

**LES ÉVALUATIONS NATIONALES :
UN « OUTIL » INSAISSISSABLE ?
Une analyse de la documentation
institutionnelle sur le dispositif**

Daniel Bart
Théodile-CIREL (équipe d'accueil 4354)
Université Charles de Gaulle – Lille 3

Entre la fin des années 1980 et le début des années 2010, l'École française a été le théâtre d'évaluations nationales qui ont concerné de manière plus ou moins continue tous les élèves de CE2 et de sixième (de 1989/1990 à 2006/2007), de cinquième (par exemple en 2002), des secondes (1992/1993 à 2001/2002) ou de CE1 et de CM2 (de 2007/2008 à 2011/2012). Sans visée explicite de classement, de notation, de certification ou d'orientation, ces évaluations, bien que donnant lieu à la publication de résultats nationaux, étaient présentées comme des « outils » destinés aux enseignants pour diagnostiquer les difficultés de leurs élèves (Ministère de l'Éducation nationale, désormais MEN, 1989, p. 3) ou plus récemment évaluer leurs « acquis » (MEN, 2009).

Généralisé avec le lancement des évaluations CE2/6^e en mathématiques et en français à la rentrée 1989 suite au rapport sur « la réussite à l'école » (Migeon, 1989), cet abord individuel de l'opération les distinguait donc nettement des évaluations par sondage visant à rendre compte des performances globales d'élèves ou du fonctionnement du système, comme par exemple dans les évaluations sur l'enseignement des mathématiques, du français et des « disciplines d'éveil » à l'école élémentaire menées dès la fin des années 1970 (MEN 1981, 1983a et 1983b). Entre autres différences, il convient aussi de noter que dans ces évaluations

nationales, ce sont les équipes pédagogiques et non des enquêteurs qui avaient la charge de la passation des épreuves, du codage des réponses et de la saisie des codes, du traitement et de l'analyse des résultats, de la communication aux élèves et aux familles, etc. sans parler bien sûr des éventuelles actions d'enseignement entreprises en conséquence.

À la suite de travaux sur la passation des évaluations nationales de français dans les classes de primaire (Menouar & Bart, 2014), nous nous proposons de revenir sur cet « outil » évaluatif, inscrivant notre approche en complément des discussions sur les problèmes méthodologiques de ces dispositifs (Hassan, 2005) ou sur leur intégration dans des formes récentes de régulation du système scolaire (Normand *et al.*, 2004), aux analyses de leurs exercices et de réponses d'élèves (Delcambre & Denizot, 1999 ; Reuter & Lahanier-Reuter, 2004) ou encore aux études sur les points de vue d'enseignants (Harlé & Josso, 2013). Les travaux sur ces évaluations institutionnelles (voir aussi *Spirale*, 1990 ; *Recherches*, 2003) paraissent en effet avoir assez rarement investigué l'amont des épreuves, leur documentation, leur « mode d'emploi ». C'est sur cet aspect que notre contribution veut apporter des éclairages complémentaires en analysant la manière dont ces évaluations et leur fonctionnement étaient expliqués dans les publications destinées à ceux chargés de les mettre en œuvre et de les exploiter dans leur classe.

Nous nous livrerons pour ce faire à une étude des discours institutionnels relatifs à ce qui était supposé être évalué, aux instructions et consignes d'effectuation, aux publications des résultats, produits pour la discipline français durant les années 1989-2006 (CE2) et 2007-2012¹ (CE1). Cette étude s'inscrira dans une tradition didactique d'interrogation critique des évaluations (Daunay, 2008), non dans le but de les disqualifier mais pour contribuer à l'éclairage d'un dispositif dont nombre de travaux ont pu montrer par ailleurs les intérêts possibles (Roubaud, 2005 ; Grangeat, 1996), y compris dans des contributions de membres de la Direction de l'Évaluation et de la Prospective (DÉP désormais ; Lévassieur, 1996 ; Émin & Lévassieur, 2007 ; Brézillon & Champault, 2008) ou de chercheurs ayant participé à ces dispositifs (Rémond, 1996).

Après avoir présenté ce protocole national d'évaluation et le corpus documentaire que nous avons construit, nous explorerons donc successivement quatre composantes majeures du dispositif discursif : la présentation des fonctions de l'évaluation tout d'abord, l'exposé des contenus disciplinaires évalués ensuite, puis les tableaux de présentation synthétique de l'opération et enfin la désignation des regroupements d'items en vue des analyses. Pour chacune de ces composantes, nous nous intéresserons tout particulièrement aux interrogations que génère le discours, aux problèmes, tensions ou dissonances qu'il véhicule au regard des certitudes institutionnelles quant à son intelligibilité ou son appropriation². Ce travail

-
1. Les évaluations CE1 occupent une place moins centrale dans notre corpus dans la mesure où les protocoles concernés ont changé annuellement dès après leur mise en place en 2007 en raison de nombreuses critiques sur la qualité du dispositif (Harlé & Josso, 2013), un changement de fonction (du diagnostic à l'évaluation des acquis) et avant qu'elles ne deviennent facultatives à la suite du changement de gouvernement de mai 2012.
 2. Notre démarche conduira à mettre en relation des discours couvrant près de 25 ans, sans pour autant que ces derniers aient été nécessairement produits avec le souci de constituer un ensemble soumis à

nous conduira notamment à nous questionner sur l'assemblage de deux perspectives fortes : d'une part la mise en avant du rôle des enseignants dans la mise en œuvre et l'exploitation des évaluations nationales et d'autre part une présentation du dispositif, de la méthodologie retenue et de ses éventuelles conséquences (intérêts, limites, difficultés, etc.) sur le mode de l'évidence.

1. QUELQUES INDICATIONS SUR LES ÉVALUATIONS NATIONALES ET LA DOCUMENTATION CONSTITUÉE

Sur une période de près de 25 ans, des millions d'élèves et leurs enseignants ont donc été concernés par la passation d'évaluations nationales à visée diagnostique à l'entrée au cycle 3 (1989-2006³) puis d'évaluation des acquis en fin de cycle 2 (2007-2012). Si le contenu de l'opération a pu varier (voir infra) son schéma général est resté à peu près identique : les épreuves « papier-crayon » en français et mathématiques étaient regroupées dans un ou deux cahiers organisés en quatre ou cinq séquences indivisibles de vingt à trente minutes d'exercices disciplinaires, soit quelques dizaines d'exercices en tout par discipline (voir un exemple d'exercice en annexe 1). Le MEN diffusait en outre aux enseignants un ou plusieurs documents rassemblant une série d'informations (présentation de l'opération, calendrier de réalisation, etc.) et stipulant les consignes de passation et de codage des réponses (cf. annexe 1). On pouvait y trouver également des commentaires concernant les tâches demandées aux élèves, leurs erreurs possibles ou encore des suggestions de remédiation face aux difficultés observées. Enfin, le protocole annuel était complété par un instrument papier et informatique (notamment les logiciels Casimir et J'Ade) permettant aux enseignants de relever les codes affectés aux réponses des élèves et de réaliser des traitements statistiques (par exemple : calculer le score moyen des élèves d'une classe en « Production de texte »). Au plan national, des résultats annuels étaient ensuite calculés par la DÉP sur un échantillon et publiés par Minitel puis Internet et repris en édition papier (par exemple : DÉP, 2003).

