

*L'appréciation des compétences
des élèves et des jeunes en lecture
et en écriture et l'évolution
de ces compétences dans le temps*

Marie-Thérèse CÉARD

Inspectrice d'académie-inspectrice pédagogique régionale honoraire

Martine RÉMOND

Chercheuse associée à l'Institut national de la recherche pédagogique

Michelle VARIER

Inspectrice de l'éducation nationale honoraire

N° 11
Décembre 2003

**Rapport établi à la demande
du Haut Conseil de l'évaluation de l'école**

Directeur de la publication : Christian FORESTIER

Secrétariat général : 3-5, bd Pasteur 75015 – PARIS

Tel : 01 55 55 77 41

Mèl : hcee@education.gouv.fr

ISSN en cours

Conception et impression : DEP/Bureau de l'édition

Sommaire

Introduction	5
Chapitre 1- Les évaluations françaises	
Historique des évaluations françaises.....	9
Les évaluations diagnostiques	12
Les évaluations diagnostiques dans le premier degré	13
Les évaluations diagnostiques dans le second degré.....	31
Les évaluations-bilan	
Les évaluations-bilans dans le 1er degré.....	67
Les évaluations-bilans dans le 2nd degré.....	73
Les évaluations spécifiques et comparatives	77
Usages attendus et usages réels des évaluations nationales	82
Evaluation hors système éducatif : la journée d'appel de préparation à la défense (JAPD)	95
Chapitre 2 - Les évaluations internationales	97
Le programme PIRLS – Progress in International Reading Literacy Study	98
L'enquête Reading Literacy de l'IEA (1991).....	103
Le programme PISA.....	104
Peut-on évaluer à l'échelle de plusieurs pays sans partir d'un corpus commun traduit ?	109
Chapitre 3 - Les évaluations nationales de la lecture dans d'autres pays	
L'évaluation britannique	111
L'évaluation américaine NAEP	114
Comparaison des évaluations nationales anglaises et américaines avec les protocoles nationaux français CE2-6 ^{ème}	116
Chapitre 4 – Fiabilité des protocoles	117
Chapitre 5 – Conclusion générale	123
Chapitre 6 – Propositions	
Propositions méthodologiques applicables à tous les protocoles	127
Propositions structurelles sur les évaluations nationales françaises	130
Annexe 1 : Liste des entretiens	136
Annexe 2 : Liste des documents utilisés et références bibliographiques	137
Annexe 3 : Evaluation des tableaux de compétences dans les protocoles de 6e.....	142
Annexe 4 : Les protocoles d'entrée en 6e permettent-ils de suivre l'évolution des compétences des élèves ?	145

Introduction

Dans le cadre de la mission qui lui a été confiée par le ministre de la jeunesse, de l'éducation nationale et de la recherche, le Haut Conseil de l'évaluation de l'école a retenu dès sa création la question de l'évaluation des compétences des élèves en lecture et en écriture. Il a décidé de traiter cette question sous le libellé suivant : «L'appréciation des compétences en lecture et en écriture des élèves et des jeunes, et l'évolution dans le temps de ces compétences».

Compte tenu de l'ampleur du sujet, l'étude ne portera pas sur les analyses de corpus de productions écrites d'élèves effectuées par des chercheurs, ni sur les appréciations que portent les maîtres sur les travaux de leurs élèves. De même, les tests psychotechniques sur la lecture n'ont pas été retenus : en effet, on n'en connaît pas de recension complète et leur utilisation est trop spécialisée pour entrer dans le cadre de ce rapport.

L'étude portera donc sur l'ensemble des dispositifs d'évaluation nationaux pris en charge par la DEP (Direction de l'évaluation et de la prospective) ou avec son concours, et des dispositifs étrangers ou internationaux. L'évolution des compétences dans le temps sera examinée à partir de 1979, date des premières évaluations mises en place par le ministère de l'éducation nationale. La population observée sera principalement composée des élèves de l'école élémentaire et du collège jusqu'au moment où les élèves viennent de le quitter, c'est-à-dire en début de seconde, même si des dispositifs portent éventuellement sur des jeunes plus âgés (comme le test de la JAPD, journée d'appel pour la préparation à la défense).

Le libellé introduit la notion de compétence. Les outils élaborés par les dispositifs d'évaluation visent précisément à donner un contenu concret et mesurable à cette notion, en fonction du niveau évalué : à tel moment du temps, qu'est-ce en effet que savoir lire et savoir écrire pour un jeune de seize ans, pour un élève de cours préparatoire ? Apprécier les compétences en lecture et en écriture suppose donc à la fois d'analyser les dispositifs construits pour les mesurer, et de dire ce qu'apprennent sur les compétences des élèves les résultats obtenus par ces mêmes dispositifs. Cette question, dont le danger de l'illettrisme, est un sujet de débat récurrent dans l'opinion publique mais aussi dans l'Education nationale. Ainsi, les instructions officielles de 1938 déploraient déjà que la lecture courante "ne fût pas complètement acquise à 10 ans par la moyenne des élèves" et qu'en écriture l'échec "restât trop fréquent."

Par ailleurs, la volonté d'apprécier les compétences des élèves en français et en mathématiques, deux disciplines de base dans la formation, n'est pas gratuite, mais répond à des besoins donc à des finalités. C'est pourquoi l'usage qui est fait des informations obtenues est inséparable de l'étude de leur élaboration.

Le rapport s'attachera donc essentiellement à cerner autant que possible, en fonction des moyens mis en œuvre pour le mesurer, ce que savent et ne savent pas faire, en lecture et en écriture, les élèves français à différentes étapes de leur cursus scolaire, à voir si leurs acquis et leurs difficultés sont stables ou non dans le temps, à s'interroger sur la fiabilité qu'on peut accorder aux outils eux-mêmes et à explorer les usages qui en sont faits ou qu'on peut en attendre.

Méthodologie.

Cette étude a été réalisée à partir :

- d'entretiens avec des personnalités françaises ou étrangères (cf. annexe 1),
- de rencontres et échanges de documents avec des collègues d'organismes étrangers, NFER et QCA (Grande-Bretagne), ETS (Etats Unis),
- d'une analyse comparative des différents protocoles, et de l'ensemble des documents de présentation des résultats publiés ou à usage interne,
- des résultats d'analyses psychométriques sur les protocoles,
- des rapports d'expertise effectuées antérieurement,

- des rapports de l'inspection générale et des plans académiques de formation,
- des textes publiés par le ministère de l'éducation nationale,
- de divers travaux existant sur les évaluations.

Les différents types d'évaluation.

Evaluer, c'est porter un jugement de valeur et situer un objet évaluable sur une échelle de valeur déterminée. Ainsi en disant qu'on veut **évaluer le savoir-lire**, on évoque un objet dont la nature et les dimensions varient selon le contexte théorique ou les définitions invoquées. Il faut donc disposer d'outils valides.

L'évaluation des compétences des élèves en lecture / écriture se situe dans le champ des acquisitions des élèves (connaissances, attitudes, capacités psychomotrices, etc.). A l'intérieur de ce champ il faut distinguer une perspective axée sur les élèves, individuellement, et une autre axée sur la population (ou des sous-populations) en vue de tirer des conclusions sur le fonctionnement du système.

La perspective axée sur les élèves individuellement

Dans cette perspective, l'évaluation peut, ou non, avoir une fonction certificative ou encore une fonction diagnostique, à laquelle pourra s'articuler une évaluation formative.

L'évaluation certificative permet d'officialiser une reconnaissance des acquis de chaque élève au cours de sa scolarité et conduit souvent à la délivrance d'un diplôme, par exemple le baccalauréat. Un problème qui se pose ici, parmi d'autres, est celui de l'introduction d'une procédure qui permette de garantir l'équivalence - au moins approximative - d'un diplôme obtenu, quel que soit le lieu géographique de la passation de l'examen.

L'évaluation diagnostique n'a pas une fonction certificative. Etablir un diagnostic (en psychologie, au moins), c'est à partir des performances, des comportements, des faits, rechercher ce qui peut les expliquer. L'évaluation diagnostique se pratique au début et en cours de formation afin de déterminer les aides à apporter. Elle peut être conçue pour permettre aux enseignants de connaître les forces et les faiblesses de leurs élèves, et ainsi d'identifier leurs lacunes en vue d'y remédier. Il s'en suivra une adaptation des activités d'enseignement aux capacités des élèves pendant la classe ordinaire, à des fins de facilitation et de régulation des apprentissages. L'enseignant pourra alors grâce à **l'évaluation formative** qu'il mettra en place, estimer l'efficacité de ses méthodes à travers l'observation de la qualité des apprentissages réalisés. Le diagnostic peut aussi conduire au choix d'aides très spécifiques apportées par différents réseaux (rééducation individuelle, par exemple).

La perspective axée sur la population

Dans cette perspective, on s'intéresse au fonctionnement du système, et dans ce cadre on cherche à établir un bilan par rapport aux objectifs terminaux d'un cycle d'étude ou d'un niveau dans le cursus. Les observations recueillies devraient alors permettre d'évaluer le rendement du système, notamment par la comparaison avec le niveau atteint antérieurement dans le cursus, et de définir l'importance relative des différentes variables d'entrée (évoquées plus haut) qui contribuent à la formation scolaire. Les comparaisons temporelles et entre systèmes éducatifs s'intègrent également dans ce cadre. Ces différentes approches paraissent indispensables à un pilotage efficace du système.

Les évaluations sommatives, aussi appelées « évaluations de **bilan** » des acquis scolaires, ont une fonction d'évaluation du rendement du système éducatif afin de tirer des conclusions sur son fonctionnement structurel ou pédagogique et de le piloter. Ces bilans **axés sur la population** visent à établir l'état des savoirs, des savoir-faire et des compétences des élèves à un moment de leur scolarité, par l'évaluation de performances. Ils sont destinés à nourrir la réflexion des politiques pour savoir si le système éducatif atteint, ou non, les objectifs qui lui sont assignés. Les évaluations de bilan sont généralement prises en charge par des organismes et portent sur des échantillons représentatifs.

Nous venons de définir les évaluations par leurs **fonctions**, en restreignant cependant le propos au champ de ce rapport. Les pratiques évaluatives des enseignants et les situations d'examen ou de concours en sont exclues.

La fonction de l'évaluation

Elle va fixer des contraintes auxquelles devra satisfaire l'élaboration d'un dispositif d'évaluation ; selon que celui-ci est prévu pour remplir une fonction diagnostique ou une fonction de bilan, ces contraintes seront largement différentes. Il y a eu lieu de **bien distinguer les fonctions** de l'évaluation des **techniques** d'évaluation. Les types d'instruments existant peuvent être classés en : empirique, diagnostiques et cognitifs.

Les instruments empiriques reposent sur la tâche en elle – même : « montrer qu'on est capable de lire, d'utiliser les informations présentes dans un texte » (Vrignaud, 2001). La plupart du temps, ces instruments se présentent sous la forme d'un texte à propos duquel sont posées des questions. Rentrant dans cette catégorie, le matériel proposé dans les évaluations de rentrée de la DEP, celui de la JAPD, le test de « La lecture silencieuse » de l'INETOP (test étalonné, élaboré dans les années 70 et révisé plusieurs fois, Aubret et al., 1981 ; Aubret & Blanchard, 1991), etc.

Les tests diagnostiques

Les évaluations de rentrée, appelées diagnostiques par leurs concepteurs, servent à détecter des lacunes des élèves, mais dans certains cas, il faudra pouvoir interpréter pourquoi un enfant a atteint de faibles performances et continuer à explorer ses difficultés grâce à d'autres outils. Il s'agit dans ce cas de le mettre devant une épreuve révélatrice des dysfonctionnements, épreuve relevant des **tests diagnostiques** dans la taxonomie des psychologues. Ces tests visent à dépister les difficultés d'apprentissage rencontrées par certains élèves : la dyslexie, par exemple. Parmi les tests connus, mais anciens, *l'Alouette* s'inscrit dans cette optique (Lefavrais, 1963). L'examineur demande à l'élève de lire à haute voix un texte renfermant des difficultés grapho – phonétique particulières et des mots dont le sens n'est probablement pas connu des élèves (« *les brignolles perfides* ») ou des mots quasi – homophones placés à une faible distance l'un de l'autre (« *Ami* », « *Annie* »). Il s'agit ici de montrer sa capacité de décodage grapho – phonétique, la compréhension du texte n'étant pas considérée.

Les tests cognitifs

Plus récemment, la batterie d'Aubret et Blanchard (années 1990) a été élaborée en se référant à un modèle **cognitif** de l'activité de lecture. Elle comporte des épreuves correspondant aux différents niveaux de traitements envisagés par son référent théorique (modèle de Craig) : grapho – phonétique, sémantique, syntaxique. Cette batterie peut être appliquée à partir de la classe de sixième.

Une batterie d'épreuves de compréhension de l'écrit pour le cycle 3 a pour point de départ une épreuve générale de compréhension qui permet de repérer les enfants présentant des déficits de compréhension. Dans un second temps, une série d'épreuves concernant des opérations particulières impliquées dans la compréhension décèle la ou les sources des difficultés, à partir de là, il est possible d'envisager les aides les plus appropriées pour l'élève. Quatre séries d'épreuves applicables au long de l'année ont ainsi été conçues et validées pour des élèves de CM1 (Rémond, en cours). Ce travail s'appuie sur les modèles de fonctionnement du lecteur élaborés dans le cadre de la psychologie cognitive et sur les modèles de la psychométrie pour valider les épreuves.

Les catégories auxquelles nous essayons de renvoyer les épreuves ne reflètent pas bien la complexité de ces outils. Certains tests empiriques reposent en partie sur des modèles de l'activité de lecture (par exemple, les épreuves pour des évaluations internationales), les tests cognitifs ont souvent des visées diagnostiques. Les classements évoqués ici (et repris de la littérature du domaine) révèlent la diversité existante dans les approches d'évaluation de la lecture. Il est parfois difficile de faire la distinction entre la fonction de l'évaluation et le type d'outils utilisé.

Nous distinguerons les évaluations en lecture – écriture totalement conçues, appliquées et analysées par la DEP, des évaluations internationales prises en charge par la DEP. Les évaluations correspondant au premier cas seront appelées dans ce document Evaluations françaises.

Chapitre 1

Les évaluations françaises

Historique des évaluations françaises

I. Les évaluations avant 1989.

NI= Note d'Information, DT= Document de Travail, DEF= Dossier Education et Formations (publications de la DEP).

Dans la perspective de l'évaluation permanente du système éducatif, un dispositif d'évaluation pédagogique a été mis en place en 1979 dans les écoles, puis en 1980 dans les collèges et en 1984 dans les lycées. Ce dispositif est conduit à la demande de la Direction des Ecoles et celle des Lycées et Collèges par la Direction de l'Evaluation et de Prospective (le Service de la Prévision, des Statistiques et de l'Evaluation devient la Direction de l'Evaluation et de la Prospective le 10.02.1987).

Dès le départ, ce dispositif s'est donné pour tâche d'évaluer les compétences des élèves aux niveaux clefs du cursus scolaire. Il devait, en mesurant l'efficacité de l'enseignement, permettre de "connaître les résultats du fonctionnement pédagogique du système éducatif au regard des objectifs qui lui sont assignés" (*Evaluation pédagogique dans les collèges*, juin 1982, page de présentation), savoir à quel degré sont atteints les objectifs assignés au système éducatif par les programmes officiels et par les réformes ministérielles. Ces évaluations se sont faites sur des échantillons représentatifs d'élèves appartenant à plusieurs académies. Très souvent, les enseignants ont donné leur avis sur les compétences évaluées et leurs pronostics sur les scores des élèves.

A - Autres évaluations du premier degré

Le dispositif de 1979 porte sur un panel important suivi du cours préparatoire à la fin du cours moyen :

En 1987, une enquête complète les données du dispositif précédent ; son but est de vérifier dans quelle mesure les élèves ont acquis les savoirs et savoir-faire correspondant aux exigences de la 6^e

En 1989, afin d'évaluer les effets de la mise en place des bibliothèques d'écoles sur les pratiques de lecture, une évaluation a été mise en place simultanément du CE1 au CM2. Des épreuves identiques ont été données d'un niveau à l'autre, permettant ainsi d'observer l'évolution des compétences réalisées en une année par les élèves.

Jusqu'en 1989, les évaluations sont des évaluations-bilans effectuées sur échantillon ; et qui ont pour fonction d'accompagner et d'évaluer les réformes .

Ces dispositifs sont présentés plus précisément au chapitre 4, portant sur les évaluations-bilans.

B - Evaluations du second degré

1. Pour "approfondir la connaissance du système éducatif aux principales étapes" du collège, un dispositif est mis en place, de 1980 à 1984. Il comprend trois opérations situées à l'entrée et à la sortie du cycle d'observation d'alors (sixième-cinquième), et à la fin du collège.

2. Un autre dispositif, de 1987 à 1988, vise à suivre "l'évolution des apprentissages au cours du cycle d'observation à partir des acquis et des difficultés des élèves à la fin du CM2 et deux ans plus tard" (DEF n°6 de juillet 1990 sur l'évaluation de 5^e 1988).

Les protocoles comportaient des parties communes au CM2 et à la cinquième.

3. De 1989 à 1992, un dispositif évalue l'effet de la rénovation des collèges, lancée en 1983 à la suite du rapport Legrand (*Pour un collège démocratique* 1982). Ce rapport demandait un assouplissement des structures internes du collège. C'est pourquoi ce dispositif porte également sur les cycles aménagés en collège (6^e-5^e en trois ans au lieu de deux).

4. En mai-juin 1988, une évaluation est conduite en fin troisième technologique. Il s'agit de la fin d'un cycle de deux ans (quatrième et troisième technologiques), existant déjà en lycée professionnel mais ouvert en collège en 1985.

En mai-juin 1995, une autre évaluation est conduite en fin de troisième, mais à la fois en fin de troisième générale et en fin de troisième technologique, que les troisièmes technologiques soient en collège ou en lycée professionnel (DEF n°86 de mai 1997). Toujours pour suivre l'évolution du système éducatif, les exercices sont partiellement repris de protocoles de troisième précédents (1984 et 1990).

5. En mai 1986, une évaluation est conduite en fin de seconde, pour expérimenter et asseoir les épreuves définitives des évaluations diagnostiques.

II. Les évaluations depuis 1989

A partir de 1989, des évaluations diagnostiques annuelles sont instaurées en même temps à l'entrée du CE2 et de la sixième. Elles ne portent plus sur des échantillons représentatifs, mais sont passées obligatoirement par tous les élèves de l'enseignement public et privé sous contrat.

Ces évaluations visent surtout à aider les professeurs à identifier, dès le début de l'année, les acquis et les lacunes de leurs élèves pour y adapter leur enseignement. Elles font en effet partie d'un dispositif qui prévoit, grâce à la formation des enseignants, une meilleure réponse aux difficultés des élèves. Elles sont poursuivies, de 1992 à 2001, par des évaluations diagnostiques et pronostiques annuelles à l'entrée en seconde (seconde générale et technologique, seconde professionnelle).

Destinées à faire chaque année un état des acquis et des difficultés des élèves, elles sont complétées, pour les besoins des professeurs en cours d'année, par des tomes d'exercices intitulés chacun *Banque d'outils d'aide à l'évaluation* qui, dès janvier 2001, ont en outre été disponibles sur Internet. L'objectif de cette banque est de mettre à la libre disposition des enseignants des "supports d'évaluation diagnostique, pour l'école, le collège et en seconde pour le lycée, dans une majorité de disciplines." Cette banque comprend actuellement environ 800 outils d'évaluation allant, par compétence, de la maternelle au lycée.

Enfin, à la rentrée 2002, une évaluation diagnostique est conduite pour la première fois à l'entrée en cinquième. Elle est reconduite à la rentrée 2003, mais à titre facultatif.

Entre temps, on peut citer certaines évaluations à fonctions particulières et réalisées sur des échantillons d'élèves.

En septembre 1997, a eu lieu une évaluation spécifique des élèves en difficulté de lecture à l'entrée en sixième. Il s'agissait d'une épreuve complémentaire à l'évaluation diagnostique d'entrée en sixième de 1997. Elle portait sur un échantillon d'élèves, venant de collèges situés en Zep et hors Zep.

Une étude comparative est réalisée en 1993 (NI 94-12).

Il s'agit de comparer les résultats des élèves à l'entrée en sixième en 1980 et en 1993, c'est-à-dire à treize ans de distance. Certains exercices sont repris, dans le protocole de 1993, du protocole de 1980 (les autres exercices viennent des groupes académiques).

Commentaires

La date de 1989 voit l'apparition d'un nouveau type d'évaluation.

Avant 1989, les évaluations ont pour fonction d'"améliorer la connaissance globale des acquis des élèves aux différents niveaux de la scolarité" (Lettre de Lionel Jospin du 28 juillet 1989). Elles sont donc principalement destinées aux responsables nationaux, régionaux et locaux et visent à éclairer le pilotage du système éducatif. Elles sont utilisées pour mesurer l'effet des réformes et des nouvelles structures. Elles assurent une fonction de bilan et peuvent se dérouler en observant une même population sur plusieurs années (suivi de cohorte), ponctuellement sur une population

particulière, sur la population d'un cursus (cycle) avec des épreuves échelonnées sur plusieurs niveaux.

Avec l'apparition des évaluations diagnostiques en 1989, la finalité est différente, ainsi que les premiers destinataires. Il s'agit plutôt de fournir aux enseignants un outil professionnel : un diagnostic sur l'état des acquis des élèves au regard des compétences censées acquises, un pronostic au regard des compétences à acquérir au cours de l'année. C'est pourquoi ces évaluations se situent au début de l'année scolaire, pour éclairer l'organisation de l'enseignement dans la classe et la mise en place éventuelle de structures de remédiation (groupes d'aide et de soutien par exemple). C'est aussi pourquoi elles sont obligatoires, passées par tous les élèves du niveau choisi (CE2, sixième, seconde), et répétées chaque année en s'adaptant à l'évolution des programmes. C'est enfin pourquoi le codage des réponses s'est affiné avec le temps pour permettre un dialogue fécond entre le professeur et chacun de ses élèves sur la source des erreurs constatées.

L'existence d'une banque d'outils à la disposition des enseignants s'inscrit dans la logique des évaluations diagnostiques. Elle a pour fonction d'aider les professeurs à évaluer, après le diagnostic initial, l'effet de leur enseignement sur les acquis de leurs élèves au cours de l'année scolaire et de réguler sans cesse la nature des apprentissages pour leur donner l'efficacité souhaitée. Elle propose également des outils d'évaluation sur des parties des programmes qui ne peuvent être pris en charge par les protocoles (comme l'oral, par exemple).

Co-existent toujours des évaluations remplissant différentes fonctions. En 2003, sont en cours, par exemple, une évaluation de cohorte suivie depuis 1997 (CP) jusqu'en classe de cinquième, des évaluations bilan fin de cycle (CM2 et 3ème), des évaluations diagnostiques.

Les évaluations diagnostiques

C'est en 1989 dans le cadre de la loi d'orientation que des évaluations diagnostiques sont créées. Dans une lettre du 28 juillet 1989 aux personnels de l'éducation nationale, Lionel Jospin les définit: « A la suite des recommandations formulées par le recteur Migeon, j'ai arrêté le principe d'une évaluation de tous les élèves à l'entrée du CE2 et de la classe de sixième.

Cette évaluation se distingue de toutes celles qui ont été réalisées jusqu'à présent dans l'éducation nationale. En effet, alors que ces dernières étaient destinées à améliorer la connaissance globale des acquis des élèves aux différents niveaux de la scolarité, celle qui est mise en œuvre aujourd'hui a pour but de vous aider à mieux connaître vos propres élèves et plus précisément à identifier, dès le début de l'année scolaire, d'éventuelles lacunes dans le champ des apprentissages fondamentaux : lecture, écriture, mathématiques. »

Modalités

La manière d'élaborer les protocoles et de faire passer les tests a été conçue et mise en place dès les premiers dispositifs d'évaluation. Elle s'est affinée et développée ensuite pour trouver son aboutissement dans les évaluations diagnostiques réalisées depuis 1989 jusqu'à nos jours.

Des groupes nationaux, disciplinaires ou pluridisciplinaires, élaborent, à partir des programmes, les instruments d'évaluation (épreuves, questionnaires aux enseignants, consignes, systèmes de codage,...) en relation avec des groupes académiques. Avant leur mise au point définitive, les épreuves sont expérimentées auprès d'élèves de diverses académies. Le protocole d'évaluation définitif consiste en un cahier, qui réunit l'ensemble des exercices avec le codage défini pour chacun d'entre eux.

Pour le déroulement de l'évaluation dans les établissements, les élèves disposent d'un cahier d'exercices dans chacune des disciplines évaluées, les enseignants d'un cahier de consignes et de codage des épreuves propre à chaque discipline et d'un cahier de relevé des réponses. Les réponses sont codées : les résultats du décodage effectué par les professeurs sont traités par des logiciels (Casimir pour la sixième et Evarem pour la seconde) qui peuvent regrouper ou croiser les données pour faire apparaître des informations utiles, comme des types d'erreurs, des domaines de compétences, des profils d'élèves. Pour chaque évaluation, les résultats d'un échantillon national représentatif d'élèves sont prélevés et disponibles sur le site du ministère sur Internet à la fin de la même année ; ils sont en outre présentés et analysés, puis édités par la DEP dans des dossiers (*Dossiers Education et formations*) et résumés dans des notes d'informations. C'est pourquoi les élèves constituant ces échantillons reçoivent, avec le cahier d'exercices, un questionnaire sur la vie scolaire ou une fiche de renseignements, et leurs professeurs, si nécessaire, des questionnaires par exemple pour recueillir leurs avis.

