

Quarante ans d'évaluation ministérielle des acquis des élèves en France : complexification et politisation

Xavier Pons

Université Paris-Est Créteil.

Mots-clés : Éducation nationale – Évaluation – Action publique.

Adoptant une perspective de sociologie historique des instruments d'action publique, cet article retrace la carrière d'un outil d'évaluation utilisé dans le champ de l'éducation en France depuis le milieu des années 1970, particulièrement emblématique des hésitations contemporaines en matière de régulation de ce système scolaire : les évaluations standardisées des acquis des élèves mises en œuvre par le ministère de l'Éducation nationale. Sur la base de matériaux rassemblés dans le cadre de trois recherches qualitatives complémentaires, l'auteur analyse la succession de quatre étapes dans la carrière de cet outil au gré des impulsions et des ruptures introduites par les décideurs, de l'évolution de ses propriétés méthodologiques et morphologiques et de sa réception par les professionnels de l'éducation. Cette perspective met en évidence deux évolutions majeures de l'outil que sont sa complexification et sa politisation.

régulation des systèmes scolaires depuis les années 1980 (Maroy, 2006) ou encore des nouveaux modes de légitimation du pouvoir selon lesquels on gouvernerait de plus en plus par des instruments (Lascoumes et Le Galès, 2004).

Mais cette étude est aussi relativement difficile à mener. Premièrement, la notion d'« évaluation » reste floue et polysémique. Les articles L. 241-1 à L. 241-11 du Code de l'éducation confient bien une mission d'évaluation à certains acteurs (principalement les corps d'inspection et les services statistiques de l'administration centrale et des rectorats) mais sans jamais expliciter la notion d'« évaluation » elle-même, les démarches à son fondement et les finalités politiques visées. De plus, l'évaluation des politiques d'éducation s'est globalement développée, jusqu'au milieu des années 2000, en marge du mouvement interministériel d'évaluation des politiques publiques à tel point que les éléments méthodologiques et conceptuels mis en avant par le Conseil scientifique de l'évaluation (1990-1998), par exemple, ont été peu repris dans le secteur éducatif, souvent au motif de la particularité de ce dernier et de la difficulté à établir un « résultat » en éducation. On retient donc ici une définition volontairement générique de l'évaluation des politiques éducatives, désignant « un processus collectif et réflexif de construction d'un retour d'information sur le fonctionnement du système éducatif permettant d'apprécier, en référence à des valeurs politiques multiples (pertinence, cohérence, légalité, efficacité, efficience ou encore équité), son organisation, ses résultats et les effets de différentes mesures mises en œuvre » (Pons, 2011:9). Deuxièmement, cette activité d'évaluation donne lieu à de nombreux outils d'action publique, souvent très techniques, qui ont leurs spécificités et leur propre dynamique

Étudier les politiques d'éducation (1) françaises sous l'angle de leur évaluation se révèle particulièrement fertile pour comprendre leurs caractéristiques et leur évolution. Comme le précise Agnès van Zanten (2011), les valeurs au fondement de l'école républicaine sont particulièrement ambitieuses de sorte qu'elles sont partagées par un très grand nombre d'acteurs à un certain niveau de généralité mais qu'elles sont également très difficiles à mettre en œuvre sans susciter la désapprobation d'une partie importante d'entre eux. Analyser les évaluations qui sont effectivement conduites est, dès lors, un bon moyen d'interroger les finalités et les valeurs réellement poursuivies par les politiques d'éducation. Par ailleurs, l'évaluation constitue à bien des égards une mesure emblématique : de la nouvelle gestion publique telle qu'elle a pu être synthétisée par certains politologues (Hood, 1991 et 1996), des changements de

(1) L'éducation désigne ici l'enseignement scolaire obligatoire primaire et secondaire.

institutionnelle. Troisièmement, l'usage de ces divers outils et leurs liens ont rarement été formalisés dans une politique d'évaluation d'ensemble, celle-ci étant le résultat de deux paramètres principaux : la concurrence que se livrent les évaluateurs entre eux en matière de production de connaissances et leurs liens de dépendance plus ou moins forts selon les contextes avec les décideurs politiques (Pons, 2010 a). Le choix a été ici de se focaliser sur la carrière d'un outil d'évaluation en particulier : les évaluations nationales standardisées des acquis des élèves conduites par la Direction de l'évaluation, de la prospective et de la performance (DEPP) et ses formes passées (2) depuis les années 1970 (3). Contribuant à l'analyse des résultats du système éducatif (à travers ceux des élèves en son sein) et des effets sur les apprentissages des mesures destinées à le réformer, cet outil est l'un des plus anciens mis en œuvre par le ministère, même si ces formes ont considérablement évolué (Normand, 2005) et sa carrière constitue un bon indicateur de l'évolution de l'action publique éducative en matière d'évaluation (Pons, 2010 a). À cet égard, il a donné lieu à plusieurs travaux mobilisés pour l'analyse. Enfin, dans la mesure où il concerne les élèves, on peut supposer, *a priori*, qu'il constitue un outil d'évaluation relativement proche de l'univers familial (4).

Comme le mettent en évidence plusieurs spécialistes de la sociologie des instruments d'action publique (Lascoumes et Le Galès, 2004), il est essentiel d'étudier ces derniers non pas seulement à travers leurs propriétés intrinsèques mais aussi en les replaçant dans l'espace social de référence qui se crée autour d'eux et qui agit sur eux. Ainsi, la carrière des évaluations standardisées des acquis des élèves sera retracée en analysant les configurations historiques générales dans lesquelles elles prennent forme.

