

LES ÉVALUATIONS EN 6^e : LES DIFFICULTÉS DE TRADUCTION D'UNE POLITIQUE NATIONALE

Dominique CHAZAL & Romuald NORMAND
UMR Education et Politique
Institut National de Recherche Pédagogique

Dans un système éducatif déconcentré et progressivement décentralisé, les évaluations nationales poursuivent un double but : éclairer le pilotage national et aider les acteurs et spécialement les enseignants à réajuster leurs pratiques. Si le premier objectif est généralement considéré comme atteint, la situation est beaucoup plus incertaine sur le second point. Le but de notre recherche était de faire le point sur cette question en étudiant l'utilisation des évaluations des connaissances des élèves qui sont menées depuis douze ans à deux niveaux du système éducatif : le CE2 et la sixième. Les enseignants mais aussi les chefs d'établissement et les inspecteurs (IEN, IA, IPR) utilisent-ils les résultats des évaluations nationales pour réajuster leurs actions locales ? Quels sont les effets de cet usage sur les élèves ?

Bien que les évaluations nationales soient le plus souvent mises en rapport avec la pédagogie et les apprentissages, parce qu'elles sont conçues pour mesurer les acquis des élèves et leur performance dans la maîtrise de certaines compétences ou connaissances, on peut considérer que leur mise en oeuvre génère également des effets importants sur les politiques éducatives, les contenus d'enseignement et les pratiques des enseignants. En effet, ces évaluations nationales influencent la nature des programmes, la formation initiale ou continue des enseignants ainsi que les procédures et les systèmes d'évaluation au niveau d'une académie, d'un établissement scolaire ou d'une classe.

A partir de cette considération des effets des évaluations nationales sur le système éducatif dans son ensemble, deux questions majeures ont guidé notre recherche : comment le système éducatif répond à l'introduction de ces évaluations

nationales en CE2 et 6^e dans le domaine des mathématiques et du français ? Quelles sont les conséquences sur les pratiques enseignantes ?

Nous avons cherché à poser une série de questions permettant d'identifier les formes de traduction de ces évaluations nationales depuis leur conception au sein du ministère de l'éducation nationale jusqu'à leur mise en œuvre par les enseignants dans leur classe. En effet, les évaluations nationales CE2/6^e en mathématiques et français participent normalement à certains changements dans la manière dont ces disciplines ont été jusqu'à présent enseignées. Mais leur impact est largement indirect et doit être apprécié au travers d'effets induits par d'autres logiques présentes au sein du système éducatif, au niveau local comme au niveau national. En effet, il existe différentes chaînes de traduction par lesquelles ces évaluations nationales influencent les pratiques enseignantes : elles sont complexes, en interaction et différentes selon les champs disciplinaires. De plus, les enseignants ont des expériences différentes de ces évaluations nationales en fonction du contexte, des ressources mises à leur disposition, et de leurs propres connaissances de l'évaluation. Cette recherche n'est pas terminée et nous réservons la publication de nos résultats à notre commanditaire. Toutefois, nous donnerons à la fin de l'article quelques éléments sur la réception des évaluations nationales par les enseignants de mathématiques au collège pour lesquelles nous disposons aujourd'hui de données significatives.

LES ÉVALUATIONS NATIONALES : UN ENJEU POLITIQUE POUR L'ÉCOLE

L'évaluation du système éducatif est opérée par la Direction de la Programmation et du Développement (DPD) qui produit régulièrement des données statistiques et les communique à travers différentes publications (l'Etat de l'Ecole, Géographie de l'Ecole, Notes d'information, etc.). Elle vise à répondre à une exigence externe (rendre compte à la Nation de l'état de son école) et interne (promouvoir une culture de l'évaluation parmi les acteurs de l'Education Nationale). Dans ce cadre statistique, les évaluations nationales sont conçues comme des « évaluations de masse » dont la fonction principale est de fournir des outils aux enseignants et aux établissements. Pour cela, sont élaborés des protocoles au sein d'un groupe de pilotage présidé par la DPD et composé de représentants du ministère, des corps d'inspection, d'enseignants et de chercheurs. Ceux-ci définissent les objectifs à évaluer après une lecture approfondie des programmes et des instructions officielles ainsi que les exercices (items) permettant de mesurer les capacités des élèves. Avant d'analyser plus en détail comment ces évaluations nationales sont mises en œuvre sur le territoire national, nous présenterons leur histoire et les conceptions qui s'y rattachent.