La rigueur dans la passation des épreuves est assurée par les consignes indiquées dans le cahier du professeur ; et la rigueur de la prise d'information que constitue chaque test est assurée par le codage des réponses, une composante essentielle de l'évaluation. Ce codage n'est pas une note, une correction analogue à celle du professeur en classe : chaque code détermine, en fonction des objectifs évalués dans les exercices, les différentes modalités des réponses possibles : réponse exacte et attendue, non attendue mais possible, exacte mais partielle, incomplète mais sans erreurs, erronée, absente, etc. C'est le professeur qui code les réponses des élèves testés dans son établissement en fonction du codage défini pour chaque exercice dans le protocole : il sera ainsi mieux à même d'explorer, avec ses élèves, la source des erreurs indiquées par ce codage.

Ces épreuves ne sont pas des examens ; et la DEP rappelle sans cesse que les informations recueillies dans le cadre de ces évaluations ne sont ni des modèles pour l'apprentissage ni des normes. Elles n'ont de sens que rapportées au contexte précis de l'évaluation.

Les évaluations diagnostiques dans le premier degré (CE2)

Le choix des compétences et les instructions officielles

Evaluer en début de CE2, c'est évaluer le cycle des apprentissages fondamentaux.

Le choix dépend de la conception qu'on a de ces apprentissages. Il faut donc examiner si ces conceptions ont évolué à travers la lecture des textes officiels de ces 25 dernières années.

Contrairement à ce qu'on pourrait attendre, vu les controverses violentes qui ont animé le débat sur les méthodes d'apprentissage de la lecture et sur les très nombreux travaux des chercheurs et didacticiens, les définitions successives ne sont pas réellement opposées : elles soulignent, toutes, la nécessité d'allier le souci de la compréhension aux techniques de décodage, mais selon le moment, les textes mettent plutôt l'accent sur tel ou tel aspect.

- En 1972, la circulaire du 4-12 rappelle que "des résultats également bons peuvent être obtenus par des démarches différentes" mais on cite Alain: le maître "*éveille l'esprit et délivre d'épeler*"

La lecture silencieuse est encouragée et on insiste sur l'importance de la compréhension: "*l'apprentissage de la lecture serait aussi inutile que pénible si l'enfant ne comprenait pas ce qu'il lit*"

- En 1977 l'arrêté du 18-3 définit ainsi le savoir lire au cycle préparatoire:

- "lire à haute voix, sans faute de lecture, en articulant convenablement sans syllaber, en respectant la ponctuation et les groupes de souffle, un texte (imprimé ou manuscrit) de 5 à 8 lignes dont le contenu correspond à l'expérience ou aux intérêts et possibilités mentales de l'enfant, et la facture (vocabulaire, syntaxe) à ses compétences linguistiques.

- lire silencieusement un texte (de même nature mais plus bref que ci-dessus) et faire apparaître par ses comportements qu'on l'a compris: exécuter les consignes.....

- raconter ou résumer un texte lu; répondre à des questions ou en poser à son sujet."

- En 1980 (arrêté du 18-7) , une définition est à nouveau donnée pour le cycle moyen:

- "lire à haute voix un texte de 15 lignes,

- adapter les modalités de lecture à la situation et au but recherché",

- "c'est aussi aimer lire".

- Pour les programmes de 1985 "lire c'est comprendre "; il est précisé que " la maîtrise de la combinatoire est nécessaire mais elle n'a de valeur qu'associée à la compréhension de l'idée exprimée, à la perception de la phrase, à l'intelligence du sens des mots" .

Dans les documents d'accompagnement, "compléments aux programmes" de 1985, on lit

"Dans l'acte de lire, il y a à la fois saisie globale et activité analytique en vue d'une synthèse ultérieure." Mais ,dans ce texte pourtant détaillé et précis, rien n'est dit sur les correspondances grapho-phonétiques .

- En 1990, suite à la loi d'orientation, les programmes sont définis par cycle et déclinés en compétences; lire/ écrire font partie d'un ensemble plus large: la maîtrise de la langue.

"A l'issue du cycle 2 l'élève peut:

- donner après lecture des renseignements ponctuels sur le texte,

- résumer sommairement ,

- identifier les personnages d'un récit et les retrouver quels que soient les procédés utilisés pour les désigner,

- repérer les indices morpho-syntaxiques,

- lire oralement en articulant correctement et en comprenant le sens de ce qui est lu

- déchiffrer des mots inconnus,

- utiliser une bibliothèque, repérer et identifier les ouvrages d'une B.C.D."

- En 1992, paraît la brochure, préfacée par le ministre, *Maîtrise de la Langue* ; elle n'a pas de statut précis mais elle constitue une référence pour les formateurs, les conseillers pédagogiques et les I.E.N.

Ce texte s'appuie sur les premiers constats des évaluations de 1989 pour signaler que " 80% des élèves de CE2 comprennent "globalement un texte mais que "66% seulement sont susceptibles d'une compréhension fine."

Cet ouvrage développe une conception de la lecture reposant sur les recherches et précise les processus d'apprentissage de la lecture: par exemple au cycle 2, lire c'est: "parvenir à une

reconnaissance automatique des mots par voie directe et indirecte"; dans "apprendre à comprendre", on distingue compréhension littérale et compréhension fine."

On peut donc dire que jusqu'en 1992, même si aucun texte n'oublie les processus d'analyse et de synthèse dans l'apprentissage de la lecture, ce qui est mis en avant, c'est la compréhension et la structure de la langue écrite.

Il faut attendre les dernières I.O. pour voir clairement indiquées la nécessité du travail grapho-phonétique et définies les deux voies qui sont nécessaires pour pouvoir automatiser les processus de lecture: le déchiffrement (les lettres sont assemblées pour constituer des syllabes) et l'identification directe qui repose sur la reconnaissance de l'image orthographique du mot.

L'apprentissage de la compréhension est aussi analysé, et particulièrement la compréhension des textes.

La conception de l'apprentissage qui prévaut aujourd'hui et qui est exposée de façon très détaillée dans les derniers programmes est résumée dans le rapport de l'ONL *Apprendre à lire " Les données issues de la recherche scientifique sur l'apprentissage de la lecture ont été rassemblées et leurs conclusions font ressortir quelques points essentiels pour l'apprentissage de la lecture au cycle 2: découverte guidée du principe alphabétique de la langue, automatisation de l'identification des mots, acquisitions de certaines régularités morpho-syntaxiques, accès à la compréhension des textes.*

II Evolution des choix et de l'organisation des protocoles

Tableau des compétences de 1989

Objectifs de l'évaluation et compétences acquises	Approche des textes	Vocabulaire	Grammaire	Orthographe
SAVOIR LIRE <ul style="list-style-type: none"> • Connaissance du code • Découverte du sens 	1. Distinguer des textes par leur forme extérieure usuelle : textes poétiques, narratifs, injonctifs, informatifs, énoncés divers.	6. distinguer des familles de mots	10. distinguer les marques du singulier/pluriel, du masculin/féminin, de conjugaison, de l'accord de verbes, de noms, d'adjectifs.	14. écrire les relations graphèmes phonèmes
	2. Comprendre une phrase, un paragraphe, un mot, un contexte, un texte.	7. Comprendre un mot dans plusieurs acceptions.	11. Repérer des indices morpho-syntaxiques simples.	15. Distinguer des homophones, des homonymes usuels.
SAVOIR ECRIRE <ul style="list-style-type: none"> • Reproduction d'écrit • Transcription d'oral • Elaboration de messages 	3. Copier un mot une phrase, un paragraphe.	8. Ecrire des mots usuels	12. appliquer des règles simples (cf. 10)	16. Orthographier des mots, des phrases dictés.
	4. Elaborer un court message adapté à une situation (types de textes). 5. Respecter les exigences de présentation	9. Choisir un lexique adapté	13. Appliquer des règles simples	17. Ecrire autrement.

Tableau des compétences de 2002

Champs	Capacités	Compétences	Composantes des tâches évaluées
Savoir lire	Comprendre un texte	Comprendre et appliquer les consignes courantes de la vie scolaire	<ul style="list-style-type: none"> • Appliquer une consigne écrite simple • Appliquer une consigne écrite complexe
		Distinguer des textes différents (récits, documents, consignes, poésies, etc.) en utilisant des indices extérieurs aux textes.	<ul style="list-style-type: none"> • Mettre en correspondance un ouvrage et sa couverture.
		Reconstituer la chronologie des événements dans des textes de statuts variés.	<ul style="list-style-type: none"> • Repérer des informations d'ordre chronologique à partir d'un récit réaliste • Déduire des informations d'ordre chronologique à partir d'un récit réaliste
		se représenter les relations spatiales et les lieux évoqués dans des textes de statuts variés	<ul style="list-style-type: none"> • Se repérer sur une image • Adopter un point de vue relatif dans l'espace • Tracer un déplacement à partir de données orales
		comprendre un texte et monter qu'on l'a compris donner après lecture des informations ponctuelles contenues explicitement ou implicitement dans le texte dans un récit, identifier les personnages et les retrouver, quels que soient les procédés utilisés construire la signification générale d'un texte simple en prenant en compte ses énoncés successifs	<ul style="list-style-type: none"> • Donner après lecture des informations ponctuelles sur un texte (récit réaliste) • Identifier des personnages par déduction • Sélectionner la phrase qui résume le mieux le récit • Donner après lecture des informations ponctuelles sur un document mêlant texte et image (affiche) • Donner des informations ponctuelles sur un texte (conte) • Choisir le titre (conte)
Outils de la langue	Maîtriser les outils de la langue	reconnaître les mots à l'écrit et à l'oral <ul style="list-style-type: none"> • savoir reconnaître des mots écrits du vocabulaire courant; • savoir déchiffrer des mots inconnus. 	<ul style="list-style-type: none"> • Eliminer le mot écrit sous une image s'il ne convient pas (image orthographique)
		repérer les usages typographiques courants et se situer dans l'espace graphique d'une page imprimée.	<ul style="list-style-type: none"> • Connaître l'organisation d'une page imprimée • Connaître les usages typographiques
		Pour mieux comprendre les mots d'un texte <ul style="list-style-type: none"> • distinguer selon le contexte le sens particulier d'un mot; • reconnaître les familles de mots • se servir d'un dictionnaire adapté. 	<ul style="list-style-type: none"> • Distinguer selon le contexte le sens particulier d'un mot

Outils de la langue		Pour comprendre un texte, identifier <ul style="list-style-type: none"> • les accords du verbe avec le sujet, de l'adjectif avec le nom, du nom avec le déterminant mettant en jeu le genre et le nombre; • les temps et les marques temporelles qui organisent le texte; • les divers indicateurs de rappel (pronoms et autres procédés) assurant la cohésion; • les principaux mots de liaison (connecteurs). 	<ul style="list-style-type: none"> • Identifier les accords du verbe avec le sujet, de l'adjectif avec le nom, mettant en jeu le genre et le nombre. • Identifier les temps et les marques temporelles qui organisent un texte.
		Copier un mot, une phrase, un texte en respectant les exigences de présentation et en écrivant lisiblement	<ul style="list-style-type: none"> • Copier un texte en vers
		Ecrire sous la dictée des syllabes, des mots courants, de petites phrases ou des textes courts. Transformer un texte en appliquant des règles simples.	<ul style="list-style-type: none"> • Ecrire une phrase sous la dictée • Ecrire des mots outils sous la dictée • Transformer un texte en passant du féminin au masculin
Savoir	Produire un texte	Savoir produire un écrit bref dans une situation de communication déterminée.	<ul style="list-style-type: none"> • Décrire un personnage en s'aidant d'un autre texte.
		Savoir produire un écrit cohérent dans une situation de communication déterminée.	<ul style="list-style-type: none"> • écrire le corps d'un récit, le début et la fin étant donnés.

La comparaison des deux tableaux montre que :

- on a dissocié champs, capacités, compétences et composantes;
- les compétences ne sont pas regroupées de la même façon: la connaissance du code devient autonome ;
- la répartition des compétences a changé : ainsi la copie a longtemps appartenu au savoir écrire;
- la décomposition en composantes, en partie commandée par la nécessité d'inscrire le protocole dans le logiciel "JADE", oblige à préciser les tâches évaluées.

Le nombre d'items affectés à chaque compétence varie aussi d'une année à l'autre .

C'est pour ces raisons que, même si les contenus restent à peu près les mêmes, les résultats globaux, exprimés en score moyen pour chaque champ ne peuvent être comparés.

Ce que savent les élèves à l'entrée au cycle 3 d'après les dernières évaluations (1997 à 2002)

Les résultats globaux, exprimés en scores moyens ne sont pas comparables pour de multiples raisons (cf. plus haut). Cependant, on peut repérer sur une même compétence des tendances générales, des niveaux stables ou en évolution.

Pour mesurer ce que les élèves savent aujourd'hui, nous avons choisi de relever les résultats de plusieurs années (1997 à 2001 : protocole relativement homogènes parce que un certain nombre d'exercices y sont repris d'une année sur l'autre) .

1. Compétences globalement maîtrisées en fin de cycle des apprentissages fondamentaux (environ par 7 à 8 élèves sur 10)

1.1 Savoir lire: comprendre un texte

1.1.1. Lire et exécuter des consignes simples

Entre 8 et 9 élèves sur 10 comprennent et savent appliquer une consigne lorsque celle-ci indique des tâches simples et courantes de la vie scolaire; ce résultat est stable.

1.1.2 Repérer et identifier des écrits à partir d'indices matériels et/ou textuels

8 élèves sur 10 savent repérer l'auteur, le titre d'un livre et savent reconnaître par exemple un documentaire, un recueil de poèmes (73,7% en 2001; 88,2 en 2002); un dictionnaire est un peu moins bien reconnu 75% en 2001 et 71% en 2002)

1.1.3 Comprendre un texte

Quand il s'agit de prélever des informations explicites dans un récit, les scores sont très bons.

En 2000, 85,2 à 92,8% des élèves réussissent à des items vérifiant les différents aspects de la compréhension littérale d'un texte de fiction et sur ces mêmes questions en 1998, les scores de réussite vont de 86 à 89%; et en 2002, sur des questions de même nature mais sur un texte plus difficile (Wann), la réussite est de 81%; pour toutes ces questions, le taux de non réponses est très faible, moins de 2%.

On observe à travers l'examen des scores qu'un relevé d'informations peut être difficile si ces informations sont trop éloignées du monde connu des enfants et/ou de leur connaissance des codes d'un type d'écrit particulier; ainsi, 88,7 % des élèves retrouvent la date d'un événement sur une affiche (2002), mais 30,9% seulement repèrent le nom de la ville et 10% des élèves ne répondent pas.

Près de 9 élèves sur 10 identifient correctement les personnages aussi bien dans les récits de fiction (87% en 1999) que dans les récits plus réalistes (86% en 2002).

Des questions qui obligent à des transformations ou à mises en relation de cause à effet sont parfois bien réussies : près de 7 enfants sur 10 comprennent et rédigent une réponse expliquant que c'est la pauvreté qui conduit, le héros, à peindre pour se nourrir. De même, en 1999 et 2000, 2 élèves sur 3 comprennent le sens métaphorique d'une phrase ("manger des étoiles") d'après le sens général du texte.

Ce sont des compétences de haut niveau pour des élèves qui viennent d'apprendre à lire.

1.1.4 se représenter les relations spatiales dans des textes de statuts variés.

En 2001, 88% des élèves identifient des marques spatiales (à droite, à gauche de...) dans un texte simple et les reportent sur un plan; ils manifestent ainsi leur connaissance d'un lexique de base ;ils étaient 83 à 90% à les comprendre dans un récit en 1999 et 2000. En 2002, 90,2% des élèves sont capables de se décentrer, de se situer à un point précis d'un plan pour adopter un point de vue adapté à la tâche qui leur est demandée.

1.1.5 reconstituer la chronologie des événements dans des textes de statuts variés

Plus de 7 élèves sur 10 retrouvent une date dans un récit (1999), une lettre (76,5% en 2000), un agenda (2001 74 à 87% de réussite).

1.2 Savoir écrire

1.2.1 Produire des écrits

Plus de 7 élèves sur 10 (77%) savent produire un écrit bref : une description de 3 phrases en 1999, 2000 et 2002, un message, en 2001, en employant une syntaxe correcte. En 2002, 86% des élèves savent décrire la tête, le costume d'un personnage en s'aidant d'un autre texte.

87% respectent au moins la moitié des critères formels quand ils rédigent une lettre (mise en page, date, signature, ..).

Près de 8 élèves sur 10 savent écrire un récit d'au moins 8 lignes en respectant la consigne; dans un exercice, où leur sont donnés la situation initiale, le héros (Croat, le corbeau), le déclenchement de l'action et la phrase de fin : "Maintenant que tu es mon ami, dit Croat, je suis le plus heureux des oiseaux.", 93% des élèves utilisent le personnage du corbeau comme héros, 80% ont pris en compte le cri pour déclencher pour l'action, 69% ont raconté comment le corbeau se fait un ami.

Environ 60% savent utiliser les temps du récit même s'ils font des erreurs de conjugaison ou d'orthographe en 1999 comme en 2002.

1.3 Utiliser ou reconnaître les outils de la langue pour lire

1.3.1 Reconnaître les mots

Ce sont des activités fondamentales qui assurent l'autonomie du lecteur; plusieurs exercices évaluent cette compétence.

EN 1995, 1997, 1999 et 2000, 3 exercices portaient sur différents aspects de cette compétence

- Les élèves devaient choisir parmi plusieurs mots écrits, celui qui ne contient pas le même son que les autres; par exemple en 1995, trouver les intrus dans les listes suivantes:

un éléphant	le chemin	du lait	un point	la foire
une cane	le train	un piano	le soin	une ardoise
un manche	le jardinier	une fraise	un coin	une noix
il attend	la peinture	la chaise	un lion	un violon

Cet exercice est très bien réussi (80 à 95%);les résultats sont stables.

- Un deuxième exercice demandait aux élèves de repérer rapidement un mot parmi des mots proches ;par exemple retrouver très rapidement les mots prononcés par le maître : l'étiquette, le visage, l'escargot, l'échelle dans les listes:

l'épuisette	village	le berlingot	l'échevelé
l'étincelle	virage	l'escorte	l'escarcelle
l'étiquette	vidage	l'escalator	l'échelon
les tirettes	vivace	les cargos	l'échelle
les tiques	visage	l'escargot	les chaînes
l'étiqueteuse	vissage	l'escalope	les chevilles

Cet exercice est également très bien réussi (82,5 à 95%).

On peut penser que les élèves qui ont une reconnaissance rapide (certains disent automatique) des mots réussissent mieux que ceux qui sont encore au stade du déchiffrage, et que ces élèves ont acquis une bonne maîtrise du code.

- Dans le troisième exercice, les élèves devaient retrouver le mot dit par le maître dans une liste de mots proches qui n'appartenaient pas à son vocabulaire proche, et comportant des graphies complexes; par exemple en 2000, il fallait retrouver et entourer chinchilla dans une liste contenant les mots: cochenille, chevrotain, chrysalide; tamandua au milieu de tarentule, tabernacle, tamarin; docimaste parmi duodénum, décimale, docilité.

Cet exercice, qui oblige au déchiffrage, est très bien réussi: 87 à 95% .

Les élèves qui ont réussi ces trois exercices semblent maîtriser les processus fondamentaux de la reconnaissance des mots.

En 1994, 1998 et 2001, c'est un autre exercice qui est proposé: l'élève doit vérifier que le mot écrit sous une image correspond exactement au référent évoqué par cette image." L'objectif de cette épreuve est de vérifier dans quelle mesure en fin de cycle 2 l'élève devient capable d'une lecture précise et ne se contente pas d'une vérification approximative." (Dossiers 128 mars 2002 page 58). C'est un exercice emprunté à un test de psychologie de l'apprentissage qui exige une très grande concentration.

La réussite globale sur trente mots, à une ou deux erreurs près, est de 52% en 1994, 58% en 2001 et 2002; pour ces élèves l'apprentissage du code semble assuré. Mais ce résultat cache de grandes disparités: en 2001 les Z.E.P. affichent un déficit de 16 points dans cette épreuve; un élève sur dix est en très grande difficulté avec huit erreurs et plus. Plus d' un élève sur trois (37,6%) à l'entrée au cycle 3 fait entre trois et sept erreurs et n'a donc pas fini son apprentissage de la lecture des mots. En 2002, l'écart n'est que de 11,2, mais il existe quand même 34% des élèves qui font de trois à sept erreurs et 7,5% qui font huit ou plus de huit erreurs.

On voit donc que selon le type d'exercice, les appréciations sur la reconnaissance de mots peuvent être divergentes: très bonne réussite des élèves (jusqu'à 95%) dans un cas; et dans une situation qui n'est pas familière, réussite de moins de six élèves sur 10.

I.3.2 Copier

74 à 85% des élèves recopient intégralement un texte plus ou moins long en respectant sa mise en page (textes rimés) ou en transformant cette présentation : par exemple copier un message et une adresse sur une maquette de carte postale; dans ce cas les résultats sont très dispersés et les élèves de Z.E.P. sont en plus grande difficulté (16 points d'écart en 2000 alors que la moyenne des écarts est de 11). On peut attribuer cet écart à une méconnaissance des codes de présentation de ce type d'écrit.

Malgré cette bonne réussite d'ensemble, 15% à 16% des élèves n'arrivent pas en 1999 , 2000 et 2002 à terminer la copie dans le temps imparti; ce qui constitue un handicap certain pour les multiples activités scolaires qui utilisent la copie. Ce constat est conforme à l'étude sur les 10% d'élèves les plus faibles (cf. l'étude 1987-1997 qui souligne que nombre de ces élèves travaillent très lentement.)

L'orthographe peut être considérée comme correcte lors de ces exercices de copie si on comptabilise les copies incomplètes: en 2002, sur un texte de huit vers comportant des mots rares 73,6% des

élèves recopient intégralement le texte, la disposition en vers et la ponctuation sont respectées par 70% d'entre eux à un oubli près. On peut noter que c'est une des épreuves qui ne discrimine pas les Z.E.P. (écart de 1,2 point) mais que les filles obtiennent un meilleur score que les garçons.

1.3.3 Ecrire sous la dictée

Quand la dictée est préparée, plus exactement quand les élèves ont eu le texte sous les yeux même brièvement, les résultats sont corrects.

Pour les mots invariables très fréquents, les écarts de réussite sont très importants entre des mots également fréquents sans qu'on puisse expliquer ces différences; ainsi *comme* est correctement orthographié par 81% des élèves, *chez* par 68%, *alors* par 52%.

1.3.4. Transformer un texte en appliquant des règles simples

La transformation d'un texte du féminin au masculin est réussie par plus de huit élèves sur dix; seul l'accord de "*ami*" (qui ne s'entend pas à l'oral) n'est réussi que par près de sept élèves.

Le passage du pluriel au singulier est plus délicat surtout avec des mots comme taureaux/taureau, animal/animaux (63% de réussite) et un piège comme "bois" est quand même réussi par plus de six élèves sur 10.

2 Compétences installées mais encore fragiles à l'entrée au cycle 3

2.1 Comprendre un texte

Dès qu'il s'agit de mettre en relation plusieurs informations et/ou de produire des inférences, les résultats sont naturellement plus faibles.

Quand il s'agit de choisir la phrase qui résume le mieux le texte parmi quatre autres phrases, en 2000, 44% des élèves choisissent la réponse juste et 30% choisissent une réponse non contradictoire avec le sens du texte mais partielle.

En 2002, dans les quatre résumés du texte qui sont proposés aux élèves, deux font contresens avec le texte, un est acceptable, un seul correspond au point de vue adopté par le texte; 50% des élèves font le bon choix, et 20% le choix acceptable. Ce sont des résultats très bons pour des élèves de CE2 qui ne sont pas habitués le plus souvent à des distinctions aussi fines.

On peut dire que 70% des élèves comprennent le sens général du texte "en gros", mais que la moitié seulement d'entre eux manifeste une compréhension rigoureuse.

2.2 Produire un texte

Lorsqu'un modèle de texte est donné, lettre, description, notice de fabrication, les élèves savent se servir de ce texte comme point d'appui et rédigent des textes globalement conformes à la structure demandée. Ce qui reste difficile, c'est de garder cette cohérence tout au long du texte produit. Il en est de même pour les temps du récit dont le système est compris par 60% des élèves, mais qui n'est maîtrisé que par 30% d'entre eux.

La capacité à choisir un registre qui est celui de l'écrit et non de "l'oral transcrit" est bien mise en oeuvre dans 30% des textes produits, 40% des élèves mêlant des formes de l'oral à celles de l'écrit.

2.3 les outils de la langue pour comprendre et lire

2.3.1. pour mieux comprendre les mots d'un texte

Choisir la bonne définition est réussi à 69% en 2001 pour le mot "carte".

De même en 2002, distinguer le mot selon le contexte est réussi à 78% pour le nom "bouton" mais à 57% pour le verbe "monter".

Les familles de mots sont encore mal perçues: associer l'adjectif "routière" dans l'expression "carte routière" à la famille "autoroute, route, routine, dérouter, dérouté" est réussi par 36,71% des élèves; 19,73% l'associant à "roue, rouler, roulade..." et 33,18% associant "routière" à une liste construite autour de "rouge".

On voit que les élèves peuvent avoir une compréhension globale du mot sans avoir conscience des correspondances sémantiques.

La valeur des préfixes négatifs est reconnue à plus de 75% lorsqu'il s'agit de "mal" mais seulement à 57% avec le préfixe "dé"; il faut noter là un très grand écart avec les Z.E.P. où le mot «malade» est

compris comme composé avec le préfixe "mal"; comme si pour ces élèves, la langue n'était pas dissociée de l'expérience, comme si la langue n'avait pas d'existence propre.

2.3.2. identifier les accords, les marques temporelles qui construisent la cohérence du texte. Sept élèves sur dix repèrent et choisissent les temps du futur quand il s'agit de choisir parmi plusieurs formes verbales celle qui convient. ((2002).

La concordance des temps est respectée par moins d'un enfant sur deux quand il faut transformer un texte du présent au futur; les marques de la 3ème personne du singulier sont mieux connues(à plus de 40%) que celles du pluriel (à moins de 30%).

2.3.3 choisir les connecteurs qui assurent la cohésion du texte.