Ces dernières sont elles-mêmes envisagées à travers trois de leurs dimensions : les liens de ces évaluations avec la politique éducative impulsée par les décideurs (dimension politique), les propriétés méthodologiques et morphologiques générales de l'outil (dimension cognitive) et les relations entre les groupes professionnels concernés par ce dernier (dimension professionnelle). Cette approche permet, dans la suite de l'article, de distinguer quatre configurations historiques et de mettre en évidence, en conclusion, deux tendances principales au cours de la période : la complexification et la politisation de l'outil (5). Pour cela, sont utilisés des matériaux rassemblés par le biais de méthodes qualitatives au cours de trois recherches, matériaux complétés, pour cet article, par l'analyse d'un corpus de soixante-dix dépêches de l'Agence Éducation Formation (AEF) (6) pour la période allant d'octobre 2009 à la mi-mai 2012. La première est une analyse critique de l'abondante littérature institutionnelle publiée en France sur les évaluations standardisées des acquis des élèves, réalisée entre 2004 et 2005 dans le cadre d'une collaboration avec Nathalie Mons (Mons et Pons, 2006). Cette littérature est composée principalement d'enquêtes, de rapports publics, de documents méthodologiques et de banques d'épreuves (disponibles en ligne à l'adresse suivante <http://educ-eval.education.fr>) La deuxième est une thèse de doctorat effectuée entre 2004 et 2008 portant sur l'évaluation des politiques éducatives en France (Pons, 2010 a) qui a été l'occasion d'effectuer trente-huit entretiens semi-directifs auprès des acteurs concernés par la conception et la mise en œuvre des évaluations standardisées, d'analyser un corpus de publications de la DEPP et de recueillir plusieurs documents internes à cette direction. La troisième est une recherche ponctuelle sur le lancement des évaluations nationales

(2) Cette direction, créée en 1987 sous une autre appellation sur les bases du service statistique ministériel, a changé plusieurs fois de nom. Par commodité, on utilise le sigle actuel.

(3) Par « évaluation standardisée des acquis des élèves », on désigne toute épreuve évaluant les acquis des élèves dont la conception, l'administration, la correction et la publication des résultats répondent à un cadre standardisé, c'est-à-dire normé, que son objectif soit formatif, sommatif, certificatif ou d'aide au pilotage du système éducatif (Mons et Pons, 2006). Même s'ils peuvent être intégrés dans cette définition, on écarte du champ d'analyse, par souci d'efficacité, le baccalauréat, plus ancien et obéissant à une autre logique de certification, les épreuves conçues dès 1967 par le ministère de la Défense (actuel test dit « des trois jours ») et les comparaisons internationales.

(4) Deux nuances doivent toutefois être apportées. D'une part, cet outil n'est pas le seul à être potentiellement mobilisé par les familles. On pense en particulier aux palmarès des lycées publiés régulièrement dans la presse depuis les années 1980 et qui ont contribué à faire émerger dans le débat public la figure des « consommateurs d'école » (Ballion, 1982). D'autre part, les recherches disponibles sur l'usage de ces évaluations (d'élèves et d'établissements) par les familles montrent que celles-ci privilégient à ces informations « froides » en général d'autres sources d'information « chaudes » pour apprécier la qualité des établissements scolaires comme la rumeur, la réputation et l'apparence générales de l'établissement, les dires d'un réseau de relations personnel proche, les propos des enseignants ou des élèves eux-mêmes, même si cette tendance est socialement différenciée (van Zanten, 2001).

(5) Ces deux termes sont employés ici dans une perspective descriptive. La complexification peut être comprise globalement comme la mise en système d'éléments s'ajoutant à une structure de départ unifiée et relativement simple. Il peut s'agir, par exemple, de l'affinement des méthodes et des techniques employées (complexification interne) comme de la mise en relation plus ou moins formalisée et volontaire de dispositifs d'évaluation aux finalités différentes (complexification externe). La politisation désigne, dans cet article, un processus par lequel des considérations d'ordre politique, au sens de « *politics* » (donc de la lutte entre des forces politiques en présence), pénètrent progressivement la mise en œuvre et les usages des outils d'évaluation.

(6) L'AEF a été créée en 1998 et regroupe actuellement une quarantaine de journalistes, dont une petite dizaine travaillant spécifiquement sur l'enseignement scolaire. Elle propose à ses clients un fil d'informations continu, synthétique et factuel sur la mise en œuvre des réformes du système éducatif et l'évolution du débat public national en matière d'éducation dont on a pu mesurer, à plusieurs reprises, l'importance dans le secteur de l'éducation au cours de différentes recherches.

en classes primaires CE1 et CM2, fondée sur une dizaine d'entretiens et un corpus réduit de dépêches de presse (Pons, 2010 b).

Développer un nouvel outil (1973-1988)

En 1973, le service statistique du ministère reçoit pour mandat de procéder à l'évaluation de l'action éducative. L'année suivante, il met en place un « bureau des évaluations pédagogiques » dont il confie la gestion à un jeune énarque. Ce dernier dispose d'une liberté totale pour organiser le service. Pendant un an, secondé par une jeune statisticienne, il sillonne les territoires pour entendre les spécialistes de l'évaluation standardisée des élèves et recruter les futurs membres de son service. Devant l'absence de commande politique claire, cette petite équipe s'adjoint la compétence de quelques chercheurs de l'Institut national de la recherche pédagogique (INRP) et expérimente, entre 1974 et 1976, la première évaluation nationale sur échantillon des acquis (7) des élèves du collège, en français et en mathématiques, en fin de classe de 5^e.