Brève histoire des évaluations nationales

La DPD anciennement dénommée DEP (Direction de l'Evaluation et de la Prospective) a été créée en 1986 et répondait au souhait du ministère de l'Education

Nationale de faire évoluer son système de statistique administrative vers un système d'évaluation avec de nouveaux objectifs. Cette évolution a commencé avec la réforme Haby de 1975 où l'on a cherché à suivre un panel d'élèves pour mesurer les effets du collège unique, notamment pour vérifier le lien entre la réussite scolaire et l'origine sociale des élèves. Le travail a été confié à des administrateurs de l'INSEE (Alain Darbel puis Claude Seibel) qui, tout en connaissant bien les travaux de Pierre Bourdieu, ont cherché à rendre compte de l'inégalité des chances à l'intérieur du système éducatif (Derouet, 2000). A partir de 1981, les débuts de la décentralisation accroissant l'autonomie des établissements scolaires mais aussi le développement du consumérisme scolaire et la publication de Palmarès concernant le baccalauréat dans la presse ont fait craindre un éclatement du service public d'éducation et ont rendu urgente une réflexion sur l'évaluation du système éducatif. A la même époque, le ministère s'est préoccupé d'élaborer des indicateurs de pilotage lui permettant de conserver une certaine maîtrise sur le système décentralisé et sur l'évaluation des inégalités entre établissements. A cela sont venues s'ajouter des préoccupations concernant le rendement des investissements éducatifs dans un contexte de réduction des dépenses publiques et une inquiétude sur la baisse du niveau de l'enseignement perçue comme une conséquence de la démocratisation scolaire. Afin de répondre à ces interrogations, le ministère a cherché à faire de l'évaluation un outil d'auto-régulation pour les acteurs du système éducatif notamment par la création d'un système IPES (Indicateurs de Pilotage des Etablissements du Second Degré). La loi d'orientation de 1989 a donné une légitimité supplémentaire à cette régulation par l'évaluation en créant des grandes évaluations nationales aux niveaux clés du système éducatif : CM2, 6^e et seconde.

L'évaluation conduite par la DEP a été conçue à l'origine selon un triptyque « coût, fonctionnement, résultats » accompagné d'indicateurs permettant de construire, organiser et comparer les données recueillies (Thélot, 1993). Les indicateurs de résultat visaient principalement la transmission des connaissances, les indicateurs de fonctionnement devaient concerner l'activité éducative, les indicateurs de coût mesurant le niveau de la dépense en éducation. L'évaluation du système éducatif par les résultats scolaires, c'est-à-dire sur la base des connaissances acquises par les élèves (telles que les mesurent les évaluations nationales), obéit à une longue tradition qui nous vient des Etats-Unis, même si les modalités pratiques et techniques en sont différentes. En 1983, fut rendu public le rapport *A Nation at Risk*, de la *National Commission of Excellence in Education* qui faisait état d'un certain nombre de préoccupations concernant le système éducatif aux Etats-Unis dans sa relation à la mondialisation. Le rapport proclamait que la nation américaine était en danger citant l'augmentation des taux d'abandon en cours de scolarité, les difficultés de lecture des élèves les plus jeunes, et l'inaptitude des diplômés à occuper les emplois. Le système éducatif était ainsi considéré comme incapable de former des diplômés de haut niveau ce qui menaçait à terme la compétitivité de l'économie américaine.