Les élèves doivent choisir entre plusieurs connecteurs dans un texte; "mais" est choisi par plus de huit élèves sur dix; "pour que " et les connecteurs qui impliquent des relations logiques comme la cause/conséquence ne sont retrouvés que par un élève sur deux; ce qui semble cohérent avec le développement cognitif d'enfants de cet âge.

3 Compétences que seuls réussissent les meilleurs

3.1.Savoir lire

3.1.1 comprendre un texte

Quand la question porte sur les intentions de l'auteur (1999), et qu'il faut être attentif à des indices fins (syntaxe, changement de temps verbaux), seuls 21% des élèves réussissent à trouver la suite qui convient. Ces élèves manifestent un sens de la langue bien au-delà de ce qu'on attend habituellement d'un enfant de cet âge.

Dans ces questions plus difficiles, le taux d'absence de réponse est élevé : de 13 à 17%, et les écarts ZEP/non ZEP se creusent (19,5 en 2002).

3.1.2.Reconstituer la chronologie des événements dans des textes de statuts variés

Trouver la date du "lendemain" quand il faut la déduire de plusieurs informations non explicites est réussi par un élève sur quatre en 1999.

4. Produire un texte

Les consignes et le type de texte demandé font varier les difficultés.

Quand la consigne exige d'inclure une phrase dans un récit: "A ce moment-là, un bruit étrange lui signala qu'il (elle) n'était pas seul (e).", la plupart des élèves respecte bien la consigne mais un élève sur deux arrive à terminer son récit.

La ponctuation, les majuscules ne sont utilisées correctement dans l'ensemble du texte que par un tiers des élèves en 1997 et par 24% en 2002.

Le système d'alternance des temps du récit est présent dans 60% des textes produits.

La syntaxe est jugée sans faute pour 46,3% des élèves (1997) .

5. Utiliser et reconnaître les outils de la langue

- Copier sans aucune faute un texte rimé de six à huit lignes n'est réussi que par 31 à 33% des élèves en 1999 et 2000 .
- Orthographier une courte phrase dictée du type: "Trois petits chats jouent dans la cour" ou " Pendant la récréation, les garçons et les filles jouent aux billes." est une épreuve difficile en particulier pour les accords du sujet avec le verbe.

- Identifier les indices syntaxiques et morphologiques.
-

Lorsqu'on donne un message, où on ne peut savoir si *Steph* ou *Claude* sont des filles que par des marques d'accord sur des adjectifs ou participes comme "*prête*" "*rentrée*", seuls 25,8% des élèves savent repérer le féminin et répondre que *Steph* est une fille .

Des exercices de ce type sont donnés régulièrement aux évaluations et les résultats sont toujours du même ordre; le pluriel est mieux repéré et interprété (44,9% des élèves en 1999 et 48,7% en 2000).

Conclusion

Les résultats sont cohérents avec les analyses précédentes.

On voit donc aujourd'hui, comme déjà des rapports antérieurs le soulignaient, que les enfants savent à plus de 80 % lire des textes simples et en extraire des informations explicites. Dès qu'il s'agit de traiter l'information, de mettre en relation plusieurs informations, voire d'inférer, les résultats sont moindres, ce qui paraît normal à 8 et 9 ans.

La reconnaissance de mots est très réussie (par plus de huit élèves sur dix) quand il s'agit de repérer des mots, d'assurer des correspondances phonie/graphie; mais quand l'exercice est inhabituel et qu'il exige des opérations cognitives plus complexes, les résultats chutent.

Tous les outils de la langue sont en cours d'acquisition (marques de genre de nombre, marqueurs syntaxiques et morphologiques, marques temporelles), ce qui est normal à la fin du cycle 2, le travail systématique sur l'analyse de la langue n'est approfondi qu'au cycle 3.

La majorité est capable d'écrire un texte de plus de 8 lignes cohérent, activité très difficile.

Les outils scolaires: lecture de consignes, copies sont aussi bien assurés.

En revanche l'orthographe en situation de dictée est très mal maîtrisée.

Bien que les compétences ne soient évaluées que sur un nombre limité d'items et dans des exercices de difficultés différentes, ces scores rendent compte d'un niveau général.

Plus de 10% des élèves n'ont pas acquis les compétences nécessaires

Les chiffres de 85 et 90%, parfois même au-delà, sur les compétences de base sont satisfaisants mais ils montrent aussi que selon les exercices, 10 à 15% des élèves sont en difficulté sur des compétences simples et qui engagent la scolarité à venir. Comment faire progresser ces élèves de telle sorte qu'ils acquièrent au plus vite ces compétences sans lesquelles ils ne peuvent satisfaire aux exigences du cycle 3? C'est ce à quoi les P.P.A.P (programmes personnalisés d'aide et de progrès), créés en 1998, veulent servir.

Les élèves de Z.E.P. sont en grande difficulté sur le traitement de l'information, sur tout ce qui semble exiger des compétences de haut niveau et sur ce qui implique la connaissance de codes sociaux et culturels; on peut penser que c'est le cycle 3 qui assurera un progrès de ces compétences.

Les écarts sont moins importants sur les tâches scolaires effectuées régulièrement et où les codes sociaux interviennent peu (lecture de consignes simples, copie, dictée, compréhension nécessitant un prélèvement d'information) sur ce qui est travaillé systématiquement à l'école et seulement à l'école.

Ceci pourrait donner à penser qu'il faut renforcer les apprentissages scolaires sur les compétences moins bien acquises.

La dispersion des scores est très grande

Depuis 2001, la DEP a raffiné ses analyses statistiques et a fait apparaître les écarts entre les 10% les plus forts et les 10% les plus faibles.

En 2002, en français, au CE2

Champs	Ensemble des élèves			10% les plus faibles			10% les plus forts		
	moyenne	borne inf	borne sup.	moyenne	borne inf	borne sup	moyenne	borne inf.	borne sup.
Savoir lire	68	71	17,1	36,2	0	64,5	90,9	80,6	100
Savoir écrire	68,4	72,2	19,6	33,9	0	75	92,1	72,2	100

Capacités									
Comprendre		71,4	17,6	36,6	0	71,4	91,1	78,6	100
Maîtriser les outils de la langue	68,6	70,4	19,1	32,8	0	61,4	91,3	77,3	100
Maîtriser les outils de la langue pour lire	66,8	70	19,7	35,4	0	70	90,4	65	100
Maîtriser les outils de la langue pour écrire	66,8	70,8	21,9	30,6	0	79,2	92	66,7	100
Produire un texte	71,9	75	22,3	40,6	0	91,7	92,5	50	100
Score global	68,2	70,4	16,9	35,4	3,1	44,9	91,3	87,8	99

A l'intérieur de chaque champ, les scores sont très dispersés. Par exemple, les 10% les plus faibles en production de textes obtiennent des scores compris entre 0 et 91,7. Ces grandes dispersions peuvent être dues au petit nombre d'items; cependant, on observe que des élèves peuvent être très faibles sur l'ensemble du protocole et réussir bien, voire très bien, dans une champ donné.

On constate que les 10% les plus faibles réussissent mieux en lecture qu'en production de textes, alors que c'est l'inverse pour les 10% les plus forts.

Les items qui sont les plus discriminants entre ces deux populations extrêmes sont en 2002:

- la représentation chronologique d'une suite d'événements (réussie à 100% pour les meilleurs et à 42% par les élèves les plus faibles) et respectivement à 100% et 37% sur un autre item évaluant la même compétence.

- la reconnaissance des mots varie de 96,4 à 22,8%; la construction de la signification générale d'un texte par le choix d'un titre de 99,1% à 23,4%.

- En production d'écrit les 10% les meilleurs ont une moyenne de 98,9% à l'item qui code l'usage d'une langue qui se situe dans le registre de l'écrit; les 10% les plus faibles restent dans le registre de l'oral; 36,1% mêlent langue écrite et langue orale.

- L'utilisation des marques pour comprendre un message: pour un item, l'écart est de 71,9 à 10,4%; et pour un autre item, il va de 89,2 à 27,7%.

Les résultats et les moyennes mêlent compétences exigibles et compétences remarquables.

Les renseignements que donnent ces évaluations sont à élucider: il faut distinguer ce qui est exigible à ce niveau de scolarité de ce qui est à développer. De ce point de vue, les moyennes ne donnent pas une image juste des compétences acquises par les élèves au regard de celles qu'ils doivent acquérir. Autant leur présence dans le protocole est intéressante pour les maîtres qu'elles engagent dans de nouveaux apprentissages et qu'elles invitent à ne pas négliger les excellents élèves, autant elles sont un mauvais indicateur pour savoir si le niveau des élèves est conforme à ce que les Instructions officielles, les parents et la nation sont en droit d'attendre.

Comparaisons ayant comme support les évaluations de 1989 à 2002

I Comparaisons des résultats

Les mêmes exercices ont été donnés plusieurs fois au cours de ces 14 dernières années, même s'il faut prendre des précautions méthodologiques (ordre des exercices, surcharge cognitive...), il est intéressant de comparer les résultats; on peut dégager certaines tendances.

1.1 sur la reconnaissance de mots

Les exercices portent sur ce qu'on appelle les fondamentaux : la reconnaissance par voie directe et indirecte des mots (cf. annexe).

Le même exercice a été donné en 1994, 1998, 2001, 2002.

La progression est globale : de 52% des élèves en 1994 à 58% en 2001 et 2002 identifient les trente mots à une ou deux erreurs près; mais la dispersion des résultats est forte:

Il faudrait pouvoir analyser si l'écart s'est creusé entre les meilleurs et les plus faibles.

1.2. sur la compréhension en lecture

Le même texte (un récit) a été proposé en 1994 et 2002 avec certaines questions identiques et d'autres légèrement modifiées.

Sur les mêmes questions, les résultats progressent de 58% à 61% et de 58,4 à 69% sur une question très difficile qui implique de bien distinguer cause/conséquence/effet.

Il faut noter aussi qu'il y a moins de "non-réponses" en 2002.

On constate une progression générale mais avec une forte dispersion; la moyenne des réussites en Z.E.P. est de 20 points inférieure à la moyenne nationale.

Le récit de "Petit Gaston" a été donné avec les mêmes questions en 1995 et 2000.

Les résultats sont stables et excellents de 85% à 92,8 % pour les trois premières questions qui vérifiaient la compréhension littérale; les questions suivantes qui supposent un traitement de l'information plus important sont réussies à plus de 60%. la question la plus difficile qui consistait à choisir le titre parmi des propositions toutes vraisemblables est la moins réussie en 1995 comme en 2000 (44%). C'est sur ces derniers items que l'écart est particulièrement important en défaveur des élèves de ZEP (- 17%).

1.3. sur la production de texte

La même situation d'écriture avec la même consigne a été donnée en 1999 et en 2002: écrire un récit dont le début et la fin sont donnés. Seule la présentation est légèrement différente (l'amorce du récit a été écrite en cursive en 2002, et non en script) et le nombre d'items est très légèrement modifié.

On constate entre 1999 et 2002 sur les mêmes items des progrès sur tous les items sauf un.

Items	scores en 1999	Scores en 2002
le récit écrit a plus de 8 lignes	51,2%	63,7%
l'élève a utilisé le héros	88,8%	92,7%
l'élève a tenu compte de la situation initiale pour démarrer son récit	65%	79,8%
il a tenu compte de la phrase finale	60,3%	69,2%
il a produit une suite d'événements cohérents entre eux	76,2%	75%
il a su utiliser les temps du récit	56,7%	60%

1.4. sur l'orthographe en situation de dictée

En 1989, 1992, 1995, 1999 et 2002, c'est la même phrase qui a été dictée:

"Pendant la récréation, les garçons et les filles jouent aux billes."

En 1999 et 2002, le texte de la dictée a été mis sous les yeux des enfants quelques minutes et l'enseignant a lu la phrase en pointant chaque mot et en demandant aux élèves de bien regarder comment les mots étaient écrits.

Le tableau des résultats comparés montre la dégradation régulière que la préparation ne parvient même pas à enrayer, sauf sur quelques accords.

pourcentage de réussite	1989	1992	1995	2002 "préparée"
Pendant	57,8%	53,3%	36,6%	51%
Récréation	73,4%	68,9%	62%	69%
Garçons	73,6%	74,2%	59,1%	62%
Et	87,7%	87,6%	85,9%	87%
Filles	77,5%	75,8%	64,4%	69%
Jouent	34,1%	53,7%	25,3%	61%

Aux	32,9%	31%	30,6%	59%
Billes	54,7%	53,6%	46,8%	61%

Conclusion

Si on peut s'aventurer à dégager une tendance générale à partir de si peu d'éléments, on peut penser que le niveau moyen s'améliore en lecture (mécanismes de base d'identification et compréhension) et en production écrite mais avec une forte, dispersion, et que le niveau moyen en orthographe s'est dégradé.

On ne peut donc comparer qu'un petit nombre d'items de 1989 à 2002, ceux qui portent sur des exercices repris d'un protocole à l'autre; cependant, il est intéressant de comparer ce qui est demandé aux élèves depuis 1989 pour voir si sous les mêmes termes, on met les mêmes contenus et dégager éventuellement des tendances.

II L'évolution des évaluations depuis 1989

Les protocoles ont évolué; la répartition des exercices et leur classement par champ.

La Lecture

1. La reconnaissance des mots

En 1989, les épreuves de reconnaissance de mots sont d'une nature différente: il s'agit de :

- relever des mots ,
- distinguer des homophones,
- repérer des graphies.

Ces exercices sont beaucoup moins complexes que ceux donnés par la suite et surtout moins déstabilisants que l'épreuve où il faut à la fois associer une image à un mot et repérer plusieurs types d'erreurs .

D'autre part, face à ce tableau comportant 30 images et 30 mots, l'élève doit se débrouiller seul, et non comme dans les exercices de 1989 étape par étape, guidé par le maître, chaque étape ne concernant qu'un type d'erreur (annexe).

2. La compréhension

1. Les textes de lecture ont évolué

- en longueur

Le texte de 89 comprend 4 phrases ,70 mots et c'est un texte qui a été repris d'un manuel ; il est donc proche de ce que les élèves fréquentent en classe. Tous les items sont réussis à plus de 80% et c'est sans doute pour cela, qu'on a choisi ensuite des textes plus difficiles .

- en nature

Les textes se sont diversifiés: documentaire sur la préhistoire, sur la vie des enfants au Mexique ; ces textes exigent des connaissances spécifiques, qui même quand elles portent sur le programme ne sont pas forcément acquises par les élèves. Le groupe national , ayant bien analysé ces difficultés, a renoncé à ce type d'écrit ces dernières années et choisi soit des contes, type familier des élèves depuis l'école maternelle, soit des textes proches de leur vie .

2. les questions sur ces textes

Elles comportent toujours un prélèvement d'informations explicites qu'il suffit de repérer; ces questions sont de difficultés équivalentes et les résultats toujours très bons.

Les questions qui obligent à un traitement de l'information sont devenues au fil du temps très complexes, par exemple anticiper la suite, ou saisir que le texte est dans le registre du merveilleux . Elles ont permis de mettre en valeur des compétences acquises par peu d'enfants encore au CE2, compétences qualifiées d'ailleurs à l'époque de " remarquables ", et d'encourager les maîtres à travailler ces compétences au cycle 3.

L'année 1994, avec le choix d'un texte comme Wan , constitue un tournant dans la conception des questions de compréhension , mettant en œuvre des processus cognitifs plus complexes et prenant en compte la dimension littéraire du texte.

3. Les formats de réponses

En 1989 , la plupart des items offrait seulement deux choix: Vrai/Faux. On s'est bien rendu compte que ce choix pouvait être aléatoire et renseignait peu sur les compétences testées; on a donc ensuite

proposé 3 voire 4 choix; et parfois même dans ces 4 choix des "pièges", c'est à dire des réponses plausibles mais non rigoureuses.

C'est dire, si les élèves peu sûrs, ceux qui doutent, peuvent se trouver désarmés devant de tels choix. Les statisticiens ont les moyens de vérifier si les résultats à ces exercices sont cohérents avec la réussite générale de l'épreuve; en général ils le sont; mais on constate aussi, entre 1989 et 2002, sur ces items difficiles, une forte augmentation des non-réponses.

On peut conclure qu'à score égal, les réussites d'aujourd'hui témoignent de compétences supérieures en compréhension.

II La production d'écrit

1. Les contraintes ont changé: ainsi en 1989, 1990 et 1991, on demande d'inventer une histoire intéressante en utilisant au moins trois animaux et/ou objets (proposés dans une liste) et de lui donner un titre.

En 2001, il faut construire un récit autour d'une phrase donnée: " A ce moment-là, un bruit étrange lui signala qu'il (elle) n'était pas seul (e) dans la maison."

En 1999, 2000, 2002 et 2003, on donne un premier paragraphe qui situe l'action et nomme les personnages de l'histoire, et on donne la phrase de fin: "Et c'est ainsi que Croa se fit un ami."

Il faut donc non seulement écrire un récit cohérent par rapport aux données de départ mais aussi faire progresser ce récit vers cette fin et celle-là seule.

Les compétences évaluées dans ce type d'écriture sont, on le voit, beaucoup plus complexes

2. les types d'écrit se sont diversifiés:

En 1994 et 1996 les élèves ont à rédiger une fiche technique ("Comment nourrir les oiseaux l'hiver?" " Comment fabriquer un masque?") ; ils doivent écrire des légendes pour décrire et expliquer des vignettes.

En 1995 et 1997, c'est une lettre qu'il faut écrire en réponse à une lettre donnée comme référent, et en outre, il faut dans cette lettre décrire un objet, expliquer comment il fonctionne et donner son avis.

On voit que dans cet exercice de nombreuses contraintes sont à l'œuvre et que la tâche demandée à des enfants de 8/9ans est très complexe.

3. un nouvel exercice a été introduit en 1998.

Il s'agit de décrire en trois phrases un personnage, un animal avec l'aide d'un texte de même sujet et même structure.

La compétence " savoir écrire" est, depuis, jugée sur deux écrits différents; cet écrit bref permet d'apprécier ce que les élèves même faibles et embarrassés parfois par une écriture "longue", savent produire.

4. Les critères de correction ont eux aussi évolué.

Les évaluations nationales de 1989 s'inspirent des évaluations précédentes en essayant de les adapter aux contraintes d'un dispositif de masse. Les items varient en nombre de 19 en 1991 à 8 en 2002; mais, ils cherchent à identifier toujours les mêmes compétences: écrire un texte de plus de 8 lignes,

- respecter la consigne et le type de texte,
- tenir compte de la situation donnée,
- produire une suite d'événements cohérents entre eux
- utiliser les temps du récit
- utiliser la ponctuation et les majuscules.

Les regroupements d'items concernent principalement la maîtrise de la langue écrite: on détaillait en 1991 et 1992, la variété du vocabulaire et son adaptation au sujet, la correction syntaxique, pour aboutir en 2002 à une formulation plus synthétique: " utiliser le registre de l'écrit".

Conclusion

La production d'écrit a d'une part évolué vers plus d'exigences dans les contraintes et, d'autre part, pris en compte les élèves qui sont encore à une étape où ils ont besoin d'un modèle, d'une structure pour écrire. Cela a permis de mieux évaluer la totalité de la population scolaire de CE2 et en particulier ceux qui sont les moins habiles sans pour autant être en échec.

III La copie

Là aussi les épreuves se sont complexifiées:

- les textes sont plus longs:

En 1989 la copie porte sur une phrase; elle est réussie à 61.4% avec la ponctuation (point et majuscule) et 28.7% des élèves copient sans faute mais en oubliant la ponctuation.

En 2002, c'est un texte de 8 vers qu'il faut recopier.

- les normes de présentation des textes se sont diversifiées.

Il faut par exemple, recopier en respectant une mise en page (textes rimés), en distinguant message et adresse (carte postale), en insérant les phrases dans les paragraphes déjà disposés d'une fiche technique.

Quand 85% réussissent cette copie assortie d'une mise en page, ils réussissent à coup sûr une épreuve plus difficile que celle proposée en 1989; et quand 51.8% réussissent la copie et mise en espace de la carte postale avec disposition de l'adresse et orthographe des noms propres, il est difficile de savoir si ces élèves sont meilleurs que les 61% qui avaient réussi la copie d'une seule phrase en 89, mais ils ont été confrontés à une tâche beaucoup plus difficile.

IV L'orthographe

Les épreuves sont restées sensiblement de même difficulté depuis 1989 :

- des mots invariables dictés et réussis de façon disjointes,
- une dictée d'une ou deux phrases du type: "Pendant la récréation, les garçons et les filles jouent aux billes." (cf. plus haut) Les résultats se sont dégradés.

En conclusion, il semble qu'en lecture, aussi bien en compréhension qu'en reconnaissance des mots, les épreuves sont plus difficiles et donc que les compétences évaluées sont d'un niveau supérieur.

En production d'écrit long, la difficulté aussi est plus grande, mais elle est équilibrée par la rédaction d'un texte bref; la nature des compétences évaluées en production d'écrit est plus complète, plus complexe et prend en compte les plus faibles comme les meilleurs.

En orthographe, les résultats en dictée se sont dégradés.

Tout fonctionne comme si les groupes nationaux, analysant au fur et à mesure les résultats, raffinaient eux-mêmes leurs analyses, ceci les poussant parfois à proposer des questions de plus en plus subtiles comme en lecture, et parfois à introduire un nouvel exercice, le texte bref, pour mieux prendre en compte l'ensemble des élèves.

Comparaisons à partir d'épreuves antérieures à 1989.

Les comparaisons sont d'autant plus difficiles que :

- les finalités des évaluations sont différentes :

"Cette évaluation (l'évaluation diagnostic) se distingue de toutes celles réalisées jusqu'à présent dans l'éducation nationale. En effet, alors que ces dernières étaient destinées à améliorer la connaissance globale des acquis des élèves aux différents niveaux de la scolarité, celle qui est mise en œuvre aujourd'hui a pour but de vous aider à mieux connaître vos propres élèves et plus précisément identifier, dès le début de l'année scolaire, d'éventuelles lacunes dans le champ des apprentissages fondamentaux: lecture, écriture, mathématiques." (L.Jospin 28/7/1989)

- la conception même de la durée de l'apprentissage de la lecture a évolué. Depuis la loi d'orientation qui définit les cycles, la lecture est enseignée pendant les 3 ans du cycle 2. Avant, le cycle préparatoire comprenait seulement le Cours préparatoire et on a donc évalué en fin de C.P. voire en fin de premier trimestre de CE1. Or quand, comme en 1989, on a fait passer aux mêmes élèves les mêmes épreuves en juin et en décembre de la même année, on a évidemment constaté des progrès importants quoique inégaux. Il est donc impossible de comparer des résultats observés à des moments différents des apprentissages. (cf. historique).

- Les programmes ont changé.

- La population scolaire a changé.

Le décret du 6 septembre 1990 dispose que la scolarité dans l'ensemble des deux derniers cycles de l'école primaire ne peut être prolongée de plus d'un an. Les effets de cette évolution s'observent dans l'analyse des résultats: ainsi en 1987, 64% des élèves "sont à l'heure" au CM2, alors qu'ils sont 76,9% en 1997. Ce qui veut dire qu'on évalue dans la même classe aujourd'hui des élèves qui auraient sans doute redoublé dans la classe précédente.

A score égal, les enfants évalués sont donc dans l'ensemble d'un meilleur niveau que ceux qui étaient évalués avant la mise en place effective de ce décret.

Dans les évaluations d'avant 1989,

1. l'équilibre langue, lecture, production d'écrit est différent ; les exercices de langue (grammaire, vocabulaire, conjugaison) sont plus nombreux.

2. les épreuves sont parfois différentes ; en lecture compréhension, les réponses sont sous forme de Vrai, Faux ou QCM, aucune question ne demande de réponse rédigée.

- la reconnaissance de mots s'apparente aux exercices les mieux réussis depuis 1989 (cf.: ce que les élèves savent).

- la production d'écrit se fait à partir de 4 images .

3. les critères de correction sont plus simples.

4. en orthographe, les dictées sont plus longues et plus difficiles.

Cependant, on retrouve des tendances lourdes constatées depuis, sur la lecture: déchiffrage et reconnaissance des mots ne sont pas maîtrisée par tous ; sur les difficultés de compréhension autre que littérale, sur la difficulté à appliquer les règles de grammaire . Seules les exigences en orthographe se sont amoindries.

L'appréciation des compétences et les pratiques pédagogiques des maîtres du premier degré

Il est intéressant de rapprocher les évaluations nationales d'autres évaluations de statut différent et en particulier lorsque ces évaluations s'accompagnent d'observations dans les classes.

En 1995, dans le cadre de son programme de travail, une mission d'évaluation sur l'apprentissage de la lecture à l'école primaire a été confiée à l'inspection générale.

Au CP, les observations ont porté sur 153 classes et selon les objectifs suivants:

- identifier et caractériser les démarches pédagogiques à propos de l'enseignement de la lecture,

- apprécier l'adéquation de ces démarches aux recommandations en vigueur,

- analyser l'efficacité de ces démarches en les rapportant aux résultats d'épreuves, organisées dans les classes à la fin de l'année scolaire.

Les conclusions sont formulées ainsi :

"Des instructions de 1985, on a essentiellement retenu les formules "lire c'est comprendre", " développer le désir de lire", "présenter des textes divers" et "la lecture silencieuse est l'objectif à atteindre". On ne perçoit pas clairement comment associer "la maîtrise de la combinatoire...à la compréhension de l'idée exprimée, à la perception de la structure de la phrase, à l'intelligence des mots."

- Les points faibles

- la brièveté du temps réel de lecture à voix haute de l'élève,

- l'absence quasi générale d'articulation entre les activités concernant la combinatoire, celles portant sur la reconnaissance des mots et leur sens, et celles portant sur la signification des textes,

- le peu d'attention accordée aux procédures utilisées par l'élève pour connaître les mots et accéder au sens,

- le peu d'importance attachée à la production de textes.

Les maîtres semblent avoir peu été touchés par les débats sur les méthodes de lecture; dans leur très grande majorité , ils associent compréhension des textes et décodage. Leurs interrogations portent essentiellement sur l'équilibre et l'articulation entre ces activités.

