

L'ÉDUCATION NATIONALE : ORGANISER SON ÉVALUATION POUR AMÉLIORER SA PERFORMANCE

Enquête demandée par le Comité d'évaluation et de contrôle des
politiques publiques de l'Assemblée nationale

Décembre 2017

Sommaire

AVERTISSEMENT	5
SYNTHÈSE	7
ORIENTATIONS ET RECOMMANDATIONS	15
INTRODUCTION.....	17
CHAPITRE I DES ACTEURS DE L'ÉVALUATION DISPERSÉS	21
I - DES INSTANCES CHARGÉES DE L'ÉVALUATION, AUX POSITIONS FRAGILES ET INSTABLES	21
A - Des instances sans existence pérenne	22
B - Le Conseil national d'évaluation du système scolaire (CNESCO) : une action mal articulée avec celle du ministère.....	23
C - Un malaise permanent sur la reconnaissance d'une instance indépendante chargée de l'évaluation du système éducatif.....	26
II - DE NOMBREUX SERVICES MINISTÉRIELS PARTICIPANT À L'ÉVALUATION, SANS COHÉRENCE D'ENSEMBLE.....	27
A - Les inspections générales : des missions partiellement axées sur l'évaluation, une confusion des rôles pour l'inspection générale de l'éducation nationale	28
B - La direction de l'évaluation, de la prospective et de la performance (DEPP) : entre service statistique et acteur de l'évaluation	29
C - La direction générale de l'enseignement scolaire (DGESCO) : une position à clarifier entre prescripteur pédagogique et acteur de l'évaluation.....	31
D - Le conseil supérieur des programmes : un champ d'évaluation non exploré.....	32
III - DES ACTEURS EXTERNES CONTRIBUANT À L'ÉVALUATION DE L'ÉDUCATION NATIONALE	33
A - Le comité de suivi de la Loi de refondation de l'École : l'évaluation comme un des critères d'application de la Loi	33
B - Le secrétariat général pour la modernisation de l'action publique : une visée autre qu'évaluative	34
CHAPITRE II DES MODES D'ÉVALUATION SANS UNITÉ DE VISION.....	37
I - UNE ÉVALUATION DES ACQUIS DES ÉLÈVES TROP DISCONTINUE POUR ÊTRE UTILISÉE.....	37
A - Une boîte à outils au contenu très changeant	37
B - Des résultats peu conciliables entre plusieurs dispositifs d'évaluation de la maîtrise du socle commun de connaissances	40
II - LES ÉVALUATIONS DES PERSONNELS ET DES ÉTABLISSEMENTS : UNE CONTRIBUTION TRÈS FAIBLE À L'ÉVALUATION D'ENSEMBLE.....	41
A - L'impact très réduit de l'évaluation pour différencier les évolutions de carrière	41
B - Une évaluation des enseignants mal articulée avec les autres dispositifs évaluatifs du système scolaire	43
C - L'évaluation des personnels de direction et des établissements : un système incomplet et imparfait.....	45
III - L'ÉVALUATION DES DISPOSITIFS SCOLAIRES : DES INITIATIVES NOMBREUSES MAIS SANS APPRÉCIATION EXPLICITE DE LEURS RÉSULTATS	49
A - L'enseignement de la lecture : des résultats contradictoires	51
B - Les classes de niveau : une position jamais clarifiée au regard des évaluations internationales	52
C - Le suivi individualisé des élèves et la différenciation pédagogique : un renoncement à évaluer.....	54

D - L'éducation prioritaire : des conclusions péremptoires et contradictoires	55
E - Un pilotage mal assuré et peu assumé des innovations et des expérimentations	57

CHAPITRE III DES DÉPENSES DIFFUSES ET MAL RÉPARTIES, UNE EFFICIENCE MAL MESURÉE65

I - L'ÉVALUATION DES ACQUIS DES ÉLÈVES : DES COÛTS INHÉRENTS À L'ACTE D'ENSEIGNER, UNE RATIONALISATION POSSIBLE65

A - L'évaluation formative : un ordre de grandeur d'un coût consubstantiel à l'instruction des élèves	65
B - Les évaluations certificatives : un coût incertain pour une portée évaluative faible	66
C - Les évaluations sommatives par test : un coût modeste et dégressif	69

II - LA DÉPENSE NON PILOTÉE DES SERVICES CHARGÉS DE L'ÉVALUATION70

A - Le coût mal cerné des inspections et évaluations locales.....	70
B - Les dépenses faibles de soutien aux évaluations, modiques pour les actions elles-mêmes.....	71
C - Une mesure de l'efficacité prévue par les lois de finances sans consistance opérationnelle	72

CHAPITRE IV DÉCIDER DURABLEMENT DES FONDEMENTS D'UNE FONCTION D'ÉVALUATION.....75

I - APPROFONDIR LES RÉSULTATS DES ENQUÊTES INTERNATIONALES POUR DES ENSEIGNEMENTS CONCRETS.....75

A - Une réaction univoque et peu opérationnelle.....	76
B - Des écarts de performance non expliqués et non étudiés	78
C - La richesse non exploitée des comparaisons internationales qualitatives	82

II - REMPLIR UNE CONDITION PREMIÈRE ET NÉCESSAIRE : MESURER SYSTÉMATIQUEMENT LES ACQUIS DE LA SCOLARITÉ OBLIGATOIRE.....83

A - Généraliser l'emploi des tests standardisés.....	84
B - Irriguer la fonction d'évaluation des données sur les acquis scolaires.....	88

III - ORDONNER LA FONCTION ÉVALUATIVE DU SYSTÈME ÉDUCATIF.....89

A - Classifier les outils d'évaluation de l'éducation nationale	90
B - Les conditions de l'établissement d'une architecture cohérente de la fonction d'évaluation.....	91

CONCLUSION GÉNÉRALE97

ANNEXES99

Avertissement

Sur proposition du Comité d'évaluation et de contrôle des politiques publiques (CEC), le Président de l'Assemblée nationale a saisi, par lettre du 6 décembre 2016, le Premier président de la Cour des comptes d'une enquête sur l'organisation et l'utilisation de la fonction d'évaluation au ministère de l'Éducation Nationale, en application de l'article L. 132-6 du code des juridictions financières. Le Premier président a accepté cette demande dans une lettre au Président de l'Assemblée nationale du 14 décembre 2016, la remise du rapport étant fixée au mois de décembre 2017 au plus tard.

Le champ de l'enquête a été précisé lors d'une réunion avec les députés rapporteurs désignés par le CEC, M. Michel Piron et Mme Valérie Corre, le 1^{er} février 2017. Un courrier du Premier président du 10 février 2017 a fixé les axes principaux des travaux de la Cour (annexe I) :

- le degré de mise en cohérence des différents outils d'évaluation quels que soient leurs objets, élèves, personnels, unités d'enseignement, pratiques pédagogiques ou système dans son ensemble ;
- l'impact des évaluations sur la gestion du système éducatif et les moyens de le renforcer s'il apparaissait qu'il fût insuffisant.

L'enquête a été conduite auprès de la direction générale de la direction générale de l'enseignement scolaire (DGESCO), du secrétariat général du ministère au titre de la direction de l'évaluation, de la prospective et de la performance (DEPP), de la direction générale des ressources humaines (DGRH) et de la direction des affaires financières (DAF), de l'inspection générale de l'administration de l'éducation nationale et de la recherche (IGAENR), de l'inspection générale de l'éducation nationale (IGEN), du Conseil national d'évaluation du système scolaire (CNESCO) ainsi que dans cinq académies (Paris, Toulouse, Versailles, Aix-Marseille et Amiens). Des échanges ont eu lieu avec le Conseil supérieur des programmes (CSP), le secrétariat général à la modernisation de l'action publique (SGMAP) et la direction de l'éducation de l'Organisation de coopération et de développement économiques (OCDE).

Un relevé d'observations provisoires a été adressé en juin 2017 aux administrations citées ci-dessus, les réponses ont toutes été reçues fin juillet 2017 et des auditions ont été provoquées par la Cour pour les quatre principales administrations du MEN concernées (IGEN, IGAENR, DGESCO, SG/DEPP) et pour le CNESCO, en septembre 2017.

Synthèse

L'évaluation constitue pour l'Éducation nationale, peut-être plus encore que pour tout autre service public, un impératif. L'école a pour mission première la transmission des connaissances. Aussi la mesure des connaissances transmises est consubstantielle à sa mission, inhérente à l'acte d'enseigner comme cette mesure est indispensable à l'évaluation du fonctionnement et des performances du service public de l'éducation.

Pourtant savoir comment cette évaluation, dont la nécessité paraît si évidente, est organisée au sein de la sphère éducative nationale, comment elle produit et diffuse ses résultats, quels enseignements en sont tirés, est un sujet mal connu car très difficile à saisir de façon synthétique.

La multiplicité des niveaux et des modes d'évaluation qui coexistent dans le système scolaire, n'est pas une cause explicative suffisante de cette perception confuse. Bien au contraire, l'éclatement des dispositifs d'évaluation aurait dû conduire à instituer un système ordonné. Or la volonté d'organiser cette fonction primordiale avec clarté et selon des modalités pérennes a fait défaut.

Les remises en cause des résultats de notre système éducatif n'ont cependant pas manqué et ne manquent pas : mauvais classement dans les comparaisons internationales en termes de niveaux et de progression pour les élèves les plus défavorisés, interrogations sur l'efficacité de l'école dans l'opinion publique nationale, débats récurrents sur l'efficacité médiocre de la dépense publique, la pertinence des réformes pédagogiques successives, les cadres de l'organisation de l'enseignement.

Si des tentatives ont été faites pour moderniser l'évaluation au sein du système scolaire, elles sont restées parcellaires et n'ont jamais envisagé cet objectif dans son ensemble, en termes institutionnels, administratifs et fonctionnels. Elles n'ont jamais eu suffisamment d'ambition pour installer une culture de l'évaluation progressivement acceptée par toutes les composantes du système scolaire.

Aussi le message de l'institution publique nationale sur l'efficacité de notre système scolaire n'est pas audible face à des évaluations internationales dont les conclusions puissamment charpentées, qualifient avec netteté la performance de chaque système d'enseignement. Tout se passe comme si la France ne disposait pas d'une fonction d'évaluation apte à éclairer les enjeux d'efficacité, d'équité et d'efficacité qu'emporte le service public de l'éducation, comme si la confrontation avec une évaluation internationale nouvelle et incisive dans ses résultats et préconisations, n'avait pas conduit la puissance publique à réagir vigoureusement en réorganisant son propre dispositif évaluatif.

Il découle de ce hiatus une grande incertitude sur ce que les citoyens comme les pouvoirs publics doivent penser de la performance collective de l'éducation nationale qui a mobilisé 68 milliards de crédits en 2017, en mobilisera 70 milliards en 2018 et emploie presque un million d'agents publics pour éduquer près de 12 millions d'élèves.

Une évaluation trop indécise entre tradition et modernité

L'évaluation des acquis des élèves par le maître dans la classe, dit formative, comme la certification du niveau atteint par des examens, certains à forte valeur symbolique comme le baccalauréat, sont des fondements constitutifs de la scolarité. Les résultats de ces évaluations formatives et certificatives ont longtemps suffi pour juger de l'efficacité du système scolaire français.

Cette vision classique, toujours indispensable, n'est plus suffisante.

L'évaluation finale de la performance du système éducatif continue à être officiellement fondée sur l'obtention des diplômes du second degré d'enseignement. Or, le lien entre ce mode d'évaluation et la mesure de la performance du système éducatif s'est relâché; sa crédibilité a été affaiblie en tant que mesure des acquis et des niveaux des élèves par la poursuite d'une cible publique de taux de réussite.

Quant aux évaluations en classe, indissociables de l'acte pédagogique, elles ne sont pas de nature à renseigner de façon normée et quantifiable sur les acquis (diagnostic) et sur les niveaux et les compétences (bilan). Elles ont une autre fonction, celles d'adapter l'apprentissage dans la classe selon les capacités et les avancées des élèves.

La massification des populations scolaires impose une sommation de la mesure des savoirs transmis, maintenant possible par l'emploi de modes d'enquêtes, plus systématiques et globaux, sous forme de tests le plus souvent numérisés. C'est ainsi que pratiquent les enquêtes de référence internationales et nombre d'institutions éducatives étrangères.

Pourtant le système éducatif français reste hésitant quant à l'utilisation de ce type de mesures. Après avoir renoncé à la systématisation de tests de diagnostic ou de bilan, l'approche nationale a restreint ses investigations à des mesures par échantillons, fragmentées par discipline selon des rythmes pluriannuels différents. Une inflexion s'amorce cependant : en 2015 des évaluations de compétence des élèves ont été testées, puis généralisées lors de la rentrée de 2017, pour les classes de cours préparatoire et sous forme numérisée pour la classe de sixième.

Par ailleurs le ministère ne s'est pas saisi de protocoles évaluatifs fondés sur des expérimentations à grande échelle. Quand ces expérimentations sont conduites, c'est à l'initiative d'autres acteurs publics qui concourent au financement de dispositifs innovants : le Fonds d'expérimentation pour la jeunesse ou les opérateurs des programmes d'investissements d'avenir. L'association d'organismes de recherche existe mais au cas par cas, de façon aléatoire, comme si les ministères chargés de l'éducation nationale et de l'enseignement supérieur, parfois réunis, peinaient à mobiliser des ressources disponibles au profit d'une politique publique majeure qui les intéresse conjointement.

Enfin les évaluations en termes d'efficience, c'est-à-dire celles qui examinent par des indicateurs inscrits en lois de finances, le rapport entre les moyens consacrés et les résultats obtenus, selon des objectifs fixés, sont de faible apport. La loi organique relative aux lois de finances (LOLF) aurait dû depuis 2002 être un vecteur puissant de cette évaluation de l'efficience. Or les objectifs stratégiques et les indicateurs les plus représentatifs retenus par le ministère pour suivre et justifier son budget, sont trop généraux, les taux d'accès au diplôme national du brevet ou du baccalauréat par exemple, pour fonder des arbitrages budgétaires.

D'autres ne peuvent être renseignés chaque année, par exemple le degré de maîtrise du socle commun de connaissances et de compétences, sanctionnant la fin de la scolarité obligatoire.

L'efficacité du système scolaire n'est pas seulement un enjeu social majeur, c'est aussi en raison de l'importance de la dépense publique induite, de l'ordre de 21 % du budget de la nation, un enjeu d'efficience qui nécessite plus que par le passé, d'évaluer les effets des politiques éducatives conduites.

Si l'évaluation de ce système n'a pas été renouvelée, c'est que l'éducation nationale doit surmonter des difficultés importantes, d'ordre culturel et de nature administrative et fonctionnelle, pour conduire ce changement.

Des obstacles culturels encore à surmonter : une acculturation nécessaire

La volonté de surmonter de nombreuses réticences culturelles n'a été ni forte, ni continue. Comme le rapport le démontre, les blocages furent et demeurent nombreux.

Le recueil systématique, dit sommatif, des données sur les résultats des élèves s'est heurté non seulement à la tradition, mais aussi à la peur que son usage ne concoure à nourrir d'autres processus évaluatifs, celui des enseignants, des établissements, à terme à fonder les dialogues de gestion au sein de l'éducation nationale. L'irruption des comparaisons internationales, fondées sur cette méthode, n'a pas été saisie pour infléchir ces réticences.

La reconnaissance de la portée des comparaisons internationales a été longue, elle demeure encore incertaine et fragile au sein de l'éducation nationale. Leur critique construite et positive qui eût été une appropriation utile de leurs apports, n'a pas eu lieu au bénéfice d'un discours parfois simpliste sur leurs conclusions, parfois inutilement dépréciatif pour le système éducatif français. L'emploi qui aurait pu en être fait de façon opérationnelle est quasi inexistant.

La réalisation d'expérimentations de grande ampleur a été entravée par des oppositions de principe fondées sur un risque de rupture d'égalité entre les élèves et par la rigidité du cadre d'emploi des enseignants. La Cour a déjà relevé ces obstacles qui rendent difficiles l'introduction en France de la différenciation des pratiques enseignantes. Enfin il existe un défaut de rigueur des autorités publiques pour respecter le cadre des protocoles expérimentaux.

L'emploi des techniques d'évaluation dites quantitatives, encourageant la critique d'être des modalités réductrices d'une réalité multifactorielle, est encore mal accepté.

L'instabilité des instances chargées de suivre l'évaluation du système scolaire peut expliquer l'incapacité à lever progressivement ces obstacles. Elles ont changé trois fois en moins de 15 ans, de composition comme de responsabilités. Chaque changement, lié à des alternances politiques, se traduit par des liens différents avec le système éducatif, marqués par des degrés variables d'indépendance ou d'interactions vis-à-vis du ministère.

Aucune d'entre elles n'a pu asseoir son autorité dans la durée, produire des référentiels méthodologiques reconnus, tenter d'ordonner progressivement la fonction évaluative de l'éducation nationale. Elles n'ont pas pu mettre en place un suivi continu et approfondi des modes d'évaluation instaurés dans d'autres pays, ce qui eût été souhaitable pour stimuler les réflexions nationales, ni s'approprier les comparaisons internationales pour les contextualiser aux réalités nationales.

Cette instabilité est un des signes du malaise qu'éprouve l'institution scolaire à faire vivre une culture de l'évaluation modernisée.

Un autre indicateur de ce malaise est la faiblesse du cadre budgétaire dans lequel s'inscrit l'évaluation de l'éducation nationale. Les deux actions budgétaires qui en rendent compte (« certification » et « soutien à l'évaluation ») ne sont pas conçues pour mettre en relief une politique d'évaluation et aucun document récapitulatif ne consolide tous les moyens qui y sont consacrés. Les montants des dépenses engagées sont parfois incertains (par exemple pour le baccalauréat) et ceux consacrés à la production d'évaluations (tests, conduite d'expérimentations) modiques, de l'ordre de moins de cinq millions par an. La mission budgétaire MIES qui réunit tous les crédits alloués à l'enseignement scolaire, n'a jamais accordé de visibilité aux instances chargées de l'évaluation du système éducatif : le parlement ne peut connaître, ni apprécier les moyens dont elles disposent, en crédits et en effectifs.

Ainsi l'éducation nationale, confrontée tant à ses besoins de mesure de son efficacité qu'aux résultats des enquêtes internationales, n'a pas réagi avec vigueur pour ressourcer sa culture de l'évaluation. Elle n'a pas non plus mis en interaction et en cohérence les différents niveaux d'évaluation qu'elle réalise.

Un éclatement des niveaux d'évaluation sans recherche d'une cohérence d'ensemble

Le ministère de l'éducation nationale dispose désormais d'un système statistique de grande qualité qui renseigne de façon fiable sur les cursus des élèves, les met en relation avec leurs caractéristiques familiales et socio-économiques (seulement pour le second degré) et leurs conditions de scolarisation (nombre d'élèves par division, d'heures d'enseignement par élève, etc.) et à un niveau relativement fin.

En revanche, sa capacité diagnostique d'ensemble est limitée car la fonction d'évaluation n'est pas organisée pour enrichir un thésaurus commun et nourrir des appréciations synthétiques. Les visées de processus évaluatifs majeurs ne sont pas intégrées et concourent peu, voire pas, à l'appréciation de la performance collective.

Il en est ainsi des modalités d'évaluation des enseignants dont la déconnexion avec les questions de performance a même pu être revendiquée. Il en est de même de dispositifs mobilisant pourtant des moyens importants (évaluations en classe, temps consacré aux inspections, examens de fin de cycle scolaire, etc.) qui ne contribuent pas à dégager une vision d'ensemble de la performance.

Ce qui est vrai de l'efficacité collective, l'est plus encore de dispositifs éducatifs particuliers pour lesquels les résultats d'évaluations nombreuses ne produisent pas, par défaut de procédures d'arbitrage entre eux, une doctrine arrêtée. Les débats confus autour de l'impact de l'éducation prioritaire, pourtant au cœur des enjeux d'efficacité et d'équité du système scolaire, en témoignent. Il en découle la quasi impossibilité d'orienter l'action publique sur le fondement des évaluations conduites.

Certes relier entre eux les résultats des différents processus évaluatifs, nécessaires au fonctionnement de l'institution scolaire, exige des précautions. Mettre en relation la mesure des résultats des élèves selon des tests normés et l'évaluation des pratiques des enseignants n'est pas chose aisée, et nécessite un effort de contextualisation et d'appréciation sur des séries

longues. Pour autant il semble légitime et utile de tisser des liens entre deux niveaux d'évaluation intimement en rapport. De nombreux pays pratiquent ces rapprochements.

De même c'est par le recueil de données sur les acquis de connaissances et de compétences des élèves que leurs besoins peuvent être mieux connus et l'allocation des moyens adaptée, comme la Cour l'a rappelé dans de nombreux rapports. L'approche par les indicateurs de valeur ajoutée des lycées illustre qu'il est possible d'analyser des données pour forger une approche de la performance d'un établissement. En quoi serait-il impossible de généraliser ce type de travaux qui renseignerait sur chaque unité scolaire ? Pourquoi ces indications sur la performance ne pourraient-elles pas alors être utilisées pour nourrir, avec d'autres paramètres, les différents niveaux de dialogue de gestion du sein du ministère ?

Si les évaluations sont peu reliées entre elles, les méthodologies employées au sein du ministère ne font pas non plus l'objet d'une doctrine, permettant de les classer, voire de les hiérarchiser selon leur valeur plus ou moins probante. La qualité du processus évaluatif requiert un tel effort. Les critiques adressées à l'instance actuelle chargée de l'évaluation du système éducatif, le Conseil national d'évaluation du système scolaire (CNESCO), illustrent les divergences de vision et d'attentes en termes de méthodologies. Sans enlever de l'intérêt à des conclusions obtenues par « consensus participatif » ou « collaboratif » tel que pratiqué par le CNESCO ou le secrétariat général à la modernisation des politiques publiques, elles n'acquièrent leur valeur probante que si elles se combinent avec des données objectives.

Les méthodes dites participatives ou collaboratives d'évaluation, pour disposer d'une base solide, ont grand besoin de données. Aussi l'appareil statistique et numérisé moderne qui peut produire beaucoup de données brutes, puis l'utilisation de l'économétrie pour les contextualiser, ne sont pas antinomiques mais complémentaires des approches dites qualitatives, plus sociétales. La coexistence de résultats évaluatifs divers, voire contradictoires, impose plus de rigueur méthodologique pour les départager et éclairer les décideurs publics. À défaut, ils ne peuvent s'appuyer sur les dispositifs d'évaluation qui ne sont alors qu'une perte de moyens sans impact.

Arrêter une cible centrale : évaluer les acquis des élèves durant la scolarité obligatoire

Pour remédier à ces difficultés culturelles et organisationnelles, il serait possible de construire un système cohérent, crédible et efficace d'évaluation à condition de lui fixer, au moins dans un premier temps, une cible prioritaire et de concevoir les instruments à cet effet

Un consensus pourrait se dégager pour estimer comme condition première et nécessaire à l'évaluation du système éducatif, la mesure du degré de maîtrise du socle commun de connaissances, de compétences et de culture, qui clôt la scolarité obligatoire.

Centrée sur un tel objectif, conforme au code de l'éducation, la mesure de la performance correspond alors à celle des niveaux d'enseignement analysés par les comparaisons internationales, au premier rang desquels l'enquête PISA (*Programme for International Student Assessment*) qui prend en compte les élèves âgés de 15 ans.

Or les instruments de mesure des acquis du socle commun mis en œuvre actuellement restent parcellaires et éclatés : les enquêtes du cycle d'évaluation disciplinaire Cèdre, sont fragmentées par champ disciplinaire différent chaque année. Il faut attendre une consolidation

en fin de cycles d'enquête (trois ou six ans, bientôt cinq ans) pour disposer d'une appréciation d'ensemble des niveaux scolaires. Le déploiement actuel des technologies numériques dans les établissements scolaires et les progrès réalisés par les tests standardisés, rendent possible la généralisation de contrôles réguliers des connaissances et compétences acquises, pour un coût raisonnable car fortement dégressif. La généralisation des tests de compétences en classe de sixième à la rentrée 2017 ouvre cette voie.

Grâce à un dispositif étendu d'évaluation des acquis des élèves, il serait non seulement possible de mesurer une performance collective, mais de décliner des données agrégées par équipes éducatives, par établissements, par dispositifs pédagogiques.

Ces indications devraient être considérées comme des appuis aux personnels et aux équipes enseignantes, elles permettraient de forger un dispositif évaluatif des écoles, qui fait en grande partie défaut, et d'améliorer celui des collèges, qui est en cours d'enrichissement. La consolidation de ces données par dispositifs éducatifs, pour ceux qui font l'objet d'un suivi particulier (par exemple l'éducation prioritaire) ou sont expérimentés, donnerait une base rigoureuse aux évaluations qui les concernent, même si elles doivent incorporer d'autres facteurs (allocation des moyens budgétaires et en effectifs, cartographie des dispositifs par exemple) lorsqu'il s'agit d'apprécier une politique publique dans son ensemble.

Fixer un objectif national : publier un rapport sur la performance du système scolaire obligatoire

Si de tels efforts étaient faits, la France disposerait des moyens de publier son évaluation de la performance de son système éducatif, au moins pour sa partie obligatoire.

Il s'agit d'un enjeu majeur comblant une carence actuelle car la seule appréciation donnée à l'opinion publique sur la performance du système scolaire français provient d'une enquête internationale d'ensemble (PISA) ou d'enquêtes thématiques portant sur les niveaux en mathématiques et en sciences (TIMSS) et la maîtrise de la lecture (PIRLS).

La France, qui est un grand pays disposant d'une culture statistique et économétrique reconnue et éprouvée, n'a pas été capable jusqu'à présent « d'émettre » sa propre opinion sur le degré de performance de son système scolaire, adaptée à la composition de sa population scolarisée, à ses programmes scolaires, à ses attentes. Il en découle que la première recommandation de la Cour est de publier un rapport équivalent à celui de PISA, ce qui implique que certaines conditions soient remplies.

Il est d'abord indispensable que des garanties soient apportées par l'instauration d'une institution indépendante chargée de réguler l'exercice de l'évaluation au sein de l'éducation nationale et de publier un tel rapport. C'est le sens de plusieurs recommandations que de définir les prérogatives et les moyens à accorder à cette instance pour qu'elle joue ce rôle d'émetteur respecté.

Il faut aussi diffuser et faire partager cette culture de l'évaluation : c'est l'enjeu des recommandations portant sur l'emploi des données sur les acquis des élèves durant la scolarité obligatoire, leur appropriation par le monde enseignant dans ses niveaux d'organisation, leur partage selon des modes d'évaluation non usités en France avec les parties prenantes que sont les parents d'élèves.

Respecter une condition indispensable : revoir l'architecture de l'évaluation du système scolaire

Pour répondre aux objectifs précédents, il est temps de sortir de compromis organique et fonctionnel de circonstances qui n'ont pas permis de donner du souffle à la culture de l'évaluation au sein du système éducatif. Le comité de suivi de la Loi de refondation de l'école appelle à cet effort dans son avis de février 2017, qui demande de mieux situer le CNESEO dans le paysage institutionnel de l'évaluation.

D'une part l'instauration d'une instance chargée de missions renforcées ne pourra produire des effets durables que si son articulation avec les inspections générales et directions du ministère est claire et précise. Les difficultés rencontrées par le passé et actuellement tiennent en partie à une insuffisante définition initiale des rôles respectifs et des moyens accordés. La Cour est convaincue que c'est la collaboration, à la fois respectueuse des missions opérationnelles des services du ministère et des prérogatives de l'instance indépendante, qui doit être recherchée plutôt qu'une position surplombante ou totalement extérieure au ministère, source d'incompréhensions et de déperdition d'efforts.

Le statut d'autorité administrative indépendante n'est pas nécessaire, ni d'ailleurs conforme aux missions de cette instance, à partir du moment où des modalités propres à garantir concrètement l'exercice indépendant de ses missions, sont mises en œuvre.

En l'état actuel, les pouvoirs publics doivent prendre en considération le malaise existant sur le positionnement du CNESEO. Si cette instance dispose de jure de nombreuses compétences, elle n'a pas vu de facto son rôle et ses apports compris et intégrés, elle ne réunit pas en outre toutes les conditions requises pour asseoir son autorité. Les cadres juridique (composition et compétences) et matériel (moyens et effectifs) dans lesquels elle s'inscrit, devront être profondément revus si l'on souhaite affermir la place et le rôle d'une instance chargée de l'évaluation du système scolaire.

D'autre part le ministère devra mieux identifier le rôle respectif de ses inspections générales et administrations centrales et les adapter à l'insertion d'une instance chargée de l'évaluation dans l'architecture d'ensemble, évitant son plaquage sur une organisation qui n'aurait pas été revue.

Divers sujets méritent d'être abordés tels que la distinction plus nette des fonctions statistiques et d'études évaluatives au sein de la DEPP en traitant la question du niveau des moyens budgétaires accordés à leur réalisation, la clarification du rôle entremêlé de prescripteur et d'évaluateur de la DGESCO, la détermination des marges de manœuvre dévolues aux rectorats (académiques ou de régions académiques) pour conduire des innovations, la définition de la place des deux inspections générales, notamment de l'inspection générale de l'éducation nationale dont une part des missions pourrait relever de l'instance chargée de l'évaluation. Sur ce dernier point, la réflexion lancée par le ministre sur le rôle des deux inspections générales, voire leur rapprochement, offre un moment propice à ce réagencement d'ensemble : il s'agirait d'éviter dorénavant la confusion des genres dans les activités de l'inspection générale de l'éducation nationale entre intervention opérationnelle dans la conduite des politiques éducatives et évaluation de leurs effets.

Orientations et recommandations

L'évaluation n'a pas été conçue au sein de l'éducation nationale pour qu'elle devienne grâce à elle, une organisation « apprenante » tirant les enseignements de ses succès comme de ses échecs, appréhendés selon des méthodes clarifiées et validées et fondés sur le recueil de données sur les acquis et les besoins des élèves.

Face à un tel enjeu systémique, la Cour propose trois orientations de réforme dont sont déclinées les recommandations suivantes:

Ordonner la fonction d'évaluation pour créer un dispositif cohérent

1. Donner pour mission à une instance chargée de l'évaluation de :
 - publier un rapport pluriannuel sur la performance du système éducatif français (à l'instar des grandes comparaisons internationales)
 - préparer un programme pluriannuel d'évaluation, en lien avec les services du ministère de l'éducation nationale .
2. Accorder à cette instance des prérogatives précises et prescriptives pour veiller à la qualité et l'exhaustivité de l'évaluation :
 - accéder directement aux données dont dispose le ministère ;
 - valider les outils d'évaluation ;
 - classer et qualifier les productions évaluatives.
3. Faire dresser par cette instance une cartographie des évaluations produites et/ou nécessaires au système éducatif et des services concernés pour établir une architecture renouvelée de la fonction d'évaluation du ministère ;
4. Redéployer les moyens du ministère :
 - pour permettre à l'instance chargée de l'évaluation, de conduire sa mission et de diligenter, en cas de défaut d'exécution du programme pluriannuel, des évaluations autonomes ;
 - pour permettre à la DEPP de conduire des protocoles évaluatifs quantitatifs.
5. Créer une action budgétaire dans le programme « Soutien de la politique de l'éducation nationale » pour identifier les moyens (crédits et effectifs) de l'instance chargée de l'évaluation.

Définir comme condition nécessaire et première de la fonction d'évaluation du système éducatif français, la mesure du niveau et des acquis des élèves durant et à la fin de la scolarité obligatoire

6. Systématiser les évaluations des élèves du socle commun de connaissances, de compétences et de culture (en début et en fin des 3 cycles, puis à l'entrée et à la sortie de chaque année) par des tests standardisés dématérialisés ;
7. Consolider progressivement ces données pour mesurer la performance du système éducatif, décliner ces données au service des enseignants, des équipes enseignantes et dirigeantes des écoles et établissements.

Développer la culture de l'évaluation, par une appropriation des modes d'évaluation par l'ensemble des acteurs du système éducatif

8. Développer les évaluations collectives, soit d'équipes enseignantes, soit d'établissements, en utilisant les évaluations des résultats des élèves, développer l'évaluation des écoles primaires ;
9. User de nouveaux modes d'évaluation des unités d'enseignement, impliquant les enseignants et personnels de direction dans des processus dirigés d'auto-évaluation, les usagers dans des formules participatives ;
10. Intégrer au processus des réformes éducatives leur évaluation par des techniques quantitatives (essais randomisés contrôlés) ;
11. Faciliter les innovations et expérimentations locales, y compris par des soutiens budgétaires ou en effectifs, dérogeant si besoin au cadre de droit commun des affectations, apporter un appui méthodologique pour prévoir leur reproduction.

Introduction

L'évaluation constitue une partie intégrante de l'acte pédagogique. Avant d'enseigner, le niveau de connaissance et de compétence de l'élève doit être connu; au cours de l'apprentissage, ses progrès doivent être suivis, enfin ses acquis être estimés et validés, selon les cycles scolaires par un examen.

Cette fonction consubstantielle à l'acte d'enseigner, a vu progressivement son champ s'enrichir. Si la visée initiale subsiste et demeure centrale, elle ne constitue plus à l'heure actuelle l'unique objet de la fonction d'évaluation du monde scolaire. Les responsables publics et l'opinion publique attendent une évaluation globale de la performance du système éducatif.

Cette appréciation exige des processus évaluatifs spécifiques et divers en raison de la nature polysémique de la mission du système éducatif, qui concerne certes en premier lieu le niveau des élèves mais aussi la qualité des pratiques enseignantes, l'efficacité comparée des écoles et des établissements scolaires, l'impact des dispositifs pédagogiques et éducatifs.

En outre l'importance des ressources budgétaires consacrées à l'éducation nationale et leur croissance rapide du fait de la massification des populations scolarisées comme de l'exigence de rehaussement des niveaux scolaires, impose une appréciation en efficacité et en efficience.

Enfin la place croissante, voire prépondérante, dans l'opinion publique des comparaisons internationales sur les résultats des systèmes éducatifs, a donné à l'évaluation un relief accru.

