprends à ce que je sais du sujet traité dans l'extrait.	1 2 3 4 5 6
Connaissance sur soi-même	8. Je sens que la compréhension d'un texte oral en français constitue pour moi un défi.	1 2 3 4 5 6
Résolution de problèmes	9. J'utilise mon expérience et mes connaissances pour m'aider à comprendre.	1 2 3 4 5 6

C. Complexité lexicale, complexité syntaxique, fluidité : des indices pour une évaluation du comportement de l'apprenant

Les recherches précédentes ont analysé ce que des élèves en L2 expriment (oralement ou à l'écrit) au cours d'une tâche de compréhension orale pour mettre en évidence les stratégies d'écoute et les processus cognitifs qui sous-tendent la compréhension. Ces études ont même permis de construire des outils d'(auto)évaluation comme le MALQ. Dans nos recherches (Roussel *et al*., 2008 ; Roussel & Tricot, 2015), nous avons souhaité aller plus loin en analysant les dimensions syntaxiques, les dimensions lexicales et la fluidité de ces protocoles verbaux. Nos résultats indiquent qu'un rappel structuré, distancé, détaillé et organisé témoigne d'une compréhension profonde du document sonore, alors qu'un rappel peu structuré et constitué d'énumérations d'éléments entendus, peu reliés entre eux par des mots du discours ou des conjonctions, montre que l'articulation du discours et donc la cohérence des idées a été moins précisément repérée. La présence de connecteurs (comme « mais », « cependant », « parce », « bien que », etc.), témoigne en surface de la mise en place de relations cohérentes entre les différentes propositions (Kintsch, 1998). Nous considérons que les omissions, les interférences, comme l'utilisation de la L2 dans un rappel en L1, la parataxe (juxtaposition de mot ou de phrase sans mot de liaison), le mot à mot, peuvent être des signes d'une compréhension dégradée. Au contraire, des rappels synthétiques, où les apprenants utilisent le mot juste, rendent compte de l'implicite, des détails, ajoutent des remarques personnelles ou des métacommentaires (voir Tableau 3), témoignent d'une régulation (ou d'un contrôle) de sa compréhension par l'auditeur. Voilà donc, de notre point de vue, des indices à évaluer et à mesurer en classe. Nous donnons quelques exemples de rappels d'élèves de seconde, extraits de Roussel *et al*. (2008).

Tableau 3 : Indices de la compréhension (Roussel et al. 2008)

Omission	*« Puis elle propose d'être manager et …* »
De l'allemand (L2) dans les rappels en français (langue maternelle)	« *Marco serait* "*Baur*" *»* : le mot « *Bauer »* (paysan) n'est pas reconnu, il est transcrit approximativement, mais pas compris.
Fausse interprétation, mais bonne exploitation des indices	Ici les éléments saisis par l'élèves sont « *journaux, télévision* » et « *énerver* ». Dans le texte, un célèbre chanteur de rock a lancé une fausse rumeur pour se moquer des médias, racontant qu'il allait se marier. L'élève, avec les mots qu'il a entendus et compris, rappelle ce qui serait tout à fait plausible dans le contexte : « *Elle (la journaliste) lui demande si cela ne l'énerve pas trop qu'on le voit partout à la télévision et dans les journaux. Mais lui ça ne le dérange pas*. »
Le mot juste, des hypothèses solides	Une élève écrit « *... Et que plus tard, elle veut être* "*Schauspieler*"*, actrice, je suppose* ». Elle déduit donc le sens d'un mot et avoue au lecteur, dans le même temps, par son métacommentaire, qu'il lui était inconnu. De la même manière, un autre élève rappelle : « *Il m'a semblé entendre les mots* "*nouvelle année scolaire*"*, donc j'en déduis que le sondage se déroule à la rentrée*. »
Des commentaires personnels	« *Le quatrième est un garçon qui est dans la même chambre que son grand frère qui aime le métal et qui a des posters de groupes de métal, mais lui préfère les voitures, encore une fois, je n'en suis pas sûre*. » (…) » *J'ai compris certains éléments et il y en a d'autres dont je ne suis pas vraiment sûre, ce qui me permet une approximation de la compréhension de l'écoute*. »

L'évaluation (ici au sens de l'analyse) par l'enseignant des rappels des élèves peut l'aider à identifier les difficultés rencontrées et donc à ajuster son enseignement et les entraînements qu'il propose dans le but de les faire progresser.

D. Quelles activités d'entraînements pour la compréhension de l'oral ?

Nous avons vu précédemment que l'entraînement systématique des processus de compréhension, et notamment des processus de bas niveau, est susceptible de créer un « cercle vertueux » : leur automatisation progressive permettrait de libérer des ressources cognitives pour la réalisation de processus « de haut niveau », lesquels favorisent la construction progressive du sens et contribuent à leur tour à la diminution du coût cognitif global de l'activité (Roussel, 2019). Du point de vue didactique, deux types de démarches peuvent donc être envisagées pour l'entraînement à la CO : entrainer les processus de bas niveau et/ou favoriser et encourager la mise en œuvre des processus de haut niveau.

Roussel, Gruson et Galan (2019) ont mesuré l'efficacité de ces deux types d'entraînement chez des élèves de seconde apprenant l'anglais, l'allemand ou l'espagnol (*cf*. Roussel, 2019). Un premier groupe d'élèves s'est vu proposer, pendant trois semaines, des entraînements visant l'automatisation de processus de bas niveau : décodage, segmentation de la chaîne phonique, transcription, repérage de mots. Un deuxième groupe a effectué, sur les mêmes supports sonores, une série d'exercices visant à favoriser la mise en œuvre de processus de haut niveau : anticipation à partir du titre du document sonore, élaboration d'hypothèses, discussion en binômes sur le sens du document sonore. Cette expérimentation montre principalement que les effets des différents types d'entraînements sur les performances en compréhension sont différents voire inverses selon le niveau initial des élèves : alors

que les apprenants les moins compétents au départ ont significativement progressé grâce à l'entraînement des processus de bas niveau, les apprenants les plus compétents se montrent quant à eux plus sensibles à un entraînement des processus de haut niveau. Ce qui nous intéresse encore davantage, c'est bien de montrer quels types d'entraînement, notamment de bas niveau, ont permis aux élèves de progresser. Nous proposons ainsi les deux tableaux suivants (Tableau 4) qui « mettent face à face » les processus cognitifs des élèves et des exemples d'entraînement et de situations d'évaluation pour en faciliter la mise en œuvre.

Tableau 4 : Activités des élèves et processus de bas niveau correspondants (d'après Roussel, Gruson & Galan, 2019)

Activités pédagogiques	Processus de bas niveau
• Entourer dans une liste les mots entendus dans un extrait du document • Entourer dans une liste de paires de mots proches celui qui est effectivement prononcé (*must/most ; thing/sing ; work/walk*) • Souligner les mots accentués dans un extrait du document.	Reconnaissance lexicale et phonologique
• Compter le nombre de mots entendus dans un extrait du document • Compter le nombre de phrases entendues dans un extrait du document • Compléter les mots manquants dans une partie du script du document sonore	Segmentation du flux sonore
• Repérer dans plusieurs phrases extraites du document le couple sujet-verbe • Identifier et compter les verbes et les sujets dans une phrase extraite du document • Identifier dans un extrait les sujets, les verbes les adjectifs et les classer dans un tableau selon leur catégorie (adjectifs, adverbes, etc.)	Catégorisation grammaticale
• Transcrire une autre partie du document sonore	Renforcement du lien graphophonologique
• Indiquer les mots qui viennent spontanément à l'esprit en L1 après l'écoute d'un extrait du document	Traitement sémantique

Pour l'entraînement des processus de haut niveau, conformément à la proposition de Vandergrift et Tafaghodtari (2010, p. 476), les élèves peuvent, à partir du titre du seul document, formuler des hypothèses sur son contenu et sur les mots et expressions qu'ils pensent entendre puis vérifier leurs hypothèses au fil des écoutes en se concertant.

Tableau 5 : Activités des élèves et processus de haut niveau correspondants (d'après Vandergrift & Tafaghodtari, 2010, p. 476, repris par Roussel, Gruson & Galan, 2019)

Activités pédagogiques	Processus de haut niveau
Activités de pré-écoute : Planification/ prédiction	
1. Une fois que les élèves ont été informés du sujet et du type de texte, ils élaborent des hypothèses sur le type d'information et les mots qu'ils sont susceptibles d'entendre.	1. Planification et attention dirigée
Première écoute : vérification	
2. Les élèves vérifient leurs hypothèses initiales, les corrigent si besoin.	2. Attention sélective, contrôle et évaluation
3. Les élèves comparent ce qu'ils ont compris/écrit avec leurs pairs, modifient au besoin, déterminent ce qui doit encore être résolu et décident des détails importants qui nécessitent encore une attention particulière.	3. Contrôle de la tâche d'écoute, évaluation, planification et attention sélective.
Deuxième écoute : deuxième vérification	
4. Les élèves vérifient les points de désaccord antérieurs, apportent des corrections et notent les détails supplémentaires qu'ils ont compris.	4. Attention sélective, contrôle de la tâche d'écoute, évaluation, planification et attention sélective et résolution de problème
5. Discussion en groupe dans laquelle tous les élèves contribuent à la reconstruction des points principaux et des détails les plus pertinents du texte, entrecoupée de réflexions sur la façon dont les élèves sont parvenus à déduire le sens de certains mots ou passage du texte.	5. Monitoring, évaluation et résolution de problème
Troisième écoute : vérification finale	
6. Les élèves écoutent attentivement l'information révélée dans la discussion en groupe qu'ils n'ont pas été en mesure de déchiffrer plus tôt.	6. Attention sélective, contrôle de la tâche et résolution de problème
7. En s'appuyant sur la discussion antérieure sur les stratégies utilisées pour compenser ce qui n'a pas été compris, les élèves se fixent des objectifs pour la prochaine activité d'écoute.	7. Évaluation, planification

Il faut souligner, dans l'entraînement à la CO, le rôle plus général des consignes d'écoute (qui prennent ci-dessus la forme d'exercices ciblés) et des activités d'anticipation. Roussel et Tricot (2015) ont montré par exemple que lorsque des étudiants effectuent avant l'écoute des activités d'anticipation à propos du lexique qu'ils vont entendre dans le document sonore, leur attention porte sur les mots difficiles et le repérage des éléments qu'ils ont anticipés. Au contraire lorsque les activités d'anticipation portent sur les idées susceptibles d'être présentées dans le document sonore, alors l'attention porte sur le sens et sur le repérage de ces informations. Aucune des deux stratégies n'est plus efficace dans l'absolu, cela dépend de l'objectif d'apprentissage poursuivi.