Cette nouvelle orientation prise au milieu des années 1970 s'explique par la congruence de quatre facteurs principaux. Le premier renvoie à l'unification formelle du premier cycle de l'enseignement secondaire en France qui débouche sur la création du collège unique en 1975, consistant à accueillir dans un même établissement tous les élèves des classes de 6^e à la 3^e pour donner à tous les collégiens le même enseignement, et ce afin d'élargir et de démocratiser l'éducation. Ce processus pose progressivement la question des effets du fonctionnement de ce système éducatif ainsi unifié sur la réussite scolaire des élèves. Le deuxième facteur est d'ordre cognitif. Plusieurs courants de recherche en éducation contribuent à diffuser l'idée d'un développement nécessaire de l'évaluation. Historiquement, le premier est sans doute la docimologie (définie globalement comme l'étude systématique des pratiques d'examen) qui, en montrant les multiples biais liés à l'évaluation des élèves par les enseignants, a pour principale conséquence de désacraliser l'évaluation professorale et d'inviter l'institution scolaire à multiplier les formes d'évaluation alternatives des élèves, dont les évaluations standardisées (Vogler, 1996). Le troisième facteur concerne l'action publique au niveau interministériel, notamment les développements de la planification et, à partir de 1968, de la procédure de rationalisation des choix budgétaires (RCB). Si la production et la diffusion par le ministère de l'Éducation nationale des travaux conduits lors de ces expériences administratives

sont limitées, ces procédures interministérielles permettent à certains réformateurs de justifier la création d'un service statistique à part entière au sein du ministère, de recruter de nombreux chargés d'étude et d'acquérir progressivement les compétences techniques nécessaires à la mise en œuvre d'outils statistiques de grande envergure. Enfin, les statisticiens du ministère ont su faire preuve d'opportunismes technique et politique. Bénéficiant des données recueillies par les premiers panels d'élèves mis en place à partir de 1972, le chef du service statistique d'alors convainc le ministre René Haby, en pleine réforme sur le collège unique, de l'utilité des évaluations de masse des élèves pour mesurer leur degré de maîtrise des nouveaux programmes (Pons, 2011).

Malgré les réticences d'une partie des inspecteurs, une deuxième évaluation des élèves en lecture et en mathématiques en fin de cours préparatoire (CP) de l'école primaire voit le jour en 1979. Dans les années 1980, ces évaluations sont de plus en plus fréquentes mais leur périodicité, les niveaux et les disciplines interrogés et leurs finalités ne sont jamais stabilisées (tableau 1 p. 12). Deux types d'évaluation émergent cependant. Les premières ont lieu en début de cycle (classe de CE2 à l'école primaire et classes de 6^e et 5^e au collège) et visent à permettre d'établir un diagnostic des acquis des élèves utile à l'enseignant pour concevoir une pédagogie différenciée (évaluations diagnostiques). Les secondes interviennent en fin de cycle (classe de CM2 au primaire et classe de 3^e au collège) pour effectuer un bilan de ces acquis (évaluations bilans). Si elles présentent un certain nombre de points communs comme leur objet (les « acquis »), le mode de composition des groupes de conception (par cooptation) ou certaines étapes méthodologiques (construction des échelles de compétences, expérimentation des épreuves l'année précédant la mise en œuvre, etc.), les évaluations relèvent de finalités cognitives et politiques différentes. L'une, exhaustive et relativement facile, est conçue comme un outil d'accompagnement professionnel de l'enseignant qui vise à détecter les difficultés des élèves et leurs principales causes. L'autre, plus sommative (mesure des connaissances acquises au terme d'un processus ou d'un cycle d'apprentissage) et fondée sur un échantillon, est destinée à être un outil de pilotage national.

Ces évaluations ne sont assorties d'aucune obligation légale et répondent d'abord aux exigences de la mise en place de nouveaux programmes à la suite de la réforme du collège unique (notamment en 1980, 1982 et 1984). Elles répondent aussi à des commandes ponctuelles des ministres selon leurs

(7) C'est-à-dire à la fois des connaissances, des compétences et des comportements.

Tableau 1

Les évaluations des acquis des élèves de 1980 à 1989

| Dates | Classes | Types d'évaluation | Disciplines |
|-------|------------------------------|--------------------|---|
| 1980 | 6 ^e | Diagnostique | Français et mathématiques |
| 1981 | CE2 | Diagnostique | Français et mathématiques |
| | CM2 | Bilan | Français et mathématiques, disciplines d'éveil |
| 1982 | 5 ^e | Bilan | Français, mathématiques, anglais, allemand, sciences physiques, sciences naturelles, vie scolaire |
| 1983 | CM2 | Bilan | Français, mathématiques, disciplines d'éveil |
| 1984 | 3 ^e | Bilan | Français, mathématiques, anglais, allemand, sciences expérimentales, histoire-géographie, vie scolaire |
| 1986 | 5 ^e | Bilan | Français, mathématiques, vie scolaire |
| | Seconde | Bilan | Français, mathématiques, anglais, allemand, histoire-géographie, sciences économiques et sociales, sciences physiques, sciences naturelles, technologie, éducation physique et sportive, vie scolaire, expression orale |
| 1987 | CM2 | Bilan | Lecture, écriture |
| | Première | Bilan | Français |
| | Terminale | Bilan | Histoire-géographie (à partir des copies de baccalauréat) |
| 1988 | 4 ^e technologique | Bilan | Non précisé |
| | 3 ^e technologique | Bilan | Non précisé |
| | 3 ^e | Bilan | Culture économique |
| | Terminale | Bilan | Anglais (à partir des copies de baccalauréat) |
| 1989 | CE1 | Bilan | Lecture |
| | CE2 | Diagnostique | Français et mathématiques |
| | 6 ^e | Diagnostique | Français et mathématiques |
| | Terminale | Bilan | Culture économique |

Source : Levasseur J., 1996, *L'évaluation nationale des acquis des élèves*, *Revue internationale d'éducation*, n° 11:101-114.

initiatives (exemple de l'évaluation de 1988 sur l'enseignement technologique) ou selon l'évolution des acquis des sciences de l'éducation au sens large. Les premières évaluations au niveau primaire, par exemple (1979, 1981 et 1983), ne concernent pas seulement les mathématiques et le français mais recouvrent également la plupart des « disciplines d'éveil » conformément aux évolutions pédagogiques souhaitées à ce moment-là. Cette configuration présente certains avantages. Les groupes de conception des épreuves sont l'occasion de faire discuter, dans un cadre nouveau et relativement souple, des acteurs dont la communication institutionnelle est parfois difficile (inspecteurs et enseignants) ou qui se connaissent peu (statisticiens et inspecteurs, par exemple). Elle suscite également dans les académies plusieurs travaux autour d'inspecteurs territoriaux volontaires. Cependant, cette appropriation spontanée est porteuse d'ambiguïtés sur l'identité et les finalités mêmes de l'outil. Elle pose, en outre, deux problèmes techniques considérables aux statisticiens, l'éparpillement des données et la nécessité de fréquemment reconstituer les échantillons, exposant ainsi les évaluateurs à plusieurs critiques de formes diverses selon les périodes : impossibilité de traiter les données de

manière longitudinale, insuffisant contrôle des variables contextuelles pouvant expliquer les résultats observés, condamnation du tout quantitatif véhiculé par l'outil, etc.