C'est à cet ensemble documentaire faisant partie de ce que le MEN (1995, p. 3) dénommait les « instruments » de l'évaluation, que nous avons choisi de nous intéresser dans cet article. Malgré l'importante quantité de documents générée et diffusée chaque année par le protocole, l'accès à la série complète en vue d'une étude un tant soit peu systématique n'est pas aisé : si le web permet de consulter les

une telle lecture (nombre de documents étudiés comportent d'ailleurs des passages communs d'une année sur l'autre). Notre méthodologie tend donc à mettre en lumière des divergences par le seul fait que les auteurs n'avaient peut-être pas envisagé une telle mise en perspective. Mais cela serait alors déjà un élément intéressant sur la conception institutionnelle de l'outil. Je remercie Isabelle Delcambre de m'avoir alerté sur ce point et plus largement pour l'aide qu'elle m'a apporté dans l'écriture de cet article. Je remercie également Bertrand Daunay, Isabelle Delcambre, Catherine Mercier et Bruno Trosseille (chef du Bureau de l'évaluation des élèves à la DÉP), pour m'avoir grandement facilité le travail de collecte des cahiers d'évaluation.

3. Pour les rentrées de 1996 et de 1997, l'obligation de procéder aux évaluations de rentrée CE2/6^e ne s'est appliquée alternativement qu'à l'un des deux niveaux d'enseignement (MEN, 1999), pour faciliter et inciter à leur utilisation (MEN, 1996, p. 3).

versions numériques des documents les plus récents⁴, seuls des documents épars sont recensés dans le système universitaire de documentation (comprenant les bibliothèques des ESPÉ) tandis que les services territoriaux du MEN contactés ont pu nous renvoyer... vers les armoires des écoles. Cela ne surprendra évidemment pas de la part d'une institution à la peine pour conserver les traces de sa propre histoire. Mais on peut aussi se demander si ne s'illustre pas là une certaine conception de ce qu'est une évaluation pour nombre de discours promouvant « accountability », culture de l'évaluation ou du résultat, etc. : l'exploitation des mesures que ces évaluations produisent et les prises de position les concernant pouvant être relativement déconnectées d'un intérêt pour la « matérialité » même des tests, leur contenu, leur évolution dans le temps, etc. C'est donc précisément à une étude des discours institutionnels sur l'évaluation construits dans les livrets de présentation des épreuves aux enseignants et les publications des résultats annuels que nous souhaitons contribuer ici.

2. UNE ÉVALUATION SANS LOGIQUE NORMATIVE NI CLASSANTE ?

L'examen des discours des concepteurs des évaluations nationales concernant les traitements statistiques et interprétations qu'elles permettaient montre qu'il est en fait assez difficile de se faire une idée nette des analyses et inférences que ces derniers encourageaient. Ainsi, tout en diffusant des scores moyens annuels qualifiés de « références nationales » (Direction de la Programmation et du Développement – DPD désormais – 2002, p. 9), les documents institutionnels insistaient largement sur le fait que ces chiffres ne rendaient compte « ni de la diversité des élèves ni de celle des situations proposées » (*ibid.*), qu'il s'agissait de « grandes tendances », loin d'« un étalonnage absolu des capacités et compétences des élèves » (DÉP, 1992, p. 5).

Ces résultats étaient donc assortis de réserves méthodologiques plus ou moins explicites, signalant par exemple que les différences de pourcentage observées n'étaient « en général » pas statistiquement significatives en raison des erreurs d'échantillonnage (DPD, 2000, p. 19) ou que l'outil était inadapté pour mener des études diachroniques sur les évolutions des performances (MEN, 1999). Dans un commentaire sur les variations des taux de réussite aux items, le livret de présentation précise ainsi : « Comme chacun le sait, les épreuves de l'évaluation nationale sont modifiées chaque année. Il est donc impossible de traiter ces résultats en les comparant par rapport aux années précédentes, la marge de fluctuation étant essentiellement liée aux variations de la difficulté des nouvelles tâches proposées. » (*ibid.*, p. 2). Au-delà des variations liées aux programmes, des exercices particulièrement bien réussis une année pouvaient en effet être volontairement abandonnés au profit de tâches plus complexes pour renforcer le pouvoir discriminant du test (DÉP, 1992, p. 5) tandis que des exercices pouvaient être repris

4. Les archives 1997-2006 des évaluations de rentrée CE2/6^e sont disponibles sur le site ministériel : <http://cisad.pleiade.education.fr/eval/>.

des épreuves antérieures tout en étant partiellement modifiés⁵ en fonction des difficultés qu'ils avaient pu poser aux élèves. Ainsi, « comparer ces résultats sur plusieurs années comme s'ils étaient comparables serait le reflet d'une mauvaise connaissance, voire d'une totale incompréhension d'une opération dont la finalité essentielle est pédagogique, et tendrait à lui donner une fonction normative qui, précisément, n'est nullement son but. » (DPD, 1999, p. 2).

Les publications annuelles des résultats ne mentionnaient d'ailleurs pas les scores des années précédentes. Autant de visées évaluatives et de choix méthodologiques qui limitaient particulièrement les usages possibles des données publiées et leur constitution en « performances » (Daunay, 2008). Ce qui est intéressant car rappelons, avec Vogler (1996), qu'en cette fin de décennie 1980 où se généralisaient les évaluations nationales, le Ministère se préoccupait du développement de ses propres dispositifs de mesure face à de vives controverses concernant des insuffisances supposées de l'École ou l'illettrisme des élèves (par exemple : DÉP, 1988). Notons toutefois que les plus hauts responsables ne s'embarrassaient pas toujours de cette prudence méthodologique : Chartier & Hébrard (2000, p. 657-659) montrent ainsi comment certains ministres ont pu, selon les nécessités politiques, chercher à utiliser ce dispositif pour rendre compte des variations de niveaux des élèves.

La large diffusion des « repères extérieurs » était donc plutôt justifiée par la DÉP en ce qu'ils étaient supposés constituer des appuis aux réflexions et à la formation des maîtres, par exemple en permettant de « relativiser ce qu'il(s) considère(nt), localement, comme étant un bon ou un moins bon résultat » (MEN, 2000, p. 8). Toutefois les écrits modéraient là encore la portée de ces appuis puisque la fin de la phrase précédente indiquait par exemple que ceux-ci « ne donnent évidemment pas de critères pour caractériser et apprécier le savoir lire des élèves » (*ibid.*). Ainsi, le discours institutionnel vantait la qualité d'un « outil » de « diagnostic individuel » qui, « indépendamment de tout jugement de valeur » (MEN, 2001, p. 8), permettait aux enseignants de mieux connaître les élèves et leurs difficultés, de conforter leur réussite, d'informer les parents, de construire des stratégies de soutien individuel ou des projets de cycle, etc. Mais il rappelait aussitôt qu'il ne s'agissait pas d'« étiqueter » les élèves (*ibid.*) et que ces résultats⁶ étaient limités par le fait que les compétences étaient en cours d'acquisition, qu'il s'agissait d'un diagnostic et non d'un bilan du cycle 2 (MEN, 2002, p. 8), que les écarts provenaient aussi des progressions variables dans les enseignements des cycles, ou que « les performances des enfants ne sont à cet âge ni stabilisées ni entièrement significatives » (DÉP, 1997, p. 19).

-
5. L'exercice en annexe 1 est ainsi une création de 1998, transformée en 2000 et reprise en 2004. Certains protocoles étaient toutefois construits spécifiquement à des fins de comparaison. Par exemple, sur les 18 exercices proposés en 2000, un seul était nouveau (MEN, 2000, p. 2). À l'inverse, la quasi-totalité des exercices de français est par exemple renouvelée en 1993 quand bien même le cadre évaluatif était resté constant (MEN, 1993, p. 1).
 6. Même si cette prudence est devenue moins nette lorsque les évaluations de rentrée de CE2 ont été abandonnées au profit d'évaluations des acquis en fin de CE1 à partir de la rentrée 2007 (Harlé & Josso, 2013).

