

## L'évaluation des innovations pédagogiques : quelles modalités de coopération entre les différents acteurs ?

Sophie Genelot and Bruno Suchaut

Volume 23, Number 1, 2000

URI: <https://id.erudit.org/iderudit/1091232ar>

DOI: <https://doi.org/10.7202/1091232ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Genelot, S. & Suchaut, B. (2000). L'évaluation des innovations pédagogiques : quelles modalités de coopération entre les différents acteurs ? *Mesure et évaluation en éducation*, 23(1), 23–42. <https://doi.org/10.7202/1091232ar>

Article abstract

Using the examples of recent research, this article presents two aspects of the assessment of innovative teaching practices. First, the article describes a certain number of methodological principles that need to be considered in using external assessment. Secondly, we present the various ways different actors can cooperate within an assessment process. The general idea is that assessment processes can become policy tools for all the actors interested by innovative practices within the educational system.

## **L'évaluation des innovations pédagogiques : quelles modalités de coopération entre les différents acteurs ?**

**Sophie Genelot**

*Université de la Martinique et IREDU-CNRS*

**Bruno Suchaut**

*Université de Bourgogne et IREDU-CNRS*

**MOTS-CLÉS :** Innovation, évaluation, coopération, acteurs

*En se basant sur des exemples de travaux récents portant sur l'évaluation de dispositifs et d'innovations pédagogiques, cet article expose, dans un premier temps, un certain nombre de principes méthodologiques à mettre en œuvre pour conduire une évaluation externe. Dans un second temps, les modalités qui favorisent la coopération des différents acteurs engagés dans le processus d'évaluation sont présentées. L'idée générale est que l'évaluation peut devenir un outil de pilotage au service de tous les acteurs intéressés par les innovations dans le système éducatif.*

**KEY WORDS :** Innovation, assessment, co-operation, actors

*Using the examples of recent research, this article presents two aspects of the assessment of innovative teaching practices. First, the article describes a certain number of methodological principles that need to be considered in using external assessment. Secondly, we present the various ways different actors can cooperate within an assessment process. The general idea is that assessment processes can become policy tools for all the actors interested by innovative practices within the educational system.*

---

Note des auteurs : Présentation au Colloque de l'ADMEE-Europe, les 11, 12 et 13 janvier 2001 à Aix-en-Provence. Adresse : IREDU/CNRS, Université de Bourgogne, BP 47870, 21078 DIJON Cedex. La correspondance par courrier électronique peut être acheminée aux adresses suivantes : [sgenelot@u-bourgogne.fr](mailto:sgenelot@u-bourgogne.fr) ou [bsuchaut@u-bourgogne.fr](mailto:bsuchaut@u-bourgogne.fr)

## Innovation et évaluation : quels intérêts ?

La question de l'évaluation des expérimentations pédagogiques se révèle particulièrement cruciale dans le contexte récent de « l'institutionnalisation » de l'innovation dans notre système éducatif. En effet, alors que dans les années 60/70 l'innovation est plutôt portée par une certaine utopie pédagogique dans une logique militante et reste relativement marginalisée, les années 80 l'ont vu progressivement être davantage valorisée par l'institution (notamment dans le cadre de dispositifs spécifiques tels que la politique des ZEP) jusqu'à ce qu'elle soit, à partir des années 90, officiellement encouragée, portée par la vague de déconcentration du système éducatif qui, face à la diversité des problèmes posés, notamment par l'hétérogénéité des publics accueillis, redonnait toute leur légitimité aux « acteurs de terrain » et à leurs capacités créatives.

À l'heure actuelle, l'innovation peut apparaître comme une condition indispensable à l'avenir de notre système scolaire ce qui fait dire à Cros (1997) qu'elle est devenue « une composante indispensable de la compétence professionnelle de tout enseignant ». La création du bureau de la valorisation des innovations pédagogiques au MEN et des services correspondants dans les académies, des dispositifs comme la « Charte pour l'école du XXI<sup>e</sup> siècle » sont autant d'exemples récents qui sont le signe que le recours à l'expérimentation est devenu un outil à part entière du pilotage de notre système éducatif.

Cette volonté institutionnelle se développe selon un schéma assez classique qui, à partir de l'encouragement à la production d'innovations diverses dans les établissements scolaires, organise leur valorisation, leur diffusion et leur intégration dans les plans de formation continue des enseignants. Cette démarche intègre l'idée sous-jacente que ce dispositif institutionnel peut permettre de transformer certaines de ces expériences innovantes en références généralisables pour le système, si ce n'est en réforme (exemple : le tutorat dans l'enseignement supérieur, l'aide individualisée en seconde sont des réformes institutionnelles récentes fortement inspirées d'expérimentations locales déjà anciennes).

Or, il manque un maillon à ce schéma de développement institutionnel de l'innovation : qui valide et à quel moment le contenu des pratiques diffusées ? Sur quels critères se base-t-on pour décider que telle innovation « marche bien avec les élèves », voire mieux que des pratiques alternatives ? Qui vérifie que

cette expérience locale peut-être transférée dans un autre contexte avec les mêmes effets sur les élèves que ceux repérés initialement? Quels repères donne-t-on aux enseignants destinataires de cette diffusion des pratiques innovantes?

Actuellement si les autres étapes de ce schéma sont institutionnellement organisées (production, valorisation, diffusion), il se trouve que la question de l'évaluation de ces pratiques innovantes est bien souvent insuffisamment traitée quand elle n'est pas simplement oubliée (même dans le cadre d'appels à projets provenant des autorités académiques ou ministérielles) ou laissée à la libre initiative des concepteurs de ces expérimentations.

Or, lorsque les membres d'une équipe éducative quelconque, qu'il s'agisse de praticiens de terrain (enseignants, conseillers pédagogiques, etc.) ou de chercheurs (dans le cadre de la didactique des disciplines, par exemple), décident d'intégrer la question de l'évaluation à leur expérimentation, ils se trouvent souvent devant un ensemble de questions difficiles à résoudre.