Dans ce contexte, les décideurs politiques, en accordant plus ou moins d'importance aux arguments de tel ou tel groupe professionnel, jouent un rôle d'arbitre. Le ministre de Alain Savary, par exemple (1981-1984), est très soucieux de la bonne mise en œuvre des nouveaux programmes. Jean-Pierre Chevènement (1984-1986) semble y accorder moins d'importance. Pour l'équipe du premier ministre de l'Éducation nationale à exercer dans un contexte de cohabitation qu'est René Monory (1986-1988), les évaluations des acquis des élèves constituent un moyen approprié – car il est indirect – d'impulser des changements dans

un système éducatif perçu comme étant bloqué. S'inspirant du cas des États-Unis, cette nouvelle équipe utilise ces évaluations comme un outil permettant de renvoyer aux enseignants les résultats de leurs pratiques en espérant que ces derniers les modifient et trouvent d'eux-mêmes les solutions, donc d'agir sur le corps enseignant sous couvert de modernisme tout en évitant la confrontation traditionnelle avec les syndicats.

Systématiser l'outil, théoriser l'évaluation (1989-1997)

L'année scolaire 1988-1989 est un tournant politique important. En mars 1989, en plein débat sur l'illettrisme, le ministre Lionel Jospin annonce que des évaluations des acquis des élèves en français et en mathématiques à l'entrée en classe de CE2 en primaire et en 6^e au collège seront conduites chaque année à partir de la rentrée suivante. Cette annonce fait suite à plusieurs semaines de négociations avec les syndicats d'enseignants pour préparer la future loi d'orientation du 10 juillet 1989 qui comportera, parmi d'autres, un titre sur « l'évaluation du système éducatif ». Au cours de ces négociations,

plusieurs syndicats rappellent le principe fondamental de leur liberté pédagogique et évoquent leur crainte de voir ces outils d'évaluation progressivement normer le comportement des élèves et des familles. Ils craignent, surtout, que ces évaluations nationales soient un moyen détourné d'évaluer les enseignants eux-mêmes. Ils obtiennent ainsi du ministre, premièrement, qu'il n'y ait aucune obligation légale faite à l'enseignant de tenir compte des résultats de ces évaluations, deuxièmement, que les résultats ne soient pas publiés à l'échelle de l'établissement pour ne pas attiser le consumérisme scolaire des parents et, troisièmement, que ces évaluations soient bien des évaluations diagnostiques, donc des outils professionnels au service de l'enseignant et de la remédiation de l'échec scolaire, et non des évaluations bilans destinées à renseigner le pilotage du système éducatif.

Malgré ces réserves, le caractère désormais systématique des évaluations CE2-6^e et leur confirmation politique par la loi de 1989 changent considérablement les rapports sociaux entre les acteurs concernés. Si la période antérieure était plutôt marquée par des tâtonnements organisationnels, des initiatives personnelles et la possibilité pour certains acteurs, notamment les enseignants et les inspecteurs, de minorer l'importance de cet outil, voire de l'ignorer, celui-ci occupe désormais une place centrale dans le calendrier de la rentrée et est intégré dans une politique globale d'évaluation du système éducatif. Une campagne de formation massive est réalisée à la rentrée scolaire de 1989. La DEPP propose deux stages de formation de deux jours à destination des inspecteurs territoriaux et de divers formateurs. Ces derniers forment ensuite, à leur tour, tous les enseignants de France, obligés d'être présents à cette journée de formation. Certains centres de documentation pédagogique territoriaux publient les actes de ces journées et plusieurs séminaires de rentrée de l'administration centrale sont consacrés à ce thème.

Les évaluations standardisées deviennent progressivement l'un des outils fondamentaux de la stratégie impulsée par Claude Thélot à la tête de la DEPP (1990-1997) dont l'ambition est de faire de cette direction une interface légitime entre les décideurs politiques, l'administration du ministère, la recherche académique et l'opinion publique, l'évaluation et, plus globalement, la statistique devant contribuer à structurer le débat public en fournissant à tous des constats chiffrés partagés. Les évaluations diagnostiques deviennent le fer de lance de la diffusion d'une nouvelle « culture de l'évaluation » auprès des professionnels de l'éducation (Thélot, 1994). Le dispositif est ainsi étendu aux lycées en 1992 avec la création d'une nouvelle évaluation à l'entrée en seconde. En mars 1994, une première enquête de la DEPP par questionnaire auprès d'un échantillon

représentatif d'enseignants de CE2 et de 6^e révèle que 93 % d'entre eux utilisent les évaluations diagnostiques pour dialoguer avec les parents d'élèves, 82 % pour mettre à jour les besoins de chaque élève et 67 % pour modifier l'enseignement qu'ils dispensent (DEPP, 1994). Même si le primat donné aux évaluations diagnostiques, exhaustives et donc coûteuses en temps et en moyens, produit plusieurs effets pervers, au premier rang desquels la mise en retrait et la conduite irrégulière d'évaluations bilans, la systématique de l'outil amène certains membres de la DEPP, dans leurs écrits, à évoquer l'existence d'un véritable « observatoire des acquis des élèves » (Levasseur, 1996). Ce mouvement débouche même sur une théorisation explicite de l'évaluation que C. Thélot a présentée dans plusieurs de ses publications et qu'il illustre par les évaluations diagnostiques : la théorie de l'effet miroir. Selon cette théorie, l'évaluation consiste à mettre à disposition des acteurs du système éducatif des données chiffrées sur les résultats de leur action en espérant que ces derniers, face au miroir qui leur est tendu, trouvent d'eux-mêmes les remédiations nécessaires (Pons, 2010 a et b).