Ainsi, en entrainant et en évaluant régulièrement (pas forcément de façon formelle !) les processus mis en œuvre par les élèves, l'enseignant peut accompagner leurs progrès en compréhension orale de manière ciblée. Nous avons en effet montré ici quelles formes peuvent prendre les exercices proposés aux élèves en fonction des processus de compréhension dont on souhaite accompagner la mise en œuvre.

La nécessité d'améliorer la CO en langue seconde (L2) est largement documentée par la littérature scientifique (Field, 2018 ; Vandergrift & Goh, 2012). Des recherches ont identifié un grand nombre de facteurs qui influencent la performance en CO en L2, comme la complexité linguistique et la longueur du discours (Brunfaut & Révész, 2015), la précision des questions de compréhension et les facteurs liés à l'auditeur qui incluent, nous l'avons vu, la conscience métacognitive (Vandergrift & Tafaghodtari, 2010), l'anxiété (Elkhafaifi, 2005), la motivation.

Il est donc difficile pour les professeurs dans leur classe de tenir compte de tous ces facteurs pour concevoir des tâches d'entraînement et évaluer progressivement les progrès de leurs élèves. Ces facteurs ont d'ailleurs, dans la littérature scientifique, été principalement examinés du point de vue des auditeurs et du document sonore écouté, sans être étendus à l'effet de l'enseignement de la compréhension. On ne saurait donc trop insister sur l'importance d'enseigner aux apprenants comment devenir des auditeurs efficaces (Turan Öztürk & Tekin, 2020 ; Roussel, 2021) et d'analyser et d'évaluer leurs stratégies d'écoute.

Nous avons montré que les descripteurs de compréhension de l'oral présentaient certains avantages dans une perspective communicative bienveillante d'évaluation de la CO, puisqu'il ne s'agit plus de tout comprendre, mais bien d'être capable de repérer certains éléments dans un contexte précis pour effectuer une tâche. Cependant, l'absence d'indications sur l'entraînement des élèves fait de ces textes des objets de critiques, en concentrant l'attention sur l'évaluation de l'objectif fixé (l'atteinte d'un certain niveau A1, A2, etc.) au détriment de l'évaluation des processus qui permettent d'atteindre cet objectif. Du côté de la recherche, la connaissance des processus cognitifs qui sous-tendent la CO peut permettre la conception d'activités d'entraînement et de situation d'évaluation pour les élèves. L'exploitation des protocoles verbaux met en évidence les stratégies utilisées par les élèves pour comprendre et a donné lieu à la construction d'outils d'auto-évaluation qui peuvent les accompagner dans le développement de leur compétence. L'exploitation de ces protocoles dans des dimensions plus syntaxiques et lexicales ouvre également des pistes pour compléter les évaluations de la CO dans les classes de langues. Il en découle cependant, et c'est bien la conclusion la plus importante, que ce qui pourrait être évalué au-delà et même indépendamment du produit, du résultat de la compréhension, concerne l'exploitation des éléments repérés : l'élève parvient-il à mettre en lien ce qu'il entend pour construire du sens ? L'élève utilise-t-il certaines stratégies, comment évoluent-elles tout au long d'une année ? Une manière d'utiliser l'évaluation pour favoriser les apprentissages, c'est d'évaluer non pas seulement l'atteinte de l'objectif de compréhension, mais les processus qui permettent d'atteindre cet objectif. L'important, ce n'est pas la destination, mais le voyage en lui-même…

Références

Bailly, N. & M. Cohen. (2009). L'approche communicative. http://flenet.rediris.es/tourdetoile/NBailly_MCohen.html (consulté en avril 2022).

Brunfaut, T., & Revesz, A. (2015). The role of task- and listener-characteristics in second language listening. *TESOL Quarterly*, *49*(1), 141-168. https://doi.org/10.1002/tesq.168

Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues – Apprendre, enseigner, évaluer.* Didier.

Conseil de l'Europe (2018). *Cadre européen commun de référence pour les langues – apprendre, enseigner, évaluer. Volume complémentaire avec nouveaux descripteurs.* Disponible en ligne : https://rm.coe.int/cecr-volume-complementaire-avec-de-nouveaux-descripteurs/16807875d5

Eduscol (2016). Evaluation cycle 4, langues vivantes. https://eduscol.education.fr/document/14833/download

Eduscol : (https://eduscol.education.fr/document/14833/download)

Elkhafaifi, H. (2005). The effects of prelistening activities on listening comprehension in Arabic learners. Foreign Language Annals, 38, 505–513.

Field, J. (2018) *Rethinking the Second Language Listening Test*. Equinox

Gaonac'h, D. et Fayol, M. (Eds.), (2003). Aider les élèves à comprendre du texte au multimédia. Hachette – Collection Profession Enseignant.

Goh, C. M. (1998). How ESL learners with different listening abilities use comprehension strategies and tactics. *Language Teaching Research*, *2*(2), 124-147.

Goh, C.M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*, 55-75.

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System*, *36*, 85-93.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Maurer B. & Puren, C. (2019). *CECR : par ici la sortie !* Éditions des archives contemporaines.

Ministère de l'Education national (2015) *Bulletin officiel n°17 du 23 avril 2015.* https://www.education.gouv.fr/bo/15/Hebdo17/MENE1506516D.htm?cid_bo=87834

Puren, C. (2019). Critique du CECRL. https://culture-fle.de/paroles-dexperts-critique-du-cecrl-avec-christian-puren/

Montaigu, R. & Nicodème, R. (2010). Rapports IGEN https://ww2.ac-poitiers.fr/langues/spip.php?article154

Roussel, S., & Tricot, A. (2015). Effet de l'élaboration d'hypothèses sur la compréhension de l'oral et sur les stratégies d'autorégulation de l'écoute en langue seconde : une étude empirique. *Alsic*, (Vol. 18, n° 1). Disponible en ligne : https://doi.org/10.4000/alsic.2788

Roussel S., Rieussec A., Nespoulous J-L., & Tricot A. (2008). Des baladeurs MP3 en classe d'allemand - L'effet de l'autorégulation matérielle de l'écoute sur la compréhension auditive en langue seconde, *Alsic*, http://alsic.u-strasbg.fr/v11/roussel/alsic_v11_04-rec1.htm

Roussel, S., & Gaonac'h, D (2017). L'apprentissage des langues. Éditions Retz.

Roussel, S. (2019). Quels entraînements permettent d'améliorer les compétences de compréhension de l'oral ? Cnesco. http://www.cnesco.fr/wp-content/uploads/2019/04/CCLV_Roussel_MEF-v2.pdf

Roussel, S. (2021) *L'approche cognitive en didactique des langues*. De Boeck.

Roussel, S., Gruson, B., & Galan, J-P (2019) What Types of Training Improve Learners' Performances in Second Language Listening Comprehension? *International Journal of listening, 33*(1), 39-52.

Tardieu, C. (2010). Votre B1 est-il mon B1 ? L'interculturel dans les tests d'évaluation en Europe, *Recherches en didactique des langues et des cultures*, 7-2. https://doi.org/10.4000/rdlc.2301

Turan Öztürk, D., & Tekin, S. (2020). Encouraging extensive listening in language learning. *Language Teaching Research Quarterly, 14*, 80-93.

Vandergrift, L., & Goh, C. (2012). *Teaching and learning second language listening. Metacognition in action*. Routledge.

Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning, 53*, 463– 496.

Vandergrift, L., Goh, C., Mareschal, C., & Tafaghodatari, M.H. (2006). The Metacognitive Awareness Listening Questionnaire (MALQ): Development and validation. *Language Learning, 56*, 431-462.

Vandergrift, L. & Tafaghodtari, M. H., (2010). Teaching students how to listen does make a difference: An empirical study. *Language Learning, 60*, 470-497.

Zoghlami, N. (2015). La compréhension de l'anglais oral (L2) : processus cognitifs et comportements stratégiques, *Recherche et pratiques pédagogiques en langues de spécialité*, *3.* Disponible en ligne : http://journals.openedition.org/apliut/5322

Comment les enseignants peuvent-ils partager la responsabilité de l'évaluation avec les élèves ?

