

Publié avec le soutien du Fonds national suisse
de la recherche scientifique



Ce texte est sous licence Creative Commons : elle vous oblige, si vous utilisez cet écrit, à en citer l'auteur, la source et l'éditeur original, sans modifications du texte ou de l'extrait et sans utilisation commerciale. Pour consulter une copie de cette licence, visitez le site internet <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Action publique, gouvernance et recherche en éducation
© Éditions HEP-BEJUNE, 2021

LA RECHERCHE ÉVALUATIVE : UN ENJEU DE TAILLE DANS LA GOUVERNANCE EN ÉDUCATION ET EN FORMATION

**Françoise Pasche Gossin, Murielle Roth
& Werner Riesen**

INTRODUCTION

Ce chapitre propose une réflexion sur la recherche évaluative dans les modes de gouvernance en éducation et en formation. Il est structuré en fonction de ce que nous retenons de nos expériences en tant que chercheuses et chercheurs lorsque nous devons répondre à une commande issue d'une instance politique et de sa gouvernance. Plus précisément, c'est au travers de quelques exemples concrets de réalisation et de notre implication dans ce type de recherche évaluative que nous avons décidé de faire reposer notre réflexion et notre argumentation. Dès lors, nous discutons de ces recherches commanditées et montrons qu'elles impliquent une prise en charge bien particulière. Il ne s'agit pas de présenter les résultats relatifs à ces recherches, mais de discuter et mettre au jour les enjeux, les limites et les tensions de ce type de recherche effectuée sur mandat. En ce sens, il s'agit d'un travail d'objectivation mené a posteriori, à partir de nos propres travaux.

Dans ce chapitre, la première partie assume une fonction de problématisation en s'appuyant sur la littérature. Précisément, nous posons quelques définitions et mettons en évidence différentes caractéristiques sur les recherches évaluatives et sur la gouvernance. Dans un deuxième temps, nous élaborons un cadre d'analyse qui fournit la trame de la convocation de chaque illustration. Nous présentons ensuite trois exemples sur lesquels s'appuient notre réflexion et développons notre argumentation. Simultanément, nous discutons des apports, des enjeux et des limites de ce type de recherche évaluative dont la finalité est de répondre à une commande. Enfin, nous concluons en proposant quelques pistes de réflexion pouvant ouvrir sur des pratiques à initier entre recherche et gouvernance dans un double contexte d'éducation et de formation.

PROBLÉMATIQUE

Depuis quelques années, on observe un regain d'intérêt à l'égard de la recherche évaluative dans les modes de gouvernance. Cet intérêt pour ce type de recherche – qui participe et œuvre à la qualité des dispositifs d'éducation et de formation – provient, entre autres, des nouvelles pratiques de gouvernance à initier qui, selon Lussignan et Pelletier (2009), reposent « sur un réseau collaboratif d'acteurs à la recherche des meilleurs compromis, en opposition aux pratiques qui caractérisent l'organisation hiérarchique traditionnelle décrites par Weber (1921, éd. 1995) où les directives viennent d'en haut et les responsabilités d'exécution sont assumées en bas en suivant des règles élaborées, formelles et impersonnelles » (p. 13). Vue sous cet angle, la nouvelle gouvernance tend à prendre appui sur les résultats de la recherche évaluative pour réguler son action, pour interroger la pertinence et la légitimité de ses choix, pour façonner ses prises de décisions, et pour favoriser le développement cohérent des systèmes d'éducation et de formation. Toutefois, ce rapprochement entre recherche évaluative « commanditée » et gouvernance n'est pas évident en raison des intérêts multiples et parfois contradictoires des actrices et acteurs impliqués, de l'utilisation des résultats et des propositions qui y sont faites (ce qui est retenu, ce qui est écarté, ce qui est reformulé), ainsi que des effets multiples sur les actrices et acteurs et sur leurs pratiques.

La recherche évaluative dans un contexte d'éducation et de formation

Nous entendons par recherche évaluative celle qui contribue, selon une méthode scientifique, à la collecte de données, à l'analyse et à l'évaluation des programmes et des dispositifs d'éducation ou de formation à des fins de préconisation, d'amélioration continue et de décision (Bedin & Talbot, 2013; Lessard & Meirieu, 2005). D'autres dénominations peuvent être considérées, telles que le suivi scientifique de type évaluatif (Ambühl & Stadelmann, 2011), la recherche-intervention à visée évaluative (Tran Van & Martin, 2005) qui est vue en ergonomie comme une conduite de projet dont le principe est de transformer la situation de travail ou encore l'évaluation commanditée (Aussel, 2015). Cette dernière est circonscrite dans une convention qui occupe une place déterminante dans l'évaluation des dispositifs (Bedin & Talbot, 2013). La recherche évaluative fournit aux responsables des politiques éducatives un certain nombre d'indicateurs susceptibles d'encadrer et d'aider la prise de décision. Cette idée - d'évaluer le rendement pour permettre un meilleur pilotage - n'est pas nouvelle. Cependant, les travaux réalisés à ce jour - principalement sous une dimension économique - ont été menés par des économistes de l'éducation ou de la formation qui se sont chargés de réaliser des audits ou des évaluations externes dans la perspective « d'optimiser le rendement des systèmes de formation » à travers les critères d'efficacité et d'efficience (Gérard, 2001, p. 54). Selon cet auteur, limiter une évaluation à une approche économique ne permet pas de rendre compte de toutes les autres dimensions telles que la dimension sociale à travers le critère d'équité, la dimension pédagogique et la dimension d'ordre conatif. Il est donc urgent de réfléchir à une recherche qui prend en compte toutes ces dimensions (Gérard, 2001).