On retrouvait également cette modalité de discours émoussant les aspects saillants et les fonctionnalités prétendues de l'évaluation, dans le fait de présenter à la fois les contenus évalués comme irréductibles à un standard, un modèle, ou un « noyau dur » du programme sur lequel il faudrait mettre l'accent » (MEN, 1995, p. 1) tout en établissant une nomenclature hiérarchisée des compétences de base à partir de 1995-1996, en faisant état d'un niveau exigible de compétences à atteindre à la fin du cycle 2 en 2001 ou en distinguant des « exercices cibles⁷ » en 2007-2008. Il s'agissait peut-être là de déjouer les controverses concernant la course à la performance et au « bachotage » que favorisait potentiellement la diffusion large des résultats et des épreuves. Plus largement, la documentation institutionnelle semblait d'ailleurs chercher à montrer cette évaluation comme dénuée de normativité. Chaque livret de présentation des épreuves aux enseignants mettait de la sorte en exergue jusqu'à 2006-2007 la diversité des contributeurs impliqués dans la construction des protocoles. Cette construction était même qualifiée par l'institution de « travail d'équipe » (Jospin dans DÉP, 1989, p. 1) ou de « démarche participative » (DÉP, 1990, p. 1). On verra peut-être dans cette recherche de « consensus » (Levasseur, 1996, p. 106) une explication possible au type de discours produit, dont nous venons de montrer qu'il proposait assez peu de logiques de fonctionnements et d'usages des évaluations fermes. Mais on y verra peut-être aussi une contribution à cette forme de réserve dans l'explicitation des contenus évalués que nous allons aborder dans la partie suivante.

3. « SAVOIR LIRE » ET « SAVOIR ÉCRIRE » MAIS « SAVOIR LIRE » D'ABORD⁸

De façon un peu surprenante, on trouve assez peu de détails sur la manière dont le français était appréhendé dans cette évaluation disciplinaire présentée, dès son lancement, comme une évaluation de la lecture et de l'écriture ou du « savoir lire » et du « savoir écrire ». C'est ainsi d'ailleurs que Lionel Jospin avait pu circonscrire le protocole au « champ des apprentissages fondamentaux : lecture, écriture, mathématiques » dans sa lettre aux enseignants du 28 juillet 1989 (MEN, 1989, p. 1). Quasi-reprise de la devise programme « lire, écrire, compter » (DÉP, 1988) dont le caractère consensuel et éprouvé semble avoir permis d'économiser des explications sur les contours, l'intérêt ou les contenus évalués au-delà des listes de compétences ou de tâches⁹. Ce sont plutôt des modalités évaluatives – mais pas les problèmes qu'elles posent (Dabène, 1994) – qui semblent avancées pour expliquer aux enseignants ce découpage puisque « plusieurs objectifs n'ont pas pu être retenus

7. Les « exercices cibles » étaient relatifs à des compétences considérées comme les soubassements des apprentissages construits en fin de cycle 2 (MEN, 2007, p. 4). Cette typologie d'exercices n'a semble-t-il été utilisée que pour l'évaluation de l'année 2007-2008.

8. Cf. le numéro 23 de *Recherches* (1995) : « Écrire d'abord ».

9. Parmi la quinzaine de compétences évaluées en 2001 (MEN, p. 9), on trouve par exemple dans le champ « Savoir lire : compréhension » : « Reconstituer la chronologie des événements dans des textes de statut varié » ou « Comprendre et savoir appliquer les consignes courantes du travail scolaire ».

pour des raisons techniques de passation (lecture et expression orales par exemple...) et d'autres [...] ne figurent pas dans le champ de l'évaluation pour des raisons matérielles de durée totale de passation » (MEN, 1989, p. 3).

Les discours semblent d'ailleurs plus soucieux d'atténuer d'abord cette partition lecture-écriture *via* la nécessité rappelée aux maîtres de penser ces deux ensembles de façon solidaire. Ainsi, les premières lignes du chapitre « Évaluation du savoir lire et du savoir écrire » de 1991 (p. 6) précisent que « les pratiques pédagogiques lient naturellement le lire et l'écrire : leur étroite interpénétration montre qu'il ne saurait y avoir aucune prédominance ni antériorité de l'un sur l'autre », affirmation qui pourrait surprendre nombre d'analystes des relations asymétriques de l'écriture et de la lecture. Relation tellement naturelle ici du reste, qu'à partir de 1995 (et jusqu'à 2003), elles seront considérées comme deux activités « indissociables » en référence, non plus aux pratiques pédagogiques mais... aux « nouveaux programmes » (MEN, 1995, p. 8). Par la suite, il sera même question d'un « savoir "lire-écrire" » au singulier (MEN, 2001, p. 7). Le problème semblait de toute façon assez simple puisque dès le protocole de 1990 (p. 7) on pouvait lire : « Il est clair en effet que celui qui écrit bien sait par essence bien lire. »

La partition lire-écrire se trouvait plutôt justifiée par des exigences pratiques de l'évaluation : « En revanche **l'évaluation rend nécessaire** une approche successive des domaines » (MEN, 1991, p. 6). Or il est intéressant d'observer l'évolution de cette formulation : elle devient en effet à partir de 1995 (p. 8) : « L'évaluation met toujours en jeu ces deux activités mais **tente d'analyser** les différentes composantes de l'une et de l'autre ». Puis la formulation devient à compter de 2002¹⁰ (p. 16) : « L'évaluation met toujours ces deux [etc...]. Cependant, chaque exercice **permet de voir**, à travers les différents items, la complexité de ces compétences ». Ainsi, comme nous le verrons pour d'autres aspects du dispositif dans la suite de l'article, une contrainte (ici méthodologique) devient subrepticement dans le discours une opportunité, sans qu'aucun élément ne soit avancé pour expliciter ce changement.

Cette approche de l'évaluation en français qui distingue l'écriture et la lecture sous-tend néanmoins tous les livrets de présentation des protocoles de CE2 et CE1¹¹, y compris lorsqu'un domaine « Espace et temps » sera temporairement adjoint (voir la partie suivante). Il est donc d'autant plus surprenant de constater qu'il n'y a pas de scores moyens calculés dans ces deux domaines¹², excepté entre 2002 et 2004 (DÉP, 2003, p. 14-15). La publication (non explicitée) de ce calcul permet toutefois d'observer que si ces deux domaines sont présentés comme également majeurs, ils ne semblent pas bénéficier d'un poids comparable : l'évaluation de la lecture représente en effet un peu plus de 60% des items du français en 2002 et 2004. Cette hiérarchisation classique semble ancienne¹³ dans le dispositif : sans que ce choix soit

10. Cette mise en avant de l'évaluation comme explication de ce découpage binaire sera moins manifeste à partir des épreuves de 2004.

11. La mise en place des évaluations des acquis en CE1 en 2008-2009 verra l'introduction d'un domaine « Étude de la langue » à côté de « Lire » et « Écrire ».

12. Il s'agissait peut-être là d'éviter les controverses politiques et médiatiques sur le niveau des élèves...

13. De même, la répartition dans notre corpus des items relatifs aux trois ensembles « Connaissance du code », « Compréhension » et « Production de texte » est la suivante : près de la moitié des items pour le premier, 35% pour le second et 15% pour le troisième.

expliqué, seule la lecture (avec le calcul) avait été en effet concernée par l'introduction d'une « nomenclature » hiérarchisée des compétences en CE2 au milieu des années 1990 (MEN, 1995, p. 2 & 9). Dès lors, la présentation des compétences, traitements et analyses possibles des résultats en lecture¹⁴ avait crû nettement par rapport à l'écriture. Ainsi le nombre de lignes consacré au chapitre « Savoir lire » est multiplié par 7 entre les livrets de présentation de 1994 et 1995 (MEN, 1995, p. 9) tandis que la taille du « Savoir écrire » reste comparable à ce qu'elle était pour chacun des domaines jusqu'ici, c'est-à-dire quelques lignes comme le montre l'intégralité du chapitre reproduit ci-après.