FERNANDO MORALES VILLABONA

HAUTE ÉCOLE PÉDAGOGIQUE DU CANTON DE VAUD (SUISSE)

Introduction

L'implication des élèves dans les processus évaluatifs n'est pas une idée nouvelle en éducation, mais elle a pris de l'ampleur grâce à l'intérêt grandissant que des perspectives développées depuis les années 1990 portent au développement de compétences pour la vie (Black & Wiliam, 1998) et à la promotion d'un apprentissage autodirigé (Davis *et al*., 2007). Lorsque les enseignants confient une partie de la responsabilité de l'évaluation aux élèves, ces derniers ont la possibilité de construire et s'approprier des connaissances en matière d'évaluation leur permettant, par exemple, de porter un jugement sur la qualité d'un travail au regard de critères multiples (Sadler, 1998). Le concept d'évaluation formatrice (Nunziati, 1990), dans le contexte francophone, et celui de l'évaluation en tant qu'apprentissage (*assessment as learning* – Earl, 2003), dans le contexte anglophone, insistent tous les deux sur le rôle actif, engagé et critique de l'élève en tant qu'évaluateur de son propre travail. Mais les élèves peuvent également être actifs en évaluant des travaux produits par autrui (par leurs pairs, notamment), l'idée de base étant toujours qu'ils deviennent des acteurs – à la place d'être des consommateurs – de l'évaluation (Sadler, 1998). Dans ce type d'approche, l'évaluation n'est plus seulement l'affaire de l'enseignant : ce qui est visé est une plus grande responsabilisation et autonomie de l'élève face à son apprentissage (Mottier Lopez, 2015).

Pour répondre à la question du partage de la responsabilité de l'évaluation entre l'enseignant et les élèves, ce texte : 1) rend compte de la place des critères d'évaluation, 2) présente les modalités d'implication de l'élève reconnues dans la littérature scientifique, 3) met en évidence les difficultés associées aux pratiques évaluatives impliquant les élèves, et 4) aborde la dimension collaborative pouvant être intégrée à ce type de pratiques.

I. Le rôle des critères d'évaluation

La littérature scientifique insiste abondamment sur l'importance de définir au mieux ce qui est attendu de la part des élèves (McMillan & Hearn, 2008 ; Sadler, 1998 ; Sebba *et al*., 2008 ; Veslin & Veslin, 2001). L'explicitation des attentes de l'enseignant passe généralement par la définition des objectifs d'apprentissage, puis des *critères d'évaluation.* Le critère d'évaluation est défini comme un « point de vue » (Scallon, 1988 ; Roegiers, 2005) à partir duquel la production de l'élève est évaluée. En ce, il est l'expression de ce qu'on valorise en référence aux dimensions essentielles d'un apprentissage en jeu (Harlen, 2005 ; Mottier Lopez, 2015). Concrètement, les critères définissent les objectifs

d'apprentissage en termes opérationnels pour les élèves (Brookhart, 2022), généralement sous la forme d'énoncés qui explicitent les qualités attendues.

Black *et al*. (2004) soulignent que les critères d'évaluation doivent être « transparents » pour les élèves afin qu'ils puissent se construire une représentation claire des objectifs visés et de ce qui sera considérée comme une production de « bonne qualité ». L'idée d'une transparence évoquée par ces auteurs est certainement à comprendre au regard de l'importance de rendre les critères explicites pour les élèves, tout en acceptant que cette tâche est loin d'être simple. Dès lors, un travail important d'accompagnement de l'enseignant semble indispensable pour aider les élèves à comprendre et à s'approprier la signification et l'utilité (ainsi que l'usage) des critères d'évaluation. Ce travail peut s'appuyer sur différentes stratégies, telles que la mise en place de discussions collectives (entre l'enseignant et la classe), la pratique régulière de feedbacks ou encore la proposition d'exemples concrets pour modéliser l'usage des critères dans l'évaluation d'une production (Andrade & Valtcheva, 2009 ; Black *et al*., 2004 ; Ross, 2006). D'ailleurs, ce partage des critères d'évaluation avec les élèves peut aussi faciliter leur introduction progressive dans des démarches d'autoévaluation ou d'évaluation par les pairs (Balan & Jönsson, 2018).

La recherche montre qu'il existe des liens positifs entre l'utilisation des critères d'évaluation par les élèves et le développement de l'autorégulation de ces derniers, que ce soit au niveau de la planification de l'activité à réaliser, du suivi de celle-ci (c'est-à-dire, du progrès vers l'objectif poursuivi) et de la réflexion métacognitive en cours ou en fin d'apprentissage (Braund & Deluca, 2018 ; Panadero *et al*., 2016), et cela dans différents contextes et niveaux scolaires. De plus, le fait de connaître et d'utiliser les critères d'évaluation est associé à une plus grande motivation des élèves dans la réalisation des tâches scolaires (Fletcher, 2016) et à une perception accrue de l'utilité des commentaires (feedback) des enseignants (Vattøy & Smith, 2019).

II. Les modalités d'implication de l'élève

Trois modalités impliquant les élèves dans les processus évaluatifs sont largement reconnues par la littérature (et ceci dans différentes langues de publication, dont le français, l'anglais et l'espagnol). Il s'agit de l'autoévaluation, l'évaluation par les pairs et la coévaluation.

Dans **l'autoévaluation**, l'élève évalue sa propre production ou ses démarches de réalisation d'une tâche (Allal, 1999). L'élève est amené à poser un regard critique sur son travail et, en plus, un regard unique compte tenu que, finalement, c'est lui le seul à savoir ce qu'il a voulu faire (Hadji, 2015). Ici, l'élève est amené à formuler des jugements conscients sur sa propre production ou processus d'apprentissage, à évaluer ses manières d'agir lorsqu'il apprend, et à identifier des stratégies et actions à entreprendre pour améliorer sa compréhension et développer ses compétences (Andrade & Valtcheva, 2009 ; McMillan & Hearn, 2008). Des revues qui synthétisent les principaux résultats des études dans la matière (Andrade, 2019 ; Brown & Harris, 2013 ; Panadero *et al.*, 2017) font état des liens positifs entre les pratiques d'autoévaluation et le développement des capacités d'autorégulation des élèves et, par ce biais, l'amélioration de leurs résultats scolaires. L'autoévaluation se met ainsi au service de la progression des apprentissages des élèves (Allal, 2007 ; Mottier Lopez, 2012). Toutefois, la recherche doit encore approfondir cette question des effets de l'autoévaluation sur les apprentissages des élèves (qui peuvent varier en fonction du contexte, des caractéristiques des élèves ou encore du type d'autoévaluation proposée), afin de mieux comprendre les mécanismes cognitifs et affectifs qui entrent en jeu (Andrade, 2019).

L'autoévaluation représente un exercice complexe pour l'élève, justement car il implique une prise de recul par rapport à la réalisation ou aux résultats de son activité scolaire, afin de les apprécier. En effet, l'élève n'a pas forcément les compétences nécessaires pour décider tout seul des buts des actions d'apprentissage et d'évaluation qui interviennent dans les situations didactiques et les tâches proposées à l'école. Or, s'autoévaluer ne signifie pas forcément pour l'élève d'être responsable de toutes les composantes de l'évaluation (définition des objectifs et critères, recueil des informations, interprétation de celles-ci, prise de décision). L'implication de l'élève peut être partielle et se faire de manière progressive. Elle peut se limiter, par exemple, à l'identification de critères pour que l'élève produise des jugements concernant son propre travail ou bien être plus exigeante, si l'élève doit, en plus, produire des jugements (sur son travail, son niveau de réussite, son processus d'apprentissage), voire prendre des décisions sur les actions qu'il doit entreprendre pour continuer à progresser (Deakin-Crick *et al*., 2005). McMillan et Hearn (2008), proposent un modèle dont la logique est celle de donner de plus en plus de responsabilité et de liberté d'action aux élèves en ce qui concerne la définition des critères d'évaluation, l'étayage concernant leur usage, la manière d'utiliser le feedback et la définition de nouveaux objectifs et de nouvelles stratégies d'apprentissage. Ce modèle, praticable dès l'école primaire, vise à soutenir les enseignants dans la mise en place de démarches d'autoévaluation dans leurs classes et il peut être adapté en fonction du niveau de scolarisation des élèves et des caractéristiques de la classe. Ainsi, par exemple, selon le niveau de compréhension des élèves à propos des critères d'évaluation, l'enseignant peut demander aux élèves d'identifier ces critères en examinant des exemples contrastés (de « bonnes » et de « mauvaises » productions) ou leur proposer directement des explications sur la signification et l'usage des critères définis.

L'évaluation par les pairs implique deux ou plusieurs élèves qui, partageant un même statut dans la situation didactique, évaluent leurs productions ou démarches respectives ou conjointes (Allal, 1999). Dans cette modalité, différentes pratiques sont possibles puisque, comme le montrent des études, il est possible de faire varier un nombre important de facteurs (Gielen *et al*., 2011 ; Strijbos *et al*., 2009 ; Topping, 1998), par exemple : la direction ou l'orientation de la relation d'évaluation (unilatérale, réciproque) ; la fonction de l'évaluation (sommative, formative) ; le caractère public, privé ou anonyme de l'évaluation ; le nombre de participants impliqués (un apprenant à la fois, des apprenants réunis en groupes).

Les travaux scientifiques sur l'évaluation entre pairs soulignent un ensemble d'aspects bénéfiques, en rapport avec cette pratique, susceptibles de contribuer à la progression des élèves (Double *et al*., 2020). Notamment, le fait que les élèves semblent accepter plus facilement (voire plus sérieusement) les remarques et commentaires (feedback) fournis par leurs pairs, comparativement à ceux qui proviennent de l'enseignant et que, par ce biais, un dialogue autour de l'évaluation serait possible car les élèves oseraient plus facilement interrompre, contredire, poser des questions, ce qui ne serait pas forcément le cas face à l'enseignant (Black *et al*., 2004 ; Hadji, 2015). Les recherches empiriques montrent aussi que la participation des élèves à des démarches d'évaluation par les pairs favorise le déclenchement de conflits sociocognitifs et la mise en œuvre de régulations métacognitives (Allal & Michel, 1993), ainsi qu'un plus grand engagement des élèves dans le processus d'apprentissage et le développement de compétences évaluatives et réflexives (Sebba *et al*., 2008). Il convient de signaler que la littérature scientifique reconnaît aussi que davantage de travaux sont nécessaires pour confirmer (ou infirmer) les bénéfices attribués à l'évaluation entre pairs en matière d'apprentissages possibles, notamment en raison de la grande variété de pratiques existantes (Strijbos & Sluijsmans, 2010).

