

Les raisons du faible usage des résultats d'évaluation externe par les enseignants. Étude croisée dans trois contextes éducatifs

Gonzague YERLY
Université de Fribourg, Suisse

Résumé

L'évaluation des apprentissages des élèves n'est aujourd'hui plus uniquement interne à la classe (au niveau micro), elle est aussi externe et à grande échelle (au niveau macro). Avec les nouveaux modes de gouvernance des systèmes éducatifs, ces deux niveaux se rencontrent. Il est attendu des enseignants qu'ils prennent en compte les résultats d'épreuves externes afin d'améliorer leurs pratiques. Toutefois, la littérature montre un faible usage de ces données externes sur le terrain. Notre analyse du regard de 41 enseignants (entretiens semi-directifs) issus de trois juridictions distinctes quant à leur dispositif et à leur politique d'évaluation externe – au Québec, en Ontario et en Suisse – permet d'identifier les raisons de ce « découplage » spécifique. Notre analyse semi-inductive s'inspire du concept de « validité unifiée d'un test » de Messick (1989) pour comprendre le point de vue des enseignants. Notre analyse montre que le faible usage des résultats d'évaluation externe peut s'expliquer par le fait que les enseignants des trois contextes étudiés accordent peu de validité(s) à ces données externes. Les enseignants remettent en cause non seulement la mesure à large échelle, mais aussi le manque d'implication pratique des résultats d'évaluation externe et leurs conséquences sur les acteurs de l'École. Des différences sont identifiées, liées à certaines caractéristiques des dispositifs d'évaluation externe (finalités, suivi, conséquences) et au profil des enseignants (ancienneté professionnelle, confrontation directe aux tests). Finalement, alors que l'évaluation des acquis des élèves est devenue un enjeu aussi bien politique que pédagogique, notre contribution montre que, aux yeux des enseignants, les différentes approches d'évaluation (interne et externe) sont trop éloignées. D'importantes dissonances les séparent tant sur le plan de leurs méthodes, de leurs usages et de leurs finalités.

Mots-clés

Évaluation externe, résultats d'épreuves externes, enseignants du primaire, validité unifiée, découplage.

Abstract

The assessment of students learning is nowadays not only internal to the classroom (at the micro level), but also external and on a large scale (at the macro level). With the new modes of governance of the educational systems, these two levels meet. Teachers are expected to take into account the results of external tests in order to improve their teaching practices. However, the literature shows little use of these external data in the classrooms. Our analysis of the perceptions of 41 teachers (semi-directive interviews) from three different contexts as to their tools and policy of external evaluation – in Quebec, Ontario and Switzerland – helps to identify the reasons for this specific "decoupling". Our semi-inductive analysis uses the concept of "unified validity of a test" by Messick (1989) to understand the point of view of teachers. Our analysis shows that the low use of external testing results is due to the fact that teachers, in the three contexts studied, give little validity to this external data. It is not only the large-scale measure that is questioned, but also the lack of practical implications of the results of external tests and their consequences on school stakeholders. Differences are identified according to certain characteristics of the external evaluation devices (purposes, follow-up, consequences) and to the profile of teachers (professional experience, direct confrontation with the tests). Finally, while the assessment of students' learning has become both a political and an educational issue, our contribution shows that the different approaches of assessment (internal and external) are too different for teachers. Significant dissonances separate them in terms of their methods, their uses and their aims.

Keywords

External testing, tests' results, primary teachers, unified validity, decoupling.

1. Introduction

Depuis plusieurs décennies, l'évaluation des apprentissages est l'objet de nombreuses réformes dans la plupart des systèmes scolaires. On attend des enseignants qu'ils mobilisent un ensemble grandissant de méthodes afin de faire le bilan des acquis des élèves, mais aussi de les soutenir dans leurs apprentissages. En outre, l'évaluation des apprentissages des élèves n'est aujourd'hui plus seulement interne à la classe, elle est également externe et à grande échelle (Allal, 2009 ; Mottier Lopez, 2009). Il est attendu des enseignants et/ou des établissements scolaires – par des mécanismes divers – qu'ils prennent en compte les résultats de leurs élèves à des épreuves externes (désormais EE) afin d'ajuster leurs pratiques (Laveault, 2009). Ainsi, ces nouveaux modes de régulation institutionnelle (Maroy, 2014) font se rencontrer les différents niveaux contextuels (De Ketele, 2006) autour de l'évaluation des acquis des élèves : le niveau micro (celui de l'enseignement et de la classe), le niveau macro (celui du système éducatif et des politiques), mais aussi le niveau méso (celui de l'établissement).