Extrait n° 1 (MEN, 1995, p. 9)

SAVOIR ÉCRIRE

Les connaissances dont nous disposons pour analyser les façons d'écrire de jeunes enfants ne nous permettent pas encore des analyses aussi fines qu'en lecture. Toutefois, là encore, nous proposons une moyenne des résultats obtenus dans toutes les épreuves mettant en jeu l'écriture. C'est un simple instrument de visualisation des résultats.

Or il est intéressant de relire le chapitre « Savoir lire » (qui précède en fait ce chapitre « Savoir écrire ») à la lumière de la première phrase de l'extrait n° 1 : on y trouve en effet peu d'éléments de connaissance (et aucune indication théorique) sur les façons de lire des « jeunes enfants », excepté deux ou trois lignes sur la « maîtrise du code graphique » et la lecture de mots connus et inconnus (*ibid.*). Il est vrai que l'introduction de ce chapitre « Savoir lire » invitait les enseignants à la prudence : « À ce point du cursus l'apprenti lecteur aura encore beaucoup à faire pour se servir des écrits de façon autonome. [...] Ce sont donc **les résultats des premiers apprentissages que l'on mesure ici plus que le savoir lire** » (*ibid.* ; nous soulignons). Invitation à la prudence dont il faut au passage souligner là aussi l'atténuation par une réécriture en 2002 (MEN, p. 16) : « À ce point du cursus [...] ce sont **les résultats des premiers apprentissages que l'on mesure ici.** »

Enfin, on voit mal en quoi le second extrait ci-dessous, toujours tiré du chapitre « Savoir lire » du même livret de 1995 et traitant des « analyses fines », ne pourrait concerner l'écriture, étant donné la généralité du propos :

Extrait n° 2 (MEN, 1995, p. 9)

Chacun des items de l'évaluation CE2 impliquant (*sic*) un acte de lecture (y compris celle des consignes) est porteur d'informations sur le savoir lire de l'élève. Il peut être analysé isolément puisqu'il met en lumière la façon dont l'enfant s'est servi de ce qu'il sait et de ce qu'il sait faire pour réagir à la situation qui lui est proposée. Cette première approche (qui permet des analyses fines et la mise en œuvre de nouveaux apprentissages adaptés) a l'inconvénient de conduire à une vision éclatée des acquis à l'issue du cycle 2.

La fin de ce paragraphe nous semble d'ailleurs à discuter. On y lit en effet que si l'approche analytique rendue possible pour la lecture génère une évaluation fine,

14. Le protocole de 1990 (p. 7) voit même un développement supplémentaire pour le chapitre du « Savoir écrire », expliquant que « la présentation du cahier des élèves tend à montrer que tout converge vers la production d'écrits ».

elle n'en constitue pas moins un « inconvénient » puisqu'elle produit un éclatement des acquis en jeu. Dans cette logique, la suite du chapitre s'attache à proposer des modalités de regroupement d'épreuves pour « permettre aux enseignants de disposer d'une appréhension plus structurée des résultats de leurs élèves » (*ibid.*). La première phrase du chapitre « Savoir écrire » (cf. extrait n° 1) est donc assez déconcertante : elle présentait défavorablement le développement réduit du « Savoir écrire » par l'impossibilité d'effectuer des analyses aussi fines qu'en lecture, alors qu'une large part du chapitre consacré à l'évaluation de la lecture porte sur les moyens de pallier les problèmes générés par ces analyses fines...

Mais le traitement du chapitre « Savoir écrire » illustre plus encore le mécanisme discursif déjà décrit qui tend à requalifier des éléments contraignants en opportunités. Précisons que les quelques lignes de ce chapitre (cf. extrait n° 1) ont été tirées du protocole de 1995 mais qu'elles se retrouvent intégralement jusque dans les épreuves de 2000. La seule variation observée, mais non moins savoureuse, concerne la remarque suivante, que l'on trouve à la conclusion de ce chapitre dans le protocole d'évaluation de 1996 uniquement (p. 11) : « Moins encore que pour la lecture, il faudrait le prendre [le simple instrument de visualisation des résultats] pour un constat précis et rigoureux. » Or, sans tenir compte de cette dernière phrase, voici ce que devient le chapitre « Savoir écrire » du côté des publications annuelles de résultats : « Les connaissances dont nous disposons pour analyser les façons d'écrire de jeunes enfants **nous permettent, là encore**, de proposer une moyenne des résultats obtenus dans toutes les épreuves mettant en jeu l'écriture. C'est un **instrument** de visualisation des résultats. » (DÉP 1996, p. 20). On observe donc d'une part que le niveau des connaissances qui apparaît comme un élément défavorable dans les protocoles (cf. extrait n° 1) devient un facteur avantageux dans les publications annuelles et, d'autre part, que la simplicité qui affecte dans les protocoles, le « simple instrument de visualisation des résultats » (MEN, 1995, p. 9, extrait n° 1) disparaît des résultats annuels. Ces quelques lignes des publications annuelles seront réécrites conformément à la version des protocoles à partir de 2000 (DPD, p. 24). Et, à l'instar de ce que nous avons observé concernant les changements plus ou moins importants apportés à l'épreuve, cette rectification et le problème qu'elle réglait ne feront l'objet d'aucune explicitation...

4. LA SYNTHÈSE DE L'ÉVALUATION : STRUCTURER L'ÉPREUVE OU LA PRÉSENTER ?

Les discours sur les domaines évalués dont nous venons de rendre compte s'appuyaient en outre sur un certain nombre de tableaux, exposant notamment systématiquement pour chaque niveau et discipline, les objectifs ou compétences testés (voir les marges du tableau de français, CE2, 1989, en annexe 2). Plus qu'un simple dispositif récapitulatif, ces tableaux ont pu être présentés comme des moyens permettant aux enseignants d'analyser les erreurs de leurs élèves et de réajuster en conséquence leur enseignement (MEN, 2003, p. 13), voire comme un rouage essentiel de fabrication du dispositif (MEN, 1989, p. 3) : « Les tableaux [...], arrêtés par l'inspection générale de l'Éducation nationale, fixent le cadre des grands objectifs évalués. Ces matrices ont servi de base à l'élaboration d'épreuves. »

La linéarité de cette relation entre tableaux et protocole est en fait difficile à retrouver comme le montre le traitement des premières évolutions notables du protocole de CE2 en 1991. À la suite des instructions relatives à l'application des cycles à l'école primaire publiées cette année-là, « quelques compétences transversales [...] ont été introduites dans le champ de l'évaluation : savoir observer, savoir se situer dans l'espace, savoir se situer dans le temps » (MEN, 1991, p. 1). Si on ne trouve pas de trace à priori d'une évaluation du « savoir observer » en tant que telle dans la suite du document (la mention restera néanmoins jusqu'à la présentation des épreuves de 1993), les deux autres compétences sont effectivement constituées en un domaine « Espace et temps » à côté du « Savoir lire et Savoir écrire » (MEN, 1991, p. 6-7). Quelques exercices et items renseignant ce domaine sont donc intégrés aux épreuves. Pour autant, le tableau de synthèse des objectifs évalués demeure inchangé, comme le précise simplement le livret : « Ces objectifs [relatifs à « Espace et temps »] ne sont pas reportés sur le tableau » (*ibid.*, p. 7). Sa structure reste donc quasiment identique à celle présentée en annexe 2 de cet article, et ce jusqu'à 1994.