Perfectionner l'outil, mais dans quel but ? (1997-2007)

L'arrivée de Claude Allègre (1997-2000) rompt clairement cette dynamique. Cette rupture se traduit par la réorganisation interne de la DEPP, par le discrédit jeté publiquement par le ministre sur les données produites par cette dernière (notamment sur la proportion d'élèves ne maîtrisant pas les compétences de base en lecture) et par la défense d'un modèle scientifique d'évaluation externe au ministère de l'Éducation nationale à l'opposé de la conception promue jusqu'alors par C. Thélot, qui démissionne à la rentrée 1997 (Pons, 2010 a et b). Si les évaluations diagnostiques en CE2 et en 6^e sont maintenues, la publication d'évaluations bilans est clairement entravée au cours de la période, à l'exception notoire de deux évaluations visant à reconduire le protocole d'évaluation de l'enquête « Lire, écrire, compter » de 1987 en fin de CM2 pour procéder à des comparaisons temporelles (en 1997 pour la lecture et en 1999 pour les mathématiques). Par ailleurs, des travaux publiés au début des années 2000 interrogent le pouvoir régulateur des évaluations standardisées (notamment les évaluations diagnostiques) et mettent en évidence des usages ambivalents de la part des professionnels de l'éducation. Si une seconde enquête de la DEPP, en 2005, permet de conclure, au prix de la construction de catégories statistiques agrégées très générales, que les évaluations diagnostiques en classe de 6^e contribuent à faire évoluer les pratiques évaluatives de la majorité des enseignants

de collège (8), une enquête ethnographique conduite dans trois académies par des chercheurs de l'INRP met en évidence que les enseignants éprouvent de grandes difficultés à intégrer ces évaluations dans leur pédagogie et dans leur communication auprès des parents (Derouet et Normand, 2003). Lise Demailly (2003) parle même de résistance des enseignants qu'elle explique, sur la base des travaux qu'elle a coordonnés pendant plusieurs années dans l'académie de Lille, par la défense d'intérêts corporatistes (peur d'une relation entre l'évaluation et le déroulement des carrières, refus de l'ingérence dans une pratique professionnelle qui veut maintenir ses zones d'autonomie), par l'affirmation de grandes valeurs (refus d'une technologie liée au management privé, d'une évaluation fondamentalement injuste) et par une distance cognitive face à l'usage de la mesure statistique, inadaptée pour saisir l'acte d'enseignement et ses effets.

Cette ambivalence prend une autre forme pour les chefs d'établissement. Si le pilotage par les résultats fait désormais partie de leurs missions et s'ils reconnaissent que ces évaluations améliorent leurs marges d'action, elles sont également sources d'une bureaucratisation supplémentaire fortement critiquée et elles s'insèrent mal dans les relations de travail quotidiennes (Barrère, 2009). On observe ainsi une intégration très inégale d'un établissement à l'autre de ces évaluations dans les projets d'établissement (Verdière, 2001). A. van Zanten (2001) montre, en outre, que quand ces évaluations sont utilisées par les établissements scolaires, c'est moins pour fonder un projet pédagogique au sein de l'établissement que pour maintenir sa réputation globale (dans le cas d'établissements d'excellence) ou se fabriquer une bonne image (cas d'établissements désireux de lutter contre une stigmatisation abusive en montrant leur bonne volonté à participer à des évaluations). Au niveau des collectivités territoriales, certains travaux montrent que les évaluations diagnostiques ont agi comme un outil puissant d'observation que les municipalités ont su s'approprier en contournant les experts traditionnels du système éducatif pour s'affirmer comme de véritables acteurs en matière de politique éducative (Dutercq, 2000 ; van Zanten, 2000). Mais ils mettent aussi en évidence un investissement très inégal des collectivités selon les territoires et une focalisation plus grande sur la dimension cognitive de l'évaluation (la production de connaissances qu'elle permet) que sur sa dimension régulatrice (sa capacité à réorienter effectivement les

pratiques). Au niveau académique, enfin, les évaluations diagnostiques seraient parfois utilisées par les responsables locaux comme des indicateurs de pilotage de leurs écoles non sans détourner ces dernières de leur finalité première (IGEN-IGAENR, 2005 a).

Sur le plan politique, la période qui s'ouvre au début des années 2000 se caractérise donc par de multiples hésitations que révèlent l'inflation de nouveaux dispositifs *ad hoc*, d'un côté, et les nombreux revirements concernant les dispositifs existants, de l'autre. En 2001, par exemple, est abandonnée l'évaluation diagnostique à l'entrée en classe de seconde dans le secondaire alors que l'année suivante est lancée une nouvelle évaluation de ce type à l'entrée en 5^e au collège (IGEN-IGAENR, 2005 b). En 2004-2005, une nouvelle évaluation diagnostique à l'entrée en CE1 en primaire est expérimentée dans le cadre du plan de prévention de l'illettrisme. Expérimentée à nouveau en 2005-2006 afin de « *cerner, au cours du premier trimestre, la nature des graves difficultés que peuvent rencontrer les élèves de CE1* » (circulaire du 24 juin 2005), elle est généralisée en 2006-2007 et abandonnée en 2008 comme l'évaluation du même type lancée l'année précédente en CM2 dans le prolongement de la loi d'orientation de 2005 et de la mise en place des programmes personnalisés de réussite éducative (PPRE). Enfin, les évaluations à l'entrée en CE2 sont arrêtées en 2006-2007.