La gouvernance dans un contexte d'éducation et de formation

Pour décrire la gouvernance de la formation suisse des enseignants, nous faisons référence au rapport d'Ambühl et Stadelmann (2011) qui la décrit comme une « scène où s'expriment quatre grands groupes d'acteurs : les institutions de formation, au centre d'un triangle formé par la politique d'éducation et de formation, la pratique de l'enseignement et

les sciences de l'éducation » (p. 63). Plus spécifiquement, la gouvernance est « une forme de théorie évoluée du pilotage » (p. 61) qui a pour enjeu « non seulement d'exercer un pilotage en fixant des règles et des objectifs », mais de « garder un œil sur l'évolution de l'environnement et de l'ensemble du système de réglementation, de manière à veiller à son efficacité, à sa durabilité » (p. 62). Autrement dit, c'est le fait « de piloter et coordonner (ou encore diriger) en gérant les interdépendances entre les acteurs (en règle générale collectifs) » (Ambühl & Stadelmann, 2011, p. 62).

La recherche évaluative devient donc un instrument indispensable à l'existence d'un mode de gouvernance dont ses finalités sont d'orienter, de guider et de coordonner. En dernier lieu, nous y voyons une volonté d'œuvrer à une forme de pilotage des institutions publiques, plus largement inscrit dans le nouveau management public (*New Public Management*). Pour ce faire, ce pilotage mobilise des expertises scientifiques internes ou externes pour analyser et évaluer les programmes de formation afin de fournir à la gouvernance des outils, des connaissances suffisantes dans la perspective d'affiner selon Demailly (2009) « les objectifs, les discours, les rhétoriques, la mise en œuvre des politiques, la connaissance des effets de celle-ci » (p. 35).

DÉMARCHE D'ANALYSE

La réflexion se fera donc à partir de trois exemples. Le premier exemple concerne le suivi scientifique d'un dispositif de formation « par l'emploi » mis en œuvre dans le cadre de la pratique professionnelle des étudiant·e·s de dernière année de formation à la Haute école pédagogique des cantons de Berne francophone, du Jura et de Neuchâtel (HEP-BEJUNE). Cette contribution analyse la commande qui est faite, ses visées, ses retombées, la posture des chercheur·e·s, les objectifs et les critères de scientificité, ainsi que les rapports complexes existant entre recherche et gouvernance. Le second exemple, mené par l'*Institut de recherche et de documentation pédagogique* (IRDP), concerne le projet EpRoCom (épreuves romandes communes) qui a connu ces dernières années des changements importants concernant ses orientations. Les différentes phases du projet sont présentées, tout en retraçant et en problématisant le travail de recherche qui a accompagné le projet. Nous

montrons le rôle fondamental joué par la recherche, et dans quelle mesure il est possible de concilier des considérations d'ordre didactique avec un mandat politique alors que les enjeux ne sont pas les mêmes. Enfin, le troisième projet s'intéresse à l'utilisation qui est faite des travaux de recherche en éducation par les instances dirigeantes. Le nombre d'enquêtes conduites dans le monde est en constante augmentation et la question de la véritable utilité de toutes les évaluations menées est abordée et illustrée par l'enquête PISA.

C'est au travers de ces trois exemples que nous discutons des espaces de relation à privilégier entre recherche et gouvernance, de la conciliation des enjeux et des rôles de la recherche évaluative et de la gouvernance, ainsi que des effets sur les actrices et acteurs, sur les institutions et sur le pilotage des systèmes d'éducation/de formation.

EXEMPLE 1 : L'ÉVALUATION D'UN DISPOSITIF DE FORMATION DANS UNE INSTITUTION DE NIVEAU TERTIAIRE

L'objet de cette recherche évaluative est un dispositif de formation par « l'emploi » mis en place en formation primaire depuis 2014 dans le cadre de la pratique professionnelle, destiné aux étudiant·e·s de troisième et dernière année de formation à l'enseignement. La recherche évaluative que nous décrivons a pour origine une demande émergeant du politique et de sa gouvernance en formation.

Contexte, origine et objectifs

En juin 2010, le Comité stratégique de la HEP-BEJUNE a chargé un groupe de travail tripartite d'étudier la question des effectifs du personnel enseignant dans l'espace BEJUNE afin de prévenir la pénurie de ce personnel et de pouvoir, le cas échéant, y faire face. Les réflexions menées par des représentant·e·s des syndicats, des cantons et de la HEP-BEJUNE ont abouti en mars 2012 à la publication d'un rapport privilégiant un modèle de formation par l'emploi dans le cadre de la troisième année de formation de la population estudiantine. Mandaté, en 2013, par le Comité stratégique pour donner suite aux conclusions du rapport, le décanat de

la formation primaire et les responsables de la pratique professionnelle ont mis sur pied et opérationnalisé le modèle de formation retenu. Tout en garantissant la qualité des quatre axes de formation, la HEP a mis des étudiant·e·s de 3^e année à la disposition des établissements scolaires depuis la rentrée d'août 2014. Selon les initiatrices et les initiateurs de ce dispositif, celui-ci devrait permettre de répondre à la problématique de la pénurie d'une part, mais aussi, et ce n'est pas le moindre de ses avantages, d'apporter des réponses à certaines questions concernant l'insertion professionnelle, notamment en contribuant au renforcement du sentiment de compétence (être prêt à prendre la classe). L'organisation des remplacements peut également être améliorée grâce à ce dispositif. C'est bien dans une perspective de durabilité et d'implantation du dispositif que le rectorat a confié au Département de la recherche de la HEP-BEJUNE la réalisation d'un suivi scientifique de ce dispositif de « formation par l'emploi » afin de s'assurer de l'adéquation des réalisations et des objectifs déclarés. Il est également demandé, par cette étude, d'apporter des éléments pour faire évoluer le dispositif (Pasche Gossin & Wentzel, 2016).

Les espaces de relation à privilégier entre recherche évaluative et gouvernance font émerger plusieurs enjeux que nous présentons ci-après.