Les travaux de synthèse traitant de l'impact des EE sur les pratiques des enseignants (Au, 2007 ; Mons, 2009 ; Rozenwajn et Dumay, 2015 ; Yerly, 2014) partagent les mêmes conclusions : indépendamment des contextes, les enseignants confrontés à ces dispositifs ne procèdent généralement qu'à des ajustements mineurs ou symboliques de leurs pratiques et/ou mettent en œuvre des pratiques de préparation intensive aux examens en cas de conséquences importantes liées aux résultats. Surtout, si le format et le contenu des tests leur permettent un certain alignement de leurs pratiques – d'enseignement et surtout d'évaluation – sur certains contenus et/ou sur certaines méthodes attendues, les résultats ne seraient que très peu mobilisés par les enseignants (Yerly, 2017a, 2017b). La présente recherche¹ propose d'identifier, dans le discours d'enseignants, les raisons de ce « découplage » (Coburn, 2004) spécifique observé entre résultats d'EE et pratiques enseignantes. Notre étude qualitative a été menée dans trois contextes scolaires primaires francophones sélectionnés pour leur variété quant à leur politique d'EE : une commission scolaire au Québec, un conseil scolaire en Ontario et un canton en Suisse.

2. Cadrage théorique et empirique

Des sondages menés dans de nombreux contextes européens ou nord-américains ont permis d'interroger le regard des enseignants quant aux EE. La synthèse de ceux-ci montre que l'attitude des enseignants face à ces dispositifs peut être qualifiée de manière générale comme « ambivalente » (Mons, 2009). Une étude, que nous avons réalisée préalablement dans quatre contextes bien distincts quant aux dispositifs et aux politiques de « gouvernance par les résultats » (ou politiques d'*accountability*) en Suisse, en Belgique, en Ontario et au Québec, a permis d'identifier et de comprendre davantage les « ambivalences » des enseignants face à ces politiques et à leurs outils (Yerly et Maroy, soumis). Elle montre que les enseignants accordent une certaine « légitimité » (morale, cognitive et pragmatique) au principe d'une évaluation externe (par exemple, pour renforcer une certaine uniformité au sein des systèmes). Néanmoins, les praticiens sont beaucoup plus critiques – également sur ces trois niveaux de légitimité – sur la mise en place (par exemple, renforce des inégalités, temps perdu) et les effets de ces politiques sur les acteurs du système (par exemple, pression, pratiques de

¹ Cette recherche a été menée dans le cadre d'une bourse postdoctorale financée par le Fonds National suisse de la recherche scientifique (P2FRP1_161705), avec le soutien des professeurs : Dany Laveault (Université d'Ottawa) et Christian Maroy (Université de Montréal). L'investigation empirique a également bénéficié du soutien du Conseil de la Recherche en Sciences Humaines (CRSH) du Canada (projet CRSH Savoir n° 435-012-0701, sous la direction de C. Maroy).

teaching to the test). Ces constats rejoignent également ceux établis par Demailly (2003) quant à l'évaluation externe des établissements scolaire en France. Elle propose « trois cadres explicatifs » qui permettent d'expliquer les résistances des enseignants face à cette forme d'évaluation institutionnelle (Demailly, 2003) : les intérêts de l'acteur (résistance stratégique, lutte de pouvoir), les valeurs de l'acteur (résistance politique, lutte entre différentes visions de l'École, la rationalité de l'acteur (résistance cognitive, distance cognitive avec la mesure statistique).

Les différentes études présentées permettent une compréhension sociologique approfondie des réactions des enseignants face aux politiques d'EE et à leurs outils. Pour cette contribution, nos objectifs sont spécifiques au faible usage des résultats des élèves aux EE par les enseignants. En effet, nos études préalables (Yerly, 2017b ; Yerly et Maroy, soumis) montrent que la plupart des enseignants interrogés relèvent la qualité de ces examens (leur forme et leur contenu). Ils se servent des (anciennes) épreuves notamment pour identifier les contenus à enseigner et pour modifier leurs propres outils d'évaluation sommative. Par contre, les enseignants interrogés en Suisse, en Belgique, au Québec et en Ontario ne mobilisent que très peu les résultats des élèves aux EE pour ajuster leurs pratiques de leur propre initiative (en Ontario, les résultats sont par contre fortement mobilisés par les conseillers pédagogiques lors de formations ; en Belgique certains résultats sont analysés en équipe-école). L'analyse des 53 entrevues montre que la comparaison entre résultats internes (ceux de la classe) et externes (ceux des EE), est réalisée de manière très superficielle par les enseignants et ne les entraînent que peu à s'autoévaluer et à procéder à des modifications de leurs pratiques d'enseignement et/ou d'évaluation. Ils ne consacrent par exemple pas ou très peu de temps à analyser les résultats de leurs élèves et en discutent de manière très informelle et rapide avec leurs collègues (échange d'avis ou de sentiment global sur la réussite de sa classe). À la rigueur, les résultats des EE sont pour eux une sorte de « baromètre » très général pour valider leur travail, s'assurer être globalement dans la « bonne ligne » en comparaison à une norme régionale.