Enfin, dans **la coévaluation**, l'apprenant confronte son autoévaluation à l'évaluation réalisée par l'enseignant. Le but est que cette confrontation conduise à un dialogue entre les acteurs concernés à propos des deux évaluations effectuées à propos du même objet (Paquay *et al*., 2001). Ce dialogue peut revêtir un caractère formatif, lorsqu'il vise l'identification des capacités à maintenir, des difficultés à surmonter, des moyens à utiliser pour progresser. Mais il faut souligner que la recherche (surtout les publications anglophones, dans des contextes d'enseignement supérieur) s'est souvent intéressée à étudier la coévaluation dans une perspective sommative, en se focalisant sur les corrélations entre les notes produites par l'autoévaluation des apprenants et celles attribuées par l'enseignant à une même production. Autrement dit, il existe peu d'études sur la coévaluation dans une perspective formative.

Or, dans certains travaux, la notion de coévaluation est synonyme d'une évaluation *collaborative*. Dans ce cas, elle implique souvent une discussion entre l'enseignant et les apprenants pour s'accorder non seulement sur la décision finale de l'évaluation mais aussi, en amont, sur les critères d'évaluation (Gouli *et al.*, 2008), voire sur les objets évalués et les rôles des acteurs (enseignant, apprenants) dans l'évaluation (Sommervell, 1993). Cet élargissement de la compréhension de la coévaluation est intéressant en ce qu'il évoque une possible négociation et une construction collective de l'évaluation associant enseignant et apprenants, ce qui représente tout un champ de possibilités dans une perspective de soutien aux apprentissages. Nous reviendrons un peu plus tard sur cette dimension collaborative de l'évaluation, pouvant concerner la relation enseignant-élève(s) mais également la relation élève(s)-élève(s).

III. Limites à l'implication des élèves dans les pratiques évaluatives

Malgré les avantages reconnus à l'implication des apprenants dans les processus d'évaluation, celle-ci peut comporter d'importantes tensions et résistances. Des études menées donnent à voir un ensemble de préoccupations de la part des apprenants lorsqu'ils sont impliqués dans des démarches évaluatives. Parmi ces préoccupations, on retrouve que :

- Les élèves indiquent souvent ne pas avoir confiance en eux-mêmes, ni en leurs pairs, en tant qu'évaluateurs et continuent à voir dans la figure de l'enseignant celle de l'expert et, donc, représentant la seule source « objective » de l'évaluation (Brown *et al*., 2009 ; Peterson & Irving, 2008). Par exemple, dans l'étude de Ross et al. (1998), des élèves de 23 classes (au primaire et au secondaire) ayant participé à des démarches d'autoévaluation ont été interviewés. Dans ce cadre, certains considéraient que l'autoévaluation ouvrait la porte à la tricherie (quand quelqu'un surévalue son travail) et d'autres s'opposaient à ce type de démarche estimant manquer d'expertise pour juger leur propre travail.

- Un certain malaise psychologique et relationnel est associé au fait d'être jugé ou de devoir juger les autres (Brown & Harris, 2013). Ce constat, mis en avant notamment par des travaux menés dans l'enseignement tertiaire (Gueldenzoph & May, 2002 ; Hanrahan & Isaacs, 2001), apparaît également dans des études réalisées à l'école. C'est le cas de celle de Vanderhoven *et al*. (2015), dans laquelle 69 élèves de secondaire (15-16 ans) participaient à des démarches d'évaluation par les pairs et en présentaient leurs perceptions à travers un questionnaire. Les résultats de cette étude confirment que les élèves peuvent expérimenter de l'inconfort, de l'anxiété, voire de la peur, quand ils doivent évaluer ou quand ils sont évalués par des pairs.

- Des effets sociaux ou de réciprocité entre élèves peuvent aussi entraver les pratiques évaluatives. Plus précisément, dans des démarches d'évaluation entre pairs, on peut observer des phénomènes de surévaluation ou sous-évaluation des travaux liés à des rapports d'amitié et/ou de pouvoir entre les élèves (Sluijsmans *et al*., 1999), d'autant plus lorsque le dispositif étudié comporte une composante certificative (notamment, en lien avec des pratiques de notation).

Les croyances et représentations des enseignants par rapport à l'implication des élèves dans le processus évaluatif peuvent également faire obstacle. Certains craignent un manque d'expérience chez l'élève et admettent se sentir mal à l'aise avec un changement de rôle qui donne plus de contrôle aux élèves (Falchikov, 2004). L'implication des élèves est souvent faite de manière superficielle, leurs introspections étant souvent utilisées pour justifier l'évaluation faite par l'enseignant, ôtant à l'autoévaluation son caractère formatif (Kelly & Brandes, 2008). Peu d'enseignants sembleraient réellement prêts à accepter de dévoluer une part de responsabilité de l'évaluation à leurs élèves (McConnell, 1999). À ce propos, l'étude de Noonan et Duncan (2005) est particulièrement parlante : 75 % des 118 enseignants de secondaire au Canada participant à une enquête de large envergure, affirment proposer (de manière plus ou moins régulière) des démarches d'autoévaluation ou d'évaluation par les pairs dans leurs classes. Toutefois, la moitié des enseignants émet des réserves quant à l'utilité de ces démarches et hésitent à les mettre en place, considérant notamment que les élèves ne possèdent pas les compétences nécessaires pour évaluer.

Malgré ces limites, dont les conséquences peuvent être lourdes, la littérature scientifique souligne l'intérêt de mettre en œuvre dans la classe des pratiques évaluatives impliquant les élèves. Pour Allal (1999), la prise de conscience de ces difficultés potentielles représente un premier pas vers une construction éclairée de démarches d'évaluation impliquant les élèves, susceptibles de soutenir la régulation des apprentissages en situation scolaire. Pour éviter certains écueils, l'auteure propose de privilégier, dans un premier temps, la participation des élèves à des démarches d'évaluation par les pairs et de coévaluation, afin qu'ils développent des compétences évaluatives à exploiter ensuite dans des situations d'autoévaluation. De manière plus générale, la recherche insiste sur l'importance d'apprendre aux élèves à s'autoévaluer (ou à évaluer des pairs). En ce sens, Panadero *et al*. (2016) soulignent que les enseignants ne peuvent pas partir du principe que les élèves sauront s'autoévaluer et, donc, ces derniers ont besoin d'expérimenter ce type de démarches et d'être accompagnés dans le processus. Les auteurs signalent aussi que de nombreuses études empiriques ont démontré que des élèves de primaire et du secondaire peuvent développer des compétences d'autoévaluation. Dans des études s'intéressant à cette « formation » à l'autoévaluation, les stratégies pédagogiques qui combinent le recours à des critères d'évaluation, l'utilisation de modèles ou d'exemples, et le feedback des enseignants, apparaissent particulièrement porteuses (Brown & Harris, 2013 ; Panadero & Alonso-Tapia, 2013). De ce fait, certaines conditions apparaissent nécessaires au développement des compétences d'évaluation des élèves, telles que : une définition claire des critères d'évaluation, l'étayage de l'enseignant pendant la démarche évaluative, l'opportunité de s'exercer régulièrement à l'autoévaluation (ou à une autre modalité) et, enfin, des occasions pour corriger et améliorer le travail évalué (Goodrich, 1996).

IV. Dimension collaborative dans les démarches évaluatives impliquant les élèves

Comme signalé plus haut, des travaux scientifiques s'intéressent en particulier à la dimension de collaboration qui est parfois associée aux modalités d'évaluation dans lesquelles l'apprenant est fortement impliqué : l'autoévaluation, l'évaluation entre pairs, la coévaluation. Falchikov (2004), par exemple, signale que les démarches d'autoévaluation et d'évaluation entre pairs peuvent inclure un certain degré de collaboration entre les enseignants et les apprenants, lorsque les deux parties discutent et s'accordent sur les critères d'évaluation. L'auteure soulève que si la plupart des travaux portant sur l'autoévaluation et sur l'évaluation entre pairs se focalisent sur le travail individuel de chaque apprenant, ces deux modalités évaluatives peuvent aussi être déployées dans le contexte du travail de groupe – par exemple, quand l'évaluation porte sur la contribution individuelle de chacun au travail du groupe ou quand un groupe évalue un autre groupe. Dans le même esprit (prise en compte des interactions dans un groupe ou entre groupes), Prins *et al*. (2005) indiquent que la collaboration dans des démarches d'évaluation entre pairs peut concerner non seulement la définition des critères d'évaluation, mais également le jugement porté sur les productions des autres et les processus de feedback qui servent (potentiellement) à y apporter des améliorations. En ce sens, le rôle de l'enseignant apparaît fondamental, pour cadrer ce processus, en fournissant des guides ou des règles sur la manière de donner et aussi d'accepter (et gérer) le feedback.

Le rapport entre évaluation et collaboration est parfois associé également aux pratiques de coévaluation, insistant ainsi davantage sur le rôle actif de l'enseignant. Dans ces cas, très souvent, les termes *coévaluation* et *évaluation collaborative* sont synonymes (Gómez Ruiz & Quesada Serra, 2017 ; Gouli *et al*., 2008). Dans ce type de conceptualisation, des dynamiques dialogiques (de négociation, d'argumentation, de construction de consensus, de responsabilité partagée du processus) caractérisent une collaboration possible entre les apprenants et l'enseignant. Dans une telle démarche (et de manière similaire aux propositions sur l'évaluation par les pairs), la collaboration concerne notamment la définition des objectifs et critères d'évaluation ainsi que la prise de décision (notamment sur le résultat de l'évaluation).