L'évaluation quelque peu « intuitive » produite par des acteurs plus ou moins extérieurs à l'expérimentation lors de sa mise en œuvre dans les classes (enseignants « expérimentateurs », « concepteurs ou observateurs » du dispositif pédagogique expérimental, etc.) ne peut suffire, en effet, à répondre à un certain nombre de questions qui se posent généralement à son propos.

Comment mesurer l'impact spécifique du dispositif sur les élèves et leurs apprentissages? Comment évaluer le dispositif expérimental comparativement à d'autres pratiques existantes?

Comment prendre en compte, dans cette évaluation, l'interaction possible entre le dispositif expérimental en lui-même (les activités, la démarche, le curriculum, etc.) et les caractéristiques des enseignants qui le mettent en œuvre? Certains auteurs (Demailly, 1991 ; Marsollier, 1998) ont montré que les enseignants ou les établissements qui mettent en œuvre des pratiques innovantes ou qui s'engagent dans des dispositifs d'expérimentation possèdent certaines caractéristiques particulières et on peut légitimement se poser la question de savoir si les effets éventuellement constatés chez les élèves sont dus aux « activités expérimentales » en elles-mêmes ou aux caractéristiques particulières de ces équipes pédagogiques ou aux deux à la fois. On peut également s'interroger de la même façon à propos des interactions possibles entre programme expérimental et caractéristiques des élèves : tel dispositif « marche-t-il mieux », par exemple, avec certains élèves qu'avec d'autres ?

Ces deux dernières questions nous apparaissent particulièrement importantes lorsqu'on se situe dans le cadre d'expérimentations que l'on veut diffuser largement, voire généraliser dans les pratiques enseignantes. De la même façon, la relation au temps, d'un dispositif expérimental est une question particulièrement importante à instruire dans le cadre d'innovations s'inscrivant dans des mesures de politiques éducatives. Les acquisitions scolaires sont, en effet, des processus dont le caractère cumulatif a été très souvent démontré ; il est donc légitime, lorsqu'on veut évaluer une expérimentation pédagogique, de vouloir examiner à la fois son rapport à la quantité de « temps passé à ... » et à la fois l'évolution de ses éventuels effets sur les élèves.

Cette communication se veut être une contribution à la réflexion sur cette relation entre évaluation et innovation pédagogique. Dans un premier temps, nous proposerons une approche méthodologique de cette question en nous appuyant pour cela sur l'analyse de plusieurs recherches réalisées ou en cours à l'IREDU, toutes étant des évaluations de dispositifs pédagogiques expérimentaux ou d'innovations pédagogiques<sup>1</sup>. Dans un second temps, nous aborderons la question de la coopération entre les acteurs intéressés par ces dispositifs d'innovation : nous nous appuyerons pour cela sur deux recherches particulières dans lesquelles cette question a été spécifiquement travaillée.

### **Évaluer les expérimentations pédagogiques : quelques orientations méthodologiques**

De l'ensemble des recherches précédemment citées, nous pouvons essayer de décrire un ensemble de principes méthodologiques pouvant guider la conception de dispositifs pour l'évaluation d'expérimentations ou d'innovations pédagogiques.

#### ***Préciser les différents objets d'évaluation***

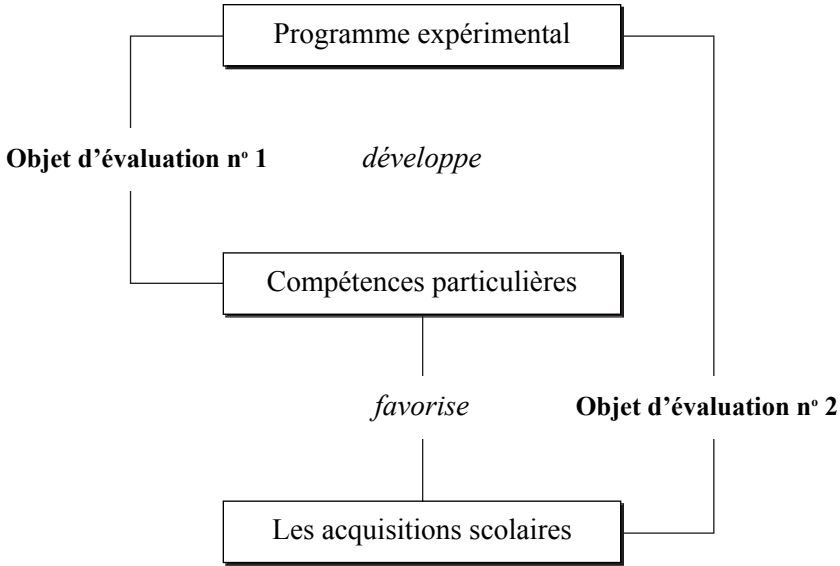
D'une façon générale, la plupart des expérimentations pédagogiques reposent sur une même hypothèse de base qui pourrait se résumer ainsi : le dispositif didactique ou pédagogique expérimental est généralement conçu pour développer des compétences particulières chez les élèves, lesquelles compétences permettant, au final, d'améliorer les acquisitions scolaires dans le domaine visé.

Ainsi, dans l'expérimentation d'activités musicales en grande section de maternelle (Mingat & Suchaut, 1994), ce qui est visé c'est le «développement de compétences cognitives transversales» qui peuvent être mobilisées par les élèves dans les apprentissages fondamentaux du cycle 2, lecture et mathématiques au CP notamment.

Dans une recherche actuelle sur l'évaluation d'activités en didactique de l'orthographe au cycle 3, l'équipe qui a conçu ces activités le fait dans le but de «développer chez les élèves des compétences métaorthographiques, lesquelles peuvent améliorer sensiblement la maîtrise orthographique des élèves aux sens scolaire et social du terme». De la même façon, le curriculum d'activités d'éveil aux langues à l'école primaire appelé «EVLANG» a été conçu par ses initiateurs comme devant, notamment, «développer des compétences métalinguistiques et métacommunicatives chez les élèves pouvant favoriser leurs apprentissages dans le domaine des langues en général, langue de l'école ou langues étrangères».