En 1994, le livret de présentation précise que le « cadre de référence des objectifs évalués [...] défini par le groupe national de pilotage, est très proche de 1993 » (p. 1). On constate cependant que les premières pages du livret de 1994 ne mentionnent plus de « compétences transversales » et qu'il n'y a plus de chapitre « Espace et temps » comme cela était encore le cas l'année précédente. Ces notions se retrouvent toutefois à la fin du chapitre titré « Évaluation du savoir lire et du savoir écrire », au détour d'une phrase signalant que « l'opération [...] reprendra l'évaluation de la structuration du temps et de l'espace, en l'articulant avec les compétences linguistiques mises en jeu (ces objectifs ne sont pas reportés sur le tableau) » (MEN, 1994, p. 7). Courte phrase dont on peut se demander si elle ne vise pas à réhabiliter une lacune en mettant en exergue des « compétences linguistiques » dont le lecteur ne saura rien de plus ; il ne saura pas plus en quoi il y a là un changement avec les années précédentes. Finalement, les objectifs relatifs à « Espace et temps », domaine qui n'existe plus en tant que tel depuis 1994 donc, se trouvent reportés dans le tableau de 1997 (p. 10) dans une case¹⁵ située à l'entrecroisement du « Savoir lire » et de « Lecture et production d'écrits ». Aucune précision ne sera donnée aux enseignants sur les raisons de ce choix, si ce n'est l'indication que « des exercices les évaluaient dans les protocoles antérieurs sans qu'ils soient présents dans la grille pour autant » (p. 7). La présentation de 1997 stipule pourtant bien (p. 7-8) que « la grille d'objectifs de l'an dernier a été légèrement modifié (*sic*) pour prendre en compte la mise en application des nouveaux programmes de CE1 à la rentrée 1996 ». Une liste (*ibid.*) récapitule en conséquence les suppressions, dédoublements, regroupements, réécritures

15. L'objectif est ainsi formulé : « Reconstituer la chronologie des événements, se représenter les relations spatiales et les lieux évoqués dans des textes de statuts variés (en s'appuyant, en particulier, sur les connaissances acquises dans les domaines des programmes de cycle 2 sous l'intitulé "Découverte du monde") ». (MEN, 1997, p. 10).

d'objectifs effectués entre les deux années et les reformulations des rubriques du tableau sans toutefois en expliciter le gain¹⁶.

L'intérêt des modifications effectuées entre 1999 et 2001 fut à l'inverse souligné dans les livrets (MEN, 2001, p. 2) : les enseignants de fin d'élémentaire étaient ainsi informés qu'une harmonisation (non détaillée) avait été entreprise avec l'évaluation à l'entrée de sixième afin de leur permettre de « recourir plus facilement à l'une et à l'autre ». Objectif qui supposait donc de parvenir à comprendre et manipuler chacun des deux tableaux... qui s'étaient désormais sur deux pages en raison de l'ajout de colonnes relatives aux numéros d'exercices et d'items (cf. la structure du tableau 2001 en annexe 3). L'intérêt affiché de ces informations supplémentaires était « de situer chaque exercice dans son contexte et de mieux appréhender l'ensemble des compétences évaluées » (*ibid.* p. 7). Travail qui semblait de plus en plus complexe puisque le livret 2001 adjoignait à ces tableaux de synthèse et aux tableaux récapitulatifs des exercices de chaque séquence (précisant la durée de l'exercice, les tâches et compétences évaluées), deux nouvelles formes de récapitulation : une liste de tous les exercices précisant la compétence, la tâche et les items correspondants (p. 11-13) et un tableau des modalités de passation des exercices (supports et consignes vu-e-s et/ou entendu-e-s par les élèves ; p. 14). Notons que ces nouveaux dispositifs textuels n'étaient proposés qu'en français sans que l'on sache pourquoi¹⁷. Au final, cette difficulté à mettre en relation les tableaux et l'ensemble de l'opération est peut-être liée non seulement à leur fonctionnalité hybride (présenter vs structurer l'épreuve) mais aussi au fait que leur logique de construction n'étant pas tout à fait convergente avec les autres formes de discours descriptif sur le dispositif, ils introduisent paradoxalement quelques incertitudes. Ce désajustement a pu d'ailleurs être amplifié par les ambiguïtés autour des catégories retenues pour leur élaboration comme nous allons l'aborder dans la partie suivante.

5. CONCEPTION, TRAITEMENT ET ANALYSE : DES SYSTÈMES DE CATÉGORIES COMPLEXES

Les protocoles et leur présentation comportaient donc différents dispositifs discursifs dont nous avons pu analyser jusqu'ici les tensions et flottements tant pour

16. Une des conséquences est alors sur le plan statistique d'intégrer les traitements des items « Espace et temps » à ceux de la lecture. Or, de 1993 à 1996, le traitement des scores relatifs au domaine « Espace et temps » était distinct de ceux relatifs à l'écriture et à la lecture, sans que cela soit explicité. Par exemple, les scores de « Espace et temps » en 1995 ne font pas l'objet d'une synthèse (DÉP, 1996, p. 12) et ne sont pas toujours intégrés dans le calcul du score global de français (*ibid.*, p. 21). Conséquemment, ce document présente différents pourcentages de réussite de la discipline sans que cela soit spécifié (par exemple, toujours dans le même document : 66,1% de score global en français à la p. 16, contre 65,7% à la p. 12).

17. Signalons une autre différence disciplinaire inexpliquée concernant les modalités de passation : de 1991 à 1994 compris, les consignes stipulent que pour les exercices de français, les enseignants ont une certaine latitude pour adapter au mieux le temps de passation à leurs élèves (MEN, 1994, p. 3). La spécification disciplinaire disparaît entre 1995 et 2001. À partir de 2002, c'est la latitude concernant la gestion du temps qui disparaît. Le temps alloué était calculé sur la base du temps requis par 80 ou 85% des élèves lors de la phase d'expérimentation (MEN, 2001, p. 3).

ce qui touche aux fonctions des évaluations, aux contenus visés, qu'à leur structuration. Mais nous allons voir qu'une autre difficulté majeure tient à la rareté des discours explicatifs concernant les catégories utilisées pour désigner les découpages de contenus disciplinaires (domaines, champs, objectifs, etc.) et leur évolution. La notion de « Compétences » qui constitue une catégorie de tous les tableaux depuis 1989 n'est ainsi jamais définie¹⁸. Par exemple encore, le protocole de 2001, que nous venons d'aborder, insiste (p. 7) sur la clarification produite par la recomposition du tableau supposée donner « une vision d'ensemble de l'évaluation » (cf. annexe 3). Or aucun élément ne vient cependant étayer cette affirmation ni expliquer la signification des rubriques employées ou le choix des nouvelles désignations, comme les « composantes des tâches évaluées¹⁹ », si ce n'est que ces dimensions vont de « la plus grande unité (le champ) à la plus petite unité significative (l'item) » (MEN, 2001, p. 7). Et aucun élément non plus ne permet de reconstruire avec cohérence les relations que ces catégories entretiennent avec les dénominations antérieures.

Ainsi, la notion de « Champ » est utilisée entre 1989 et 1996 dans la désignation de la marge supérieure en « Champs d'application » (cf. annexe 2) du type « Vocabulaire », « Compréhension et production des textes », etc. Puis elle réapparaît en 2001 pour dénommer les marges de gauche « Savoir lire » et « Savoir écrire » (cf. annexe 3) classées jusque-là plutôt parmi les objectifs et compétences (cf. annexe 2)... Puis la notion de « champ » disparaît en 2004 et réapparaît en 2005 (p. 12-13) pour désigner des catégories comme « Reconnaissance de mots » ou « Écriture et orthographe ». Difficile de reconstruire le sens de tels réaménagements descriptifs, d'autant plus que la première page des livrets de présentation comportait jusqu'à 2000 un chapitre dont le titre mentionnait également le mot « champ » (« Nature et champ de l'évaluation ») et qui décrivait globalement l'opération (sa construction et ses fonctions) sans pour autant aborder les contenus...

Et pour complexifier l'ensemble, il faut ajouter à ces catégorisations conceptuelles, des désignations de traitement des réponses qui ne convergeaient pas nécessairement avec les premières. Sans que les livrets de présentation des protocoles précisent cette distinction, le terme de « Champ » était par exemple également employé dans la méthodologie de ces évaluations pour désigner des regroupements d'items réalisés dans la saisie et le calcul des résultats. Ces regroupements d'items pouvaient être créés par un enseignant en utilisant localement le logiciel Casimir, y compris en croisant les disciplines, par exemple pour calculer les scores de ses élèves aux items concernant la maîtrise des repères spatiaux. Mais il existait aussi des regroupements d'items disciplinaires définis nationalement et dénommés « Champs standards » (par exemple : « Production de textes »).