Notre compréhension de l'évaluation collaborative tient compte, à la fois, de la collaboration possible entre élèves (notamment lorsqu'ils travaillent à plusieurs) et entre l'enseignant et les élèves. Nos travaux (Morales Villabona, 2019, 2020 ; Morales Villabona & Mottier Lopez, 2020) montrent que les dynamiques d'interaction entre élèves sont diverses quand on leur demande de formuler une évaluation partagée sur leurs productions propres ou sur celles d'autrui. Malgré leur manque d'expérience en tant qu'évaluateurs, les élèves apparaissent capables d'expliquer et de justifier leurs décisions évaluatives. Leur participation à l'évaluation apparaît favorable à l'autorégulation de l'élève, notamment en soutenant le développement du regard critique des élèves (y compris la mise à distance nécessaire pour juger une production – propre ou d'autrui) et de la compétence à repérer, interpréter et confronter des informations dans le but de prendre des décisions évaluatives et de les justifier. En revanche, pour nous, la question de la participation de l'enseignant dans ce type de démarche reste encore à explorer, notamment pour tenter de comprendre comment l'enseignant peut gérer à mieux la double contrainte de collaborer avec les élèves pendant l'évaluation et de les accompagner pour assurer la collaboration entre pairs dans le groupe.

Références

Allal, L. (1999). Impliquer l'apprenant dans le processus d'évaluation : promesses et pièges de l'autoévaluation. In C. Depover & B. Noël (dir.), *L'évaluation des compétences et des processus cognitifs : modèles, pratiques et contextes* (pp. 35-56). De Boeck.

Allal, L. (2007). Régulations des apprentissages : orientations conceptuelles pour la recherche et la pratique en éducation. In L. Allal & L. Mottier Lopez (dir.), *Régulation des apprentissages en situation scolaire et en formation* (pp. 7-23). De Boeck.

Allal, L. & Michel, Y. (1993). Autoévaluation et évaluation mutuelle en situation de production écrite. In L. Allal, D. Bain & P. Perrenoud (dir.), *Evaluation formative et didactique du français* (pp. 239-264). Delachaux et Niestlé.

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*(87). https://doi.org/10.3389/feduc.2019.00087

Andrade, H. & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into practice, 48*(1), 12-19. https://doi.org/10.1080/00405840802577544

Balan, A. & Jönsson, A. (2018). Increased explicitness of assessment criteria: Effects on student motivation and performance. *Frontiers in Education, 3*(81), 1-13. https://doi.org/10.3389/feduc.2018.00081

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 8-21. https://doi.org/10.1177/003172170408600105

Braund, H. & Deluca, C. (2018). Elementary students as active agents in their learning: An empirical study of the connections between assessment practices and student metacognition. *Australian Journal of Educational Research, 45*, 65-85. https://doi.org/10.1007/s13384-018-0265-z

Brookhart, S. (2022). Criteria connect grading with other assessment practices. *La Revue LEeE, 6*. https://doi.org/10.48325/rleee.006.01

Brown, G. T. L. & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (dir.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). SAGE.

Brown, G. T. L., Irving, S. E., Peterson, E. R. & Hirschfeld, G. H. F. (2009). Use of interactive–informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction, 19*(2), 97-111. https://doi.org/10.1016/j.learninstruc.2008.02.003

Davis, N. T., Kumtepe, E. G. & Aydeniz, M. (2007). Fostering Continuous Improvement and Learning Through Peer Assessment: Part of an Integral Model of Assessment. *Educational Assessment, 12*(2), 113-135. https://doi.org/10.1080/10627190701232720

Deakin-Crick, R., Sebba, J., Harlen, W., Guoxing, Y. & Lawson, H. (2005). *Systematic review of research evidence of the impact on students of self- and peer-assessment. Protocol.* EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Double, K. S., McGrane, J. A. & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review, 32*(2), 481-509. https://doi.org/10.1007/s10648-019-09510-3

Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximise student learning*. Corwin.

Falchikov, N. (2004). Involving students in assessment. *Psychology Learning & Teaching, 3*(2), 102-108. https://doi.org/10.2304/plat.2003.3.2.102

Fletcher, A. K. (2016). Exceeding expectations: Scaffolding agentic engagement through assessment as learning. *Educational Research, 58*(4), 400-419. https://doi.org/10.1080/00131881.2016.1235909

Gielen, S., Dochy, F. & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137-155. https://doi.org/10.1080/02602930903221444

Gómez Ruiz, M. Á. & Quesada Serra, V. (2017). Coevaluación o evaluación compartida en el contexto universitario: la percepción del alumnado de primer curso. *Revista Iberoamericana de Evaluación Educativa, 10*(2), 9-30.

Goodrich, H. (1996). *Student self-assessment: At the intersection of metacognition and authentic assessment* (thèse de doctorat inédite). Harvard University.

Gouli, E., Gogoulou, A. & Grigoriadou, M. (2008). Supporting self-, peer-, and collaborative-assessment in e-learning: the case of the PEer and Collaborative ASSessment Environment (PECASSE). *Journal of Interactive Learning Research, 19*(4), 615-647.

Gueldenzoph, L. E. & May, G. L. (2002). Collaborative peer evaluation: Best practices for group member assessments. *Business Communication Quarterly, 65*(1), 9-20. https://doi.org/10.1177/108056990206500102

Hadji, C. (2015). *L'évaluation à l'école, pour la réussite de tous les élèves*. Editions Nathan.

Hanrahan, S. J. & Isaacs, G. (2001). Assessing self-and peer-assessment: The students' views. *Higher Education Research & Development, 20*(1), 53-70. https://doi.org/10.1080/07294360123776

Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal, 16*(2), 207-223. https://doi.org/10.1080/09585170500136093

Kelly, D. M. & Brandes, G. M. (2008). Equitable classroom assessment: Promoting self-development and self-determination. *Interchange, 39*(1), 49-76. https://doi.org/10.1007/s10780-008-9041-8

McConnell, D. (1999). Examining a collaborative assessment process in networked lifelong learning. *Journal of Computer Assisted Learning, 15*(3), 232-243. https://doi.org/10.1046/j.1365-2729.1999.153097.x

McMillan, J. H. & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons, 87*(1), 40-49. http://www.jstor.org/stable/42923742

Morales Villabona, F. (2019). *Comprendre l'évaluation collaborative en classe : analyses de deux démarches intégrées à la production de textes écrits à l'école primaire genevoise* (thèse de doctorat inédite). Université de Genève.

Morales Villabona, F. (2020). Collaboration dans l'évaluation et processus de régulation : étude de cas d'une démarche d'évaluation collaborative intergroupe à l'école primaire. *La Revue LEeE, 3*. http://revue.leee.online/index.php/info/article/view/59

Morales Villabona, F. & Mottier Lopez, L. (2020). Dans un contexte d'évaluation collaborative : Étude à l'école primaire d'une démarche d'autoévaluation de groupe en production écrite en français. Dans S. Cartier & J.-L. Berger (dir.), *Prendre en charge son apprentissage. L'apprentissage autorégulé à la lumière des contextes* (p. 115-143). L'Harmattan.

Mottier Lopez, L. (2012). *La régulation des apprentissages en classe*. De Boeck.

Mottier Lopez, L. (2015). *Évaluations formative et certificative des apprentissages. Enjeux pour l'enseignement*. De Boeck.

Noonan, B. & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research and Evaluation, 10*(17), 1-8. https://doi.org/10.7275/a166-vm41

Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers pédagogiques, 280*, 47-64.

Panadero, E. & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551-576. http://dx.doi.org/10.14204/ejrep.30.12200

Panadero, E., Brown, G. T. & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational psychology review, 28*(4), 803-830. https://doi.org/10.1007/s10648-015-9350-2

Panadero, E., Jonsson, A. & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review, 22*, 74-98. https://doi.org/10.1016/j.edurev.2017.08.004

Paquay, L., Darras, E. & Saussez, F. (2001). Les représentations de l'autoévaluation. In G. Figari & M. Achouche (dir.), *L'activité évaluative réinterrogée, regards scolaires et socioprofessionnels* (pp. 119-133). De Boeck Université.

Peterson, E. R. & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction, 18*(3), 238-250. https://doi.org/10.1016/j.learninstruc.2007.05.001

Prins, F. J., Sluijsmans, D. M. A, Kirschner, P. A. & Strijbos, J. W. (2005). Formative peer assessment in a CSCL environment: A case study. *Assessment & Evaluation in Higher Education, 30*, 417-444. https://doi.org/10.1080/02602930500099219

Roegiers, X. (2005). L'évaluation selon la pédagogie de l'intégration : est-il possible d'évaluer les compétences des élèves ? In K. Toualbi-Thaâlibi & S. Tawil (dir.), *La refonte de la pédagogie en Algérie. Défis et enjeux d'une société en mutation* (pp.107-124). UNESCO-ONPS.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation, 11*(10), 1-13. https://doi.org/10.7275/9wph-vv65

Ross, J. A., Rolheiser, C. & Hogaboam-Gray, A. (1998). Skills training versus action research in-service: impact on student attitudes to self-evaluation. *Teaching and Teacher Education, 14*(5), 463-477. https://doi.org/10.1016/S0742-051X(97)00054-1

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in education: Principles, Policy & Practice, 5*(1), 77-84. https://doi.org/10.1080/0969595980050104

Scallon, G. (1988). *L'évaluation formative des apprentissages*. Presses Université Laval.