L'enjeu éthique

Le premier enjeu est d'ordre éthique. Dans notre cas, pour répondre à une commande, la première phase de la démarche a consisté à établir une convention fixant les modalités de l'engagement. Celle-ci vise à légitimer l'accès au dispositif à évaluer et à fixer le cadre du suivi scientifique. Ainsi, il s'agit d'un premier jalon permettant de s'entendre avec les parties prenantes et de mettre au jour les enjeux de la commande, de préciser les repères pour l'action et les modalités d'intervention susceptibles d'apporter une plus-value à ses bénéficiaires. Abstraction faite d'aspects relatifs aux modalités de fonctionnement, la commande permet également de préciser la place et la posture des chercheuses et chercheurs, de leur donner un rôle actif tout en limitant les biais inhérents à ce type de recherche. Pour répondre à une commande, les chercheuses et chercheurs ne peuvent qu'être autonomes et guidés par les seules règles de leur méthode scientifique. Cette activité scientifique ne s'accommode

pas toujours avec l'objet de la demande d'étude tel que perçu par les commanditaires. Le risque est réel de détourner le processus ou l'objet vers d'autres fins que celles initialement conçues, comme le souligne Aballéa (1995, p. 18) : « le souci d'opérationnalité immédiate des réponses apportées à des questions d'évidence mais fortement prédéterminées par la réponse attendue accentue, dans de nombreux cas, les risques de soumission et de dérives » (cité par Bedin & Talbot, 2013, p. 171).

L'enjeu temporel

Le deuxième enjeu est temporel, celui-ci correspond à la question du temps nécessaire pour effectuer une activité de recherche. Cette temporalité est à mettre en lien avec la temporalité de la commande. Dans notre cas, il s'agit d'une étude longitudinale, laquelle a permis de recueillir le maximum d'informations permettant de comprendre les situations vécues et perçues par les actrices et acteurs concernés par le dispositif. Cette temporalité longue ne correspond pas toujours à la temporalité des commanditaires qui attendent des résultats rapides pour prendre des décisions. Cet enjeu temporel peut provoquer quelques tensions, voire impliquer des situations de fortes pressions temporelles pour les chercheuses et chercheurs et pour la qualité de la recherche. D'où l'importance de prendre le temps de formaliser la commande afin d'éviter des malentendus entre les parties prenantes.

L'enjeu participatif

Dans notre cas, la recherche issue d'une commande a joué dans le débat et l'orientation de l'action. La restitution des résultats a été accueillie comme un cadre de référence critique lors de débats. Elle a permis d'aboutir à une recherche de consensus et de compromis sur le sens de l'action à mener. Le recueil d'information sur l'état de fonctionnement du dispositif et la restitution des résultats ont permis aux trois parties prenantes – les décisionnaires, les responsables du dispositif et les chercheurs – d'interagir. Autrement dit, la recherche évaluative apporte une légitimité scientifique et un « rendre compte » aux commanditaires afin qu'elles et ils puissent faire reposer leur décision sur une expertise scientifique. Nous constatons que celle-ci joue de plus en plus un rôle déterminant dans les politiques de formation et les prises de décisions. Les résultats de la recherche procurent

des données et des avis aux commanditaires, ceux-ci les perçoivent comme des apports précieux pour confirmer ou contester une décision. Les informations (résultats, interprétations, recommandations) peuvent être utilisées dans la prise de décision ; cela est en lien avec l'expression « mieux connaître pour mieux décider ». En ce sens, la recherche évaluative reste un outil privilégié pour parvenir à construire des consensus et des compromis sur le sens de l'action à mener, à condition d'enrôler les parties prenantes et de prendre en compte leurs points de vue. Les décisions ont fait, dans notre cas, l'objet de négociation en concertation. Cet enjeu est relatif au mode partenarial à insuffler dans une institution. Ceci peut se comprendre comme un moyen de contre-pouvoir dans le sens de proposer une alternative aux décisions d'une seule autorité. Nous adhérons aux propos de Bedin et Talbot (2013) lorsqu'ils relèvent qu'idéalement la démarche d'évaluation de dispositifs devrait « s'appuyer sur une intelligence collective dotée d'un pouvoir d'action au bénéfice des compétences professionnelles de ses participants, des intérêts des bénéficiaires des actions prévues et de l'institution dans laquelle l'évaluation est vouée à se réaliser » (p. 167).

L'enjeu pragmatique

Le quatrième enjeu est pragmatique. Nous avons pour mission d'évaluer le dispositif de formation par l'emploi, puis de communiquer et de restituer des résultats correspondant à l'idée d'amélioration continue. Notre recherche a permis d'identifier les forces et les faiblesses du dispositif et, dans un même élan, à aidé à surmonter les difficultés de jeunesse du dispositif au travers de recommandations et de pistes réelles d'amélioration du fonctionnement conduisant à une pérennisation du dispositif. En ce sens, la recherche à visée évaluative est un instrument d'amélioration continue qui met en avant les éléments avantageux et les insuffisances. Il s'agit de résoudre des problèmes et de trouver des solutions à des fins d'amélioration ou d'adaptation d'une pratique/d'un dispositif. Autrement dit, les résultats de recherche donnent des repères pour remédier aux faiblesses et aux insuffisances du dispositif.

Quelques éléments de synthèse

Dans notre cas, la recherche évaluative a permis d'établir du lien, d'unifier, de rapprocher la recherche et la gouvernance. En effet, d'un point

de vue structurel, ces deux entités semblent suivre leur propre logique, tant leurs démarches, leurs instruments et leurs objets sont spécifiques. Or, la recherche évaluative peut constituer un levier pour positionner la recherche au sein des formations. Elle peut créer du lien et occuper ainsi une place déterminante dans les pratiques de gouvernance.

Tout au long de cette section, nous avons montré que la recherche évaluative répond à un double impératif. La première obligation consiste à souligner les contraintes et les effets du dispositif de formation sur les actrices et les acteurs, de s'assurer de l'adéquation des réalisations et de l'atteinte des objectifs déclarés, alors que la deuxième est orientée vers l'amélioration du dispositif et des pratiques qui en résultent lorsque qu'il s'agit d'apporter des recommandations dans la perspective d'optimisation, de durabilité et d'implantation du dispositif. En réponse à la commande, l'évaluation permet de se situer à différents niveaux (institutionnel, organisationnel, pratique, etc.) et auprès de différents acteurs (étudiant·e·s, formatrices, formateurs, mentor·e·s, directrices, directeurs, etc.).