- Les points forts

- l'importance accordée à la place des textes, à leur diversité;

- le soin apporté par la majorité des maîtres à rechercher un équilibre entre le travail sur le sens et le travail de déchiffrage;
- Une assez bonne liaison entre lecture et écriture.

"L'équation personnelle du maître, son niveau d'exigence, son savoir-faire sont signalées comme des facteurs essentiels de la réussite de la classe observée."

- De ces observations, les inspecteurs généraux ont dégagé des facteurs scolaires de réussite:
 - une expérience de l'enseignement au cours préparatoire,
 - une utilisation du manuel scolaire enrichi par des textes divers,
 - un enseignement menant de front apprentissage du code grapho-phonologique et accès au sens.
 - dans les classes où on obtient les moins bons résultats, on privilégie nettement le code par rapport au sens ou l'inverse,
 - une articulation entre pédagogie de la lecture et pédagogie de l'écriture,
 - une explicitation systématique des procédés utilisés par les élèves.

"Il y a un déficit quasi-général de notre pédagogie sur l'analyse des erreurs et il serait absolument nécessaire d'entraîner les maîtres à réagir correctement aux productions de leurs élèves dans les grands types d'apprentissage" précise le rapport.

 - une dimension culturelle de la lecture.

La présence de la lecture sous toutes ses formes et la très forte implication personnelle des maîtres dans la lecture à haute voix de contes et d'histoires *"semblent bien caractériser"* les classes où on obtient les meilleurs résultats.

 - une meilleure prise en compte de la continuité grande section de maternelle/ cours préparatoire.
 - dans la gestion de la classe: adaptation des tâches à la diversité des élèves et sollicitation constante de tous.

Il est intéressant de noter que, lorsqu'on interroge les maîtres sur leurs difficultés, ils répondent plutôt sur les difficultés des élèves; c'est souvent une approche globalisante de ces difficultés qui mêle problèmes des élèves et didactique de la lecture; ce qui empêche une analyse rigoureuse des obstacles cognitifs à l'apprentissage de la lecture, ou les problèmes d'adaptation de l'enseignement à la diversité des élèves.

Au CM2, les observations ont porté sur 79 classes. Les conclusions sont plus sévères:

- Les objectifs poursuivis par les maîtres sont formulés en termes généraux: accéder à la compréhension d'un texte, développer le plaisir de lire...

En général, la définition des objectifs annoncés ne passe pas par une analyse précise de l'acte de lire, de ses composantes et des compétences qu'il suppose chez les élèves. Elle ne fait pas l'objet d'un travail organisé selon une progression établie.

- La nature des activités reste superficielle.

L'étude porte souvent sur ses textes très divers photocopiés; il s'agit souvent de "balayer rapidement l'ensemble du texte et "d'en dégager des informations qu'il est d'ailleurs possible de prélever sans bien le comprendre. On répond à un questionnaire écrit ou à une interrogation orale; la lecture à voix haute est rare et n'est pas intégrée à la démarche.

La lecture suivie de romans est minoritaire et très localisée.

L'entraînement à la lecture rapide part du principe que vitesse et compréhension vont de pair,

Les conclusions sont claires :

L'enseignement de la lecture au CM2 est insuffisant; deux domaines sont particulièrement négligés:

- l'exploration systématique du code alphabétique et notamment des relations qui ne sont pas totalement acquises au cycle 2
- la lecture attentive de textes de moins en moins simples, de plus en plus longs, impliquant une progression.
- il est nécessaire de poursuivre l'enseignement de la lecture au cycle 3.

Cette étude pointe bien les écueils et fait des propositions simples; elle rejoint les conclusions des études faites en 1983

Le cours préparatoire peut être amélioré mais les efforts doivent aussi et surtout porter sur le cycle 3 où les maîtres ne considèrent plus l'apprentissage du code comme de leur ressort, et où la lecture de textes longs et plus complexes n'est pas considérée comme un apprentissage, et donc pas organisée comme tel.

En 1997 est reconduite une enquête qui avait été réalisée en 1990 sur les pratiques déclarées de l'enseignement de la lecture au C.P. et au CE1.

Cette étude qui porte sur 1500 enseignants, constate que :

- pour la lecture

- Le temps d'enseignement hebdomadaire de la lecture est très variable

Un quart des classes y consacre moins de 5 heures et un quart plus de 8 heures au CP.

En CE1, le temps consacré à la lecture diminue : en moyenne les élèves ont 4 heures de lecture. Mais les situations sont très contrastées: un quart des classes a moins de 2 heures 30 et un quart des classes au moins 5 heures.

Par rapport à 1990, le temps moyen consacré à l'enseignement du français semble avoir diminué: en 1990, plus de six enseignants sur dix déclaraient y consacrer 12 heures ; en 1997 la moitié des enseignants disent y consacrer moins de 10 heures. En 1990, 30% des enseignants disaient consacrer plus des trois quarts du temps d'enseignement du français à la lecture; en 1997, ils sont 15% .

- Au CP, les activités les plus fréquentes mettent l'accent sur le mot, la phrase et la correspondance phonie/graphie; en fin d'année, elles sont davantage situées sur la phrase et le texte.

96% des maîtres vérifient la compréhension en posant des questions sur le texte.

En CE1, 85% placent en tête les activités de lecture "libre" et lecture de texte avec questions. Viennent ensuite les activités sur des phrases entières ; en fin d'année, ce sont presque exclusivement ces activités qui sont proposées.

On constate donc que l'apprentissage du code est peu présent dès la fin du CE1., parfois même dès la fin du CP.

- sur les conceptions des maîtres

Les compétences classées comme premières par les enseignants de CP et de CE1. sont d'abord la conscience des fonctions de l'écrit, la maîtrise des correspondances entre les lettres et les sons, et l'anticipation des mots dans une phrase.

L'utilisation du manuel décroît entre le début du CP et la fin du CE1; les manuels sont plus utilisés en 1997 qu'en 1990.

- sur la production d'écrit

On observe un changement d'attitude sur les activités d'écriture qui sont moins considérées comme des moyens de vérification des acquis des élèves et plus pour sensibiliser les élèves aux productions ultérieures ; l'écriture en classe sert à montrer l'utilité du texte et à conserver des traces.

Les dictées de texte sont moins fréquentes ; ceci au profit des dictées de mots.

- sur les difficultés des élèves:

Les élèves sont perçus comme plus indisciplinés , moins motivés.

La grande majorité des enseignants(deux tiers) pense en 1997 que les difficultés des élèves sont dues en premier lieu à des problèmes personnels, affectifs et relationnels; les problèmes médico-psycho-pédagogiques (intellectuel et sensoriel) sont plus souvent évoqués qu'en 1990.

Pour aider les élèves en difficulté, deux tiers des enseignants déclarent privilégier le soutien individuel mais celui-ci est moins fréquent qu'en 1990 au bénéfice du soutien hors classe ce qui est cohérent avec l'analyse qui est faite des causes. Enfin le soutien individuel est moins fréquent au CE1. Le regroupement des élèves faibles pour des activités spécifiques est pratiqué par moins de la moitié des enseignants.

Il est difficile à partir de cette comparaison de définir des évolutions claires. Cependant on note un problème récurrent: l'étude du code décroît du CP au CE1; le soutien individuel des élèves faibles décroît aussi entre le CP et le CE1, alors que toutes les évaluations montrent que l'apprentissage du code est loin d'être terminé en CE2. Les difficultés graves des élèves sont perçues comme ne relevant pas de l'école.

Les évaluations diagnostiques dans le second degré

Evaluations diagnostiques à l'entrée en sixième

Les compétences évaluées en lecture

Les protocoles, tels qu'ils apparaissent dans les cahiers d'élèves, comprennent des exercices regroupés en séquences : en moyenne deux ou trois séquences de lecture, et une de production écrite avec éventuellement des consignes pour améliorer le brouillon. Chaque séquence, indivisible, est conçue pour une durée approximative de quarante minutes, le cahier du professeur précisant la durée indicative proposée pour chaque exercice. Les exercices de lecture portent sur des textes ou des documents de nature et de longueur variées. On trouve des extraits discursifs (lettre, énoncé téléphonique, dialogue romanesque ou téléphonique, bulles dans des images, phrases,...), narratifs (extraits de romans ou de contes, courts récits,...), informatifs (articles de presse ou de revue, extraits de manuels,...), documentaires (fiche horaire SCNF, tableaux extraits de manuels ou de modes d'emploi, ...), iconographiques (images, photos, dessins, schémas...). Il arrive que les questions portent non sur un écrit lu par l'élève, mais sur un texte lu aux élèves par l'enseignant, afin d'évaluer seulement la compréhension d'un message en particulier quand ce message est constitué de consignes orales.

Un premier groupe de compétences est d'ordre linguistique ; elles relèvent, dans les protocoles, de la *maîtrise des outils de la langue*, expression qui montre bien que la langue est considérée comme un instrument qui permet de comprendre l'écrit. Le choix des connaissances à contrôler répond à la question suivante : à partir de quelles connaissances et savoir-faire linguistiques l'élève lecteur peut-il construire des informations pour comprendre un texte ou un document ? On mesure donc la maîtrise d'éléments de la langue comme la correspondance graphie/phonie, l'ordre alphabétique, le vocabulaire, les déterminants, les pronoms personnels, les groupes de mots désignant le temps, le lieu ou les personnages, les formes verbales, les marques d'accords, les mots de liaison logique et chronologique, les types et formes de phrases, les fonctions dans la phrase, la ponctuation, pour ne citer que les plus courants.

Comme il s'agit de savoir dans quelle mesure l'élève maîtrise assez pour avoir accès au sens, on lui demande soit de les insérer dans les blancs d'un texte lacunaire qu'il doit donc comprendre pour le compléter, soit de les en extraire pour répondre à une question sur le texte. Ces questions portent sur les aspects fondamentaux de la compréhension d'un écrit : son genre et sa fonction, son organisation spatiale, logique ou chronologique, le cadre spatio-temporel, l'identification des personnes et des personnages, le thème du texte et les idées qui le développent, etc.

Un second groupe de compétences porte sur les opérations que l'élève exerce sur le support écrit, texte ou tableau de données, pour le comprendre soit globalement soit partiellement : est-il capable d'explorer le texte pour en faire une lecture globale qui aille à l'essentiel ? Pour répondre à une question sur le texte, peut-il y prélever des informations qui y figurent explicitement ou bien, en reliant, en mémorisant et en recoupant les indices explicites qu'il relève, peut-il construire et synthétiser des informations, et éventuellement avoir accès à l'implicite ? Ces opérations sont généralement regroupées, dans les protocoles récents, sous l'intitulé *compréhension*.

Selon les protocoles, la part respective des savoir-faire linguistiques et des opérations de compréhension est plus ou moins étendue. Puisqu'on ne peut pas tout évaluer, la question qui se pose est la suivante : pour mesurer l'aptitude d'un élève entrant en sixième à lire un écrit c'est-à-dire à le comprendre, doit-on principalement mesurer la maîtrise qu'il a de la langue qui est comme le matériau de cet écrit, ou plutôt l'efficacité des activités qu'il met en œuvre pour construire et utiliser, à partir de la langue qu'il lit, des informations qui le mèneront au sens ?

Cette question suppose une définition de la lecture et une idée de ce que l'élève de sixième doit déjà maîtriser de cette capacité. Ainsi, les processus qu'on pourrait nommer neurophysiologiques de la lecture (opérations de perception, d'identification et de mémorisation des mots, activités de déchiffrement) sont censés maîtrisés à l'entrée en sixième donc rarement évalués, sauf dans des protocoles spécifiques portant sur les élèves en grande difficulté (1997). Dans les tout premiers protocoles de sixième, la maîtrise de la langue est largement explorée, les exercices de compréhension d'un écrit se limitant généralement à reconstituer la structure d'un texte et à répondre à des questions sur un texte. La lecture y est donc conçue, du moins au niveau d'un élève entrant en sixième, d'abord comme une démarche de compréhension de la langue, surtout dans les domaines du vocabulaire et de la grammaire et au niveau des mots, groupes de mots et phrases. Progressivement et surtout à partir de 1997, les activités de compréhension se développent et se spécifient, et parallèlement sur un texte et sur un tableau de données : compréhension globale, compréhension de la structure interne (appelée compréhension logique), prise d'indices explicites, construction et traitement de l'information, éventuellement communication de l'information. L'évaluation de la maîtrise de la langue s'en trouve limitée à quelques savoir-faire considérés comme essentiels dans toute activité de ce type. La lecture y est donc conçue d'abord comme une activité de construction du sens d'un texte ou d'un document, à partir des éléments de signification donnés par la langue et des informations construites sur ce texte ou ce document.

Conclusion.

Dans tous les protocoles de sixième, la lecture est conçue comme un processus cognitif propre à construire du sens ; mais ce sens se déplace progressivement, dans les protocoles, des éléments de signification élaborés à partir du fonctionnement de la langue aux informations construites sur la totalité de l'écrit. Cette évolution reflète les recherches sur l'activité du lecteur dans l'acte de lecture et correspond à la définition de cette compétence dans les programmes : «Savoir lire, c'est savoir accéder, à travers l'écrit, de manière autonome, à la pensée d'autrui.» (Complément aux *Programmes et Instructions de 1985* intitulé *Lire au collègue*).

D'autres aspects de cette activité du lecteur sont absentes des protocoles de sixième, parce que prématurées, mais présentes plus tard et, surtout, en début de seconde : comment par exemple, grâce à sa familiarité avec toutes sortes d'écrits, l'élève utilise-t-il ses attentes sur ce qu'il va lire pour orienter son analyse du texte, faire des hypothèses sur le sens, les vérifier ou les modifier au fil de sa lecture ? Comment passe-t-il de l'observation à l'interprétation ? Enfin, d'autres fonctions de la lecture sont absentes sans doute parce qu'elles sont difficilement évaluables, comme l'interaction entre la lecture et les intérêts, les émotions, l'imaginaire du lecteur. Le goût de la lecture, l'un des objectifs du collège, tient pourtant beaucoup à cela, car un élève qui aime lire s'attache plus aux héros et aux expériences qu'il rencontre dans un livre qu'il n'est heureux d'avoir construit des informations sur lui. C'est peut-être l'absence de ce dernier aspect de la lecture qui vaut parfois aux protocoles le qualificatif péjoratif de technicistes.

Les compétences évaluées en production écrite

Si «savoir écrire» consiste à élaborer par écrit une phrase ou un énoncé construits, dont la longueur dépasse celle du mot ou du groupe de mots, on y a très peu recours, dans les séquences de lecture, pour répondre à une question sur un texte ou sur un document. En effet, dans les exercices de lecture, les libellés sont faits pour écarter autant que possible ce qui pourrait gêner l'interprétation des réponses. On demande donc le plus souvent aux élèves de répondre par des chiffres, des signes, des flèches reliant plusieurs mots, des croix dans un tableau à compléter ou pour cocher leur réponse parmi plusieurs propositions (QCM), des soulignements ou des encadrements dans le texte, des mots ou des phrases à relever dans le texte et à recopier, ce qui est le format de réponse le plus courant. Quand les élèves doivent produire d'eux-mêmes une phrase ou un mot qui ne se trouvent ni dans le texte ni dans la question, ce qui est évalué est l'exactitude de la réponse, non d'autres aspects de la production écrite comme la syntaxe ou l'orthographe, qui sont alors légitimement considérés comme extérieurs à la compétence évaluée.

De la même manière, un exercice apparu en 1997 (*La vie du cerisier*, extrait d'un manuel de sciences) demandait à l'élève d'élaborer un texte pour expliquer un schéma. Mais cet exercice a été repris dans les protocoles suivants sans la production de cet énoncé explicatif. Un aspect de la production écrite qu'on pourrait appeler l'écriture fonctionnelle, destinée à rendre compte d'une autre activité qu'elle-même comme ici d'une lecture, n'apparaît donc pour ainsi dire pas dans les protocoles de sixième. Elle prendra un peu plus de place, mais jamais une place centrale, dans les protocoles de cinquième et de seconde.

Les compétences d'écriture sont donc évaluées principalement à partir de deux types de productions écrites complémentaires : l'une est un texte informatif brut que l'élève doit recopier en choisissant une mise en page qui en révèle la structure ; l'autre, qui a toujours été présent dans les protocoles, est un récit d'une vingtaine de lignes, que l'élève doit composer à partir d'une situation d'écriture précisée dans la consigne. Par ailleurs, la maîtrise des «outils de la langue» nécessaires à la production écrite est évaluée dans des exercices ponctuels.

La conception de l'écriture qui apparaît ici met l'accent sur ce qui, dans toute production écrite, visualise sur la page même la structure de la pensée et la hiérarchie des informations (paragraphe, retraits et alinéas, titres et sous-titres, ou tout autre moyen approprié) et sur l'écriture de fiction. Le premier exercice (analogue à *La vie du cerisier*) lie étroitement la lecture et l'écriture, données dans les programmes comme complémentaires. Le second, la production d'un récit, est le seul à demander à l'élève de mobiliser ses savoirs et ses savoir-faire pour construire et rédiger complètement un texte écrit. Si la copie lisible et sans erreurs d'un texte et l'élaboration d'un récit sont considérées comme devant être maîtrisées à l'entrée en sixième, la visualisation de la structure d'un texte, surtout informatif, ne l'est pas.

Ces productions ne sont pas évaluées globalement, mais à partir d'un nombre important d'items qui les évaluent chaque fois sous un angle particulier. Ces items sont classés selon une typologie qui s'est simplifiée avec le temps et qu'on retrouve, même sous des intitulés différents, dans les protocoles de seconde. Le texte produit par l'élève est-il pertinent, c'est-à-dire répond-il à la consigne et au genre demandé ? Le texte est-il cohérent, c'est-à-dire est-il complet, construit, sans ruptures ni ambiguïtés ? Le texte est-il lisible, c'est-à-dire respectueux des contraintes lexicales, grammaticales, graphiques et orthographiques de la langue ?

Le processus d'écriture n'est pas évalué. La tentative d'évaluer le retour sur son écrit par le biais du brouillon a montré que les élèves améliorent surtout des erreurs ponctuelles de langue.

Conclusion

La production écrite évaluée dans les protocoles de sixième est principalement une écriture de fiction, et précisément de narration, la plupart du temps indépendante des textes de lecture proposés ailleurs dans le protocole. L'écriture de commentaire ou d'explication, qui n'est pas demandée pour répondre par un énoncé plus ou moins long aux questions de compréhension, n'est pas évaluée. Le processus d'écriture, comme planification et régulation constante de ce qu'on écrit, n'est pas évalué non plus, parce qu'il est sans doute trop complexe pour donner lieu à des épreuves standardisées.

Choix des compétences et programmes

Programmes et instructions pour l'école élémentaire 1985.

Horaires et programmes d'enseignement de l'école primaire, arrêté du 25.01.2002.

Collège, Programmes et instructions, 1985, arrêté du 14.11.1985.

Programmes de 6^o, arrêté du 22.11.1996.

La présentation générale des protocoles rappelle toujours aux professeurs que les évaluations ne sont pas exhaustives et ne reprennent pas tous les éléments des programmes, mais seulement ceux qui semblent déterminants pour mesurer les compétences évaluées. Éléments qui, en outre, doivent pouvoir donner lieu à des épreuves standardisées, ce qui n'est pas le cas de tout ce que préconisent les programmes.

En lecture, les protocoles de sixième ont retenu des compétences communes aux programmes de 1985 destinés à l'école élémentaire (en particulier au cycle 3) et au collège (en particulier à la

sixième). Certaines y sont énumérées comme compétences transversales, communes à toutes les disciplines. Il s'agit en particulier des concepts fondamentaux d'espace et de temps, et du traitement de l'information. Les exercices des protocoles demandant de sélectionner des informations dans un texte ou un tableau de données, de reconstituer un texte ou de construire des informations sur des aspects essentiels d'un texte (temps et lieux, personnages, situations, etc.) relèvent de ces compétences transversales. D'autres relèvent plus spécifiquement de la compréhension d'un texte éventuellement littéraire lu silencieusement par l'élève, et sont reprises dans les protocoles : savoir exécuter une consigne (à partir également d'un message oral dans les protocoles) ou répondre à des questions sur le texte.

Ce que les programmes de 1985 pour la sixième apportent de nouveau par rapport au dernier cycle de l'école c'est, outre «la maîtrise de l'image», le choix d'œuvres de genres et d'époques divers propres à former une culture. Il est clair que les protocoles d'entrée en sixième visent à évaluer l'efficacité des opérations intellectuelles mobilisées lors de la lecture silencieuse et non la culture des élèves. Ils choisissent donc bien comme supports des textes de genres divers, mais plutôt récents que lointains, et plutôt brefs ou par extraits que longs et intégraux, avec une quasi absence du texte poétique comme support d'étude (il ne sert généralement qu'à la reconnaissance du genre).

Enfin, les protocoles ont retenu plus ou moins durablement l'évaluation d'éléments de la langue nécessaires à la compréhension, énumérés dans le Complément aux programmes de 1985 intitulé *Lire au collège* pour le cycle d'observation d'alors (sixième et cinquième) : l'élève devra apprendre à observer les éléments et la construction d'un texte, dont «les registres de langue, les marques de l'énonciation, le statut du narrateur, les repères spatiaux et temporels ; l'emploi des pronoms, les temps et les modes verbaux ; la présence de l'implicite, ... ». Les nouveaux Programmes de 6^o de 1996, en précisant les compétences acquises et à acquérir en lecture, distinguent clairement ce qui peut être considéré comme acquis à l'entrée en sixième et ce qui sera plus particulièrement développé au cours de l'année : «A l'entrée en sixième, les élèves devraient avoir atteint la compréhension immédiate des mots et des phrases (saisie du thème et du propos global d'un texte) et être entrés dans la compréhension logique (saisie des relations qui structurent un texte). Le but de la classe de sixième en ce domaine est que l'élève maîtrise la compréhension logique et reconnaisse la présence de l'implicite.»

En production écrite, les protocoles de sixième ont également retenu des exercices communs au cycle 3 de l'école et à la sixième, repris dans les programmes récents : structurer un texte en utilisant la présentation (paragraphe, graphies...), construire et rédiger un récit de fiction. Les programmes de 1996 pour le collège citent entre autres les compétences suivantes : assurer la lisibilité d'un texte (mise en page...; graphie claire et régulière ; maîtrise des différences de code ...; orthographe), et produire un récit qui suive l'ordre chronologique. Ils précisent ce qui doit être acquis à l'entrée en sixième (la production de phrases, l'articulation des idées, l'organisation des paragraphes) et ce qui revient à la sixième (...que les élèves soient capables de produire un texte narratif cohérent d'une page environ). Enfin, le souci de lier l'étude de la grammaire, du vocabulaire et de l'orthographe aux activités de lecture et d'écriture conduit à intituler le chapitre correspondant dans le programme « outils de la langue pour la lecture, l'écriture et la pratique de l'oral, » où l'on reconnaît la logique des protocoles récents.

Cependant, l'éventail des exercices écrits proposés par ces programmes est plus vaste que les sujets retenus dans les protocoles. Aussi bien pour l'école primaire que pour la sixième, on recommandait aussi de nombreux exercices d'imitation, de reformulation ou de réécriture à partir de textes, ou encore des exercices d'écriture fonctionnelle, comme présenter son point de vue sur un événement, rédiger un compte rendu de visite, d'expérience ou de lecture, sans compter l'écriture d'invention comme celle de textes poétiques.

Il est certain que, pour faciliter l'évaluation, les protocoles ont choisi de ramener la variété des exercices écrits proposés par les programmes de l'école et du collège à la priorité affichée dans les deux cas, composer un récit, en éliminant les autres types d'écriture, plus fonctionnels ou plus étroitement liés à la lecture. Mais certaines des fonctions sociales de l'écriture se trouvent ainsi négligées dans les évaluations : la formulation ou reformulation pour soi ou pour autrui, la description et la mémorisation éventuellement synthétiques, d'idées, d'opinions, d'impressions, de documents, de situations. En outre, il serait utile d'indiquer explicitement aux professeurs que tout ce qui relève des procédures et des démarches mises en œuvre dans la production écrite pour traiter simultanément des contraintes de natures diverses en vue de l'effet final à obtenir ne peuvent être évaluées, et ne le sont pas. Les évaluations ne peuvent donc fournir d'indications sur la maîtrise de la capacité elle-même,

mais sur certaines des composantes de cette capacité, et dans des situations qui ne sont pas les situations d'apprentissage.

Parmi les éléments de la langue, qu'on ne peut pas évaluer en totalité, on a choisi les priorités, dont certaines figuraient déjà dans les activités de lecture et d'écriture, et toujours, comme cela est recommandé dans les programmes, en relation avec le sens :

- dans la maîtrise du vocabulaire, "distinguer, grâce au contexte, différents sens d'un mot, des homonymes, des mots de sens proche ou de sens contraire", "l'enrichissement par familles de mots", plus rarement le sens propre et le sens figuré, évalué en 1989. Les niveaux de langue, évalués également en 1989, ont disparu ensuite.

- En grammaire, la connaissance de la phrase verbale (non de la phrase nominale également au programme), des types et des formes de phrase (surtout impérative, interrogative et négative, les plus fréquentes, plus rarement déclarative, exclamative et affirmative). La phrase passive n'est pas évaluée en tant que telle, ni la transformation passive.

"Analyser les structures de quelques phrases complexes" se fait parfois, par exemple quand l'élève doit transformer une phrase. Dans la connaissance des classes de mots et des éléments du groupe nominal, priorité est donnée aux déterminants et aux pronoms (non à la proposition relative et à l'apposition qui ne sont pas évaluées en tant que telles). La syntaxe des accords est bien représentée, même l'accord du participe passé employé avec les auxiliaires être et avoir, pourtant en cours d'acquisition à l'issue du cycle 3. L'identification des modes et des temps usuels des verbes précisés dans les programmes apparaît souvent, mais pas toujours.

Conclusion.