Dans une telle conception de l'expérimentation pédagogique, il est donc utile de distinguer assez rapidement deux objets possibles de l'évaluation qui peuvent être schématisés sous la forme ci-dessous. Ou bien on s'intéresse à vérifier que le dispositif expérimental développe bien les compétences «intermédiaires» de type cognitif, décrites dans les hypothèses des concepteurs comme étant des conditions de base pour améliorer les acquisitions scolaires des élèves (relation 1 du schéma), ou bien on essaie de vérifier l'impact du dispositif expérimental sur les acquisitions scolaires des élèves (relation 2 du schéma).

Dans les recherches précitées, différentes options ont été choisies de ce point de vue : dans le cas des activités musicales à la maternelle, les deux objets d'évaluation ont été successivement explorés puisque les compétences «cognitives» des élèves ont été évaluées en fin de grande section de maternelle (objet 1) ainsi que leurs acquisitions en français et en mathématiques à la fin du CP (objet 2). Dans le cas d'EVLANG, les deux objets d'évaluation sont également visés mais plus concomitamment puisque les compétences métalinguistiques des élèves sont mesurées à la fin de la mise en œuvre du programme d'activités (fin du CM2) parallèlement à leurs acquisitions en langue de l'école. Dans le cas du programme didactique en orthographe, seul l'objet 2 a été exploré : ce ne sont que les acquisitions scolaires en orthographe des élèves qui ont été évaluées.



À cet égard, il est à noter qu'il est parfois difficile d'évaluer ces compétences « intermédiaires » invoquées par les concepteurs de ces programmes expérimentaux car elles sont rarement définies avec précision. Par ailleurs, lorsqu'elles sont décrites, il est souvent assez difficile de trouver ou de concevoir des épreuves compatibles avec les conditions de passation exigées par ce type de dispositif d'évaluation. Ainsi, les compétences de type « discrimination auditive » sont-elles, par exemple, assez difficiles à mettre en place, de façon satisfaisante, dans des situations de passation collective. C'est pour cela que dans certains cas, le choix est fait de se « contenter » d'évaluer la relation de type 2 : l'efficacité du dispositif expérimental éventuellement observée étant alors attribuée, par défaut, au développement supposé de compétences transversales.

### *L'approche comparative*

La seconde hypothèse commune à la plupart des expérimentations ou innovations pédagogiques, même si elle n'est pas toujours explicitement formulée comme telle par leurs concepteurs, est que l'efficacité du programme expérimenté est supposée être supérieure aux autres méthodes ou pratiques existantes.

La seule option méthodologique qui permette d'instruire cette question est de disposer à la fois d'un échantillon de classes dites «expérimentales» dans lesquelles se déroule l'innovation pédagogique et d'un échantillon de classes «témoins». Ou bien l'expérimentation porte sur l'introduction d'une activité nouvelle, ce sera alors l'absence de cette activité qui différenciera les classes témoins des classes expérimentales (comme dans le cas de l'introduction de l'enseignement d'une langue étrangère à l'école ou dans le cas d'EVLANG, par exemple).

Ou bien l'innovation porte sur «une pratique alternative» de certaines activités traditionnelles des classes, on comparera alors cette approche particulière spécifique aux classes expérimentales à des pratiques «ordinaires» que, par ailleurs, on ne cherchera pas à décrire particulièrement mais qui seront considérées par défaut : c'est le cas, par exemple, dans l'expérimentation d'activités musicales en maternelle dont l'impact du curriculum expérimental sur les élèves est comparé à celui des activités musicales «tout venant» des classes témoins.

Dans tous les cas, cette option méthodologique va consister à comparer les performances des élèves des classes expérimentales, dans le domaine visé par l'expérimentation, à celles des élèves des classes témoins, à l'issue de la mise en œuvre du dispositif d'innovation dans les premières.

Lors de la mise en place de l'expérimentation, on devra veiller particulièrement au choix des classes expérimentales afin de ne pas introduire un biais bien connu en sciences sociales (effet Hawthorne). La solution idéale serait l'affectation au hasard des enseignants dans les groupes témoin et expérimental, ce qui n'est que rarement possible dans le domaine de l'éducation. Pour réduire ce biais, deux conditions doivent être respectées. Le groupe témoin et le groupe expérimental devront être très proches en ce qui concerne les caractéristiques des enseignants, des classes et des élèves. En second lieu, on choisira des échantillons importants du point de vue numérique, ce qui permettra de conduire des analyses statistiques solides compte tenu du grand nombre de facteurs à analyser et de la variété d'ensemble de l'échantillon.

### ***L'approche longitudinale***

Pour autant, cette approche comparative entre classes «expérimentales» et classes «témoins», si elle permet d'instruire la question de l'efficacité particulière du dispositif innovant par rapport à des pratiques alternatives, ne résout pas entièrement les questions qui se posent à cet égard. En effet, si des



différences d'acquisitions sont observées à l'issue de la mise en œuvre de l'innovation au bénéfice des élèves des classes « expérimentales », il est difficile de dissocier la part qui provient de différences déjà présentes chez les élèves avant le début de l'innovation de celles qui se sont développées au cours de la période d'expérimentation.

La dimension temporelle des acquisitions scolaires ne peut, en effet, être ignorée : dans ce cas, une approche longitudinale, dans laquelle on évalue les compétences de l'ensemble des élèves (expérimentaux et témoins) avant puis après la mise en œuvre du dispositif innovant, peut être utile pour résoudre cette question.

Dans ce cas, on ne compare plus seulement les différences d'acquisitions entre élèves expérimentaux et témoins à l'issue de la période d'expérimentation mais bien les différences de progression dans le domaine visé entre leur niveau à l'entrée et celui atteint à la fin de cette période. On dissocie alors, dans l'analyse, les acquisitions qui se sont constituées au cours de la période considérée de celles qui constituent en quelque sorte le « capital » de l'élève dont on sait, de par le caractère cumulatif des acquisitions scolaires, qu'elles conditionnent assez fortement le niveau atteint pas les élèves à l'issue d'une période donnée<sup>2</sup>.