Les Etats-Unis ont par la suite mis en œuvre une réforme radicale de leur système éducatif à la suite de la publication du rapport *A Nation at Risk*. Ce dernier utilisait énormément d'indicateurs dans ses conclusions tout en effectuant des comparaisons internationales. Il eut pour effet d'attirer le regard de la communauté

internationale sur les pratiques d'éducation aux Etats-Unis mais aussi de révéler la variété de l'information statistique pouvant être utilisée pour évaluer les systèmes éducatifs. Un an plus tard, le Secrétariat de l'Education américain publiait sa première série de *league tables* permettant de comparer la performance des Etats dans le domaine de l'éducation, ce qui souleva de vives réactions dans l'opinion publique et le monde politique américain habitués à une grande autonomie locale. Malgré cela, le mouvement s'étendit jusqu'à la mise en œuvre d'évaluations nationales dans le cadre du *National Assessment of Educational Progress* (l'équivalent de nos évaluations nationales). Celles-ci furent conçues pour établir des comparaisons entre les résultats scolaires des élèves américains au niveau fédéral. Compte tenu des efforts réalisés en ce domaine, le gouvernement américain exerça par la suite une forte pression auprès de l'OCDE (Organisation de Coopération et de Développement Economique) pour obtenir que l'efficacité des systèmes éducatifs devienne un thème prioritaire dans la réflexion et l'activité de cet organisme international qui joue un rôle important dans l'orientation des politiques éducatives des pays industrialisés. Les responsables américains souhaitaient que l'on mette en œuvre une statistique s'inspirant de l'expérience du *National Assessment of Educational Progress* aux Etats-Unis, un système où l'on mettait au point des indicateurs permettant la comparaison entre Etats. Cette pression était liée également au développement des comparaisons internationales dans un climat marqué par la concurrence, où *l'Association Internationale pour l'Evaluation du rendement scolaire* (IEA) jouait un grand rôle. A partir de là, l'OCDE a procédé à la mise en place d'indicateurs de résultats. Les pays membres ont décidé d'échanger quelques indicateurs-clés utilisés dans un cadre national à partir desquels ont travaillé quelques experts. En 1992, l'organisme publiait son premier ensemble d'indicateurs de l'enseignement destiné à orienter les politiques éducatives des pays membres et à fournir une base de comparaisons internationales dans le sens d'une plus grande qualité et efficacité des systèmes éducatifs. En France, les évaluations nationales ont été introduites à la même époque selon une architecture sensiblement similaire et depuis, chaque année, la DPD comme l'OCDE publie une étude comparée des résultats scolaires des élèves sur la base de l'évaluation de leurs « compétences de base ».

De la sphère de décision à la classe : une chaîne complexe de traduction

Comment les évaluations nationales sont-elles accueillies et interprétées par les acteurs du système éducatif (enseignants, responsables de l'encadrement, inspecteurs) ? Si ces évaluations sont conçues en fonction de certains objectifs, notamment pour aider les enseignants dans leurs pratiques pédagogiques, elles sont mises en œuvre dans des cadres variés donnant lieu à de multiples interprétations. Les évaluations nationales peuvent induire des changements ou être tout simplement ignorées. Nous avons donc cherché à suivre la manière dont les corps d'inspection, les formateurs, les évaluateurs et d'autres comprennent ces évaluations et s'en font ou non les porte-parole. Il nous paraissait important également de comprendre comment sont modifiées les pratiques des enseignants, en étant attentifs aux effets des évaluations nationales sur la pédagogie dans la classe et la façon dont les outils

et les documents d'accompagnement étaient mobilisés dans les situations d'apprentissage.