Dans la logique de leur réflexion sur la nature de l'activité de lecture et sur les modalités de son évaluation, les groupes d'élaboration des protocoles de sixième ont repris, dans les programmes de l'école élémentaire, ce qui relevait de la sélection et du traitement de l'information à partir de supports diversifiés. En cela, ils sont fidèles à ce qui, dans les programmes de sixième, poursuit le travail de l'école. La construction d'informations, en particulier quand elle met en jeu l'implicite, *a fortiori* l'interprétation, sont du ressort de la sixième et même du collège. La lecture de l'image, la lecture d'un texte long ou intégral, absentes des protocoles de sixième mais non des programmes, seront possibles plus tard, par exemple dans les protocoles de seconde. L'écriture poétique, elle, est très peu représentée dans l'ensemble des protocoles.

C'est en écriture que les épreuves des protocoles réduisent le plus l'éventail des exercices proposés dans les programmes en choisissant cependant l'exercice écrit le plus représentatif du cycle 3 et de la sixième.

Il est évident enfin que tout ce qui relève de l'oral, de la lecture personnelle ou de l'autonomie dans le travail relève de l'évaluation que l'enseignant élabore personnellement pour sa classe. Un dispositif d'évaluation de l'oral figure d'ailleurs dans le tome 3 d'*Aide à l'évaluation* pour la seconde.

Choix des compétences et pratiques des élèves.

Que les compétences soient choisies en fonction des programmes permet de mesurer le degré d'atteinte des objectifs fixés par ces programmes, ce qui est conforme aux finalités du dispositif d'évaluation. Mais on pourrait également s'inspirer des pratiques des élèves, et éventuellement des pratiques langagières des plus faibles, pour mesurer les difficultés que leur langage oral ou écrit révèle.

Par exemple, en sixième et encore en seconde, les élèves maîtrisent mal, en lecture mais surtout en écriture, la proposition relative, en particulier introduite par *que* et *dont*. Ils maîtrisent mal la phrase complexe, sa structure et la hiérarchie des informations qu'elle organise, surtout dans l'expression du temps et de la causalité. Ils sont facilement perdus quand s'accumulent des négations, que ce soit des structures négatives ou des termes de sens négatif. Les plus faibles ne saisissent pas toujours l'emploi du passif. Enfin, ils ne maîtrisent pas la construction des verbes, surtout quand il s'agit de verbes à constructions multiples (apprendre quelque chose de / à quelqu'un, dire quelque chose à quelqu'un sur quelqu'un, etc.). La proposition relative, la négation et le passif figuraient d'ailleurs dans d'anciens protocoles (1980 et 1982 par exemple).

Enfin, pour ce qui relève de la lecture, ce sont moins les opérations de prélèvement ou de

construction des informations qui gênent les élèves au fur et à mesure qu'ils avancent dans leurs études. C'est l'univers culturel dont dépendent les textes qui leur manquent et qui est la source de leurs erreurs. Il en est de même en écriture, où l'absence de culture générale conduit souvent à utiliser un vocabulaire constamment impropre, donc inintelligible, dans une phrase pourtant bien construite. Mais peut-on évaluer, par des évaluations standardisées, cette carence-là ?

L'élaboration des protocoles

Elaborer un protocole, c'est classer et organiser entre elles les compétences à évaluer, et construire des exercices pour les évaluer.

Connaître la façon dont ces compétences sont reliées entre elles est nécessaire pour comprendre la finalité et parfois le codage des exercices. Cela revient à comprendre la logique du processus même d'évaluation. Car un exercice, en soi, est neutre : il propose à l'élève une tâche à effectuer sur un support, avec un type de réponse (QCM, relevé, etc.). Pour interpréter cette tâche et en rapporter les résultats à un savoir-faire global comme la lecture ou l'écriture, il faut situer l'exercice dans une sorte de typologie des compétences : de quelle compétence l'exercice dépend-il ? De quel savoir-faire plus général cette compétence dépend-elle à son tour ? L'organisation des compétences est donc présentée dans un tableau figurant chaque année au début du livret destiné aux professeurs.

Des regroupements d'items effectués par le logiciel Casimir pour dégager des scores partiels ont été présentés dans le protocole de 1998. Ils constituent trois catégories : compréhension, connaissance et maîtrise du code (ou de la langue), production de textes. Ces trois catégories seront reprises dans les tableaux des protocoles suivants sous le nom de capacités, dont dépendent les compétences évaluées et les exercices proposés pour les évaluer.

Ces tableaux traduisent la logique de l'évaluation des compétences. Ils se sont fortement modifiés avec le temps, soit pour s'adapter aux changements dans les programmes de sixième (protocole de 1997), soit par suite d'une réflexion toujours plus approfondie sur l'évaluation, et sur les capacités de lecture et d'écriture. Cette évolution a eu deux conséquences. L'analyse des tableaux successifs est un excellent outil de formation à l'évaluation de la lecture et de l'écriture, ce qui était l'une des finalités du dispositif d'évaluation lui-même. Mais, même si l'on retrouve d'un protocole à l'autre des exercices souvent proches et portant sur des activités formulées d'une manière équivalente, leur finalité n'est pas toujours identique. La modification des tableaux a entraîné l'apparition ou la disparition de certains exercices c'est-à-dire de certaines compétences, tandis que d'autres exercices étaient redistribués sous d'autres compétences. Certains exercices passent par exemple, sans se modifier, de l'évaluation de l'écriture à celle de la lecture, ou de l'évaluation de la compréhension logique d'un texte à celle de la maîtrise de la langue (cf. annexe 3). Si bien qu'il est très risqué de comparer entre eux les résultats de protocoles qui ne relèvent pas de la même logique, et donc d'interpréter les résultats des élèves sur une longue période.

D'autre part, les modifications introduites dans les exercices ne permettent pas non plus de mesurer rigoureusement l'évolution du niveau des élèves sur plusieurs années. Soit les supports changent pour un même type d'exercice, ce qui entraîne parfois des résultats imprévus. Soit les codages et les items sont modifiés pour un exercice et un support identiques. Il arrive qu'un exercice apparemment semblable d'un protocole à l'autre adopte chaque fois une logique en réalité différente qui rend les scores difficilement comparables. Pour les professeurs et les formateurs, cette variété est un avantage, qui répond à la finalité affichée de l'évaluation. Pour ceux qui voudraient suivre l'évolution des acquis des élèves, c'est un handicap ; on ne peut noter, dans le meilleur des cas, que des tendances ou éventuellement des constantes (voir des exemples en annexe 4).

Ce que savent les élèves à l'entrée en sixième

Une étude comparative des acquis des élèves à l'entrée en sixième de 1980 à 1993 a été effectuée à partir d'exercices communs à ces deux protocoles et absolument identiques. Les résultats montrent une progression en lecture mais une régression en orthographe. En lecture, les taux de réussite ont augmenté de plus de onze points, grâce surtout à la compréhension explicite. L'emploi des déterminants est également toujours mieux maîtrisé en 1993, sauf pour le démonstratif dont on sait

que l'emploi reste encore problématique en cinquième et au-delà. Sur les items mal réussis par les élèves en 1980, on a dix points de plus en 1993. Les résultats de la dictée, en revanche, sont inférieurs, bien que les mots les mieux et les moins bien orthographiés soient sensiblement les mêmes en 1980 et en 1993. La progression globale calculée par la DEP est de 1%, donc le niveau est à peu près identique à treize ans d'intervalle. L'étude montre également que la dispersion des résultats (écarts entre les scores des plus faibles et des plus forts) est inchangée. Les plus forts ont en français des résultats deux fois plus élevés que les plus faibles.

Une telle analyse serait très utile sur tous les exercices qui se retrouvent d'un protocole à l'autre, surtout à partir de 1997 où les protocoles sont construits selon la même logique. Cependant, on peut déjà regarder quelles sont les compétences qui arrivent toujours, malgré la variation des scores, au même niveau de réussite dans tous les protocoles, et en déduire les acquis des élèves. Ce classement repose sur les protocoles de 1989 à 2002, avec une attention plus particulière aux protocoles de 1997 à 2002, qui se reprennent en tout ou partie les uns les autres.

En 1992, une étude a été menée sur les résultats aux évaluations depuis 1989 : selon le degré de réussite, elle classe les compétences en trois groupes et calcule les écarts entre les scores moyens des 20% d'élèves les plus faibles et le score moyen général. Les résultats de cette étude sont utilisés dans le classement qui suit.

Enfin, il est important de préciser la nature de l'activité à partir de laquelle on déduit le savoir-faire acquis, la compétence des élèves. Car il faut se souvenir que cette tâche n'est qu'une des possibilités de mesurer la compétence qu'elle représente ; sur une activité différente, les acquis des élèves, pour mesurer la même capacité, seraient peut-être différents.

Lecture

Compétences globalement maîtrisées à l'entrée en sixième

Comprendre l'essentiel d'un texte et l'organisation d'un tableau de données en y prélevant les informations explicites demandées

Huit à neuf élèves sur dix y parviennent pour le texte, surtout quand il s'agit d'un conte, genre bien connu des élèves, environ neuf sur dix pour un tableau de données ; et les scores sont équivalents quand l'exercice est repris d'une année sur l'autre.

C'était déjà le cas de 1989 à 1992, ce qui faisait alors classer cette compétence dans le groupe de celles qui sont les mieux réussies, même par les élèves les plus faibles.

Quand il faut rechercher dans un texte une information pertinente et la sélectionner parmi les informations explicites, les résultats peuvent baisser selon la nature de l'indice à trouver (le lieu ou le temps par exemple). C'est aussi le cas quand les élèves sont surpris par un tableau de type nouveau. Mais ces baisses sont rares, et les scores atteignent encore environ 70%.

Enfin, huit à neuf élèves sur dix comprennent un message oral en mémorisant et en exécutant les consignes qu'il énonce explicitement.

Identifier le genre d'un texte lu ou entendu

De sept à neuf élèves sur dix y parviennent selon les genres à identifier. Dans le genre narratif, le conte est reconnu par presque tous les élèves ; les réponses partielles ou certaines questions montrent que des élèves qui ne donnent pas la réponse attendue ont pourtant une idée de ce qu'est un récit et reconnaissent les situations initiale et finale constitutives de ce genre. En revanche, tous montrent une certaine difficulté à identifier le genre documentaire.

En 1992, 60% des élèves savaient identifier différents types d'écrits. En 1997, où la question est posée pour la première fois, le genre est identifié, selon les textes, par 60% ou 70% des élèves.

Quand la question est posée, la fonction d'un texte (le but qu'il vise) est un peu moins bien identifiée, par plus de six élèves sur dix en moyenne, et par environ 70% des élèves en 1997. Mais c'est une compétence délicate et en cours d'acquisition en sixième.

Parmi les savoir-faire linguistiques qui aident à comprendre un écrit, environ 8 élèves sur 10 réussissent à :

* Reconnaître les types et formes de phrases, donc savoir si le personnage conseille, interroge, déclare, affirme ou nie, etc., avec des différences constantes selon les phrases depuis 1989 : la phrase interrogative est toujours mieux reconnue que les phrases impérative, déclarative et négative (environ six élèves sur dix).

* Utiliser l'ordre alphabétique quand les termes à classer ne concernent que les initiales. Mais de six à sept élèves sur dix seulement maîtrisent les exercices quand les termes portent sur plusieurs lettres après l'initiale : la compétence est donc à poursuivre et à consolider en sixième.

De 1989 à 1992, cette compétence était classée parmi les compétences moyennement réussies (70% environ) par tous les élèves, et très mal réussie par les élèves faibles. Les scores sont environ de 60% en 1995 et 1997 dès qu'on dépasse les deux premières lettres des mots à classer.

Compétences installées mais encore fragiles à l'entrée en sixième

Saisir l'essentiel d'un texte lu (récit avec dialogue ou texte informatif) quand il s'agit de reconnaître, parmi plusieurs propositions, le titre qui convient à la totalité du texte. Procéder donc à une lecture globale en synthétisant les informations.

Cette compétence est fragile, parce qu'elle dépend encore trop de la nature du texte ou de sa présentation (quand les élèves ne sont pas égarés par des termes du libellé des questions). L'exercice peut être, selon le texte, réussi par huit élèves sur dix, ou seulement par six à plus de sept élèves sur dix. Les élèves peuvent avoir généralement compris le sujet du texte sans choisir toujours le titre qui lui convient. De 10 à 15% des élèves font encore des lectures partielles. Et, quand il est arrivé que le texte soit présenté en paragraphes distincts, presque sept élèves sur dix ont bien identifié l'idée principale de chaque paragraphe alors qu'ils ne sont qu'environ 40% à avoir choisi le titre convenable, avec un nombre important de réponses erronées : apparemment il leur était difficile de passer de la multiplicité des lectures partielles à une lecture synthétique et globale.

Utiliser des savoirs grammaticaux pour aller au sens, précisément les marques d'accord (adjectifs et participes passés) pour identifier les personnes dans un message (une lettre), ils sont six à sept élèves sur dix à y réussir, parfois moins, avec un taux important de réponses erronées (souvent un tiers, parfois plus de 40 ou de 60%).

Il suffit en effet de quelques difficultés du texte (un adjectif isolé en tête de phrase ou ayant la même forme aux deux genres comme « méconnaissables ») pour que le score de réussite tombe à moins de 50%. Peut-être ces élèves ont-ils été inattentifs ; peut-être aussi n'ont-ils pas une maîtrise assez mécanique et intériorisée de la langue pour devenir attentifs à ces finesses du texte. Il est significatif que les élèves soient généralement plus nombreux à identifier celui qui envoie le message et celui qui le reçoit, que les personnages dont on parle dans le message : pour ces derniers en effet, il faut recouper un plus grand nombre d'indices du texte.

En 1992, cette compétence était classée parmi les compétences moyennement réussies (70% environ), et très difficile pour les élèves les plus faibles (de 42 à 82 points d'écart avec les autres).

Utiliser le contexte pour trouver le sens des mots en particulier quand ils sont polysémiques

* Six à sept élèves sur dix en moyenne savent utiliser le contexte pour trouver le sens d'un mot inconnu : cependant, malgré de nettes réussites de 1998 à 2001, les scores parfois faibles sur cet item montrent la fragilité de cette compétence qui était classée, en 1992, parmi les compétences médiocrement réussies (60%) par tous les élèves et très mal réussie par les élèves les plus faibles (écart de 42 à 82 points avec les autres).

* Environ sept élèves sur dix, en s'aidant du contexte, trouvent le sens d'un mot polysémique, ou identifient dans une liste le mot dont les différents sens conviennent aux divers cas d'un texte à compléter (milieu, régime, point, par exemple) ; et ils sont presque huit sur dix à reconnaître la définition qui convient à chacun de ces sens. C'était déjà le cas en 1995 (82% de réussites). Mais l'exercice dépend encore beaucoup du support. Quand le texte support et les définitions ne comportent pas de mots qu'on puisse aisément rapprocher, même s'ils sont encore sept sur dix à réussir l'exercice, les réponses sont beaucoup plus dispersées, signalant l'embarras des élèves : absence de réponse ou réponses erronées, mais aussi réponses incomplètes ou partiellement exactes.

Trouver dans un texte le synonyme d'un mot, même quand le sens de ce mot a été reconnu par la grande majorité des élèves, est plus difficile encore : seul un peu plus d'un tiers des élèves y

parvient, avec un nombre important de non réponses (plus de 20%). En 1995, les scores de réussite sont de 56% sur cet item. Et, en 1992, la maîtrise de la polysémie est classée parmi les compétences moyennement réussies (de 70 à 80%) et diversement selon le degré de fréquence des mots.

Enfin, quand la question est posée, moins d'un élève sur deux trouve le mot de base sur lequel sont formés les autres mots de la même famille : les scores sont équivalents en 1989 et en 2002. Les élèves ont donc dans l'ensemble de bonnes compétences lexicales ; mais elles sont très variables selon les situations et les exigences, donc à poursuivre et à consolider.

Maîtriser la correspondance entre graphie et phonie

Les résultats sont différents selon ce qui est évalué ; peu d'exercices sont proposés et peu sont repris d'un protocole à l'autre. En 1989, un peu plus d'un élève sur deux reconnaissait les différentes graphies du son /s/, mais plus de six sur dix celles du son /è/ ; en 1995, ils étaient environ sept sur dix à distinguer les sons /p/ et /b/, /f/ et /v/. Cela signifie que de trois à cinq élèves sur dix environ ne maîtrisaient pas ces phonèmes essentiels pour la lecture.

Quand ils doivent déchiffrer en identifiant rapidement les mots (exercices de 1999 et 2000), 42% des élèves ont montré la rapidité attendue, mais plus de huit sur dix l'ont fait avec le degré d'exactitude attendu.

Compétences en cours d'acquisition

La construction et la représentation du temps

Les élèves qui entrent en sixième ont une représentation du temps, comme le montrent des réussites ponctuelles dans certains exercices (prise en compte de connecteurs temporels comme alors que / avant que, relevés de termes circonstanciels, utilisation d'indices temporels en production écrite, par exemple). Mais leur réussite est très variable quand ils doivent justement repérer les indices de lieu et de temps dans un texte (55% de réussite en 1991, 60 à 70% en 1995, par exemple).

Quand il s'agit de comprendre la succession des faits qui sont évoqués dans un texte en les situant sur un calendrier, les élèves y parviennent généralement, mais plus ou moins bien selon le texte : les scores peuvent aller de 80 à 90% de réussite à 50% selon l'exercice. En 1992, retrouver la chronologie était classé parmi les compétences les moins bien réussies par l'ensemble des élèves (40% à 50%).

Quand enfin il s'agit de reconstruire la succession des faits sans le support concret du calendrier et en se fondant sur les seules indications du texte, ils ne sont plus alors qu'environ un sur deux à y parvenir. Presque un tiers des élèves confond l'ordre des événements et l'ordre du récit, en reproduisant l'ordre d'apparition des actions dans le texte. L'écart entre ZEP et non ZEP est ici important : de 15 points environ en 1999 et 2002. Ils semblent avoir du mal à isoler et à relier les indices temporels du texte pour réordonner les faits qui y sont évoqués. Surtout, il leur arrive de remplacer, par un ordre chronologique plausible pour eux, l'ordre chronologique du texte. Si lire, c'est en premier lieu se soumettre au texte pour le comprendre, ils n'ont pas encore cette attitude quand il leur faut reconstituer une temporalité. Le rapport des non-réponses (qui varient de 0,9 à 4,4 en 2002) et des réponses erronées (presque 30%) montre que les élèves, en se trompant, n'ont sans doute pas eu conscience de la difficulté et ont cru répondre correctement, ce qui met en évidence l'absence de maîtrise.

La construction du temps est donc en cours d'acquisition à l'entrée en sixième, et la maîtrise très incomplète de la conjugaison le confirme : un bon tiers des élèves identifie mal les formes verbales du passé à l'indicatif (imparfait, passé défini et passé composé), et moins d'un élève sur deux est capable d'analyser les formes verbales, compétences déjà classées, en 1992, parmi celles qui sont les moins bien réussies (40 à 50%).

La cohérence textuelle

Il s'agit principalement de deux compétences :

- saisir ou retrouver l'enchaînement logique et chronologique d'un texte (et la difficulté de reconstruire le temps en fait partie) ;
- comprendre la chaîne référentielle, c'est-à-dire se repérer parmi les pronoms et les groupes lexicaux qui reprennent différemment, de phrase en phrase, les éléments du texte (personnes, personnages, objets, événements, lieux, etc.).

Quand les élèves doivent rétablir l'ordre d'un texte proposé en phrases désordonnées, ils ont une activité très complexe à exercer : il leur faut établir et mémoriser le sens de chaque phrase isolée ;

y chercher des indices linguistiques pertinents et de nature diverse (déterminants, verbes, groupes lexicaux, connecteurs logiques et temporels, entre autres) et capitaliser ces informations ; relier ces indices et recouper ces informations pour rétablir la cohérence perdue. Environ un élève sur deux y parvient et les scores changent peu d'une année sur l'autre. C'est pour cette compétence que l'écart est souvent le plus grand entre les ZEP et les non ZEP : 20,4 points en 1999, 19 points en 2001 ; il s'atténue en 2002 (11,9 points). Elle était classée, en 1992, parmi les compétences les moins bien réussies avec l'écart de points le plus important pour les élèves les plus faibles (de 42 à 82 points).

Quand le traitement des questions le permet, on voit cependant que deux-tiers des élèves peuvent rétablir partiellement l'ordre logique du texte, et plus de la moitié y parvient en ZEP. En 2001, plus de 20% des élèves n'ont fait qu'une erreur (omission ou inversion) dans l'ordre à rétablir ; et plus de neuf élèves sur dix ont trouvé la phrase initiale et un enchaînement exact. Les élèves ne sont donc pas totalement démunis dans ce domaine en entrant en sixième ; ils ont des possibilités qui se développeront au collège.

La chaîne référentielle (second exercice) est généralement reconnue par environ six élèves sur dix selon les années. C'est une compétence très fragile, comme le montrent les scores de réussite parfois bas et généralement très variables d'une année sur l'autre. En outre, les difficultés en production écrite confirment généralement les scores en lecture : 61% de réussites en 98, mais 45% en 2001 par exemple. Enfin, les écarts importants (environ 17 points en 2002) entre élèves en ZEP et hors ZEP confirment la difficulté de cette compétence. En 1992, elle était classée parmi les compétences moyennement réussies (avec un écart de 23 à 31 points entre les scores des élèves faibles et celui des autres élèves).

Cependant les élèves ne sont pas non plus totalement démunis, puisqu'ils sont en moyenne de 15 à 20% à n'avoir fait qu'une erreur dans leur réponse. Mais ils restent trop sensibles aux "pièges" des textes : par exemple quand le hérisson est repris par la « petite bête » elle-même reprise par « elle », ou encore quand le même personnage est désigné par « vous » et « moi » selon qu'on est en récit ou en dialogue.

Là encore, ils ne peuvent s'appuyer en toute sécurité sur certains éléments linguistiques indispensables à cette compétence mais dont la maîtrise complète se développe au cours du collège : les pronoms personnels et les déterminants (articles, adjectifs démonstratifs, possessif et indéfinis). Les pronoms personnels sont assez bien maîtrisés ; mais les réussites très inégales et très variables aux exercices qui portent sur les déterminants montrent que le système global des déterminants n'est pas installé. Certes les habitudes langagières socio-culturelles interfèrent avec les connaissances grammaticales et les élèves savent majoritairement employer les articles défini et indéfini, ainsi que les adjectifs possessifs. Mais l'usage par exemple de l'indéfini (67,8% de réussites en 2001, 55,5% en 2002), et secondairement du démonstratif (seul un bon tiers des élèves l'utilise de manière adéquate) semble encore mal assuré. Or le démonstratif, par exemple, joue souvent un rôle important dans la chaîne référentielle. Si les scores de réussite ont remonté à 70% environ en 2001 et 2002, ils sont néanmoins inférieurs à ceux de 1992 à 1997 (autour de 80%).

La construction et le traitement des informations

Il ne s'agit plus dans ce cas d'explorer le texte pour y prélever une information qui y figure explicitement. Il s'agit d'utiliser le document écrit pour construire une information éventuellement implicite, ou pour valider une affirmation : en s'orientant par la consigne ou la question, surtout si c'est une question ouverte (Pourquoi... ? Qu'espèrent...?) il faut sélectionner et mémoriser les indices, les recouper, exécuter sur ces indices les opérations nécessaires pour en déduire les informations demandées. Quand la réponse est proche d'un prélèvement (chasse-neige dans un texte permettant de déduire qu'il s'agit de l'hiver), les élèves n'ont pas de difficulté. Mais quand il s'agit d'opérations plus complexes, les élèves en ont, aussi bien pour un tableau de données que pour un texte, avec des différences nettes dans les scores de réussites : à peu près six élèves sur dix en moyenne réussissent à sélectionner des informations dans un tableau de données alors qu'ils sont 80 à plus de 90% à y prélever sans faute des informations explicites. Quand il faut croiser et recouper les informations, les scores de réussite peuvent être plus bas : 58,7% en 1995 sur un tableau de données, moins de 50% en 1998 sur un texte informatif, par exemple. En 1995, 55% des élèves dégagent du texte une information implicite, compétence classée, en 1992, parmi les moins bien réussies (40 à 50% de réussite).

Quand on leur demande de dire si telle affirmation sur le contenu du texte, mais qui ne figure pas telle quelle dans le texte, est vraie ou fausse, la disparité des scores montre à quel point ils restent

dépendants de leur attention à la langue (sens des mots, formes verbales, référent de déterminants, ...) pour construire la réponse exacte : selon les cas, ils sont huit sur dix ou un sur deux à réussir. L'écart entre le nombre des non-réponses (généralement inférieur à 5% pour ces items) et celui des réponses erronées (de 10 à 45%) où les élèves ont cru bien répondre, confirme leurs difficultés. Il ne faut cependant pas exclure que les élèves, quand ils doivent choisir entre deux possibilités, jouent la réponse "à pile ou face"...

La justification et la formulation des réponses

Certains exercices cherchent à évaluer si l'élève a conscience des procédés qu'il met en œuvre pour lire, s'il peut justifier sa réponse en ayant recours à des citations.

D'une manière générale, les scores chutent quand les élèves doivent indiquer en les recopiant les mots ou expressions qui les ont conduits à donner une réponse, et cela même quand la réponse donnée montre que la compétence évaluée est acquise. Selon le support ou l'exercice, la chute est plus ou moins sévère mais elle est constante. De 1995 à 2002 quand les questions le demandent, les scores peuvent baisser de trente à quarante points entre la réponse donnée et sa justification. On trouve couramment le cas où six élèves sur dix environ donnent une réponse juste mais deux sur dix seulement une justification exacte, avec souvent un nombre important d'élèves qui ne fournissent aucune justification.

Formuler la réponse au lieu de la reconnaître parmi plusieurs propositions (QCM), soit en trouvant le mot juste soit en rédigeant une phrase, accentue visiblement les difficultés des élèves. S'informer leur est plus facile qu'informer et communiquer.