La recherche évaluative est une activité rigoureuse et scientifique qui a avantage à s'intégrer à la gouvernance. Il s'agit, sans aucun doute, d'une avancée pour l'intégration de la recherche dans les hautes écoles pédagogiques, d'une nouvelle forme de légitimation des décisions lorsqu'il s'agit de faire preuve de transparence, d'exigence en matière d'amélioration continue. Cependant, il faut considérer la question de son développement institutionnel et de ses retombées concrètes avec prudence. S'il s'avère nécessaire de la distinguer des autres pratiques d'évaluation – telles que les démarches de contrôle, de gestion ou d'audit de performance – tant ses finalités, ses instruments méthodologiques, ses méthodes d'analyse sont différentes, nous devons également prendre en compte le type d'objet soumis à son expertise, l'environnement offert, ainsi que ses retombées dans le processus décisionnel. Tout ne peut ou ne doit pas être évalué par la recherche. Plus précisément, la recherche évaluative se doit d'être prudente face aux réformes publiques, aux impératifs de rationalisation budgétaire, à l'esprit de plus en plus concurrentiel qui tend à émerger dans le paysage des hautes écoles. Le risque encouru est qu'elle s'éloigne de ses préoccupations scientifiques, d'innovation, de réflexivité et d'ouverture pour devenir un « simple » instrument de régulation, pouvant aller jusqu'à devenir un outil « destructeur » de pratiques.

EXEMPLE 2 : L'ÉVALUATION DE L'ATTEINTE DES OBJECTIFS DU PLAN D'ÉTUDES ROMAND (PER). UNE RÉFLEXION DIDACTIQUE DANS LE CADRE D'UN MANDAT POLITIQUE

Dans le cadre scolaire, l'évaluation a, de tout temps, fait partie des activités menées en classe. Les élèves réalisent des productions que l'enseignant·e va examiner afin de les évaluer. Ainsi, l'évaluation est un outil au service, à la fois du personnel enseignant pour les aider à situer leurs élèves et guider leur enseignement, et de la formation des élèves pour leur permettre de savoir où ils en sont dans leurs acquisitions ; elle a fondamentalement une fonction formative. Aujourd'hui, cette évaluation, inhérente à toute didactique, a non seulement tendance à être confondue avec la mesure mais on assiste également à une prolifération de différents types d'épreuves. La croyance est forte que tout peut être mesuré et que les résultats de ces mesures permettent d'apporter des réponses indiscutables sur les orientations à prendre en matière de politique éducative.

En Suisse, cette évolution vient, d'une part, d'une volonté d'harmoniser davantage la scolarité obligatoire, par la définition de *standards minimum de formation*, l'élaboration de plans d'études et d'instruments d'évaluation communs à chaque région linguistique, et d'autre part, d'une certaine influence du monde de l'économie qui conduit à une nécessité de rentabilité et d'efficacité du système scolaire comme dans une entreprise. Une des conséquences a été la multiplication des dispositifs d'évaluation dit *externes* destinés à mesurer l'efficacité et l'efficacité de l'école sur la base des performances des élèves, comme par exemple les tests PISA ou les tests de référence basés sur les standards de formation HarmoS. Le projet d'épreuves romandes communes – EpRoCom –, dont il va être question dans cette section, se situe pleinement dans cette mouvance (de Pietro *et al.*, 2016).

Tout en retraçant l'historique du projet EpRoCom avec ses différentes phases et ses changements importants d'orientations, nous montrerons le travail de recherche qui a accompagné le projet tout au long de la démarche et les problèmes rencontrés. Puis nous reviendrons sur les enjeux de recherche en examinant dans quelle mesure il est possible de les concilier avec ceux de la politique éducative.

De 2001 à 2007, le concordat HarmoS a été préparé au sein de la *Conférence suisse des directeurs cantonaux de l'instruction publique (CDIP)*. Dans celui-ci, il est question d'harmonisation entre les cantons concernant des éléments fondamentaux du système éducatif, tels que la durée des niveaux d'enseignement, l'âge d'entrée à l'école et les principaux objectifs de formation des élèves. Pendant que cet accord était négocié, en 2004 déjà, des groupes de spécialistes issus des cantons romands et de l'*Institut de recherche et de documentation pédagogique (IRDP)* ont mené une première réflexion sur les conditions de création d'une épreuve commune aux sept cantons romands. L'implication de l'IRDP dès les premiers travaux s'explique premièrement par son rattachement – en tant que structure scientifique – à la *Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP)*, pour laquelle elle conduit des travaux de recherche, orientés vers l'aide et la décision politique. Deuxièmement, l'Institut travaille dans la continuité puisqu'il mène, depuis de nombreuses années, des activités de recherche sur les questions d'évaluation (Weiss, 1991; Pochon, 2006; Behrens, 2010).

Toujours dans une volonté de plus d'harmonisation, en 2005, la CIIP déclarait vouloir créer un véritable *Espace romand de la formation* avec pour base un plan d'études commun à toute la Suisse romande. Dans la foulée, le canton de Berne, appuyé par les cantons de Neuchâtel, du Jura et de Fribourg, commencent d'élaborer un plan d'études cadre. Les autres cantons – Vaud, Genève et Valais – rejoignent petit à petit le projet qui devient romand en 2007 et aboutit au Plan d'études romand (PER). Celui-ci a été introduit dans les cantons dès la rentrée scolaire 2011 avec une possibilité de mise en œuvre jusqu'en 2015. En 2007, le groupe de travail qui avait été constitué pour réfléchir à la potentialité d'une épreuve au niveau romand a remis son rapport en soulignant la nécessité d'avoir des synergies intercantionales pour développer une culture commune en matière d'évaluation: élaborer une méthodologie commune, une banque d'items, etc.