Concernant les évaluations bilans, un nouveau cycle émerge en 2003 pour comparer à échéance régulière les compétences acquises en fin d'école primaire et en fin de collège dans plusieurs domaines. Ce cycle d'évaluations disciplinaires réalisées sur échantillon (CEDRE) s'inspire beaucoup plus que lors des périodes précédentes de la méthodologie des comparaisons internationales et des acquis de la psychométrie (usage des modèles de réponse à l'item, construction d'échelles de performance, technique des cahiers tournants etc.) et se poursuit à ce jour (9). Néanmoins, ce nouveau cycle ne permet pas de renseigner les indicateurs retenus en parallèle dans la loi organique relative aux lois de finances (LOLF) concernant le degré de maîtrise des compétences de base par les élèves, si bien que la DEPP a dû concevoir d'autres évaluations sur échantillon en fin de CM2 et de 3^e en français et en mathématiques. Expérimentées en 2005 et 2006, elles ont été généralisées en 2007, au départ à titre provisoire.

(8) 56,8 % des professeurs de lettres en classe de 6^e déclarent que les évaluations diagnostiques font évoluer leurs pratiques évaluatives soit « un peu » (33,5 %), soit « assez » (17,9 %), soit « beaucoup » (5,4 %), contre 54,2 % pour les professeurs de mathématiques, les pourcentages composant ce résultat global étant respectivement de 36,2 %, 15,6 % et de 2,4 % (DEPP, 2005:77).

(9) Pour une présentation de la méthodologie au fondement de ces évaluations CEDRE, voir la *Note d'information* n° 10-22 de la DEPP publiée en décembre 2010.

Rationaliser l'outil ou l'instrumentaliser ? (2008-2012)

Le 13 novembre 2007, dans un document d'orientation pour la réforme de l'école primaire, le ministre de l'Éducation nationale Xavier Darcos (2007-2009) envisage la création de deux évaluations nationales témoins destinées à mesurer les acquis des élèves de CE1 et de CM2 (Pons, 2010 b). Cette annonce constitue une nouvelle inflexion politique majeure. D'une part, elle sonne définitivement le glas de l'ancien dispositif dit des « évaluations CE2-6^e ». Si les premières sont arrêtées dès l'année scolaire 2006-2007, les secondes étaient encore en vigueur et sont désormais laissées à la libre mise en œuvre des rectorats à partir de la rentrée 2008. D'autre part, elle débouche sur une évaluation standardisée d'un nouveau type qui s'éloigne des canons statistiques et psychométriques habituels (dans sa première version en tout cas) et dont la mise en œuvre est confiée non pas à la DEPP mais à la Direction générale de l'enseignement scolaire (DGESCO) (10). Enfin, de façon inédite, ces évaluations suscitent une polémique qui a cours tout au long de la période et qui en constitue indéniablement l'originalité.

Initialement, les évaluations sont supposées pallier plusieurs manques de l'organisation précédente comme en témoigne la pluralité des finalités confiées à ces évaluations qui doivent « *mesurer les acquis des élèves français à des moments clefs de leur scolarité* », « *donner aux parents toutes les informations auxquelles ils ont droit, en évitant une mise en concurrence incontrôlée des écoles* », constituer « *un nouvel indice de l'efficacité des enseignements à l'école* » et fournir « *un instrument de pilotage du système éducatif, du niveau local de l'école jusqu'au niveau national* » (extraits du discours de X. Darcos du 2 octobre 2008). Deux arguments principaux sont avancés par le gouvernement pour justifier ce changement. Les évaluations CE1-CM2 sont décrites comme le prolongement logique des nouveaux programmes d'enseignement primaire présentés le 20 février 2008 par le ministre, un moyen de mesurer leur efficacité dès leur première année d'existence. Elles constitueraient, en outre, un complément indispensable aux évaluations internationales pour mesurer plus finement la baisse de performance observée dans plusieurs de ces comparaisons, mais aussi pour ne pas dépendre uniquement des comparaisons internationales pour évaluer l'enseignement français. Deux autres facteurs d'ordre structurels peuvent être évoqués. Le premier est l'état du système d'information statistique sur l'enseignement primaire. La difficulté à faire remonter des données depuis les inspections académiques,

la grève administrative, entamée en 1999, des directeurs d'école, qui refusent de transmettre les enquêtes de rentrée ou les contestations suscitées par la nouvelle « Base élèves » à partir d'octobre 2006, affectent sensiblement la qualité et la diversité des informations statistiques disponibles sur l'enseignement primaire, ce qui rend difficile, entre autres, le renseignement des indicateurs de la LOLF. Le deuxième facteur structurel est le hiatus, depuis les années 1970, entre d'un côté, des évaluations bilans qui aident au pilotage mais qui sont fondées sur des échantillons nationaux et qui ne permettent pas un pilotage territorial du système éducatif par les autorités académiques et, de l'autre, des évaluations diagnostiques qui, elles, sont systématiques et permettent donc des exploitations à différentes échelles (classes, établissements, circonscriptions, académies) utiles aux pilotes intermédiaires (recteurs, inspecteurs d'académie et inspecteurs de l'enseignement primaire en charge d'une circonscription), qui ne sont pas conçues dans cette perspective de pilotage mais pour repérer les difficultés des élèves. Dans la mesure où elles sont exhaustives, on attend ainsi des nouvelles évaluations – initialement administrées en janvier (CM2) et en mai (CE1) – qu'elles permettent de calculer la médiane des résultats des élèves à différents niveaux de pilotage (Pons, 2010 b).