En suivant les chaînes de traduction des évaluations nationales nous avons été amenés à prendre en compte trois grands domaines : le curriculum, la formation des enseignants, l'évaluation. En effet, la mise en œuvre des évaluations nationales est dépendante d'un certain nombre de décisions et conditions ayant trait à la discipline enseignée (mathématiques ou français) à différents niveaux du système éducatif. Cela concerne au premier chef les programmes d'enseignement mais aussi la configuration des manuels scolaires, les orientations de la didactique, les choix et les priorités des enseignants dans les contenus enseignés. De même, les évaluations nationales peuvent avoir un effet en retour sur la manière dont est défini et structuré le curriculum selon qu'il accorde une place plus ou moins importante à l'évaluation. Parallèlement, des références aux évaluations nationales dans la formation initiale ou continue des enseignants peuvent avoir des effets sur la manière dont ils conçoivent l'évaluation dans leurs pratiques pédagogiques mais aussi dans la fabrication des compétences professionnelles qu'ils seront à même de développer. C'est pourquoi nous nous sommes intéressés à la nature et aux dispositifs de formation mis en œuvre autour des évaluations nationales (à l'IUFM ou au sein d'un Plan Académique de Formation) afin de comprendre comment étaient ensuite reproblématisés un certain nombre de concepts et de techniques opératoires. Enfin, il existe de fait une pluralité des formes d'évaluation servant à informer une diversité d'acteurs (élèves, parents d'élèves, enseignants, responsables du système éducatif) ce qui oblige l'enseignant à mobiliser des outils et des dispositifs hétérogènes. C'est la raison pour laquelle nous avons jugé utile d'étudier la manière dont il parvenait à réduire cette complexité en construisant son propre système d'évaluation dans la classe. Nous nous sommes intéressés au type de décisions qu'il était amené à prendre dans le choix des outils d'évaluation, dans la conduite de sa classe, dans sa communication avec les élèves ou avec des tiers et à la nature de ces attentes en termes de connaissances et résultats à atteindre. Pour traiter ces questions majeures, nous avons considéré qu'il fallait ouvrir une « boîte noire » entre d'un côté la conception des évaluations nationales au niveau national et de l'autre son impact sur les pratiques enseignantes.

Toutefois, il était impossible pour notre équipe de recherche de suivre l'ensemble de la chaîne de traduction des évaluations nationales depuis le centre (au niveau du ministère) jusqu'à la périphérie (au niveau des dispositifs observés dans les classes). C'est pourquoi nous avons choisi de donner quelques coups de projecteur sur certains segments de la chaîne tout en travaillant sur l'articulation entre les différents niveaux. Chaque niveau peut être appréhendé comme un espace d'intéressement où différents acteurs en partant de logiques différentes visent à élaborer un dispositif, une politique, ou un plan d'action conduisant à la mise en œuvre ou au rejet des évaluations nationales. Le premier niveau est constitué par la sphère de décision et de conception des évaluations nationales, c'est-à-dire au sein du ministère et des groupes de travail chargés de la construction des items et des outils. Le deuxième niveau intermédiaire est celui de la traduction de ces évaluations nationales et des directives ministérielles qui les accompagnent au niveau d'une académie (enseignement secondaire) ou d'un département (enseignement primaire).

Sont concernés les dispositifs d'accompagnement ou de pilotage visant à sensibiliser les enseignants à ces évaluations. Le troisième niveau est celui de la traduction de ces évaluations nationales au sein des établissements scolaires et dans les classes où ont été conduits le suivi et l'observation d'enseignants. Cette traduction ne va pas de soi comme le montrent nos premiers constats : du fait de la perte d'information en ligne liée à la multiplication des niveaux intermédiaires mais aussi parce que cette information et les instruments d'évaluation prennent sens en fonction du contexte jugé pertinent par les acteurs. De plus, ces instruments portent en eux-mêmes un certain nombre de tensions internes résultant des compromis entre des logiques qu'ils sont censés incarner, logiques relevant de sphères politique, économique et scientifique très différentes.