Une première tentative de concrétisation

En juin 2008, la CIIP confie le projet EpRoCom à l'IRDP en lui attribuant un mandat de recherche pour trois ans, afin d'étudier la possibilité de créer une épreuve romande commune pour le français et les mathématiques basée sur le Plan d'études romand (PER). Pour mener à bien ce projet,

l'équipe en charge du mandat crée un Consortium romand, composé de délégués cantonaux chargés de l'évaluation, et des groupes de travail, constitués de spécialistes de l'évaluation – chercheur·e·s, didacticien·ne·s, formatrices, formateurs et enseignant·e·s – dans les deux disciplines concernées. Le Consortium, dont un des rôles est de définir une conception commune pour l'évaluation dans le cadre de ces épreuves, rencontre assez vite un obstacle de taille et se rend compte que les dispositifs d'évaluation ne sont pas les mêmes dans tous les cantons et qu'il ne peut donc pas décider seul du type des futures épreuves romandes. Ainsi, les travaux du Consortium sont suspendus et, en 2009, celui-ci est dissous avec un renvoi des discussions auprès des chefs de département eux-mêmes. Cette question de la finalité des épreuves, en d'autres termes à quelle fin les résultats obtenus à de telles épreuves vont servir, est une question éminemment politique avec des enjeux différents d'un canton à l'autre. La même année, le concordat HarmoS entre en vigueur et la *Convention scolaire romande (CSR)* est ratifiée. Celle-ci concrétise la volonté de créer un *Espace romand de la formation*, en mettant notamment en œuvre les tâches que le concordat HarmoS délègue à la CIIP, comme par exemple l'harmonisation des plans d'études ou l'élaboration d'épreuves communes (Concordat HarmoS – ch.4, art.8, al.2⁷⁸). Les plans d'études, les moyens d'enseignement et les instruments d'évaluation ainsi que les standards de formation sont coordonnés entre eux (CSR – ch.2, sect.2, art. 15, al.1 et 2⁷⁹). De plus :

- La CIIP organise des épreuves romandes communes à l'*Espace romand de la formation*, en vue de vérifier l'atteinte des objectifs du plan d'études.
- En fin de cycle ou à la fin du degré secondaire I, si la discipline choisie pour l'épreuve romande commune correspond à celle d'un test de référence vérifiant un standard national, le test de référence peut servir d'épreuve commune.

De 2008 à 2010, les groupes de travail disciplinaires du projet EpRoCom ont tout d'abord recensé différents types d'évaluations cantonales afin d'en proposer un panorama montrant les grandes disparités d'un

⁷⁸ <http://www.edk.ch/dyn/11737.php>

⁷⁹ <http://www.ciip.ch/documents/showFile.asp?ID=2518>

canton à l'autre, tant concernant le nombre d'évaluations passées que de la fonction de ces évaluations. Forts de ce constat, les groupes ont cherché à stabiliser les définitions des différentes démarches d'évaluation – somative, diagnostique – en clarifiant la fonction de chacune d'elle. Ensuite, ils ont échafaudé des scénarios possibles pour une évaluation adaptée au PER et ont expérimenté l'un de ceux-ci basé sur une évaluation diagnostique. Les rapports qui en ont résulté (Marc & Wirthner, 2012, 2013) soulèvent plusieurs questions non résolues qui concernent – en particulier – la place de telles épreuves dans un environnement déjà bien occupé par d'autres évaluations (PISA, HarMoS, épreuves cantonales, épreuves en classe) avec un risque de surcharge pour les élèves, les liens entre les tests de référence HarMoS et les EpRoCom, la visée de ces épreuves, la démarche et la méthodologie à adopter et les éléments du PER à retenir pour l'évaluation. Les instances politiques en ont pris acte mais aucune réponse à ces questions cruciales n'a été apportée, en raison notamment de profondes divergences quant aux attentes des différents cantons.

Le chantier se poursuit

Dès 2011, face à ces questions non résolues, ni décisions concrètes prises par les autorités, les collaborateurs de l'IRDP en charge du dossier⁸⁰ ont – tout en commençant à développer la banque d'items, comme suggéré déjà en 2007 par le premier groupe de réflexion – poursuivi certains travaux scientifiques préparatoires, en travaillant en amont de la réalisation effective des EpRoCom (Sánchez Abchi *et al.*, 2016). Ils ont proposé des pistes concrètes en relation avec certaines questions en suspens, telles que l'approfondissement de l'analyse du PER afin de clarifier davantage les notions et concepts en jeu (compétence, connaissance, références culturelles) et leur articulation (Roth & de Pietro, 2013), l'analyse didactique des objectifs d'enseignement et d'apprentissage du PER qui peuvent/doivent faire l'objet d'évaluation, etc. (Roth *et al.*, 2014). Ces nouvelles analyses ont apporté des éléments supplémentaires à la compréhension de la problématique de l'évaluation, mais ont aussi soulevé des questions sur la manière dont chaque discipline considère et articule les notions en jeu, les types de tâches qui leur sont liées, le contenu des évaluations, etc.

⁸⁰ Dès 2011, ce projet a été mené – pour la discipline du français – par Murielle Roth et Jean-François de Pietro, qui a contribué aux réflexions développées dans ce chapitre.

(de Pietro, 2015). Un certain nombre d'autres questions évoquées précédemment, et qui relèvent de décisions politiques, ont été remises à nouveau sur le devant de la scène dans un rapport (Roth & de Pietro, 2015) et avec l'espoir que des décisions soient prises.

Lors du bilan de la CSR de 2015, la question des EpRoCom a été rediscutée, étant donné que l'article de la convention, précisant que la CIIP devait organiser des épreuves romandes communes, n'avait toujours pas été mis en œuvre. Devant l'absence de consensus intercantonal pour les questions – soulevées à maintes reprises – indispensables à la concrétisation des épreuves romandes communes, les responsables politiques ont décidé de poursuivre le développement de la banque d'items, faisant déjà l'objet de réflexions scientifiques et techniques, qui contiendra des matériaux pour évaluer les apprentissages des élèves, et sera mis à disposition du personnel enseignant et des départements cantonaux.

La question de l'évaluation : difficultés scientifiques et politiques

Depuis les prémisses du projet jusqu'à aujourd'hui, plus de dix ans se sont écoulés sans qu'une épreuve ne soit élaborée, peut-être un mal pour un bien ! Le projet dépend de décisions politiques sur lesquelles la recherche n'a pas prise et qui ne relèvent pas de la science. Par exemple, si les didacticien·ne·s peuvent certes en exposer les tenants et aboutissants, ce n'est pas sur des bases scientifiques qu'on peut finalement décider de la visée d'une épreuve – avec un intérêt pour les acquisitions des élèves par rapport au PER, ou pour l'instrument PER et dans quelle mesure il permet aux élèves d'atteindre les objectifs. De même, la science permet de décrire les contenus d'enseignement et les observables, mais ce n'est pas à elle de définir ce qu'on attend comme acquisitions de la part des élèves de Suisse romande et pourquoi.