L'objectif de cette contribution est de comprendre davantage quelles sont les raisons du faible usage des résultats d'EE par les enseignants. C'est la question de la validité d'un test – concept que nous empruntons pour un temps à la psychométrie – qui a servi à analyser le point de vue des enseignants. Plus précisément, nous nous inspirons du concept de « validité unifiée » de Messick (1989). Selon ce concept, la validité d'un test n'est pas seulement à considérer quant à la qualité de la mesure, mais aussi quant à ses implications pratiques et à ses conséquences sociales. Ainsi, trois questions ont guidé notre analyse du regard des enseignants : les résultats d'EE sont-ils une mesure pertinente des acquis des élèves (*meaningfulness*) ? Les résultats d'EE permettent-ils des implications pratiques (*usefulness*) ? Les résultats d'EE ont-ils des conséquences sociales (*appropriateness*) ?

3. Contextes de l'étude

Notre analyse porte sur trois contextes de scolarité primaire : une commission scolaire au Québec, un conseil scolaire en Ontario et un canton en Suisse². Depuis les années 2000, l'évaluation externe des acquis des élèves tient une place centrale dans ces trois juridictions, suite à l'instauration de politiques de « gouvernance par les résultats » qui visent à améliorer les résultats du système scolaire. Ces réformes ont été amorcées et sont inspirées en grande partie par les enquêtes internationales sur les acquis des élèves (par exemple, PISA, TIMMS).

² Le contexte belge qui a été étudié dans les travaux précédents n'a pas été mobilisé dans la présente analyse.

Pour atteindre leurs buts, les administrations ont misé principalement sur l'EE. Elle doit offrir aux acteurs des différents niveaux (micro, méso, macro) davantage d'information sur la performance des élèves et ainsi orienter leurs pratiques. C'est ainsi que se sont multipliées les EE évaluant les acquisitions des élèves à divers moments de la scolarité³. Si les finalités générales et les instruments principaux (curriculum uniforme, EE) sont plutôt semblables, les politiques et les dispositifs d'EE des trois juridictions étudiées se différencient sur certains points que nous identifions dans le tableau 1.

Juridiction	Dans un canton en Suisse	Dans un conseil scolaire en Ontario	Dans une commission scolaire au Québec
Modèle (Maroy et Voisin, 2014)	Responsabilisation douce	Reddition de comptes et responsabilisation réflexives	Reddition de comptes néo-bureaucratique
Mécanisme	Responsabilisation interne de l'enseignant	Reddition de comptes interne (équipe-enseignants → direction) et externe (école → conseil scolaire → ministère → public)	Reddition de comptes interne (équipe-enseignants → direction) et externe (école → commission scolaire ministère → public)
Instruments	Curriculum uniforme EE cantonales (en 2 ^{ème} , 4 ^{ème} et en 6 ^{ème} année) ⁴	Curriculum uniforme EE provinciales (en 3 ^{ème} et en 6 ^{ème} année) Plan d'action des écoles et du conseil scolaire	Curriculum uniforme EE provincial (en 4 ^{ème} et 6 ^{ème} année) Contractualisation avec cibles de réussite pour les écoles et la CS
Contenus évalués	Maths, L1, L2, sciences naturelles et humaines (en alternance pour les EE de 2 ^{ème} et 4 ^{ème})	Maths, L1 (écriture et lecture)	L1 (en 4 ^{ème}); L1, L2 et maths (en 6 ^{ème})
Rôle des enseignants	Passation, correction (en partie pour l'EE de 6 ^{ème}) et communication des résultats aux élèves EE de 2 ^{ème} et 4 ^{ème} = faibles	Passation (mais ne peuvent pas voir les copies des élèves)	Passation, correction et communication des résultats aux élèves
Conséquences des résultats pour les élèves	(prise en compte libre par l'enseignant – passation parfois facultative) EE de 6 ^{ème} = élevées (critère pour l'orientation des élèves dans les filières du secondaire)	Les résultats des EE ne comptent pas au bulletin, mais ils sont publiés et servent aux parents dans le choix de l'école.	Modérées (20% de la note du bulletin) Les résultats des tests sont publiés et servent aux parents dans le choix de l'école.
Conséquences pour l'établissement et les enseignants	Aucune conséquence formelle. Toutefois, la réputation de l'enseignant est engagée fortement pour l'EE de 6 ^{ème} qui confirme l'évaluation interne des élèves	Suivi et support pour les écoles qui performant le moins. Classement des écoles dans les médias selon les résultats d'EE	Aucune conséquence formelle. Toutefois, les écoles sont redevables au niveau supérieur. Classement des écoles dans les médias selon les résultats d'EE
Feedback et soutien suite aux résultats des EE	Aucun – enseignant individuellement (N.B. il n'y a pas de direction dans les établissements)	Présentation des résultats à l'équipe-école (par la direction) Suivi des écoles par le CS (formation spécifique) sur la base des résultats	Présentation des résultats à l'équipe-école (par la direction) Rédaction d'un plan d'action école par les enseignants (supervisé par la direction et le CS)

Tableau 1 : Description des dispositifs d'EE des trois contextes étudiés

³ Certains dispositifs d'EE au Québec et en Suisse sont antérieurs à ces réformes. Ils servaient alors uniquement à la certification des élèves à la fin de certaines étapes du cursus.