Le cas est courant quand il s'agit de construire une information sur un texte et de la transmettre en produisant un écrit pour répondre à une question ouverte. Sans doute les deux difficultés, construire une information et rédiger la réponse, se cumulent-elles. C'est aussi cette difficulté qui creuse l'écart entre les ZEP et les non ZEP : il est environ de 15 points en 1999 au bénéfice de ces derniers.

Mais le simple fait de devoir nommer pose déjà problème. Ainsi, la plupart des élèves qui entrent en sixième savent reconnaître le genre d'un texte ; cependant, ils sont un peu moins nombreux à réussir quand ils doivent nommer le genre d'extraits différents au lieu de le choisir parmi plusieurs propositions (QCM). On peut penser également que la même difficulté est présente quand on leur demande d'analyser des formes verbales. D'une manière générale, l'analyse des formes verbales, qui demande de nommer et non seulement de reconnaître, obtient des scores beaucoup plus faibles (aux environs de 40% en 1998) que la seule reconnaissance des formes dans un texte (autour de 60% en 1998).

Conclusion

Les élèves qui entrent en sixième sont des enfants en cours de formation. Les compétences évaluées dans les protocoles se fondent sur des savoirs mais supposent aussi des démarches de recherche et des raisonnements qui sont eux-mêmes en cours d'acquisition, sans négliger la concentration que demandent certains exercices à plusieurs items ou à plusieurs textes supports (comme les exercices de vocabulaire). Une erreur ou une omission peuvent altérer des scores de réussite par simple manque de concentration, et non par manque de capacité.

Il semble que les élèves soient capables de concentration, d'attention, de mémorisation quand ils doivent, en explorant un texte ou un tableau de données, isoler les informations explicites qui leur sont demandées et qui sont importantes pour la compréhension du texte ou du tableau. Ils montrent les mêmes compétences quand il s'agit d'un message oral et d'un texte qui leur est lu. Ils maîtrisent certains outils nécessaires à la compréhension, comme la reconnaissance des genres, et, éventuellement, la démarche qui permet d'utiliser le contexte pour donner un sens à des mots inconnus ou, ce qui leur est plus difficile, polysémiques.

Quand il s'agit de synthétiser les informations pour parvenir à une lecture globale qui va à l'essentiel d'un texte, à son idée principale, ils y parviennent généralement, mais sont encore trop dépendants des textes ou de leur présentation. L'attention fine à la langue, dans la mesure où elle pourrait les aider à identifier les personnes ou objets du discours par exemple, est encore fragile.

La capitalisation des informations, leur construction et leur traitement par sélection et recoupements, leur transmission par une désignation ou un énoncé écrit sont en revanche difficiles et

en cours d'acquisition. Enfin, certains résultats invitent à se demander si les élèves savent vérifier leurs réponses, ne serait-ce qu'en les confrontant à la consigne.

Quant à la hiérarchisation des informations, par exemple pour dégager l'organisation d'un texte informatif, elle est en cours d'acquisition en sixième (voir ci-dessous, en production écrite). Il est normal que les élèves de fin de CM2 la maîtrisent mal et ne parviennent à la mettre en lumière par la mise en page que très inégalement, selon la nature et la présentation du texte. Alors qu'un élève sur deux y parvient en 1998 pour un texte dont ils sont six sur dix à comprendre la structure et moins d'un tiers à ne pas la comprendre du tout, on constate l'inverse en 1999-2000 sur un texte différent : moins d'un tiers des élèves comprend la structure du texte et la met en lumière par la mise en page.

Production écrite

Rappelons que des exercices spécifiques évaluent pour la lecture puis pour l'écriture les mêmes savoirs linguistiques (types de phrases, accords lexicaux et grammaticaux, emploi des déterminants et des pronoms, orthographe, vocabulaire,...) : les élèves peuvent-ils les utiliser pour comprendre la pensée d'un autre, pour exprimer leur propre pensée? Les résultats comparés mettent en évidence les difficultés spécifiques des élèves en production écrite. En effet, la plupart du temps, les scores sont moins bons en écriture qu'en lecture. C'est que la tâche demandée n'est pas de même nature chaque fois. Dans les exercices de lecture, les élèves doivent le plus souvent reconnaître ou relever les faits de langue, par QCM, soulignements ou encadrements. Dans les exercices d'écriture, ils doivent trouver ou inventer la réponse qui convient au sens du texte-support de l'exercice ; ils sont moins nombreux à y réussir.

D'autre part, l'emploi pertinent des ressources de la langue est mesuré par ces exercices spécifiques, mais aussi dans la rédaction d'un récit. Les deux types d'activité ne demandent pas la même concentration ni les mêmes démarches. La comparaison des résultats dans les deux cas est souvent éclairante sur la solidité ou la fragilité des compétences observées.

Graphie, segmentation du discours et mise en page d'un texte

Les élèves qui entrent en sixième ont généralement une graphie lisible : ils sont environ huit sur dix à former correctement les lettres et à les attacher pour former des mots isolés dans la phrase.

Quand la segmentation du discours leur pose des problèmes, ce n'est donc pas au niveau du mot : c'est au niveau du texte, en particulier informatif, quand il s'agit d'en analyser la structure pour la mettre en évidence par la mise en page et les paragraphes ; ou au niveau de la phrase, quand il s'agit de la délimiter par la majuscule et la ponctuation. En 1992, faire des paragraphes et ponctuer faisaient déjà partie des compétences les moins réussies (40 à 50% de réussites). Ensuite et jusqu'en 2002, cinq à sept élèves sur dix, selon les textes, reconnaissent et distinguent dans la page les parties du texte et les éléments de chaque partie (en particulier les énumérations). C'est là une compétence en cours d'acquisition.

La majuscule et le point final qui, dans le code écrit, limitent et visualisent la phrase, ne sont pas encore senties par la majorité des élèves comme constitutives de la phrase. Un bon tiers des élèves souligne ou relève des phrases sans se sentir obligé de noter la majuscule ou le point d'interrogation. Dans les récits, ces signes ne sont respectés que par un élève sur deux sans changement notable de 1998 à 2001, avec de 15% à 20% des élèves, selon les années, qui ne les utilisent pas. Cependant un progrès constant de 1998 à 2001 fait passer le nombre des élèves qui acquièrent, même encore approximativement, cette compétence d'un peu plus de 22% en 1998 à presque 30% en 2001. En outre, quand ils rédigent un texte, les élèves sont plus nombreux à ne pas négliger la ponctuation que lors d'un exercice isolé dont ce n'est pas le premier objectif. Cela signifie peut-être que la ponctuation ne paraît pas encore vraiment constitutive de tout écrit pour ces élèves à l'entrée en sixième.

La maîtrise de la langue : inégale et toujours fragile

La phrase

Les exercices spécifiques qui portent sur les phrases de types et de formes différentes montrent que cet outil linguistique est mieux maîtrisé que les autres. Et c'est heureux puisque c'est par lui que les élèves comprennent ou indiquent à quoi sert le langage pour celui qui l'emploie (affirmation ou déni, déclaration, conseil ou ordre, interrogation, etc.). Mais les réussites dépendent encore de l'activité proposée. Pour prendre un exemple parmi d'autres, en 1998, l'interrogative est correctement produite par huit élèves sur dix, l'impérative par environ six élèves sur dix. En 2001, sur un texte nouveau, un

récit présenté en trois phrases isolées, environ un élève sur deux produit une interrogative (40% de réponses inadéquates) et quatre sur dix une impérative (40,9% de réponses inadéquates).

La transposition d'un énoncé de l'affirmative à la négative est maîtrisée par six à sept élèves sur dix. Et on peut noter que, même dans les réponses qui concernent les autres questions, l'omission de ne, si courante à l'oral, est négligeable à l'écrit (de 1 à 3%), ce qui confirme que, même fragile et en cours d'acquisition, la conscience des contraintes du code écrit existe chez les élèves à la fin du CM2.

Il reste que, de 1998 à 2002, si la production d'une phrase négative obtient des scores qui varient mais à l'intérieur d'une fourchette stable (de 60 à 80% de réussites), la production d'une interrogative et d'une impérative ont vu les scores de réussite décliner avec le temps pour se rétablir un peu en 2002. Il serait intéressant d'évaluer cette compétence dans la rédaction d'un texte entier, ce qui n'est pas le cas pour le moment.

Dans l'élaboration d'un récit, environ un élève sur deux construit des phrases grammaticalement correctes sans que ce score change beaucoup de 1998 à 2001. Mais si l'on y ajoute le score des élèves qui ne font pas plus de trois erreurs dans l'ensemble de leur texte, plus de huit élèves sur dix y parviennent, avec des résultats en progrès continu. Variable selon les années, le pourcentage d'élèves qui ne sait pas construire des phrases grammaticalement correctes en rédigeant est en moyenne de 15%. Les élèves ont donc majoritairement conscience de l'unité de sens que constitue la phrase, ce que confirment, on l'a vu, leurs résultats plutôt honorables dans l'exercice spécifique qui concerne les types et formes de phrases.

Les accords lexicaux (nom et adjectif) et grammaticaux (accord du verbe et du sujet)

Cette compétence est considérée comme normalement acquise à la fin du CM2. Les scores montrent cependant qu'elle ne l'est pas vraiment. En lecture, déjà, les élèves ne sont pas assez sûrs de leurs connaissances pour ne pas être à la merci de la moindre difficulté du texte. En écriture, quand un exercice demande aux élèves de faire eux-mêmes les accords en transposant par exemple un texte du singulier au pluriel, on constate des réponses très dispersées ou un taux élevé des réponses inadéquates. Si l'on prend en compte, avec les réponses exactes, celles qui ne comportent qu'une erreur, cinq à six élèves sur dix savent faire les accords nécessaires dans un texte, avec un progrès en 2002 où le score global de cet exercice est de 63,6%, et de 73,8% pour l'accord des formes verbales. En 1989 et 1992, faire les accords était classé parmi les compétences les moins réussies (autour de 70%), en particulier chez les élèves faibles (écart de 42 à 82 points avec les autres élèves).

L'accord du nom et de l'adjectif, surtout quand il est attribut, l'accord du verbe et du sujet restent insuffisamment maîtrisés. Le participe passé n'est pas le plus maltraité, mais les réponses sont très inégales selon le texte : sept élèves sur dix l'accordent correctement en 1998, mais seulement cinq à six sur dix en 1999-2000. Enfin, lorsque cette compétence est évaluée dans la production d'un récit, elle n'est également maîtrisée que par un élève sur deux.

La maîtrise de l'orthographe

*Le respect de l'orthographe quand il s'agit de recopier un texte à mettre en page est honorable : environ sept élèves sur dix y parviennent (25% environ d'élèves ne respectent pas l'orthographe du texte qu'ils recopient ; ce score tombe à 7,9% en 2002). En 1992, le score était de 82,6% et cette compétence était considérée comme une des mieux réussies par tous les élèves. Ce fait illustre la part de la concentration et de l'attention dans la maîtrise de l'orthographe qui en outre, dans cet exercice, n'est peut-être pas spontanément considérée par les élèves comme l'objet principal de leurs efforts.

* L'évaluation de l'orthographe lexicale par des exercices spécifiques donne des résultats qui peuvent être convenables, mais très inégaux selon les années. De 1998 à 2002, les scores montrent que sept à huit élèves sur dix orthographient correctement une dizaine de mots courants ou à finale muette dictés et à transcrire dans les blancs d'un texte. En 1999 et 2000, ils sont même neuf sur dix à y réussir, avec 6% d'erreurs. Certes, l'exercice correspondant du protocole de 2001 montre une baisse inattendue et l'orthographe des mots outils n'a été maîtrisée que par 30,2% des élèves en 2002. Mais la réussite globale montre que les élèves ne sont pas tout à fait sans compétence dans ce domaine.

De même, ils sont plus nombreux avec le temps à reconnaître les préfixes et les suffixes dans un mot : les scores passent de 30-40% à 60-70% de 1998 à 2002. Les progrès sont sensibles ; mais, dans tous les cas, les taux élevés de réponses inadéquates et de non réponses traduisent leur embarras.

Cette difficulté à repérer en particulier les préfixes et les termes de base creuse l'écart entre ZEP et non ZEP : entre 15 à 21 points de 1999 à 2002 au bénéfice de ces derniers.

* L'orthographe grammaticale, elle, reste très fragile : on l'a vu par exemple pour l'exercice spécifique portant sur les accords. C'est également le cas dans le récit élaboré par les élèves : ces dernières années, ils sont moins d'un sur deux à les maîtriser. En 2002, un tiers des élèves ne maîtrise pas l'accord du verbe et du sujet dans le récit produit.

En effet, c'est précisément dans la production d'un texte, quand les élèves doivent traiter l'ensemble des problèmes d'orthographe qui se présentent à eux au lieu de ne concentrer leur attention que sur les quelques termes d'un exercice, que la maîtrise de l'orthographe lexicale et grammaticale se révèle médiocre sans que les scores changent beaucoup au fil des années. Un peu plus d'un élève sur deux montre une maîtrise convenable (dont environ 15% une bonne maîtrise). Mais 40% à 50% des élèves ne maîtrisent pas l'orthographe lexicale ou grammaticale alors que, lorsqu'ils rédigent un texte, ils choisissent eux-mêmes les termes et les structures linguistiques dont ils ont besoin. C'était déjà le cas en 1992 (40% à 50% de réussite de 1989 à 1992 pour la maîtrise de l'orthographe grammaticale en production écrite).

Le vocabulaire

L'utilisation d'un vocabulaire approprié ou "pertinent", quand il est évalué dans les textes produits, n'est le fait que d'un peu plus d'un élève sur deux. En 1997, 21% des élèves utilisaient un vocabulaire pertinent et, en outre, varié. La part des récits dont le vocabulaire est globalement impropre est infime, autour de 3 à 4%. Mais ils sont 30% en 1998 à employer un vocabulaire "approximatif" dans leur récit, 51,4 % en 2001 et 23,7% en 2002 à ne pouvoir diversifier les substituts, dont les groupes nominaux, qui renvoient aux personnages de leur texte.

La maîtrise du texte : assez bien installée et plutôt en progrès

Il s'agit principalement de trois compétences :

- construire un cadre spatio-temporel cohérent et, en particulier, savoir utiliser les temps verbaux pour maintenir l'homogénéité temporelle du texte et éviter les incohérences ;
- rédiger un texte où les mots qui désignent les personnages ou qui y renvoient (déterminants, pronoms personnels, groupes lexicaux) soient suffisamment clairs pour qu'il n'y ait pas d'ambiguïté sur les personnages (chaîne référentielle),
- faire en sorte que la progression des informations au fil du texte soit claire et sans rupture (progression thématique).

Ces trois compétences sont également évaluées par des exercices isolés des protocoles, en lecture et en écriture.

Le cadre spatio-temporel du récit

A l'entrée en sixième, les élèves ont une représentation du temps et de l'espace ; ils sont capables d'utiliser des repères temporels et des indices spatiaux clairs et cohérents. Avec le temps, cette compétence se confirme ou progresse nettement. De 1998 à 2001, de sept à plus de huit élèves sur dix construisent un cadre temporel cohérent, de cinq à plus de huit élèves sur dix un cadre spatial cohérent. Le pourcentage d'élèves qui ont des déficiences dans ce domaine varie de 12% à 20% selon les années pour le temps, mais s'est réduit d'environ 20% pour l'espace. Si donc ils ont du mal, en lecture, à utiliser les éléments du texte qu'ils lisent pour reconstituer la succession des faits, ils en ont beaucoup moins quand il s'agit d'utiliser eux-mêmes ces éléments de la langue pour traduire un ordre temporel qui est le leur.

La valeur des temps (dont le passé simple) et leur emploi pertinent du début à la fin du récit sont maîtrisés par plus de sept élèves sur dix, avec là encore des progrès continus : moins de 50% de réussites en 1998, plus de 60% en 1999 et 2000, plus de 75% en 2001.

Ces réussites et ces progrès reflètent sans doute le travail continu accompli en CM2 sur la cohérence spatio-temporelle du récit. Elles ne sont pas contradictoires avec la maîtrise très insuffisante de la morphologie verbale par les élèves entrant en sixième (moins d'un sur deux), et avec le fait qu'ils n'étaient en 1998 que cinq à six sur dix à reconnaître les formes verbales du passé de l'indicatif dans un texte. Les élèves ne maîtrisent pas la conjugaison, la structure du système verbal, ni le métalangage

de l'analyse des formes verbales, mais ils connaissent la valeur des temps dans un récit et savent l'utiliser.

La chaîne référentielle et les mécanismes de reprise

En rédigeant un récit, sept à huit élèves sur dix savent employer des substituts qui renvoient sans ambiguïté à ce qu'ils représentent ; cette compétence est en progrès depuis 1998 même si la situation d'écriture proposée en 2001 provoque une légère baisse des scores. La proportion d'élèves qui ne peuvent éviter des ambiguïtés gênantes s'est réduite de plus de la moitié en 1999 et 2000 (8,2% et 6,8%) par rapport à 1998 (20,4%) ; elle remonte à 23,7% en 2001 sans doute à cause d'une situation d'écriture moins favorable, qui n'annule cependant pas les progrès passés. Le pourcentage des élèves qui ne savent pas du tout utiliser les substituts et rendent à peu près impossible l'identification des référents est toujours de l'ordre de 3 à 5%.

Cependant, moins d'un élève sur deux sait varier les substituts, et en particulier employer des substituts lexicaux sans se contenter des seuls pronoms personnels. Cette compétence dépend du lexique dont dispose l'élève et du traitement de l'information puisque, souvent, le substitut lexical ajoute une information à la simple reprise. Elle se développe au cours de l'année de sixième.

Les exercices spécifiques des protocoles portant sur la chaîne référentielle montrent que six élèves sur dix environ savent identifier les référents des pronoms personnels et des substituts lexicaux dans un texte qu'ils lisent. Cette légère différence dans les scores de réussites pourrait signifier que, en situation de production écrite, les élèves savent traiter leurs difficultés pour éviter l'ambiguïté, en particulier en évitant les substituts lexicaux. Ce serait un signe très encourageant dans l'acquisition de l'écrit.

Les exercices des protocoles portant sur l'emploi des déterminants et des pronoms (autres que les pronoms personnels) pour compléter un texte dont on doit par conséquent comprendre la cohésion grammaticale, montrent une maîtrise très inégale selon les déterminants et pronoms et selon le contexte, avec des pourcentages de réponses erronées importants (de plus d'un tiers à presque 40% selon les cas). Ces difficultés expliquent que les élèves aient encore du mal à les utiliser comme termes de reprise.

L'emploi constant d'une même personne, et en particulier de la troisième personne, sans changement de narrateur au cours du récit est maîtrisé par la majorité des élèves, avec un progrès continu depuis 1998.

La progression thématique

Presque sept élèves sur dix rédigent un texte où les informations progressent clairement du début à la fin, et ce score change peu d'une année sur l'autre. De 5% à 8% des élèves, selon les années, ne maîtrisent pas la progression thématique et accumulent inutilement les répétitions d'informations.

Les ruptures dans la progression thématique, qui signalent une compétence en cours d'acquisition, semblent le fait d'un pourcentage d'élèves qui s'accroît avec le temps : il passe de 13% en 1998 aux environs de 20% en 1999 et 2000, pour atteindre presque 30% en 2001 et encore 25% en 2002. Mais cette compétence complexe n'est pas censée être acquise à l'entrée en sixième.

Conclusion

Environ sept élèves sur dix produisent donc des récits cohérents : cela montre qu'ils savent ce qu'est un récit ; mais aussi qu'ils savent traiter les problèmes de nature diverse (organisation du temps et de l'espace, usage des temps verbaux et des personnes, mécanismes de reprise, progression des informations) que pose la rédaction du texte à mesure qu'ils l'élaborent. Et cette compétence est en général en progrès depuis 1998.

Leurs difficultés ou leurs insuffisances correspondent souvent à une maîtrise de la langue en cours d'acquisition. Encore savent-ils parfois gérer leurs déficiences en choisissant le moindre mal, sacrifier la richesse à la clarté par exemple dans l'emploi des substituts, ce qui est une attitude intéressante en production écrite où l'on est responsable de ses choix.

On peut rappeler qu'en lecture, au contraire, la compréhension ou la reconstitution de l'organisation logique et temporelle d'un texte, de sa cohérence interne, est une compétence difficile à laquelle est consacrée la classe de sixième mais qui est déjà maîtrisée, à l'entrée en sixième, par un peu plus d'un élève sur deux.

Le respect de la consigne et de la situation d'écriture

Elaborer un récit d'une longueur convenable

Plus de sept élèves sur dix sont capables d'élaborer un texte d'une vingtaine de lignes. Si l'on ajoute à cette proportion le nombre de ceux qui produisent un texte de longueur plus réduite (cinq à quinze lignes), ils sont plus de neuf sur dix à y parvenir. Le pourcentage des élèves qui ne sont pas capables de produire un texte (quelques lignes ou rien du tout) est très faible et en baisse régulière.

Les textes produits sont des récits. Le nombre de ceux qui ne respectent pas la consigne et produisent des textes qui ne sont pas des récits se réduit fortement avec le temps : de 10,9% en 1998 à 2,9% en 2001.

Tenir compte des éléments imposés ou proposés par le libellé et, éventuellement, par les documents iconographiques qui l'accompagnent

Les résultats semblent montrer que tous les éléments proposés dans la situation d'écriture n'ont pas le même caractère de nécessité aux yeux des élèves. Si le destinataire d'un texte pourrait leur paraître indispensable dans un discours, une lettre par exemple, il n'est pris en compte que par moins d'un élève sur deux pour le récit en 1999, malgré un libellé qui le définit explicitement. Il est possible que cette contrainte leur ait paru gratuite. Une situation d'écriture différente, par exemple un récit à insérer dans un texte pour justifier le propos d'un personnage, rendrait peut-être plus nécessaire aux yeux des élèves la prise en compte du destinataire dans un récit en donnant une fonction au récit (illustrer ou expliquer une prise de position par exemple).

Que le libellé soit ou non accompagné d'une image qui fournit les éléments du récit à écrire, un peu plus de sept élèves sur dix respectent dans la situation proposée ce qui relève à leurs yeux de l'essentiel d'un récit : les personnages et un événement ou une histoire qu'ils vivent. Mais si le libellé est un peu plus complexe, proposant par exemple une rencontre suivie d'une aventure, 25,5% se contentent de la rencontre. De la même manière, tout en élaborant un récit cohérent dans ces domaines, ils ne sont qu'environ un sur deux à utiliser les indices spatiaux et les repères temporels fournis par l'image qui accompagne le libellé en 2000. Enfin, ils ne se sentent pas obligés de déduire du libellé ce que ce dernier ne dit pas explicitement. Ainsi, rencontrer un animal qui va vous faire vivre "une aventure extraordinaire" n'implique pas pour tous les élèves qu'on identifie cet animal et *a fortiori* qu'on le caractérise pour justifier ce que cette aventure aura d'extraordinaire : presque 7% des élèves ne font ni l'un ni l'autre, presque 30% ne le caractérisent pas, seulement un élève sur deux raconte la rencontre en caractérisant suffisamment l'animal.

Quand la situation d'écriture propose à l'élève de faire ses propres choix (temps et lieu, narrateur et personnage), huit élèves sur dix les respectent, dont plus de la moitié en les insérant avec pertinence dans le récit.

La proportion des élèves qui ne tiennent pas vraiment compte de la situation d'écriture proposée, même en élaborant un récit qui reste cohérent, se montre très inégale selon la tâche proposée, de 4% à 11% environ, sans dépasser le tiers des élèves. Sans doute restreignent-ils le terme de récit à quelques éléments essentiels qu'ils retiennent du libellé (des personnages, une histoire), et qu'ils respectent à près de 70%.

Conclusion

Plus de sept élèves sur dix savent donc élaborer un récit d'une vingtaine de lignes, et respecter l'essentiel de la situation d'écriture proposée par le libellé, bien qu'un élève sur deux en moyenne prenne des libertés avec la consigne, en omette une partie ou n'en apprécie pas toute la portée. La proportion de ceux qui sont incapables d'élaborer un texte est infime.

Ils savent majoritairement écrire un récit cohérent, sans ruptures dans la progression des thèmes et des propos, avec des termes de reprise qui permettent au lecteur d'identifier les personnages, même si la variété de ces termes ne permet pas toujours au récit d'éviter la monotonie. La plupart des élèves ont également le sens de la situation d'énonciation du récit, aussi bien pour l'emploi des personnes verbales que pour la valeur des temps. Les progrès continus dans ce domaine semblent refléter l'efficacité de l'apprentissage sur ce sujet.

C'est dans la maîtrise de la langue que les difficultés apparaissent surtout : seul un élève sur deux en moyenne est capable de construire des phrases syntaxiquement correctes, de faire les accords nécessaires et d'utiliser convenablement la ponctuation pour segmenter l'énoncé ; un tiers utilise un vocabulaire approximatif dans le récit, et quatre élèves sur dix n'en maîtrisent pas l'orthographe, que seuls 15% en moyenne maîtrisent vraiment bien. C'est dans ce domaine aussi qu'on constate le plus fréquemment un écart entre les scores de lecture et d'écriture. En effet, la production écrite demande à l'élève de traiter en même temps et tout au long du texte les différents problèmes linguistiques que

soulève son discours ; cela demande de la vigilance et une mobilisation continue des savoirs et savoir-faire, c'est plus difficile.

Les compétences des élèves de ZEP à l'entrée en sixième.

Ces dernières années (2000, 2001, 2002), l'écart moyen entre les scores des élèves en ZEP et hors ZEP est d'environ dix points (8,8 en 2000, 9,4 en 2001, 10,9 en 2002). A titre de comparaison, il était de 10 points en 1994 dans l'académie de Créteil. En observant les cas où cet écart est plus ou moins éloigné de l'écart moyen, on peut avoir une idée des compétences sur lesquelles les élèves de ZEP ne rencontrent pas plus de difficultés que les autres, et celles sur lesquelles ils sont notablement plus faibles.

I. Compétences les mieux réussies par les élèves de ZEP.

Ce sont celles où les scores des élèves de ZEP sont proches ou peu éloignés de ceux des autres élèves, que ces scores soient élevés et signalent des compétences installées, ou qu'ils soient bas pour tout le monde et signalent des compétences non acquises ou en cours d'acquisition à l'entrée en sixième. Ce n'est pas à travers ces compétences que les élèves de ZEP se distinguent des autres.