Or cette évaluation est particulièrement délicate car les effets des politiques éducatives sont différés, peuvent être affectés de nombreux autres paramètres que ceux strictement pédagogiques et s'étendent, pour les individus, sur une vie entière, et pour un pays, sur plusieurs générations. Délicate également car les objectifs assignés à un service public éducatif, notamment dans les pays occidentaux, peuvent comporter de nombreuses exigences, empreintes d'une vision extensive de la mission d'instruction.

Pour toutes ces raisons, si une société estime que la fonction d'évaluation de son système éducatif est un enjeu primordial, qu'il est souhaitable que la production et la diffusion de ses résultats soient acceptés, aient un impact et soient en définitive utiles, elle doit veiller à ce que son organisation soit robuste, compréhensible et stable.

**

En France, le code de l'éducation nationale témoigne en effet de l'importance donnée à cette fonction au sein du système scolaire, le terme d'évaluation étant présent plus de 180 fois dans ses différents articles.

L'article L 221-1 du code de l'éducation assigne à l'État cette mission en lui donnant comme ambitieuse visée de conforter la cohérence du service public : « L'État assume, dans le

cadre de ses compétences, des missions qui comprennent le contrôle et l'évaluation des politiques éducatives, en vue d'assurer la cohérence d'ensemble du système éducatif ».

Le rapport annexé à la Loi du 10 juillet 1989 d'orientation sur l'éducation explicite la place de la fonction d'évaluation dans le système éducatif : « L'évaluation s'applique à tout le système éducatif : les élèves, les étudiants, les personnels, les établissements, les services extérieurs, l'administration centrale. Loin de conduire à une mise en concurrence des établissements et de leurs enseignants, elle contribue à l'amélioration du système éducatif en vérifiant la mise en œuvre des objectifs éducatifs nationaux, en les adaptant aux différents publics auxquels ils s'adressent et en opérant une régulation permanente de l'ensemble du système éducatif ».

*
**

Au regard de son importance systémique comme en témoigne la loi précitée, la Cour des comptes a déjà émis dans de nombreux rapports des observations et des recommandations, sur la conformité de l'évaluation telle que pratiquée par le ministère de l'éducation nationale avec les prescriptions du code de l'éducation.

Les rapports publics thématiques suivants : L'éducation nationale face à l'objectif de la réussite de tous les élèves¹, L'orientation à la fin du collège : la diversité des destins scolaires selon les académies², Gérer les enseignants autrement³, Le suivi individualisé des élèves : une ambition à concilier avec l'organisation du système éducatif⁴ et enfin, Gérer les enseignants autrement : une réforme qui reste à faire⁵, ont tous insisté sur la nécessité de structurer et renforcer les évaluations dans les domaines qu'ils traitaient.

Maintes observations de la Cour ont mentionné des failles, voire des béances, dans la fonction d'évaluation, par exemple pour les dispositifs de suivi individualisé des élèves. Elles ont surtout insisté pour que l'évaluation soit plus utilisée pour adapter les moyens aux besoins grâce à une connaissance plus fine de ces derniers : les deux rapports publics portant sur la gestion des enseignants (Gérer les enseignants autrement (Mai 2013 et octobre 2017)) ont notamment recommandé la mise en place d'un « système de mesure et d'analyse assurant une connaissance précise et fiable des besoins des élèves » et de « développer l'évaluation collective au niveau d'un établissement ou d'une équipe pédagogique ».

¹ Cour des comptes, *Rapport public thématique : L'éducation nationale face à l'objectif de la réussite de tous les élèves*. La Documentation française, mai 2010, 216 p., disponible sur www.ccomptes.fr

² Cour des comptes, *L'orientation à la fin du collège : la diversité des destins scolaires selon les académies*. Communication à la commission des finances, de l'économie générale et du contrôle budgétaire de l'Assemblée nationale, décembre 2012, 243 p., disponible sur www.ccomptes.fr

³ Cour des comptes, *Rapport public thématique : Gérer les enseignants autrement*. La Documentation française, mai 2013, 211 p., disponible sur www.ccomptes.fr

⁴ Cour des comptes, *Rapport public thématique : Le suivi individualisé des élèves : une ambition à concilier avec l'organisation du système éducatif*. La Documentation française, mars 2015, 173 p., disponible sur www.ccomptes.fr

⁵ Cour des comptes, *Rapport public thématique : Gérer les enseignants autrement : une réforme qui reste à faire*. La Documentation française, octobre 2017, 137 p., disponible sur www.ccomptes.fr

En outre de nombreuses limitations au développement de l'évaluation trouvent leur origine dans des modes de fonctionnement et d'organisation du système éducatif dont la Cour a critiqué les trop grandes rigidité et uniformité.

Ainsi les expérimentations ne peuvent pas toujours se déployer aisément en raison d'un cadre d'exercice du métier d'enseignant qui manque de souplesse et d'une autonomie restreinte des établissements qui en limite l'audace. Cette faible autonomie rend également inopérantes les conséquences qui pourraient être tirées par les chefs d'établissement et les équipes enseignantes de l'évaluation des unités d'enseignement, leur faisant perdre une part de leur intérêt opérationnel. La Cour a plusieurs fois appelé à assouplir ces contraintes statutaires et institutionnelles.

Les modes de gestion de la carrière des enseignants comme d'allocation des moyens restent trop peu différenciés pour que les gestionnaires ressentent le besoin de disposer de critères plus fins, ce qui incite peu à développer l'évaluation pour des motifs de gestion. Là aussi la Cour a maintes fois recommandé de faire une plus large place à des critères issus de l'appréciation des résultats des élèves et des besoins qui en découlent.

En définitive la Cour a déjà observé que les modes d'organisation du service public de l'éducation n'incorporent pas naturellement, ni ne facilitent en son sein, la culture de l'évaluation.

**

Le présent rapport est ordonné en quatre chapitres qui examinent successivement

- **Les acteurs de l'évaluation du système éducatif**, pour observer l'instabilité des instances vouées à l'évaluation du système éducatif, mises en place sans doctrine pérenne quant à leur place, leurs missions, leurs moyens, ainsi que la parcellisation de la fonction d'évaluation au sein du ministère entre ses deux inspections générales et ses directions centrales.
- **Les outils et modalités d'évaluation**, en ce qui concerne les élèves, les enseignants, les établissements, les politiques éducatives et les expérimentations pédagogiques, pour noter ou le défaut de dispositifs systématiques (les élèves), ou la faiblesse du lien entre processus évaluatif et performance du système éducatif (les enseignants), ou l'absence de cohérence et d'arbitrage (les politiques éducatives), enfin la place réduite des expérimentations.
- **Les coûts budgétaires de cette fonction**, pour relever leur caractère disparate, l'absence d'une vue budgétaire d'ensemble et la faiblesse des objectifs et indicateurs prévus par la Loi organique relative aux lois de finances (LOLF) comme leviers d'évaluation.
- **L'édification d'une fonction évaluative cohérente du système éducatif**, pour constater une réaction insuffisante au défi posé par les comparaisons internationales, pour appeler à un dispositif de connaissance systématisée des acquis des élèves, enfin pour analyser les conditions d'une refonte organisationnelle de la fonction évaluative du système éducatif.

Chapitre I

Des acteurs de l'évaluation dispersés

Que l'évaluation de l'enseignement scolaire, qui porte sur des objets variés (élèves, personnels, établissements, méthodes, dispositifs pédagogiques), soit réalisée par de nombreuses entités dont les procédures et les méthodes diffèrent, n'est pas anormal tant que ces intervenants multiples effectuent leurs évaluations dans un cadre ordonné et cohérent, soumis à une procédure d'arbitrage quand leurs positions divergent.

Aussi l'existence ou non d'une fonction régulatrice de l'évaluation est primordiale, c'est le sujet des instances indépendantes de l'évaluation du système éducatif. La cohérence des structures administratives qui interviennent ne l'est pas moins, c'est le sujet de l'architecture de la fonction d'évaluation au sein du ministère.

I - Des instances chargées de l'évaluation, aux positions fragiles et instables

La Commission du débat national sur l'avenir de l'école, dont les travaux se déroulèrent de septembre 2003 à mars 2004, avait suggéré la création d'une « haute Autorité indépendante » chargée de développer la culture de l'évaluation au sein du système éducatif.

Le rapport issu de la concertation sur le thème « Refondons l'école de la République », publié en octobre 2012, a appelé à « une évaluation de l'École recrédibilisée », il mentionnait entre autres aspects que « l'évaluation concerne le niveau central, qui doit se doter d'une vision globale du fonctionnement et de l'efficacité du système éducatif... L'évaluation, pour être reconnue et utile, doit être cohérente, scientifiquement légitime et participative ». Aussi était-il recommandé de « Créer une instance indépendante qui contribuera à rendre lisible et transparent l'ensemble du processus d'évaluation ».

Malgré ces préconisations convergentes à dix ans d'écart, les instances successives qui ont été créées, sont demeurées précaires et leurs missions incomplètement précisées et exercées.

A - Des instances sans existence pérenne

1 - Le Haut Conseil pour l'évaluation de l'école (HCéé) : l'évaluation de l'évaluation

Créé par le décret du 27 octobre 2000, le HCéé, composé de 35 membres⁶ avait une triple mission : l'expertise des évaluations produites et diffusées par le ministère, la synthèse des études et des recherches existantes, la proposition d'évaluations spécifiques .

Outre un rapport annuel, il a publié entre sa création et son remplacement par le Haut conseil de l'école (HCE) en 2005, une vingtaine d'avis évaluatifs portant sur de multiples et très variés sujets (l'effet de la réduction de la taille des classes sur les progrès des élèves, l'évaluation des lycées, l'évaluation des enseignants et des pratiques enseignantes des premier et second degré, l'orientation à la fin du collège et au lycée).

Le HCéé était une instance d'évaluation majoritairement interne au système éducatif, tous ses membres étant nommés par le ministre de l'éducation nationale. Il a pu être considéré comme une instance chargée d'évaluer l'évaluation. Il a d'ailleurs émis des avis sur l'organisation de l'évaluation au sein du ministère (octobre 2001) estimant que « l'appréciation des forces et des faiblesses de l'évaluation du système éducatif français peut être considérée comme la tâche permanente du Haut Conseil de l'évaluation de l'école ».

Dans son rapport « L'éducation nationale face à l'objectif de la réussite de tous les élèves » (mai 2010), la Cour avait déploré sa disparition, considérant que l'évaluation de l'évaluation n'était pas assurée par l'instance qui lui a succédé (le HCE). Elle tenait pour paradoxal que cette suppression intervienne alors que la Loi organique relative aux lois de finance (LOLF) renforçait la mesure de l'efficacité des politiques publiques.

2 - Le Haut Conseil de l'Éducation (HCE) : un comité des sages

La loi d'orientation et de programme pour l'avenir de l'école du 23 avril 2005 a créé le Haut Conseil de l'Éducation, composé de neuf membres dont aucun n'était désigné par le ministre de l'éducation (trois par le Président de la République, deux respectivement par les présidents du Sénat, de l'Assemblée nationale et du CESE), la plupart n'était pas directement liée à la sphère de l'éducation nationale. Organisme consultatif, seul son rapport annuel était obligatoirement public.

Le HCE devait dresser, selon une formule cursive et vague, un bilan des résultats du système éducatif (« *Le Haut Conseil de l'éducation adresse, chaque année, un bilan des résultats obtenus par le système éducatif* » décret du 22 août 2005), ainsi que des expérimentations menées sur le terrain. Il pouvait formuler des propositions sur la pédagogie, les programmes, l'organisation, les résultats du système éducatif et la formation des enseignants. Sa saisine était obligatoire dans deux domaines : les contenus du socle commun

⁶ Élus, personnels de l'Éducation nationale, parents d'élèves, lycéens, étudiants, salariés et employeurs, personnalités françaises et étrangères reconnues pour leurs compétences dans le domaine de l'évaluation du système éducatif (12).

de connaissances et de compétences instauré par la loi du 23 avril 2005 et du cahier des charges de la formation des enseignants.

Le HCE a produit, jusqu'à son remplacement en 2012, sept « *bilans des résultats du système éducatifs* », rapports remis au Président de la République, plus vastes dans leurs champs (le collège, l'école primaire, l'orientation, l'enseignement professionnel, etc.) que les avis du HCcé et conclus par des recommandations parfois générales (« *Impulser le changement* »⁷), parfois plus restreintes (créer un statut du directeur d'école primaire, permettre aux chefs des établissements les plus difficiles d'émettre un avis sur l'affectation d'un enseignant dans leur établissement).

La logique et le registre de cette instance relevaient plutôt d'un comité des sages permanent que d'un organisme régulant l'évaluation du système éducatif.

B - Le Conseil national d'évaluation du système scolaire (CNESCO) : une action mal articulée avec celle du ministère

Le CNESCO a été créé par l'article 33 de la loi pour la refondation de l'École de la République du 8 juillet 2013 avec une mission ample, plutôt orientée cependant sur les questions méthodologiques : « évaluer en toute indépendance l'organisation et les résultats du système scolaire ...réaliser ou faire réaliser des évaluations » et de « se prononcer sur les méthodologies et les outils des évaluations conduites par le ministère chargé de l'éducation nationale ainsi que sur les résultats de ces évaluations... donner un avis sur les méthodologies, sur les outils et sur les résultats des évaluations des systèmes éducatifs conduites dans le cadre de programmes de coopération européens ou internationaux....formuler toute recommandation utile au regard des résultats des évaluations »⁸.

1 - Une pratique évaluative voulue comme participative

Son indépendance par rapport au ministre de l'éducation nationale a été nettement affirmée. Le CNESCO est composé de deux députés, deux sénateurs, deux représentants du CESE et de huit personnalités qualifiées, parmi lesquelles les sociologues sont particulièrement représentés. Son comité consultatif fait une large part aux représentants syndicaux des personnels de l'enseignement (et des salariés en général) et des parents d'élèves et des lycéens. C'est plus le registre de la participation des « parties prenantes » que celui d'un appui scientifique et méthodologique qui a été privilégié. Ce choix est cohérent avec l'orientation générale donnée aux méthodes de travail du CNESCO, qui se revendique comme porteur d'une évaluation « *participative* ».

Aussi, est-il logique que dans les 6 formats d'activités que le CNESCO a retenus, l'approche participative ait été nettement privilégiée par l'organisation de conférences sous diverses formes, soit « *de comparaisons internationales* », soit « *de consensus* », soit de « *conférence virtuelle interactive* », soit enfin de « *forum en région* », consacrées à des sujets

⁷ Rapport 2012.

⁸ Art. L. 241-12 du CE.

très divers, allant d'objets strictement pédagogiques (redoublement et alternatives, compréhension en lecture) à des thèmes plus « sociétaux » (Les mixités à l'école, alliances éducatives, citoyenneté, etc.). Ces conférences qui se distinguent assez peu de celles qu'organiseraient des universités ou centres de réflexion, sont fondées sur des « *évaluations scientifiques* » existantes, sur des synthèses de publications, parfois très générales (par exemple une note d'actualité sur les grands débats du baccalauréat en juin 2016, un rapport international dénommé « la mixité sociale à l'école en juin 2015 », une note d'actualité sur l'école et l'immigration en janvier 2015, une « revue documentaire sur l'école et le handicap » en janvier 2016...), enfin sur des travaux évaluatifs commandés à cette occasion (souvent grâce aux nombreux partenariats tissés par le CNESCO avec des équipes de chercheurs).

Ces conférences produisent des « *synthèses* » et des « *préconisations* » selon la méthodologie suivante (*source CNESCO*):



Cette option participative a permis à environ un millier de personnes appartenant aux « parties prenantes » (enseignants, parents, représentants de collectivités locales, du monde syndical ou associatif, etc.) de contribuer à partir des travaux, à des « *co-constructions* » de synthèses et de préconisations. Allant de pair avec ce travail d'appropriation, une large politique de diffusion a été menée.

2 - Des critiques dues à des perspectives et des attentes trop divergentes

Un malaise diffus s'est instauré entre cette instance et les administrations du ministère, fondé sur deux types de critiques : d'une part la production d'évaluations en tant que telle est restée faible, d'autre part la mission d'appui méthodologique du CNESCO a été sous investie.

Au-delà de facteurs de circonstances, la raison principale de ce malaise tient à un défaut de vision commune et initiale entre le conseil et le ministère sur son mode de travail, son type de collaboration avec les services, plus fondamentalement sur son rôle et son apport.

Une illustration de cette incompréhension mutuelle est l'insatisfaction manifestée sur les apports à la méthodologie de l'évaluation. Si des préconisations ont été systématiquement émises lors des conclusions des études, elles portaient souvent sur le contenu à infléchir des dispositifs éducatifs étudiés (ce qui est aussi une mission du CNESCO), moins sur le contenu des outils d'évaluation et les conditions de leur mise en œuvre opérationnelle. À contrario, le CNESCO estime avoir produit de nombreuses préconisations méthodologiques, par exemple « la mise en place d'un appareil statistique robuste sur l'enseignement primaire, la remise en place des évaluations standardisées nationales sur tous les niveaux d'enseignement, la mise en place d'un modèle d'évaluation des académies, au-delà des pratiques actuelles du dialogue de gestion, la mise en place d'un appareil d'évaluation, national et local, sur les débouchés des formations professionnelles (accessible aux familles, la rénovation des outils d'évaluation des

acquis des élèves pour mieux intégrer le handicap... ». Or leur caractère très général ne correspond pas aux attentes du ministère.

La production d'une synthèse par le CNESEO sur les enquêtes PISA⁹ et TIMSS¹⁰ en fin 2016, aurait pu être un moment important pour apprécier leur portée. En fait il s'est agi d'une explicitation et d'une vulgarisation des modes d'enquêtes de ces études (certainement utile pour en faire comprendre les ressorts complexes) plus que de leur analyse critique, par exemple en les contextualisant aux réalités françaises. La synthèse présentée dans le rapport d'activité de 2016 est descriptive, assortie de précautions d'interprétation très générales. La conclusion du rapport spécifique¹¹ est de la même tonalité: « Les analyses qui figurent dans ce rapport aideront à la compréhension de tous des résultats des enquêtes PISA 2015, TIMSS 2015 et TIMSS Advanced 2015, et montrent combien la réception de leurs résultats, souvent présentés sous la forme réductrice de palmarès unidimensionnels, doit être prudente, combien la nécessité de croiser les analyses est essentielle pour porter un regard averti sur les acquis des élèves, et la nature de ces acquis ».

La recommandation finale n'est pas suffisamment explicite pour être opérationnelle : « Elles révèlent aussi les investigations supplémentaires à mener en termes de mise en perspective à long terme des résultats de ces enquêtes internationales, d'analyses comparatives internationales des curriculums, en mathématiques et en sciences, et de recherches sur les activités réelles des élèves pendant des évaluations standardisées ».

Les avis successifs du Comité de suivi de la Loi de refondation de l'école, instauré pour en suivre l'application, témoignent d'une désillusion certaine et d'une attente déçue, fondée sur une incompréhension de l'orientation prise par l'instance.

Le comité de suivi de la loi d'orientation et de programmation pour la refondation de l'école de la République du 8 juillet 2013, a rendu deux avis sur la mise en place du CNESEO dans ses deux rapports au Parlement en 2015 et 2016 :

- en 2015 : « *Le CNESEO, par les synthèses qu'il réalise des études et évaluations existantes et une politique de communication et d'animation très active, contribue à la diffusion des connaissances sur le système éducatif auprès des acteurs. Il est attendu maintenant qu'il engage une réelle évaluation du système éducatif, de ses résultats et de son efficacité, par une activité d'expertise des évaluations existantes et de production de cette expertise. L'objectif est également de bâtir les outils d'une diffusion des compétences d'évaluation à l'intérieur du système éducatif, susceptibles de devenir partie intégrante de la culture professionnelle de chacun des personnels et à chaque niveau de l'action publique* »

- en 2016 : « *D'une manière générale, les études consistent en des compilations d'études existantes, auxquelles s'intègrent parfois des enquêtes dites « inédites » (comme celle auprès d'étudiants dans l'étude sur l'attractivité du métier d'enseignant) et des contributions scientifiques nouvelles, notamment dans le cadre des conférences. Les données de la DEPP et de l'OCDE sont massivement utilisées. L'hypermédiatisation des synthèses de recherche est sans doute utile à la diffusion des questions sur l'éducation dans la société, mais est-elle utile à l'évolution et l'évaluation du système éducatif, comme le demande la loi, telle est la question que pose le Comité de suivi* ».

⁹ Programme international pour le suivi des acquis des élèves, désigné par l'acronyme PISA, Programme for International Student Assessment.

¹⁰ Programme international pour le suivi des acquis en sciences, désigné par l'acronyme *Trends in International Mathematics and Science Study*.

¹¹ Rapport Comparaison des évaluations PISA et Timss, novembre 2016.

De façon générale, l'objectif politique de « recréabiliser » l'évaluation, notamment par l'affermissement des méthodes, ne paraît pas atteint. Le CNESCO pour la seconde partie de son mandat (2017-2020) aurait prévu de s'investir davantage sur les questions méthodologiques.

C - Un malaise permanent sur la reconnaissance d'une instance indépendante chargée de l'évaluation du système éducatif

Le sentiment d'insatisfaction n'est ni nouveau, ni propre au mode de fonctionnement de l'instance actuelle. De multiples raisons expliquent ce malaise récurrent et l'instabilité constatée.

a) Le statut de l'institution chargée de l'évaluation est un sujet en débat perpétuel.

Dans son dernier rapport, le HCE avait appelé, faisant ainsi écho à la suggestion de la Commission du débat national de 2003, à la création d'une instance indépendante, assurant la transparence des travaux évaluatifs, capable de s'autosaisir, ayant accès sans restriction à l'ensemble des bases de données et des évaluations menées par les services du ministère de l'Éducation nationale. Elle aurait eu le statut d'autorité administrative indépendante (AAI).

Pourtant deux ans plus tard, la loi qui crée le CNESCO en 2013 ne lui accorde pas ce statut. La loi du 20 janvier 2017, qui fixe le statut des autorités administratives indépendantes, n'a pas retenu non plus le CNESCO dans la liste annexée des instances répondant aux critères requis. En effet ni le critère d'autonomie juridique affirmée, ni celui de détention de pouvoirs prescriptifs ne sont remplis. À l'inverse de l'instance chargée de l'évaluation dans les domaines de l'enseignement supérieur et de recherche, le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres), dont la loi fondatrice avait prévu ce statut, le CNESCO ne détient pas un pouvoir direct et obligatoire d'évaluation alors que les établissements d'enseignement supérieur et de recherche sont soumis à celui du Hcéres.

Aussi la question du statut de ces instances est demeurée ouverte, laissant l'impression d'une démarche institutionnelle incertaine et inachevée. L'hypothèse d'une agence de l'évaluation a pu aussi être évoquée.

Ce rapport reviendra sur la question institutionnelle quand il abordera dans le dernier chapitre les conditions nécessaires à une architecture ordonnée de la fonction d'évaluation.

b) Les moyens n'ont jamais été accordés à hauteur de l'enjeu

Les moyens accordés à ces instances ont toujours été mesurés et non jamais été individualisés dans la présentation du budget du ministère de l'éducation nationale, ni par une action budgétaire propre, encore moins par une dotation dont un responsable de l'instance eût été l'ordonnateur de fait, même si le responsable de programme demeurait l'ordonnateur de droit.

Les hauts conseils successifs n'ont disposé qu'une structure d'animation restreinte. Le CNESCO quant à lui disposait en 2016 et 2017 de sept emplois et de crédits, y compris pour

les études, de l'ordre de 350 000 euros. Ces moyens étaient intégrés sans distinction dans l'action « évaluation et contrôle » du programme 214 « soutien de la politique de l'éducation nationale ». Le conseil national n'avait accès aux données statistiques du ministère que par la voie d'une convention (octobre 2014) signée avec la direction compétence (DEPP).

c) Les conditions de la collaboration avec les services du ministère de l'éducation nationale ont toujours été imprécisées

Les articulations entre ces instances et les administrations du ministère chargées de l'évaluation n'ont jamais été précisées, lors de leur instauration, nécessitant de gérer au jour le jour leurs relations, ce qui a été la cause de nombreuses frictions. Seul le HCéé semblait avoir trouvé sa place en intervenant sur les méthodes de l'évaluation pratiquées par le ministère.

Une des recommandations du dernier rapport du Comité de suivi de la Loi de refondation de l'école (février 2017) exprime l'impasse actuelle, qui ne se réduit pas à une appréciation du bilan du CNESCO mais traduit une attente insistante pour combler un manque récurrent, qui est d'ordre systémique et structurel : « Situer le CNESCO dans le paysage global institutionnel de l'évaluation et par rapport aux instances d'évaluation qui existent à l'intérieur de ce système et à l'extérieur ».

Au total, malgré trois tentatives, les instances chargées de l'évaluation du système éducatif n'ont pas pu ou su concilier une fonction interne de conception, d'accompagnement et de stimulation de l'évaluation au sein du système éducatif et une fonction externe de transparence et de reddition des résultats aux différentes parties intéressées du système scolaire.

Les options prises par les autorités politiques, n'ont jamais été poussées jusqu'à leur logique ultime, tant aux plans institutionnel, fonctionnel que matériel. Il est fort probable qu'une forme de réticence, confinant parfois à la résistance, de la part des administrations centrales du ministère, ont freiné la mise en place de ces instances supérieures, les ministres successifs n'ayant pas en outre envisagé une refonte des compétences de leurs administrations pour tenir compte de leur création.

II - De nombreux services ministériels participant à l'évaluation, sans cohérence d'ensemble

De multiples administrations du ministère participent de la fonction évaluative, avec des légitimités diverses, selon des angles différents, en intrication avec d'autres missions et en interaction entre elles. Certaines directions peuvent être acteur et sujet de l'évaluation, ce qui n'est pas de bonne pratique.

A - Les inspections générales : des missions partiellement axées sur l'évaluation, une confusion des rôles pour l'inspection générale de l'éducation nationale

Les deux inspections générales du MENESR exercent des responsabilités en matière d'évaluation. L'inspection générale de l'éducation nationale (IGEN), qui compte un peu plus de 150 inspecteurs généraux, « a pour mission d'évaluer les types de formation, les contenus d'enseignement, les programmes, les méthodes pédagogiques, les procédures, les moyens mis en œuvre (et prend part à celle des) personnels d'inspection, de direction, d'enseignement, d'éducation et d'orientation ». L'inspection générale de l'administration de l'éducation nationale et de la recherche (IGAENR) qui compte un peu moins de 90 inspecteurs généraux évalue « l'efficacité et les performances du système éducatif à tous les niveaux d'enseignement ». L'une comme l'autre « procèdent, en liaison avec les services administratifs compétents, à des évaluations départementales, académiques, régionales et nationales ».

La part des travaux évaluatifs dans les activités des deux inspections générales depuis 2010 peut être estimée à environ 15 % pour l'IGEN et 6 % pour l'IGAENR. Elles sont fréquemment associées dans des travaux communs : sur les 31 rapports explicitement évaluatifs de l'IGEN, la moitié a été réalisée en collaboration avec l'IGAENR, sur les 25 rapports explicitement évaluatifs de l'IGAENR, 11 ont été réalisés en collaboration avec l'IGEN. Les évaluations des inspections générales relèvent plutôt du « dire d'experts » et comportent très rarement des estimations chiffrées, issues des méthodes évaluatives dites quantitatives.

La dualité des inspections est une particularité du ministère, qui résulte de son histoire. Pourtant la spécificité des compétences de l'IGEN, essentiellement de nature disciplinaires, ne constitue pas un argument décisif de la dualité des corps. Ainsi l'inspection générale des affaires sociales (IGAS) accueille-t-elle aussi bien des administrateurs civils que des médecins. Il ne saurait évidemment être question de confier à des administrateurs civils par exemple, des missions pédagogiques ou disciplinaires, mais la spécialisation des agents n'est pas incompatible avec l'unicité de leur statut dans un corps d'inspection unique.

Une autre singularité réside dans la dualité des missions de l'IGEN. Elles participent, pour nombre d'entre elles, à la gestion directe du système : élaboration des programmes et des réformes pédagogiques, validation de sujets d'examen, gestion de ressources humaines (choix et inspection des professeurs de classes préparatoires, présidence de concours de recrutement, gestion des agrégés, etc.), collaboration aux côtés de la DGESCO et de la DEPP à la création de nombreux dispositifs éducatifs et pédagogiques. Ainsi cette inspection générale est partie intégrante de la gestion et du pilotage du système éducatif, ce qui rend délicate sa position d'évaluateur de sa performance. Une clarification du rôle de l'IGEN serait bienvenue.

Ces deux particularités ont justifié qu'une lettre de mission des deux ministres chargés de l'éducation nationale et de l'enseignement supérieur et de la recherche, soit adressée aux deux chefs de corps le 28 août 2017, pour leur demander d'étudier « les modalités d'un exercice renouvelé de la mission d'inspection » au sein du système éducatif.

B - La direction de l'évaluation, de la prospective et de la performance (DEPP) : entre service statistique et acteur de l'évaluation

La DEPP assume deux grandes fonctions : celle de service statistique ministériel (arrêté du 28 octobre 2016), celle de service d'études, dont certaines peuvent être évaluatives. Elle demeure plutôt une direction chargée de produire, de diffuser et d'étudier des données statistiques.

1 - Un exercice de la mission statistique heureusement enrichi

L'activité de la direction est centrée sur la publication annuelle de documents statistiques de référence : « *Repères et références statistiques sur les enseignements, la formation et la recherche* », « *L'éducation nationale en chiffres* » et « *L'état de l'école* ». Ces documents regroupent une trentaine d'indicateurs, dont certains dits « *indicateurs annoncés* », sont diffusés chaque année à date fixe tels que les prévisions d'effectifs scolaires à deux ans (avril), le pourcentage de bacheliers dans une génération (juillet), le constat des effectifs scolaires (décembre), la dépense intérieure d'éducation (novembre). Le programme annuel de la DEPP est rendu public sur le site du ministère.

Elle est le correspondant pour le ministère de l'éducation nationale des organisations internationales, qui conduisent des études comparées sur les résultats et les modes de fonctionnement des systèmes éducatifs.

La DEPP a progressivement élaboré des indicateurs qui ont enrichi son travail statistique et rehaussé la qualité des instruments d'évaluation de l'enseignement scolaire. À titre principal, il s'agit depuis 1994 d'un indicateur de valeur ajoutée des lycées (IVAL) qui mesure, par établissement, la différence entre les résultats obtenus et les résultats attendus, compte tenu des caractéristiques scolaires et socioprofessionnelles des élèves. Il s'agira demain d'un indicateur pour les collèges (ICAL).

Elle a mis en place depuis 2003 un cycle national d'évaluations-bilans par échantillon pour mesurer les performances des élèves par rapport aux objectifs fixés par les programmes scolaires (Cycle des évaluations disciplinaires réalisées sur échantillons, Cèdre). Elle a mis en place depuis 2015 des tests standardisés de compétences des élèves, d'abord expérimentés puis généralisés en classe de 6^e depuis la rentrée scolaire 2017 (*cf. infra*)

Enfin, progressivement la DEPP a ouvert ses données aux acteurs de la recherche, qu'elle accompagne également par son appui méthodologique dans leur utilisation.

2 - Une intervention dans les processus évaluatifs plus en retrait, en raison de sa position administrative

La DEPP est le dépositaire d'un savoir-faire statistique et d'un trésor de données sur le système scolaire, primordiaux pour asseoir sur des bases solides la fonction évaluative. La valorisation de cette richesse s'est améliorée, produisant des indicateurs de performance.

Son action proprement évaluative est cependant en retrait. Les textes lui confèrent des missions qui en font plus un contributeur qu'un acteur premier de la fonction d'évaluation :

« Elle contribue à l'évaluation des politiques conduites par le ministère de l'éducation nationale (et)... conçoit et met à disposition de l'ensemble des acteurs du système éducatif des outils d'aide à l'évaluation, au pilotage et à la décision ».

Aussi cette direction lance peu d'opérations strictement évaluatives sur des politiques et dispositifs éducatifs précis. En 2016, trois expérimentations seulement ont été inscrites à son programme et sont la poursuite de travaux en cours. Pour l'essentiel, ces études sont de nature qualitative, la DEPP n'ayant jamais lancé d'évaluations selon des modalités d'investigations quantitatives, certainement faute de moyens budgétaires. Les enquêtes faisant appel à ces méthodes, au coût élevé, ont toujours été supportées par des financeurs extérieurs, le Fonds d'expérimentation pour la jeunesse et les Programme d'investissements d'avenir. La DEPP peut être selon les cas plus ou moins associée : elle l'est de plus en plus comme en témoigne son rôle central pour l'évaluation des initiatives numériques dans le domaine éducatif¹². A l'heure actuelle, la conduite par la DEPP de l'évaluation du dédoublement des cours primaires en éducation prioritaire, qui est une avancée (*cf. infra*), n'en reste pas moins un compromis méthodologique entre suivi opérationnel et mesure d'impact.

Ce relatif retrait peut s'expliquer de diverses façons.

D'une part sa mission première est celle d'un service ministériel statistique et ses moyens sont proportionnés à cet objectif. D'autre part elle reste une direction d'administration centrale qui, hors les garanties accordées à un service statistique, ne peut ni arrêter librement les axes de son programme d'évaluation, ni décider seule de l'opportunité de la diffusion des résultats de ses études. En effet, si l'indépendance de la DEPP est protégée pour la production statistique, en revanche, pour ses activités d'étude, elle ne bénéficie pas juridiquement de la même indépendance. Sa protection n'est qu'indirecte : il faudrait que l'Autorité de la statistique publique estime qu'une intervention extérieure (ministérielle par exemple) sur ses travaux d'études soit susceptible d'altérer son indépendance dans le domaine statistique. Cette position ambiguë peut expliquer que si peu de politiques ou dispositifs éducatifs fassent l'objet d'évaluation de sa part.

En l'état actuel, les textes concernant l'organisation du ministère ne prévoient pas que cette direction soit dépositaire de la fonction d'évaluation, son statut administratif ne lui confère pas dans ce domaine précis l'indépendance nécessaire. Cette question n'a pas évolué depuis le rapport de la Cour de 2010 (*cf. supra*) faisait état des mêmes constats.

¹² Programme d'évaluation de l'action « initiative numérique pour l'excellence éducative » proposé par la DEPP validé par le comité de pilotage des PIA le 8 mars 2017.