Une première tension que l'on peut qualifier d'« épistémologique » est liée au fait que les évaluations nationales tendent à concilier deux espaces de référence de l'évaluation dans le domaine de la recherche en éducation. La première conception se rattache au modèle de la pédagogie de la maîtrise et à la taxinomie des objectifs pédagogiques de Bloom : elle s'inscrit dans une conception néo-béhavioriste de l'apprentissage assez proche du *mental testing* très développée dans le monde anglo-saxon. Elle vise à faire de l'évaluation un outil de mesure des résultats ou de l'atteinte d'objectifs préalablement fixés. La deuxième conception est d'inspiration plus piagétienne. Elle a été développée à Genève autour du groupe Rhapsodie et de Linda Allal. Elle conçoit l'évaluation comme un outil d'aide à l'apprentissage et à l'accompagnement pédagogique des élèves en difficultés. Ces deux conceptions se retrouvent dans les finalités assignées aux évaluations nationales. En effet, elles visent à mesurer les acquis des élèves afin que l'enseignant puisse établir une classification des erreurs et y remédier en se fixant des objectifs pédagogiques à atteindre. Mais elles sont également conçues comme un outil de médiation vis-à-vis des difficultés des élèves. Toutefois, leur fort équipement cognitif et instrumental ne leur permet pas d'avoir la souplesse et la flexibilité nécessaire pour remplir cette deuxième fonction. La complexité du dispositif et l'algorithmique dans l'élaboration de profils d'élèves s'opposent de fait au travail réflexif et rétroactif que nécessiterait la seconde forme d'évaluation appelée généralement « formative ».

Une deuxième tension que l'on qualifiera de « politique » est liée aux différents objectifs politiques que visent à remplir ces évaluations nationales. On pourrait dire que se côtoient un modèle anglo-saxon et un modèle français. Le premier est inspiré des grandes comparaisons internationales de l'OCDE qui, comme on l'a vu précédemment, ont incité les pays industrialisés à développer leur système d'évaluation nationale des acquis des élèves. Les items ont donc été construits à des fins de statistique publique et de comparaison des systèmes éducatifs. Selon cette orientation, l'évaluation y est conçue comme une fonction de production qu'il faut chercher à optimiser en termes de coût-efficacité (sous l'influence de la théorie du capital humain). L'autre modèle est celui de la statistique française. Cette statistique a accompagné le développement de l'Etat centralisé en visant dans un premier temps à mesurer l'inégalité des chances à l'école, avant de mesurer l'inégalité des acquis. S'il a subi l'influence de l'OCDE, ce modèle s'inscrit dans une visée d'égalité de traitement des élèves (qui accompagne la massification) et se rattache fortement à l'élaboration de programmes nationaux visant une démocratisation par les savoirs.

Cette tension politique est également perceptible dans deux orientations du ministère qui ont fait jouer aux évaluations nationales des finalités différentes. Selon la première orientation, on a cherché à constituer un instrument de pilotage de l'action publique en éducation dans la tradition de l'INSEE, la publication des résultats devant éclairer l'opinion publique et les décideurs. Dans le second cas, on a souhaité promouvoir une certaine diversification pédagogique rendue nécessaire par la démocratisation scolaire tout en maintenant les exigences en matière de contenus des programmes d'enseignement.

Une troisième tension de nature « pédagogique » sera évoquée dans la deuxième partie consacrée à l'usage des évaluations nationales par des enseignants de mathématiques en 6^e.

L'USAGE DES ÉVALUATIONS DANS LA CLASSE : PREMIERS CONSTATS

Dans cette partie, nous nous intéresserons à la critique formulée par les enseignants de mathématiques concernant le dispositif des évaluations nationales. Notre but n'était pas de récupérer une dénonciation facile mais d'essayer de resituer ces éléments de critique en fonction du contexte et de l'expérience des enseignants afin de comprendre leurs points de vue et leurs motivations pour refuser d'utiliser ces évaluations à des fins pédagogiques.