Sebba, J., Deakin-Crick, R., Yu, G., Lawson, H., Harlen, W. & Durant, K. (2008). *Systematic review of research evidence of the impact on students in secondary schools of self and peer assessment. Technical report.* EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Sluijsmans, D. M. A., Dochy, F. & Moerkerke, G. (1999). Creating a learning environment by using self-peer- and co-assessment. *Learning Environments Research, 1*, 293–319. https://doi.org/10.1023/A:1009932704458

Somervell, H. (1993). Issues in assessment, enterprise and higher education: The case for self-, peer- and collaborative assessment. *Assessment & Evaluation in Higher Education, 18*(3), 221-233. https://doi.org/10.1080/0260293930180306

Strijbos, J. W. & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269. https://doi.org/10.1016/j.learninstruc.2009.08.002

Strijbos, J. W., Ochoa, T. A., Sluijsmans, D. M., Segers, M. S. & Tillema, H. H. (2009). Fostering interactivity through formative peer assessment in (web-based) collaborative learning environments. In C. Mourlas, N. Tsianos & P. Germanakos (dir.), *Cognitive and emotional processes in web-based education: Integrating human factors and personalization* (pp. 375-395). IGI Global.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276. https://doi.org/10.3102/00346543068003249

Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T. & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education, 81*, 123-132. http://dx.doi.org/10.1016/j.compedu.2014.10.001

Vattøy, K. & Smith, K. (2019). Students' perceptions of teachers' feedback practice in teaching English as a foreign language. *Teaching and Teacher Education, 85*, 260-268. https://doi.org/10.1016/j.tate.2019.06.024

Veslin, O. & Veslin, J. (2001). Évaluation formatrice et critères de réalisation. In G. Figari & M. Achouche (dir.), *L'activité évaluative réinterrogée, regards scolaires et socioprofessionnels* (pp. 89-101). De Boeck Université.

Comment la formation des personnels peut-elle favoriser le développement d'une évaluation au service de l'apprentissage ?

SYLVIE FONTAINE

UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS (CANADA)

Introduction

Ce texte traite des situations de formation en évaluation des apprentissages qui sont proposées aux étudiants en enseignement et aux enseignants en fonction. La présentation de résultats de recherches ayant pour but de former les personnels scolaires permettra de situer l'apport de la formation dans le développement d'une évaluation au service des apprentissages des élèves.

Nous abordons premièrement la formation reçue par les étudiants, futurs enseignants, dans les programmes universitaires du Québec en ciblant principalement le développement des compétences en évaluation des apprentissages. Puis, la formation continue en évaluation des apprentissages proposée aux enseignants est illustrée à l'aide de cas issus de la recherche. Trois modalités de formation seront abordées.

I. La formation des futurs enseignants

Dans le but de cibler uniquement ce que les futurs enseignants devraient maîtriser dans le domaine de l'évaluation des apprentissages des élèves, un comité américain formé de professionnels de l'éducation a proposé une liste de sept compétences pour établir les fondements de l'expertise en évaluation (AFT, NCME, & NEA, 1990). Cette liste a été mise à jour avec l'ajout d'une huitième compétence par Brookhart (2011). Ces compétences sont présentées dans le Tableau 1.

Tableau 1 : Compétences pour le développement d'une littératie en évaluation (Traduction libre de Xu & Brown, 2016, p. 5)

	Les compétences en évaluation des apprentissages
1	Choisir la méthode d'évaluation en lien avec les décisions éducatives
2	Développer la méthode d'évaluation en lien avec les décisions éducatives
3	Administrer, noter et interpréter les résultats d'évaluation obtenus à des évaluations internes et externes
4	Utiliser les résultats de l'évaluation lors de la prise de décisions concernant les élèves, la planification de l'enseignement, le curriculum et le progrès dans l'école
5	Développer des méthodes de notation valides
6	Communiquer les résultats de l'évaluation aux diverses parties prenantes
7	Reconnaître l'intégrité, la légalité et la pertinence des méthodes et des informations entourant l'évaluation
8	Prendre en compte les développements récents concernant les politiques et les pratiques en évaluation (ex. l'évaluation formative et la reddition de compte)

Ces huit compétences s'inscrivent dans une démarche de développement de l'« *assessement literacy* », traduit en français par *littératie en évaluation* (Pasquini & Morales Villabona, 2022). La littératie en évaluation des apprentissages peut se définir de la façon suivante : une bonne compréhension de l'évaluation ; des connaissances sur les démarches, concepts, etc. en évaluation ; la capacité de mettre en œuvre ces connaissances et cette compréhension dans des tâches d'évaluation des apprentissages (Stiggins, 1991 dans Xu & Brown, 2016). Toutefois, avoir une liste de compétences prescrites ne garantit pas que les futurs enseignants développent ces compétences et les mobilisent dans leur contexte scolaire et dans les matières qu'ils enseignent (Xu & Brown, 2016). D'autant plus que la plupart des programmes de formation en enseignement, du moins au Québec, offrent, au mieux, 45 heures de cours sur l'évaluation des apprentissages. C'est donc peu de temps consacré au développement de ces huit compétences.

En fait, la recherche montre que, malgré la formation reçue, les jeunes qui débutent leur carrière en enseignement ne se sentent pas suffisamment formés pour évaluer leurs élèves (DeLuca & Bellara, 2013 ; DeLuca, 2012 ; Fontaine, Kane, Duquette, Savoie-Zajc, 2012 ; Kane, Jones, Rottmann et Conner, 2010). De plus, l'évaluation des apprentissages serait une difficulté qui persiste dans le temps et qui serait associée à un désir de quitter la profession (Fontaine *et al*., 2012). Outre le peu de temps accordé à l'évaluation des apprentissages dans les programmes de formation, trois autres raisons peuvent être évoquées pour expliquer le sentiment d'incompétence vécu par les futurs enseignants.

- La première a trait aux nombreux changements qui ont eu lieu dans le domaine de l'évaluation et, plus particulièrement, au passage d'une conception de l'évaluation **de** l'apprentissage à une conception de l'évaluation **pour** l'apprentissage (Stiggins, 2009). Il s'agit là d'un changement de paradigme que les jeunes enseignants doivent intégrer à leur pratique enseignante alors que le contexte scolaire, et parfois même universitaire, peine à s'approprier cette conception de l'évaluation.

- La deuxième raison concerne l'écart possible entre ce qui est enseigné dans les cours consacrés à l'évaluation, la documentation officielle (programmes d'enseignement, politique d'évaluation des apprentissages, etc.) et ce qui se passe réellement en salle de classe (Ogan-Bekiroglu & Suzuk,

2014 ; Laurier, 2014). Un certain malaise et de la confusion peuvent être engendrés par ce manque de cohérence.

- Enfin, la troisième raison se situe dans la confusion qui semble exister entre deux tendances actuelles en évaluation. D'une part, il y a une prise de conscience de l'importance que l'évaluation faite par les enseignants a sur l'apprentissage de ses élèves (Rey & Feyfant, 2014). Conséquemment, une place plus grande est accordée à l'évaluation intégrée à l'enseignement. Cela se traduit par la présence de l'évaluation formative, dont l'enseignant est le principal responsable, dans les documents ministériels et les politiques relatives à l'évaluation des apprentissages. D'autre part, la présence de plus en plus grande des examens externes à forts enjeux (examens du ministère, des commissions scolaires, etc.) vient brouiller, et certainement réduire, le rôle de l'enseignant au regard de l'évaluation de l'apprentissage de ses élèves (Fontaine *et al*., 2017). Il résulte de ces deux tendances que l'enseignant en salle de classe est responsable d'une partie seulement de l'évaluation de l'apprentissage de ses élèves. En effet, les résultats à ces examens sont intégrés aux notes inscrites au bulletin de l'élève. Dans certains cas, par exemple en secondaire 4 et 5[102], la pondération des examens standardisés a une valeur de 50 % de la note de l'élève. Cette dualité quant à la responsabilité de l'évaluation peut possiblement causer certaines tensions pour les enseignants.

Que faire alors pour mieux former les futurs enseignants et s'assurer qu'ils mettent en œuvre des pratiques d'évaluation qui faciliteront l'apprentissage de leurs élèves ? La recherche montre que l'évaluation formative[103] fait une différence sur l'apprentissage des élèves tout en favorisant aussi leur réussite aux examens sommatifs (Birenbaum *et al*., 2015). En fait, selon ces mêmes auteurs, ce constat expliquerait la présence de l'évaluation formative dans les documents et les politiques ministériels. Par ailleurs, le fait d'apprendre à évaluer les élèves en favorisant une démarche de soutien à leur apprentissage lors de la formation initiale des enseignants influencerait leur décision d'intégrer l'évaluation formative dans leur enseignement (Box, Skoog & Dabbs, 2015).

Le rôle premier de l'enseignant est de faire apprendre et de vérifier que cet apprentissage a bien eu lieu pour chaque élève. Pour ce, le futur enseignant doit développer sa littératie en évaluation. Doit-il développer nécessairement les huit compétences proposées dans le Tableau 1 ? Le peu de temps consacré à ces compétences dans les programmes de formation initiale, les retombées positives de l'évaluation formative sur l'apprentissage des élèves (Birenbaum *et al*., 2015) et le lien entre l'enseignement de pratiques d'évaluation formative en formation initiale et leur intégration ultérieure dans l'enseignement des jeunes enseignants (Box, Skoog & Dabbs, 2015) nous portent à croire qu'il serait plus pertinent de cibler les compétences 1, 2, 4 et 8. Les deux premières compétences concernent le lien entre la méthode d'évaluation et les décisions à prendre. À ce sujet, Black & Wiliam (2018) insistent sur la complémentarité entre l'évaluation formative et l'évaluation sommative. Ainsi, les mêmes instruments peuvent être utilisés, mais la visée de l'évaluation sera distincte. En ce sens, les résultats de l'évaluation (compétence 4) seront utilisés en cohérence avec la visée de l'évaluation.