18. Notons qu'entre 1994 et 1998, apparaissent des renvois vers la brochure du MEN, *Maîtrise de la langue à l'école élémentaire* (1992) pour des références sur les compétences langagières mises en œuvre dans l'écriture et la lecture.

19. La compétence « Comprendre et savoir appliquer les consignes courantes du travail scolaire » que nous avons déjà citée comporte par exemple la « composante de tâche évaluée » suivante : « Appliquer une consigne » (MEN, 2001, p. 9).

Or ces « Champs standards » ne correspondaient pas tout à fait aux catégories utilisées dans les tableaux ou dans les textes de présentation des épreuves, sans pour autant que ces différentes logiques de construction soient indiquées. Ainsi, nous avons vu que dès le lancement de l'opération en 1989 le texte de présentation de l'épreuve mettait en avant le « Savoir Lire » et le « Savoir écrire », catégories auxquelles le tableau de l'évaluation en question ajoutait : « Approche des textes », « Vocabulaire », « Grammaire », « Orthographe » (cf. annexe 2). De son côté, le dispositif de relevé des codes et de calcul des scores fourni aux enseignants était organisé selon les ensembles d'items suivants (dénommés alors « activités », MEN, 1989, p. 47) : « Lecture 1 (mécanismes) », « Écriture (production de texte) », « Lecture 2 », « Orthographe-Grammaire-Vocabulaire ».

La publication que la DÉP produira à partir de cette évaluation de rentrée 1989 désignera encore différemment ces regroupements : tout d'abord « Compréhension de textes », « Mécanismes », « Orthographe-Grammaire-Vocabulaire », « Production écrite » (DÉP, 1990, p. 13-14) et quelques pages après « Lecture (mécanismes de base) », « Production de texte », « Lecture », « Orthographe-Grammaire-Vocabulaire » (*ibid.*, p. 15-19). À partir de 1990, le dispositif de relevé des réponses et la publication des résultats seront organisés de manière identique en trois ensembles : « Compréhension », « Connaissance du code » et « Production de texte » (DÉP, 1991, p. 25). Ces ensembles constitueront les « champs standards » dont les dénominations seront dès lors globalement stables²⁰ et ne correspondront jamais totalement aux découpages des livrets de présentation des épreuves et des tableaux des compétences.

Et il n'est pas impossible de penser que la réintégration de la notion de « Champ » dans le tableau de 2001 que nous signalions au début de cette partie visait justement à réduire cette disjonction. En effet, le livret de 2001 indiquait que l'organisation du tableau avait été mise « en adéquation avec les champs standards effectués par le logiciel Casimir » (MEN, 2001, p. 2). Mais on peut se demander quel est l'effet de clarification produit dans la mesure où, comme le montre le tableau que nous avons construit en annexe 4, les désignations des champs relevées dans les trente premières pages de la publication nationale des résultats en français de cette évaluation 2001 diffèrent (DPD, 2002).

CONCLUSION

On le voit donc, l'analyse de la documentation destinée aux utilisateurs de cet « outil » évaluatif soulève nombre d'interrogations, dont la moindre n'est pas de savoir ce qui en advenait au final. Car après s'être vus formuler les visées générales, les contenus et la structure de l'épreuve, puis notifier les catégories de traitement à l'œuvre, les enseignants étaient logiquement invités, dans le respect des consignes, à la passation et au codage des réponses en vue de l'« exploitation des résultats »

20. Les désignations de ces trois catégories ont varié de façon mineure : la « Connaissance du code » a ainsi été désignée « Outils de la langue » entre 1999 et 2004 puis « Étude de la langue » à partir de 2008.

(MEN, 2003, p. 27-30). Aucun doute en effet sur l'intérêt de l'outil : « Les causes possibles des erreurs étant définies grâce aux codes et à l'analyse des productions, les enseignants sont alors en mesure de mettre en œuvre des remédiations appropriées. » (*ibid.*). Bien sûr, il fallait pour cela « passer des résultats statistiques à l'interprétation pédagogique », calculer et observer la distribution des différents codes de réponse, étudier plus avant les cahiers des élèves si besoin en menant des « entretiens individualisés », etc. (*ibid.*).

Mais là encore, le logiciel Casimir fournissait un appui considérable, ses fonctionnalités permettant notamment rien moins que « le pilotage de l'école, l'information aux familles ; le repérage des difficultés éventuelles rencontrées par les élèves ; [...] l'adaptation des projets d'enseignement à la réalité des besoins recensés » (*ibid.*). Ainsi donc les outils nécessaires étaient fournis aux enseignants, et c'est au titre de leur efficacité affirmée que cette évaluation semblait prescrite (MEN, 2001b). Efficacité potentielle qui reposait néanmoins dans les livrets, comme nous venons de le voir, sur un travail d'appropriation et d'exploitation qui restait largement du ressort des enseignants²¹ (traitements, analyses, projets d'équipe, etc.).

On peut donc se demander s'il n'y avait pas un paradoxe entre cette nécessité d'une implication enseignante forte et une documentation relative à cet « outil professionnel » (Brézillon & Champault, 2008, p. 84) assez peu explicative voire proche des modalités discursives des injonctions ou prescriptions de l'institution dont ce dispositif relevait (Daunay, 2000). Nous avons aussi souligné à plusieurs reprises dans nos analyses, les mécanismes discursifs de neutralisation des tensions générées par la conception de l'outil (par exemple entre généralisation et individualisation ou entre logiques comptables et formatives) et de requalification des dysfonctionnements possibles (par exemple la partition lecture-écriture ou le traitement du domaine « Espace et temps »).

On peut voir là les effets d'une incorporation de multiples visées plus ou moins congruentes (pédagogiques, disciplinaires, managériales, etc.) dans un dispositif aux enjeux politiques, sociaux, et professionnels importants (Pons, 2012). On peut aussi penser que, fût-ce au prix de quelques approximations, ce traitement en discours de l'évaluation cherchait à mettre en exergue la responsabilité professionnelle des enseignants dans la réussite de l'opération ou encore ambitionnait de favoriser la passation de l'évaluation en évitant de s'attarder sur les lacunes, les doutes ou les problèmes possibles. Mais on peut aussi penser que cette apparente assurance applicationniste était effectivement générée par l'idée selon laquelle l'outil en question n'était pas tributaire de certaines conceptions des enseignements et des disciplines ou marqué par des orientations méthodologiques²², et qu'il était tout à fait

21. Ce type d'opération a pu d'ailleurs être dit à « trois volets » : « évaluation des élèves/formation des enseignants/réponses aux élèves » (MEN, 1990, p. 1). Notons que le MEN a pu afficher une politique de soutien aux enseignants dans l'exploitation de ces évaluations par des formations, séminaires, publications, etc. (Hébrard, 2000). Une banque en ligne d'outils d'aide à l'évaluation diagnostique a également été mise à la disposition des enseignants.

22. Levasseur (1996, p. 106) n'écrivait-elle pas que dans ce dispositif « la participation de toutes les responsabilités engagées dans l'action éducative, d'une part, constitue une garantie de neutralité à l'égard de toute théorie, conception, expérience ou projet pédagogique, d'autre part, elle est une

possible pour un enseignant d'utiliser et d'intégrer aisément, telle quelle et sans trop d'explication sur ces choix, une évaluation générique ou externe, conçue en dehors de sa classe et de sa dynamique pédagogique. Bref, un abord de l'évaluation à rebours des travaux de recherches, principalement didactiques (Delcambre, 2013), qui ont montré de longue date en quoi l'évaluation constituait un rouage majeur du fonctionnement ordinaire de la classe, des enseignements et des disciplines... et qui permettent par conséquent d'envisager des pistes possibles d'explication à cette forme d'insaisissabilité des évaluations nationales.