En lecture,

- saisir globalement le sens d'un texte en choisissant le titre qui convient à la totalité du texte, bien que les résultats varient sensiblement selon la nature du texte ;
- reconnaître et nommer le genre de textes différents (avec une difficulté, partagée par tous les élèves, à identifier un texte documentaire) ;
- prélever des informations ponctuelles explicites dans un texte lu ou entendu, ou dans un tableau de données;
- comprendre un message oral et en exécuter les consignes ;
- construire des informations pour évaluer la vérité ou la fausseté d'affirmations portant sur des éléments explicites d'un texte lu ;
- traiter les informations d'un tableau de données quand la sélection et le recoupement des indices textuels ne sont pas trop nombreux et le tableau de données relativement simple ;
- relever les indices textuels qui justifient les réponses (les scores des élèves de ZEP baissent pour ces items, mais pas significativement plus que ceux des autres élèves).

En écriture,

- Recopier un texte informatif lisiblement et sans erreur, en proposer une mise en page qui mette en évidence la structure du texte (selon les textes, la hiérarchisation des informations est plus ou moins bien comprise. mais, en 2002, l'exercice est globalement mieux réussi qu'auparavant) ;
- construire un récit d'une longueur convenable (environ vingt lignes) et prendre en compte les éléments proposés dans la consigne et le libellé ;
- respecter jusqu'au bout du récit le choix énonciatif (troisième personne) et construire un système pronominal cohérent ;
- utiliser d'une manière cohérente et relativement variée les substituts pronominaux et lexicaux qui renvoient aux personnages.

Deux compétences importantes pour la cohérence interne d'un texte, utiliser avec pertinence les temps du récit et construire un texte dont les thèmes et les propos s'enchaînent sans faire naître de rupture, sont en ZEP fragiles avec un écart qui se creuse en 2002.

Dans la maîtrise des outils de la langue,

- Reconnaître et produire des phrases de types différents (phrase interrogative ou impérative) ;
- orthographier correctement, en les insérant dans un texte lacunaire, quelques mots courants dictés.

Quand il s'agit plus particulièrement d'utiliser la ponctuation pour segmenter en phrases un texte lu, de connaître et de savoir utiliser les déterminants et les pronoms, ou de faire les accords nécessaires dans le groupe nominal et dans le groupe verbal, les élèves de ZEP ne se distinguent pas des autres par un écart constamment très éloigné de leurs scores avec l'écart moyen (sauf pour la constitution de paragraphes dans le texte). Mais cela signifie qu'ils ont autant de difficultés que les autres sur ces compétences où les scores globaux sont généralement bas. En outre, sur certains

accords (accord du participe passé, de l'adjectif possessif et du pronom personnel, par exemple), il semble que leurs connaissances soient trop peu sûres pour que l'inattention ou la lenteur ne soient pas sources d'erreurs selon la nature de l'exercice.

II. Compétences où les élèves de ZEP rencontrent des difficultés particulières.

Ce sont des compétences où, quel que soit le score global, l'écart se creuse entre les scores des autres élèves et ceux des élèves de ZEP.

En lecture,

- comprendre, par une lecture attentive et non une intuition globale, le sens des différents paragraphes ou des différentes parties d'un texte (identifier la phrase qui résume le plus exactement un texte est plus difficile pour eux que d'identifier celle qui pourrait lui servir de titre) ;
- construire des informations à partir d'un texte lu, quand il s'agit de confronter plusieurs indices textuels explicites pour trouver l'information demandée, ou d'appliquer sur eux une opération (déduction par exemple) ;
- utiliser l'information construite en l'appliquant à un cas différent ;
- traiter (et non seulement repérer ou prélever) des informations à partir d'un tableau de données, en particulier quand il s'agit de sélectionner et de recouper plusieurs informations pour trouver la réponse ;
- comprendre, en la reconstituant, la structure logique et chronologique d'un texte (compétence spécifiquement au programme de la sixième) ;
- établir abstraitement la succession temporelle des actions évoquées dans un texte ou les situer sur un calendrier (compétence difficile chez tous les élèves, très dépendante en ZEP des situations : l'écart est important en 2001 et 2002, négligeable en 2000).

En écriture,

Dans la rédaction d'un récit,

- segmenter en phrases le texte qu'on écrit grâce à la ponctuation, le constituer en paragraphes (avec un écart toujours plus important de 2000 à 2002) ;
- construire des phrases syntaxiquement correctes,
- accorder correctement le verbe et le sujet,
- maîtriser l'orthographe,

Dans la maîtrise des outils de la langue,

- En vocabulaire, identifier en s'aidant du contexte le sens d'un mot ou le sens particulier d'un mot polysémique (compétence fragile : les écarts, négligeables en 2000, l'étaient moins en 2001 et se creusent en 2002) ;
- connaître et utiliser la composition des mots et le métalangage adapté (préfixe, suffixe, terme de base, famille de mots) ;
- utiliser les accords verbaux ou nominaux pour identifier les personnages d'un texte (écarts constamment très importants) ;
- identifier les référents des substituts lexicaux et pronominaux ;
- en orthographe lexicale et grammaticale, écrire correctement les mots à finale muette, faire les accords nécessaires dans les groupes nominal et verbal en particulier dans une transformation de phrase ou dans une phrase dictée ;
- transformer une phrase affirmative en une phrase négative (cas, dans la production de phrases, où est généralement utilisé un métalangage : la *forme négative*).

Ces compétences font souvent partie des compétences en cours d'acquisition à l'entrée en sixième, comme l'identification des référents des substituts, par exemple, qui n'est généralement réussie que par six élèves sur dix. La composition des mots et la connaissance du système lexical n'est également maîtrisée, selon les items, que par quatre à six élèves sur dix.

Conclusion.

Les élèves de ZEP montrent, comme les autres élèves, plus de facilité à prélever des informations explicites dans un texte qu'ils lisent ou qu'ils entendent ou dans un tableau de données, et même à construire des informations quand ces dernières ne supposent pas de recoupements trop compliqués. Ils comprennent un message oral et en exécutent les consignes. Ils reconnaissent le genre de la plupart

des textes. Comme on peut s'y attendre, les scores les plus faibles des élèves de ZEP portent sur les exercices où généralement l'ensemble des élèves a déjà des difficultés et qui relèvent de compétences «pointues» : le traitement ou la construction des informations, la compréhension de la temporalité d'un texte, la saisie de son enchaînement logique et chronologique c'est-à-dire sa cohérence interne, compétences en cours d'acquisition.

Dans la production écrite, ils ne se distinguent pas notablement des autres élèves pour rédiger un récit écrit lisiblement et d'une longueur convenable, qui respecte la consigne et la situation d'écriture, dont les propos s'enchaînent et dont les substituts sont clairs.

C'est dans la maîtrise de la langue, celle des déterminants et des pronoms, mais aussi de l'orthographe grammaticale (accords, morphologie verbale) et du vocabulaire que leurs résultats semblent fragiles, parce que très variables d'une année sur l'autre, et dépendants de la situation proposée (phrase ou texte, activité demandée). Les écarts sont particulièrement élevés, ou moyennement élevés mais avec constance, dans tout ce qui relève de la maîtrise de la langue, dans ce qui suppose une attitude réflexive sur la langue, qu'il s'agisse de l'observer pour construire le sens ou de l'utiliser pour produire un texte.

A titre de comparaison, dans l'académie de Créteil en 1994, les élèves de ZEP montraient des difficultés non pour reconnaître, mais pour produire ou analyser en utilisant un vocabulaire spécifique. Les items qu'ils avaient le moins bien réussis concernaient également la langue (maîtrise du paragraphe et de la ponctuation, orthographe, vocabulaire, grammaire, accords).

C'est pourquoi, peut-être, les élèves de ZEP sont-ils relativement plus dépendants que les autres de la situation d'écriture ou de lecture qui leur est proposée. Leur intuition du sens global d'un texte peut leur permettre de choisir le titre convenable alors qu'ils ne disposent pas d'informations justes sur les idées principales de chaque partie du texte ; au contraire, les élèves qui disposent de ces informations ont du mal à les synthétiser pour choisir un titre (2001). Sur un autre texte, dont ils comprennent la situation d'énonciation, ils se trompent lourdement sur le titre (2000). De même, selon la situation d'écriture, ils peuvent utiliser avec cohérence les temps du récit (2001), ou ne pas y parvenir (2000, 2002).

Il se peut qu'ils soient plus négligents ou plus instables que les autres. Mais peut-être aussi leur maîtrise de la langue est-elle trop fragile pour qu'ils puissent la mobiliser sans effort et sans perte de temps dans la lecture et la production écrite.

NB. En 1989-1990, une étude a été menée sur les élèves entrant en 6° et en particulier en 6° «aménagée» (classe de remise à niveau). 29% des élèves de 6° aménagée réussissaient moins de 50% des items, contre 7% des élèves de 6° «standard». L'écart sur certaines compétences pouvait être de 45 à 65 points (comme en 1992, où les 20% d'élèves les plus faibles avaient des scores inférieurs de 23 à 82 points aux scores des autres élèves). La difficulté la plus grande à tous les élèves de 6° concernait la conjugaison, l'identification et l'effectuation des accords. (*DT n°398*, DEP 10, décembre 1991).

Les 10% d'élèves les plus faibles

En 1993, des grilles de compétences en lecture ont été élaborées, à partir des évaluations à l'entrée du CE2 et de la 6° de septembre 1992, par la DEP et l'Inspection générale (*Les niveaux de lecture à l'entrée au CE2 et en 6°*, DEP 10 n° 93-320 du 1° octobre 1993). Par recouvrements d'items, elles hiérarchisent les compétences de lecture en trois groupes :

- les compétences de base, nécessaires à un élève pour profiter pleinement de son année,
- les compétences approfondies, qui mesurent les avancées des élèves sur les acquisitions de l'année,
- les compétences remarquables, qui préfigurent les savoir-faire et les stratégies qui seront développés dans l'année.

Pour chacune des trois compétences répertoriées dans ces grilles, le seuil de réussite représente «le nombre d'items auxquels un enfant aura dû répondre correctement pour maîtriser la compétence concernée», c'est-à-dire 75% des items constituant cette compétence (9 items sur 12 en CE2, 7 items sur 9 en sixième, par exemple). Maîtriser le second ou le troisième bloc de compétences (approfondies ou remarquables), suppose donc la maîtrise préalable des blocs de compétences précédents **et** celle d'au moins 75% des items qui constituent le bloc de compétences auquel on accède. Même calcul pour les sous-compétences : «Un élève maîtrise...une sous-compétence du deuxième bloc (compétences approfondies) s'il a répondu correctement à 75% de cette sous-compétence **et** s'il maîtrise les compétences de base.»

Cette analyse des évaluations, abandonnée plus tard parce que considérée comme trop empirique, était conçue pour «permettre, a priori, d'établir une base de comparaison, un indicateur d'évolution de niveau» : à partir des évaluations de septembre 1994, «les protocoles seront construits de manière à mesurer, dans cette nomenclature pertinente, les compétences en lecture des élèves de ces deux niveaux.» Elle présentait surtout deux avantages : hiérarchiser les compétences, besoin exprimé par certains de nos interlocuteurs qui estiment les protocoles moins clairs et moins utilisables sans hiérarchisation des compétences ; et indiquer quelles compétences sont ou ne sont pas maîtrisées par les pourcentages d'élèves donnés. C'est pourquoi nous disposons d'une indication chiffrée des niveaux de compétences des élèves en lecture dès 1992, avec leur suivi en CE2 et en 6°.

**Liste des compétences de lecture (1992)
Niveau CE2**

Compétences	Pourcentages d'élèves
1. Compétences de base: comprendre les mots et les phrases	80,1 %
1.1. Comprendre les différents sens d'un mot dans des contextes usuels	75,4%
1.2. Comprendre des énoncés et des consignes simples et usuels	84,6%
1.3. Prélever des informations explicites et simples dans 1 texte court. Les restituer.	75,7%
2. Compétences approfondies: prélever des informations pertinentes.	25,6%
2.1. Distinguer des types d'écrit dans lesquels les indicateurs peuvent être diffus.	50,4%
2.2. Retrouver une définition simple dans une page de dictionnaire.	50,7%
2.3. Reconnaître l'ordre alphabétique avec ou sans utilisation du dictionnaire.	3,4%
3. Compétences approfondies: discriminer de façon pertinente.	10,7%
3.1. Prélever une information qui nécessite induction et mise en relation.	19,6%
3.2. Discriminer les indices morpho-syntaxiques pour accéder à un sens précis.	10,4%
3.3. Utiliser le dictionnaire pour trouver le sens d'un mot.	8,2%

**Liste des compétences de lecture (1992)
Niveau sixième**

Compétences	Pourcentages d'élèves
Compétences de base de lecture d'un texte: saisir l'information explicite de l'écrit.	88,5 %
1. Comprendre de qui (de quoi) on parle.	67,6%
2. Tirer des informations ponctuelles d'un écrit.	98,6%
3. Maîtriser l'ordre alphabétique.	58,3%
Compétences approfondies de lecture: reconstituer l'organisation de l'explicite.	60,5 %
1. Retrouver les enchaînements de l'écrit.	61,9%
1.1. l'enchaînement logique de causalité.	82,4%
1.2. l'enchaînement chronologique.	49,5%
1.3. l'enchaînement référentiel.	54,0%
2. Maîtriser les règles principales du code écrit.	46,4 %
2.1. les accords.	65,0%
2.2. les types de phrase.	64,0%
2.3. les formes verbales.	19,5%
3. Utiliser les ressources du contexte.	71,1%
3.1. Déjouer les effets de la polysémie.	71,1%
Compétences remarquables: découvrir l'implicite.	19,1 %

4.1. Mettre en relation deux informations.	30,1%
4.2. Dégager le présupposé d'un énoncé.	13,1%
4.3. Dégager du contexte le sens d'un mot inconnu.	32,3%

Répartition des élèves selon les niveaux de lecture (1992)

Niveaux de lecture	Entrée au CE2	Entrée en 6 ^o
Elèves maîtrisant les compétences de base	80,1%	88,5%
<u>ne maîtrisant pas les compétences de base</u>	<u>19,9%</u>	<u>11,5%</u>
ne maîtrisant que les compétences de base	54,5%	28,0%
maîtrisant les compétences de base et approfondies	14,9 %	41,4%
maîtrisant les compétences remarquables, donc les 3 blocs de compétences.	10,7%	19,1%

NB. La première ligne du tableau a été ajoutée au tableau initial, ainsi que les soulignements.
Les niveaux de lecture à l'entrée au CE2 et en 6^o. DEP 10 n° 93-320. 1.X.1993.

Ces tableaux permettaient à la DEP de dire en 1992, qu'à l'entrée en sixième, **«un peu plus d'un élève sur dix ne maîtrise pas les compétences de base permettant de saisir l'information explicite d'un écrit, trois élèves sur dix ne maîtrisent pas plus que ces compétences de base.»** pour «quatre élèves sur dix qui savent reconstituer l'organisation de l'explicite ... et environ deux sur dix (qui) comprennent l'implicite d'un texte.»

En 1998, en fonction des mêmes grilles, l'analyse de la DEP aboutissait aux résultats suivants :

- 14,7% des élèves entrant en sixième ne maîtrisent pas les compétences de base en lecture (23,4% à l'entrée au CE2) ;
- ces 14,7% d'élèves qui ne maîtrisent pas les compétences de base en lecture à l'entrée en sixième réussissent pourtant en moyenne 58% des items de base, 32% des items approfondis, et 22% des items remarquables ;
- les 10% d'élèves les plus faibles réussissent quatre questions sur dix, avec des résultats plus dispersés que les autres; ils sont très faibles sur les compétences complexes, c'est-à-dire la compréhension globale et logique, la construction d'informations.

En 1998, on pouvait donc penser que les 10% d'élèves très faibles sont loin d'être démunis de toute compétence de lecture ; pour autant, il leur manque la maîtrise de plus de la moitié des compétences essentielles en lecture.

A partir de 1998, les notes d'information et les dossiers annuels de la DEP (*Evaluations CE2-sixième, repères nationaux*) fournissent des graphiques montrant la répartition des élèves selon leurs scores en français.

Années	moins de 30% des items	moins de 40% des items	moins de 50% des items	score global	score 10% forts	score 10% faibles
2000	2,4% (1/40)	5,7 (1/18)	13% (1/8)	68,5%	91,7%	35,5%
2001	1,6% (1/60)	3,2% (1/30)	7,8% (1/13)	72% (66,2 items/92)	90,2%	52,2%
2002	2,3% (1/40)	6,4%(1/ 16)	17,7% (1/6)	65,5% (57/87 items)	89,1%	34,9%

D'après ces graphiques, en 2000, 5,7% des élèves en début de sixième avaient un score compris entre

30 et 40 en français, c'est-à-dire avaient réussi en moyenne 30% à 40% des items du protocole. 13%, un peu moins d'un élève sur huit, avaient réussi moins de 50% des items du protocole.

Les commentaires qui accompagnent les graphiques dans les notes d'information observent en 2002 que «les 10% les plus faibles à l'ensemble de l'épreuve ne sont pas obligatoirement faibles dans tous les champs ou toutes les capacités» (ce que notait déjà la DEP en 1993) et que les performances de «cette population hétérogène» «fluctuent beaucoup d'un champ à l'autre et d'une capacité à l'autre», avec quelques difficultés particulièrement sérieuses dans le champ de la compréhension dès qu'il ne s'agit plus seulement de prélever directement des informations dans un document.

On peut noter cependant que la comparaison des scores globaux des 10% des élèves les plus faibles et des 10% des élèves les plus forts montre que la force de ces derniers réside surtout dans la production écrite. Peut-être est-ce aussi ce qui fait ou ce qui accentue la faiblesse des premiers.

Années	capacités	1. score global	écart 1 moins 2	2. 10% faibles	écart 3 moins 2	3. 10% forts	écart 3 moins 1
2001	<i>compréhension</i>	78,1%	- 19,6 points	58,5%	-34,5 points	92,7%	+14,6 points
2001	<i>outils de la langue</i>	65,3%	-25,9 points	39,4%	-48,5%	87,9%	+22,6 points
2001	<i>production de textes</i>	70%	-25,6 points	44,4%	-50 points	94,4%	+24,4 points
2002	<i>savoir lire</i>	69,1%	-29,6 points	39,5%	-51,2 points	90,7%	+21,6 points
2002	<i>outils de la langue</i>	62,2%	-32,3 points	29,8%	-57,7 points	87,5%	+25,3 points
2002	<i>savoir écrire</i>	60,4%	-31,9 points	28,5%	-58,3 points	86,8%	+26,4 points

L'ensemble des élèves maîtrise mieux la lecture que l'écriture, et mieux l'écriture que les outils de la langue ; c'était aussi le cas les années précédentes et les résultats de 2002, légèrement supérieurs en maîtrise de la langue qu'en production écrite, ne contredisent pas cette constante. C'est vrai aussi pour les élèves les plus faibles et, en général, pour les élèves les plus forts. Les scores plus particuliers des outils de la langue pour lire d'une part, pour écrire d'autre part, montrent entre eux des différences légères pour les trois catégories d'élèves.

Si l'on regarde le rapport des trois compétences entre elles, on voit que les scores entre maîtrise de la langue et production écrite forment en général un bloc de scores proches, éloigné du score de la lecture. Ainsi pour les élèves les plus faibles en 2001, 5 points séparent les scores de maîtrise de la langue et d'écriture, contre 14,1 points entre lecture et écriture et 19,1 points entre lecture et maîtrise de la langue. Ce n'est pas le cas chez les élèves les plus forts, qui ont en général des scores beaucoup plus ramassés. Peut-être une maîtrise insuffisante de la langue est-elle solidaire, chez les plus faibles, de la faiblesse en écriture.

Si l'on regarde d'autre part l'écart des scores entre les différentes catégories d'élèves, on voit que c'est généralement en écriture qu'il est le plus important : en 2001, 25,6 points séparent les élèves les plus faibles de l'ensemble des élèves en écriture pour 19,6 en lecture, et 50 points les séparent des élèves les plus forts en écriture pour 34,5 en lecture. En 2002, où les scores de tous les élèves sont moins dispersés entre les trois compétences, 31,9 points séparent encore les élèves faibles de l'ensemble des élèves en écriture pour 29,6 points en lecture, et 58,3 points les séparent des élèves les plus forts en écriture pour 51,2 en lecture. Notons d'ailleurs qu'en 2001, les scores d'écriture étaient, chez les plus forts, d'1,7 point supérieurs aux scores de lecture.

Conclusion. Si une évaluation ciblée portait sur les compétences et les difficultés de ces 10% d'élèves les plus faibles (que savent-ils vraiment ? D'où viennent les difficultés qu'ils rencontrent ?), il serait intéressant de voir dans quelle mesure une maîtrise plus solide de la langue leur permettrait d'une part

de fournir à leur intelligence plus de points d'appui sur lesquels fonder la compréhension des textes, d'autre part de traiter plus aisément les multiples problèmes qu'implique la production d'un écrit.

Enfin, un nouvel indicateur de niveau serait bienvenu : il devrait permettre de hiérarchiser les compétences, donc de mesurer plus facilement l'évolution dans le temps du pourcentage d'élèves privés des compétences attendues à l'entrée en sixième, tout en précisant de quelles compétences il s'agit.

Les évaluations diagnostiques de seconde

Les compétences de lecture dans les protocoles de seconde GT (générale et technologique)

Les évaluations diagnostiques nationales à l'entrée en seconde ont existé annuellement de septembre 1992 à septembre 2001, dans les séries générale et professionnelle. Destinées à la fois aux élèves et aux enseignants, elles visent à éclairer les uns sur leurs possibilités et leurs besoins, à aider les autres à organiser leur enseignement et particulièrement les modules (groupes d'aide et de remédiation).

Les évaluations ne fournissent donc pas un bilan des acquis du collège, mais un diagnostic sur les acquis et difficultés des élèves, et un pronostic sur leurs possibilités à atteindre les objectifs de la classe où ils viennent d'entrer. Comme ce pronostic, fait en début d'année, ne peut suffire à guider la constante réorganisation des modules au cours de l'année scolaire, l'évaluation prend un double aspect à partir de septembre 1995 :

- évaluation à la rentrée, «nationale, systématique et obligatoire» ,
- utilisation libre, en cours d'année, de la banque d'exercices dont on annonce la parution à partir de janvier 1996 (Présentation générale, cahier du professeur, septembre 1995).

En 1998, la banque d'exercices était par exemple constituée, pour la seconde GT, de trois tomes appelés *Aide à l'évaluation* , avec en particulier une évaluation de l'oral. Les exercices seront ensuite informatisés.

Par ailleurs, les cahiers destinés aux professeurs de 2^oGT proposent, outre les réponses attendues aux exercices du protocole et les consignes de codage, des séquences complètes d'enseignement qui peuvent servir à exploiter, en classe, les résultats des évaluations.

Les protocoles de sixième et de seconde GT en lecture ont été construits dès l'origine selon des logiques différentes.

Dans les protocoles de seconde GT, les items de lecture ne portent pas, comme en sixième, sur des supports changeant la plupart du temps avec chaque exercice, afin d'évaluer chaque fois une composante déterminée et isolée de la capacité de lecture. Les protocoles de seconde proposent soit un petit nombre de textes relevant, en tout ou en partie selon les années, de genres différents (informatif, narratif et descriptif ou argumentatif), soit un groupement de textes représentant différents genres comme les programmes de lycée en recommandent la pratique. Tous les items de lecture portent alors soit sur chacun de ces textes, soit sur les textes et l'ensemble du groupement. Il s'agit d'évaluer la manière dont les élèves savent mettre en œuvre les diverses compétences acquises au collège pour comprendre le texte qu'ils ont sous les yeux. Si un item porte sur un élément de la langue ou du style dans le texte, c'est que la compréhension de cet élément concourt à la compréhension d'ensemble du texte. Du coup, alors que, dans les protocoles de sixième, les exercices paraissent juxtaposés à l'intérieur d'un cadre ordonné, la succession de ceux des protocoles de seconde semblent constituer une démarche de lecture, qui passe par des observations ou analyses ponctuelles pour aboutir généralement à la formulation d'une synthèse ou d'une interprétation d'ensemble.

Les activités demandées en seconde consistent, comme auparavant, à comprendre globalement les textes, à en saisir la structure interne, à construire sur eux des informations et des significations. Mais, même à partir de libellés apparemment équivalents, les tâches sont souvent plus complexes.

Cette complexité des tâches est due en grande partie à la complexité des textes : cela représente, par rapport aux protocoles de sixième, le seuil qui sépare le collège du lycée. Une tâche apparemment simple et présente dans les protocoles de sixième, comme extraire une information explicite, suppose en réalité, en seconde, la compréhension totale du texte.

La complexité vient aussi de formats de réponse comme les tableaux. Dans les protocoles précédents, les tableaux sont des supports à partir desquels les élèves doivent relever des indices explicites ou construire des informations. En seconde GT, les tableaux sont des formats de réponse : construits spécifiquement pour la question et souvent sans titre, ils doivent être complétés totalement ou partiellement par l'élève. Cela suppose que l'élève, avant de répondre, ait compris la logique qui a présidé à leur construction et en déduise ce qu'on lui demande. Un grand nombre de réponses erronées

viennent de là.

La complexité vient enfin du fait que les réponses demandent, plus qu'auparavant, de produire des phrases ou de rédiger des énoncés. Il peut s'agir de formuler synthétiquement une thèse à partir des arguments précédemment dégagés, ou l'interprétation globale d'un texte narratif. Il peut s'agir de produire un énoncé explicatif fondé sur des citations du texte. Parfois, un item demande de formuler en une phrase ce qu'on peut déduire des relevés demandés à l'item précédent : la production écrite joue ici le rôle de vérification et de synthèse.