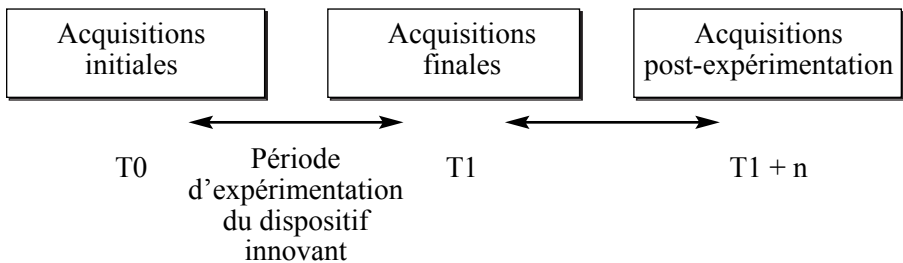
Concrètement, des épreuves sont administrées aux élèves au début et à la fin de la période évaluée sous la forme de tests standardisés. Le pouvoir discriminant de ces épreuves sera considéré comme une qualité, car dans la majorité des cas, on ne cherchera pas à établir un score absolu (en référence un une norme pédagogique ou institutionnelle), mais plutôt un score relatif qui permettra une comparaison entre les élèves de l'échantillon.

Par ailleurs, ce dispositif longitudinal de l'évaluation peut être également étendu au-delà de la période d'expérimentation elle-même. En effet, les concepteurs d'innovations font souvent référence, dans leurs hypothèses fondatrices, à la possibilité que la démarche conduite puisse avoir des effets différés dans le temps, sur les apprentissages ultérieurs des élèves.

Dans les exemples de recherche déjà cités, on retrouve cette position très clairement formulée dans l'expérimentation d'activités musicales en maternelle dont les concepteurs attendent qu'elles aient des effets positifs sur les acquisitions des élèves au cours du CP et même au-delà. De la même façon, l'introduction de l'enseignement d'une langue étrangère à l'école primaire vise explicitement (dans les textes officiels qui l'organisent) l'amélioration des performances des élèves dans cette langue au collège, ainsi donc dans le

dispositif d'évaluation de cette expérimentation, les acquisitions des élèves en anglais ont été successivement mesurées en fin de sixième et en fin de cinquième pour l'ensemble de l'échantillon, c'est-à-dire à la fois pour les élèves témoins ayant débuté l'apprentissage de cette langue au collège et pour les élèves expérimentaux ayant bénéficié d'une initiation dans cette langue à l'école primaire. Enfin, une évaluation en cours cherchera à mesurer les effets d'un dispositif d'accompagnement scolaire («Coups de pouce») pendant l'année de CP, mais aussi de façon différée en fin de CE1.

On obtient alors un dispositif d'évaluation dont l'organisation temporelle peut être schématisée sous la forme suivante :



dans laquelle T0 constitue le moment précédant la mise en œuvre du dispositif innovant, T1 le moment de la fin de l'expérimentation et T1 + n, un moment dont il convient de définir le terme en fonction des hypothèses des concepteurs de l'expérimentation (sachant qu'on peut multiplier les mesures «post-expérimentation»).

### ***La modélisation***

Cependant à ce stade de conception du dispositif d'évaluation, il reste encore à pouvoir isoler l'impact spécifique du dispositif innovant sur la progression des élèves expérimentaux.

En effet, dans l'éventualité de différences observées entre la progression des élèves ayant bénéficié de l'innovation et celle des élèves témoins (au cours de la période expérimentale ou après celle-ci), on peut s'interroger sur l'origine de ces différences : elles peuvent certes provenir de la mise en œuvre du dispositif innovant, mais elles peuvent être également être le résultat des différents contextes scolaires en présence (niveau scolaire des classes visées, compétences particulières de certains maîtres, etc.) ou encore de certaines caractéristiques des élèves (leur âge, leur sexe, leur milieu social, etc.). Autant de facteurs dont on sait, par ailleurs (les références de la documentation

scientifique en la matière sont nombreuses et convergentes), qu'ils exercent également un impact significatif sur la progression des élèves dans différents domaines des acquisitions scolaires.

Ce qui fait que l'on peut avancer, sans grand risque de se tromper, que les différences de progression observées, le cas échéant, entre élèves expérimentaux et témoins sont la résultante des effets respectifs d'un grand nombre de facteurs, sans pouvoir les réduire à la seule conséquence de la mise en œuvre de l'innovation pédagogique évaluée.

Comme nous l'avons évoqué auparavant, devant l'impossibilité qui nous est faite, dans le cadre d'expérimentations pédagogiques «en situation» (à la différence de dispositifs expérimentaux en «laboratoire»), de disposer de deux échantillons d'élèves et de classes expérimentaux et témoins qui ne diffèrent que par la seule présence ou absence du dispositif innovant, il nous faut, pour conclure sur l'impact spécifique de l'innovation considérée, trouver un cadre d'analyse qui nous permette de dissocier les effets respectifs de ces différents facteurs sur la progression des élèves.

La modélisation multivariée permet cette approche, son principe repose sur l'explication de la variété des acquisitions des élèves à T1 (puis à T + n si nécessaire) sous la forme d'une relation mathématique entre le niveau d'acquisition T1 et l'ensemble des facteurs que l'on peut et que l'on choisit «d'objectiver» (en recueillant un certain nombre de données auprès des élèves, des classes et des enseignants).

Le fait d'avoir ou non bénéficié de l'innovation est alors considéré comme une caractéristique particulière de chaque élève de l'échantillon dont on mesure l'impact spécifique sur son niveau d'acquisition à T1, indépendamment de l'influence des autres facteurs dont on peut également isoler l'impact respectif (clause du «toutes choses égales par ailleurs»). On peut présenter ce type d'analyse sous la forme de l'équation «type» suivante :

$$AF = f(AI, CE, CS, \text{Expé})$$

dans laquelle AF représente le niveau d'acquisition «finale» (aux différents moments T1, T + n considérés), AI le niveau d'acquisition «initiale» à T0, CE un ensemble de caractéristiques des élèves et CS un ensemble de caractéristiques des contextes scolaires en présence et Expé, la participation ou non au dispositif innovant.