RÉFÉRENCES

- BRÉZILLON, Gérard, CHAMPAULT, Françoise (2008), « Les évaluations diagnostiques en France : le diagnostic, une notion et une exploitation en classe qui évoluent depuis 1989 », dans Baillat, Gilles, de Ketele, Jean-Marie, Paquay, Léopold, Thélot, Claude (dir.) *Évaluer pour former*, Bruxelles, De Boeck, p. 83-98.
- CHARTIER, Anne-Marie, HÉBRARD, Jean (2000), *Discours sur la lecture (1880-2000)*, Paris, BPI-Centre Georges Pompidou, Fayard.
- DABÈNE, Michel (dir.) (1994), « L'évaluation de la lecture : approches didactiques et enjeux sociaux », *LIDIL* n° 10, Grenoble, ELLUG.
- DAUNAY, Bertrand (2000), « L'oral au rapport ! Ou comment l'Institution s'empare de l'oral », *Recherches* n° 33, *Oral*, Lille, ARDPF, p. 7-27.
- DAUNAY, Bertrand (2008), « Performances et apprentissages disciplinaires », *Les Cahiers Théodile* n° 8, Villeneuve d'Ascq, Université Lille 3, p. 7-23.
- DELCAMBRE, Isabelle (2013/2007), « Évaluation », dans Reuter, Yves (éd.), *Dictionnaire des concepts fondamentaux des didactiques*, Bruxelles, De Boeck, p. 101-105.
- DELCAMBRE, Isabelle, DENIZOT, Nathalie (1999), « Nadège ou les infortunes de l'écriture », *Recherches* n° 31, *Violences culturelles*, Lille, ARDPF, p. 115-142.
- DÉP (1988), « Lire-écrire-compter au sortir de l'école élémentaire », *Éducation et formations* n° 14, Paris, MEN.
- EMIN, Jean-Claude, LEVASSEUR, Jacqueline (2007), « L'évaluation dans le système éducatif en France », *Revue Suisse des Sciences de l'éducation* vol. 29 n° 1, *Mesure des performances et évaluation des systèmes éducatifs – présupposés et contextes*, p. 65-78.
- GRANGEAT, Michel (1996), « Tirer parti de l'évaluation d'entrée en sixième », *Revue française de pédagogie* n° 115, p. 89-98.
- HARLÉ-GIARD, Isabelle, JOSSO-PERROCHAUD, Anne (2013), « Les évaluations nationales en CE1 et CM2 en France : entre enjeux institutionnels et tensions professionnelles », *Spirale* n° 51, *Éducation et mondialisation*, Villeneuve d'Ascq, ARRED, p. 45-58.

démonstration que l'évaluation nationale n'est pas une fin en soi, ni un "hobby" de technocrates, mais que l'évaluation est au service de l'acte éducatif ? »

- HASSAN, Rouba (2005), « Évaluations nationales en CE2 et 6^e : évolutions, mises en perspectives, problèmes », dans Reuter, Yves (dir.), *Démarches pédagogiques et lutte contre l'échec scolaire. Rapport de Recherche, tome 1*, Villeneuve d'Ascq, Université Lille 3, p. 365-374.
- HÉBRARD, Jean (2000), « Définition des niveaux de base en lecture à la fin des cycles 2 et 3 », dans MEN-DESCO, *L'exploitation de l'évaluation nationale en CE2. La lecture. Actes du séminaire national des 9 et 10 octobre 2000*, Paris, MEN-DESCO, p. 7-19.
- LAHANIER-REUTER, Dominique, REUTER, Yves (2004), « Décrire une figure géométrique. Éléments d'interrogation à propos de l'écriture en mathématiques et en français », *Les Cahiers Théodile* n° 5, Villeneuve d'Ascq, Université Lille 3, p.79-92.
- LEVASSEUR, Jacqueline (1996), « L'évaluation nationale des acquis des élèves », *Revue internationale d'éducation de Sèvres* n° 11, *L'évaluation des élèves*, Sèvres, CIEP, p. 101-114.
- MEN (1981), « Évaluation de l'enseignement à l'école primaire, année 1979, CP », *Études et Documents* n° 80-3, DTEE/SEIS, Paris, MEN.
- MEN (1983a), « Évaluation pédagogique dans les écoles et les collèges, classes de CM2/6^e », *Éducation et formations* n° 3, SIGES, Vanves.
- MEN (1983b), « Évaluation de l'enseignement à l'école primaire, année 1981, CE2 », *Éducation et formations* n° 7, SIGES, Vanves.
- MEN (2001b), « Évaluations en CE2, en sixième et en seconde – année 2001-2002 », Circulaire n° 2001-143 du 18 juillet 2001, parue au *BOEN* n° 30 du 26 juillet 2001.
- MENOUAR, Élisabeth, BART, Daniel (2014), « Les évaluations nationales de CM2 : modes de passation et spécificités disciplinaires », *Recherches en éducation* n° 19, p. 184-194.
- MIGEON, Michel (1989), *La Réussite à l'école. Rapport au Ministre de l'éducation nationale*, Paris, CNDP.
- NORMAND, Romuald, GARNIER, Bruno, REMOND, Martine, DEROUET, Jean-Louis (2004), *L'émergence d'une culture de l'évaluation dans l'Éducation Nationale : comment les enseignants utilisent-ils les résultats des évaluations nationales ?* Lyon, Commissariat au Plan, DPD, INRP.
- PONS, Xavier (2012), « Quarante ans d'évaluation ministérielle des acquis des élèves en France : complexification et politisation », *Politiques sociales et familiales* n° 110, *Les outils de l'évaluation des politiques sociales et leurs usages*, p. 9-18.
- Recherches* (2003), « Évaluations et examens », n° 38, Lille, ARDPF.
- RÉMOND, Martine (1996), « Pour une pédagogie de la compréhension au cycle 3 », dans Observatoire national de la lecture, *Regards sur la lecture et ses apprentissages*, Paris, MEN, p. 158-172.
- ROUBAUD, Marie-Noëlle (2005), « Reconsidérer l'erreur », *Les cahiers pédagogiques* n° 438, *L'évaluation des élèves*, p. 31-32.
- Spirale* (1990), « Évaluation CE2-6^e », n° 4, Lille, APEDIEN.
- VOGLER, Jean (1996), *L'évaluation*, Paris, Hachette.

CORPUS

- DÉP (1990), « Évaluations CE2-6^e : résultats nationaux, septembre 1989 », *Éducation et formations hors-série*, Paris, Vanves, MEN, DÉP.
- DÉP (1991), « Évaluation des acquis des élèves », *Éducation et formations* n° 27, Paris, MEN.
- DÉP (1992), « Évaluations CE2-6^e : résultats nationaux, septembre 1991 », *Éducation et formations hors-série*, Paris, Vanves, MEN, DÉP.
- DÉP (1996), « Évaluations CE2-6^e : repères nationaux, septembre 1995 », *Les Dossiers éducation et formations* n° 65, Paris, Vanves, MEN, DÉP.
- DÉP (1997), « Évaluations CE2-6^e : repères nationaux, septembre 1996 », *Les Dossiers éducation et formations* n° 79, Paris, Vanves, MEN, DÉP.
- DÉP (2003), « Évaluations CE2-6^e-5^e : repères nationaux, septembre 2002 », *Les Dossiers éducation et formations* n° 141, Paris, Vanves, MEN, DÉP.
- DPD (1999), « Profil des élèves en début de CE2. Évaluation de septembre 1998 », *Note d'information*, n° 99-33, Paris, MEN.
- DPD (2000), *Évaluations à l'entrée au CE2 et en sixième : repères nationaux, septembre 1999*, Paris, MEN, DPD.
- DPD (2001), « Évaluations CE2-6^e : repères nationaux, septembre 2000 », *Les Dossiers éducation et formations* n° 124, Paris, Vanves, MEN, DPD.
- DPD (2002), « Évaluations CE2-6^e : repères nationaux, septembre 2001 », *Les Dossiers éducation et formations* n° 128, Paris, Vanves, MEN, DPD.
- MEN (1989), *Évaluation à l'entrée au CE2, mathématiques, français. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1990), *Évaluation à l'entrée au CE2, mathématiques, français. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1991), *Évaluation à l'entrée au CE2, mathématiques, français. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1992), *Évaluation à l'entrée au CE2, mathématiques, français. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1993), *Évaluation à l'entrée au CE2, mathématiques, français. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1994), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1995), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1996), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1997), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1998), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (1999), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DPD.
- MEN (2000), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DPD.