Mais ce projet soulève aussi de nombreuses questions scientifiques difficiles – dont on ne peut faire l'économie –, et qui ont trait à la validité de l'évaluation : quels sont les éléments du PER qui peuvent être observés, évalués ? Qu'évalue-t-on vraiment ? Est-ce que les questions qui sont posées à l'élève évaluent vraiment sa compréhension globale de l'histoire ou sa capacité à recopier, à repérer ?

Pour gouverner, les instances politiques ont besoin de données valides sur le plan scientifique, c'est pourquoi la recherche joue un rôle fondamental non seulement en mettant en lumière des mécanismes invisibles, en montrant et démontrant, mais aussi en critiquant ce que font les politiques éducatives. Dans le cas du projet EpRoCom, qui est le fruit d'une décision politique, une réflexion didactique est indispensable, d'une part pour que les matériaux d'évaluation proposés, basés sur le PER, qu'ils soient utilisés par les enseignant·e·s pour évaluer les acquisitions de leurs élèves ou dans le cadre d'une future épreuve romande, aient une validité de contenu, autrement dit qu'ils évaluent bien ce qu'on veut évaluer, et d'autre part, pour prévenir toute dérive techniciste et économiste de l'évaluation.

EXEMPLE 3 : DONNÉES RÉCOLTÉES ET UTILITÉ DES RECHERCHES

S'il apparaît de manière évidente que les recherches ne sont pas toujours inutiles, force est de constater qu'elles ne servent pas toujours avantageusement celles et ceux à qui elles sont prioritairement destinées et les résultats ne fournissent pas forcément des réponses satisfaisantes aux questions que se posent les mandants. Dans le monde de l'éducation, innover est devenu indispensable, c'est même un impératif de fonctionnement efficace de l'école moderne en raison de la somme des connaissances et la masse des savoirs qui augmentent constamment, alors que dans le même temps les exigences institutionnelles deviennent souvent insupportables.

La recherche pédagogique, par les interrogations qu'elle suscite en relation avec sa fonction, et l'implication des chercheur·e·s lors de modifications scolaires, devrait permettre de mettre en évidence les lacunes, voire les inepties de certaines réformes adoptées pour une soi-disant amélioration des structures scolaires. Dans notre réflexion, le terme de recherche doit être compris dans son sens le plus large, comme étant l'ensemble des activités et des dispositifs qui sont mis en place pour étudier une démarche, observer un comportement, recueillir des informations ou mettre en place des expérimentations. Dans cette conception, la préoccupation majeure des chercheur·e·s devrait être de faire le meilleur usage possible des données recueillies, en mettant leur travail en valeur par une présentation simple et surtout compréhensible des résultats mis en évidence. Si les objectifs et les attentes liés aux démarches

évaluatives envisagées sont clairement précisés, alors les recherches permettront d'accompagner les processus d'innovation de manière efficiente. Dès lors, il devient évident que les évaluations pédagogiques influenceront notablement sur les modifications touchant le monde scolaire et en ayant pour objectif prioritaire l'amélioration de l'organisation et du fonctionnement des structures éducatives.

Les enquêtes qui s'intéressent plus particulièrement aux apprentissages des élèves, dans différentes disciplines, essayent avec plus ou moins de succès de faire le lien entre le monde de la recherche pédagogique et celui de la politique (Riesen, 1994). Et, au-delà de la curiosité engendrée par ce monde si particulier et presque exotique, les discussions autour des résultats obtenus abordent fréquemment la question de leur véritable influence sur les pratiques pédagogiques quotidiennes et subsidiairement sur les changements qu'ils vont inévitablement engendrer. L'inexploitation d'une grande quantité de données récoltées génère, à juste titre, des critiques souvent judicieuses concernant le manque de valorisation de certaines grandes enquêtes décidées unilatéralement par les instances politiques, qui souvent refusent de donner les moyens nécessaires aux scientifiques pour pouvoir procéder à des analyses approfondies.

Les grandes enquêtes

L'idée d'effectuer des comparaisons entre différents systèmes éducatifs, par le biais d'études menées à grande échelle (De Landsheere, 1982), est apparue dans le monde de la recherche pédagogique il y a plus de cinquante ans. Elles s'intéressaient alors aux similitudes et aux différences que les multiples et diverses pratiques scolaires pouvaient engendrer à l'école. Ainsi, les enquêtes internationales s'intéressant aux apprentissages des élèves sont apparues il y a près de cinquante ans. Une des premières grandes enquêtes a été initiée par l'UNESCO⁸¹ en 1961 et portait sur les acquis des élèves par une évaluation du fonctionnement de l'enseignement et des structures scolaires. Ensuite, se sont mises en place d'autres grandes enquêtes internationales telles que PIRLS⁸², TIMSS⁸³ ou PISA⁸⁴.

⁸¹ United Nations Educational, Scientific and Cultural Organization.

⁸² PIRLS : Progress in International Reading Literacy Study, tous les cinq ans depuis 2001.

⁸³ TIMSS : Trends in international Mathematics and Science Study, tous les quatre ans depuis 1995.

⁸⁴ PISA : Programme international pour le suivi des acquis des élèves, tous les trois ans depuis 2000.

Depuis maintenant plus de vingt ans, le nombre de recherches menées annuellement dans le monde de l'éducation, en général, et de l'école, en particulier, a augmenté de manière spectaculaire. Les travaux menés par les chercheur·e·s sont souvent entrepris sur mandat des instances en charge de l'éducation et les motivations qui sous-tendent à la délivrance de tels mandats de recherche reposent assurément sur la volonté affirmée de pouvoir disposer d'indicateurs fiables (Hubermann & Gather Thurler, 1989) qui doivent permettre un pilotage aussi efficace que possible du système éducatif sans toutefois vouloir produire trop d'informations dérangeantes pouvant faire l'objet de critiques auxquelles il serait difficile de répondre. En élaborant des indicateurs qui mettent en évidence les réussites et les difficultés du fonctionnement des écoles et des institutions de formation, les enquêtes permettent d'effectuer des comparaisons entre les différentes politiques éducatives en ne s'intéressant pas uniquement à la mesure des performances des élèves, mais en observant également le fonctionnement des systèmes éducatifs et parfois le contenu des programmes scolaires (Riesen, 2014).