Ce type de modélisation des acquisitions scolaires des élèves dans le domaine visé a l'avantage, par ailleurs, de pouvoir instruire des questions relatives à des hypothèses sous-jacentes ou explicitement formulées dans bon nombre d'expérimentations pédagogiques.

Il en va ainsi de l'objectif d'équité revendiqué par de nombreuses innovations : non seulement les concepteurs de tels programmes expérimentaux recherchent une efficacité particulière de leur démarche par rapport aux pratiques alternatives existantes mais ils en espèrent également souvent un bénéfice spécifique pour les élèves en difficulté.

À partir de la relation «type» décrite plus haut on peut alors instruire cette question de l'équité de l'innovation en modifiant la spécification du modèle sous la forme suivante :

$$AF = f(AI, CE, CS, \text{Expé}+, \text{Expé}-)$$

dans laquelle la situation des élèves expérimentaux de niveau initial faible (Expé-) est distinguée de celle des élèves expérimentaux de niveau initial fort (Expé+).

Dans les exemples de recherche cités plus haut, on a pu montrer, par exemple, qu'au-delà de l'impact positif des activités musicales conduites en grande section de maternelle sur les résultats au CP (en mathématiques et en français) des élèves qui en avait bénéficié, on observait un avantage plus grand pour les élèves les plus faibles en début de grande section. Grâce au même type d'analyse, on constate, dans cette même recherche, que l'expérimentation en activités musicales profite également davantage aux élèves qui sont entrés tardivement à l'école maternelle (à 3 ou 4 ans comparés à ceux qui ont été scolarisés à 2 ans). De même, lors de l'évaluation d'une démarche pédagogique innovante en lecture au CP, on a pu mettre en évidence que l'objectif premier était atteint, à savoir que ce sont les élèves qui étaient les plus «fragiles» dans leurs acquisitions en lecture au début du CP qui ont bénéficié le plus de l'innovation.

L'ensemble de ces résultats permet donc de conclure non seulement à l'efficacité de l'innovation considérée, mais également à relever ses dimensions égalisatrice et compensatrice par rapport à différentes inégalités présentes entre les élèves avant la mise en œuvre du dispositif expérimental.

Dans le cas de l'évaluation de l'enseignement de l'anglais à l'école primaire, à l'inverse, si l'on constate un impact positif de cette initiation, bien que faible, sur les acquisitions des élèves en fin de la classe de sixième,

l'analyse complémentaire en fonction du niveau scolaire initial des élèves fait apparaître son caractère fortement discriminant. L'initiation à l'anglais au primaire ne profite qu'exclusivement aux élèves scolairement les plus forts: le niveau en anglais des élèves faibles n'est en effet pas significativement différent, en fin de sixième, de celui des élèves témoins.

Par ailleurs, l'analyse longitudinale, qui permet d'évaluer l'impact de l'initiation conduite au primaire sur les acquis en fin de la classe de cinquième, fait apparaître qu'à ce stade, même son impact moyen sur le niveau atteint par les élèves en anglais a disparu. L'ensemble de ces résultats conduit donc à relever un intérêt assez limité de cette expérimentation par rapport aux avantages initialement escomptés par ses concepteurs.

Ce type d'analyse permet également, au-delà de l'impact moyen de l'innovation considérée, d'en évaluer l'efficacité éventuellement différenciée des différentes conditions de sa mise en œuvre, la plus aisément appréhendable d'entre elles étant le temps «d'exposition» au programme expérimental. Dans le cas des activités musicales en grande section de maternelle, les concepteurs de cette innovation ont expérimenté le curriculum sous deux formes différentes (2 ou 4 heures hebdomadaires) de sorte que le dispositif d'évaluation a pu prendre en compte cette distinction, l'analyse ayant été conduite en deux étapes :

$$1) AF = f(AI, CE, CS, \text{Expé})$$

$$2) AF = f(AI, CE, CS, \text{Expé}_2, \text{Expé}_4)$$

L'impact de l'innovation a d'abord été appréhendée d'un point de vue global (équation 1), puis les deux organisations horaires ont été ensuite distinguées (équation 2). Il ressort ainsi que deux heures hebdomadaires d'activités musicales en grande section sont suffisantes pour obtenir un impact significatif sur les acquisitions des élèves en fin de maternelle et en fin de CP et que l'on ne gagne rien de très significatif à augmenter le volume horaire au-delà de cette durée. Ce résultat est intéressant pour donner des indications concrètes quand à sa mise en œuvre dans le cadre de la diffusion de cette expérimentation.

Ceci dit, indépendamment de la durée de l'expérimentation, d'autres conditions de mises en œuvre des innovations considérées peuvent être analysées quand à leurs effets différenciés sur les progressions des élèves.

Ainsi, dans les hypothèses qui sous-tendent la conception du programme expérimental d'éveil aux langues à l'école primaire (EVLANG), ses auteurs déclarent que «l'efficacité de ce programme sera d'autant plus importante que les activités seront mises en œuvre dans le cadre d'une démarche de pédagogie «active». Pour tenter d'évaluer cette hypothèse, un dispositif d'observation des pratiques pédagogiques des enseignants a été mis en place sur une partie de l'échantillon expérimental : plusieurs outils d'observation ont été construits (interactions maître/élèves, indicateurs du degré «d'activité» des élèves, interactions élèves/élèves dans de petits groupes de travail) devant conduire à établir une typologie des pratiques de mises en œuvre d'EVLANG. Sur cet échantillon restreint de classes, cette typologie sera alors introduite dans les modèles d'analyse comme une caractéristique individuelle des élèves («a un enseignant qui pratique un enseignement de type X, Y, etc.»):

$$AF = f(AI, CE, CS, Evlang X, Evlang Y)$$

Il va de soi que ce type d'analyse alourdit considérablement le dispositif d'évaluation dans la mesure où il nécessite le recours à une observation directe des classes, à un premier traitement des données issues de ces observations pour constituer une typologie, avant que d'intégrer le modèle général d'analyse. Il ne peut être conduit que sur une partie seulement de l'échantillon expérimental compte tenu des contraintes (notamment temporelles et matérielles) imposées par l'observation directe tout en restant sur un échantillon assez grand de façon à pouvoir intégrer la modélisation.