C - La direction générale de l'enseignement scolaire (DGESCO) : une position à clarifier entre prescripteur pédagogique et acteur de l'évaluation

La direction générale de l'enseignement scolaire élabore la politique éducative et pédagogique, assure la mise en œuvre des programmes d'enseignement et définit la politique relative à l'éducation prioritaire dont elle « anime et évalue la mise en œuvre »¹³.

La DGESCO est le prescripteur principal des dispositifs d'évaluation des élèves (contrôles continus, cadre des notations, contenu des examens), même si elle ne peut conduire ce travail qu'avec l'appui de l'IGEN. Le rôle de l'IGEN, majeur dans ce domaine, illustre l'intrication complexe entre ces deux entités et la confusion des rôles possible. D'une part l'IGEN participe à l'élaboration des dispositifs évaluatifs des niveaux des élèves, d'autre part elle émet des analyses critiques sur ces dispositifs.

Les actions conduites par la DGESCO s'apparentent plutôt à la mise en place de contrôles de gestion et au suivi opérationnel de la mise en application de réformes et dispositifs éducatifs.

Pour autant la frontière entre ces leviers de gestion et le pilotage de processus d'évaluation est ténue. Par des enquêtes demandées aux académies, la DGESCO participe non seulement à la commande d'évaluation, mais la dirige. Il s'agit le plus souvent de dispositifs limités, par exemple l'évaluation du dispositif « *la mallette des parents- 2012-2013* » ou celui des « *cordées de la réussite* », mais parfois aussi de politiques à fort enjeu. Le comité national de suivi du dispositif « *Plus de maîtres que de classes* » (PDMQDC), chargé de définir ses modalités d'évaluation, le comité de suivi des nouveaux rythmes scolaires à l'école primaire, ayant pour mission d'en « *identifier les réussites et les points d'amélioration (et) de valoriser les bonnes pratiques* », ont été mis en place par la DGESCO et travaillent sous son égide. Quant à la politique majeure de l'éducation prioritaire, la position de la DGESCO est ambiguë : elle n'en assurerait que le suivi opérationnel grâce à la remontée de nombreuses données qu'elle demande aux rectorats de transmettre. En pratique, elle détient les clés du dispositif d'évaluation d'une politique primordiale pour répondre aux critiques sur l'inégalité du système scolaire français.

En outre, plus la nécessité d'intégrer l'évaluation comme paramètre de pilotage du système éducatif s'accroît, plus la question de l'articulation de la DGESCO avec l'IGEN et la DEPP se posera. Il sera, par exemple, indispensable de traiter ce sujet avec une nouvelle génération de contrats entre l'administration centrale et les académies, adossés à une démarche d'auto-évaluation des académies, et suivie d'une évaluation externe à laquelle participeront les inspections générales. L'IGEN ne pourra pas alors être associée en amont à l'élaboration de ces contrats.

En l'état actuel, deux clarifications sont nécessaires.

Créatrice de nombreuses normes scolaires, opérateur de maints dispositifs éducatifs, la DGESCO a la responsabilité légitime d'assurer leurs suivis par des retours d'expérience

¹³ Décret n° 2014-133 du 17 février 2014 fixant l'organisation de l'administration centrale des ministères de l'éducation nationale et de l'enseignement supérieur et de la recherche.

opérationnels. Or les frontières entre suivis opérationnels et dispositifs d'évaluation n'ayant pas été précisés au sein du ministère, la DGESCO se trouve alors dans la situation d'être l'évaluateur de ses prescriptions, quand elle estime nécessaire d'en approfondir l'observation. Sa position centrale au sein du MEN lui permet non seulement de prendre ce type d'initiatives, mais aussi de les piloter alors qu'il pourrait revenir à la DEPP de le faire.

Les rapports entre la DGESCO, l'IGEN et la DEPP devraient être revues, pour éviter une confusion entre apports opérationnels de celle-ci à la définition et la gestion des politiques éducatives et évaluation de ces mêmes politiques.

D - Le conseil supérieur des programmes : un champ d'évaluation non exploré

Le conseil supérieur des programmes CSP a pour responsabilité principale d'émettre des avis et des propositions sur la conception générale des enseignements, la nature et le contenu des épreuves des examens conduisant aux diplômes nationaux. Le décret du 24 juillet 2013 dispose que le conseil élabore une « charte des programmes » qui présente ses principes de travail et définit son objet. Adoptée en avril 2014, celle-ci prévoit une « *évaluation régulière des programmes, de façon transparente, pour permettre à tous de savoir dans quelle mesure et comment ils sont effectivement mis en œuvre dans les classes, et avec quel bénéfice pour les élèves en termes d'acquis effectifs* ».

Le CSP n'a lui-même procédé à aucune évaluation de ces programmes, ce qui aurait nécessité des compétences et des moyens dont il ne dispose pas. Il a été précisé à la Cour que « de facto, le CSP n'a pas disposé d'évaluation explicite d'ensemble, mais de données qu'il a collectées, issues de différentes sources (Depp, Pisa, recherches universitaires) ». Aussi les propositions de modification de programme qu'il a formulées¹⁴ ne sont pas fondées sur l'évaluation de l'impact des programmes scolaires précédents.

Prenant acte du fait que les commandes ministérielles de refonte de programmes scolaires qu'il a reçues n'ont pas été précédées d'évaluations explicites, le CSP a souhaité dans son dernier rapport d'activité que les programmes soient régulièrement évalués selon des procédures à définir, en précisant le rôle des différentes instances (recherche universitaire, IGEN, DEPP, CNESCO, DGESCO...) et qu'il puisse faire état des résultats de ces évaluations dans son rapport annuel.

Lors de la présentation des nouveaux programmes de la scolarité obligatoire devant le conseil supérieur de l'éducation (CSE) en octobre 2015, il a été décidé de mettre en place dès 2016 un comité de suivi ayant notamment pour objet, « dans le respect des missions de chaque instance (IGEN, DGESCO, DEPP, ESPE) » et dans le cadre d'un programme de travail défini en commun, de « préconiser une éventuelle évolution des textes programmatiques au regard de leur pertinence, de leur degré de mise en œuvre effective, des effets sur les apprentissages des élèves, des attentes de la communauté éducative ».

¹⁴ Par exemple pour l'école élémentaire et le collège (cycles 2-3-4), le socle commun de connaissances, de compétences et de culture, l'école maternelle, l'enseignement moral et civique, le parcours d'éducation artistique et culturelle.

III - Des acteurs externes contribuant à l'évaluation de l'éducation nationale

Ce rapport ne recense pas les entités universitaires et de recherche qui peuvent concourir à l'évaluation du système scolaire. Il observe cependant que le recours à des organismes de recherche dans les processus évaluatifs sont rares, souvent plus du fait de prescripteurs extérieurs de l'évaluation tels que le Fonds d'expérimentation pour la jeunesse FEJ que du ministère lui-même.

A - Le comité de suivi de la Loi de refondation de l'École : l'évaluation comme un des critères d'application de la Loi

Le comité de suivi de la loi d'orientation et de programmation pour la refondation de l'école de la République, déjà mentionné, est « chargé d'évaluer l'application de la présente loi » en vertu de l'article 88 de celle-ci. Ce comité a vocation à porter une appréciation critique (assortie de recommandations) sur le degré d'effectivité des principales mesures prévues par la Loi de refondation de l'école. Pour remplir sa mission, il a défini 10 critères de suivi de l'application des dispositions de la loi, classés selon l'échelle « appliqué », « cours d'application », « différé », et dans lesquels figure « l'intégration de l'évaluation dans l'application de la loi ».

L'insertion de l'évaluation est ainsi un des critères renseigné pour toutes les dispositions prévues par la Loi de refondation de l'école. Il est le plus souvent classé dans la catégorie « en cours d'application », par exemple en 2016 pour le service public du numérique éducatif, la priorité au premier degré, la réforme de la formation initiale des enseignants.

En ce qui concerne la mise en place du CNESCO, les items « Mise en mouvement des pratiques » et « Logique de mise en œuvre » sont classés comme « différés » dans les rapports de 2015 et 2016, les autres items qui portent sur la mise en place opérationnelle du CNESCO (« affectation des moyens », « organisation nationale et locale », « communication », « perception du changement ») ont été classés en cours d'application ou appliqués. Quant au critère portant sur l'insertion de l'évaluation (En pratique l'auto-évaluation du CNESCO), il a été classé comme « appliqué », en raison de la publication du rapport d'activité.

Le comité de suivi, même s'il n'a ni vocation, ni moyens d'investigation pour apprécier l'impact de la Loi sur les résultats des élèves, est un puissant catalyseur de l'évaluation en vérifiant si celle-ci est intégrée à l'application de chacune des mesures prévues.

B - Le secrétariat général pour la modernisation de l'action publique : une visée autre qu'évaluative

Le SGMAP « coordonne, favorise et soutient, au niveau interministériel, les travaux conduits par les administrations en vue d'évaluer et de moderniser l'action publique¹⁵ ».

Si ce service du premier ministre n'est pas spécifiquement chargé de l'évaluation des politiques éducatives, leur importance dans les politiques publiques l'a conduit à s'y investir. Il a notamment coréalisé avec la DGESCO des évaluations « partenariales » sur la lutte contre le décrochage scolaire, la politique d'éducation prioritaire, la scolarisation des enfants en situation de handicap et la médecine scolaire ou la formation des enseignants du primaire.

La spécificité des évaluations « partenariales » est d'associer étroitement les administrations porteuses de ces politiques pour susciter leur implication et s'assurer que les conclusions de l'évaluation, concertées avec elles, seront suivies d'effet. Ces arguments, qui relèvent d'un souci d'opérationnalité, ne sont pas convaincants au plan méthodologique car le concepteur d'une politique participe à son évaluation, même si c'est au stade final, et peut en orienter le sens en anticipant par exemple ses conséquences au plan budgétaire. Le SGMAP fait valoir que l'approche partenariale des travaux du SGMAP (avec les administrations concernées), qui pourraient être qualifiés de bilans d'étape ou de retours d'expérience, correspond aux standards de l'évaluation des politiques publiques, qui recommande la pluralité des parties prenantes.

Quant aux données nécessaires à ces travaux, le SGMAP a estimé que ses équipes ont été confrontées à des difficultés liées à « la qualité, l'exhaustivité et la disponibilité des données quantitatives » dont disposaient les responsables des politiques de lutte contre le décrochage scolaire et de l'éducation prioritaire. Sur celle-ci, a été mentionnée une forte limitation : « L'indisponibilité des données a eu des conséquences sur la capacité de la mission d'évaluation à mesurer l'impact de la politique faute de liens de causalité établis ».

Pour autant le SGMAP n'a pas tenté de mesurer lui-même les effets directs de cette politique, considérant que sa mission est d'animer et de coordonner les moyens et travaux existants.

¹⁵ Idem.

CONCLUSION

Le recensement des organes qui concourent à l'évaluation du système éducatif, croisé avec celui de leurs compétences et de leurs approches, décrit un ensemble insuffisamment ordonné et articulé. Des champs restent inexplorés comme celui de l'impact des programmes scolaires, dont les réformes sont décidées sans évaluations préalables.

Les compétences des administrations concernées font presque toutes référence à l'évaluation, sans en préciser nettement la portée ou leur laissant une grande latitude d'interprétation et d'initiative.

Leurs modes d'intervention, qu'il s'agisse des rapports d'inspections, des études et travaux à vocation évaluative, des productions statistiques et de leur valorisation enrichie, des suivis opérationnels de mise en place de politiques et dispositifs éducatifs, se confondent car mises sur le même plan. Ces modes d'appréciation, qui ont des vocations diverses et légitimes quant à leur objet propre, ne sont ni classés, ni hiérarchisés, pour nourrir un processus évaluatif d'ensemble. La responsabilité de valider la méthodologie de ses différents travaux n'est pas attribuée, ni exercée.

L'absence d'une instance chargée de l'évaluation du système scolaire solidement reconnue dans son rôle explique certainement cette profusion d'intervenants et cette confusion dans la qualification de leurs travaux.

Celles qui se sont succédé, n'ont pas été mises, dans leur positionnement, leurs attributs et leurs moyens, en état de s'inscrire durablement et fortement dans une architecture d'ensemble, qui est restée inchangée lors de leur création. Aussi ces instances ont été plus juxtaposées qu'intégrées.

Chapitre II

Des modes d'évaluation sans unité de vision

Le système éducatif a besoin de disposer d'appréciations critiques sur des champs divers selon des outils adaptés: les acquis et les compétences des élèves, la qualité de l'instruction donnée par les enseignants, la performance des personnels de direction et des établissements scolaires, l'impact des dispositifs pédagogiques, les enseignements à tirer des expérimentations.

Au-delà des observations sur la stabilité de ces outils, leur portée plus ou moins évaluative, les garanties méthodologiques prises pour assurer leur fiabilité, la question de savoir s'ils forment un corpus cohérent de la mesure de la performance du service éducatif, est centrale. À défaut les efforts déployés risquent de perdre leur visée d'ensemble et de n'apporter aux décideurs publics qu'une image parcellaire et floue.

I - Une évaluation des acquis des élèves trop discontinuée pour être utilisée

Depuis longtemps, des hésitations nombreuses ont marqué les modalités d'évaluation des élèves en France alors qu'elle constitue la mesure de la finalité du système éducatif : évaluer ce qui est transmis et acquis en connaissances et en compétences, au moins durant la scolarité obligatoire.

A - Une boîte à outils au contenu très changeant

Deux catégories d'évaluation des niveaux scolaires sont distinguées, l'une dite formative, réalisée lors de l'enseignement pour déterminer le niveau atteint, identifier les difficultés et définir les prochains objectifs, l'autre dite sommative, dans le but de mesurer et de valider les acquis. Si la première est le plus souvent réalisée en classe dans le cadre de l'autonomie pédagogique des enseignants, la seconde, notamment lorsqu'elle est certificative, prend la forme d'examens dont le diplôme national du brevet (DNB) et le baccalauréat sont les exemples les plus connus.

Comme les grandes évaluations certificatives n'ont pas fondamentalement évolué et sont limitées à deux rendez-vous (le diplôme national du brevet DNB et le baccalauréat), il est devenu indispensable de disposer d'une information intercalaire sur les niveaux acquis

(évaluation sommative), notamment de vérifier l'atteinte progressive du socle commun de connaissances par le plus grand nombre. Plusieurs dispositifs peuvent être mobilisés à cet effet :

- des évaluations nationales, dites « diagnostique » ou « bilan », qui fournissent des résultats utilisables également aux niveaux des classes, dès l'établissement ou des académies ;
- des évaluations sur échantillon qui produisent des données dont il est possible d'inférer les résultats pour l'ensemble de la population étudiée, avec une marge d'erreur liée notamment à la taille de la population testée. Ces évaluations ne sont pas exploitables localement.

Il en résulte un champ très ouvert d'évaluations possibles, plus ou moins contextualisées (prises en compte des caractéristiques socio-professionnelles des parents), portant ou sur des disciplines, ou sur l'ensemble des connaissances et compétences liées à des cycles scolaires, selon des périodicités diverses (rythme annuel ou pluriannuel des enquêtes).

Dans un tel contexte méthodologique, la stabilité du choix d'un type d'évaluation est indispensable pour forger une vision crédible du niveau des élèves. Cette stabilité a fait défaut en France.

1 - Une instabilité des mesures nationales, diagnostiques et de bilan, qui témoigne d'une doctrine incertaine du ministère

Alors qu'au diagnostic est associée l'idée de début de séquence, les bilans sont généralement réalisés en fin ou en milieu pour mesurer les effets de l'instruction.

Réalisées chaque année de 1989 à 2006 en début de CE2 et de 6^e, les premières évaluations diagnostiques nationales furent ensuite remplacées par des évaluations en début de CE1 et de CM2, puis, à compter de 2009, par des évaluations bilan à la fin de ces dernières classes, évaluations dont l'ambition était à la fois de situer les acquis de chaque élève par rapport aux objectifs de fin de cycle et de permettre aux enseignants d'apporter une aide personnalisée aux élèves qui en avaient besoin. En complément de l'utilisation de ces évaluations par l'enseignant de CE1 ou de CM2, une synthèse départementale, académique puis nationale des données avait pour fonction de fournir aux responsables de ces différents niveaux un instrument de pilotage. Poursuivant un double objectif d'évaluation-bilan et d'évaluation diagnostique, le caractère obligatoire de ces évaluations a été supprimé pour la rentrée 2013.

Par ailleurs, des évaluations nationales de fin de 5^{ème} ont été organisées en 2012 et 2013 puis suspendues. A la rentrée 2015, il a été mis fin à ces évaluations nationales. Elles ont été remplacées par des évaluations en début de CE2, en principe obligatoires, mais réalisées à la seule initiative des enseignants, « au moment choisi par eux et en fonction des objectifs poursuivis au sein de la classe »¹⁶ c'est-à-dire sans possible remontée nationale des données. La circulaire de rentrée 2017 ne les a plus mentionnées.

¹⁶http://www.education.gouv.fr/cid262/evaluation-des-acquis-des-eleves-ce1-cm2.html#Evaluer_le_niveau_de_chaque_eleve_en_francais_et_en_mathematiques_au_debut_de_la_classe_de_CE2_pour_identifier_ses_besoins_et_y_apporter_des_reponses_adaptees

2 - Les évaluations par échantillon, des données parcellaires sur les acquis disciplinaires

Des évaluations sur échantillon dites Cédre (Cycle des évaluations disciplinaires basés sur les programmes d'enseignement) sont réalisées depuis 2003 pour mesurer les acquis des élèves selon les 3 cycles du socle commun de connaissances (de l'école à la fin du collège). Il couvre la plupart des domaines disciplinaires : maîtrise de la langue (école), compétences générales et langagières (collège), langues vivantes étrangères, histoire-géographie et éducation civique, sciences expérimentales, mathématiques. Ces évaluations situent les performances des élèves sur des échelles de niveau, selon des intervalles de 6 ans (bientôt de 5 ans). Ces résultats sont disponibles par disciplines et académies pour la communauté enseignante.

Les évaluations nationales Cédre ne représentaient jusqu'en 2015 qu'environ 45 000 élèves, pour des items limités et des champs différents, selon des périodicités diverses, sur un effectif d'élèves de l'ordre de 7,5 millions (écoliers et collégiens).

Quelle que soit la valeur statistique des échantillons utilisés, ces évaluations s'étendent sur des périodes allant de 3 à 6 ans (bientôt 5 ans) avant de pouvoir donner une vue d'ensemble des niveaux scolaires pour le socle commun de connaissances.

Sont également réalisées depuis 2014 des évaluations sur échantillon sur « la maîtrise des éléments essentiels aux compétences du socle commun de connaissances ». Elles ont lieu selon un cycle de 3 ans par palier d'acquisition du socle commun de connaissances (CE1, CM2, 6^e). Elles ne portent que sur certaines compétences attendues de l'acquisition du socle commun de connaissances (maîtrise de la langue, principaux éléments de mathématiques).

3 - L'amorce de la reprise d'une évaluation généralisée sur les niveaux de compétences

En 2015, pour la première fois, la DEPP a mis en place une évaluation standardisée par ordinateur sur un échantillon beaucoup plus large (160 000 élèves) pour mesurer le niveau de compétences du socle commun en 6^e. Elle n'a concerné que deux compétences (langue française et mathématiques).

À la rentrée 2017, le ministre a généralisé ce dispositif à tous les élèves de sixième. L'objectif de cette évaluation nationale est de permettre selon le ministère : « aux équipes pédagogiques de disposer d'un diagnostic standardisé des compétences de chaque élève et d'accompagner le pilotage pédagogique dans les établissements. Au niveau académique, ce dispositif a pour visée d'enrichir et de compléter les outils de pilotage déjà existants ». Les collèges devraient disposer de façon quasi immédiate des résultats individuels de chacun de leurs élèves selon quatre degrés de maîtrise, la remontée nationale sera, elle, anonymisée. L'évaluation sera entièrement réalisée sur support numérique entre novembre et décembre 2017.

Par ailleurs une évaluation généralisée de compétences en cours préparatoire a été réalisée lors de la rentrée 2017, mais sur support papier et corrigée par les enseignants.

Il ressort une grande instabilité du système de recueil des données sur le niveau des élèves. Ces inflexions continuelles sont dues à l'absence d'un consensus sur la nécessité de mesurer hors la classe, hors la tradition des évaluations formatives et certificatives. Elles sont aussi les conséquences des craintes des enseignants quant à leur utilisation par l'institution scolaire.

Ces résistances ont pour effet un système d'évaluation des élèves composé d'une myriade de protocoles, dont la cohérence d'ensemble est mal assurée et les résultats beaucoup moins lisibles que ceux des grandes évaluations internationales.

B - Des résultats peu conciliables entre plusieurs dispositifs d'évaluation de la maîtrise du socle commun de connaissances

1 - Des estimations contradictoires du niveau de maîtrise du socle commun

Le socle comportait, jusqu'à la rentrée 2016, 7 compétences, 26 domaines et 98 items qui pouvaient être chacun « acquis, non acquis, voire en cours d'acquisition ou partiellement acquis », sans compensation entre les compétences. Deux exemples éclairent la difficulté de concilier les données disponibles pour établir une opinion solide sur la maîtrise de ce niveau scolaire.

- Le premier indicateur de performance des programmes budgétaires 139 (enseignement privé) et 141 (second degré de l'enseignement public) est la « proportion d'élèves maîtrisant en fin de troisième les compétences 1 et 3 du socle commun », soit les maîtrises de la langue française et des principaux éléments de mathématiques.

Selon les estimations fournies par l'indicateur, les taux d'atteinte sont compris entre 50 % et 80 %. Cependant, en ce qui concerne la compétence n°2 (« pratique d'une langue vivante étrangère »), les estimations disponibles font état d'une maîtrise par seulement 40 % des élèves à l'écrit et 26 % à l'oral. Dès lors qu'il n'y a pas de compensation entre les différentes compétences, le socle ne serait maîtrisé que par un quart des élèves au maximum ce que ne traduit pas l'indicateur retenu dans les documents budgétaires.

- Le taux de réussite au Diplôme national du brevet (DNB) qui clôt ce cycle d'enseignement, est nettement supérieur à la fourchette comprise entre 50 % et 80 % depuis près de 10 ans. Selon les documents budgétaires de 2016, « toutes séries confondues, 91 % des élèves en classe de troisième présents au diplôme national du brevet (DNB) ont validé l'ensemble des sept compétences en 2013, soit 0,7 point de plus qu'à la session 2012 ». L'explication des écarts avec les données précédentes tiendrait au fait que « le taux de 91 % de candidats ayant validé les sept compétences correspond à une évaluation continue au cours de la scolarisation au collège » alors que « le renseignement de l'indicateur repose sur des évaluations ponctuelles et standardisées ». Il s'agit cependant d'écarts significatifs, de plusieurs dizaines de points.

Apparaissent ainsi des contradictions majeures entre plusieurs mesures d'une même réalité, le degré de maîtrise du socle commun de connaissances qui clôt les cycles de l'enseignement obligatoire.

2 - Un instrument de consolidation à affermir : le livret scolaire unique

Les évaluations de niveau et de compétence effectuées par les enseignants en classe sont collationnées dans le livret personnel de compétences (LPC), devenu Livret scolaire unique (LSU), informatisé pour le collège depuis 2011, et renseigné par les chefs d'établissement en concertation avec l'équipe pédagogique.

L'analyse des bases informatiques pour les années 2014 et 2015 (Annexe V) montre une très variable complétude des données : ce ne sont en moyenne que 31¹⁷ des 98 items prévus qui sont renseignés, dans des proportions très différentes selon les académies : 22 pour celle de Rennes, 46 pour celle de Strasbourg. La même dispersion se retrouve pour les établissements.

Le ministère estime que cette situation devrait s'améliorer : d'une part le déploiement du LSU dans les établissements scolaires est désormais achevé, d'autre part des applications informatiques devraient permettre que se déversent dans le LSU à la fin de chaque période (trimestre, année) les évaluations réalisées. En réponse à la Cour, le ministère indique : « Cela assurerait la prise en compte des résultats des évaluations sommatives (connaissances et compétences) comme cela est recommandé par la Cour, pour un coût raisonnable ».

II - Les évaluations des personnels et des établissements : une contribution très faible à l'évaluation d'ensemble

L'évaluation des enseignants peut être envisagée de deux points de vue, celui de la part de leur notation liée à leur évaluation pédagogique et disciplinaire, à côté d'autres critères plus mécaniques tels que l'ancienneté, celui de l'interaction entre leur évaluation et celle du système éducatif. Les mêmes questions se posent pour le personnel de direction, dans le cadre incertain de l'autonomie des établissements scolaires

A - L'impact très réduit de l'évaluation pour différencier les évolutions de carrière

Les enseignants font l'objet d'inspections pédagogiques périodiques bien que les statuts des corps n'établissent pas de lien explicite entre l'inspection individuelle et l'attribution de la note, ni entre l'évaluation des enseignants et leur avancement d'échelon ou de grade.

¹⁷ Ce défaut de documentation traduit un défaut d'estimation précise et non un défaut de communication des estimations réalisées ; si les compétences précises étaient vraiment testées mais inégalement communiquées, il y aurait des établissements avec 100 % d'items renseignés et des établissements à 0 %. Or ce que l'on constate c'est que les établissements ne communiquent qu'une partie des items et il est dès lors vraisemblable que c'est parce qu'ils n'en renseignent qu'une partie et valident globalement les autres. De la même façon, les items renseignés varient d'un établissement à l'autre : l'appréciation des acquisitions est donc réalisée partiellement et sur des items variant selon les établissements.

Le dispositif d'évaluation des enseignants doit être réformé à partir de septembre 2017, en lien avec le protocole « *Parcours Professionnels, Carrières et Rémunérations* », sans que pour autant ses fondements (critères et modes d'évaluation) soient profondément modifiés.

1 - Les conséquences limitées de l'évaluation pour l'avancement et les rémunérations des enseignants

Les conséquences pour l'enseignant d'une bonne ou d'une mauvaise appréciation sont restreintes. Dans le système en passe d'être réformé, un enseignant exceptionnellement performant pouvait être promu au grand choix et accéder à la hors classe dès le 7^e échelon tandis qu'un enseignant très insuffisant serait toujours promu à l'ancienneté et, en principe, jamais à la hors classe. En réalité, aucun enseignant n'a été promu à la hors classe avant le 10^e échelon et presque tous les enseignants finissent par accéder à la hors classe.

Les positions administratives finales des enseignants en 2015 montrent que 85 % des professeurs certifiés accède à la hors classe, que 65 % des professeurs des écoles y accède aussi. Les carrières courtes sont les seules à ne pas atteindre la hors classe (Ce qui explique le taux pour les professeurs des écoles) quel que soit le mérite des personnes, témoignant de l'importance de l'ancienneté.

L'enjeu de l'appréciation de la performance est limité sur le plan pécuniaire. La comparaison des carrières extrêmes avec la carrière médiane montre des écarts maximaux de traitement perçu sur toute la vie, situés entre 17 % et 12 % selon les catégories, ce qui est faible par rapport aux autres agents publics de catégorie A (le fait que la part indemnitaire des rémunérations des enseignants, vecteur de différenciation, soit très modeste explique notamment cet écart). Enfin les écarts de rémunération ne sont plus affectés par les effets des évaluations après le passage au grade de la hors classe.

2 - Un rythme d'évaluation lié au déroulement de carrière plus qu'au besoin d'inspection et d'appui

La Cour a analysé les notes obtenues tout au long de leur carrière par les 13 886 enseignants partis à la retraite en 2015, à un âge moyen de 60,4 ans. Leur ancienneté moyenne à la date de leur retraite était de 35 ans (36 pour un professeur des écoles, 34 pour un agrégé). En moyenne ils avaient fait chacun l'objet de 5,2 notations, soit une tous les 7 ans (5 ans pour l'enseignement primaire). Dans le secondaire le nombre de notes varie considérablement d'une matière à l'autre, 5 à 6 notes par carrière pour l'enseignement professionnel contre 3 à 4 pour les langues et les matières générales.

L'essentiel des inspections est programmé en fonction de la gestion de l'avancement des personnels, c'est-à-dire centré sur ceux en situation d'être promus. Si cette sélection paraît logique en termes de gestion de corps, ne sont pas prioritairement inspectés ceux pour lesquels ce serait le plus nécessaire. Dans ces conditions, l'inspection systématique à intervalles irréguliers ou réguliers (*cf. infra sur la réforme prévue*) ne constitue pas un bon usage de l'expertise pédagogique des inspecteurs chargés du primaire et du secondaire (IEN et des IA/IPR).

Pour être utile, l'inspection devrait être en priorité réservée à l'évaluation-accompagnement des jeunes enseignants et des enseignants en difficulté, sur signalement par le chef d'établissement ou à leur propre demande.

3 - La réforme prévue de l'évaluation des enseignants : un risque de déconnexion accrue entre évaluation et carrière

La réforme projetée répond à certaines des critiques émises : le cadencement des inspections devient régulier (quatre rendez-vous de carrière), deux accélérations de carrière sont possibles (30 % de proposition au choix par l'administration), une évaluation qui serait désormais fondée sur une nouvelle grille de critères, une appréciation d'ensemble qui se substitue à la notation et fait un place plus importante au chef d'établissement, une différenciation de carrière accrue par la création d'un grade nouveau, la classe exceptionnelle.

Le ministère de l'Éducation nationale réfléchit à un projet de grille d'évaluation professionnelle des enseignants qui comporte neuf thèmes d'évaluation, tels que "coopérer au sein d'une équipe" ou "construire, mettre en œuvre et animer des situations d'enseignement et d'apprentissage prenant en compte la diversité des élèves". Les critères liés à la participation collective et à l'individuation des pratiques apparaissent ainsi plus nettement.

Cependant ce projet prévoit que tous les enseignants ayant une carrière complète, atteindront la hors classe. Si les accélérations de carrière seront plus fortes qu'auparavant et plus déterminées par les autorités académiques, le critère de l'ancienneté demeure très prégnant. Le projet du ministère dispose que les propositions annuelles de promotion à la hors classe seront basées sur un "barème" fondé sur deux éléments, d'une part l'appréciation finale du troisième rendez-vous de carrière de l'enseignant, d'autre part le nombre d'années de présence de l'enseignant dans la plage d'appel statutaire à la hors-classe.

Aussi cette réforme est loin de garantir une connexion plus forte entre carrière et évaluation. De surcroît apparaît le risque que s'instaure effectivement une déconnexion entre évaluation et avancement sans que pour autant les missions de conseil et d'appui des inspections se mettent en place au profit des enseignants qui auraient besoin, les moyens des inspections étant toujours autant mobilisés par la gestion de la procédure.

B - Une évaluation des enseignants mal articulée avec les autres dispositifs évaluatifs du système scolaire

1 - Un modèle français d'évaluation atypique

À la différence de nombreux pays étrangers, les progrès des élèves, qui ne sont pas mesurés autrement que par l'appréciation des enseignants, ne sont pas un des critères de l'évaluation de ceux-ci.

L'OCDE a conduit une enquête sur l'utilisation des évaluations des élèves¹⁸ auprès des chefs d'établissement. Parmi les réponses proposées, il ressort en moyenne OCDE que 80 % des élèves fréquentent un établissement où les évaluations des élèves servent à identifier les aspects de l'enseignement ou des programmes pouvant être améliorés ; 63 % des élèves sont scolarisés dans un établissement qui a recours aux évaluations des élèves pour comparer la performance de l'établissement à la performance régionale ou nationale et 50 % des élèves fréquentent un établissement qui utilise les évaluations des élèves pour évaluer l'efficacité des enseignants.

La France est très décalée puisque seulement 23 % des élèves français fréquentent un établissement dont le chef d'établissement indique que les évaluations des élèves sont utilisées pour évaluer l'efficacité des enseignants.

De même, l'incorporation de l'opinion des « parties prenantes » (parents d'élèves, élèves eux-mêmes) est peu pratiquée : seuls 13 % des élèves français fréquentent un établissement pratiquant un « *retour d'information écrit* » des usagers contre plus de 60 % en moyenne dans l'OCDE.

À l'inverse, le recours à des inspections pédagogiques pour suivre les pratiques des enseignants est beaucoup moins pratiqué à l'étranger. Ainsi seulement 27 % des élèves sont-ils en moyenne OCDE scolarisés dans des établissements où cette méthode est utilisée contre 69 % qui ont recours au seul chef d'établissement¹⁹.

2 - Un impact limité sur les pratiques d'enseignement

Au regard de l'objectif d'amélioration de l'enseignement, par le contrôle et le conseil, il est estimé que l'évaluation « a peu d'effets sur les pratiques individuelles et collectives et ne débouche pas – ou tout au moins pas souvent – sur des éléments d'amélioration et de régulation du système éducatif »²⁰.

Au-delà du poids prépondérant des aspects liés au déroulement de carrière, certaines observations expliquent ce constat préoccupant.

Les rapports d'inspection individuelle consultés par la Cour montrent que si les notes attribuées sont encadrées par une grille qui impose une prise en compte de l'ancienneté, aucun standard n'a été défini pour le déroulement des inspections, ni les thèmes abordés, sauf pour les enseignants débutants à l'occasion de leur titularisation qui a récemment fait l'objet d'un « outil d'accompagnement : descripteurs des degrés d'acquisition des compétences à l'entrée dans le métier ». La continuité des rapports n'est pas assurée, il s'agit d'exercices rédactionnels à chaque fois renouvelés et qui ne font apparaître aucun suivi explicite d'une inspection à la suivante pour un même enseignant.

¹⁸ OCDE (2014), Résultats du PISA 2012 : Les clés de la réussite des établissements d'enseignement : Ressources, politiques et pratiques (Volume IV), Évaluations et responsabilisation (chapitre IV/page 139 et suivantes) PISA, Éditions OCDE, Paris.

¹⁹ Idem.

²⁰ L'évaluation des enseignants avril 2013, IGEN, Rapport - n° 2013-035, avril 2013.

Cette situation n'est pas satisfaisante pour un processus qui se reproduit environ 170 000 fois par an, qui pourrait aisément être fondé sur le référentiel des compétences enseignantes défini en 2010 et révisé en 2013.