- MEN (2001), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DPD.
- MEN (2002), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DPD.
- MEN (2003), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (2004a), *Évaluation à l'entrée au CE2, français, mathématiques. Présentation. Document à l'attention du maitre*, Paris, MEN, DÉP.
- MEN (2004b), *Évaluation à l'entrée au CE2, français. Cahier de l'élève*, Paris, MEN, DÉP.
- MEN (2005), *Évaluation à l'entrée au CE2, français, mathématiques. Document à l'attention de l'enseignant*, Paris, MEN, DÉP.
- MEN (2006), *Évaluation à l'entrée au CE2, français, mathématiques. Document à l'attention de l'enseignant*, Paris, MEN, DÉP.
- MEN (2007), *Dispositif d'évaluation CE1. Livret de l'enseignant*, Paris, MEN, DGESCO.
- MEN (2009), *Évaluation nationale des acquis des élèves en CE1, français, mathématiques. Livret de l'enseignant*, Paris, MEN, DGESCO.
- MEN (2010), *Évaluation nationale des acquis des élèves en CE1, français, mathématiques. Livret de l'enseignant*, Paris, MEN, DGESCO.
- MEN (2011), *Évaluation nationale des acquis des élèves en CE1, français, mathématiques. Livret de l'enseignant*, Paris, MEN, DGESCO.
- MEN (2012), *Évaluation nationale des acquis des élèves en CE1, français, mathématiques. Livret de l'enseignant*, Paris, MEN, DGESCO.
- MEN (2013), *Évaluation nationale des acquis des élèves en CE1, français, mathématiques. Livret de l'enseignant*, Paris, MEN, DGESCO.

ANNEXES

Annexe 1 : Exercice 13, CE2, 2004, français

(Adaptation de l'exercice 13 de 1998, 14 de 2000)

Exercice 13													
Dictée.													
.....	<table border="1"><tr><td>1</td><td>2</td><td>8</td><td>9</td><td>0</td></tr><tr><td colspan="5" style="text-align: center;">63</td></tr></table>	1	2	8	9	0	63						
1	2	8	9	0									
63													
.....	<table border="1"><tr><td>1</td><td>2</td><td>8</td><td>9</td><td>0</td></tr><tr><td colspan="5" style="text-align: center;">64</td></tr></table>	1	2	8	9	0	64						
1	2	8	9	0									
64													
.....	<table border="1"><tr><td>1</td><td>2</td><td>7</td><td>8</td><td>9</td><td>0</td></tr><tr><td colspan="6" style="text-align: center;">65</td></tr></table>	1	2	7	8	9	0	65					
1	2	7	8	9	0								
65													
.....	<table border="1"><tr><td>1</td><td>9</td><td>0</td></tr><tr><td colspan="3" style="text-align: center;">66</td></tr></table>	1	9	0	66								
1	9	0											
66													

Fac-similé du cahier des élèves (MEN, 2004b, p. 20)

Extrait du livret de l'enseignant : consignes de passation (MEN, 2004a, p. 65-66)

Dites aux élèves : « Vous vous souvenez, tout à l'heure je vous ai montré une phrase, maintenant je vais vous la dicter. »

Dictez la phrase segment par segment.

Chaque segment ne sera dicté qu'une seule fois.

La ponctuation sera dictée.

Des roses jaunes / parfument le salon. /

Dites aux élèves : « Passez maintenant à l'exercice suivant. »

Extrait du livret de l'enseignant : consignes de codage (Ibid.)

Item 63 : Orthographier correctement un nom au pluriel (Des roses jaunes)

« roses » code 1

Orthographe grammaticale correcte, mais avec une faute d'usage code 2

Absence de marquage du pluriel..... code 8

Autres cas code 9

Mot non écrit code 0

Item 64 : Orthographier correctement un adjectif au pluriel (Des roses jaunes)

« jaunes » code 1

Orthographe grammaticale correcte, mais avec une faute d'usage..... code 2

Absence de marquage du pluriel..... code 8

Autres cas code 9

Mot non écrit code 0

Item 65 : Orthographier correctement un verbe conjugué à la 3^èe personne du pluriel (parfument)

- « **parfument** » code 1
- Orthographe grammaticale correcte, mais avec une faute d'usage**..... code 2
- « parfumes » code 7
- Absence de marquage du pluriel..... code 8
- Autres code 9
- Mot non écrit code 0

Item 66 : Orthographier correctement un nom simple au singulier (le salon)

- « **salon** » code 1
- Erreur d'orthographe code 9
- Mot code 0

Annexe 2 : Marges du tableau de l'évaluation CE2, français, 1989 (MEN, p. 6)

Champs d'application	Approche des textes	Vocabulaire	Grammaire	Orthographe
Objectifs de l'évaluation et compétences acquises				
SAVOIR LIRE <ul style="list-style-type: none"> • Connaissance du code (combinatoire, fonctionnement de la langue) • Découverte du sens 				
SAVOIR ÉCRIRE <ul style="list-style-type: none"> • Reproduction d'écrits • Transcription d'oral • Élaboration de messages 				

**Annexe 3 : Marges du tableau de l'évaluation CE2, français, 2001
(MEN, p. 9-10)**

CHAMPS	COMPÉTENCES	COMPOSANTES DES TÂCHES ÉVALUÉES	EXERCICES	ITEMS
SAVOIR LIRE : COMPRÉHENSION				
SAVOIR LIRE : LES OUTILS DE LA LANGUE				
SAVOIR ÉCRIRE : LES OUTILS DE LA LANGUE				
SAVOIR ÉCRIRE : PRODUIRE DES TEXTES				

Annexe 4 : Les systèmes de catégories impliquant le terme « champ » pour les résultats de l'évaluation 2001 en français (DPD, 2002)

Pages	p. 13	p. 14	p. 23	p. 27	p. 28-29	p. 30
Objet	Tableau des scores moyens nationaux	Description synthétique des scores moyens	Rappel des scores moyens	Liste des compétences évaluées	Tableau des compétences évaluées	Rappel des scores moyens
Mode de catégorisation	Selon « trois champs »	Dans « chaque champ »	« Par champ »	« Pour chacun des quatre champs »	Colonne « Champs »	
Dénomination des catégories	« Compréhension »	« Compréhension »	« Compréhension »	« Compréhension »	« Savoir lire : compréhension »	« Compréhension »
	« Outils de la langue pour lire, pour écrire »	« Outils de la langue » (dont deux « sous-champs » : « Outils de la langue pour lire », « Outils de la langue pour écrire »)	« Outils de la langue »	« Outils de la langue pour lire »	« Savoir lire : les outils de la langue »	« Outils de la langue pour lire, pour écrire »
	« Production de textes »	« Production d'écrits »	« Production de textes »	« Outils de la langue pour écrire »	« Savoir écrire : les outils de la langue »	« Production d'écrits »
				« Production d'écrits »	« Savoir écrire : Produire des textes »	