---

[102] Au Québec, plusieurs examens ministériels nécessaires à l'obtention du diplôme d'études secondaires ont lieu les deux dernières années du secondaire.
Voir à ce sujet : http://www.education.gouv.qc.ca/eleves/bulletin-releve-diplome/releve-des-apprentissages/

[103] Le terme *évaluation formative* est souvent utilisé comme synonyme à *évaluation pour l'apprentissage* ou *évaluation en soutien à l'apprentissage* dans les écrits scientifiques (pour un développement de ces notions et de leurs nuances, voir la revue de littérature également produite dans le cadre de la conférence de consensus (Fagnant, 2023)).

Apprendre à poser les gestes évaluatifs qui s'inscrivent dans une démarche au service de l'apprentissage des élèves, à utiliser les résultats de l'évaluation pour guider l'élève et son enseignement et enfin, s'assurer d'être à jour dans son expertise (compétence 8) pourrait, à notre avis, être la cible principale de la formation initiale qui, faut-il le rappeler, est plutôt limitée en termes de durée. Si tel était le cas, les autres compétences pourraient être développées dans un deuxième cours en évaluation des apprentissages dans le programme de formation ou encore lors de la formation continue destinée aux enseignants.

II. La formation continue des enseignants

La compétence 8 du Tableau 1 va au-delà de la formation initiale des futurs enseignants. En effet, si les étudiants en formation doivent développer la capacité de mettre à jour leurs connaissances et compétences en évaluation, les enseignants en fonction sont supposés poursuivre cet effort dans le cadre d'activités de formation continue. La formation continue des enseignants en évaluation des apprentissages est offerte par divers intervenants (ministère, commission scolaire, école, consultants, etc.) et sous divers formats (ateliers, conférences, sessions intensives, etc.). Notre objectif n'est pas de faire le bilan des formations en évaluation destinées aux enseignants en exercice. Nous ciblons plutôt trois types de formation continue en évaluation des apprentissages et situons les retombées de ces formations sur la mise en œuvre de démarches d'évaluation en soutien à l'apprentissage.

A. Ateliers de formation

Le premier type de formation consiste à offrir des ateliers sur une période qui peut varier de quelques jours à quelques semaines. Dans le cadre d'une recherche, Mertler (2009) a proposé une formation intensive de deux semaines sur l'évaluation des apprentissages. Les participants à cette formation prenaient part à des discussions, à des exercices pratiques et ils devaient réaliser des tâches en lien avec différents aspects de l'évaluation. Au début et à la fin de la formation, les participants ont dû répondre à un questionnaire visant à faire un état de leurs connaissances. Ils devaient aussi tenir un journal réflexif pour documenter leur expérience. L'auteur arrive à la conclusion que la formation intensive a eu un effet important sur le développement des connaissances des participants ainsi que sur leur compréhension en évaluation des apprentissages. Cet effet est illustré par l'écart entre les résultats obtenus au questionnaire au début et à la fin de la formation. L'apport positif de la formation est aussi décrit dans les commentaires issus des journaux réflexifs des participants qui se disent maintenant prêts à intégrer ce qu'ils ont appris dans leurs classes futures. La difficulté avec ce type de formation est l'absence de suivi des participants lors de leur retour en classe. En effet, il est impossible de savoir si les enseignants vont effectivement mettre en œuvre ce qu'ils ont appris lors de la formation. C'est donc la capacité de transfert des enseignants qui est à questionner. Par ailleurs, il est évident que l'absence de suivi des participants lors du retour en classe ne permet pas de voir si cette formation en évaluation à des retombées sur l'apprentissage des élèves.

B. Formation professionnelle

Dans une recension d'écrits qui traitent de la formation continue, Timperley (2011) propose des conditions pour que les formations continues en évaluation fassent une différence sur l'apprentissage des élèves. Elle précise d'entrée de jeu qu'une « formation professionnelle efficace se caractérise entre autres par une connaissance de l'intérêt et des limites de tel ou tel test, mais aussi par une utilisation d'évaluations plus informelles. » (p. 33). Elle fait donc référence aux savoirs théoriques que les enseignants doivent maîtriser et insiste sur le fait que ces savoirs « doivent être intégrés dans une théorie de l'apprentissage cohérente ». (p. 33) Nous supposons que ces savoirs font partie de toutes les activités de formation continue et qu'ils correspondent aux compétences énumérées dans le Tableau 1. À l'instar de la recherche de Mertler (2009), les études recensées par Timperley (2011) soulignent aussi l'importance d'accompagner l'enseignant pendant ou après la formation pour que le transfert de la théorie puisse se faire dans la pratique de salle de classe.

À la lumière de cette recension, Timperley (2011) a mis au point une formation professionnelle en évaluation des apprentissages dans le domaine de la littératie (écriture et lecture) destinée aux enseignants néo-zélandais. Le but ultime de cette formation des enseignants dans le domaine de l'évaluation était d'améliorer les résultats scolaires des élèves en difficulté. La formation qu'elle propose comporte une étape d'identification des besoins d'apprentissage des enseignants participants et ce, tant au niveau didactique et pédagogique qu'en ce qui a trait à la démarche d'évaluation. Selon l'auteure, cette étape se justifie par la nécessité de comprendre en profondeur la matière enseignée avant de pouvoir l'évaluer. À l'aide d'une méthodologie de type enquête, des tuteurs accompagnent les enseignants dans une co-construction de la démarche d'évaluation qui met l'accent sur l'importance pour les élèves de bien comprendre les objectifs et la tâche d'apprentissage à réaliser. Cette pratique de validation de la compréhension des élèves est documentée afin que les enseignants puissent en constater l'efficacité. Cette formation continue a eu un impact important sur les résultats des élèves en difficulté comme le démontrent les résultats de la recherche de Timperley (2011). L'auteure souligne toutefois que les enseignants en exercice se doivent d'assumer positivement leur responsabilité de formation continue, mais également leur responsabilité de vérifier si la mise en œuvre des compétences développées lors des formations a un bénéfice pour leurs élèves. Ce deuxième type de formation, qui cible un aspect spécifique de l'évaluation, soit la compréhension commune des consignes par l'enseignant et par les élèves, s'inscrit dans un effort d'évaluation au service des apprentissages des élèves.

C. Communauté d'apprentissage professionnelle

Le troisième type de formation vise également le développement d'une démarche d'évaluation au service des apprentissages, mais aussi, la création d'un groupe de soutien pour les enseignants. Fontaine, Savoie-Zajc et Cadieux (2013) ont mis en œuvre une communauté d'apprentissage professionnelle (CAP) pour consolider les connaissances en évaluation, soutenir les enseignants dans l'ajustement de leurs pratiques et susciter une dynamique de co-formation entre enseignants. Les rencontres se sont échelonnées sur une période de deux ans. Les chercheurs se sont assurés de suivre les enseignants tout au long du processus par le biais de rencontres régulières et de journaux de bord. Les enseignants devaient mettre en œuvre une pratique évaluative auprès de leurs élèves après chacune des rencontres de groupe et documenter cette mise en œuvre dans leur journal de bord. Ils devaient de plus conserver des traces concrètes de cette mise en œuvre pour les élèves. Par exemple, la mise en œuvre d'une démarche d'évaluation formative lors d'une activité de rédaction donnait lieu

à deux traces, deux textes : le premier avec des commentaires et des questions de l'enseignant pour aider l'élève à bonifier son texte et le deuxième, le texte final et son évaluation réalisée à l'aide d'une grille d'évaluation.

Selon les enseignants participants, la CAP leur a permis de consolider leurs connaissances, d'établir plus de liens entre la théorie et la pratique, mais plus important encore, de partager leurs préoccupations professionnelles, de réfléchir sur leurs pratiques et de les partager avec leurs collègues. Contrairement à l'étude de Timperley (2011), les résultats des élèves n'ont pas été pris en compte dans cette formation. L'intérêt était plutôt de suivre leur apprentissage et leur progrès à la suite d'une démarche d'évaluation en soutien à l'apprentissage expérimentée par leur enseignant.

D. Points communs des formations professionnelles

On retient de ces divers types de formation, l'importance de soutenir les enseignants dans leur changement de pratiques, car même s'ils se disent prêts à mettre en œuvre leurs nouvelles compétences, le retour dans la salle de classe donne souvent lieu au retour des « vieilles pratiques » ! Donc, un accompagnement qui va au-delà de la formation ponctuelle est nécessaire. De plus, comme le souligne Timperley (2011), les enseignants en formation continue sont aussi responsables de vérifier si les changements qu'ils apportent à leurs pratiques évaluatives font une différence pour les élèves. C'est là l'objectif ultime de ces formations, former les participants afin qu'ils puissent utiliser l'évaluation pour soutenir l'apprentissage des élèves.

Conclusion

La formation initiale et la formation continue des enseignants en fonction font l'objet de nombreuses recherches et nous n'avons abordé qu'une infime partie de ces recherches[104]. Le développement de la littératie en évaluation des apprentissages s'avère une responsabilité importante, mais difficile à bien maîtriser. Il importe que des efforts soient consentis en ce sens afin de permettre aux étudiants, futurs enseignants, et aux enseignants, novices et expérimentés, de se sentir plus compétents et de recourir à des activités d'évaluation qui font vraiment une différence sur l'apprentissage des élèves dont ils sont responsables.