Il est anormal que non seulement les inspections aient peu de conséquences sur la gestion du corps enseignant, mais également sur les aspects pédagogiques qui constituent sa justification primordiale, alors même que la mission d'inspection mobilise une part importante des moyens. En comptant qu'une inspection représente entre 50 % et 60 % d'une journée de travail (temps de déplacement, d'inspection et d'entretien, de rédaction du rapport), 170 000 inspections représentent de 85 000 à 102 000 jours de travail soit 400 à 490 ETP, c'est-à-dire entre 15 % à 20 % du temps d'inspection disponible (IEN et IA-IPR).

C - L'évaluation des personnels de direction et des établissements : un système incomplet et imparfait

1 - Les évaluations des personnels de direction marginales dans le primaire, redondantes dans le secondaire et de faible impact

a) Dans le premier degré, les directeurs d'école sont évalués en tant qu'enseignants

L'évaluation des enseignants du premier degré repose sur une appréciation pédagogique et fait l'objet d'une note²¹. Lorsque l'enseignant est le directeur de l'école, l'évaluation de la partie « direction » de l'activité a une forme et une portée différentes. Malgré l'annonce de la « la définition d'un protocole d'inspection spécifique » pour les directeurs d'école²², les protocoles d'évaluation, qui relèvent de la responsabilité de l'académie, restent à l'heure actuelle inchangée.

En pratique, l'inspection des directeurs d'école en tant que tels, et dans une approche collective qui englobe l'unité pédagogique que forme l'école, reste marginale, voire fait défaut, ce qui est regrettable.

b) Dans le second degré, des procédures à simplifier

Elle vise le double objectif (protocole d'accord relatif aux personnels de direction du 16 novembre 2000) d'« apporter à ces personnels l'aide, l'accompagnement et la valorisation de leurs pratiques » et de « permettre aux opérations de gestion collective et individuelle qui impliquent comparaison et action discriminante (promotions, mutations) de se dérouler dans un contexte d'objectivité ». L'introduction de l'indemnité de fonctions, de responsabilités et de résultats (IF2R), arrêtée selon les résultats de l'entretien professionnel, a constitué une étape nouvelle dans le processus de l'évaluation.

L'évaluation des chefs d'établissement du secondaire est fondée sur un examen triennal de leur performance qui s'appuie sur une lettre de mission assise sur un diagnostic des points

²¹ Décret n° 90-680 du 1er août 1990 relatif au statut particulier des professeurs des écoles.

²² Fiche 1 du groupe de travail n° 1 sur la modernisation du métier d'enseignant (juin 2016).

forts et faibles de l'établissement, cette évaluation détermine si les objectifs fixés sont atteints et le montant de l'IF2R.

La portée de ce mécanisme est altérée par un deuxième dispositif qui préside à la mobilité de ces personnels, en fait plus déterminante en termes de carrière et de rémunération car il conditionne l'affectation à la direction d'un établissement de catégorie supérieure. Cette seconde évaluation phagocyte²³ la première. Pour ne pas perdre le fruit des évaluations, le ministère a décidé qu'à compter du mouvement 2017, le compte-rendu d'entretien professionnel serait également utilisé dans la procédure de mobilité des personnels. C'est une modeste avancée alors que la fusion des deux processus s'imposerait.

L'évaluation des personnels de direction a de faibles conséquences en termes de rémunération. L'IF2R qui varie en principe de 0 € à 6 000 € par période triennale est dans la plupart des cas arrêtée à 2 000 €. Les appréciations portées se caractérisent par une inflation du niveau « excellent » ce qui a conduit la DGRH du ministère à une régulation, par l'imposition d'un quota maximal par académie de 10 % d'excellence (depuis 2014, dans une fourchette de 10 à 15 %). En outre, tous les personnels sont réputés atteindre au moins partiellement leurs objectifs.

Certains cas de dysfonctionnement grave, pris toutefois en compte pour la mobilité professionnelle, n'excluent pas le versement de l'IF2R même si elle est réduite. A contrario, pour un chef d'établissement « sans doute trop jeune pour avoir des objectifs dépassés » mais qui a pourtant obtenu des « résultats exceptionnels notamment en termes de réussite au DNB dont le taux est passé de 50 à 80 % », l'évaluation très positive n'a abouti qu'au bénéfice d'une IF2R standard à 2 000 €²⁴.

Ainsi, comme pour les enseignants, le temps consacré à la procédure est sans commune mesure avec les effets produits, aussi bien financiers qu'en termes de reconnaissance professionnelle ; les indicateurs de performances contextualisés, les indicateurs de valeur ajoutée des lycées, sont disponibles mais non utilisés dans les évaluations des chefs d'établissement.

2 - L'évaluation des établissements : une démarche à développer, dépendante du degré d'autonomie des unités d'enseignement

L'évaluation collective, par unité pédagogique ou par unités d'enseignement (écoles, collèges et lycées) reste peu développée. La circulaire du 11 décembre 2015 qui actualise les missions des corps d'inspection ne fait pas explicitement référence à l'évaluation collective, elle se contente de préciser que l'inspection individuelle « s'inscri[t] dans le contexte de l'école ou de l'établissement » et est réalisée « en référence à un travail d'équipe ».

²³ Dans leur rapport n° 2015-060 de juillet 2015 « L'évaluation des personnels de direction », les inspections générales utilisent le terme de « dévoiement ».

²⁴ Idem.

a) Des données absentes pour l'école, nombreuses pour les établissements du secondaire

L'évaluation des unités d'enseignement, c'est-à-dire des 53 000 écoles et des 11 000 établissements du secondaire EPLE, est l'objet d'une myriade de dispositions éparées. Il en résulte un paysage très contrasté entre :

- d'une part une école primaire qui, ne disposant à ce jour²⁵ d'aucun outil statistique pour mesurer ses résultats ou ses moyens, reste très peu concernée par une démarche évaluative, hormis le rôle de suivi des inspecteurs de l'éducation nationale IEN ;
- d'autre part, des EPLE pour lesquels il existe de nombreux indicateurs réunis par la DEPP sur la réussite aux examens (BAC et DNB), la fluidité des parcours (taux de redoublements, taux de passage dans la classe supérieure, taux d'orientation vers les différentes filières du lycée), les ressources (effectifs enseignants, rotation, ancienneté, DGH, H/E, E/S) et des éléments contextualisés (indicateurs de valeur ajoutée des lycées IVAL et des collèges ICAL).

Le rapport consacré en 2012 par l'IGEN à l'évaluation des établissements²⁶ constatait « *une image confuse* », avec une appropriation faible par les établissements des instruments mis à leur disposition et notamment de l'outil d'auto-évaluation et de pilotage de l'établissement de la DGESCO (OAPE), application informatisée qui comporte un ensemble d'informations destinées à faciliter des démarches d'auto-évaluation participative. La faible utilisation d'OAPE, son site n'est consulté que par environ 2 000 visiteurs par an depuis plusieurs années, témoigne d'un faible intérêt pour les démarches évaluatives.

Quant à la base APAE qui comprend des données brutes (résultats au bac) et des indicateurs de valeur ajoutée (IVAL pour les lycées et ICAL pour les collèges, en cours d'élaboration), rendus publics et repris par la presse, elle n'a qu'une vocation informative.

b) Des indicateurs de valeur ajoutée : progrès incontestable mais à conforter au plan méthodologique

La démarche suivie par la DEPP pour évaluer les unités d'enseignements consiste à déterminer les connaissances des élèves, si possible à l'entrée (en 6^e et en 2nde) et à la sortie (en 3^e et en terminale) des établissements et à les contextualiser en intégrant des données socio-économiques, avec une indication sur des taux d'accès aux examens pour mettre en lumière d'éventuelles pratiques d'écramage. Ces indicateurs souffrent cependant de plusieurs défauts :

- les niveaux des élèves sont grossièrement approximatifs ; les résultats en fin de terminale sont écrasés par la seule prise en considération de l'obtention d'un diplôme de plus en plus fréquemment obtenu quand les données sur l'entrée en 6^e ne sont intégrées que par la prise en compte du taux de redoublants.

²⁵ Un projet d'APAE 1^{er} degré serait en cours de réalisation avec un objectif d'atterrissage au 4T 2016 mais il se heurte à de nombreuses obstacles et notamment la méconnaissance des niveaux des élèves et de leurs caractéristiques sociales, de l'éclatement d'établissements 5 fois plus nombreux que les EPLE et de la charge de travail supplémentaire que cela impose aux directeurs d'école.

²⁶ « L'évaluation des unités d'enseignement Pour une démarche méthodologique et déontologique », Rapport n° 2012-065, juin 2012.

- ces indicateurs sont très instables ; la valeur ajoutée d'un établissement peut varier fortement d'une année sur l'autre sans que ni la composition de son public, ni les pratiques de l'équipe éducative n'aient changé. D'après le ministère, les raisons de ces variations sont inhérentes à la méthode d'estimation, ce qui mériterait une enquête méthodologique.

Ils n'en constituent pas moins pour les parents un réel apport évaluatif en leur permettant de cerner, au-delà des résultats bruts aux examens, ce que les établissements font réellement pour les élèves qu'ils accueillent, ils sont le produit d'une démarche de qualité et de transparence envers les usagers, quasi unique au sein de l'éducation nationale.

Pour les équipes enseignantes, ces indicateurs sont un retour de la mesure des efforts collectifs faits alors que les démarches collectives demeurent rares. Ils sont aussi pour l'éducation nationale la mesure des efforts d'un établissement pour porter un projet adapté aux besoins des élèves.

Les inspections générales estiment que les IVAL ne dessinent pas un modèle unique vertueux d'établissement, facilement reproductible : « la valeur ajoutée, positive ou négative, ne peut jamais être imputée à un seul élément déterminant. C'est toujours une conjonction de facteurs qui se renforcent mutuellement, qui peut expliquer la valeur ajoutée d'un établissement. Il n'y a pas de conjonction modèle, pas de conjonction standard »²⁷.

Cependant une analyse devrait être entreprise pour faire ressortir plusieurs modes efficaces de fonctionnement d'établissements ; à défaut le partage des bonnes pratiques au sein du système scolaire demeurera faible, l'apprentissage commun inexistant et l'apport des indicateurs de valeur ajoutée indiscernable.

De même ils devraient être étudiés pour préciser les marges d'autonomie qu'il serait efficace d'accorder aux établissements.

c) Hors IVAL, l'évaluation de « l'effet établissement » limitée par la faiblesse de leur autonomie

« L'effet établissement » n'est pas un des facteurs majeurs de la réussite scolaire : selon la plupart des études disponibles, les connaissances acquises lors d'une séquence d'enseignement dépendent avant tout du niveau initial des connaissances (50 % de la variance inter-élèves environ), de l'environnement familial dans ses composantes socio-économique et géoculturelle (25 % de la variance), de la classe à laquelle l'élève appartient (10 % à 15 %) et, pour une partie du solde seulement (5 %), de l'établissement. Pour autant il conditionne beaucoup l'optimisation des autres facteurs : relations avec l'environnement familial, exercice du métier d'enseignant, adéquation des moyens aux besoins des élèves, mise en perspective collective du projet éducatif en termes d'orientation et d'ambition scolaires.

²⁷ Rapport n° 2015-065 précité.

Or cet « effet-établissement » qui a un impact de catalyseur, est contraint par la faiblesse des moyens d'action du chef d'établissement²⁸ et par une absence de définition (et de contenu précis) de la notion d'autonomie.

Dans cette situation ambiguë, l'évaluation des établissements se heurte, pour avoir un sens, à la question de leur degré d'autonomie et au niveau des prérogatives des personnels de direction. Ainsi l'évaluation des établissements est assez peu pratiquée en comparaison de l'étranger²⁹ comme l'IGEN l'a constaté : « La France est un des rares pays européens qui n'a pas encore généralisé l'évaluation des établissements d'enseignement »³⁰.

En définitive le contraste entre l'injonction de généralisation de l'évaluation³¹ et la pauvreté des pratiques évaluatives concernant les établissements est saisissant.

Ce défaut d'évaluation trouve d'abord sa source dans la limitation des moyens disponibles. Les éléments dont disposent les services pour procéder aux évaluations demeurent restreints. Pour ce qui concerne les résultats des élèves, ils consistent principalement dans ceux obtenus aux examens de fin de premier (DNB) et de second cycle (BAC) contextualisés par la prise en compte de l'origine sociale (IVAL et ICAL) mais les progressions annuelles des élèves, prises dans leur ensemble, sont ignorées. Dans le primaire, ne sont disponibles que les livrets de compétence qui, même lorsqu'ils sont dématérialisés, ne donnent pas lieu à une exploitation statistique. Aussi l'évaluation des écoles primaires reste un angle mort.

De surcroît, l'utilisation de méthodes innovantes, à la différence de nombreux pays étrangers, comme les inspections collectives qui embrassent la totalité de l'établissement scolaire, ou l'évaluation associant les parties prenantes (élèves, parents, collectivités locales), est inexistante.

III - L'évaluation des dispositifs scolaires : des initiatives nombreuses mais sans appréciation explicite de leurs résultats

Fondées sur des méthodologies concurrentes ou réalisées dans des environnements différents, les évaluations des dispositifs et pratiques pédagogiques et éducatifs, aboutissent souvent à des préconisations éparpillées, voire contradictoires, sans qu'existent les moyens

²⁸ Le rapport des inspections générales n° 2015-065 consacré aux facteurs de valeur ajoutée des lycées note par exemple dans un développement consacré aux « fausses évidences » que « des lycées qui avaient été confrontés à un déficit de chef d'établissement suite à une absence prolongée consécutive à une maladie ou pour d'autres raisons et qui produisaient néanmoins de la valeur ajoutée positive ».

²⁹ « En France, où le système d'inspection est principalement axé sur le personnel de l'école, il n'existe pas de protocole standardisé définissant le contenu et les procédures d'évaluation externe des établissements scolaires », Commission européenne/EACEA/Eurydice, 2015. « Assurer la qualité de l'éducation : politiques et approches de l'évaluation des écoles en Europe ». Rapport Eurydice. Luxembourg : Office des publications de l'Union européenne.

³⁰ « L'évaluation des unités d'enseignement : Pour une démarche méthodologique et déontologique », IGEN n° 2012-065.

³¹ L'article R241-19 du CE qui définit les responsabilités des inspecteurs territoriaux en la matière est représentatif de cette aspiration à l'universalité de l'évaluation qui comprend « le travail individuel et le travail en équipe des personnels enseignants, d'éducation et d'orientation des écoles, des collèges et des lycées ...l'enseignement des disciplines, des unités d'enseignement, des procédures et des résultats de la politique éducative ».

d'arbitrer et de trancher entre elles dans le cadre d'une fonction d'évaluation mal ordonnée. La Cour en donne les exemples suivants.

Les outils méthodologiques de l'évaluation des dispositifs publics

Deux grands types de méthodes existent pour évaluer des actions publiques. Elles peuvent être schématiquement caractérisées de la façon suivante.

Les méthodes quantitatives, issues de la statistique et de l'économétrie, s'intéressent à l'efficacité des politiques publiques pour répondre aux questions suivantes : quels sont les impacts des politiques publiques ? La politique évaluée permet-elle d'atteindre les objectifs qui lui sont assignés par les pouvoirs publics ? Elles s'appuient sur l'exploitation de données relatives à de grands échantillons d'individus, par exemple des tests standardisés et numérisés.

Les méthodes qualitatives, très utilisées en sociologie, s'intéressent au fonctionnement des politiques publiques et tentent de répondre à la question suivante : comment le dispositif évalué produit-il ces effets ? Quelle est la perception par les personnes concernées de la conduite de ces politiques et de leurs effets ? Elles reposent sur des enquêtes approfondies auprès d'un nombre réduit d'acteurs.

Les deux approches requièrent des précautions méthodologiques, qui assurent leur crédibilité et leur transparence.

- La méthode considérée comme la plus fiable des évaluations quantitatives est celle des « expériences aléatoires » (*randomized experiments*), c'est-à-dire une expérimentation avant la généralisation d'une politique. Il s'agit de constituer au hasard deux groupes d'individus, un groupe de personnes « traitées » par cette politique expérimentale (le groupe de traitement), et un groupe d'individus non concernés (le groupe de contrôle). Il est alors possible, une fois que la mesure testée a produit ses effets, de comparer directement les variables d'intérêt (les « revenus ») de ces deux groupes de personnes pour en mesurer les impacts. Ceci implique d'avoir des échantillons d'une importance telle que la diversité de la population soit représentée.

Ces expériences présentent un certain nombre d'inconvénients tenant à leur coût ou la lourdeur des dispositifs de suivi à mettre en œuvre. Lorsqu'ils ne peuvent y recourir, les évaluateurs se tournent vers des méthodes d'estimation issues de l'économétrie. Celles-ci tentent de reconstituer ex post le cadre de l'expérience en identifiant un groupe de traitement et un groupe de contrôle grâce à des données « observationnelles » (par opposition aux données issues de protocoles expérimentaux qui sont constituées in vivo).

Cette méthode mobilise les résultats du système statistique public, issus d'enquêtes (dont les données ne concernent que quelques milliers d'individus au plus) et de bases de données administratives (c'est-à-dire des données exhaustives qui peuvent concerner des millions d'individus). Le travail de l'évaluateur consiste alors à appliquer à ces données les méthodes les plus couramment utilisées en économétrie pour mesurer des effets de causalité entre application d'une mesure et les modifications des variables caractérisant une population.

- Les évaluations qualitatives s'appuient sur un corpus de méthodes issues des sciences sociales, qui comprend des entretiens semi-directifs ou ouverts, collectifs ou individuels, des sondages sur des échantillons d'individus. L'évaluateur peut recourir à une enquête par immersion dans les structures et auprès des populations concernées : pour une politique éducative, par exemple, observer en temps réel dans un établissement les interactions entre les enseignants et les élèves qui expérimentent un nouveau dispositif. L'appréciation évaluative finale est parfois arrêtée grâce à l'association des « parties prenantes » de la politique considérée (acteurs administratifs, représentants des populations concernées) à la communauté des chercheurs ayant participé à l'évaluation, pour tenter de dégager un consensus.

Qu'elles soient de nature quantitative ou qualitative, les méthodes de l'évaluation sont complémentaires.

A - L'enseignement de la lecture : des résultats contradictoires

Les pratiques d'enseignement de la lecture varient considérablement³² d'un pays à l'autre sans que cela puisse être expliqué par des spécificités linguistiques.

La mesure des performances comparées des différents systèmes est fragile. Ainsi, si dans l'enquête PISA 2009, centrée sur la compréhension de l'écrit, la France a obtenu des scores moyens plutôt satisfaisants (496 contre 490 pour l'ensemble des 27 pays européens) avec un étalement important (écart-type de 106 contre 96), l'enquête PIRLS³³ 2006 lui avait attribué des scores mauvais (522 contre 533 pour l'ensemble des 27 pays européens) et faiblement étalés (écart type de 66,6 contre 74,4).

1 - L'expérimentation PARLER

Le dispositif « Parler, Apprendre, Réfléchir, Lire Ensemble pour Réussir » (P.A.R.L.E.R.) a été conçu par une équipe de recherche de l'université Pierre-Mendès-France de Grenoble. Ce programme a été testé entre 2005 et 2008 sur la durée du cycle des apprentissages fondamentaux³⁴ dans des écoles de cette agglomération situées en zone d'éducation prioritaire. Sur les classes auxquelles le protocole avait été appliqué, le niveau atteint était estimé, en fin d'expérience, nettement supérieur à celui du groupe témoin et supérieur à la moyenne nationale d'appropriation des codes d'expression.

La DGESCO lança une expérimentation de plus grande ampleur en 2010/2011 dans 44 classes de l'éducation prioritaire de l'académie de Lyon, expérimentation étendue en 2011/2012 sous l'appellation « projet Lecture³⁵ » par l'association « Agir pour l'école » à 59 classes dans toute la France et poursuivie en 2012/2013. Doté d'un budget global proche de 2 M€, quasi-intégralement financé par l'association, le projet a été évalué par une équipe externe composée de la DEPP, de l'institut de recherche sur l'éducation (IREDU) de l'université de Bourgogne et du laboratoire d'étude des mécanismes cognitifs (EMC) de l'université Lyon II.

Dans une note d'étape de mars 2013, les évaluateurs concluaient que les « effets du dispositif expérimental sont positifs et statistiquement significatifs...+ 17 % en reconnaissance des lettres, + 22 % en habiletés phonologiques et + 41 % en lecture par voie non lexicale...en phonologie (et) en compréhension l'effet est significatif pour les élèves les plus faibles ».

Saisie en septembre 2012, l'IGEN avait remis au ministre, sans disposer des éléments précédents, des conclusions d'une tonalité très dubitative et critiques sur la procédure « Les élèves ne peuvent être réduits à un statut de « cobayes » sur lesquels on exerce une action pour en voir les effets ».

Ainsi, deux avis contradictoires furent produits par deux services du MEN utilisant des méthodologies différentes. Depuis aucune position d'ensemble n'a été clairement et explicitement dégagée. L'absence d'une instance chargée de l'évaluation, non impliquée dans

³² L'enseignement de la lecture en Europe : contextes, politiques et pratiques, EACEA; Eurydice, Mai 2011.

³³ Elle concerne les écoliers de CM1 alors que PISA porte sur des élèves de 3^{ème}.

³⁴ De la grande section de maternelle à la fin du CE1.

³⁵ Sur la partie strictement scolaire du protocole.

des choix opérationnels mais chargée de valider des résultats et proposer leur synthèse, peut expliquer cet état de fait.

2 - L'enquête « Lire et Ecrire CP »

Lancée en 2011 avec le soutien de la DGESCO, l'enquête « Lire et Écrire CP » avait pour ambition d'étudier l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages.

131 enseignants et 2507 élèves ont été observés au début et à la fin du CP, puis à la fin du CE1, selon un protocole défini par 60 enseignants-chercheurs exerçant dans 13 universités, par 190 enquêteurs (enseignants-chercheurs, docteurs ou doctorants, conseillers pédagogiques en activité mis à disposition par l'Éducation nationale, enseignants ou formateurs à la retraite) et évalués par une centaine d'étudiants en psychologie embauchés sur crédits de la DGESCO et formés par l'équipe de chercheurs. L'observation en classe a duré trois semaines (une par trimestre) et a porté sur toutes les activités d'enseignement de la lecture et de l'écriture, soit 3 000 heures de classe codées et filmées, complétées par des documents écrits (questionnaires d'entretiens par exemple). Ce travail d'une ampleur exceptionnelle a eu un coût élevé (plus de 700 000 €).

Lors de la présentation de ce projet au comité scientifique de la DGESCO en juin 2014, maints aspects de l'étude furent contestés, tant au plan méthodologique que des résultats, ce qui n'a pas empêché qu'une présentation des résultats soit faite lors de la conférence de consensus (« Comment soutenir le développement de compétences en lecture ? ») organisée par le CNESCO en mars 2016³⁶.

Confrontée à des appréciations aussi contradictoires, la DGESCO a conclu ce débat « en affirmant que la diversité des méthodologies de recherche est positive », et espérant que « si les chercheurs ne peuvent s'accorder sur tout ...ils puissent cependant produire un message clair, cohérent, le plus loin possible dans l'accord pour les pilotes de l'école, les politiques, qui sont les décideurs, et vis-à-vis de la sphère enseignante qui a besoin d'être rassurée sur sa légitimité, sa professionnalité et sur sa capacité à remplir ses objectifs de formation »³⁷.

À partir des conclusions d'une étude couteuse, aucune position pour orienter des pratiques pédagogiques courantes n'a pu être dégagée.

B - Les classes de niveau : une position jamais clarifiée au regard des évaluations internationales

La France se place sur la question des classes de niveau, très différemment des autres pays de l'OCDE. D'après les études publiées par l'OCDE (PISA 2012), elle a un indice faible de « stratification horizontale », qui mesure la variation des modes d'instruction dans un même

³⁶ « Quelles sont les pratiques enseignantes qui ont un effet positif sur l'apprentissage initial de la lecture ? », GOIGOUX Roland, Université Blaise Pascal, Clermont-Ferrand.

³⁷ Déclaration de la directrice générale citée dans le compte rendu du Conseil scientifique de la DGESCO du 12 juin 2014.

niveau d'enseignement, ce qui traduit l'absence de regroupement des élèves par niveau au bénéfice d'une organisation uniforme par classe dans chaque établissement.

Face à cette situation décalée, la France ne s'est pas mise en état de disposer de données scientifiques solides et concordantes qui garantissent que la mixité scolaire soit plus ou moins efficace que l'homogénéité scolaire.

Les effets de la mixité scolaire sur les acquisitions des différents publics sont controversés, comme en témoigne le rapport rédigé à l'occasion de la conférence de comparaisons internationales organisée par le CNETCO en juin 2015. D'une revue de la littérature existante, le rédacteur tire un bilan nuancé : « les élèves moins performants bénéficient du contact d'élèves plus performants, mais seulement si le niveau scolaire de ces derniers se rapproche du leur. En d'autres mots, un certain niveau d'inégalité des résultats scolaires dans la classe est bénin pour les élèves les plus performants et bénéfique pour ceux qui le sont moins. Mais si l'écart est trop important, les conséquences deviennent négatives pour les meilleurs élèves et non significatives pour les plus faibles... ».

La position de certains chercheurs³⁸, fondée sur une synthèse des principales recherches internationales, est plus catégorique, estimant que « dans l'enseignement secondaire, comme dans l'enseignement primaire, l'effet spécifique du groupement des élèves par classes est nul ». La position du ministère est alors de privilégier une forme d'attentisme, selon lequel « il vaut mieux conserver les classes hétérogènes dès lors qu'il n'y a aucune preuve de l'efficacité de groupes homogènes ». Durant l'enquête, la DGETCO a précisé à la Cour que « la doctrine officielle est de ne pas constituer de groupes de niveau qui installeraient des processus plus ou moins explicites de sélection, souvent trop précoces et apparaîtraient contraire au principe d'inclusion désormais inscrit dans la loi. Cette consigne est régulièrement répétée et résiste aux changements politiques ». La réponse à l'hétérogénéité des niveaux est la « différenciation des pédagogies ».

En revanche certains travaux ont conclu en sens inverse : « Tout d'abord, les groupes de niveau sont bénéfiques pour tous les enfants : en moyenne (à taille de classe et à chaque type d'enseignant donnés), les enfants ont des résultats supérieurs de 14 % de l'écart type dans les écoles où ils ont été répartis selon leur niveau, par rapport à ceux qui ont été répartis au hasard. Plus important, cet effet existe aussi bien pour les enfants les plus faibles que pour les enfants les plus forts... »³⁹. Ce résultat est produit par une universitaire française⁴⁰ membre du comité scientifique de la DGETCO.

Les personnels de direction se déclarent très majoritairement (84 %) hostiles à ces pratiques, ils sont beaucoup moins, un sur deux, à considérer que les enseignants sont plus efficaces dans des classes à niveau scolaire hétérogène⁴¹. En revanche l'opinion des enseignants sur le rapport entre les classes de niveau et l'efficacité de l'enseignement n'est pas connue. Pourtant dans plusieurs établissements secondaires visités par la Cour, l'empirisme a conduit à

³⁸ « L'école peut-elle être juste et efficace ? » Crahay.

³⁹ « Le développement humain : Lutter contre la pauvreté », Esther Duflo.

⁴⁰ Esther Duflo est professeure au MIT et ancienne titulaire de chaire au Collège de France.

⁴¹ « Les classes de niveau sont perçues comme non pertinentes pour un bon fonctionnement de l'établissement et des bons résultats scolaires dans l'établissement par 84 % des chefs d'établissement, collèges et lycées confondus », in « La constitution des classes : pratiques et enjeux Enquête auprès des chefs d'établissement des collèges et lycées », CNETCO, Septembre 2015.

créer des groupes homogènes, ou pour permettre aux élèves en grande difficulté de combler leur retard, ou en raison de l'affectation des élèves selon leurs options (parcours bi-langue ou latin en 5^e).

Au total le système éducatif français semble avoir pris une option sans se saisir de la position excentrée de la France en matière d'homogénéité des classes, comme levier pour conduire une approche évaluative approfondie sur ce sujet.

C - Le suivi individualisé des élèves et la différenciation pédagogique : un renoncement à évaluer

La Cour a fait le constat d'une quasi-inexistence d'évaluation des effets du suivi individualisé des élèves (RPT de mars 2015).

1 - Des inspections générales qui dénoncent la carence d'évaluation mais restent en retrait

Les inspections générales dressent un constat pessimiste de la situation dans leur rapport d'octobre 2010 sur la mise en place du suivi individualisé, considérant que l'évaluation de la portée des dispositifs se trouve « *entre désert théorique et désarroi pratique* ». Ces constats sévères ne sont que la réitération des conclusions d'un rapport de juillet 2008, particulièrement éclairantes : « cette partie du rapport sera malheureusement assez brève, tant les formes d'évaluation ont peiné à se mettre en place et à se développer... On en reste donc à des formes de constat, généralement positives...mais ces éléments ne s'accompagnent ni de données chiffrées ni de bilans comparés des résultats des élèves fréquentant ou ne fréquentant pas l'accompagnement éducatif. Des formes d'évaluation plus systématiques devront être envisagées ».

Pour autant, les inspections générales s'estiment démunies pour évaluer les effets des dispositifs de suivi individualisé des élèves, ne pouvant mettre en œuvre des protocoles évaluatifs lourds.

2 - Une direction de l'évaluation face à des obstacles de principe jugés insurmontables

Aucun des dispositifs de suivi individualisé des élèves n'a fait l'objet d'une évaluation. Interrogés par la Cour, le ministère comme les académies reconnaissent ne disposer d'aucun élément d'évaluation des effets de ces dispositifs sur les élèves qui en bénéficient.

Un certain nombre d'obstacles sont mis en avant par la DEPP: sa sollicitation tardive qui ne lui permet pas de récupérer des informations essentielles en amont des changements ou ne peut orienter l'expérimentation de façon à pouvoir l'évaluer, une organisation des dispositifs qui limite leur évaluation scientifique par défaut d'échantillon témoin, dit « contrefactuel ».

Ces obstacles méthodologiques sont la résultante de questions de principe qui ne sont pas levées.

Par exemple l'exhaustivité à laquelle s'oblige le ministère dans l'application de ces dispositifs entrave la réalisation d'expérimentations préalables sur un groupe restreint et empêche de mesurer la plus-value d'un nouveau dispositif entre bénéficiaires et non-bénéficiaires de la mesure (*cf. Encadré sur les méthodologies évaluatives*).

Or le ministère a parfois dérogé à cette contrainte : une expérimentation des « programmes personnalisés de réussite éducative » (PPRE) à l'école et au collège a été conduite en 2005-2006 dans quelques établissements⁴² et les inspections générales en ont dressé le bilan⁴³. Nombre de pays étrangers ne s'obligent pas à l'exhaustivité : deux évaluations de référence, l'une aux États-Unis (le programme STAR de l'État du Tennessee en 1985), l'autre au Danemark (expérimentation « Deux enseignants dans la classe » en 2012-2013), reprise d'ailleurs dans la littérature nationale, reposent sur des expériences comparant les résultats des élèves d'écoles-tests et d'écoles-témoins.

Le second obstacle invoqué réside dans l'impossibilité d'isoler les effets d'un dispositif par rapport à un autre : la situation est en effet rendue difficile en France par la simultanéité et la multiplicité des réformes entreprises. Il suffit de rappeler que les PPRE ont été créés en 2005, que l'accompagnement éducatif a été introduit en 2007, que l'aide personnalisée a été en vigueur dans les écoles de 2008 à 2013, avant d'être remplacée par les activités pédagogiques complémentaires, ou encore qu'ont été introduits en même temps par la réforme du lycée l'accompagnement personnalisé, les stages de remise à niveau, les stages de langue, les stages passerelles et le tutorat.

Il ne tient qu'au ministère de limiter le nombre de réformes conduites simultanément dans un même niveau d'enseignement. La mesure de leur efficacité s'en trouvera facilitée et renforcée.

Il est remarquable qu'une politique de différenciation des pédagogies, largement revendiquée et appuyée par les ministres successifs, n'ait jamais fait l'objet d'évaluation, globalement ou dispositifs par dispositifs. Certes leur instabilité rend peu aisée une approche rigoureuse sur une période d'observation suffisante. Mais c'est justement ce qu'il eût été nécessaire de faire avant toute modification ou substitution d'un dispositif par un autre.

Cet exemple illustre également les défauts de réaction par rapport à des carences évaluatives repérées par les inspections générales, ainsi que la réticence à s'engager dans des expérimentations différenciant durablement le traitement des élèves.

D - L'éducation prioritaire : des conclusions péremptoires et contradictoires.

Sur une politique éducative majeure, relativement ancienne, qui emporte de forts enjeux de cohésion et d'insertion sociales, d'amélioration du classement de la performance de la France par une résorption de l'échec scolaire des élèves défavorisés, d'adaptation de l'action publique

⁴² Voir le « guide pratique pour l'expérimentation à l'école et au collège pour l'année scolaire 2005-2006 » publié par la DGESCO.

⁴³ Rapport n° 2005-048, de l'inspection générale de l'éducation nationale, de juin 2006 : « Programmes personnalisés de réussite éducative ».

aux territoires, et dont le coût est élevé (de l'ordre d'un milliard), il serait attendu que le MEN disposât d'évaluations aux fondements et conclusions robustes.

Or les études produites par les entités publiques chargées de l'évaluation, ont une valeur fragile car mal étayée par des travaux spécifiques, fondés sur les résultats des élèves ou la mesure de l'impact de certains dispositifs sur la performance du dispositif (par exemple la création d'un régime indemnitaire spécifique garantie-t-elle la stabilité des équipes enseignantes ? réduit-elle l'affectation des primo-enseignants ?). Elles emportent de surcroît des conclusions contradictoires.

Comme évoqué, les travaux du SGMAP de mai 2014 ne comportaient aucune appréciation sur l'efficacité de l'éducation prioritaire. Faisant état de constats très généraux, qui ne relèvent pas d'une méthodologie précise, l'orientation générale du propos était plutôt positive sur l'impact d'une telle politique, en appelant à sa poursuite au prix de certains aménagements (40 recommandations) qui selon le ministère « ont constitué les axes principaux de la réforme de la refondation de l'éducation prioritaire en 2015 ».