À ce titre, ce modèle d'analyse constitue un essai très intéressant d'articulation concrète d'une démarche quantitative de l'évaluation d'un dispositif d'innovation visant à en mesurer son efficacité externe avec une démarche plus intégrée au processus même de l'innovation, davantage qualitative et visant une description fine de la réalité de la mise en œuvre de l'expérimentation.

Les quelques principes méthodologiques qui viennent d'être énoncés, loin de n'être qu'un cadre conceptuel formel, constituent, à notre sens, une base d'action réelle pour concevoir des dispositifs d'évaluation pouvant efficacement accompagner le développement de pratiques innovantes ou expérimentales. Leur formulation et leur articulation sont issues de l'analyse d'un assez grand nombre de recherches conduites ces dernières années, sur des objets assez variés tant sur le plan des domaines d'acquisitions considérés (lecture, français, mathématiques, langue étrangère, etc.) que sur celui des niveaux scolaires visés (de l'école maternelle au collège, notamment).

Si la mise en œuvre de ces principes méthodologiques requiert des compétences techniques relativement spécialisées (la modélisation mobilise, il est vrai, des techniques statistiques spécifiques) et des aspects matériels assez contraignants (échantillon d'élèves et de classes de taille suffisamment importante, dimension temporelle qui «diffère» la production des résultats souvent bien au-delà de la période même d'expérimentation du dispositif innovant, etc.), nous avons pu constater que ces contraintes ne constituaient pas des obstacles insurmontables à l'intégration d'un tel dispositif d'évaluation dans des dynamiques de pratiques innovantes. Il suffit que les acteurs intéressés par cette démarche y trouvent leur intérêt propre et sachent, par ailleurs, y définir leurs rôles respectifs. C'est cette dimension que nous allons maintenant aborder en nous appuyant notamment sur deux recherches dans lesquelles cette coopération entre les différents acteurs a été particulièrement structurée.

### **Des modalités qui favorisent la coopération entre les acteurs**

Les propos qui suivent sont principalement le résultat de l'analyse de la mise en œuvre de deux recherches citées plus haut – celle concernant les activités musicales en maternelle et celle concernant les activités d'éveil aux langues à l'école primaire (EVLANG) – qui ont donné lieu à une coopération étroite entre les acteurs de l'innovation, concepteurs de ces programmes expérimentaux (conseillers pédagogiques dans un cas, didacticiens dans l'autre) et les acteurs de l'évaluation d'autre part.

L'objet de notre réflexion a consisté à tirer de ces deux expériences, de leurs points communs comme de leurs différences, les conditions qui, de notre point de vue, pouvaient permettre la coopération la plus étroite et la plus efficace entre les différents acteurs concernés.

#### ***Distinguer les fonctions de conception de l'innovation de celle de conception de l'évaluation***

Une des caractéristiques communes à ces deux expériences réside dans le fait que dans chacune d'elle, les acteurs qui ont conçu le programme expérimental étaient différents de ceux qui ont conçu le dispositif d'évaluation. Dans les deux cas, les activités/pratiques innovantes sont d'abord le résultat d'un travail collectif d'un ensemble d'acteurs spécialistes du domaine considéré : conseillers pédagogiques en activités musicales dans le premier cas, didacticiens des langues et linguistes dans le second cas.

Ce travail d'élaboration de l'innovation pédagogique a demandé, dans les deux cas, une année de travail. Dans le cas du programme EVLANG, c'est sans doute en raison de la dimension internationale du projet (cinq pays, huit équipes de recherches) et du caractère encore assez « exploratoire » de la démarche didactique considérée<sup>3</sup> ; dans le cas de la musique, la durée de conception est due à la richesse du programme d'activités à mettre en place<sup>4</sup>. Cette élaboration collective conduite (dans les deux cas), à partir d'un ensemble de références théoriques, pédagogiques ou didactiques, a donné lieu à la formulation d'un corpus d'hypothèses très précises quand aux effets attendus de la mise en œuvre de ces activités auprès des élèves visés.

Dans le cas d'EVLANG, les activités ont d'abord été « testées » dans un petit nombre de classes par des enseignants volontaires dont certains ont été associés à la conception même de ces activités : l'équipe de concepteurs parle à ce propos de phase de « validation » permettant de régler certaines adaptations concrètes à la réalité des classes, relatives aux contenus, aux supports pédagogiques et à la durée de ces activités. Pour les activités musicales, le programme d'activités découle de l'expérience de terrain accumulée par les conseillers pédagogiques au cours des années antérieures dans des contextes pédagogiques très variés.

Dans les deux cas, la volonté des concepteurs était d'obtenir un programme d'activités pouvant être aisément mis en œuvre dans des classes « ordinaires » par des enseignants non spécialistes de ces domaines d'activités (la musique d'une part, les langues étrangères d'autre part). C'est ce qui a conduit ces équipes à vouloir expérimenter ces activités dans un cadre dépassant largement celui d'enseignants volontaires – voire militants – et d'inscrire cette démarche expérimentale dans une perspective de diffusion large de ces innovations.

C'est dans cette dynamique que s'est trouvée posée, dans ces deux équipes, la question de l'évaluation de la mise en œuvre de ce programme expérimental et c'est pour dépasser le cadre d'une évaluation interne, centrée sur les activités en elle-même, que le recours à des acteurs « spécifiques » de l'évaluation s'est imposé.