Par ailleurs, certains auteurs insistent sur l'importance pour les directions d'établissements scolaires de développer leur compétence à soutenir les enseignants afin que ceux-ci puissent recourir à des pratiques d'évaluation formative qui visent l'amélioration de l'apprentissage de leurs élèves. Par exemple, dans le cadre d'une recherche de type étude de cas, Holingworth (2012) a implanté un programme d'évaluation formative dans une école secondaire américaine. La mise en œuvre de formations, l'octroi de temps et d'un local ont permis aux enseignants et au personnel de la direction de discuter de l'enseignement, de l'apprentissage et de l'évaluation. L'auteur arrive à la conclusion que la direction doit faire preuve de leadership en évaluation auprès des enseignants pour qu'ils puissent développer leurs connaissances et leurs pratiques en évaluation formative. Or, les directions ont des lacunes en ce qui a trait à leur formation en évaluation des apprentissages (Stiggins, 2002). La recherche de Grigsby & Vesey (2011), une étude des plans de cours destinés à la formation des directions dans 30 universités américaines, montre clairement que les directions reçoivent très peu de formation sur la façon d'utiliser les résultats de l'évaluation pour améliorer l'apprentissage des

[104] Voir entre autres les travaux de Baribeau, 2020 ; DeLuca & Volante, 2016; DeLuca & Johnson, 2017; Nizet & Leroux, 2015; Nizet, 2016; Durand, 2021)

étudiants et pour mieux cerner les besoins de formation des enseignants. On peut donc penser qu'il serait difficile pour elles de soutenir les enseignants dans leurs pratiques qui impliquent la prise en compte des résultats d'évaluation pour soutenir l'apprentissage des élèves.

La responsabilité des directions quant à leur formation professionnelle se distingue de celle des étudiants ou des enseignants en ce sens qu'elles doivent non seulement assurer leur propre développement professionnel (dans ce cas, des connaissances et une compréhension de l'évaluation), mais aussi apprendre à soutenir le développement de leurs enseignants, donc faire preuve de leadership, dans le cas présent, en évaluation des apprentissages (Fontaine *et al*., 2022). C'est avec ce double objectif que les directions d'établissement scolaire pourront faire une différence intéressante et susciter la création d'un leadership et d'une culture de l'évaluation au sein de leur école. Cela implique toutefois que ces personnels soient intégrés dans des formations en évaluation des apprentissages ce qui semble plutôt rare pour le moment (Stiggins, 2002; Fontaine *et al.*, 2022).

Références

American Federation of Teachers, National Council on Measurement in Education & National Education Association (AFT, NCME, & NEA) (1990). Standards for teacher competence in educational assessment of students. Educational Measurement: Issues and Practice, 9(4), 30-32.

Baribeau, A. (2020). La professionnalité de l'agir évaluatif de l'enseignant du secondaire dans le contexte québécois. *Administration & Éducation*, 165, 233-239. https://www.cairn.info/revue-administration-et-education-2020-1-page-233.htm

Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L., Wyatt-Smith, C. & Hung, R. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education,* 13(1), 117-140.

Black, P. & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575. https://doi.org/10.1080/0969594X.2018.1441807

Box, C., Skoog, G. & Dabbs, J.M. (2015). A case study of teacher personal practice assessment theories and complexities of implementing formative assessment. *American Educational Research Journal*, 52(5), 956-983.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12.

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, 34(5-6), 576-591. DOI:10.1080/01626620.2012.730347

DeLuca, C. & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, 64(4), 356-372.

DeLuca, C., & Volante, L. (2016). Assessment for learning in teacher education programs: Navigating the juxtaposition of theory and praxis. *Journal of the International Society for Teacher Education, 20*(1), 19-31.

DeLuca, C. & Johnson, S. (2017). Developping assessment capable teachers in this age of accountability. *Assessment in Education: Principles, Policy & Practice,* 24:2, 121-126. https://doi.org/10.1080/0969594X.2017.1297010

Durand, M.-J. (2021). L'évaluation des compétences à l'école, dans la formation universitaire et en milieu professionnel. 40 ans d'expériences novatrices ici et ailleurs. Dans C. Barroso Da Costa, D. Leduc & I. Nizet (dir.), *40 ans de mesure et d'évaluation* (p.173-192). Québec : Presses de l'Université du Québec.

Fontaine, S., Kane, R. G., Duquette, O. & Savoie-Zajc, L. (2012). New teachers' career intentions: Factors influencing new teachers' decision to stay or to leave the profession. *Alberta Journal of Educational Research*, 57(4), 353-378.

Fontaine, S., Savoie-Zajc, L. & Cadieux, A. (2013). L'impact des CAP sur le développement de la compétence des enseignants en évaluation des apprentissages. *Éducation et francophonie*, 41(2), 10-34.

Fontaine, S., Kane, R. et Kovinthan, T. (2017). The assessment of students' learning: A major challenge for beginning teachers in Quebec and Ontario. Dans B. Kutsyuruba et K. D. Walker (dir.), *The bliss and blisters of early career teaching: A pan-Canadian perspective* (p. 227-246). Word & Deed Publishing Inc.

Fontaine, S., Savoie-Zajc, L., Cadieux, A. & Smith, J. (2022). Trajectoires de directions d'établissement scolaire dans le cadre d'une démarche d'accompagnement au soutien d'enseignants dans leurs pratiques évaluatives. *ERAdE,* 15(1), 7-24.

Grigsby, B. & Vesey, W. (2011). Assessment training in principal preparation programs. *Administrative Issues Journal Education Practice and Research*, 1(2), 18-31.

Hollingworth, L. (2012) Why leadership matters: empowering teachers to implement formative assessment, *Journal of Educational Administration*, Vol. 50 Issue: 3, pp. 365-379. doi.org/10.1108/09578231211223356

Kane, R. G., Jones, A., Rottmann, J. & Conner, M. (2010). *The Evaluation of the New Teacher Induction Program*. Phase Three Final Report to the Ontario Ministry of Education.

Laurier, M. (2014). La politique québécoise d'évaluation des apprentissages et les pratiques évaluatives. *Éducation et francophonie*, 42(3), 31-49.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101-113. DOI: 10.1177/1365480209105575

Nizet, I., & Leroux, J. L. (2015). La construction de savoirs conceptuels en évaluation: enjeux pour un développement professionnel en contexte de formation continue. e-JIREF, 1(2), 15-29.

Nizet, I. (2016). Le développement de compétences professionnelles par des enseignants en exercice: le cas de l'évaluation des apprentissages. Éducation et francophonie, 44(2), 109-125.

Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers, assessment literacy and its implementation into practice. *Curriculum Journal*, 25(3), 344-371. DOI:10.1080/09585176.2014.899916

Pasquini, R. & Morales Villabona, F. (2022). Comment appréhender une littératie en évaluation pour décrire et comprendre le développement des compétences des enseignants ? *Revue canadienne de l'éducation*, 45(2), 512-554.

Rey, O. & Feyfant, A. (2014). Évaluer pour (mieux) faire apprendre. *Dossier de veille de l'IFÉ*, 94, 1-44.

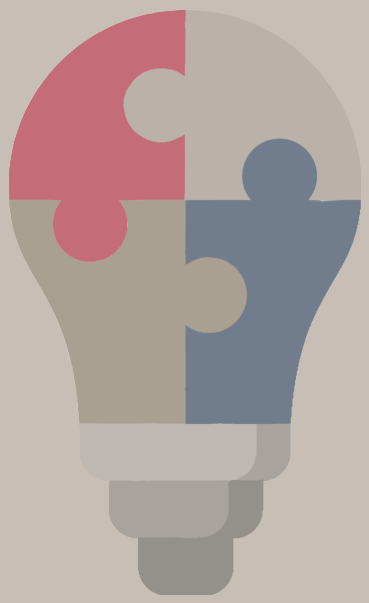
Tatto, M.T., Lerman, S. & Novotna, J. (2009). Overview of teacher education systems across the world. Dans *The professional education and development of teachers of mathematics.* Boston: Springer, 15-23. DOI: 10.1007/978-0-387-09601-8_3

Timperley, H. (2011). Le développement professionnel des enseignants et ses effets positifs sur les apprentissages des élèves. *Revue française de pédagogie* [En ligne], 174 janvier-mars 2011. DOI :10.4000/rfp.2910

Stiggins, R. J. (2009). Assessment FOR Learning in Upper Elementary Grades. *Phi Delta Kappan*, 90(6), 419-421.

Stiggins, R.J. (2002), Assessment Crisis: The Absence of Assessment for Learning, *Phi Delta Kappan* 83(10), 758–765.

Xu, Y. & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162. DOI:10.1016/j.tate.2016.05.010



le cnam
Cnesco
Centre national d'étude des systèmes scolaires

CENTRE NATIONAL D'ÉTUDE DES SYSTÈMES SCOLAIRES
CONSERVATOIRE NATIONAL DES ARTS ET MÉTIERS
41 rue Gay-Lussac - 75005 PARIS
06 98 51 82 75 - cnesco@lecnam.net
www.cnesco.fr

UNIVERSITÉ Clermont Auvergne
INSPÉ Institut national supérieur du professorat et de l'éducation Clermont Auvergne

UNIVERSITÉ CLERMONT AUVERGNE
INSPÉ DE L'ACADÉMIE DE CLERMONT-FERRAND
36 avenue Jean-Jaurès - 63407 CHAMALIÈRES
04 73 31 71 50
https://inspe.uca.fr

RETROUVEZ LES DERNIÈRES ACTUALITÉS DU CNESCO :

www.cnesco.fr Cnesco Cnesco Cnesco