A contrario les travaux du CNESCO⁴⁴ publiés en septembre 2016 ont conclu de façon négative sur l'efficacité de cette politique pour réduire les inégalités de réussite et de destins scolaires, en termes très sévères « Faire de l'éducation prioritaire aujourd'hui, c'est accepter la ségrégation scolaire, au lieu de lutter contre, on essaie d'agir sur ses effets. On met plus de moyens, on sait que ça ne marche pas et on continue » (Conférence de presse du 24 septembre 2016).

Le rapport du CNESCO porte sur le rôle joué par l'école dans la formation des inégalités de réussite scolaire. S'il s'appuie sur 22 études pour formuler ses observations et recommandations, seule une petite partie d'entre elles (deux exactement) sont explicitement et entièrement consacrées à l'éducation prioritaire. La première résume les principaux résultats des évaluations menées sur ce sujet au cours des quinze dernières années. La seconde retrace les principales évolutions de la politique d'éducation prioritaire depuis 1981 et interroge la conception même de cette politique. Si les contributions citées sont utiles pour prendre rapidement connaissance de la littérature scientifique consacrée à l'éducation prioritaire, il ressort de l'examen des différents articles publiés par le CNESCO qu'aucun d'entre eux n'apporte d'enseignements véritablement nouveaux.

Il est écrit par exemple que la politique d'éducation prioritaire « est de moins en moins efficace » sans qu'aucune étude particulière sur les résultats des élèves ne retrace les impacts de cette politique depuis 1981 jusqu'à aujourd'hui. Sans évaluation quantitative des réformes de l'éducation prioritaire (au moins celle de 2011), se prononcer sur la tendance à la baisse ou à la hausse de l'efficacité de la politique d'éducation prioritaire manque de fondements.

Quant au ministère, il reconnaît que « Le sujet de l'éducation prioritaire est complexe et ne peut être traité succinctement... La difficulté ne tient pas tant à un problème de rigueur méthodologique, - la Depp et des chercheurs de haut niveau s'y sont confrontés-, qu'à l'impossibilité de disposer d'un contrefactuel ». Les mêmes obstacles que pour le suivi individualisé des élèves sont mis en avant : « il ne s'agit pas d'expérience aléatoire où l'on pourrait comparer la situation d'un élève qui aurait dû être en EP et qui n'y est pas pour une

⁴⁴ *Inégalités sociales et migratoires à l'école : comment l'école amplifie-t-elle les inégalités ? (Septembre 2016).*

raison aléatoire et un autre qui y est, car le dispositif touche tous les établissements présentant les mêmes caractéristiques ». Interrogée par la Cour, la DGESCO a précisé que « la politique d'éducation prioritaire au travers des dispositifs successifs, RAR (Réseau Ambition Réussite à partir de 2006) puis ECLAIR (à partir de 2011), n'a pas réussi à démontrer un effet global permettant de réduire le poids des origines sociales sur la réussite ».

En conclusion, au-delà d'approches évaluatives perfectibles qui ont plus relevé jusqu'à présent du registre de l'évaluation des politiques publiques au sens large, que de la mesure de l'impact de l'éducation prioritaire, il est regrettable qu'un tel enjeu d'efficacité et d'équité n'ait pas donné lieu à la construction d'outils méthodologiques solides et constants. Le ministère se contente d'une appréciation générale plutôt mitigée, alors qu'il lui serait très utile de savoir si cette absence éventuelle d'impact est la conséquence d'efforts budgétaires insuffisants, de modalités pédagogiques inappropriées, voire d'une impossibilité de réduire significativement ce phénomène.

Les obstacles méthodologiques (absence de « contrefactuels » par exemple) relevés à plusieurs reprises sont pourtant surmontables si une impulsion politique est donnée, comme en témoigne l'évaluation lancée sur les dédoublements de classe.

Évaluation prévue à partir de la rentrée 2017 pour le dispositif CP dédoublés

Le ministère a mis en place en amont une évaluation du dispositif « Cours Primaires dédoublés en REP+ » afin d'étudier l'impact de cette mesure comparé à celui du dispositif « Plus de maîtres que de classes » (PMQDC) qui continuera d'exister dans certaines classes. La DEPP a constitué, dès la phase de préparation de l'évaluation, un groupe consultatif de chercheurs pour en concevoir les différents volets.

Il s'agit d'évaluer l'impact immédiat du dédoublement des classes de CP par rapport à des CP « ordinaires », et selon les différentes modalités : CP à 12 élèves, CP dédoublés, PDMQDC. Il s'agit aussi d'évaluer l'impact à moyen terme (N+1 et N+2) des CP dédoublés par la mise en place d'évaluations standardisées en cours élémentaires dans les mêmes écoles que celles de l'échantillon pour disposer d'une mesure des performances des élèves avant et après la mise en place du dispositif.

E - Un pilotage mal assuré et peu assumé des innovations et des expérimentations

L'amélioration de l'efficacité du système éducatif suppose qu'elle incorpore, comme tous autres domaines, les progrès techniques et les acquis des sciences humaines (pédagogie, didactique, psychologie, économie, sociologie). Ce processus d'innovation requiert le plus souvent des phases d'expérimentation, avant toute application générale.

Cette combinaison d'innovation et d'expérimentation n'est pas naturelle à la culture traditionnelle de l'école française. Dans le rapport annuel des inspections générales de 2015, centré sur « L'innovation et l'expérimentation et leurs incidences sur l'évolution du système éducatif », les deux inspections générales notaient après un titre évocateur de la rupture culturelle induite par ces processus, « De l'interdiction au pilotage de l'innovation : la transgression contrôlée » que « Si, historiquement, l'École s'est construite autour d'un modèle

d'unicité et d'uniformité (destiné entre autres à promouvoir un sentiment national), l'évolution des attentes de la société a (re)mis au goût du jour la nécessité du changement ».

L'innovation généralisée, voulue par l'institution lors d'une réforme nationale, qui s'applique à tous les élèves et à tous les professeurs, reproduisant un schéma uniforme, pose un problème d'évaluation comme les exemples précédents l'ont montré. Or la phase d'expérimentation permet de surmonter cette difficulté, même si elle « transgresse » le principe d'uniformité qui continue de régir l'organisation du système éducatif.

Les appels à l'innovation expérimentée, quand bien même elle impliquerait une différenciation des pratiques, ne manquent pas. Le Haut Conseil de l'évaluation de l'école recommandait en 2001 que « le ministère développe une politique systématique d'expérimentations contrôlées. Seules de telles expérimentations sont susceptibles, à condition de prendre le temps nécessaire à leur évaluation objective, de permettre de faire des choix motivés et efficaces et d'arrêter des priorités de politique éducative ».⁴⁵

Le législateur en a consacré le principe : la loi d'orientation de 2005, après avoir défini les contours de la liberté pédagogique, a expressément prévu, dans son article 34 (article L.401-1 du code de l'éducation), sous réserve de l'autorisation préalable des autorités académiques, la réalisation d'expérimentations locales.

Dans la foulée de la loi de 2005, de multiples expérimentations furent lancées : « Cours le matin, sport l'après-midi », « Enseignement Intégré de Sciences et Technologie au collège », « Internats d'excellence », « Apprentissage de la lecture par petits groupes », « Livret de compétences expérimental », « Mallette des parents », ou plus récemment « Classes sans notes ».

1 - Une culture de l'expérimentation en progrès, un pilotage peu rigoureux des expérimentations de grande ampleur

La conduite des expérimentations à faible enjeu politique est mieux assurée que celle des dispositifs importants, dont la généralisation est un enjeu politique.

La participation d'acteurs extérieurs, le fonds d'expérimentation pour la jeunesse FEJ, les opérateurs des programmes d'investissements d'avenir PIA, catalyse les évaluations car promoteurs d'innovations qu'ils jugent expérimentales, ils y incorporent d'emblée des dispositifs évaluatifs, plutôt de nature quantitative. Le ministère est associé de façon diverse, beaucoup plus actuellement que par le passé comme en témoigne le suivi des projets numériques éducatifs. La faiblesse des moyens budgétaires (*cf. infra*) destinés aux expérimentations n'est pas étrangère à cette situation singulière pour un ministère d'aussi grande importance.

a) L'expérimentation de « la mallette des parents » a été conduite de façon pérenne, progressive et maîtrisée

Mise en œuvre pour la classe de sixième en 2008-2009 dans l'académie de Créteil (37 collèges en zones d'éducation prioritaire pour plus de 5 000 élèves), le projet « *mallette des*

⁴⁵ Avis d'octobre 2001 sur les « forces et faiblesses de l'évaluation du système éducatif ».

parents 6^{ème} » a été sélectionné dans les appels de projet « Pour de nouvelles expérimentations sociales » du FEJ. L'académie de Créteil a alors élargi l'approche et mené une deuxième expérimentation, accompagnée par l'École d'économie de Paris (EEP). En décembre 2011, le rapport d'évaluation finale des deux expérimentations est adressé au FEJ.

En novembre 2014, « la mallette des parents 6^{ème} » est intégrée au plan de lutte contre le décrochage scolaire, avec pour objectif la généralisation de son déploiement. De janvier à mars 2016, elle est diffusée dans toutes les académies (10 000 exemplaires), en ligne et accessible à l'ensemble des équipes pédagogiques.

L'intervention d'un décideur et financeur extérieur, le FEJ, a facilité le processus évaluatif, enrichi de données économétriques.

b) L'évaluation d'expérimentations plus emblématiques d'une politique scolaire, reste imparfaite en délais, en cadrage initial, en méthodes même

Les internats d'excellence

Cette politique est présentée d'emblée en 2008 comme expérimentale avec la création de deux lycées d'État sous forme d'internat.

Or le ministère de l'éducation nationale ne prend aucune disposition pour en assurer l'évaluation. C'est le FEJ, financeur d'une partie de ce projet, qui initie dès 2008 une évaluation dite quantitative des résultats du premier internat d'excellence, situé à Sourdun (comparaison sur trois ans des résultats de deux cohortes d'élèves, l'une sélectionnée pour l'internat d'excellence, l'autre scolarisée dans des établissements de droit commun), confiée à l'École d'économie de Paris, pour un montant de 725 000 € (540 000 € consommés). C'est encore le conseil scientifique du FEJ qui souhaite en 2012, approfondir le processus évaluatif qui lui paraissait insatisfaisant, par une enquête nationale sur les innovations pédagogiques (pour un montant de 170 000 €).

Les résultats de l'évaluation sur l'établissement expérimental, attendus en 2012, n'ont été disponibles qu'en mai 2013. Or la généralisation du dispositif avait déjà été décidée et entreprise dès 2010 par un plan ambitieux de 20 000 places en internats d'excellence (12 000 places dans des établissements spécifiques et 8 000 places labellisées dans des internats existants). Ce plan était doté d'un budget de 300 M€, puis porté à 400 M€, financé par le programme des investissements d'avenir (PIA)

Dans ce cas, alors que des arbitrages significatifs étaient à prendre en termes budgétaires et de cibles d'effectifs, les processus de la décision politique et des évaluations ont été déconnectés.⁴⁶

Le dispositif « Plus de maîtres que de classes » PDMQDC.

Ce dispositif, prévu dès 2012, a fait l'objet d'une disposition de la loi de refondation de l'école, 7 000 postes devant y être consacrés pour « *renforcer l'encadrement et accompagner*

⁴⁶ *Rapport annuel de la Cour des comptes de 2014* : Des internats d'excellence à ceux de la réussite : la conduite chaotique d'une politique éducative et sociale. La Documentation française, février 2014, disponible sur www.ccomptes.fr.

les organisations pédagogiques innovantes dans les réseaux d'aides spécialisées aux élèves en difficulté (RASED) ».

Le choix d'affecter un maître surnuméraire à une ou plusieurs classes peut paraître un dispositif complexe plutôt que de diminuer la taille des classes. Aussi une expérience de maîtres surnuméraires avait été réalisée ⁴⁷ entre 2001 et 2003 dans 91 écoles maternelles et élémentaires volontaires. Selon l'évaluation réalisée par la DEPP⁴⁸, les effets de leur présence étaient jugés positifs par les enseignants mais ces constats étaient « *très empreints de subjectivité* »⁴⁹.

Dans ce contexte, un suivi de cette expérimentation s'avérait particulièrement nécessaire, elle a été prévue dans la circulaire du 18 décembre 2012 instaurant ce dispositif. Elle a souffert de deux défauts.

D'une part, le cadrage du projet est resté vague dans ses objectifs (« prévenir la difficulté scolaire, tout au long de la scolarité primaire, et y remédier si elle n'a pu être évitée ») et ses modalités, d'autre part les outils méthodologiques sont demeurés confus. S'il fut recommandé d'évaluer les projets grâce « à des indicateurs et à des cibles précis » sans toutefois « corrélérer les résultats de l'évaluation à la décision de maintenir ou non le poste », s'il fut question « de pouvoir apprécier les résultats des élèves », il fut aussitôt précisé qu'une telle démarche ne devait pas reposer sur une approche « uniquement quantitative ». Dans les académies visitées, la Cour n'a trouvé trace d'aucune appréciation quantitative des effets du dispositif sur les résultats des élèves⁵⁰.

Enfin le comité national de suivi du dispositif PDMQDC ne fut institué qu'en janvier 2014 par la DGESCO, avec pour objet « la définition des modalités d'évaluation de l'efficacité du dispositif ».

En janvier 2015, alors que près de 1 850 enseignants étaient déjà affectés, ce comité concluait en accord avec la DEPP à la « difficulté d'évaluer précisément et rapidement l'efficacité du dispositif », il recommandait une évaluation limitée à une simple « prise d'information »⁵¹. Dans son rapport de septembre 2015, tout en rappelant que l'évaluation faisait « partie intégrante du projet », qui comptait désormais plus de 2200 maîtres, le comité déplorait qu'il fût « difficile d'isoler dans un ensemble de moyens mis au service d'une politique un facteur pour l'évaluer séparément ».

⁴⁷ La DGESCO avant la DGESCO.

⁴⁸ Direction de l'évaluation et de la prospective, préfiguration de la DEPP.

⁴⁹ Des maîtres surnuméraires dans des écoles maternelles et élémentaires, Note évaluation 04.04 avril, DEP.

⁵⁰ Le rapport de recherche IFé-ENS/Lyon II mentionne la réaction révélatrice d'un enseignant à ce sujet : « Localement, le conseiller pédagogique nous a confié sa crainte concernant cet aspect évaluatif : « je sentais venir qu'on allait nous demander de travailler sur l'effet sur les élèves, et ça je ne le sentais pas du tout ». C'est surtout la mesure « objective » de ces résultats et la possible multiplication des protocoles d'évaluation qui inquiètent les divers acteurs, principalement les enseignants ».

⁵¹ Comité de suivi PDMQDC, Note d'étape, janvier 2015.

La seule évaluation réalisée a été confiée par la DEPP à deux équipes de chercheurs dont l'une s'est intéressée, non pas au progrès des élèves, mais aux effets sur les enseignants⁵² et l'autre aux enseignements dispensés dans trois écoles. Par leur ampleur limitée, ni l'une ni l'autre de ces évaluations ne peut prétendre à la significativité statistique.

Dans sa note de janvier 2017, le comité de suivi du dispositif estimait que celui-ci « dans sa double dimension de prévention et d'aide aux élèves, notamment ceux les plus éloignés de la culture scolaire, contribue à la réduction des inégalités, à l'amélioration du climat scolaire et à la réussite de tous » sans préciser les données sur lesquelles étaient fondées ce jugement général. Aucune appréciation de l'impact sur les résultats des élèves n'a été portée.

Cependant, le dispositif a été généralisé en 2015 dans les réseaux de l'éducation prioritaire, processus achevé durant l'année scolaire 2016/2017, avant d'être en partie remis en cause lors de la rentrée 2017/2018.

Si ces deux exemples illustrent les difficultés à conduire des expérimentations au niveau national, ils ne peuvent servir à un jugement général négatif sur la capacité du MEN à conduire des évaluations sur expérimentations.

Pour que la culture de l'expérimentation se diffuse au sein du système éducatif comme le vecteur d'innovations, il est indispensable qu'elle soit assise sur une doctrine mieux établie et des protocoles plus respectés. Cela impose aux autorités administratives comme politiques de déterminer les objectifs et les modalités (même nombreuses) d'une expérimentation, de se donner les moyens de les évaluer et de se tenir à un calendrier avant toute décision d'abandon ou de généralisation.

2 - Le foisonnement des innovations locales : transformer des expériences en expérimentations aux enseignements transférables

Les innovations locales sont nombreuses, elles sont plus des expériences que des expérimentations.

La base Expérithèque, qui n'en recueille qu'une partie, en recensait près de 4 600 en 2016 dont les 2/3 étaient des expérimentations dites de l'article 34, c'est-à-dire requérant une autorisation rectorale, quand le reste relevait de la catégorie « innovations » car ne modifiant ni les programmes, ni les horaires institutionnels, ni les structures de l'établissement.

Ces initiatives locales sont très inégalement réparties entre les académies ce qui peut traduire des différences de dynamisme ou des disparités de recensement : 91 à Aix-Marseille mais 384 à Nice, 61 à Toulouse mais 283 à bordeaux, 12 en Martinique mais 295 à la Réunion. Les 4/5 sont réalisées sans moyens financiers ou humains supplémentaires. Plus de la moitié concernent les collèges et près de 40 % l'éducation prioritaire. Pour ce qui concerne l'évaluation, la rubrique n'est pas renseignée dans 25 % des cas ; lorsqu'elle l'est, l'auto-évaluation et l'évaluation interne représentent les 4/5 des évaluations⁵³.

⁵² Le laboratoire Éducation, Cultures, Politiques (Lyon 2, Ifé-ENS, UJM Saint-Etienne), agissant dans le cadre d'une convention avec la DEPP, a conduit sa en l'intégrant à des projets déjà en cours sur les évolutions de la professionnalité des enseignants tout au long de leur vie professionnelle.

⁵³ La base recense 600 évaluations externes.

Un tiers des descriptifs de suivi ne font pas référence aux résultats des élèves. Par exemple, des expériences de « Classes sans note » se sont récemment multipliées⁵⁴. Dans un collège visité lors de l'enquête, l'effet des classes sans note, progressivement introduites durant 4 années par niveau, était mesuré par une augmentation des orientations vers la seconde générale, sans référence aux acquis des élèves pourtant directement appréhendables par les résultats du DNB.

D'une manière générale, si l'exigence d'une évaluation est souvent mentionnée, la pratique en est plus limitée et la transposition des résultats difficile car la taille critique des échantillons est rarement atteinte. Un appui méthodologique s'avère nécessaire pour transformer ce qui relève souvent plus d'expériences innovantes que d'expérimentations encadrées par un protocole reproductible, leur valorisation ultérieure est à ce prix.

La DGESCO a indiqué à la Cour que « puisque les évaluations aléatoires, pour des raisons de coût, ne sont pas toujours possibles, il est proposé de réfléchir à la mise au point d'une évaluation adaptée, mais néanmoins rigoureuse, dont les résultats permettraient aux cadres académiques de produire un discours solide sur le dispositif construit. ». Il serait aussi nécessaire d'accompagner ces expérimentations d'un dispositif de tests diagnostiques et de bilan, sous forme numérique, mis à la disposition des équipes éducatives concernées.

CONCLUSION

L'évaluation du système éducatif ne repose pas sur une organisation hiérarchisée, classifiée et cohérente des outils d'évaluation disponibles.

La base de toute fonction évaluative de l'école, c'est-à-dire la connaissance la plus exhaustive des niveaux et des acquis des élèves, en priorité pour les enseignements du socle commun, du début du primaire à la fin du premier cycle du secondaire, n'existe que par échantillon et de façon discontinue dans le temps.

L'évaluation des enseignants obéit avant tout aux enjeux de gestion des corps enseignants, ce qui est aussi sa fonction, mais avec des effets de différenciation des carrières très retreints. Elle exclut la prise en compte de l'efficacité dans la mission d'instruction comme critère. Par sa nature tronquée, elle empêche d'intégrer les résultats de l'évaluation des enseignants dans l'ensemble plus vaste de celle du système éducatif. En raison de la part déterminante qui revient aux enseignants dans la performance collective de l'école, les interactions entre leurs évaluations et d'autres dispositifs d'appréciation devraient être considérablement renforcées. La réforme en cours ne répond pas à cet objectif.

L'évaluation des personnels de direction comme celle des écoles et établissements se heurte à la faible autonomie des unités d'enseignement, comme à l'absence d'une culture de l'évaluation collective (équipe pédagogique, communauté enseignante autour d'un projet d'établissement, évaluation d'une école, d'un collège ou d'un lycée).

⁵⁴ 75 recensées dans Expérithèque.

L'absence d'ordonnement des outils d'évaluation se fait sentir dans le suivi de maints dispositifs pédagogiques et politiques éducatives comme dans celui des expérimentations, qui souffrent encore de réticences au nom du principe d'égalité de traitement des élèves. Il existe des failles dans la conduite des processus évaluatifs : l'impossibilité de mobiliser les résultats des élèves en raison de l'absence de leur sommation grâce à des tests de niveaux et de compétences, des méthodologies hésitantes à la valeur probante non hiérarchisée, des conclusions divergentes non arbitrées, des expérimentations aux objectifs mal définis et trop soumises pour leur évaluation à des impératifs autres que ceux fondés sur une connaissance de leurs impacts.

Chapitre III

Des dépenses diffuses et mal réparties, une efficience mal mesurée

La fonction « évaluation » du Ministère de l'Éducation Nationale n'étant pas identifiée en tant que telle dans les systèmes d'information budgétaire et comptable de l'État, la Cour a procédé à une estimation de l'ordre de grandeur des dépenses correspondantes.

I - L'évaluation des acquis des élèves : des coûts inhérents à l'acte d'enseigner, une rationalisation possible

L'évaluation des élèves revêt trois formes qui emporte des coûts pour le budget de l'éducation nationale ⁵⁵.

A - L'évaluation formative : un ordre de grandeur d'un coût consubstantiel à l'instruction des élèves

L'évaluation en classe par l'enseignant, dite évaluation formative, représenterait de beaucoup le coût principal de l'évaluation puisque les enseignants y consacraient 30 % à 35 % de leur temps que ce soit en classe (20 %) ou par la correction des devoirs (15 %), selon une enquête déjà ancienne et non actualisée⁵⁶ de l'IGEN, portant sur la seule classe de seconde. Les pratiques pédagogiques ayant peu évolué, ses conclusions devraient rester valides.

Dans une approche analytique et à vocation purement indicative, ces taux peuvent être rapportés à la masse salariale des enseignants en admettant qu'ils constituent une moyenne valable pour tous les niveaux d'enseignement. Ce calcul établit l'ordre de grandeur de

⁵⁵ - formative, pour celle réalisée en classe par l'enseignant

- certificative, pour celle qui certifie les acquis par des examens nationaux organisés en fin de cycle

- sommative pour celle qui estime les acquis des élèves en se fondant sur l'analyse des performances d'un échantillon représentatif d'entre eux.

⁵⁶ « *L'évaluation des élèves en classe : ses effets sur l'enseignement et son utilité pédagogique ; sa contribution à la progression des apprentissages* », IGEN, 1998.

l'évaluation formative autour 17 milliards d'euros, dont 7 milliards pour le premier degré et 10 milliards pour le second degré.

B - Les évaluations certificatives : un coût incertain pour une portée évaluative faible

1 - Une certification par les examens qui renseignent peu sur les acquis de connaissances et de compétences

Pour le primaire, il n'existe plus d'examen qui aurait pour vocation de certifier un niveau atteint par les élèves avant leur entrée au collège. Dans le secondaire, les deux principaux examens certificatifs⁵⁷ sont le diplôme national du brevet (DNB) passé à la fin du collège, qui a pour vocation de garantir la maîtrise du socle et le Baccalauréat (BAC) passé en fin de lycée, réputé assurer la capacité des élèves à poursuivre des études supérieures. Il est à la fois le diplôme validant la fin du second cycle du second degré scolaire, et le 1^{er} grade universitaire.

Les examens aboutissant à la délivrance de diplômes par l'État, sont régis par les principes de neutralité, de probité, de confidentialité et d'égalité de traitement des candidats, de la conception des épreuves à la correction des copies et l'attribution des notes. En ce qui concerne les épreuves écrites, cela emporte l'anonymat des copies. Des commissions d'harmonisation présidées par l'inspecteur d'académie de la discipline se réunissent préalablement à la correction pour commenter le barème applicable et veiller à l'homogénéité de correction, ce qui n'entame pas la souveraineté des jurys d'examens une fois les copies corrigées.

Si ces dispositifs ont pour objet de faire respecter les principes présidant à la délivrance des diplômes, le système éducatif poursuit également des objectifs exprimés en taux de succès aux différents examens (80 % d'une génération pour le Bac, 100 % pour le DNB). Les indicateurs⁵⁸ qui enregistrent les taux de réussite à ces examens, renseignent sur la performance de la mission budgétaire MIES depuis la loi de finances 2015.

Alors que les résultats scolaires sont caractérisés par un déterminisme social dont la réalité a été maintes fois objectivée, les taux de succès des académies aux différents examens nationaux sont très peu corrélés avec leur situation socio-économique. Ce constat est d'autant plus étonnant qu'il existe une forte corrélation entre les indicateurs socio-économiques des régions et le taux de jeunes éprouvant des difficultés de lecture tels qu'ils ressortent des tests passés à l'occasion des JDC, tests corrigés par ordinateur de manière identique quels que soient les publics.

⁵⁷ Aux termes de l'article L331-1 du CE, « l'État sanctionne par des diplômes nationaux les formations secondaires ». Les six principaux diplômes nationaux sont le diplôme national du brevet (DNB), le certificat d'aptitudes professionnelles (CAP), le brevet d'études professionnelles (BEP), le baccalauréat général, le baccalauréat technologique, et le baccalauréat professionnel. Il n'existe plus de procédure nationale de sanction des formations primaires, ni des différents paliers du socle.

⁵⁸ Notamment pour le programme n° 141 (secondaire public), proportion dans une génération de diplômés du brevet, dans une génération de diplômés du baccalauréat, et, au sein de la population des jeunes de 20 à 24 ans, des diplômés du second cycle du secondaire (CAP, BEP ou baccalauréat).

L'évolution des performances des élèves français aux évaluations internationales constitue aussi une source d'interrogations par rapport à la valeur certificative uniforme des examens.

Le taux de succès aux épreuves du DNB est ainsi passé de 78 % en 2000 à 87,3 % en 2016 avec des phases de ralentissement (2000-2005, 2010-2011) et des phases d'accélération (2012-2016). Cette amélioration significative ne trouve pas d'écho dans les résultats des enquêtes PISA, qui concernent des enfants du même âge, et qui montrent au mieux une stagnation des élèves français. Dans le même temps les résultats des différentes académies sont devenus plus homogènes : alors que l'écart des taux de réussite entre l'outre-mer et la métropole avoisinait 10 % en 2003 (69 % contre 78 %), il est aujourd'hui pratiquement nul : 87 % pour la métropole⁵⁹ contre 85,6 % pour la Martinique, la Guadeloupe et la Réunion, résultat décalé du taux de personnes en difficultés de lecture qui n'a que très peu varié sur la période⁶⁰.

Les résultats du plus récent *TIMSS advanced*⁶¹ qui estime les compétences en mathématiques et en physique des élèves de terminale sont en tendance particulièrement inquiétants : la France affiche une performance intermédiaire en 2015 et qui se dégrade, sauf pour les élèves de terminale S ayant choisi la spécialité Mathématiques et ceux souhaitant intégrer une classe préparatoire aux grandes écoles scientifique, qui affichent de très bonnes performances.. Entre 1995 et 2015, le score moyen est passé de 570 à 463 (-19 %) pour une moyenne internationale de 500. Cette dégradation affecte autant les 25 % d'élèves faibles, passés en moyenne de 511 à 413, que les élèves forts, passés de 603 à 511⁶². Or, dans le même temps, le taux de réussite au baccalauréat de la série S est passé de 78,5 % à 92 %⁶³. Ces données peuvent seulement être rapprochées pour fonder une interrogation sur la valeur certificative d'un examen par rapport à un indicateur international, mais non corrélées, les résultats au baccalauréat intégrant d'autres matières et faisant l'objet d'une compensation des notes.

Dès lors, les évaluations fondées sur les taux de succès ou les notes moyennes obtenues au DNB ou au Bac ne sauraient être appréhendées qu'avec précaution, pour certifier un niveau d'acquisition des connaissances.

⁵⁹ Il s'agit de la moyenne simple des taux de réussite des collèges.

⁶⁰ Les chiffres ne sont disponibles pour les DOM que depuis 2009 et ils sont stables alors qu'ils sont passés de 9,6 % à 8,6 % pour la métropole ; Les données 2004 pour la France métropolitaine font état de 11 % en 2004 contre 8,6 % en 2014 mais avec une rupture méthodologique dans l'intervalle.

⁶¹ Dans l'édition 2015, l'échantillon représentatif de la France métropolitaine comportait 281 classes de terminale scientifique de 144 établissements. Dans chaque classe, la moitié des élèves participait à l'épreuve de mathématiques et l'autre moitié à l'épreuve de physique, soit 3 967 élèves ayant passé l'épreuve en mathématiques et 3 958 élèves ayant passé l'épreuve en physique.

⁶² « Les performances des élèves de terminale S en mathématiques Évolution sur vingt ans », DEPP, Note d'information n° 35 - Novembre 2016.

⁶³ MENESR DEPP/Système d'information OCEAN et enquête n°60 sur les résultats définitifs du baccalauréat (RERS 2016).

2 - Une inscription budgétaire partielle pour une dépense totale mal connue.

a) La valeur relative des dépenses budgétaires

La dépense budgétaire liées aux examens des élèves est présentée dans l'action « certification » du programme n° 214 « Soutien de la politique de l'éducation nationale ». L'action « certification » regroupe les frais d'organisation des examens de l'enseignement général et technologique (DNB, bac général et technologique, BTS, etc.), des examens de l'enseignement professionnel (CAP/BEP, bac professionnel)⁶⁴.

En 2016, ces dépenses s'élevaient à 164,14 M€, en hausse de 3,9 % par rapport à 2015 (158 M€). S'agissant du baccalauréat qui représente à lui seul 57 % des dépenses de certification (79 M€), le ministère indique un coût moyen par élève présent de 77,80 € pour l'année 2015 (session 2014) avec une prévision pour 2017 inférieure à 85 €. Cette prévision a été ramenée à 80 € dans les projets annuels de performances des exercices budgétaires 2016 et 2017. Le coût moyen du DNB par candidat est faible et stable sur la période (2,50 € en 2005 et 2,80 € en 2015).

Ces montants sont des estimations construites à partir d'une enquête réalisée auprès des seules académies métropolitaines et sont restreintes aux indemnités de jurys et aux frais de déplacement et d'organisation. Les dépenses relatives à la rémunération des personnels chargés de l'organisation des examens sont exclues.

Les rapports d'inspection générale sont circonspects sur la validité des chiffrages budgétaires, même sans aborder la question de la prise en compte des rémunérations des enseignants, qui relève d'une logique de comptabilité analytique et non plus budgétaire. Le rapport des inspections générales IGF, IGAEN, IGEN, CGIET (Conseil général de l'industrie, de l'énergie et des technologies) de décembre 2011 "Proposition pour une évolution du baccalauréat" conclut de façon très prudente : « le coût du baccalauréat s'établirait plutôt entre 90 et 100 M d'euros ». Sur le sujet des indemnités de jurys et du remboursement des frais de déplacement, pour tous les examens, un audit comptable et financier de l'IGAENR et l'IGF en mai 2015 a une tonalité générale sévère sur la qualité des bases d'établissement de la dépense⁶⁵.

b) Un coût analytique non disponible, un débat sur le coût d'opportunité

Les coûts exposés précédemment ne rendent compte ni des dépenses de rémunération des enseignants faisant passer les examens, ni des effets de l'organisation des examens sur l'ensemble des élèves et des enseignants qui n'y participent pas (coût de l'inactivité), soit au minimum 2 jours pour le DNB et 5 pour le Bac qui entraînent la fermeture des collèges et des

⁶⁴ Dont les dépenses de rémunération des personnels chargés de l'organisation de ces examens, les frais de déplacement des membres des jurys, les fournitures et les locations de salles...

⁶⁵ « Un processus complexe, non prioritaire pour les services, servi par des outils informatiques inaboutis », « une réglementation trop complexe », « un décalage perturbateur mal pris en compte entre année scolaire (session des examens) et année budgétaire », « un contrôle du service fait pénalisé par le manque d'information et la multiplicité des données », « une liquidation des dépenses difficile à contrôler tant par l'ordonnateur que le comptable », « des risques de paiement au mauvais créancier », « un contrôle interne à développer ».

lycées. D'autres estimations, fondées sur la valorisation de la perte de 3 semaines d'enseignement, avancent des montants beaucoup plus élevés⁶⁶.

Par ailleurs, la Cour a déjà observé que le coût du baccalauréat subissait l'incidence d'une maquette d'options très ouverte, dont la contribution à sa valeur certificative se limitait souvent à l'obtention de mentions⁶⁷.

C - Les évaluations sommatives par test : un coût modeste et dégressif

Les enquêtes par échantillon, soit à portée nationale, soit pour les comparaisons internationales, sont réalisées par la DEPP, qui conçoit souvent les épreuves, organise leur passation et effectue leur correction, le plus souvent électroniquement. La participation aux comparaisons internationales est plus onéreuse que la gestion des dispositifs nationaux.

L'enquête PISA implique deux types de dépenses: une cotisation annuelle liée à la richesse de chaque pays (289 900 € en 2015 pour la France) et les coûts annuels de l'enquête qui varient selon que l'année est ou non une année de test. Lissés, ils représentent environ 250 000 € par an. Au total le coût annuel moyen de la participation à PISA s'élève à 540 000 €. L'échantillon PISA étant de 6 500 élèves, la charge annuelle par élève est de 83€.

Le coût des différentes enquêtes nationales sur le niveau des élèves est d'environ 6 euros par unité. Au total, la DEPP dépense en moyenne moins d'un million d'euros par an pour la réalisation de tests.