### ***Intégrer le dispositif d'évaluation au dispositif d'innovation***

Distinguer les fonctions de conception de l'innovation de celle de son évaluation ne signifie pas pour autant qu'aucune articulation ne doive venir réguler ces deux fonctions. La position des « acteurs de l'évaluation », dans



cette situation, si elle peut être qualifiée «d'externe», n'en demande pas moins un travail important d'intégration du cadre général d'évaluation proposé (qui dans ces deux cas se basait sur les principes méthodologiques décrits précédemment) au contexte de l'innovation considérée.

De ce point de vue, plus le corpus d'hypothèses qui fonde la démarche d'innovation est précis et plus le cadre général de l'évaluation est formalisé, plus ce travail d'adaptation est facilité: il consiste notamment à repérer comment chacune des hypothèses peut être instruite par le dispositif d'évaluation et à décrire les conditions d'expérimentation et de recueil des données qu'il faut réunir pour ce faire.

Lors de ce travail, il se peut que l'on arrive au constat que certaines hypothèses ne peuvent pas être testées par ce type de dispositif. Dans le cas d'EVLANG, l'hypothèse suivante: «Ces avantages (du programme EVLANG) ne peuvent être obtenus qu'à partir d'un certain degré de diversité des langues à propos desquelles les activités s'exercent» a été exclue, d'un commun accord, du dispositif d'évaluation: les conditions de son examen s'avéraient trop contraignantes et peu réalistes (il aurait fallu, pour ce faire, faire varier, selon ce critère, les curriculum d'activités proposés aux enseignants).

Il se peut également que certaines des hypothèses au fondement de la démarche d'innovation trouvent des conditions d'examen dans une démarche d'évaluation alternative. Dans le cas d'EVLANG, par exemple, l'hypothèse suivante: «Dans le contexte de classes plurilinguistiques et pluriculturelles, ces activités sont susceptibles de contribuer à une valorisation de chaque langue et culture aux yeux des élèves qui en sont originaires comme à ceux des autres» a été traitée par le recours à une démarche qualitative concernant un petit nombre d'enseignants et d'élèves sous la forme d'entretiens semi-directifs.

Dans le cas de l'expérimentation des activités musicales, et au regard de travaux antérieurs sur la question (d'ailleurs peu nombreux), les effets supposés de la musique concernaient plusieurs dimensions des apprentissages des élèves (scolaires et cognitives); il fallait donc construire des épreuves d'acquisitions qui permettaient une mesure du caractère transversal de la musique. Ainsi, en maternelle, les épreuves concernaient la maîtrise de concepts de base (discrimination spatio-temporelle), mais aussi la mémoire, la discrimination auditive et des compétences graphiques. En fin de CP, les acquisitions des élèves ont été évaluées dans les deux disciplines dites fondamentales, à savoir, les mathématiques et la lecture-écriture.

C'est à l'issue de ce travail de mise en relation entre les hypothèses sous-jacentes à la conception du dispositif innovant et le cadre méthodologique d'évaluation proposé que peut s'élaborer le dispositif d'expérimentation. En effet, si le travail de conception du dispositif innovant peut être antérieur à celui du dispositif d'évaluation, les acteurs de l'évaluation pouvant même en être totalement exclus (comme c'est le cas ici), il est par contre nécessaire que le dispositif d'expérimentation, c'est-à-dire les conditions concrètes de mise en œuvre de cette innovation dans les classes (définition de l'échantillon, déroulement dans le temps, conception des outils de recueil des données, etc.) soit élaboré en commun. C'est à ce propos que nous parlons d'intégration du dispositif d'évaluation au dispositif d'innovation.

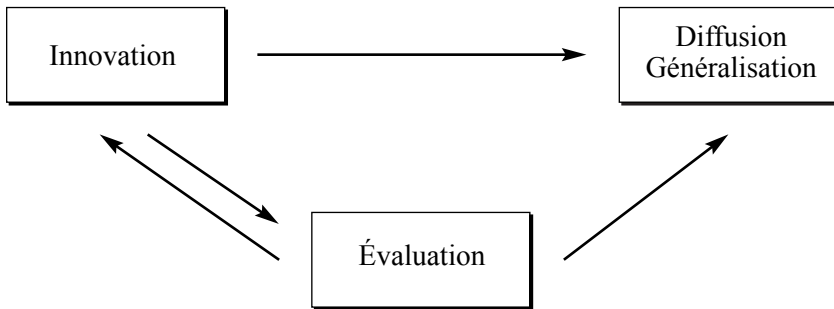
Si l'expérimentation se met en place avant que le dispositif d'évaluation ne soit élaboré on risque en effet de se retrouver dans l'impossibilité matérielle de tester certaines hypothèses ; par exemple, dans le cas où il n'y pas eu d'évaluation du niveau initial des élèves dans le domaine d'acquisition considéré avant la mise en œuvre de l'expérimentation dans les classes. C'est ce même phénomène qui a pu, par exemple, conduire à limiter la portée des résultats de certaines recherches que nous avons conduites par ailleurs<sup>5</sup>.

Notons enfin qu'on peut très bien concevoir que ce travail de mise en relation entre les cadres conceptuels respectifs du dispositif d'innovation et de celui de l'évaluation soit conduit par les mêmes acteurs. À ce titre, le fait que dans ces deux recherches ces acteurs soient distincts ne constituent pas, de notre point de vue, une condition incontournable de l'efficacité du dispositif d'évaluation. Cependant, dans ces deux cas, cette situation nous a semblé particulièrement féconde en ce qu'elle oblige les acteurs en présence à un travail d'explicitation réciproque de chacun de leurs champs d'actions ou de recherche (par exemple dans le cas d'EVLANG entre didactique et sciences de l'éducation) et à dépasser la logique propre à chacun de ces champs pour arriver à une construction scientifiquement négociée.