Il faut toutefois relever que, de nos jours, le monde de la pédagogie et particulièrement celui de la recherche, préfère parler de gouvernance pour qualifier ce qui est traditionnellement compris comme relevant du pilotage du système éducatif. Cette notion, qui peut être définie comme étant la mise en œuvre d'un ensemble de dispositifs qui permettent d'assurer une meilleure coordination entre les différents partenaires impliqués dans une structure, apparaît telle quelle dès les années 80. En ce qui nous concerne, nous retenons plus volontiers le concept de gouvernance académique évoqué par Bouvier (2012) et qui stipule qu'elle devrait envisager « *l'obligation de passer de l'égalité des droits à l'égalité des résultats* ». Ainsi, cette gouvernance, souvent évoquée et pointée du doigt lorsqu'il y a un embarras lié à la publication de résultats, est critiquée et clairement remise en cause lorsque les performances d'une organisation étatique, comme par exemple celle de l'école, sont clairement identifiées comme étant inefficaces. Comme il apparaît que la pertinence de certaines recherches n'est pas forcément évidente car, alors que les responsables politiques attendent des informations indispensables à une *bonne gouvernance* du système éducatif, il s'avère que les résultats ne sont souvent véritablement utiles qu'aux seul·e·s chercheur·e·s, que ce soit en raison de leurs présentations ou des objets sur lesquels les observations

portent. De plus, dans le monde de l'éducation, d'aucuns se montrent sceptiques concernant la véritable utilité pédagogique de ces évaluations.

Alors que le corps enseignant subit d'importantes perturbations dans sa planification lorsque des administrateurs de tests débarquent dans les établissements scolaires, le déficit de retour d'informations légitime la question de la véritable utilité de toutes les évaluations planifiées. Si les autorités en charge de l'éducation attendent des résultats et des conclusions définitives, la réalité des analyses ne permet de dresser qu'un portrait imprécis de l'organisation et du fonctionnement des structures éducatives, en raison des caractéristiques des indicateurs produits qui sont censés identifier les particularités des cursus scolaires tout en identifiant leurs forces et leurs faiblesses. Subséquemment, les résultats relevés mettent en évidence le fait que les écoles pourraient et devraient être plus efficaces tout en faisant preuve de plus d'équité. Les efforts consentis par les autorités politiques devraient alors trouver une certaine légitimité en prouvant que les investissements consentis, qu'ils soient matériels ou financiers, n'ont pas été vains et que la ligne de conduite politique est la bonne.

L'exemple de PISA

Pour illustrer nos propos concernant les analyses effectuées ainsi que l'utilisation lacunaire des résultats, considérons l'enquête PISA qui s'intéresse aux compétences des élèves et plus uniquement aux savoirs ou maîtrise des apprentissages. L'OCDE⁸⁵ précise les objectifs de cette dernière de la manière suivante : « *L'enquête PISA cherche non seulement à évaluer la capacité des élèves à reproduire ce qu'ils ont appris, mais aussi à déterminer dans quelle mesure les élèves sont capables de se livrer à des extrapolations à partir de ce qu'ils ont appris et d'utiliser leurs connaissances dans des situations familières ou originales et dans des contextes en rapport ou non avec l'école.* » (OCDE, 2000)

Cette enquête est particulièrement intéressante par l'émergence de trois types d'indicateurs : ceux qui concernent les performances (les

⁸⁵ Organisation de coopération et de développement économiques.

résultats obtenus), ceux qui s'intéressent aux caractéristiques de la population testée (les variables contextuelles), enfin ceux qui présentent l'évolution des résultats dans le temps grâce à la répétition périodique de l'évaluation. De plus, l'originalité de l'enquête est que, en plus d'attribuer un score aux élèves pour chaque domaine testé, les résultats sont distribués dans des niveaux de compétences préalablement définis. PISA a ainsi permis de produire une importante quantité de données concernant les conditions de scolarisation, les modalités d'apprentissage ou encore l'environnement social et familial des élèves, qui complètent utilement celles récoltées sur les capacités et les aptitudes des jeunes. Le questionnaire des élèves ainsi que celui destiné aux responsables d'établissements permettent l'identification de plusieurs facteurs qui contribuent à expliquer partiellement les résultats obtenus, de même que les variances observées. De plus, certains constats, concernant le rôle joué par le genre des élèves, leur origine, la langue parlée à la maison ou encore l'environnement socioéconomique de leurs parents, ont pu être établis (Nidegger *et al.*, 2014).

Différentes critiques sont émises au sujet de PISA ; elles portent sur l'éventuelle faiblesse de la représentativité des populations d'élèves en raison des échantillonnages réalisés, sur les biais statistiques induits par les types d'analyses effectuées, sur les cadres théoriques sous-jacents ou sur les problèmes liés aux traductions des questions dans les différentes langues. Si ces reproches peuvent parfois se justifier partiellement, le principal grief que nous retenons concerne la présentation des performances des pays sous forme de classements en fonction des scores moyens.

Il est évident que les médias, de même qu'une partie de la population, se délectent de l'établissement de ces hiérarchies graphiquement séduisantes mais qui représentent souvent un cauchemar pour les autorités politiques lorsque leurs élèves ne figurent pas dans le haut du classement. Et, de surcroît, les analyses et les interprétations publiées concernant les résultats dévoilés sont souvent lacunaires, erronées et malheureusement définitives. La taxonomie, établie selon des moyennes calculées de manière savante, est souvent interprétée comme étant le reflet de la qualité des systèmes de formation considérés et engendre parfois des réformes politiques et institutionnelles non souhaitables ou totalement inefficaces. Or, il nous apparaît qu'un autre type de classement, par exemple par niveaux de compétences, produit une hiérarchie sensiblement différente

qui permet une interprétation originale des résultats en établissant une autre distribution des pays participants. Finalement, il faut admettre que les conclusions retenues n'aident pas forcément à prendre des décisions politiques adéquates car elles ne disent pas grand-chose de la réalité scolaire vécue en classe ou des difficultés auxquelles sont confrontés les enseignant·e·s dans leurs pratiques quotidiennes. Dès lors, il apparaît que les velléités affirmées des instances dirigeantes de tenir compte des travaux menés par les chercheur·e·s en éducation ne semblent pas être toujours suivies de décisions ingénieuses.