Néanmoins, dès le lendemain de leur lancement, les syndicats d'enseignants du primaire, un syndicat d'inspecteurs et la Fédération des conseils de parents d'élèves (FCPE) demandent, sans succès, soit leur amélioration, soit leur report, soit leur abrogation. La polémique est particulièrement vive (soixante-deux protestations ou réserves officielles sont recensées par l'AEF entre le 3 octobre 2008 et le 11 mai 2012 inclus) et s'étend jusqu'aux élections présidentielles de 2012. Trois temps forts peuvent être mis en évidence :

- Le premier temps fort est celui de la critique méthodologique (octobre 2008-mai 2011). Les protestations suscitées par les évaluations CE1-CM2 concernent en effet moins leurs finalités que leur mode d'élaboration (discret et effectué dans l'urgence par des conseillers proches du cabinet du ministre) et leur méthodologie globale (date de passation, logique binaire de correction, manque de comparabilité dans le temps, questions posées sur des aspects à peine envisagés dans les programmes, hésitation des évaluations CM2 entre deux types d'évaluation, bilan ou diagnostique, etc.). Sur ce plan, le soutien de la Fédération des parents d'élèves de l'enseignement public (PEEP) ainsi que de certains recteurs intervenant publiquement pour défendre l'outil, le travail d'exploitation des résultats

(10) Créée en mai 2006, la DGESCO est responsable de trois des six programmes LOLF de la mission « enseignement scolaire » ; elle est souvent perçue comme une administration plus politique que technique ou scientifique dans la mesure où elle est chargée d'élaborer la politique éducative et pédagogique du ministère.

Tableau 2

Les configurations historiques des évaluations standardisées des acquis des élèves

| Périodes | Dimension politique | Dimension cognitive | Dimension professionnelle |
|-----------|--|---|--|
| 1973-1988 | Intérêt inégal de la part des décideurs. | Inégalité et instabilité des évaluations. | Conflit au niveau central entre statisticiens et inspecteurs. Investissement de l'outil par différents groupes professionnels. |
| 1989-1997 | Mesures de Lionel Jospin. Théorisation par l'effet miroir de Claude Thélot. | Dispositif « des évaluations CE2-6 ^e ». Pratique irrégulière des évaluations bilans. | Réticences de départ des syndicats d'enseignant. Rhétorique du développement nécessaire d'une « culture d'évaluation ». |
| 1997-2007 | Rupture Claude Allègre. Captation de l'outil selon des besoins politiques multiples. | Dispositif « des évaluations CE2-6 ^e ». Pratique irrégulière des évaluations bilans. Inflation de nouveaux outils et retraits multiples. | Usages ambivalents des évaluations diagnostiques à tous les niveaux du système éducatif. |
| 2007-2012 | Polémique des évaluations CE1-CM2. Instrumentalisation de l'outil et « du » gouvernement par la contrainte institutionnelle. | Accumulation de cinq dispositifs différents (évaluations CE1-CM2 de la DGESCO, évaluations CM2-3 ^e de la DEPP sur échantillon, indicateurs CM2-3 ^e pour la LOLF, évaluations bilans ponctuelles de la DEPP et évaluation annoncée en fin de 5 ^e). | Conflit corporatiste classique entre le ministère et les professionnels. |

Source : tableau réalisé par l'auteur.

DGESCO : Direction générale de l'enseignement scolaire ; DEPP : Direction de l'évaluation, de la prospective et de la performance ; LOLF : loi organique relative aux lois de finances.

du débat sur les évaluations CE1-CM2, qu'une nouvelle évaluation nationale à la fin de la classe de 5^e soit expérimentée sur le modèle de ces dernières en mai 2012 dans des établissements volontaires. Cette évaluation doit permettre à la fois d'affiner les besoins de chaque élève et de disposer d'un instrument de pilotage pédagogique à toutes les échelles du système éducatif. Autre exemple : dans son rapport consacré aux indicateurs relatifs aux acquis des élèves de septembre 2011, le Haut Conseil de l'éducation (HCE) estime que les indicateurs utilisés dans la LOLF pour évaluer le degré de maîtrise des différents paliers (fin des classes de CE1, de CM2 et de 3^e) du socle commun de connaissances instauré en 2006 ne sont pas fiables sur le plan méthodologique,

entamé dans certaines académies par les corps d'inspection, la formulation ponctuelle par le ministre de nouvelles propositions destinées à amender le dispositif (comme l'octroi d'une prime aux enseignants mobilisés par les évaluations) et l'usage récurrent de ces évaluations dans la communication gouvernementale (pour inciter les enseignants à adhérer aux derniers dispositifs mis en œuvre par le gouvernement, pour souligner les progrès de certaines catégories d'élèves ou pour mettre en évidence les effets positifs des programmes de 2008) ne suffisent pas à infléchir le cours du débat public. Ce processus débouche en mai 2011 sur la décision du ministère d'administrer les évaluations en CM2 en fin d'année, donc d'en faire clairement des évaluations bilans sommatives.

- Le deuxième temps fort se caractérise par l'agrégation, autour de ce débat sur les évaluations CE1-CM2, de considérations relatives à d'autres outils d'évaluation qui suscitent à leur tour plusieurs critiques (mai 2011-décembre 2011). La circulaire de rentrée 2011 prévoit ainsi, à rebours de l'évolution

que ceux-ci émanent de l'agrégation des attestations de maîtrise des compétences du socle déclarées par les enseignants (pour le CM2 et la 3^e) ou de l'évaluation conduite en fin de CE1 (11).

- Le troisième temps (janvier-mai 2012), moment fort de la campagne présidentielle, se caractérise par une multiplicité d'annonces contradictoires de la part des futurs gouvernants potentiels et des acteurs du système éducatif. La majorité en place confirme ses intentions et précise, dans la circulaire de rentrée 2012, envoyée dans les établissements scolaires avant le premier tour de l'élection, les modalités de la future évaluation en fin de 5^e tandis que les porte-parole en éducation de François Hollande annoncent rapidement, après le second tour, l'intention du nouveau président de suspendre les évaluations CE1-CM2. Les deux principaux syndicats d'enseignants, quant à eux, plaident pour un retour à des évaluations diagnostiques au service du développement professionnel des enseignants sur le modèle des évaluations précédentes à l'entrée en CE2.