### ***Intégrer les résultats de l'évaluation au processus de diffusion de l'innovation et de formation des enseignants***

Outre l'intérêt autonome que peut susciter l'évaluation d'une innovation pour le milieu scientifique, il est essentiel que les résultats de cette évaluation puissent servir au processus de diffusion de l'innovation, voire à une généralisation. Cet objectif a été largement atteint dans le cas de l'évaluation des activités musicales puisque les résultats positifs de l'évaluation ont permis la validation et la reconnaissance institutionnelle du programme élaboré au

départ par les conseillers pédagogiques. Concrètement, et après quelques ajustements consécutifs à l'expérimentation, le programme a fait l'objet d'une publication nationale, «labellisée» par le Ministère et diffusée dans toutes les Inspections académiques en France<sup>6</sup>. En outre, depuis sa parution, la publication rencontre un succès constant auprès des enseignants. Cet exemple montre comment l'évaluation peut servir d'élément régulateur à une innovation (schéma suivant).



Il n'apparaît pas souhaitable en effet que le processus de diffusion ou de généralisation intervienne de façon directe sans qu'une mesure des effets de l'innovation ait eu lieu antérieurement. C'est malheureusement assez souvent le cas en matière de politique éducative quand les innovations sont liées à des réformes institutionnelles visant le territoire national.

Il faut toutefois reconnaître que lorsque les évaluations fournissent des résultats négatifs sur une innovation pédagogique particulière, il est plus difficile de cerner les avantages de l'évaluation ; ils sont néanmoins réels. En premier lieu, l'évaluation peut servir à faire «changer de cap» une mesure de politique éducative qui n'a pas montré les effets escomptés, ce qui pourrait se traduire par l'arrêt du processus d'innovation ou d'expérimentation ou, de manière moins radicale, par une nouvelle réflexion sur les modalités d'application de cette innovation : on se situe alors dans la sphère du décideur. En second lieu, les acteurs de l'innovation, en s'appropriant les résultats de l'évaluation, peuvent être amenés à remettre en cause des hypothèses et à en émettre de nouvelles : on est alors dans un référentiel théorique ou didactique.

Outre l'intérêt de l'évaluation comme élément préalable à la diffusion des innovations, celle-ci peut avoir un impact plus large et plus durable par le biais de la formation des enseignants. La connaissance des résultats des évaluations conduites et de leurs applications directes sur le plan pédagogique peut participer à l'élaboration d'une culture de l'évaluation externe chez les enseignants en formation initiale et continue. Dans cette perspective, il conviendrait de considérer l'évaluation comme une étape essentielle de tout dispositif innovant. Par exemple, l'évaluation d'une démarche innovante en lecture-écriture a fourni à ses concepteurs des éléments factuels qui vont pouvoir être intégrés dans les actions de formation accompagnant le développement de cette innovation sur de nouveaux sites.

Au final, l'évaluation peut devenir un outil de pilotage au service de tous les acteurs intéressés par les processus d'innovation. Si en effet dans les paragraphes précédents nous avons plus particulièrement développé les relations entre «concepteurs» et «évaluateurs», il se trouve que dans le cadre du développement de l'institutionnalisation de l'innovation, les acteurs intégrés dans de tels dispositifs sont de plus en plus nombreux et variés. À cet égard, les collectivités locales, les associations et, plus largement, toutes les structures qui participent à la mise en œuvre de tels dispositifs dans le temps périscolaire par exemple, sont potentiellement touchées par la problématique de l'évaluation, même si leurs intérêts à son encontre diffèrent. Certains (les «opérateurs») sont sans doute plus intéressés par les dimensions liées à l'efficacité pédagogique, d'autres («les financeurs») par des analyses de type «coût/efficacité»...

Plus globalement encore, dans le cadre plus général de l'institutionnalisation de l'évaluation des politiques publiques, il serait souhaitable que les résultats des évaluations puissent être disponibles pour le plus grand nombre d'utilisateurs du service public d'éducation.

## NOTES

1. Recherches publiées : activités musicales en maternelle (Mingat & Suchaut, 1994), introduction de l'enseignement des langues étrangères à l'école primaire (Genelot, 1995), pratiques innovantes en lecture (Le Bastard & Suchaut, 2000). Recherches en cours : éveil aux langues à l'école primaire (Genelot), pratiques en orthographe au cycle 3 (Genelot), évaluation du dispositif d'accompagnement scolaire au CP «Coups de pouce» (Suchaut).
2. Dans la plupart des recherches conduites à l'IREDU sur les acquisitions scolaires (dans différents domaines) à l'école primaire, on relève qu'en général, le niveau initial des élèves explique entre 30 et 50 % de la variété du niveau final considéré.
3. Les pratiques dites «d'éveil aux langues» constituent une approche de la didactique des langues encore assez peu structurée sur le plan européen, même si on trouve dans de nombreux pays des équipes de recherche qui ont expérimenté localement des programmes plus ou moins développés relevant de cette démarche. C'était d'ailleurs un des objectifs du programme Socratès – Lingua «EVLANG» que de contribuer à «fédérer» ces différentes expérimentations.
4. Le programme d'activités musicales regroupe une centaine de fiches pédagogiques détaillées avec leur support audio.

## RÉFÉRENCES

- Cros, F. (1997). L'innovation en éducation et en formation. *Revue française de pédagogie*, 118, 127-156.
- Demailly, L. (1991). *Le collège: crise, mythes et métiers*. Lille: PUL. (Collection mutations/sociologie)
- Genelot, S. (1995). L'enseignement des langues à l'école élémentaire: éléments d'évaluation des effets au collège. *Les cahiers de l'Iredu*, 58, 209 p.
- Genelot, S., & Tupin, F. (1999). *L'éveil aux langues en Europe: à la recherche d'une efficacité*. Communication au congrès international de l'actualité de la recherche en éducation et en formation, Bordeaux.
- Marsollier, C. (1998). Ouverture ou résistance des enseignants du premier degré à l'égard de l'innovation. *Les sciences de l'éducation pour l'ère nouvelle*, 31 (4), 95-121.
- Mingat, A., & Suchaut, B. (1994). Évaluation d'une expérimentation d'activités musicales en grande section de maternelle. *Les cahiers de l'Iredu*, 56, 200 p.
- Le Bastard, S., & Suchaut, B. (2000). Lecture-écriture au cycle II: évaluation d'une démarche innovante. *Les cahiers de l'Iredu*, 61, 138 p.