⁴ Correspond aux 4^{ème}, 6^{ème} et 8^{ème} années selon la terminologie HarmoS.

Ces contextes représentent d'ailleurs trois des quatre catégories de politiques *d'accountability* définies par Maroy et Voisin (2014)⁵. Quelques nuances existent sur les contenus et la passation des tests externes. Les principales différences entre les contextes se situent surtout dans le mécanisme des politiques d'EE, leurs conséquences et le feedback et le soutien apportés suite aux résultats d'EE.

4. Méthodologie

Notre étude suit une approche qualitative et compréhensive (Albarello, 2003). Les données sont issues d'un projet de recherche plus large¹ portant sur l'impact des politiques de gouvernance par les résultats sur les pratiques des enseignants. Quarante et un enseignants du primaire ont répondu à un entretien semi-directif individuel de 60 à 90 minutes. Pour chacun des trois contextes, l'échantillon (cf. Tableau 2) a été formé afin de couvrir les différents degrés de scolarité et de varier le profil des enseignants surtout quant à leur ancienneté professionnelle et donc au nombre d'EE passées avec leurs classes.

Juridiction	Dans un canton en Suisse	Dans un conseil scolaire en Ontario	Dans une commission scolaire au Québec
Nombre d'enseignants	16 enseignants	14 enseignants	11 enseignants
Degrés de scolarité	1 ^{ère} à 6 ^{ème} année	1 ^{ère} à 6 ^{ème} année	3 ^{ème} à 6 ^{ème} année
Ancienneté	2 à 35 ans	10 à 27 ans	1 à 25 ans
Nombre d'établissements	12 établissements	2 établissements	2 établissements

Tableau 2 : Description de l'échantillon

Pour tous les enseignants que nous avons interrogés, les thématiques interrogées étaient les mêmes : le profil professionnel de l'enseignant, le profil de son établissement, le récit de la mise en œuvre de la politique d'EE dans son établissement, les perceptions de l'enseignant quant à la politique d'EE, les effets perçus sur ses pratiques. Les entretiens ont été intégralement retranscrits, puis traités au moyen du logiciel Nvivo10 pour mener une analyse de contenu en trois étapes : analyse de cas (chaque enseignant), analyse de chaque contexte, analyse croisée. Notre analyse est « semi inductive » (Maroy, 1995). Elle s'appuie sur l'analyse approfondie d'un certain nombre de cas pour arriver à la formulation d'explications quant à la question étudiée. Nos constats émergent des données recueillies sans hypothèses *a priori* mais avec l'aide d'un modèle d'analyse, le concept de « validité unifiée d'un test » de Messick (1989). Pour ce faire, une grille catégorielle a servi au traitement des données. Elle comprend les trois dimensions de la validité selon ce modèle : pertinence de la mesure ; implications pratiques des résultats d'EE ; conséquences sociales des résultats d'EE. Cet outil a permis d'identifier, dans les entretiens, les raisons évoquées par les enseignants quant à l'usage et leur non usage des résultats d'EE. Des synthèses d'étapes (par enseignant, par contexte, croisées) ont ensuite permis de répondre à notre questionnement spécifique.

⁵ Le quatrième type est celui de « reddition de compte dure » (par exemple, Angleterre et certains états américains) dans lequel des conséquences importantes pour les acteurs sont prévues suite aux résultats d'EE (par exemple, primes, licenciements).

Afin de compléter notre connaissance des contextes et suite à l'analyse de la documentation officielle, des entretiens ont été menés avec des responsables locaux et/ou intermédiaires (directions d'établissement, inspecteurs, conseillers pédagogiques).

5. Résultats

En s'inspirant de la notion de « validité unifiée d'un test » de Messick (1989), nous pouvons dégager trois catégories de raisons qui permettent d'expliquer – souvent simultanément – le faible usage des résultats d'EE par les enseignants : (a) le manque de pertinence de la mesure, (b) les implications pratiques limitées des résultats d'EE et (c) les conséquences indésirables des résultats d'EE. Nous concentrons la présentation des résultats sur les différences entre contextes. Certaines différences dues au profil des enseignants (ancienneté professionnelle, année scolaire) sont simplement soulignées, faute d'espace.