Quelques réflexions

Il apparaît souvent que les évaluations menées à grande échelle sont surtout utiles aux chercheur·e·s en raison de la valorisation proposée et de la diffusion insatisfaisante des résultats. Les principaux intéressés, les responsables d'établissements, les membres du corps enseignant ou les élèves eux-mêmes, n'ont accès qu'à des informations partielles et doivent se contenter de résultats généraux, globaux et lacunaires qui donnent parfois l'impression que d'aucuns cherchent à dissimuler des informations importantes en ne dévoilant pas toute la vérité. Les analyses disponibles révèlent de manière flagrante une sous-utilisation évidente des données disponibles, et ceci par manque de temps ou de moyens si ce n'est par absence de volonté politique.

Aujourd'hui, la question que se posent de nombreux responsables politiques est de savoir comment augmenter la moyenne globale des scores obtenus par les élèves en se demandant s'il faut soutenir activement les élèves les moins performants ou s'il est plus judicieux de favoriser l'augmentation de la proportion d'élèves qui obtiennent des scores élevés. Les deux options sont difficilement défendables et la réelle question que doivent se poser les responsables de l'école est celle de la véritable définition d'un système éducatif performant qui se doit d'assurer, à tous les jeunes, l'acquisition d'un ensemble de connaissances, de favoriser la maîtrise de certains apprentissages et d'encourager l'utilisation de leurs compétences acquises afin de faciliter leur intégration dans le monde d'aujourd'hui.

À l'avenir, il faudra absolument tenir compte de la complexité des nombreuses structures scolaires différentes et intégrer dans les réflexions

des politiciens les exigences croissantes de la société ainsi que de l'économie. Cette dernière considère encore trop souvent que l'école ne répond pas aux attentes du monde professionnel et que les acquis des jeunes ne correspondent pas aux besoins actuels de la société. Alors que les autorités scolaires prennent connaissance des propositions et des recommandations émises par les chercheur·e·s, les enseignant·e·s doivent encore patienter dans leur classe avant de percevoir concrètement les effets d'une réforme, adoptée par les politicien·ne·s mais issue des réflexions des scientifiques. Il y a plus de septante ans, Dottrens (1946) affirmait que « dès que l'on oublie les fins supérieures de l'éducation, dès que l'on perd de vue le but, la ou le pédagogue devient alors la proie de la routine, d'une pratique qu'aucun idéal n'éclaire, d'une technique qui n'est plus qu'un dressage de l'intelligence ». Finalement, pour éviter de multiplier les enquêtes et les évaluations, il faudrait peut-être se poser la question des réelles possibilités d'introduction de réformes à visées pédagogiques avant de se lancer aveuglément dans des démarches possiblement inutiles pour le quotidien des pédagogues en classe. Des recherches et des résultats, parfois beaucoup de sueur pour finalement peu de changements, mais des démarches tout de même utiles !

EN GUISE DE CONCLUSION

La recherche évaluative est une activité institutionnelle émergente faisant partie des institutions de formation et d'éducation. À ce titre, elle a une place importante à occuper dans les débats. Pour éviter de nombreuses dérives, il est recommandé de faire preuve de prudence. Ainsi, avant tout engagement, il est bon de définir les territoires et les frontières propres aux pratiques de recherche et aux pratiques de gouvernance. Cela passe par une précision des rôles, des droits, des pouvoirs et des responsabilités des uns et des autres. Cela passe également par l'idée d'établir un contrat précis dans le sens d'une mise en évidence claire des relations entre les commanditaires et les chercheur·e·s. Les travaux de Bedin et Talbot (2010) sont très explicites à ce sujet.

L'évaluation des programmes et des dispositifs de formation n'est pas une pratique nouvelle. De nombreux travaux relèvent déjà de multiples pratiques dont les diverses méthodologies ont été mises au point

pour évaluer le rendement des formations et fournir des informations aux responsables. Toutefois, ces travaux ont été réalisés la plupart du temps par des économistes de l'éducation ou de la formation dans une visée d'optimisation du rendement. Nous constatons, au fil des années, un développement de l'évaluation des programmes et des dispositifs de formation dans le pilotage de systèmes de formation et d'éducation. Cette pratique – prenant en compte diverses dimensions qui vont au-delà d'une dimension économique – devient un instrument ou un outil de la gouvernance soit pour diagnostiquer, réguler, accompagner et améliorer l'action, soit pour contrôler ou prescrire le changement. En ce sens, elle devient un instrument de pilotage des institutions d'éducation et de formation de niveau tertiaire. L'évaluation peut être interne ou externe, elle peut se réaliser sous diverses formes, selon les buts et les enjeux visés, elle peut s'organiser par différents services (service de pilotage interne avec l'aide des départements de la recherche, service externe mandaté). Néanmoins, la recherche, en tant qu'organe institutionnel et stratégique dans le mode de gouvernance, est un maillon essentiel dans la gouvernance et occupe une place sur ce terrain lorsqu'il est question d'expertise scientifique ou de suivi scientifique. Compris au sens de Bedin et Talbot (2010) : « l'évaluation de dispositifs complexes (projet éducatif, action de formation, innovation, etc.) s'inscrit le plus souvent dans une démarche de grande amplitude temporelle, mobilise différents groupes d'acteurs et nécessite la mise en œuvre de méthodologies diversifiées à la hauteur des enjeux initialement posés dans ce type d'évaluation » (p. 2). Nous ne pouvons qu'espérer que la recherche évaluative soit appelée avec prudence à devenir un réel partenaire de taille dans les modes de gouvernance en éducation et en formation.