Les raisons d'un usage limité dans la pédagogie

Tout d'abord, il apparaît que ces évaluations nationales sont perçues comme étrangères à la pédagogie. En effet, les enseignants pensent qu'elles sont essentiellement destinées à réaliser des statistiques pour faire des comparaisons en matière d'efficacité du système éducatif avec d'autres pays ou pour opérer des classements d'établissements ou encore pour que les responsables du ministère se fassent une idée du niveau des sixièmes. Certains jugent que ces évaluations sont une perte de temps. Pour les enseignants interrogés, les évaluations ne sont pas conçues dans le cadre d'un accompagnement pédagogique. Elles ne présentent pas d'intérêt direct pour le cours car les erreurs-type des élèves leur sont déjà connues et se répètent d'année en année (notamment celles concernant la confusion entre parallèles et perpendiculaires). Les évaluations ne semblent rien dire sur le cheminement de l'élève et ce qui va le conduire à l'erreur. Dans le meilleur des cas, elles servent à déterminer les effectifs qui n'ont pas réussi à tel item pour juger de manière très générale d'une éventuelle remédiation concernant certains groupes d'élèves. De manière assez unanime, les enseignants reprochent aux évaluations d'être un frein à un réel démarrage de l'activité avec les élèves. En effet, pour la plupart d'entre eux, la planification du programme est une contrainte très précise parce que les enseignants veulent dispenser un enseignement égalitaire à tous les élèves et atteindre les objectifs qu'ils se sont fixés pour l'année scolaire.

Parallèlement, nous avons pu observer l'absence de dispositifs clairs au niveau des établissements encadrant ces évaluations nationales. Généralement, la passation

des évaluations, qui se déroule selon une logique toute administrative, a lieu lorsque les enseignants ont déjà commencé leur programme. Ainsi, en mathématiques, les enseignants qui commencent par l'écriture des nombres ou par les éléments de géométrie pensent que les résultats sur les items de numération sont faussés. De plus, les dates de passation peuvent être variables d'un établissement à un autre en fonction du contexte. La saisie des codes est généralement réalisée par les enseignants mais nous avons aussi vu des saisies réalisées par des employé-jeunes ou par les responsables de l'administration. Dans les établissements que nous avons observés, il ne semble pas y avoir de concertation préalable entre les enseignants sur le codage ou sur le contenu de l'évaluation. Chaque enseignant procède à un codage de manière très automatique, un travail jugé long et fastidieux qui semble faire perdre beaucoup d'intérêt à l'évaluation. Les enseignants passent ainsi beaucoup de temps au codage car les consignes sont strictes mais ensuite ils ne se donnent plus le temps nécessaire à un travail réflexif sur le codage et sur l'évaluation.

La communication des résultats aux familles varie d'un établissement à un autre, d'un enseignant à un autre. Certains établissements convoquent l'ensemble des familles en donnant le résultat de l'enfant, la moyenne de la classe et celle de l'établissement. D'autres enseignants refusent de donner un caractère institutionnel aux évaluations et vont convoquer uniquement les parents dont les enfants ont un résultat inférieur à un certain seuil. Ce seuil est déterminé selon des critères très variables, qui peuvent être définis par une concertation entre collègues (par exemple on convoque les parents pour les élèves ayant un score inférieur à 50% ou on limite volontairement le seuil pour ne rencontrer que quelques parents), mais aussi en fonction de la propre perception de l'enseignant, de ce qu'il juge difficile dans les évaluations nationales. Ce seuil est aussi fonction du niveau des classes : dans certains cas il faudrait convoquer tous les parents, dans d'autres seulement un ou deux. D'une manière générale, les réunions avec les parents d'élèves ont lieu un mois après la passation. Entre temps les enseignants ont mis en place leur propre système d'évaluation, les évaluations 6^{ème} n'étant plus une référence. Parfois, celles-ci sont perçues négativement comme un instrument de classement prématuré des élèves avant que les liens soient tissés entre l'enseignant et sa classe. Enfin, les possibilités de remédiation ou de soutien faisant suite aux évaluations nationales dépendent fortement de la politique de l'établissement, le nombre d'heures étant variable d'un établissement à l'autre.