5.1. Le manque de pertinence de la mesure

Les enseignants québécois, ontariens et suisses que nous avons interrogés donnent généralement peu de « valeur » aux données produites par les EE. S'ils considèrent que les résultats sont un indicateur complémentaire « intéressant » (par exemple, un regard de l'extérieur, une autre façon d'évaluer), ils font généralement davantage confiance à leurs propres résultats d'évaluation et à leur connaissance des élèves. Même en cas d'écart importants entre évaluation interne et externe, la plupart des enseignants (et sans différences observée selon leur profil) disent ne pas remettre en question l'évaluation des apprentissages de leurs élèves ou ajuster leur degré d'exigence par rapport au « standard » régional. Les enseignants justifient les éventuels écarts de performances des élèves – à la hausse comme à la baisse – par divers facteurs :

- *État particulier des élèves lors de la passation.* Les enseignants interrogés justifient généralement des différences de résultats entre évaluation interne et externe par un état particulier des élèves lors de la passation d'une EE (par exemple, la fatigue, le stress). Cette situation d'examen est peu habituelle, trop longue et/ou revêt trop d'importance. Au contraire, en Ontario, les résultats n'ayant aucun impact sur le cursus des élèves, les enseignants justifient des résultats plus bas par le manque de motivation et d'implication des élèves lors des examens uniformes.
- *Degré de difficulté de l'examen.* Le degré de difficulté des examens peut être jugé inapproprié : trop élevé ou trop bas et/ou trop fluctuant d'une année à l'autre. Au Québec, les examens du ministère sont souvent considérés par les enseignants comme non pertinents car leur niveau est trop fluctuant d'une année à l'autre.
- *Trop grande standardisation des EE.* La pertinence de la mesure est souvent critiquée par les enseignants car elle est à leurs yeux « injuste ». Celle-ci ne prend pas en compte les différences entre les élèves et les milieux (par exemple, élèves allophones, écoles de milieux défavorisés) et crée ainsi des inégalités entre les élèves et les établissements.
- *Inégalité des conditions de passation et de correction.* En Ontario et au Québec, les enseignants remettent également en question la pertinence des résultats d'EE quant aux conditions de passation et de correction qui ne seraient pas identiques pour tous (par exemple, aide des enseignants, correction plus clémente). L'importance de ces EE

pour les écoles et les enseignants (par exemple, reddition de comptes aux autorités, palmarès dans les médias) entraînerait des pratiques « frauduleuses » permettant de s'extraire de ces pressions.

5.2. Implications pratiques limitées

Si les enseignants n'utilisent que peu les résultats des élèves aux EE, ce n'est pas uniquement parce qu'ils ne croient pas en la pertinence intrinsèque des données produites. C'est également qu'ils n'y voient pas ou peu d'opportunités pour développer leurs pratiques et/ou que le dispositif ne le permet pas :

- *Information trop générale.* Pour les enseignants interrogés, l'EE est une mesure trop générale et ponctuelle des acquis des élèves pour pouvoir agir. Elle n'est pas aussi approfondie que leur évaluation en continu des apprentissages de l'élève. Les résultats des classes ou des écoles sous forme de moyennes de pourcentage (Québec et Ontario) ou de note (en Suisse) sont des informations trop peu approfondies pour être prises en compte. En Ontario, par exemple, au moment de la réception des résultats par la direction de l'école certains enseignants demandent à voir les résultats individuels de leurs (anciens) élèves. Pourtant, même les résultats individuels – lorsqu'ils sont disponibles – ne représentent pour les enseignants des divers contextes qu'une performance d'un élève à un moment donné sur un ensemble très (trop) vaste de contenus. Le plus souvent, les enseignants souhaiteraient avoir accès à des résultats davantage détaillés (par exemple, par compétence ou par item).
- *Feedback insatisfaisant.* Les enseignants interrogés regrettent pour la plupart le manque de feedback qui accompagne les résultats d'EE, voire remettent en question la pertinence de celui-ci lorsqu'il existe. En Suisse et au Québec, aucun feedback – si ce n'est les résultats chiffrés – n'est proposé aux enseignants sur la base des résultats, ni individuellement ni collectivement. En Suisse, personne ne présente les résultats aux enseignants⁶. En Ontario, le suivi est plus important. Les résultats d'EE sont exploités dans des rencontres de formation (3 à 4 par année) axées sur les domaines dans lesquels les élèves de l'école (ou du conseil scolaire) ont moins bien performé. Mais même là, les pistes didactiques proposées sont souvent critiquées par les praticiens : pas assez personnalisées, peu opérationnelles, peu pertinentes. Certains enseignants ontariens considèrent même ce soutien davantage comme une sanction pour les écoles et les enseignants. Ils ont le sentiment de perdre une partie de leur autonomie.
- *Trop de différences entre les méthodes d'évaluation interne et externe.* Dans les trois contextes, les enseignants reconnaissent que les EE « couvrent » bien le programme. En outre, en Suisse et au Québec, les enseignants reconnaissent de manière générale que les examens externes sont proches des pratiques évaluatives du terrain (par exemple, questionnement, mise en forme). En revanche, en Ontario, les enseignants ne reconnaissent que peu leurs pratiques évaluatives. Les EE y sont composées en partie de questions à choix multiples, type de question que les enseignants jugent peu pertinentes pour évaluer les acquis des élèves (par exemple, aucune analyse possible, recours à la chance). Il n'est donc pas possible pour eux de comparer les résultats issus d'instruments trop différents.

⁶ Pour l'EE de 6^{ème} année, les enseignants reçoivent tout de même un document qui fait la comparaison par discipline entre la moyenne des résultats de leur classe et la moyenne des résultats de tous les élèves du canton.