Le coût des tests standardisés et numérisés de compétences est beaucoup plus faible. Hors coût de développement initial, la réalisation des évaluations exhaustives prévues en sixième en 2017, portant sur 860 000 élèves, est estimée à 430 000 euros, soit 0,5 euro par élève (soit la charge de la mise en ligne de l'évaluation, la correction des tests, la diffusion des résultats (élèves et établissement)). Le coût de développement a été faible car l'application utilisée est disponible en « open source » ce qui évite le paiement de licence, son adaptation a coûté 40 000 euros (le coût de développement de l'expérimentation en 2015/2016 s'est élevé à 20 000 euros).

Si ces tests étaient généralisés à tous les élèves relevant du socle commun de connaissances, soit environ 7,5 millions d'élèves, le coût serait de l'ordre de 3,75 M€. Si ces évaluations étaient biannuelles, c'est-à-dire, de diagnostic puis de bilan, les coûts seraient doublés.

Les dépenses engagées pour évaluer le niveau des élèves, formatives et certificatives, n'ont jamais été isolées et objectivées en tant que telles par le MEN. En l'état des informations budgétaires, elles apparaissent au moins incomplètes.

Si une indetermination des coûts pose d'indéniables difficultés théoriques et pratiques pour les évaluations « en classe », une approche de ce sujet ne serait pas sans intérêt. S'il est admis que l'utilisation d'enquêtes par tests, à certains niveaux des cycles scolaires, améliorerait la connaissance des enseignants des niveaux des élèves, faciliterait leur travail et libérerait du

⁶⁶ « Comment sauver le Bac ? », Juin 2016.

⁶⁷ RPT « le coût du lycée » septembre 2015.

temps pour l’instruction, tout en renseignant mieux le dispositif évaluatif d’ensemble et ceci à des coûts raisonnables, ce sujet mérite une vive attention.

En ce qui concerne les évaluations certificatives, leur coût demeure également incertain, très dépendant des optiques retenues pour leur calcul. En tout état de cause, les moyens budgétaires consacrés pour « certifier » ne concourent que peu à une évaluation qui valide les performances intrinsèques du système éducatif.

II - La dépense non pilotée des services chargés de l’évaluation

Les enseignants, les personnels de direction et les établissements font l’objet d’évaluations, en principe systématiques, qui présentent des coûts.

A - Le coût mal cerné des inspections et évaluations locales

- Les 1 500 inspecteurs de l’Éducation Nationale (IEN) qui pilotent le réseau des écoles primaires et maternelles des circonscriptions scolaires du premier degré, consacrent en moyenne, selon les entretiens conduits par la Cour, 25 % de leur temps de travail aux inspections des enseignants du premier degré. Appliqué à leur masse salariale, ce ratio donne un coût annuel des inspections des enseignants du premier degré d’environ 33 M€. Les 1 800 inspecteurs Pédagogiques Régionaux (IA-IPR) et inspecteurs de l’Éducation Nationale du second degré (IEN ET-EG), qui assurent l’évaluation des enseignants du second degré, y consacrent environ 20 % de leur temps. Selon la même approche, le coût serait d’environ 37 M € chaque année. Le coût total des évaluations d’enseignants représenterait environ 70 M€.
- Les directeurs académiques des services de l’éducation nationale (DASEN) et leurs adjoints sont responsables de l’évaluation des personnels de direction qu’ils réalisent en trois temps⁶⁸ : l’ensemble de ces évaluations représenterait entre cinquante et soixante journées de travail dans chaque département, soit un coût global d’environ 4 M€.

De façon très approximative, les dépenses locales d’évaluation des enseignants et personnels de direction avoisineraient 80 M€ par an.

⁶⁸- en recevant pendant 45 mn environ les personnels de direction concernés par la mobilité

- en recevant les personnels nouvellement arrivés, pour l’élaboration des lettres de missions
- en réalisant l’évaluation proprement dite trois ans après la prise de poste.

B - Les dépenses faibles de soutien aux évaluations, modiques pour les actions elles-mêmes

Les dépenses relatives à l'évaluation et au contrôle sont retracées dans l'action n° 2 du programme n° 214 « Soutien de la politique de l'éducation nationale ». L'action « évaluation et contrôle » regroupe :

- les crédits de rémunération et de fonctionnement de la direction de l'évaluation, de la prospective et de la performance (DEPP), du Conseil national d'évaluation du système scolaire⁶⁹ (CNESCO) et des services statistiques des académies ;
- les crédits de rémunération des inspections générales (IGEN et IGAENR). Les missions d'évaluation représenteraient environ 15 % (IGEN) et 6 % (IGAENR) de la capacité des inspections générales.

En 2016, les crédits consommés au titre de cette action se sont élevés à 90,5 M€ (+ 3 % entre 2012 et 2016). Elles sont à plus de 95 % constituées de dépenses de personnel. Aucun indicateur de performances ne se rapporte à cette action, dont le traitement dans les documents budgétaires est succinct. Comme cela a été évoqué, le CNESCO ne fait pas l'objet d'une action budgétaire propre, identifiant les moyens qui lui sont alloués.

Les dépenses de fonctionnement, d'un montant de 3,8 M€ en 2016 (4,3 M€ en 2015), financent principalement des études et des opérations d'évaluation et de prospective conduites par la DEPP, notamment la participation aux enquêtes internationales et la réalisation des tests de niveaux des élèves (*cf. supra*).

Face à ces crédits, le FEJ a apporté entre 2009 et 2015 son soutien financier à 256 projets d'expérimentation dans le domaine scolaire soit pour soutenir des expérimentations elles-mêmes, soit pour financer leur évaluation, pour un coût total de 8,5 M€. De même les programmes d'investissements d'avenir apportent une contribution financière importante à l'évaluation du système éducatif. La partie évaluative (suivi et traitement des données, paiement des chercheurs) de l'expérimentation sur l'apport des techniques numériques aux formations professionnelles, bâtie avec la DEPP, mobilisera pendant 3 ans une vingtaine de chercheurs sur 120 lycées professionnels et 20 000 élèves. Elle sera intégralement financée par les crédits du PIA pour 4 M€, soit 1,3 M€ par an, ce qui équivaut au tiers des crédits annuels de la DEPP pour l'ensemble des actions d'évaluation du ministère.

Les crédits des services supports au cœur de la fonction d'évaluation du MEN sont modestes, ceux consacrés à la conduite d'actions d'évaluation plus encore. Le niveau de celles-ci, toujours inférieures à 5 M€ par an, n'est pas en ligne avec les besoins d'un service public aussi massif, confronté à des débats majeurs sur ses capacités et ses résultats.

Dépasser par redéploiement la modicité des moyens consacrés aux enquêtes évaluatives, ne devrait pas présenter de difficultés majeures au sein de la MIES.

⁶⁹ Créé par la loi d'orientation et de programmation pour la refondation de l'École de la République du 8 juillet 2013.

C - Une mesure de l'efficacité prévue par les lois de finances sans consistance opérationnelle

La loi organique sur les lois de finances (LOLF) prévoit un dispositif d'évaluation de la performance de la mission budgétaire de l'enseignement scolaire (MIES), comme pour toutes les missions budgétaires. Il a vocation à travers des objectifs stratégiques de la mission, des objectifs et des indicateurs par programme budgétaire, dont certains sont classés comme « les plus représentatifs » de la mission, à être l'outil de suivi de la performance des crédits alloués au service public de l'éducation. Cette innovation dans la présentation des lois de finances, instaurée par la Loi organique, aurait dû être un catalyseur de l'évaluation.

1 - Des objectifs stratégiques instables, au caractère opérationnel peu avéré

Les objectifs stratégiques des projets annuels de performance varient chaque année (annexe III) ; un seul des objectifs du PAP 2015 est repris à l'identique dans le PAP 2016 (« Combattre les inégalités tout au long de la scolarité »). Le PAP 2017 comporte 4 objectifs stratégiques, tous nouveaux, quoique situés dans des thématiques proches des objectifs des années précédentes, le PAP 2018 en reprend 2 d'entre eux.

Dans ces notes annuelles d'analyse de l'exécution budgétaire, la Cour a fait plusieurs fois observer que « la variabilité de ces objectifs révèle qu'il n'existe pas d'objectif stable, unifié et global fixé au système éducatif »⁷⁰.

Ces objectifs sont exprimés de façon si générale qu'ils ne permettent pas une déclinaison opérationnelle aisée, la correspondance entre eux et les objectifs et indicateurs par programmes budgétaires, axés sur les cibles quantitatives du parcours scolaire (le taux de redoublement par exemple) est loin d'être assurée de façon exhaustive même si la présentation budgétaire s'y efforce parfois. Par exemple elle relie dans sa présentation l'objectif « Au cœur de la République, une école mobilisée pour ses valeurs » avec l'objectif n° 1 du programme consacré à la vie scolaire « Faire respecter l'école, améliorer le climat scolaire et favoriser l'apprentissage de la responsabilité et de la citoyenneté », suivi par trois indicateurs : le taux d'absentéisme des élèves, la proportion d'actes de violence grave et le taux de participation des lycéens aux élections des « Conseils des délégués pour la vie lycéenne ».

Si la présentation par objectifs stratégiques obéit à des exigences politiques, liées à l'annonce de mesures nouvelles, encore faudrait-il que le corpus des objectifs et indicateurs s'y adapte pour mesurer l'effet de décisions politiques à fort impact budgétaire. Il n'en a pas été ainsi lors de la création de 54 000 emplois supplémentaires. Aucun indicateur n'a été mis en place pour évaluer leur effet, leur affectation précise n'a jamais été renseignée dans les lois de finances comme la Cour l'a observé dans ses notes d'analyse de l'exécution budgétaire.

Une stabilité des orientations affichées serait souhaitable, ne serait-ce que pour suivre l'évolution de moyen terme des indicateurs, quand ils peuvent leur être rattachés, et conclure sur l'efficacité des moyens pour remplir les objectifs. À défaut de cibles stables, cet exercice

⁷⁰ Par exemple NAEB 2016.

qui mobilise chaque année l'administration centrale risque d'être plus formel qu'opérationnel alors qu'il constitue au plan des exigences démocratiques d'évaluation une évolution positive.

2 - Des indicateurs jugés « représentatifs », plus stables mais partiellement renseignés

Cette stabilité est assurée par les objectifs et indicateurs de programme jugés comme « les plus représentatifs » de la mission (« conduire tous les élèves à la maîtrise des connaissances et compétences du socle commun exigibles au terme de la scolarité primaire », assorti de deux indicateurs, et « conduire le maximum de jeunes aux niveaux de compétences attendues en fin de formation initiale et à l'obtention des diplômes correspondants », assorti de trois indicateurs). Ils sont identiques depuis 2014. En outre 3 de ces indicateurs sont depuis 2015 transversaux à tous les programmes scolaires.

Or ces indicateurs qui portent sur des enjeux primordiaux d'efficacité du système éducatif, sont ou la reprise d'objectifs politiques généraux (taux d'accès au baccalauréat), ou le suivi de niveaux de connaissances, non renseignés chaque année. Cette carence tient au caractère pluriannuel de l'enquête sur la maîtrise du socle (étalée sur 6 ans *cf. supra*), elle n'est pas jugée significative par le ministère qui estime la variabilité annuelle de ces indicateurs certainement très faible.

Tableau n° 1 : indicateurs « représentatifs » de la mission

Objectif	Indicateurs dits « représentatifs » de la mission	Référence (PAP 2016)	Cible 2016 (PAP 2016)	Réalisation 2016
1	Proportion d'élèves maîtrisant, en fin de CM2, les compétences 1 et 3 du socle commun – enseignement privé	139.1.2 – langue française 139.1.2 – mathématiques, science et technologies	non renseigné non renseigné	non renseigné non renseigné
	Proportion d'élèves maîtrisant, en fin de CM2, les compétences 1 et 3 du socle commun – enseignement public	140.1.2 - Total	non renseigné	non renseigné
2	Taux d'accès au diplôme national du brevet	Indicateur de mission	84	non renseigné
	Taux d'accès au baccalauréat	Indicateur de mission	78,6	78,8
	Proportion de jeunes âgés de 18 à 24 ans ne possédant ni CAP, ni BEP	Indicateur de mission	8	non renseigné

Sources : RAP 2015, PAP 2016, PAP 2017

Au-delà de ces indicateurs primordiaux, les PAP 2016 et 2017 comptent 55 indicateurs, quand le PAP 2014 en comptait 82⁷¹. Parmi ces indicateurs, 20 sont communs à l'enseignement public et privé. Au total, 36 indicateurs de la MIES se rapportent à des objectifs d'efficacité, 15 à des objectifs d'efficience de la gestion et 7 à des objectifs de qualité de service. La Cour recommande depuis de nombreuses années d'en limiter encore le nombre pour privilégier des

⁷¹ dont 76 à périmètre identique, le programme provisoire 408 - *Internats de la réussite* ayant été assorti de six indicateurs.

indicateurs d'efficience financière qui restent sous représentés (coût de l'enseignement d'une discipline par rapport à son importance certificative d'un cursus par exemple)⁷².

CONCLUSION

Les moyens budgétaires consacrés à la fonction d'évaluation du système scolaire ne sont ni consolidés par le MEN, ni aisés à isoler et à strictement catégoriser. Ils peuvent varier, selon les approches retenues, de plusieurs milliards d'euros à seulement quelques dizaines de millions.

Deux constats méritent une grande attention et devraient conduire à une réflexion approfondie du ministère.

D'une part la lourde architecture des objectifs et indicateurs de performance du budget de l'éducation nationale ne produit pas les fruits attendus. Elle devrait pourtant être le centre de convergence du système évaluatif pour apprécier l'efficience des moyens selon les objectifs définis et offrir ainsi aux autorités politiques un outil de pilotage. Pour maintes raisons, notamment un défaut d'informations de base sur le niveau des élèves, tel n'est pas le cas. On en reste à des cibles « institutionnelles » comme les taux d'obtention du baccalauréat, qui de facto renseignent très imparfaitement sur l'efficacité réelle du système scolaire.

D'autre part la plupart des crédits de l'action budgétaire correspondant à l'évaluation sont en pratique affectés au soutien de la fonction statistique ou à des processus d'inspection ou d'évaluation, dont il a été démontré qu'ils sont peu « connectés » à une fonction évaluative organisée, mais beaucoup plus à la fonction de gestion des ressources enseignantes du ministère.

Surtout le montant des crédits finançant des dispositifs d'évaluation au sens strict (enquêtes sur échantillon, soutien à des expérimentations, etc.) est modique et n'est pas à hauteur des enjeux. À un horizon très proche, la mise en place et l'extension de certains dispositifs, comme les tests de compétences, qui renforceraient les données exploitables pour l'ensemble des acteurs de l'évaluation du ministère, ne sont pas des charges budgétaires hors de portée. Les nouvelles technologies numériques offrent des opportunités d'outils d'évaluation à un coût supportable, adapté à la taille des populations scolaires.

⁷² Cour des comptes, *Rapport public thématique sur le coût du lycée*. La Documentation française, septembre 2015, 133 p., disponible sur www.ccomptes.fr

Chapitre IV

Décider durablement des fondements d'une fonction d'évaluation

Les cadres organique et fonctionnel de l'évaluation du système éducatif français sont instables et éclatés, obérant fortement la capacité des pouvoirs publics à disposer et à émettre une opinion construite et audible sur les performances de l'école.

Cette situation a pour effet de soumettre l'appréciation de l'efficacité du système scolaire français aux seules comparaisons internationales, qui elles, reposent sur une vision et une organisation lisibles pour l'opinion publique. La logique qui préside à ces évaluations comparatives sont facilement appréhendables à la fois sur leurs objectifs : mesurer le niveau des élèves selon les conditions socio-économiques de chaque pays en usant de tests standardisés par discipline et cycles d'enseignement, qualifier et classer selon une échelle partagée la performance collective des systèmes éducatifs.

Le système français n'a pas réagi à cette stimulation bienvenue pour ériger sa propre vision de l'évaluation et désigner la cible centrale, plus prioritaire qu'unique, de l'évaluation du système éducatif, alors que l'école se voyait assigner de très multiples objectifs, pour certains peu mesurables. Aussi la question de savoir ce qu'il fallait évaluer pour éclairer décideurs et opinion publics, est restée sans réponse.

Cet attentisme s'est étendu à l'architecture de la fonction d'évaluation, qui n'a pas été revue. À défaut d'unifier un système par nature polysémique, son ordonnancement aurait été souhaitable et fait défaut. Il doit y être remédié.

I - Approfondir les résultats des enquêtes internationales pour des enseignements concrets

Les conclusions constantes de l'enquête PISA « *Programme for International Student Assessment* » mettent en relief, pour les élèves âgés de 15 ans, le classement moyen des acquis scolaires des élèves français et l'inaptitude du système éducatif à améliorer les résultats des élèves les moins performants, qui en France particulièrement, seraient issus des milieux les plus défavorisés.

Se trouve ainsi posée la question de l'efficacité du système éducatif, qui à travers la recherche d'une plus grande équité, rehausserait sa performance collective. C'est à cette équation complexe que PISA confronte avec insistance le système français.

Or les résultats de PISA n'ont pas été investigués pour en tirer des conclusions opérationnelles. Un effort critique d'appropriation et d'approfondissement s'impose.

A - Une réaction univoque et peu opérationnelle

Dès lors que les scolarités des différents pays laissent persister des écarts de réussite entre catégories d'élèves, une approche comparatiste est nécessaire pour apprécier les motifs propres à chaque système éducatif. Elle peut être conduite en distinguant

- les effets de rendement du système éducatif, c'est-à-dire sa plus ou moins efficacité intrinsèque, tenant par exemple à la pédagogie appliquée, aux programmes mis en œuvre, à l'organisation des parcours scolaires (l'usage du redoublement, l'importance du suivi individualisé, l'existence ou non de classes de niveau, la qualité de l'orientation, etc.), par rapport aux autres systèmes scolaires⁷³,
- les effets de composition de chacune des populations concernées, qui imposent à chaque service éducatif de s'adapter aux déterminismes sociaux et culturels de leur population scolarisée⁷⁴. Cette prise en compte revient à contextualiser les résultats des évaluations des élèves ou des pratiques scolaires au regard principalement des différences de catégories socio-professionnelles.

Un tel effort d'analyse comparative est ardu et requiert des précautions méthodologiques rigoureuses. Mais, à défaut, il est difficile de tirer des enseignements utiles des comparaisons internationales, répondant aux questions suivantes : Sur quoi doit-on agir ? Sur quel type de populations scolarisées l'effort doit-il être concentré ?

1 - Une réaction inutilement dépréciative plus qu'une appropriation approfondie

Les données PISA ont suscité de multiples appréciations péjoratives sur le système éducatif français reprises à son compte par l'institution scolaire de façon très générale.

Les documents budgétaires, pour citer cet exemple emblématique, portent ce discours, le PAP 2015 consacré à l'enseignement scolaire indique que « l'école ne se contente pas de reproduire les inégalités mais qu'elle les aggrave »⁷⁵. Or aucune analyse de PISA, effectuée par le ministère et contextualisée à la situation française, ne vient justifier une telle assertion d'aggravation, ou en général ou pour une population scolarisée particulière. En effet si les

⁷³ L'effet de rendement explique l'évolution des résultats scolaires par l'évolution de la façon dont les caractéristiques des élèves jouent sur leur scolarité. Si la totalité des inégalités scolaires était due à l'effet de rendement, ce n'est pas la société mais l'école qui serait inégalitaire.

⁷⁴ L'effet de composition explique les résultats des élèves par leurs caractéristiques socio-économiques, familiales ou ethnoculturelles et celles de leur environnement éducatif. Si la totalité de la différence des inégalités scolaires était due à l'effet de composition, cela signifierait que ce n'est pas l'école mais la société qui est plus inégalitaire, les inégalités scolaires ne seraient que le miroir de la société.

⁷⁵ Annexe PAP au PLF 2015 consacrée à l'enseignement scolaire, p.84.

résultats ne s'améliorent pas, voire se détériorent, rien n'atteste que l'école en soit à l'origine par un « effet de rendement » qui accentuerait les inégalités.

Selon l'OCDE elle-même, les résultats de PISA contredisent une telle assertion : « Il est juste de mentionner pour la France que les données PISA montrent une amélioration des performances quand la comparaison porte sur les élèves immigrés de la seconde génération en comparaison à ceux de la première génération. L'OCDE a d'ailleurs mentionné dans la Note de PISA 2012 cet aspect qui n'a pas été repris par les médias » (réponse de l'OCDE à la Cour des comptes lors de l'instruction).

Le CNESCO adopte lui-même sur la question des rapports entre déterminismes sociaux et système éducatif une formulation particulièrement ambiguë⁷⁶, allant jusqu'à estimer que des politiques visant à compenser des déséquilibres socio-économiques (éducation prioritaire) constitueraient des « discriminations négatives ».

2 - Clarifier les attentes nationales entre efficacité et équité

Pour sortir d'appréciations si générales qui apportent peu aux autorités politiques pour orienter concrètement leurs actions, une clarification des enjeux posés par PISA serait souhaitable. La poursuite de l'équité soulève en effet des difficultés de principe sur lesquels il eût été nécessaire que le MEN se fixât une doctrine.

Les résultats de l'enquête PISA peuvent être différemment interprétés ce que le secrétaire général de l'OCDE fait observer en avant-propos des résultats de PISA 2012 : « Certains systèmes d'éducation ont montré qu'il était possible de progresser rapidement sans sacrifier l'équité et la qualité. Sur les 13 pays et économies qui ont sensiblement accru leur performance en mathématiques entre 2003 et 2012, 3 ont aussi amélioré le degré d'équité dans l'éducation durant la même période, et 9 autres ont aussi préservé leur degré élevé d'équité – ce qui prouve bien qu'il ne faut pas nécessairement consentir à une baisse du niveau de performance pour progresser sur la voie de l'égalité ».

En effet, dès lors que l'on définit l'efficacité comme la capacité du système scolaire à permettre à chaque enfant de réaliser le maximum de ses potentialités, si celles-ci ne sont pas identiques, l'efficacité peut se traduire par d'importants différentiels de connaissances et de compétences qui ne peuvent être interprétés comme d'inévitables inégalités.

À l'inverse, la parfaite équité pourrait s'accommoder d'un certain degré d'inefficacité. L'indicateur des programmes budgétaires qui vise à mesurer l'évolution de l'équité (« écart de taux de réussite au diplôme national du brevet (DNB) entre éducation prioritaire EP et hors EP réduit à moins de 10 % ») est aussi favorablement affecté par l'amélioration des résultats des élèves en EP que par la détérioration des résultats des élèves hors EP. Dans ce dernier cas, l'équité s'améliore tandis que le « rendement » du système scolaire décroît.

Il est donc indispensable de sortir d'un discours général pour faire émerger de cette enquête des enseignements pratiques.

76 « Comment l'école amplifie-t-elle les inégalités sociales et migratoires ? », Rapport scientifique, CNESCO, Septembre 2016.

La Cour ne méconnaît pas les difficultés d'un tel travail d'analyse. La DEPP, qui conduit un important travail d'explication de la méthode PISA (ainsi que le CNESCO), a souligné auprès d'elle, sans contester les défauts de la situation actuelle, les limites auxquelles se heurte un approfondissement opérationnel : « En termes de comparaisons internationales, ce sont les risques de mauvaise interprétation des résultats liés à une absence de contextualisation qui peuvent conduire à des messages tronqués voire faussés. Parfois, des inférences excessives sont franchies dans les comparaisons internationales lorsque des corrélations sont postulées entre variables par simple fait de rapprochement spontané et/ou lorsque l'on assigne à ces corrélations un pouvoir explicatif...C'est précisément l'un des défis des comparaisons internationales : comment passer du constat à l'analyse causale et aux recommandations. Et c'est souvent là que le bât blesse ».

La DEPP a fait également observer durant l'enquête de la Cour que tels travaux requièrent des moyens plus importants : « la DEPP n'a pas vocation (ni les moyens) de commenter toutes les analyses réalisées à partir des comparaisons internationales. La DeEPP prend cependant compte de la demande de la Cour du besoin d'expertise sur des analyses basées sur des comparaisons internationales et va s'efforcer de développer cette approche ».

Cependant la France dispose d'un savoir-faire et de compétences éprouvés dans les domaines statistique et économétrique, qui lui permet certainement de maîtriser une analyse approfondie de la question posée avec constance par les conclusions de PISA au système français : savoir dans quelle mesure le système éducatif est à l'origine d'inégalités de niveau et d'acquis et pourquoi. De mieux distinguer ainsi ce qui relevait dans les résultats français de l'efficacité intrinsèque du système éducatif (effet de rendement) et de la composition de la population scolarisée par rapport à ceux des autres pays.

Une telle analyse aurait permis de sortir d'un discours univoque et global, inutilement négatif, pour le système éducatif français, discours qui en revanche n'apporte aucune conclusion opérationnelle sur les objectifs à cibler et les modes de remédiation à mettre en œuvre.

B - Des écarts de performance non expliqués et non étudiés

1 - Un défaut d'élucidation de la dégradation de l'équité moyenne du système français

Pour apprécier l'équité des différents systèmes, la direction de l'éducation et des compétences de l'OCDE se fonde sur trois indicateurs :

- La mesure selon laquelle le niveau socio-économique (SESC)⁷⁷ des élèves est une variable prédictive de leur performance. Elle est appréciée par la valeur du coefficient de corrélation

⁷⁷ Ce niveau est défini par l'indice PISA de statut économique, social et culturel (SESC) dérivé des trois indices suivants : le statut professionnel le plus élevé des parents (HISEI), le niveau de formation le plus élevé des parents, converti en années d'études d'après la CITE (PARED) et le patrimoine familial (HOMEPOS). L'indice de patrimoine familial (HOMEPOS) englobe les items des indices WEALTH, CULTPOSS et HEDRES, et inclut la bibliothèque familiale recodée en une variable à quatre catégories (0-10 livres, 11-25 ou 26-100 livres, 101-200 ou 201-500 livres, plus de 500 livres).

entre performance scolaire et niveau socio-économique⁷⁸ d'une part et par la sensibilité moyenne de la performance scolaire au niveau socio-économique⁷⁹ d'autre part. C'est le premier indicateur qui est le plus souvent repris et le plus négatif pour le système français (57 contre une moyenne de 39). *Lignes 1 et 2 du tableau n° 2.*

- La différence moyenne de performance entre les élèves favorisés, c'est-à-dire qui appartiennent au quartile supérieur des scores de niveau socio-économique et défavorisés qui appartiennent au quartile inférieur. Cet indicateur décrit une performance en équité médiocre. *Lignes 3 et 4 du tableau n° 2.*
- La part des élèves résilients, c'est-à-dire les élèves de condition modeste qui sont très performants, figurant dans le quartile supérieur des performances. Là encore les résultats comparés ne sont pas satisfaisants pour la France. *Lignes 5 et 6 du tableau n° 2.*
- Les trois dernières lignes du tableau n° 2 donnent les scores moyens par SESC, permettant d'apprécier en valeur absolue les écarts.

Selon ces différents indicateurs, la performance relative du système éducatif français est inférieure à celle de pays qui lui sont comparables (Allemagne, Royaume Uni, Italie) ou qui sont souvent donnés en exemple (Finlande, Canada). Comme le montre le tableau suivant, l'intensité de la relation entre performance scolaire et catégorie socio-professionnelle est plus forte en France tandis que la résilience y est moindre :

Tableau n° 2 : la performance apparente de la France en matière d'équité (indicateurs PISA 2012)

	France	Allemagne	Italie	Royaume Uni	Finlande	Canada	Moyenne OCDE
<i>Intensité de la relation entre la performance en mathématiques et l'indice SESC en 2012</i>	22,5	16,9	10,1	12,5	9,4	9,4	14,8
<i>Evolution par rapport à 2003</i>	2,2	-6,9	-2,2	ND	-1,1	-0,8	-2,2
<i>Écart de score en mathématiques associé à l'augmentation d'une unité de l'indice SESC</i>	57	43	30	41	33	31	39
<i>Evolution par rapport à 2003</i>	14,4	-1,3	-0,8	ND	5,4	1,3	0,5
<i>Pourcentage d'élèves résilients parmi les élèves défavorisés 2012</i>	4,9	7,5	6,4	5,8	7,4	7,5	6,4
<i>Evolution par rapport à 2003</i>	-2,5	1,9	2,3	ND	-3,3	-1,6	0,3
<i>Score moyen premier quartile SESC</i>	442	467	447	458	488	486	452
<i>score moyen dernier quartile SESC</i>	561	569	522	545	555	558	542
<i>Écart</i>	119	102	75	87	67	72	90

Source : OCDE (2014), Résultats du PISA 2012

⁷⁸ Dans PISA « Intensité de la relation entre la performance en mathématiques et l'indice SESC »

⁷⁹ Dans PISA « Écart de score en mathématiques associé à l'augmentation d'une unité de l'indice SESC »

Cette performance s'est ainsi significativement dégradée depuis 2003 (enfants nés en 1988 scolarisés à partir de 1994 comparés à ceux nés en 1997 scolarisés à partir de 2003) alors que, dans le même temps, celle de l'Allemagne s'est, à l'inverse, significativement améliorée ce qui est d'autant plus surprenant que certaines caractéristiques fortes du système allemand, comme l'orientation précoce⁸⁰ des élèves dans différentes filières ou types d'établissements, sont réputées inéquitables.

La détérioration de la performance française ne laisse pas de surprendre car le système n'a pas subi sur la période d'évolution institutionnelle majeure si ce n'est la mise en place progressive du socle commun de connaissances et de compétence qui aurait dû avoir un effet inverse.

Interrogée par la Cour, la DEPP a mentionné la baisse tendancielle de la scolarisation précoce des enfants comme facteur causal potentiel. Cela ne peut toutefois expliquer l'évolution constatée entre 2003 et 2012 puisque les élèves qui avaient 15 ans en 2003 et en 2012 n'ont pas été concernés par cette baisse, la chute des taux de scolarisation précoce étant postérieure.

Le CNESCO a consacré à cette question une étude⁸¹ qui conclut à la responsabilité, d'une part, de la politique d'éducation prioritaire, par conjonction de la baisse des moyens alloués et de l'effet de stigmatisation lié à l'étiquetage des établissements concernés et, d'autre part, au recul du redoublement réservant désormais cette situation aux élèves les moins performants, pratique considérée comme moins propice aux acquisitions.

Toutefois, les éléments de preuve rassemblés paraissent faibles, pour une conclusion au ton très affirmé. Ainsi après avoir précisé que « l'absence de mesures de compétences passées des élèves [...] empêche d'identifier un lien de causalité entre les changements de politiques éducatives et l'évolution des compétences des élèves (et qu'il) faut donc bien se garder d'interpréter nos résultats de manière causale », l'étude conclut « Ces résultats peuvent s'expliquer par le fait que les effets positifs de l'éducation prioritaire seraient, désormais, contrebalancés par ses effets pervers. »

Au total ces propositions d'explication appartiennent plus au registre des appréciations qu'à celui de l'évaluation fondée sur une démarche scientifique, qu'elle soit de nature qualitative ou quantitative.

2 - Des écarts de performance non expliqués par les paramètres de PISA

A l'occasion de la version 2015 de l'enquête PISA, l'OCDE a rendu publique une analyse dite de régression (ensemble de méthodes statistiques utilisées pour analyser la relation d'une variable par rapport à une ou plusieurs autres), intégrant tous les facteurs jugés pertinents par ses services⁸². L'objectif était de mesurer la part explicative des facteurs retenus par PISA dans la production de ses résultats.

⁸⁰ Rapport annuel IGEN-IGAENR, 2014.

⁸¹ "Why have Inequalities in 15-year-old Cognitive Skills Increased so much in France? CNESCO, 2016.

⁸² Soit 41 facteurs décrivant les caractéristiques des élèves et de leurs classes (score SESC (catégories socio-culturelles), sexe, statut migratoire, parcours, motivation, climat, temps d'enseignement, pédagogie utilisée) et celles des établissements qu'ils fréquentent (Taille, localisation, climat, gouvernance, gestion de l'hétérogénéité, ressources financières et humaines).

Tableau n° 3 : le pouvoir explicatif des données PISA

	Coefficient de détermination⁸³	Coefficient du score SESC
<i>France</i>	55,90 %	16,17
<i>Allemagne</i>	51,10 %	11,86
<i>Italie</i>	39,80 %	6,48
<i>Royaume Uni</i>	30,50 %	15,76
<i>Finlande</i>	28,60 %	25,94
<i>Canada</i>	25,70 %	18,10
<i>Moyenne OCDE</i>	39,80 %	15,06

Source : OCDE, *Factors associated with science performance, PISA 2015*

Il en ressort :

- que le pouvoir explicatif des données PISA ne représente que 40 % de la variance en moyenne mais 56 % en France, laissant 44 % de la variance non expliquée
- que, dans la variance des résultats des élèves français, le facteur socio-culturel n'est pas significativement plus important en France que dans la moyenne des pays de l'OCDE.

En admettant que cette étude soit fondée solidement, le champ d'investigation pour les acteurs français de l'évaluation de l'école est large : élucider près de 44 % des facteurs qui expliquent les résultats de PISA, puis s'attacher à savoir, si l'impact des données socio-culturelles sont dans la moyenne OCDE, quels sont les autres paramètres qui expliquent les écarts de résultats globaux de la France.

Ces différents éléments montrent que les facteurs explicatifs des résultats de l'enquête PISA devraient être approfondis pour en extraire les ressorts propres à la situation du système éducatif français.

Les débats récurrents sur l'efficacité des dispositifs scolaires destinés à réduire les écarts de performance des élèves les moins bons issus des milieux défavorisés, pourraient être utilement éclairés par une véritable appropriation de l'enquête PISA. Ces débats restent, par rapport aux analyses des enquêtes internationales, trop souvent empreints de sommaires positions de principe.

⁸³ Mesure de la qualité de la prédiction d'une régression : le coefficient de détermination varie entre 0 et 1. Lorsqu'il est proche de 0, le pouvoir prédictif du modèle est faible et lorsqu'il est proche de 1, le pouvoir prédictif du modèle est fort.

C - La richesse non exploitée des comparaisons internationales qualitatives

Le MEN s'est peu emparé des éléments d'ordre qualitatif produits par l'OCDE. Il n'est pas possible de retenir l'étude PISA comme seule référence d'une approche comparative, sa dimension statistique n'épuise pas la dimension polysémique de l'évaluation des systèmes scolaires. Qui plus est, elle se limite aux élèves de 15 ans.