Du diagnostic à l'aide individualisée : une étape difficile dans le traitement de l'information

Nous nous sommes intéressés également à la façon dont les enseignants pouvaient utiliser les résultats des évaluations nationales comme outil d'aide à la décision concernant le traitement des difficultés des élèves à partir des informations qu'ils pouvaient lire et interpréter sur la base des items. Pour certains enseignants, les exercices proposés sont parfois jugés trop difficiles ou pour le moins très éloignés de ce qu'un élève entrant en 6^{ème} a besoin de maîtriser alors que les erreurs fondamentales sont connues depuis fort longtemps (celles relatives au calcul décimal par exemple). De plus, les programmes de l'enseignement primaire sur lesquels

reposent les évaluations sont jugés trop étendus et peu précis, ne permettant pas une maîtrise réelle des savoirs et savoir-faire en mathématiques. Certains items sont jugés particulièrement imprécis et peuvent conduire à faire douter les élèves qui choisissent alors de ne pas répondre aux questions (comme l'a montré un item sur les droites perpendiculaires où les élèves doivent vérifier avec leur équerre). D'autres items ont des questions trop ouvertes ce qui rend par la suite le codage très difficile pour l'enseignant. Pour certains enseignants de mathématiques, quelques problèmes posés par les évaluations nationales nécessitent une réelle maîtrise du calcul numérique (comme par exemple la division). Or, certains élèves répondent faux tout simplement parce qu'ils ne maîtrisent pas le calcul et non parce qu'ils sont incapables de trouver la démarche appropriée.

L'une des difficultés majeures pour mettre en place une aide individualisée est que les items ne permettent pas d'inférer la démarche de l'élève et son cheminement. Dans ces conditions, le travail de l'enseignant ne peut guère partir du livret d'évaluation parce que le résultat de l'élève compte moins que l'explicitation de sa démarche. Ceci nécessite de la part des enseignants une analyse approfondie des erreurs. Or, si les enseignants sont formés à l'analyse d'erreurs au cours de séances de formation, ils dénoncent l'ampleur de la tâche dans des classes avec des effectifs élevés ou avec des profils très hétérogènes dans des établissements qui n'offrent pas toujours les moyens de travailler par petits groupes.

Par ailleurs, les évaluations nationales semblent concurrencées par le système d'évaluation personnel de l'enseignant qu'il juge plus efficace et plus approprié. En effet, au début de l'année, l'enseignant de mathématiques, par une succession de choix en temps réel, cherche à se rapprocher, par essais et erreurs, de la moins mauvaise solution possible selon ses propres critères et ceci dans un environnement complexe (hétérogénéité des classes, contrainte des programmes, relations avec l'administration et les familles). Il oscille donc entre planification de l'action et capacités d'improvisation. Or, pour certains enseignants, les évaluations nationales arrivent trop tôt avant que l'enseignant ait eu le temps de mettre en place un contrat pédagogique avec sa classe. En fait, le travail d'évaluation de l'enseignant cherche à faire progresser les élèves sur la maîtrise d'un certain nombre d'outils mathématiques, et il lui suffit de quelques semaines pour repérer les élèves en difficultés, les erreurs de base étant connues et les évaluations nationales ne lui semblant pas apporter d'informations complémentaires.

Une information technique et complexe

Les évaluations nationales sont jugées trop technicistes. Le codage génère un comportement quasi-taylorien de saisie, les enseignants décrivant ce moment comme pénible, sans intérêt et machinal. L'utilisation du logiciel Casimir permet de sortir de nombreux graphiques statistiques, d'établir de nombreuses corrélations mais les enseignants sont démunis face à cette inflation statistique. L'usage du logiciel nécessite de passer un temps considérable pour dégager des profils d'erreurs nécessaires à la remédiation mais ceux-ci n'apportent aucune connaissance réelle sur la démarche de l'élève. Généralement, l'enseignant, quand il s'intéresse aux résultats, se contente de donner le score moyen de l'élève et ses résultats globaux

dans les principaux domaines de compétences. Lorsqu'ils sont pris en compte, ces résultats servent principalement à constituer des groupes de remédiation en ciblant un type d'erreur mais les cahiers ne sont généralement pas réutilisés. Il n'y a quasiment pas de recours aux banques d'exercices proposées par le site du Ministère parce que les enseignants ignorent leur existence ou que le nombre d'exercices trop restreint ne présente aucun intérêt par rapport à leur propre banque personnelle. Enfin, les évaluations nationales sont utilisées ponctuellement parce que la mise en œuvre d'une remédiation demande du temps et que certains enseignants doivent gérer simultanément plusieurs classes de sixième.