- *Responsabilité des enseignants.* Les enseignants interrogés remettent en cause leur (trop grande) responsabilité dans les résultats d'EE. Pour eux, d'autres facteurs expliquent autant – si ce n'est davantage – la réussite ou l'échec des élèves : caractéristiques du milieu, niveau de la cohorte d'élèves, compétences intrinsèques des élèves, suivi des parents, apprentissages réalisés – ou non – lors des années scolaires précédentes. Ainsi, les résultats montrent avant tout la performance de l'élève et non celle de l'enseignant. L'implication pratique des résultats d'EE est ainsi très faible surtout que les examens ont lieu en fin d'année (et en fin de cycle), juste avant que les élèves ne changent de classe. Les enseignants ne pourront alors plus avoir d'influence sur les apprentissages des élèves concernés.
- *Destinataires des résultats d'EE.* Souvent, les praticiens interrogés ne se voient pas comme les destinataires des résultats d'EE. Ceux-ci ne servent donc pas à faire leur autoévaluation. Pour eux, les résultats d'EE sont destinés avant tout aux élèves et aux parents (pour se situer, pour leur orientation) ou à l'administration. En Suisse, les enseignants ignorent souvent à qui sont dédiés les résultats⁷ (à un mystérieux « ils » dans leur discours). Ils associent par exemple de manière erronée ces données comme constitutives des résultats PISA. Au Québec et en Ontario, certains enseignants estiment que les EE sont plutôt l'affaire des « gestionnaires » (directions d'écoles, conseil et commission scolaire).

Selon nos analyses, ces constats peuvent être nuancés selon l'ancienneté des enseignants. Les praticiens novices ou peu expérimentés que nous avons interrogés souhaiteraient pouvoir se servir davantage des résultats d'EE. Les plus anciens, eux, sont plus critiques et se contentent bien souvent de leurs propres pratiques qui ont « fait leurs preuves ». Aussi, les plus expérimentés jugent souvent les méthodes évaluatives des EE trop différentes de leurs propres pratiques (par exemple, évaluation des compétences vs des connaissances).

5.3. Conséquences indésirables des résultats d'EE

Le faible usage des résultats d'EE par les enseignants peut également être expliqué par les conséquences indésirables qu'entraînent les résultats d'EE – mais aussi plus largement les politiques d'EE – sur les différents acteurs de la vie éducative et sur toute l'institution École :

- *Pression sur les élèves et les enseignants.* Les élèves et les enseignants ressentent parfois de fortes pressions face aux EE et à leurs résultats. Ce stress se traduit notamment par des pratiques didactiques et pédagogiques contraires à l'idéal des enseignants, peu efficaces à long terme et peu motivantes pour les élèves (*teaching to the test*). Les enjeux liés aux résultats d'EE sont en effet souvent perçus comme importants pour les élèves et les enseignants. Dans le canton suisse, les résultats des EE de 6^{ème} année sont un des indicateurs pour l'orientation des élèves dans les filières du secondaire. Ils confirment (ou non) l'évaluation interne, celle de l'enseignant. Au Québec, les résultats comptent pour 20% dans le bulletin de l'élève. Aussi, au Québec et en Ontario, les résultats des écoles sont publiés sur les sites Internet des écoles, mais aussi dans la presse sous forme de palmarès. Dans ces deux provinces canadiennes, les écoles doivent également « rendre des comptes » au niveau supérieur (conseil/commission scolaire). Surtout au Québec, les écoles ont des objectifs chiffrés de réussite à atteindre. En Ontario, les écoles « moins performantes » font l'objet d'un

⁷ Pour les EE de 2^{ème} et 4^{ème} année.

suivi plus important, parfois considéré comme une « mise sous tutelle » par les enseignants. En Suisse, la pression ressentie par les enseignants est également engendrée par un manque d'informations transmises aux enseignants. Ils craignent ainsi d'éventuelles conséquences.

- *Image faussée de l'École.* Pour les enseignants, l'EE et ses résultats donnent une image trop partielle de l'École. Ce seul résultat quantitatif ne montre pas l'ensemble des missions des enseignants qui ne s'arrête pas à une performance chiffrée à un moment donné. Au Québec et en Ontario surtout, les résultats publiés et utilisés par les parents pour faire le choix d'une école pour leurs enfants donnent selon eux un « mauvais signal » à la société.
- *Impact sur le fonctionnement de l'École.* Pour certains enseignants, l'EE a un impact important sur le fonctionnement de l'École. Elle prend beaucoup de place dans les relations entre les différents acteurs (enseignants, administrateurs, parents). En Ontario et au Québec surtout, les EE et plus largement les politiques de reddition de comptes dans lesquelles elles s'inscrivent sont souvent perçues comme une « intrusion » de l'administration dans la sphère pédagogique autrefois réservée aux seuls enseignants. Certains enseignants interrogés se sentent ainsi atteints dans leur autonomie, étant poussés à axer leurs pratiques sur la réussite aux examens et se sentant jugés voire contrôlés par rapport aux résultats de leurs élèves.