Aussi l'OCDE enrichit-elle cette approche d'une multitude de travaux complémentaires (par exemple étude TALIS sur les enseignants, PIAAC sur les compétences des adultes, revues par pays pour évaluer les réformes). Ces travaux sont nourris par la partie statistique de l'enquête PISA, qui en sus des performances des élèves, livre nombre d'informations sur les pratiques pédagogiques à l'intérieur des classes, le climat de discipline, le bien être des élèves (chaque élève passe 30 minutes à répondre à un questionnaire sur ces thèmes).

Les observations de ces études sont peu prises en compte pour orienter l'évaluation du système français. Par exemple a-t-il été envisagé de comprendre pourquoi les élèves français sont les plus anxieux avec les Japonais et les Coréens vis-à-vis de l'enseignement des mathématiques (données collectées en 2003 et 2012) alors même que leurs résultats dans cette discipline se dégradent ?

L'étude PISA donne aussi des informations par pays sur les écoles primaires grâce aux questionnaires remplis par les chefs d'établissement et les parents des élèves. Or la remontée d'informations et l'évaluation des unités du secteur primaire fait défaut au dispositif français, qui reste aveugle sur ce degré décisif de la scolarité.

L'OCDE a publié une synthèse de ces travaux sur la France « PISA 2015 Les défis du système éducatif français et les bonnes pratiques internationales ». Cette revue avait pour but d'aider les décideurs en identifiant les défis et en proposant des solutions à partir des réformes mises en place dans des pays où les résultats se sont améliorés.

Les questions de méthode sont également abordées par l'OCDE, qui relève que si la France est reconnue pour la qualité de ses chercheurs en éducation et leurs travaux, ceux-ci sont peu sollicités pour « évaluer les réformes éducatives, pour créer un dialogue constructif et scientifique avec les politiques et la population ».

Des exemples du levier de l'évaluation dans des pays étudiés par PISA,

- **Le cas allemand** : « Lors des épreuves PISA en 2000 le niveau d'équité sociale en matière d'éducation figurait parmi les plus faibles de tous les pays de l'OCDE.../...le gouvernement fédéral a commencé à subventionner l'enseignement pendant toute la journée. En 2002, un établissement sur dix proposait un enseignement à temps plein ; en 2012, c'était plus de la moitié.../...des évaluations (des élèves) ont été créées afin de mesurer les progrès réalisés par rapport aux normes de référence au niveau fédéral et des Länder. Elles sont menées dans les 16 états fédéraux afin de suivre la performance des différents établissements et du système d'éducation dans son ensemble.../.... Lors de l'enquête pisa 2012, les élèves défavorisés et les élèves issus de l'immigration avaient un indice PISA de statut économique, social et culturel plus élevé que leurs homologues en 2003, et le pourcentage global d'élèves de la première ou de la deuxième génération avait diminué.../...ces changements signifient principalement que l'écart de niveau socio-économique entre les élèves défavorisés et favorisés et entre les élèves issus de l'immigration et les élèves autochtones a diminué entre 2003 et 2012. Ces résultats indiquent que l'amélioration des résultats de l'Allemagne aux épreuves PISA au fil du temps pourrait être due à la mutation du profil social et démographique des élèves » ;
- **Le cas polonais** : « Depuis 2003, sa performance en mathématiques a augmenté à un taux annuel de 2.6 points, passant d'un score de 490 points en 2003 (inférieur à la moyenne de l'OCDE) à un score de 518 points en 2012 (supérieur à la moyenne de l'OCDE). En neuf ans, le pays a réussi à faire baisser le pourcentage d'élèves peu performants de 22 % à 14 %, et a augmenté celui des élèves très performants de 10 % à 17 %.../... chaque étape de l'éducation se termine par une évaluation (dans l'enseignement primaire) ou un examen (dans le premier et le deuxième cycle de l'enseignement secondaire), tous deux conçus à l'échelle nationale et normalisés. Ces évaluations et ces examens donnent un retour d'information aux élèves, aux parents et aux enseignants. Les décideurs politiques aux niveaux national, régional et local peuvent également utiliser les résultats pour suivre la performance du système d'éducation » ;
- **Le cas portugais** : « Au Portugal, c'est en 2006 que les scores ont commencé à augmenter en mathématiques, en compréhension de l'écrit et en sciences. Les scores de mathématiques de 2012 sont supérieurs de 21 points à ceux de 2003 et de 2006.../... En 2003, un tiers environ des élèves fréquentaient un établissement où l'efficacité des enseignants était estimée sur la base des résultats des élèves à des tests ou à des évaluations ; en 2012, la quasi-totalité des élèves fréquentent un tel établissement.../...Le système d'éducation a gagné en efficience par le biais de la réduction de l'absentéisme des enseignants et du remplacement des enseignants absents ».

L'appréhension des résultats des enquêtes de l'OCDE est réductrice. Une exploitation plus large, plus systématique, parfois directement opérationnelle pour guider le choix des évaluations lancées en France ou celui des expérimentations à entreprendre, est à envisager.

II - Remplir une condition première et nécessaire : mesurer systématiquement les acquis de la scolarité obligatoire

Si les buts assignés au système scolaire sont multiples, leur hiérarchisation implicite renvoie aux résultats de la scolarité obligatoire (à partir de 6 ans et jusqu'à l'âge de 16 ans révolus), dont la qualité doit assurer à tous la possibilité d'avoir un avenir scolaire,

professionnel et citoyen, conformément à l'article L122-1-1 du code de l'éducation : « la scolarité obligatoire doit garantir à chaque élève les moyens nécessaires à l'acquisition d'un socle commun de connaissances, de compétences et de culture, auquel contribue l'ensemble des enseignements dispensés au cours de la scolarité. Le socle doit permettre la poursuite d'études, la construction d'un avenir personnel et professionnel et préparer à l'exercice de la citoyenneté ».

Le socle commun n'est pas une idée spécifiquement française, comme le montrent les « *Common core standards* »⁸⁴ aux États-Unis ou les 8 « Compétences clés pour l'éducation et la formation tout au long de la vie » arrêtées par les institutions européennes⁸⁵.

Il est assez logique d'estimer que l'évaluation de l'éducation devrait d'abord privilégier la vérification de l'atteinte de l'objectif de maîtrise des connaissances liées à la scolarité obligatoire. Il est d'autant plus convaincant de privilégier ce niveau qu'il est le plus proche de celui qui sert de fondement à l'enquête PISA.

Pour autant cet objectif central préconisé par la Cour n'est pas exclusif de la nécessité de mieux évaluer d'autres cycles scolaires (par exemple les acquis de connaissance et de compétences atteints au niveau du baccalauréat au sein du parcours « Bac-3, Bac + 3 » qui s'enchaîne avec le socle commun de connaissances). Il n'est pas non plus exclusif de l'obligation d'évaluer des expérimentations éducatives et des innovations pédagogiques. Cependant en l'état actuel de l'évaluation du système éducatif français, ordonner ses cibles en donnant priorité à celle-ci, comme condition première et nécessaire d'une mise en cohérence, présente beaucoup d'intérêt.

Les acquis au titre du socle commun de connaissances sont connus de façon fragmentée, les évaluations formatives réalisées par les enseignants sont peu précises, l'évaluation certificative qui clôt ce cycle (le DNB) a une valeur probante très restreinte et aucun examen ne sanctionne la fin de l'enseignement primaire. Par ces méthodes, les niveaux atteints à la fin de la scolarité obligatoire sont mal cernés.

A - Généraliser l'emploi des tests standardisés

Le système éducatif français n'a réinvesti que très récemment (en 2016 et 2017) le champ des évaluations standardisées systématiques, après leur remise en cause des années 2010. Or de façon paradoxale, il vit sous le jugement de comparaisons internationales nourries par ces pratiques évaluatives.

⁸⁴ “*Common Core is what students need to know and be able to do... a set of clear college- and career-ready standards for kindergarten through 12th grade in English language arts/literacy and mathematics*”, <http://www.corestandards.org/>

⁸⁵ Recommandation 2006/962/CE du Parlement européen et du Conseil, du 18 décembre 2006, sur les compétences clés pour l'éducation et la formation tout au long de la vie [Journal officiel L 394 du 30.12.2006]

1 - Un usage français très en retrait d'expériences étrangères convaincantes

Si la France est à l'origine de la méthode des tests standardisés puisque le premier test d'« intelligence » a été créé par le français Alfred Binet en 1904, elle apparaît aujourd'hui en retrait.

La position de la France dans PISA reste marginale. Dans l'enquête 2012, elle n'a diligenté que deux experts contre cinq pour les Pays-Bas et quatre pour l'Australie et l'Allemagne et aucun en mathématiques alors que c'était l'un de ses domaines traditionnels d'excellence. Dans l'enquête 2015, il n'y avait qu'un seul expert français mais trois allemands, deux britanniques, deux hollandais et huit américains. Elle n'est pas partie au Consortium PISA qui compte deux institutions allemandes, deux belges et deux américaines alors que celui-ci est en charge de la mise en œuvre de l'enquête (« design » des tests).

Son usage de ces tests est excentré par rapport à celui de nombreux pays.

Aux États-Unis, des tests sont systématiquement pratiqués depuis le début des années 60 à tous les niveaux du système éducatif car considérés comme le moyen le plus économique pour mesurer des acquis.

Au Royaume-Uni aussi, les tests sont d'une utilisation ancienne avec d'importantes variations selon les différents territoires. Pour ce qui concerne l'Angleterre, les deux tests obligatoires réalisés en 2e et 6e années ont pour triple objet d'informer les parents, d'aider à évaluer les enseignants et les établissements scolaires.

Les Pays Bas ont mis en place un système de suivi des élèves (LVS, Leerling Volg Systeem) formé d'un ensemble cohérent de tests normalisés à l'échelon national pour l'évaluation des résultats des élèves tout au long du primaire, ainsi que d'un dispositif d'enregistrement de leurs progrès. Le système couvre le langage (compréhension de l'écrit incluse), les mathématiques, les compétences en géographie, en histoire et en biologie, le développement socio-affectif, l'anglais, les sciences et la technologie. Pour chaque élève, un rapport communiqué aux parents illustre graphiquement les progrès accomplis au fil des années.

Si le système éducatif danois est souvent pris en exemple pour l'absence de notation avant 15 ans, il est aussi remarquable par le déploiement de tests adaptatifs c'est-à-dire de tests dont le déroulement est conditionné par les performances des élèves : une réponse incorrecte amène une question plus facile, et une réponse correcte, une question plus difficile. Ceci permet de détecter aussi bien les élèves en difficulté que les élèves précoces, qui peuvent aussi être sujets à l'échec scolaire. Ces tests sont systématiquement réalisés : Danois en 2e, 4e, 6e and 8e années, Anglais en 7e, Mathématiques en 3e and 6e, Géographie Biologie Physique/Chimie en 8e. Les résultats ne sont publics qu'au niveau national agrégé et sont communiqués à la famille, sans les résultats au niveau de la classe ou de l'école.

Au Luxembourg, depuis l'année scolaire 2008-2009, des évaluations externes des acquis des élèves, appelées « Épreuves Standardisées » (ÉpStan), sont conduites qui concernent aujourd'hui toutes les classes de 1ere, 3e, 6e et 9e années. Il comprend, selon l'année d'études considérée, des tests de compétence en compréhension de l'oral et de l'écrit et en mathématiques, un questionnaire destiné aux élèves portant sur les caractéristiques familiales (langue parlée à la maison, origine sociale, etc.) et sur des dimensions motivationnelles spécifiques (concept de soi, anxiété, intérêt, climat de classe, etc.) et un questionnaire complété par les parents (niveau d'études, profession, pays de naissance des parents). Les résultats obtenus par les élèves aux tests de compétence n'entrent pas en ligne de compte pour déterminer le parcours scolaire car l'objectif principal est de fournir à tous les acteurs des informations fiables pour fonder des décisions de gestion du système.

2 - Une vision française négative à dépasser mais des précautions à prendre

Se dessine alors une singularité du modèle français, qui d'une certaine façon, s'autocensurerait quant à l'opportunité et la faisabilité du déploiement d'un tel dispositif, alors qu'il les pratique déjà, ou pour le compte d'organisations internationales, ou pour son propre recueil de données.

Un exemple des réticences à lever est illustré par le rapport distendu entretenu par l'éducation nationale avec le dispositif des Journées Défense et Citoyenneté JDC (dénommé jusqu'en 2011, Journée d'appel de préparation à la défense (JAPD) et instauré en 1998), qui pratiquent des tests de compréhension de la lecture.

Ce n'est que très tardivement, à l'occasion d'un protocole de mai 2016, que les tests passés lors des JDC, sont pris en compte : *« Les données relatives aux jeunes repérés en décrochage scolaire, lors de cette journée, sont transmises aux services chargés de l'orientation et de la remédiation, en particulier les plates-formes de suivi et d'appui aux décrocheurs...Le partenariat sera développé pour faire évoluer les tests de détection de difficulté de lecture auprès des quelques 780 000 jeunes de 16 à 25 ans effectuant chaque année la JDC ».*

La réticence à se saisir des évaluations par test : l'exemple des tests des Journées Défense et Citoyenneté (JDC).

Les services de l'éducation nationale ont longtemps accordé un faible intérêt aux signalements de difficultés d'élèves encore scolarisés qui leur sont adressés par l'institution militaire à la suite des tests réalisés à l'occasion des JDC. Le ministère ne leur demande plus aucune remontée d'information depuis 2012, sans leur avoir jamais donné de directives sur la conduite à tenir.

Pourtant, les investigations de la Cour ont mis à jour des écarts importants d'évaluation qu'il eût été intéressant, voire instructif, d'explorer. Sur le fondement des investigations menées dans cinq académies, elle a constaté un certain nombre de divergences entre les résultats des tests de lecture de la JDC et les avis des chefs d'établissements sur les élèves identifiés : un nombre relativement important d'élèves figurant dans les listes des JDC n'auraient éprouvé, selon les chefs des établissements dans lesquels ils sont scolarisés, aucune difficulté particulière.

Très peu de dispositifs de remédiation sont dès lors proposés aux élèves à la suite des résultats de la JDC. Ainsi, en 2015, sur les 310 élèves identifiés par les JDC dans le département des Hauts de Seine, seuls 18 (6 %) ont pu bénéficier de l'accès à une formation correctrice. Ils n'étaient que 24 (7 %) sur un total de 534 dans le département du Val d'Oise.

La Cour a estimé dans un rapport sur le décrochage scolaire⁸⁶ que l'utilisation de la détection des difficultés scolaires en JDC est très perfectible « En 2011, le ministère de l'éducation nationale a mis en place avec ses partenaires un « système interministériel d'échange d'information » (SIEI) autorisé par la CNIL... Il n'est pas non plus interconnecté avec le système d'information du ministère de la défense, qui organise les journées défense et citoyenneté (JDC) ».

⁸⁶ RPT « Les dispositifs en faveur des jeunes sortis sans qualification du système scolaire » janvier 2016.

a) Des critiques de nature diverse.

Une des critiques porte sur la « *centration des enseignements sur les tests* », c'est-à-dire l'organisation de l'instruction pour répondre aux tests dont les résultats conditionnent moyens des établissements et carrières des enseignants. Cette critique ressort en particulier du modèle anglais de mesure de la performance, désigné sous le terme de « *teaching to the test* ».

Cette pratique, connue depuis longtemps sous le vocable de « *bachotage* », ne disqualifie pas les tests comme outils de connaissance, la critique porte plus sur l'usage fait du test, qui, trop déterminant d'avantages matériels, pourrait être survalorisé dans le processus d'instruction.

La généralisation de la passation de tests standardisés à une grande partie de la population scolaire est aussi l'objet d'une double critique en utilité ; d'une part, il suffirait pour disposer d'une évaluation générale du système de réaliser des tests sur échantillon ; d'autre part, la connaissance du niveau atteint par un élève n'est pas suffisante pour diagnostiquer l'origine de ses difficultés d'apprentissage.

C'est réduire l'utilité des tests à la seule mesure des acquis, en faisant abstraction de leur emploi « opérationnel » par l'enseignant, par l'établissement. En effet, s'ils sont des outils évaluatifs d'une performance collective, ils sont aussi des instruments de connaissance mis à la disposition de la communauté enseignante. Les échantillons, en donnant une indication sur la performance collective, ne permet pas cette information personnalisée, dont l'enseignant et la communauté enseignante ont besoin.

b) Une utilité des tests à garantir par des précautions quant à leur utilisation.

Les écueils existent cependant et imposent des précautions pour garantir une harmonieuse et durable insertion de ce mode d'évaluation dans le système scolaire, notamment son acceptation par les enseignants.

L'inspection générale de l'éducation nationale rappelle que si le caractère systématique de l'enseignement permet précisément d'organiser son évaluation systématisée⁸⁷, pour autant « l'acte d'éduquer comporte un résidu (sic) qu'aucun système d'évaluation ne peut totalement saisir, ni résorber, ni même expliciter. Le risque est de ne nourrir le débat public sur l'éducation qu'à l'aune de ce qui est quantifiable et mesurable, en négligeant les aspects qui ne peuvent l'être (alors que) certains excès des approches des psychométriciens et un risque réel de dérive nominaliste semblent justifier une résistance des enseignants ». Il est donc vital de préciser d'emblée, lors de leur instauration, le cadre d'emploi des évaluations pour en cantonner la portée et les compléter par des approches qualitatives, sur les pratiques et les innovations pédagogiques.

Il s'agit ainsi d'éviter des « tests à forts enjeux » pour piloter des réformes ou des processus d'allocation de moyens. C'est en effet la mise en œuvre dans un certain nombre de réformes éducatives étrangères, de tests « à forts enjeux » qui ont le plus suscité d'inquiétudes. Ainsi, par exemple, certains districts américains ont-ils conditionné le passage dans la classe

⁸⁷ Rapport IGEN n°20120-056 précité

supérieure à l'atteinte d'un niveau minimum⁸⁸, ou fait dépendre la rémunération des enseignants ou les ressources financières, voire la pérennité des établissements, des performances de leurs élèves⁸⁹, mesurés grâce à ces instruments standardisés.

Ces politiques ont eu des résultats contrastés, très positifs selon certaines études, beaucoup moins selon d'autres⁹⁰. Il est surtout apparu que la force de l'enjeu rendait nécessaire une précaution particulière dans l'administration des tests, notamment pour éviter des tricheries rares mais avérées, et dans l'interprétation des résultats, qui pouvaient entraîner des décisions de trop court terme.

Ainsi la définition de leurs conséquences, et notamment des incitations financières fondées sur les performances mesurées, ne saurait être envisagée par principe et d'emblée.

B - Irriguer la fonction d'évaluation des données sur les acquis scolaires

La mise à disposition consolidée de données sur les acquis des élèves, progressivement exhaustives pour toutes les classes du socle commun de connaissances, permettait de multiples déclinaisons dans les composantes du système scolaire qui apprendrait de ses propres résultats.

1 - Privilégier une vocation de soutien des acteurs du système scolaire

- L'aide au diagnostic pour l'enseignant : il s'agit de fournir aux enseignants des outils afin d'enrichir leurs pratiques pédagogiques en évaluant mieux les acquis de leurs élèves leur permettant de diagnostiquer les compétences acquises ou non acquises par chaque élève.
- L'aide au pilotage de proximité : il s'agit de doter les autorités déconcentrées du ministère (recteurs, DASEN, IEN) d'indicateurs leur permettant de mieux connaître et comprendre les résultats des établissements. Les échelons de proximité peuvent ainsi utiliser les évaluations standardisées pour suivre et accompagner une école ou un collège.
- L'aide à la mesure des performances nationales et académiques du système scolaire, en évolutions historiques et par rapport aux comparaisons internationales).

2 - Intégrer les données sur les acquis dans tous les processus d'évaluation

Les modalités de l'évaluation des enseignants ne font, par comparaison avec les pratiques internationales, que peu référence aux acquisitions des élèves, notamment dans le secteur public (*cf. supra*).

Autour des résultats contextualisés des élèves, dès lors qu'ils seraient connus en début et en fin de séquence et qu'ils seraient consolidés sur des séries longues, il serait possible de construire une évaluation individuelle ou collective (équipe de cycle ou d'école dans le primaire, équipe disciplinaire ou par niveau dans le secondaire) qui soit également plurielle.

⁸⁸ Chicago

⁸⁹ Etat du Texas

⁹⁰ "The myth of the Texas miracle in education", Walt Haney, Boston college, Education policy analysis, Volume 8 Number 41 August 19, 2000.

Elle réunirait en sus de l'opinion du chef d'établissement, l'opinion des « parties-prenantes », au premier rang celle des parents recueillis par des enquêtes annuelles de satisfaction, l'auto-évaluation sur la base de questionnaires formalisés. De telles pratiques peuvent contribuer à l'évaluation statutaire sans en constituer pour autant le nouveau paradigme.

La déclinaison des résultats des élèves est également le noyau solide autour duquel pourrait s'élaborer l'évaluation des établissements.

À l'heure actuelle, les auto-évaluations des écoles et des collèges sont rarement pratiquées dans le secondaire et exceptionnellement dans le primaire en raison notamment de la lourdeur de l'exercice et mais aussi de l'absence de données directement utilisables dans le primaire et de leur caractère parcellaire dans le secondaire. Dès lors que seraient disponibles des instruments de suivi des acquisitions des élèves et de la satisfaction des usagers, il pourrait être fait obligation à tout unité d'enseignement du socle commun de connaissance de procéder au moins une fois tous les 3 ans à une auto-évaluation qui servirait de base au projet d'établissement ainsi qu'à l'évaluation de son chef.

La question se pose différemment pour les lycées où la multiplicité des matières et des options, de leur combinaison et des filières, empêche de facto de construire des référentiels d'acquis par élève, standardisés et généralisables. Aussi l'enrichissement des IVAL reste un enjeu important.

Pour assurer la mesure d'un objectif crucial comme les degrés d'acquisition du socle commun de connaissances, le MEN devrait se ressaisir de la possibilité de réaliser des tests standardisés de niveaux et de compétences, au moins dans un premier temps en début et fin des trois cycles du socle commun, puis pour chacune de ses années scolaires. Il lui appartient aussi de réexaminer les questions liées au caractère de diagnostic ou de bilan de tels tests, afin d'éviter les confusions passées.

Ce dispositif renseignerait toute la fonction d'évaluation du MEN sur l'acquisition du socle commun de connaissances, retenu comme un des objectifs premiers du service public d'instruction et comme un de ceux qui garantit le mieux son équité.

Enfin les utilisations qui sont faites des résultats des tests standardisés peuvent grandement varier d'un pays à l'autre, les enjeux pouvant être forts ou faibles pour les élèves, les enseignants ou les établissements. Aussi il est raisonnable de distinguer leur vocation de mesure de la performance collective et d'appui aux enseignants, de tout autre usage, et de les privilégier.

III - Ordonner la fonction évaluative du système éducatif

Renforcer l'impact des évaluations suppose d'en accroître la rigueur et d'en stabiliser le mode de production pour en assurer la cohérence. Il faut aussi en garantir l'indépendance et la transparence pour en améliorer la crédibilité. Il faut enfin établir durablement l'architecture de la fonction d'évaluation en arbitrant un certain nombre de questions, arbitrages qui ont jusqu'à présent fait défaut.

A - Classifier les outils d'évaluation de l'éducation nationale

Au-delà de la question première de disposer d'un référentiel de niveaux scolaires solide et stable, dont l'emploi est déclinable dans l'ensemble du système éducatif, l'éducation nationale est confrontée à la nécessité de conduire de nombreuses évaluations spécifiques. Pour ce faire, elle se cantonne aujourd'hui dans un relativisme qui ne hiérarchise pas les méthodologies.

Or la classification des modes d'évaluation est non seulement possible, elle est un préalable à la rigueur et à la qualité des conclusions. C'est la position adoptée dans un autre champ de l'action publique, celui de la santé, pour lequel une gradation des preuves a été adoptée aussi bien au niveau national qu'international.

Un exemple de gradation des preuves par une autorité chargée de l'évaluation

La Haute Autorité de Santé HAS a défini une échelle⁹¹ comportant quatre niveaux de preuve scientifique :

- Les essais comparatifs randomisés (ECR) de forte puissance (effectifs suffisants) ou leur méta-analyse (Niveau 1, Preuve scientifique établie, Prouvé) ;
- Les essais comparatifs randomisés de faible puissance (effectifs insuffisants), les études de cohortes (Niveau 2, Présomption scientifique, Probable) ;
- Les études de cas-témoins (Niveau 3, Faible niveau de preuve, Accepté) ;
- Les études comparatives comportant des biais importants, les études rétrospectives (Niveau 4, Faible niveau de preuve, Accepté).

En l'absence d'études, les recommandations sont fondées sur un accord professionnel. Si un niveau de preuve suffisant est considéré comme atteint sur un sujet donné, la HAS adopte une recommandation alors que dans le cas contraire, elle ne peut, le cas échéant, qu'arrêter une recommandation par consensus formalisé.

Pour l'éducation, la quasi-totalité des études sont de niveau 4. Si les besoins d'évaluation du système éducatif ne peuvent être assimilés à ceux du domaine de la santé par exemple, il existe indéniablement un besoin de classification des différents processus d'évaluation, qui manquent actuellement d'un référentiel reconnu. L'absence d'une instance extérieure, garante des méthodologies, est une des causes des confusions nombreuses dont le rapport fait état entre plusieurs types d'évaluations conduites au sein du système scolaire.

Certaines formes d'évaluation, telles que les essais comparatifs randomisés, demeurent inusitées, ce qui est regrettable alors que leur valeur probante est élevée. Cette situation s'explique en partie pour des raisons éthiques, par rapport à l'égalité de traitement des élèves. Là encore une autorité chargée de l'évaluation pourrait aider à surmonter des réticences légitimes en adoptant des principes éthiques propres à l'école.

⁹¹ HAS, Niveau de preuve et gradation des recommandations de bonne pratique, Avril 2013.

B - Les conditions de l'établissement d'une architecture cohérente de la fonction d'évaluation

Diverses institutions chargées de l'évaluation se sont succédé depuis une dizaine d'années, dont la composition comme les attributions ont varié, traduisant des hésitations récurrentes sur leur vocation et leurs moyens. Leur positionnement envers les services du ministère a toujours été confus et en définitive source de frictions avec eux. Enfin jamais les différentes entités du ministère concourant à l'évaluation n'ont vu leurs compétences précisées et aussi adaptées à la création de ces structures.

1 - Asseoir la légitimité durable d'une institution chargée de l'évaluation

L'impasse actuelle, constante depuis plusieurs décennies, dans l'établissement d'une architecture stable et claire de la fonction d'évaluation du système éducatif trouve partiellement son origine dans l'absence d'arbitrages sur les conditions d'exercice d'une instance chargée de l'évaluation dont le besoin est pourtant ressenti, quels que soient les gouvernements. Ces débats non arbitrés portent sur :

- la nature de l'organisme, entre instance scientifique (plutôt orientée vers la production d'évaluations et/ou leur validation et qualification pour en crédibiliser la valeur) ou collège d'orientation de haut niveau (plutôt tourné vers l'appréciation des politiques publiques et des recommandations stratégiques), entre régulateur opérationnel de l'évaluation (chargé notamment de veiller à la cohérence de l'organisation de cette fonction au sein du ministère) ou diffuseur d'évaluations à l'ensemble des parties prenantes, associées par ailleurs à ses travaux, et se tenant éloigné des questions d'organisation administrative ;
- le degré d'indépendance accordé, avec des pouvoirs autonomes en matière de programmation de l'évaluation, de validation (saisine automatique de protocoles évaluatifs), d'investigations et de publications ;
- le degré d'autorité prescriptive par rapport au système éducatif, au moins au plan méthodologique dans la conduite des évaluations et expérimentations (avis conforme ou simple, obligatoire ou consultatif, faculté à imposer des évaluations au ministère ou à mener de façon autonome) ;
- La disposition des moyens correspondants aux missions, qu'ils soient nouveaux ou issus de transferts des administrations existantes, ce qui implique leur recomposition, la reconnaissance au libre accès aux données du ministère.

a) La bonne distance à trouver entre indépendance de l'instance et collaboration avec les services du ministère

Si la Cour estime que la légitimité et l'effectivité du rôle d'une telle instance impliquent impérativement la garantie de son indépendance, elle mesure également qu'une définition trop tranchée de celle-ci, pour satisfaisante qu'elle serait au plan des principes, manquerait de réalisme. Elle risquerait de provoquer une coupure entre elle et les services, alors qu'au contraire la montée en puissance de l'évaluation et la diffusion de sa culture exigent une

collaboration constante et confiante. Il semble que celle-ci ait manqué ces dernières années ; dans un tel cas, l'indépendance court le risque de devenir stérile.

Pour ce faire, il faut que la détermination des prérogatives et des compétences de l'instance soit sans ambiguïtés, comprise par les services qui doivent être associés à leur définition et détaillées dans les modalités de leur exercice. Toutes choses qui ont aussi fait défaut lors de l'instauration du CNESCO.

Deux missions clairement définies semblent indispensables.

b) Une instance au cœur de la régulation de l'évaluation éducative

Une instance de cette nature devrait obligatoirement avoir pour mission de forger un cadre fonctionnel et une doctrine méthodologique de la fonction d'évaluation (son premier travail pourrait être après consultation de toutes les parties prenantes à l'évaluation au sein du ministère, d'en proposer au ministre la structuration), de garantir la rigueur scientifique des outils et des processus mis en œuvre, de certifier leurs résultats et de proposer les analyses qui peuvent en être faites, enfin d'assurer leur diffusion, en veillant à une large disponibilité publique des données pour chaque évaluation.

Ces compétences de régulation pourraient être assorties d'un degré plus ou plus moins prescriptif de leur exercice: des saisines obligatoires lors de l'élaboration d'outils (tests par exemple), de protocoles (expérimentations par exemple) ou le lancement de processus évaluatifs (dispositifs évaluatifs d'une nouvelle politiques), assortis d'avis publics, des saisines obligatoires avant la publication des résultats pour en certifier le respect des règles éthiques et méthodologiques.

Cette instance pourrait être saisie par le ministre (ou se saisir) des cas d'évaluations aux résultats divergents pour en donner une analyse publique.

Sa fonction de régulation devrait comporter un rôle central dans la définition d'un programme pluriannuel d'évaluation du ministère, qui serait d'un grand appui pour donner une visibilité de long terme aux évaluations conduites. Il pourrait éviter des remises en cause hâtives de dispositifs et assurer une stabilité des méthodes et des domaines soumis à évaluation. L'élaboration de ce programme, assorti d'un avis public et motivé, pourrait lui être confiée, en concertation avec l'ensemble des services concernés, le ministre restant dans ce cas l'autorité décisionnaire.

Au total sa mission première devrait être de veiller à la qualité et la cohérence des évaluations produites, requérant pour ces domaines une forte dimension prescriptive et un degré élevé d'indépendance.

Si cette mission a été en principe dévolue au CNESCO lors de sa création, Il doit maintenant être procédé à analyse critique de sa mise en œuvre effective, tant il est patent que le CNESCO n'a pas pu jouer ce rôle.

c) Une instance chargée d'apprécier la performance collective du système éducatif, publiant un rapport du niveau de PISA

Un des manques criants de l'organisation de l'évaluation du système éducatif français, est son incapacité à élever son discours en clarté et en portée, au niveau de celui des grandes enquêtes internationales.

Le rôle d'une instance devrait être de forger un discours d'ensemble sur les résultats de la production évaluative française, fondé d'abord sur l'agrégation des résultats des élèves, notamment sur leur degré de maîtrise du socle de connaissances et de compétences requis à la fin de la scolarité obligatoire.

Cette instance devrait devenir un « émetteur » aussi reconnu que le sont aujourd'hui les instances internationales, par exemple par la publication d'un rapport d'ensemble à périodicité pluriannuelle, calée sur la fin des cycles d'évaluation des élèves (Une période de 5 ans pour Cédre à l'avenir, moins à moyen terme).

Une telle ambition requiert une grande indépendance pour que s'instaure progressivement la crédibilité nécessaire. Elle requiert aussi que le ministre mette en place les outils et moyens de cette connaissance (*cf. supra*).

2 - Préciser l'architecture de la fonction d'évaluation

a) Les conditions d'une indépendance et d'une reconnaissance accrues de l'instance chargée de l'évaluation

Garantir l'indépendance de l'instance est d'abord une question institutionnelle.

Si ses missions doivent être définies avec le plus de clarté possible et dans l'esprit d'assurer une collaboration entre elle et les services ministériels, son existence indépendante doit être assurée par une composition de l'équipe dirigeante de haut niveau en veillant à ce que sa présidence bénéficie d'une autorité scientifique reconnue, à ce que ses comités (conseil d'administration ou conseil scientifique) incorporent une représentation internationale importante et une composante significative du monde de la recherche dans des disciplines diversifiées.

La nomination de son président par le président de la république serait par exemple un signe de reconnaissance fort.

La personnalisation juridique n'est pas une nécessité si les conditions de son existence matérielle sont garanties et publiques, par exemple sur le plan budgétaire.

Certes la haute autorité de santé HAS et le Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres) se sont vus reconnaître le caractère d'autorité administrative indépendante, même d'autorité publique indépendante, par la Loi du 20 janvier 2017. La première dispose de prérogatives de droit public (pouvoirs réglementaires, capacités d'ester en justice par exemple) qui ne sont pas nécessaires à l'exercice de ses missions par une instance chargée de l'évaluation du système éducatif. Elles interviennent toutes les deux dans des secteurs où leur pouvoir affecte directement des acteurs économiques de droit privé ou des entités publiques autonomes (les universités) ou indépendantes (les centres de recherche), ce qui n'est pas le cas d'une instance chargée de l'évaluation du système scolaire.

Il n'est donc ni nécessaire pour l'exercice de ses missions, ni justifié au plan juridique, de recommander que cette instance soit une autorité administrative indépendante au sens juridique du terme.

En revanche le fait que les instances successives, y compris le CNESCO, n'aient jamais bénéficié d'une reconnaissance budgétaire autonome, leurs moyens exposés et soumis ainsi au vote du parlement, est une anomalie qu'il convient de corriger. L'action budgétaire « *Évaluation et contrôle* » doit être complétée. La création d'une nouvelle action budgétaire pourrait être étudiée.