Une difficulté majeure du travail de remédiation avec les élèves concerne l'appropriation du logiciel Casimir. Nous avons repéré trois seuils de complexité dans son usage par les enseignants et l'importance d'une dynamique de familiarisation nécessaire à l'élaboration d'un diagnostic des difficultés des élèves. D'abord un seuil « ergonomique » parce que le logiciel Casimir possède des fonctionnalités assez incompréhensibles pour un novice en informatique qui doit s'appuyer en plus sur un mode d'emploi hyper-technique. Ensuite, un seuil « statistique » : les enseignants n'ont pas forcément les compétences requises en statistiques pour croiser les données et analyser les résultats même s'il s'agit d'enseignants de mathématiques ! Certes, l'enseignant expert peut normalement extraire des dizaines de graphiques et tableaux en obtenant des scores de réussite et d'échecs de manière détaillée pour l'ensemble des items. Mais les enseignants ne voient généralement pas ce qu'ils peuvent obtenir comme information et ce qu'ils sont capables d'en faire. Ils s'en tiennent à des éléments généraux sur l'élève et la classe. Ils regardent ainsi principalement les scores de réussite des élèves, de la classe et de l'établissement sans considérer d'autres variables permettant de formuler un diagnostic. Confrontés à des tableaux de nombres et de nombreux graphiques, les enseignants, même parvenus jusque-là, jugent ces informations dans la plupart des cas illisibles et inutilisables. Enfin, il existe un troisième seuil que nous avons qualifié d'« inférentiel ». En effet, une fois les erreurs détectées par les codes, le passage de la codification à la pratique se révèle très difficile même pour des enseignants expérimentés et familiarisés avec les évaluations nationales et la didactique des mathématiques.

En conclusion, il faut bien reconnaître que la fonction « diagnostic » des évaluations nationales, c'est-à-dire celle qui devrait servir à mieux évaluer les erreurs des élèves de façon à résoudre leurs difficultés pour les faire progresser, est loin de remplir son rôle. On pourrait être tenté d'imputer ce phénomène à la résistance au changement des enseignants, comme tente de le faire une certaine vulgate sociologique en reprenant à son compte la critique du « cercle vicieux bureaucratique » (Crozier, 1964). Toutefois, il semble que l'explication du non-usage des évaluations nationales, tout au moins au collège, soit plus complexe. Sans vouloir donner entièrement raison aux enseignants, ces causes sont à chercher dans le contexte dans lequel ces évaluations prennent sens et dans les difficultés inhérentes au travail pédagogique (Chatel, 2001). Par ailleurs, les évaluations nationales répondent à une logique institutionnelle partiellement étrangère aux enseignants parce qu'ils ne perçoivent pas forcément l'intérêt de comparer les résultats des élèves à un niveau national ou international ou d'en faire des

instruments de pilotage du système éducatif. Les évaluations nationales leur semblent conçues dans un autre monde, sans lien avec leur expérience, leurs pratiques de formation, ou la culture propre à leur discipline d'enseignement. C'est donc sans doute entre ces deux mondes, celui de la pédagogie et celui de la politique, qu'il semble nécessaire aujourd'hui de tisser des liens plus étroits, par un travail de reproblématisation et de traduction des enjeux de l'évaluation, afin de mettre fin au conflit des interprétations.

Bibliographie

A Nation at risk, 1984, National Commission on Excellence in Education, Cambridge (Mass.), USA Research.

Chatel E., 2001, *Comment évaluer l'éducation. Pour une théorie sociale de l'action éducative*, Paris, Delachaux et Niestlé.

Crozier M., 1964, *Le phénomène bureaucratique*, Paris, Seuil, 1964.

Derouet J.-L., 2000, « Administration, science de l'administration et sociologie en éducation. Quelques exercices de traduction » in *Revue de l'Institut de Sociologie*, n°1997/1-4.

Thélot C., 1993, *L'évaluation du système éducatif : coûts, fonctionnement, résultats*, Paris, Nathan.