Nos analyses, montrent que ces pressions peuvent être plus importantes pour les enseignants peu expérimentés. Ils doivent encore prouver leurs compétences et connaissent moins les plans d'études, les méthodes attendues, etc. Aussi, ces pressions sont bien plus importantes pour les enseignants qui sont confrontés directement aux EE, c'est-à-dire ceux qui enseignent dans les années concernées par la passation des tests. Ces enseignants se sentent particulièrement mis « sous le spot », contrôlés et jugés par les parents, leurs supérieurs hiérarchiques, voire même leurs collègues.

6. Discussion

Dans les trois contextes étudiés, malgré des différences importantes entre leurs dispositifs et leurs politiques d'EE (cf. Tableau 1), le faible usage des résultats d'EE par les enseignants peut s'expliquer par des raisons plus ou moins similaires. Notre analyse montre que les praticiens interrogés ne leur accordent en effet que peu de validité(s) (Messick, 1989). Premièrement, cette mesure externe n'est pour eux pas une « preuve » pertinente du degré d'acquisition des élèves. Elle est marquée de nombreux biais : état particulier de l'élève lors de la passation des EE, degré de difficulté des EE peu pertinent ou trop fluctuant, iniquités lors de la passation et de la correction des EE. Deuxièmement, c'est l'implication pratique, l'usage (potentiel) de ces données externes, qui est remis en question par les enseignants. Ceux-ci mettent donc plus largement en doute la pertinence du dispositif : information trop générale, feedback insuffisant et/ou peu opérationnel, comparaisons non pertinentes, manque de communication. Troisièmement, les enseignants rejettent les conséquences négatives des résultats d'EE sur les enseignants, les élèves mais aussi sur les établissements scolaires et le système : stress, compétition, transformation de l'École.

Ainsi, notre analyse de l'expérience des enseignants montre que la « validité » accordée – ou non – par les enseignants aux résultats d'EE va au-delà de la simple « valeur » accordée à la mesure, au chiffre. En s'inspirant du concept de « validité unifiée » de Messick (1989), notre

analyse permet de démontrer que la validité des résultats d'EE est également liée à ses implications pratiques et à ses conséquences sociales. Certes, il serait pertinent de quantifier et d'analyser les différences et les corrélations de ces trois « facettes » de la validité, mais il nous paraît impossible de les isoler. Comme l'écrit Messick (1989), elles sont « inséparables ». Il est donc nécessaire de les analyser toutes les trois pour comprendre comment les enseignants « construisent » la validité des résultats d'EE.

Si de nombreuses similitudes sont observées quant aux raisons de ce « découplage » (Coburn, 2004), certaines particularités semblent également se dessiner entre les contextes. Dans les deux juridictions canadiennes (Québec et Ontario), l'usage institutionnel des résultats d'EE est bien plus important qu'en Suisse : ils participent grandement aux mécanismes de « reddition de comptes » et à la compétition entre les écoles ou entre les juridictions à l'intérieur d'une même province (conseils/commissions scolaires). Selon nos analyses, cet important usage des résultats d'EE à l'extérieur de la salle de classe diminue encore leur validité aux yeux des enseignants aussi bien sur le plan de la mesure (la pression/compétition entraîne davantage de biais), de son implication pratique (les enseignants se considèrent encore moins comme les destinataires des résultats) que de ses conséquences (la pression/compétition entraîne davantage d'effets néfastes). Certaines raisons de ce découplage sont également liées au profil des enseignants. En effet, certaines tendances, d'ailleurs communes aux trois contextes, ont été soulignées. Elles sont liées à l'ancienneté professionnelle (quant aux implications pratiques des résultats d'EE et aux conséquences des résultats d'EE) et à la confrontation directe ou indirecte avec les EE, selon le niveau scolaire enseigné (conséquences des résultats d'EE). Ces différences seraient également à investiguer plus finement et avec un échantillon d'enseignants plus large et plus diversifié.

7. Conclusion

L'évaluation des apprentissages des élèves est devenue un outil aussi bien politique que pédagogique. Il semble cependant que les niveaux micro (la classe) et macro (le système) (De Ketele, 2006) soient très hermétiques en ce qui concerne l'évaluation des acquis des élèves. Chaque niveau protège ses pratiques, ses intérêts et ses valeurs, sans forcément reconnaître ceux de l'autre. Alors que les systèmes éducatifs ne font pas (plus) tout à fait confiance à l'évaluation réalisée par les enseignants, il semble que les enseignants de leur côté ne soient pas enclins à partager une compétence qui leur était alors réservée. Notre contribution montre que, aux yeux des enseignants, chaque approche d'évaluation (externe et interne) – propre à chaque niveau du système – est trop différente pour permettre un « dialogue ». Il y a pour eux une dissonance trop importante aussi bien sur les méthodes (évaluation standardisée à grande échelle vs évaluation en continue approfondie pour chaque élève), sur les usages (traitement statistique et comparaisons vs analyse pour chaque élève) et sur les finalités (améliorer les résultats du système vs améliorer l'apprentissage de chaque élève).