La question est aussi de savoir si une telle instance doit disposer de moyens propres en relation avec ses missions ou rester un collège aux moyens de fonctionnement restreints, de facto limités à sa collégialité et à une petite équipe de direction et d'animation. Jusqu'à présent, c'est toujours la seconde solution qui a toujours été retenue.

Deux voies sont possibles pour tenter de sortir d'arrangements de circonstances imprécis, instables et en définitive peu efficaces.

Des transferts de parties de services du ministère à cette autorité sont envisageables au moins pour deux d'entre eux (l'IGEN et la DEPP) dans des limites qu'il appartient à l'autorité politique de déterminer. La mise à disposition fonctionnelle de ces services pour remplir les missions dévolues à cette autorité est également une voie possible, à la condition expresse qu'elle soit définie initialement et précisément. En tout état de cause, maintenir une autorité chargée de l'évaluation du système éducatif dans le format actuel des moyens accordés n'est pas compatible avec l'ampleur de la tâche.

Enfin devrait être envisagée la participation d'autres services ou opérateurs de l'État, celles de l'INSEE et de l'INED notamment, afin que les composantes méthodologique et scientifique soient plus affirmées que par le passé et élargies hors la sphère des administrations chargées de l'éducation nationale. Une représentation du Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres) pourrait être envisagée pour assurer le lien avec l'évaluation de l'enseignement supérieur, notamment au titre du *continuum* Bac -3, Bac +3.

b) La recomposition des fonctions évaluatives internes au ministère de l'éducation nationale

Concomitamment il s'avère indispensable que le MEN conduise une réflexion large sur la recomposition des fonctions évaluatives en son sein, à un moment propice où a lieu une réflexion sur le rapprochement des deux inspections générales.

Dans cette perspective, devaient être abordés au moins les points suivants : la distinction plus nette des fonctions statistiques et d'études évaluatives au sein de la DEPP, l'encadrement du rôle d'évaluateur de la DGESCO, trop confondu avec ses responsabilités de suivi de la mise en place des politiques éducatives, les marges de manœuvre accordées aux rectorats (académiques ou de régions académiques) pour conduire des expérimentations, la reconfiguration des deux inspections générales, qui interviennent souvent conjointement dans le champ évaluatif et qui y occupent une place importante, notamment l'IGEN qui de facto est un acteur opérationnel du système éducatif tout en étant un opérateur majeur de son évaluation.

La première mission d'une autorité chargée de l'évaluation pourrait être d'établir une cartographie des responsabilités et des services au sein de la fonction d'évaluation, pour permettre au ministre de la réagencer.

De même l'établissement d'un programme pluriannuel serait un vecteur puissant de cette recomposition, en faisant arbitrer les champs et les sujets de l'évaluation, les moyens à y consacrer et en donnant une vue d'ensemble des travaux conduits, sur une période donnée.

À titre d'illustration de ce que pourrait être une architecture ordonnée et clarifiée de la fonction d'évaluation, la Cour présente le schéma suivant à partir de la création d'un outil nouveau, un audit de performance des établissements scolaires.

L'architecture fonctionnelle de l'évaluation à partir de la création d'un outil nouveau d'évaluation : l'audit de performance des établissements scolaires.

- Préconisation de l'instance chargée de l'évaluation de mettre en place ce type d'audit,
- Élaboration du dispositif d'audit (pour ses parties pédagogique et d'affectation des moyens) par la DGESCO en tant que tête de réseau des académies et par la DEPP (pour les outils d'évaluation des élèves et des établissements qu'elle gère et pour son soutien méthodologique dans la création de l'outil),
- Validation du dispositif par l'instance chargée de l'évaluation (avis public assorti de recommandations, label...),
- Mise en œuvre par la DGESCO avec les rectorats, selon un ciblage fondé sur les résultats des élèves et des établissements (travaux de la DEPP précités),
- Contrôle de la qualité de la mise en œuvre, puis des effets sur l'allocation des moyens, par l'IGAENR, missionnée par le ministre,
- Évaluation à terme de l'impact sur la performance du système éducatif par l'instance chargée de l'évaluation.

Conclusion générale

Au terme de l'examen de l'ensemble des niveaux et outils d'évaluation concernés ou mis en œuvre au sein du système scolaire, il apparaît que ce service public ne s'est pas organisé pour apprendre des évaluations diverses qu'il effectue ou de celles, venues de l'extérieur, qui l'affectent frontalement, comme les enquêtes internationales.

Les productions évaluatives ne contribuent que marginalement à faire évoluer le système éducatif qui tire peu d'enseignements d'une profusion mal ordonnée de leurs résultats, elles éclairent peu les autorités publiques qui ne peuvent fonder leur action sur des apports trop indécis et peu synthétisés, elles n'atteignent pas un niveau élevé d'expression publique, leur donnant de l'audience dans la société.

Une des raisons de cette situation qui n'est pas à hauteur de l'enjeu, tient certainement à une adhésion faible de tous les acteurs du monde éducatif, à l'idée que les résultats des processus évaluatifs peuvent servir à améliorer leur performance collective. Aussi, parmi les maints obstacles relevés dans le rapport pour développer une culture partagée de l'évaluation, doit figurer au premier chef une volonté publique trop hésitante pour lui donner de la force et une insuffisante détermination pour en faire un levier opérationnel et continu de l'amélioration du système scolaire.

Affirmée dans les principes qui régissent le bon fonctionnement de l'éducation nationale, l'évaluation demeure en pratique d'une portée très restreinte, quand il s'agit de guider l'action publique. En cela l'objectif assigné par le code de l'éducation à l'évaluation des politiques éducatives, celle d'aider l'État à assurer la cohérence du système éducatif, n'est pas tenu.

Aussi la Cour des comptes appelle l'État à rehausser la reconnaissance de la fonction d'évaluation de son système éducatif, à l'organiser concrètement et utilement.

Pour ce faire, elle juge indispensable de se saisir vigoureusement des évaluations internationales et d'en tirer des conclusions pratiques, comme d'autres pays l'ont fait et comme le système d'analyse statistique et scientifique de notre pays le permet. Elle estime que des objectifs précis doivent être assignés à la fonction d'évaluation du système scolaire : consciente de la nécessaire diversité des besoins d'évaluation, le recueil exhaustif et régulier des résultats des élèves durant leur scolarité obligatoire, selon des méthodes sommatives numérisées, lui paraît cependant être la cible première du système évaluatif d'un service public de l'éducation.

Elle considère enfin que la réorganisation des compétences concourant à l'évaluation dans la sphère de l'enseignement scolaire, est une réforme d'ensemble, à entreprendre sans attendre davantage. La dernière partie de ce rapport décrit ce que pourrait être les cadres d'une telle réforme qui ne se limiterait pas à dégager les termes imprécis et précaires d'un compromis, plus ou moins de circonstances, entre une instance chargée de l'évaluation, réputée indépendante sans en avoir les moyens effectifs, et les services du ministère. Privilégiant une optique plus opérationnelle qu'institutionnelle et juridique, la Cour recommande d'instaurer des relations de collaboration et de confiance entre cette instance et le ministère à travers des mécanismes concrets qui garantissent son indépendance matérielle et fonctionnelle, sa capacité

d'expression autonome, sans en faire une autorité ressentie comme étrangère aux services de l'éducation nationale. Aussi, dans le même souci, la Cour donne-t-elle in fine un exemple concret des articulations qui devraient se nouer entre les différentes entités qui conduisent les processus d'évaluation de l'école.

De la clarification des responsabilités des acteurs de l'évaluation dans des processus complexes, inhérents aux caractéristiques et au fonctionnement même d'un aussi grand service public, dépend la crédibilité et de la stabilité sur le long terme de sa fonction d'évaluation. C'est pourquoi une refonte profonde des conditions d'existence actuelles du CNESCO est indispensable.

Enfin cet effort de mise en cohérence entre les outils d'évaluation et les acteurs qui les mettent en œuvre, ce souci d'apporter des garanties d'indépendance et de bonne organisation, ce besoin de disposer de données sur les acquis des élèves, seront mieux acceptés s'ils concourent à la réalisation d'un objectif national que chacun peut comprendre et qui correspond à un besoin pour la société : disposer d'une expression nationale régulière sur le degré de performance de notre système scolaire et son évolution.

La publication d'un tel rapport, qui donnerait du sens et de la visibilité à la fonction d'évaluation du système éducatif, aurait certainement un puissant rôle de catalyseur de la culture de l'évaluation de l'éducation nationale.

Annexes

Annexe n° 1 : échange de courrier entre le Premier président de la Cour des comptes et le Président de l'Assemblée nationale	100
Annexe n° 2 : glossaire.....	103
Annexe n° 3 : objectifs stratégiques de la mission Enseignement scolaire	105
Annexe n° 4 : méthodologies de l'enquête PISA	108
Annexe n° 5 : traitements de la base des livrets personnel de compétences(LPC/LSU).....	111

Annexe n° 1 : échange de courrier entre le Premier président de la Cour des comptes et le Président de l'Assemblée nationale

KCC A1607608 KZZ
06/12/2016

RÉPUBLIQUE FRANÇAISE
LIBERTÉ • ÉGALITÉ • FRATERNITÉ

ASSEMBLÉE NATIONALE

LE PRÉSIDENT

PARIS, LE 6 DEC. 2016

Monsieur le Premier président,

Lors de sa réunion du 24 novembre dernier, le Comité d'évaluation et de contrôle des politiques publiques (CEC) a décidé, en application de l'article L. 132-5 du code des juridictions financières, de demander l'assistance de la Cour des comptes pour réaliser deux enquêtes, la première sur la prise en charge de l'autisme, la seconde sur l'organisation de la fonction d'évaluation du système éducatif.

Je vous saurais gré de me faire connaître dans quels délais la Cour des comptes serait en mesure de transmettre les rapports traduisant les résultats de ces deux enquêtes.

Je vous prie de croire, Monsieur le Premier président, à l'assurance de ma considération distinguée.



Claude BARTOLONE

Monsieur Didier MIGAUD
Premier président
Cour des comptes
13 rue Cambon
75001 PARIS

HÔTEL DE LASSAY 128, RUE DE L'UNIVERSITÉ 75007 PARIS - TÉL. 01 40 63 50 00

Cour des comptes

**Le Premier président**

Le 10 FEV. 2017

Monsieur le Président,

Ainsi que je vous l'indiquais dans mon courrier du 14 décembre 2016, la Cour est en mesure de réaliser l'enquête sur l'organisation et l'utilisation de la fonction d'évaluation au ministère de l'Éducation Nationale que vous avez demandée en application de l'article L.132-5 du code des juridictions financières.

Le rapport sera préparé par la troisième chambre de la Cour. L'instruction a été confiée à M. Pascal Samaran, conseiller maître et M. André Barbé, président de section, assurera le contre-rapport. La Présidente de la troisième chambre, Mme Sophie Moati et les magistrats concernés ont rencontré le 1er février dernier M. Michel Piron et Mme Valérie Corre, députés désignés par le Comité d'évaluation et de contrôle des politiques publiques pour assurer le suivi de l'enquête, afin d'en préciser le champ et le calendrier.

A l'issue de ces échanges, je vous propose que cette enquête aborde les points suivants :

- Une analyse de l'organisation des structures institutionnelles et administratives, tant internes qu'externes au ministère de l'Éducation Nationale, qui concourent à l'évaluation de sa gestion et des politiques éducatives qu'il conduit ;
- Un recensement des dispositifs évaluatifs mis en œuvre et des moyens qui y sont consacrés ;
- Un examen de l'impact de ces dispositifs sur la conduite des politiques éducatives, de leurs expérimentations et de leurs réformes ;
- Enfin l'exposition d'évolutions possibles et nécessaires pour accroître l'efficacité de la fonction d'évaluation, en explorant des voies de réformes tant en ce qui concernent l'organisation des structures que les objectifs recherchés et les méthodes mises en œuvre.

Monsieur Claude Bartolone
Président de l'Assemblée Nationale
Assemblée Nationale
126 rue de l'Université
75355 Paris 07 SP

Le champ de l'enquête portera sur les évaluations relatives :

- aux élèves (examens nationaux, tests sur échantillons, évaluations sommatives et formatives, etc.) ;
- aux personnels, enseignants et chefs d'établissement (inspections individuelles; notations, conséquences sur la rémunération et la carrière, etc.) ;
- aux unités d'enseignement, écoles, collèges et lycées (auto-évaluations, indicateurs de valeur ajoutée, etc.) ;
- aux dispositifs éducatifs spécifiques ou expérimentaux et aux pratiques en tant que sujets d'évaluation, afin de déterminer l'impact de celle-ci sur leur pilotage et les décisions prises sur leur avenir ;
- au système scolaire dans son ensemble (indicateurs LOLF comme enquêtes internationales de type PISA). La prise en compte dans la fonction d'évaluation du système scolaire français de l'enquête PISA sera un point important de cette analyse.

• Pour chacun des dispositifs d'évaluation susmentionnés, le rapport s'attachera à :

- en décrire l'articulation avec les autres ;
- en estimer dans la mesure du possible le coût ;
- en préciser les référents (normes, standards, objectifs, etc.) et les résultats (notes, rapports, etc.) ;
- en analyser la méthodologie (significativité des résultats, portabilité, auto-évaluation, évaluation externe, ex-post ou ex-ante, etc.) et les impacts (sur la carrière des personnels, sur l'orientation des élèves, sur les politiques menées, sur les pratiques pédagogiques, etc.).

• Outre des entretiens et des questionnaires adressés aux principales directions et services concernées du ministère (DGESCO, DEPP, IGEN, IGAENR), au secrétariat général de la modernisation de l'action publique et au CNESCO, des contacts seront pris avec les principales organisations syndicales représentatives des personnels et avec des experts (chercheurs, OCDE, etc.).

Ces diligences sont complétées par des visites en académies (Paris, Lille, Amiens, Toulouse et Marseille) aussi bien services rectoraux que départementaux et unités d'enseignement.

L'enquête s'appuiera également sur l'exploitation des bases de données disponibles portant sur les résultats des élèves (PISA), les inspections des enseignants (personnels partis en retraite en 2015), les caractéristiques ou les performances des établissements (APAE), les descriptifs des expérimentations (Expérithèque) et les statistiques et travaux publiés par la DEPP (RERS, état de l'école, géographie de l'école, notes d'information).

Le rapport vous sera remis au plus tard le 15 décembre 2017.

Je vous prie d'agréer, Monsieur le Président, l'expression de ma haute considération.



Didier Migaud

Annexe n° 2 : glossaire

APAE	Base d'aide au pilotage et l'auto-évaluation des établissements BEP Brevet d'études professionnelles
BEPC	Brevet d'études du Premier cycle
CAP	Certificat d'aptitudes professionnelles
CAPES	Certificat d'aptitude au professorat de l'enseignement du second degré
CARDIE.....	Conseiller académique en recherche-développement, innovation et expérimentation
CC	Contrôle continu
CE.....	Code de l'éducation ou Communauté européenne
CEDRE.....	Cycle des Evaluations Disciplinaires Réalisées sur Échantillon
CLMSAM	« Cours le matin, sport l'après-midi »
CNESCO	Conseil national d'évaluation du système scolaire
CNIRE.....	Conseil national de l'innovation pour la réussite éducative
CNP	Conseil national des programmes
CPES	Classe préparatoire aux études supérieures
CPGE	Classes préparatoires aux grandes écoles
CSE	Conseil supérieur de l'éducation
CSP.....	Conseil supérieur des programmes
DASEN	Directeur académique des services de l'éducation nationale
DEPP	Directrice de l'évaluation, de la prospective et de la performance
DGESCO	Direction générale de l'enseignement scolaire
DGH.....	Dotation globale horaire
DGRH	Direction générale des ressources humaines
DNB	Diplôme national du brevet
DOM	Département d'outre-mer
ÉE.....	Épreuves écrites
E/S.....	Nombre moyen d'élèves par structure (groupe ou division)
EP	L'éducation prioritaire
ESPE	École supérieure du professorat et de l'éducation
FEJ	Fonds d'expérimentation pour la jeunesse
H/E	Ratio global d'heures d'enseignement par élève
HAS.....	Haute Autorité de santé
HCE.....	Haut conseil de l'école
HCéé.....	Haut Conseil de l'évaluation de l'École
IA/IPR	Inspecteurs d'académie - inspecteurs pédagogiques régionaux
ICAL	Indicateurs de valeur ajoutée pour les collèges
IEN	Inspecteur de l'Éducation nationale du 1er degré
IF2R	L'indemnité de fonction, de responsabilité et de résultats
IGA.....	L'inspection générale de l'administration
IGAENR	L'inspection générale de l'administration de l'éducation nationale et de la recherche
IGAS	L'inspection générale des affaires sociales
IGEN	L'inspection générale de l'éducation nationale

IGF	L'inspection générale des finances
IREDU	L'institut de recherche sur l'éducation
IVAL	Indicateurs de valeur ajoutée pour les lycées
JDC.....	Journées défense et citoyenneté
LOLF.....	Loi organique relative aux lois de finances
LPC	Livret personnel de compétences
MCLM.....	Mots correctement lus par minute
MEN.....	Ministère de l'éducation nationale
MENESR.....	Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche
OAPE	L'outil d'auto-évaluation et de pilotage de l'établissement de la DGESCO
OCDE.....	Organisation de Coopération et de Développement Économiques
P.A.R.L.E.R. .	« Parler, Apprendre, Réfléchir, Lire Ensemble pour Réussir »
PCS.....	Professions et catégories socioprofessionnelles
PDMQDC.....	Dispositif « Plus de maîtres que de classes »
PIRLS	Programme international de recherche en lecture scolaire
PISA	Program for International Student Assessment
PLP.....	Professeur de lycée professionnel
PRE	Programmes de Réussite Educative »
PSAD	Plates-formes de suivi et d'appui aux décrocheurs
QCA	Qualification and Curriculum Authority
RASED	Réseaux d'aides spécialisées aux élèves en difficulté
RBP	Recommandations de bonnes pratiques
REP	Réseau d'éducation prioritaire
REP+	Réseau d'éducation prioritaire plus
RERS.....	Repères et références statistiques
RTC.....	Réduction de la taille des classes
SEGPA	Sections d'enseignement général et professionnel adapté
SESC	Statut économique, social et culturel
SGMAP	Secrétaire générale pour la modernisation de l'action publique
STAR	Student-teacher achievement ratio
TAC.....	Essais contrôlés randomisés
TIMSS.....	Trends in International Mathematics and Science Study

Annexe n° 3 : les objectifs stratégiques de la mission Enseignement scolaire

	Nombre d'objectifs stratégiques	Libellés des objectifs stratégiques
PAP 2011	3	1. « Conduire tous les élèves à la maîtrise des compétences de base exigibles au terme de la scolarité primaire » 2. « Conduire le maximum d'élèves aux niveaux de compétence attendus en fin de scolarité et à l'obtention des diplômes correspondants »
PAP 2012	3	3. « Assurer un enseignement général, technologique et professionnel conduisant à la réussite scolaire et à une bonne insertion sociale et professionnelle »
PAP 2013	8	1. « Une ambition : la réussite de tous » 2. « Priorité à l'école primaire » 3. « Consolidation du collège unique et réforme du lycée » 4. « Mieux former et évaluer les enseignants » 5. « Une école plus juste sur le territoire » 6. « La réussite éducative pour tous les élèves » 7. « Un pilotage rénové pour un système efficace » 8. « L'enseignement agricole, un réseau éducatif également en mutation »

PAP 2014	7	1. « Conduire tous les élèves à la maîtrise des connaissances et compétences du socle commun exigibles au terme de la scolarité primaire » 2. « Conduire le maximum de jeunes aux niveaux de compétences attendues en fin de formation initiale et à l'obtention des diplômes correspondants » 3. « Favoriser la poursuite d'études supérieures et l'insertion professionnelle des jeunes à l'issue de leur scolarité secondaire » 4. « Disposer d'un potentiel d'enseignants qualitativement adapté » 5. « Promouvoir un aménagement équilibré du territoire » 6. « Faire respecter l'école, améliorer le climat scolaire et favoriser l'apprentissage de la responsabilité » 7. « Promouvoir la santé des élèves et contribuer à améliorer leur qualité de vie »
PAP 2015	5	1. « Accompagner les élèves dans leurs apprentissages et la construction de leur parcours pour une meilleure insertion sociale et professionnelle » 2. « Combattre les inégalités tout au long de la scolarité » 3. « Former, soutenir et accompagner les équipes pédagogiques et éducatives » 4. « Promouvoir une école à la fois exigeante et bienveillante » 5. « Refonder la politique éducative sociale et de santé en faveur des élèves »

PAP 2016	4	1. « <i>Un cadre rénové pour une école plus juste qui accompagne les élèves dans leurs apprentissages et la construction de leur parcours</i> » 2. « <i>Combattre les inégalités tout au long de la scolarité</i> » 3. « <i>Réussir l'insertion professionnelle grâce à l'obtention des diplômes correspondant aux formations initiales et à une orientation choisie en lien avec le contexte économique</i> » 4. « <i>Favoriser la prise de responsabilité pour améliorer le climat scolaire et respecter l'école</i> »
PAP 2017	4	1. « <i>Priorité au 1^{er} degré et continuité des apprentissages de l'école au collège</i> » 2. « <i>Une école plus juste, au service de l'ambition de la persévérance scolaire et de l'insertion professionnelle</i> » 3. « <i>L'école numérique au service des apprentissages</i> » 4. <i>Au cœur de la république, une école mobilisée pour ses valeurs.</i>
PAP 2018		1 <i>Renforcer l'attractivité du métier de professeur</i> 2 <i>L'école inclusive</i> 3 <i>Une orientation repensée</i> 4 <i>L'école numérique au service des apprentissages</i> 5 <i>Au coeur de la République, une école mobilisée pour ses valeurs</i>

Source Cour des comptes d'après PAP

Annexe n° 4 : présentation de la méthodologie PISA

Plutôt que d'évaluer des connaissances liées à des programmes scolaires, l'enquête PISA (Program for International Student Assessment) se concentre sur l'évaluation de compétences qui seront utiles aux élèves dans leur vie d'adulte. Réalisée tous les trois ans depuis 2000, l'enquête évalue les compétences des élèves en « compréhension de l'écrit », « culture mathématique » et « culture scientifique ».

Chaque fois que les résultats de l'enquête sont publiés, l'attention des commentateurs et des citoyens tend à se focaliser sur le classement des pays publié par l'OCDE. Pourtant, le classement réalisé par l'OCDE doit être interprété de manière très prudente sur le plan statistique. Les scores de chaque pays sont en effet sujets à des marges d'erreurs qui peuvent être plus ou moins importantes selon la taille de l'échantillon d'élèves ayant passé les épreuves de l'enquête dans chaque pays : plus la taille de cet échantillon par rapport à la population considérée est élevée et plus l'intervalle de confiance autour du score de chaque pays est faible, ce qui signifie que le chiffre moyen mis en avant est plus fiable. Dans l'enquête PISA, la marge d'erreur est en général de plus ou moins cinq points : les scores de deux pays peuvent être séparés de quelques points sans pour autant l'être réellement⁹².

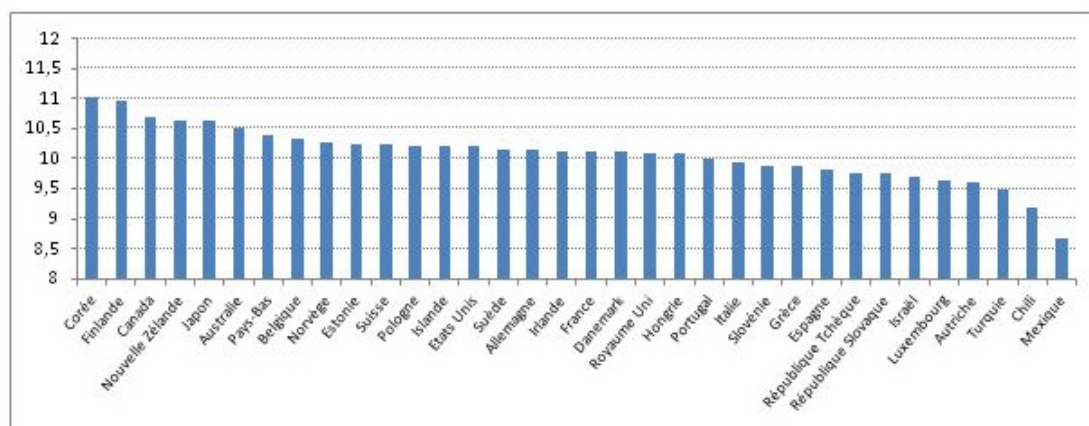
Ainsi il est très peu probable que deux pays séparés par quelques points dans le classement se distinguent l'un de l'autre en termes de performance. Dans l'enquête 2012, la France, qui a obtenu un score de 495, apparaît à un rang plus éloigné que celui de l'Irlande, qui a obtenu un score de 501. Mais les performances de ces deux pays ne sont pas significativement différentes sur le plan statistique. En revanche, la Slovénie, qui a également obtenu un score de 501, pourrait faire significativement mieux que la France car la marge d'erreur autour de son score est plus petite.

Du reste, plutôt que de concentrer son attention sur un classement des pays, l'OCDE préfère identifier trois groupes de pays aux performances comparables : un premier groupe de pays dont les performances sont statistiquement significativement supérieures à la moyenne de l'OCDE (Corée, Japon, Suisse...), un second groupe de pays dont les performances ne sont pas statistiquement significativement différentes de la moyenne de l'OCDE (dont la France fait partie) et, enfin, un dernier groupe de pays dont les performances sont statistiquement significativement plus faibles que la moyenne des pays de l'OCDE (Italie, Espagne, Etats-Unis...).

Les résultats de l'enquête ne doivent donc pas être utilisés pour classer un à un les pays en fonction de leur score car les différences entre les pays sont minimales. Il est vrai que la lecture des scores peut être perturbée par l'échelle des scores choisie, une échelle à laquelle les commentateurs ne sont pas habitués. Par convention, les concepteurs de l'enquête ont choisi de fixer la moyenne des scores à 500 et l'écart-type des notes à 100, de telle manière que les deux-tiers des élèves participant aux épreuves aient un score compris entre 400 et 600.

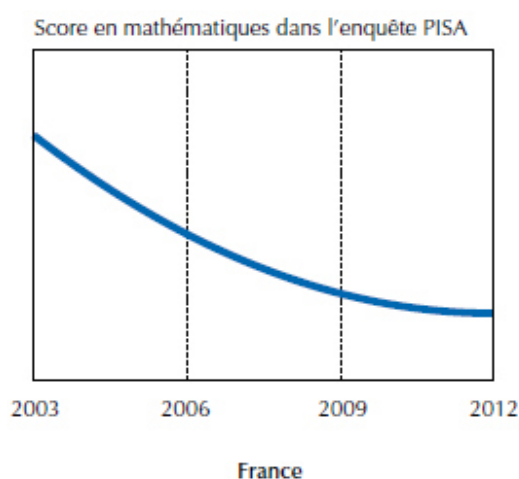
⁹² Les statisticiens diraient plutôt « significativement différents l'un de l'autre sur le plan statistique »

Dans un article publié en 2013⁹³, B. Suchaut montre qu'en ramenant l'échelle des notes de l'enquête PISA à l'échelle française de notation (un classement des notes de 0 à 20 avec une moyenne fixée à 10), les pays ayant participé à l'enquête en 2009 auraient tous un score compris entre 8,7 et 11.



Source : Résultats du PISA 2012 : savoirs et savoir-faire des élèves, performance des élèves en mathématiques, en compréhension de l'écrit et en sciences, Volume 1

Si les comparaisons entre pays s'avèrent délicates, la comparaison des résultats d'un même pays au cours du temps doit aussi obéir à des conditions tout aussi strictes⁹⁴. Mais, grâce à la qualité de la méthodologie de l'enquête, il est souvent possible de procéder à certaines comparaisons temporelles, notamment pour les pays qui ont participé à toutes les enquêtes depuis 2000. Ainsi, il apparaît que les performances de la France en mathématiques se sont dégradées entre 2000 et 2012 et que cette baisse est statistiquement significative. Néanmoins, le profil de cette baisse est moins prononcé ces dernières années.



Source : Résultats du PISA 2012 : savoirs et savoir-faire des élèves, performance des élèves en mathématiques, en compréhension de l'écrit et en sciences, Volume 1

⁹³ PISA à l'épreuve des notes, une autre lecture de la comparaison entre les pays, B. Suchaut, 2013

⁹⁴ Résultats du PISA 2012 : savoirs et savoir-faire des élèves, performance des élèves en mathématiques, en compréhension de l'écrit et en sciences, Volume 1

Une nécessaire contextualisation des résultats

Si les scores des pays à l'enquête PISA sont le résultat d'une méthodologie rigoureuse, il est possible que ceux-ci ne reflètent pas uniquement les performances relatives des systèmes éducatifs car les pays participant à l'enquête se distinguent par de nombreux aspects économiques, sociaux, culturels et institutionnels.

La lecture des résultats ne saurait faire l'économie d'une analyse des spécificités des pays, au risque de mal interpréter les données de l'enquête.

Il est difficile de distinguer ce qui, dans les résultats de l'enquête, est dû à telle ou telle spécificité d'un pays. De telles distinctions doivent être étayées par des études économétriques solides qui demanderaient des investigations spécifiques. Néanmoins, il est possible de formuler quelques hypothèses qui permettent d'expliquer les résultats moyens de la France à l'enquête PISA.

Les travaux de l'OCDE⁹⁵ mettent ainsi en évidence une corrélation positive entre l'indice PISA de statut économique, social et culturel (indice SESC) et les scores des élèves. Cette corrélation semble moins vraie entre les pays. C'est qu'il convient également de mettre en lumière le rôle que l'origine des élèves peut avoir sur leurs résultats. De manière générale, l'OCDE constate que les performances scolaires des élèves d'origine étrangère sont moins bonnes que celles des élèves autochtones. Mais le poids de l'immigration est différent selon les pays : d'après les données PISA, le pourcentage d'élèves issus de l'immigration atteint 3 % en Finlande alors qu'il est de 15 % en France.

Il ne faut pourtant pas en conclure que les performances des pays sont nécessairement influencées de manière négative par le poids de l'immigration. Le Royaume-Uni offre un contre-exemple saisissant : dans ce pays, les performances des élèves issus de l'immigration sont très peu différentes de celles des élèves autochtones. Il est possible que ce résultat s'explique par le fait que les immigrés arrivant au Royaume-Uni sont plus diplômés que ceux qui arrivent en France.

Enfin, il est nécessaire de tenir compte des spécificités institutionnelles de chaque pays dans l'interprétation des résultats. En France, la population des élèves en âge de passer le test PISA se répartit entre plusieurs classes scolaires du fait de la politique, aujourd'hui nettement moins affirmée, du redoublement. Selon J. Grenet⁹⁶, 40 % des élèves ayant passé les tests PISA étaient en retard d'au moins une année sur le plan scolaire. Dans d'autres pays, comme le Royaume-Uni, le passage à la classe supérieure se fait de manière automatique. Il est donc possible que la France soit désavantagée par rapport à d'autres pays en raison des redoublements d'une partie de ses élèves qui passent les tests PISA. Certaines études ont également montré que le redoublement pouvait affaiblir la confiance en soi des élèves qui le subissent et influencer de manière négative leurs performances scolaires.

Les développements précédents rappellent une série de faits qui pourraient expliquer les performances relativement décevantes de la France à l'enquête PISA, qu'il s'agisse d'efficacité ou d'efficacité différentielle. Ils montrent la nécessité d'une étude plus approfondie.

⁹⁵ *Résultats du PISA 2012 : savoirs et savoir-faire des élèves, performance des élèves en mathématiques, en compréhension de l'écrit et en sciences*, Volume 2

⁹⁶ *Pisa, une enquête bancale ?* J. Grenet, La vie des idées, 2008

Annexe n° 5 : traitements sur la base du livret personnel de compétences LPC (LSU)

Le livret personnel de compétences (LPC) est un outil d'enregistrement aujourd'hui électronique de l'évaluation des élèves qui les suit tout au long de leur scolarité entre le CP et la troisième. Les sept compétences⁹⁷, qui forment ensemble le socle commun de connaissances et de compétences, sont : la maîtrise de la langue française (compétence 1), la pratique d'une langue vivante étrangère (compétence 2), les principaux éléments de mathématiques et la culture scientifique et technologique (compétence 3), la maîtrise des techniques usuelles de l'information et de la communication (compétence 4), la culture humaniste (compétence 5), les compétences sociales et civiques (compétence 6) et, enfin, l'autonomie et l'initiative (compétence 7).

Seules les compétences 1, 3 et 6 sont testées en classe de CE1 qui représente le premier palier de l'évaluation. Aux paliers 2 (CM2) et 3 (troisième), toutes les compétences sont évaluées.

L'évaluation est conçue de la manière suivante. Chaque compétence est divisée en domaines qui comprennent eux-mêmes les items de l'évaluation. Par exemple, au palier 2, la compétence « maîtrise de la langue française » se compose de plusieurs « domaines » qui constituent chacun une dimension de la compétence évaluée comme « dire » (pour l'expression orale), « lire », « écrire », etc. Le domaine « dire » regroupe plusieurs « items » d'évaluation comme « s'exprimer à l'oral comme à l'écrit dans un vocabulaire approprié ou précis » ou « prendre la parole en respectant le niveau de langue adapté ». Ce sont sur ces items, qui sont au nombre de 98, que portent les appréciations des équipes pédagogiques.

La Cour a eu accès aux données d'évaluation du LPC de l'année 2014 au niveau de la classe de troisième. L'ensemble de ces données représentent environ 584 000 élèves et 10 500 établissements dans les 32 académies de France métropolitaine et d'outre-mer. Les taux de renseignement des items varient considérablement selon les établissements où sont inscrits les élèves et selon les académies.

Sur 98 items, seulement 21 sont validés pour l'académie de Rennes (taux le plus faible), 46 pour l'académie de Strasbourg (taux le plus élevé). En moyenne métropolitaine, 31 items sont validés, soit à peine un tiers de l'ensemble.

⁹⁷ À compter de la rentrée 2016, il y aura 5 domaines : les langages pour penser et communiquer ; les méthodes et outils pour apprendre ; la formation de la personne et du citoyen ; les systèmes naturels et les systèmes techniques ; les représentations du monde et l'activité humaine.