(11) Parmi les critiques énoncées à l'ensemble des outils de mesure disponibles, le HCE évoque, tour à tour, l'inadéquation des niveaux d'exigence attendus (moindres dans l'évaluation que dans les textes fondateurs du socle commun), la confusion entre évaluation diagnostique et évaluation bilan ou encore une couverture du champ des compétences à évaluer insuffisante.



Conclusion : complexification et politisation

Au final, loin d'être linéaire, la carrière des évaluations standardisées des acquis des élèves mises en œuvre par le ministère de l'Éducation nationale depuis les années 1970 se révèle particulièrement instable, montrant, en creux, les hésitations récurrentes en France en matière de régulation du système éducatif. Néanmoins, deux tendances fortes émergent de la synthèse présentée dans le tableau 2 des configurations historiques dans lesquelles cet outil a pu être mis en œuvre. La première est la complexification de l'outil. Cette complexification concerne aussi bien la méthodologie des évaluations, qui s'est globalement

affinée au cours du temps malgré quelques inflexions et retours en arrière (comme pour les évaluations CE1-CM2 récentes), que leur architecture institutionnelle d'ensemble qui, loin d'avoir connu un mouvement de rationalisation technique, s'est plutôt traduite par un empilement de dispositifs aux finalités multiples, souvent pensés de manière cloisonnée en fonction des nombreux dossiers instruits isolément par les gouvernements successifs. La deuxième tendance est la politisation progressive de l'outil. Celle-ci est visible dans l'intérêt croissant que les décideurs ont manifesté à l'égard de cet outil au cours de la période, au point d'en faire un outil de communication clé sous les ministères de Luc Chatel (2009-2012), mais aussi dans la transformation des rapports entre les groupes professionnels concernés par ces évaluations qui culmine avec un conflit corporatiste aussi long que classique.

Références bibliographiques

- Ballion R., 1982, *Les consommateurs d'école*, Paris, Stock.
- Barrère A., 2009, *Les directions d'établissement scolaire à l'épreuve de l'évaluation locale*, *Carrefours de l'éducation*, n° 28:199-214.
- Demailly L., 2003, *L'évaluation de l'action éducative comme apprentissage et négociation*, *Revue française de pédagogie*, n° 142:114-127.
- DEPP (Direction de l'évaluation, de la prospective et de la performance), 2005, *Les pratiques d'évaluation des enseignants au collège*, *Éducation et formations. Les dossiers*, n° 160.
- DEPP, 1994, *L'opinion des enseignants sur l'évaluation dans l'enseignement primaire*, *Note d'information*, n° 94-14.
- Derouet J.-L. et Normand R., 2003, *Le développement d'une culture de l'évaluation dans l'Éducation Nationale : comment les enseignants utilisent-ils les résultats des évaluations nationales ?*, Lyon, INRP.
- Dutercq Y., 2000, *Politiques éducatives et évaluations : querelles de territoires*, Paris, PUF.
- Hood C., 1996, *Exploring variations in public management reform in the 1980's*, in *Civil service systems in comparative perspective* [Bekke H. A., Perry J. L. and Toonen T. A. (ed.)], Indiana University Press, 1996.
- Hood C., 1991, *A public management for all seasons?*, *Public administration*, n° 69-1:3-19.
- IGEN-IGAENR (Inspection générale de l'Éducation nationale-Inspection générale de l'administration de l'Éducation nationale et de la Recherche), 2005 a, *L'usage des outils de pilotage élaborés par les académies ou mis à leur disposition*, Paris, ministère de l'Éducation nationale.
- IGEN-IGAENR (Inspection générale de l'Éducation nationale-Inspection générale de l'administration de l'Éducation nationale et de la Recherche), 2005 b, *Les acquis des élèves, pierre de touche de la valeur de l'école ?*, Paris, ministère de l'Éducation nationale.
- Lascoumes P. et Le Galès P., 2004, *Gouverner par les instruments*, Paris, Presses de Sciences Po.
- Levasseur J., 1996, *L'évaluation nationale des acquis des élèves*, *Revue internationale d'éducation*, n° 11:101-114.
- Maroy C., 2006, *École, régulation, marché*, Paris, PUF.
- Mons N. et Pons X., 2006, *Les standards en éducation dans le monde francophone. Une analyse comparative*, Neuchâtel, IRDP.
- Normand R., 2005, *La mesure de l'école : de la tradition statistique à la modernité économétrique*, *Éducation et Sociétés*, n° 16:209-226.
- Pons X., 2011, *L'évaluation des politiques éducatives*, Paris, PUF.
- Pons X., 2010 a, *Évaluer l'action éducative. Des professionnels en concurrence*, Paris, PUF.
- Pons X., 2010 b, *L'urgence des évaluations CE1-CM2. Réflexion sur une méthode de gouvernement*, *Administration et éducation*, 125-1:15-21.
- Thélot C., 1994, *L'évaluation du système éducatif français*, *Revue française de pédagogie*, n° 107:5-28.
- Van Zanten A., 2011, *Les politiques d'éducation*, Paris, PUF.
- Van Zanten A., 2001, *Le rôle des évaluations dans les stratégies concurrentielles des établissements et dans les stratégies de choix des parents en France et en Grande-Bretagne*, in *L'évaluation des politiques éducatives* (sous la dir. de Demailly L.), Bruxelles, De Boeck.
- Van Zanten A., 2000, *L'école, l'état des savoirs*, Paris, La Découverte.
- Verdrière J., 2001, « Les pratiques d'évaluation du travail d'enseignement », thèse de doctorat de sociologie, sous la direction de Lise Demailly, université Lille I.
- Vogler J. (dir.), 1996, *L'évaluation*, Paris, Hachette.