En conclusion, à l'instar de ce que propose Allal (2009), il semble que c'est au niveau méso – celui de l'établissement – que se situe le plus grand défi si l'on entend faire « dialoguer » les niveaux micro et macro au sujet de l'évaluation des acquis des élèves et dépasser les dissonances que soulèvent les enseignants interrogés dans cette contribution. C'est à ce niveau intermédiaire que les différences entre ces deux approches évaluatives (interne et externe) peuvent être exploitées. Du côté des enseignants, il paraît important que les rôles, les spécificités et l'apport de chaque approche évaluative leur soient régulièrement exposés. Ensuite, des opportunités – individuelles et collectives – devraient leur être fournies quant à l'usage des résultats d'EE comme un indicateur utile (parmi d'autres) pour développer leurs

pratiques. Encore faut-il que les enseignants soient accompagnés et aient accès – dans un idéal – à des résultats robustes, détaillés, opérationnels et aux conséquences néfastes limitées. Enfin, dans l’optique d’un « dialogue » entre les niveaux micro et macro, une question nous paraît encore importante : comment l’évaluation des acquis des élèves réalisée par les enseignants peut-elle, dans le sens inverse, contribuer au pilotage et à l’amélioration du système ?

Références bibliographiques

- Albarello, L. (2003). *Apprendre à chercher : l'acteur social et la recherche scientifique*. Bruxelles : De Boeck.
- Allal, L. (2009). Pratiques évaluatives des enseignants face aux méthodologies des recherches évaluatives portant sur le système scolaire. Dans L. Mottier Lopez et M. Crahay (dir.), *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. (p. 29-45). Bruxelles : De Boeck.
- Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Coburn, C. E. (2004). Beyond Decoupling: Rethinking the Relationship between the Institutional Environment and the Classroom. *Sociology of Education*, 77(3), 211-244.
- De Ketele, J.-M. (2006). La recherche en évaluation : propos synthétiques et prospectifs. *Mesure et évaluation en éducation*, 29(1), 99-118.
- Demilly, L. (2003). L'évaluation de l'action éducative comme apprentissage et négociation. *Revue française de pédagogie*, 142(1), 115-129.
- Laveault, D. (2009). L'évaluation en classe: des politiques aux pratiques. *Mesure et évaluation en éducation*, 32(3), 3-22.
- Maroy, C. (1995). L'analyse qualitative d'entretiens. Dans L. Albarello, F. Digneffe, J. Hiernaux, C. Maroy, D. Ruquoy et P. Saint-Georges (dir.), *Pratiques et méthodes de recherche en sciences sociales*. Paris : Armand Colin.
- Maroy, C. (2014). L'évaluation externe et les indicateurs comme instruments d'action publique : fonctions instrumentales et symboliques. Dans J. Morrissette et M.-F. Legendre (dir.), *Enseigner et évaluer. Regards sur les enjeux éthiques et sociopolitiques*. (p. 165-184). Québec : Presses universitaires de Laval.
- Maroy, C. et Voisin, A. (2014). Une typologie des politiques d'accountability en éducation : l'incidence de l'instrumentation et des théories de la régulation. *Éducation comparée - Nouvelle série*, 11, 31-57.
- Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher*, 18(2), 5-11.
- Mons, N. (2009). Effets théoriques et réels des politiques d'évaluation standardisée. *Revue française de pédagogie. Recherches en éducation*, 169, 99-140.
- Mottier Lopez, L. (2009). L'évaluation en éducation : des tensions aux controverses. Dans L. Mottier Lopez et M. Crahay (dir.), *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. (p. 7-25). Bruxelles : De Boeck.
- Rozenwajn, E., et Dumay, X. (2015). Les effets de l'évaluation externe sur les pratiques enseignantes : une revue de la littérature. *Revue française de pédagogie*, 189(4), 105-138.
- Yerly, G. (2014). *Les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants. Analyse du regard des enseignants du primaire*. Thèse de doctorat en Sciences de l'éducation. Université de Fribourg, Fribourg.

- Yerly, G. (2017a). Quel impact des politiques de responsabilisation douce sur les pratiques enseignantes ? Dans Y. Dutercq et C. Maroy (dir.), *Professionnalisme enseignant et politiques de responsabilisation* (p. 121-141). Bruxelles : De Boeck Supérieur.
- Yerly, G. (2017b). Évaluation en classe et évaluation à grande échelle. Quel est l'impact des épreuves externes sur les pratiques évaluatives des enseignants ? *Mesure et évaluation en éducation*, 40(1), 33-60.
- Yerly, G. et Maroy, C. (soumis). La gouvernance par les résultats est-elle un mode de régulation de l'école légitime aux yeux des enseignants ? Une enquête qualitative dans quatre systèmes scolaires. Article soumis à la *Revue française de pédagogie*.