Les protocoles de seconde demandent donc aux élèves, par des libellés qui ne sont pas fondamentalement différents de ceux des protocoles antérieurs, des opérations sur les textes plus complexes, qui mettent souvent en jeu la compréhension globale du texte et qui exigent parfois une réflexion préalable sur le libellé. Le diagnostic permet de savoir dans quelle mesure les élèves y sont préparés à la fin du collège.

Les tableaux de compétences en lecture dans les protocoles de seconde GT

Dès le début, la démarche de lecture évaluée a été semblable quel que soit le genre du texte. Dans les premiers tableaux, elle va de la capacité (comprendre un texte) aux compétences (observer, mettre en relation, construire un sens, interpréter,) puis aux composantes évaluées. Elle va «du général au particulier, du global au spécifique» (Document du professeur, Présentation générale, septembre 1995). Le choix des compétences est explicitement rapporté aux démarches mises en œuvre dans la méthode de lecture préconisée par les programmes de seconde : « l'observation objective, précise, nuancée des formes ou des systèmes de formes,... l'analyse de l'organisation de ces formes,... la construction progressive d'une signification, ... la constatation, dans une synthèse, de ce qui fait l'unité complexe et profonde du texte» (*Instructions officielles pour la classe de seconde*, B.O. spécial n°1 du 5 février 1987).

Mais, avec le temps, on s'aperçoit que les élèves ont du mal à observer, à prélever des indices et à interpréter, parce qu'on ne repère pas un indice («traces textuelles de ce qu'on cherche») sans avoir justement une idée de ce qu'on cherche, une idée fournie par une hypothèse de lecture, par une perception globale du genre et de l'enjeu du texte. C'est pourquoi, à partir de 1996, dans l'esprit d'une méthode de lecture mieux comprise, on abandonne les items qui, dès 1993, évaluaient systématiquement la langue, le vocabulaire ou l'organisation matérielle du texte. On adopte, pour tout texte, une suite équivalente de compétences qui donneront, selon le genre du texte, des «composantes» spécifiques :

- *adapter la stratégie de lecture*, c'est-à-dire prendre des informations qui orienteront l'observation du texte (genre, thème ou problème posé, enjeu du débat, ...)
- *analyser*, c'est-à-dire repérer, identifier, interpréter, isolés ou organisés en systèmes, des faits de langue et des procédés d'écriture qui permettent d'explicitier, selon le texte, le processus d'argumentation, les instances narratives, l'organisation de l'information, etc. ;
- *dégager l'essentiel*, c'est-à-dire rendre compte du parcours argumentatif, proposer une interprétation d'ensemble.

Le nombre des items, réduits à l'essentiel, passe ainsi de quinze ou dix-huit à une dizaine par type de texte. Les savoirs et savoir-faire sont évalués dans la mesure où ils sont utilisés, «finalisés» pour la compréhension d'un texte particulier.

Comme pour les protocoles de sixième, l'évolution des tableaux de compétences de seconde reflète donc une réflexion continue sur la démarche de lecture à évaluer, et sur la spécificité des genres dont relèvent les textes supports. Le rapport entre les deux évaluations n'est pas anodin. Au début de la sixième, il s'agit de voir dans quelle mesure les savoirs, démarches et savoir-faire nécessaires à la lecture sont acquis ; au début de la seconde on constate, comme par exemple en 1998, que les élèves, même lorsqu'ils les ont maîtrisés au cours du collège, ne savent pas utiliser ces faits de langue pour valider une hypothèse et pour interpréter, c'est-à-dire définir leur rôle et leur signification dans le sens global du texte. C'est la tâche que devra accomplir le lycéen.

Ce que savent faire les élèves en lecture au début de la seconde GT, c'est-à-dire au sortir du collège.

Année	score global	lire un texte argumentatif	lire un texte narratif	lire un texte informatif
1996	50%	39%	50%	24%
1997	63,5%	56%	63,4%	/

Malgré la variété des scores d'une année sur l'autre, les élèves maîtrisent mieux la lecture d'un texte narratif / descriptif à l'entrée en seconde, ce qui est normal puisqu'ils l'étudient depuis plusieurs années. Pour le texte informatif, évalué par un très petit nombre d'items, les résultats montrent que les élèves n'ont pas de stratégie de lecture adaptée à ce type de texte. Quant au texte argumentatif, dont l'étude n'est qu'abordée au collège, 40% à 50% des élèves disposent pourtant déjà de compétences pour le lire à l'entrée en seconde.

On peut se demander pourquoi la lecture des textes informatifs n'est pas mieux maîtrisée par les élèves. En effet, ils devraient les rencontrer dans d'autres disciplines que le français, plus spécialisé dans les textes littéraires, et par conséquent ils devraient les rencontrer assez fréquemment compte tenu du nombre des disciplines concernées (sciences humaines et sciences de la vie et de la terre, au moins). C'est peut-être que ce type de texte est alors plus considéré comme un document où puiser des informations que comme un texte à étudier en tant que tel.

Lecture d'un texte argumentatif ou à visée argumentative.

Les textes supports sont des discours écrits (préface) ou ayant été prononcés, des discussions dans des extraits romanesques, des articles de presse ou de revue, y compris dans les groupement de textes.

Un élève sur deux en moyenne le comprend ce type de texte. Ce résultat est intéressant dans la mesure où l'étude approfondie de l'argumentation, commencée au collège, s'approfondit surtout au cours du lycée.

Pour les différentes compétences, les scores de réussite vont du tiers à presque 50%, avec des résultats très contrastés. Ce contraste, ainsi que le nombre parfois important de réponses partielles ou erronées (souvent plus de 50%) pour un faible taux de non réponses (en général moins de 10%, sauf pour quelques items), trahit la fragilité de compétences en tout début d'acquisition.

	saisir la spécificité du texte	analyser	identifier les thèses défendue/ rejetée	identifier arguments et exemples	dégager l'essentiel
1996	42% re : 48,3% nr : 9,7%	38%	40,3% re : 58% nr : 1,7	*12,2% re 35,5% * 44%, re 57,9%	36,7% re : 57,4% nr : 5,9%
1997	73%	47,6%	38,1% reformulation . 45,8%	35,6%	71,7%

re= réponses erronées. nr= absence de réponse, non réponse.

Les scores des composantes évaluées montrent que les élèves ont surtout des difficultés dans l'analyse, qui consiste en général à synthétiser et interpréter les faits de langue retenus parce qu'ils sont essentiels au débat (articulations logiques, pronoms personnels, mots exprimant le thème, procédés stylistiques,...). Ces faits de langue peuvent leur être donnés ou doivent d'abord être identifiés et recherchés par les élèves dans le texte (pour dégager les arguments par exemple). Les scores diffèrent beaucoup selon les textes et les faits à analyser. Mais les taux de réponses erronées concernent souvent plus de la moitié des élèves, et ceux des non réponses de 15% à 25% d'entre eux.

Quatre élèves sur dix savent repérer les arguments et les exemples, à peu près autant peuvent identifier les thèses défendue et rejetée. Notons que la reformulation de la thèse est réussie par 21,2% des élèves (45,8% avec les codes 2 signalant une réussite partielle) quand ils doivent synthétiser les arguments qu'ils viennent de recueillir, et par 38% d'entre eux quand la reformulation est seule : les élèves qui sortent du collège ont encore de la peine à rédiger un écrit qui synthétise des informations, comme l'avaient déjà montré les évaluations de 6°.

La saisie de la spécificité du texte, qui ouvre la démarche de lecture et permet d'anticiper sur

le débat qui va suivre, est trop fragile pour ne pas dépendre fortement du texte. Elle semble liée à l'interprétation d'ensemble qui ferme la démarche : en 96, un quart des élèves saisissent la spécificité du texte, presque 37% dégagent l'essentiel du texte. En 97, où 73% des élèves perçoivent le thème principal du texte, l'interprétation globale est bien réussie (avec pourtant 25% de réponses erronées). Ce fait confirme la validité de la démarche de lecture présentée dans les tableaux de compétences.

Les difficultés ou les échecs des élèves tiennent souvent à la nature des consignes, des libellés ou des tableaux qui les accompagnent. Ces items ne permettent donc pas de mesurer vraiment la compétence visée, et les élèves auraient sans doute mieux réussi avec d'autres consignes. Mais cela montre peut-être, inversement, qu'ils ont du mal à réussir quand le format de la réponse est nouveau pour eux, qu'ils ne savent pas adapter leur comportement à une situation inédite. C'est le cas de tableaux sans titre dont ils doivent comprendre la logique avant de les compléter.

D'autres difficultés viennent d'un métalangage qu'ils maîtrisent encore mal, ce qui les empêche d'interpréter celui de certains libellés dans le sens qui est attendu : quand «raisons» est employé à la place d'arguments», «exemples» à la place de «phrases du texte données comme exemple», «positions» quand ils attendraient «thèses», «termes exprimant un jugement» c'est-à-dire un jugement implicite puisqu'un jugement est une proposition ou une phrase et non un terme, etc. Ils n'ont pas encore compris assez profondément le fonctionnement d'une argumentation pour ne pas s'en tenir à la seule dénomination qu'ils connaissent. Cette difficulté est inhérente aux nombreuses notions théoriques à la base de toute argumentation, ce que les groupes d'élaboration expliquent d'ailleurs en ces termes : « ...Tous les élèves ne sont pas familiarisés de la même manière avec des termes dont il est bien difficile de se passer si l'on veut poser des questions concises et rigoureuses. Pour éviter cet écueil, on a fait le choix d'utiliser des termes neutres (par exemple jugement, modification) sans parvenir pour autant à satisfaire tous les enseignants ni, ce qui est plus ennuyeux, à faciliter la tâche de tous les élèves.» (*Résultats globaux par capacités et par compétences*, DEP 1997).

Enfin, beaucoup ne distinguent pas nettement la reformulation de l'interprétation ou du commentaire.

Conclusion. A l'entrée en seconde, la nature de l'argumentation et la logique du débat sont encore obscurs pour environ la moitié des élèves, qui ne maîtrisent ni leur fonctionnement ni le métalangage qui permet de les analyser. Ils ont donc du mal à les reconnaître dans un texte et à interpréter dans ce sens les faits de langue et de style qui visent à persuader ou à convaincre. Ces confusions se reflètent dans les écrits de formulation ou de reformulation qu'ils doivent élaborer.

Lecture d'un texte narratif ou descriptif.

	score global	adapter la stratégie de lecture	Analyser	Dégager l'essentiel	
1996	50%	33%	51%	77%	
1997	63,4%	82%	58,1%	60,5%	

La connaissance que les élèves ont depuis longtemps du genre narratif fait qu'ils perçoivent apparemment bien le sens global des textes qui en relèvent : de six à plus de sept élèves sur dix proposent une interprétation globale satisfaisante du texte présenté.

Cependant, la complexité des textes fait que certaines compétences sont encore fragiles bien qu'elles aient été travaillées continûment au cours du collège. La reconnaissance du genre et de la spécificité du texte, généralement bien réussie aux évaluations de sixième, en fait partie. A l'entrée en seconde en 1997, sept à plus de huit élèves sur dix savent analyser un début de roman en fonction de ce qu'ils en attendent, disposant ainsi de repères précis pour son étude. Mais, en 1996, ils n'étaient qu'un peu plus du tiers à identifier le genre du texte, avec environ 65% de réponses erronées. Même constatation pour la délimitation des passages narratifs et descriptifs : en 1996, un quart à 40% des élèves se trompent. Enfin, 30 à 40% des élèves ne saisissent pas la progression chronologique d'ensemble, ce qui était déjà le cas des élèves entrant en sixième, avec 36,6% de réponses erronées.

Comme pour le texte argumentatif, c'est l'analyse et l'interprétation des faits de langue qui pose le plus de problèmes, surtout quand elles sont couplées avec l'obligation de les formuler. Un quart ou moins de 50% des élèves, selon les cas, ont du mal à interpréter des indices textuels qu'on a recueillis pour eux. Les items qui demandent à la fois de recueillir les indices, de les interpréter puis

de formuler ou de justifier cette interprétation connaissent des scores de réussite bas (39%, 47%) avec les taux de non réponses les plus élevés pour cette compétence (de l'ordre de 36%).

Conclusion. Bien que le genre narratif soit le mieux connu des élèves, ils sont encore loin d'en maîtriser la compréhension, surtout quand cette dernière requiert une lecture attentive. Les difficultés les plus évidentes portent sur l'autonomie de l'observation, le passage à l'interprétation, surtout l'écrit d'analyse et de commentaire en réponse à des questions ouvertes.

Notons que ces observations, présentes par exemple dans les commentaires sur les évaluations de 1996 et 1997, coïncident avec les résultats que donnera plus tard l'évaluation internationale PISA sur les compétences de lecture des élèves de quinze ans, dont beaucoup en France sortent du collège et un certain nombre de seconde. Nos évaluations montraient déjà qu'à la sortie du collège nos élèves ont plus de mal à interpréter qu'à recueillir ou à construire des informations, et manquent relativement d'autonomie quand il s'agit de recueillir eux-mêmes les informations qu'ils devront interpréter ; mais qu'ils ont surtout des difficultés à élaborer un écrit portant sur leur lecture pour expliquer, commenter ou justifier une réponse, pour formuler une interprétation ou une analyse, et plus généralement à argumenter. La tendance des élèves à renoncer à répondre en cas d'embarras est souvent notée, et cela dès la sixième ; c'est seulement lors de comparaison avec les élèves d'autres pays qu'elle apparaîtra comme une sorte de spécificité française.

Les compétences d'écriture dans les protocoles de seconde GT

Les protocoles de sixième et de seconde sont proches pour la conception de la production écrite

La production écrite y est principalement évaluée par des productions indépendantes des séquences de lecture, bien que ces dernières, en seconde, donnent une plus grande part à la production de phrases ou d'énoncés comme format de la réponse. Certes, les programmes signalent que lecture et écriture sont des activités interdépendantes. Le scripteur est le premier lecteur de ce qu'il écrit et, quand il améliore son «brouillon», il accomplit, pour écrire, une activité de lecture. Beaucoup de sujets d'écriture naissent ou s'inspirent de lectures, et c'est le même matériau linguistique qui construit l'une et l'autre. L'écriture est souvent un moyen de développer pour soi, plus ou moins brièvement, un sentiment ou une opinion nés d'une lecture.

Il serait donc possible que les protocoles, de collège comme de seconde, demandent des productions écrites de genre et de longueur divers à l'intérieur des séquences de lecture. Mais ce n'est pas le cas. A la production écrite évaluée en sixième par la rédaction d'un récit, correspond, en seconde, la rédaction d'un texte argumentatif, parfois de deux au choix, avec l'obligation de produire un nombre déterminé d'arguments et d'exemples à l'appui d'une prise de position sur un sujet général. L'évolution des programmes de seconde provoquera quelques modifications, sans que disparaisse la visée argumentative du texte à produire. La plupart du temps, le libellé précise la nature de l'écrit (un article de journal, un discours,...), sa longueur (trente lignes au moins), la situation d'énonciation, les destinataires, l'objet à traiter et le but à atteindre. Pour lier dans le travail des élèves tout ce qui relève de l'argumentation, les protocoles ont rapidement proposé de déplacer, lors de la passation des épreuves, la séquence de production écrite pour qu'elle suive directement la lecture du texte argumentatif.

Les compétences des élèves en production écrite

Une quinzaine d'items permet d'évaluer trois compétences relevant des trois domaines à distinguer dans l'écrit demandé et qu'on trouvait déjà dans les protocoles de sixième. A partir de 1997, leur intitulé prend en compte les distinctions inscrites dans les nouveaux programmes : le discours, le texte, le langage. Ces compétences sont donc intitulées: produire une argumentation (le discours), écrire un texte cohérent (le texte), respecter les contraintes linguistiques (la langue).

Année, score global	produire une argumentation	écrire un texte cohérent	respecter les contraintes ling.
1996 : 62%	63%	58%	68%
1997 : 69,4%	65,4%	80%	74,1%

La variabilité des scores sur une durée d'évaluation aussi limitée ne permet pas d'établir de constantes. On observe cependant que plus de six élèves sur dix produisent un texte qui est bien

l'argumentation demandée par le sujet, et qui respecte les contraintes orthographiques et syntaxiques de la langue. Si les scores très élevés de 1997 (80% et 74,1%) étonnent les commentateurs de la DEP qui les attribuent à la brièveté des productions évaluées, il n'en reste pas moins que plus d'un élève sur deux construit un texte cohérent, c'est-à-dire dont les articulations logiques et les reprises nominales ou pronominales sont claires. Cela signifie que 50% à 60% des élèves maîtrisent les cadres formels d'un écrit argumentatif à l'entrée en seconde.

Le détail des items montre que la difficulté porte surtout sur la construction de l'argumentation elle-même : de 25% à 40% des élèves ne réussissent pas à choisir avec pertinence, à développer et à organiser en un parcours construit les arguments et les exemples ; dans le meilleur des cas, un élève sur deux seulement formule explicitement la thèse ou la position qu'il présente ou qu'il réfute. Ces difficultés correspondent à celles que révèle la lecture d'un texte argumentatif : la moitié au moins des élèves qui entrent en seconde ne maîtrise pas le type de raisonnement qui est à l'œuvre dans une argumentation. C'est en effet l'une des tâches dévolues au lycée.

Evolution des compétences des élèves de seconde

Poursuivant l'opération d'évaluation pédagogique de l'enseignement commencé en 1979 dans les écoles et en 1980 dans les collèges, une évaluation des élèves de fin de seconde a eu lieu en mai 1986, dans les quarante-trois lycées de l'échantillon constitué, pour expérimenter et préparer les épreuves définitives. Les compétences évaluées se fondaient déjà sur les Instructions officielles de 1981 (arrêté du 26.01.1981). L'épreuve de français consistait à répondre à vingt-trois questions sur un texte narratif de Zola (*Germinal*), et à rédiger un texte argumentatif de deux pages au maximum (soixante items de correction). Les questions posées sur le texte évaluaient la compréhension du fonctionnement d'un texte, la connaissance de la langue, les compétences littéraires et la culture générale (lire un texte dans une perspective historique).

Les élèves testés alors avaient déjà suivi une année de seconde et le texte argumentatif ne figurait pas comme support des tests de lecture. Les composantes évaluées en 1986 et depuis 1992 ne sont pas exactement identiques. Cependant, les performances de ces élèves sur des compétences comparables montrent qu'ils étaient autant qu'aujourd'hui à rencontrer des difficultés analogues.

En lecture, moins d'un élève sur deux savait tirer une conclusion d'observations préalables, moins d'un sur quatre savait justifier cette synthèse : un élève sur cinq seulement savait «passer de l'observation analytique à une synthèse justifiée.» Un peu plus de quatre élèves sur dix pouvait caractériser les personnages, mais plus d'un quart n'avaient pas repéré une réponse figurant explicitement dans le texte. Les commentaires soulignent que «les exigences qu'on est en droit d'attendre d'élèves de seconde sont loin d'être satisfaites lorsqu'on essaie d'entrer dans l'analyse d'un texte», surtout en faisant preuve d'autonomie dans la recherche. Les points réussis, comme la reconnaissance de la situation initiale et de son évolution (81%), ou la «lecture guidée d'un texte» le sont encore aujourd'hui. Les erreurs étaient en partie attribuées à une lecture superficielle des consignes et à une grande méconnaissance du métalangage littéraire, ce qui avait conduit le groupe de travail à éliminer le plus possible des libellés définitifs les termes techniques, même courants (champ lexical, implicite, utopie, emphase,...).

En expression écrite, les textes produits, d'une longueur très disparate, respectaient généralement l'organisation traditionnelle en paragraphes distinguant l'introduction, le développement, la conclusion, acquis du collège. Mais cette présentation était largement formelle : la majorité des élèves ne savait pas poser un problème, produire une argumentation qui ne soit pas un «catalogue d'idées ou d'exemples», rédiger une conclusion qui réponde au problème posé (un tiers seulement de réussites sur ce point).

La maîtrise de la langue, surtout, était défailante, et déjà dans les textes produits, où on soulignait entre autres l'impropriété du vocabulaire, et l'emploi déficient de la ponctuation. Mais une partie du protocole était en outre consacrée à la compétence «connaître la langue» (niveaux et registres de langue, lexique, grammaire, syntaxe). D'après les réponses, un tiers des élèves seulement maîtrisait le vocabulaire, en particulier abstait (l'un des objectifs du lycée), moins d'un quart savait employer et orthographier correctement les temps et les modes, domaine censé acquis à la fin du collège. La terminologie grammaticale était «gravement ignorée» (4 à 14% des élèves avaient su analyser les

formes verbales proposées). Le taux de non réponse pouvait monter à 59% sur certains items de ce champ. Quant aux manières dont les paroles sont rapportées dans un texte (directement ou indirectement) moins de 20% des élèves pouvaient nommer et moins d'un quart reconnaître le discours indirect, environ 40% avaient su repérer les interventions au discours direct.

Conclusion.

Les difficultés pointées en mai 1986 portent sur des compétences en cours d'acquisition dès la seconde et au cours du lycée, comme la maîtrise du raisonnement argumentatif et du langage technique qui le constitue, l'exposé écrit d'une argumentation, la conquête d'une autonomie dans l'analyse d'un texte, l'écriture synthétique. Les scores des élèves aux évaluations récentes du début de seconde conduisent à des observations identiques.

La maîtrise très insuffisante de la langue pouvait souligner, en 1986, la pauvreté du vocabulaire et l'ignorance d'une terminologie grammaticale élémentaire, parce que cette maîtrise était spécifiquement évaluée dans une partie du protocole. C'est peut-être ce qui manquait aux évaluations plus récentes à l'entrée en seconde.

Notons que certaines de ces carences sont encore celles qui pénalisent les étudiants à leur entrée à l'université et même après :

- des ignorances en vocabulaire qui conduisent à des contresens en lecture ;
- en production écrite, une maîtrise insuffisante de certains éléments de la langue (emploi de *dont*, de la double négation, de l'interrogation indirecte, des accents non grammaticaux, entre autres), une difficulté à formuler un titre, à distinguer la neutralité qui est de mise quand on reformule la pensée d'un autre de l'engagement qu'on attend dans un commentaire ou une interprétation ;
- en lecture, une difficulté à saisir l'essentiel d'un texte et à hiérarchiser les idées, et tout ce qui relève de la chronologie (notions d'antériorité, de postériorité, d'anticipation pour construire le temps dans un texte, situation d'un texte dans le temps).

Les protocoles de 2^o professionnelle et les compétences des élèves

Ces protocoles évaluent les élèves de première année de BEP (brevet d'études professionnelles) en lycée professionnel, issus des classes de troisième (et, en 1992, de troisième technologique). D'emblée, le protocole de 1992 indique vouloir «mettre l'accent sur la capacité à mettre en œuvre des savoir faire, à mobiliser des compétences intellectuelles plutôt que des connaissances proprement dites.» La grille des compétences évaluées comprend deux *capacités générales*, comprendre un message et produire un message, et trois *capacités transversales*, respecter les consignes, maîtriser le code écrit, établir et analyser une relation logique. La maîtrise du code écrit comprend elle-même trois compétences : connaître et respecter l'orthographe, employer et élargir le vocabulaire usuel, observer et construire des phrases. L'originalité du protocole de 1992 vient de ce qu'il propose deux séries équivalentes de cinq activités chacune, afin que le professeur ou l'équipe pédagogique choisisse, tout en respectant l'ordre des activités, le parcours le plus approprié aux élèves. Chaque exercice porte sur un support spécifique, et la production écrite demande de reconstituer l'ordre d'une phrase dont les membres sont présentés en désordre, puis de rédiger quatre phrases inscrites dans le cadre argumentatif proposé.

La réflexion continue du groupe de travail modifie ensuite la logique des protocoles. On évalue désormais quatre capacités, comprenant chacune deux ou trois compétences : s'informer (rechercher, organiser l'information), réaliser (choisir, traiter, exécuter), communiquer, rendre compte (présenter, structurer), apprécier (contrôler, valider, critiquer). Ces compétences entraînent les composantes évaluées. Enfin, à partir de la rentrée 1999, un seul cahier d'élève regroupe l'ensemble de activités proposées dans chaque discipline, avec un thème fédérateur commun et un choix des capacités qui peut changer annuellement. Ainsi, en septembre 2000, le cahier évalue les deux capacités s'informer et réaliser ; il s'organise «autour d'une histoire policière et d'une énigme à résoudre, sur le thème des télécommunications». Là encore, la logique des protocoles se modifie donc profondément avec le temps.

Les résultats en français de l'évaluation de seconde professionnelle en 1996 (score global : 50,4%) montrent que les plus forts taux de réussite se trouvent rassemblés dans la capacité s'informer, bien qu'elle n'obtienne que 52% de réussite globale : «la majorité des élèves possède de réelles compétences en lecture et sait repérer des informations ponctuelles dans un texte.» Pour mettre en relation des informations, la réussite reste satisfaisante et en progrès sur les évaluations précédentes.

Les élèves sont aussi mieux préparés à la lecture attentive des énoncés et des consignes. Mais l'exercice d'écriture est moyennement réussi, surtout à cause de la pauvreté du vocabulaire.

De même, les résultats de l'évaluation de seconde professionnelle BEP en Economie et gestion de la même année (1996) montrent que les élèves ont les meilleurs résultats dans la capacité s'informer (67% de réussite), les moins satisfaisants dans la capacité communiquer et rendre compte (41% de réussite, surtout pour la compétence structurer qui obtient 23% de réussite avec un nombre élevé de non réponses). Parmi les difficultés des élèves, les commentateurs soulignent le manque de maîtrise du vocabulaire courant et ce qui relève de la communication écrite ; ils notent que la compétence critiquer-valider (47% de réussite) est encore en cours d'acquisition.

L'évaluation des élèves à l'entrée en 5^o

Pour la première fois en septembre 2002, tous les élèves entrant en cinquième ont été soumis à des épreuves standardisées de français et de mathématiques. Comme pour la sixième, ces épreuves visent à identifier les difficultés que les élèves peuvent avoir à s'adapter aux exigences du cycle auquel ils accèdent, afin que les professeurs et les chefs d'établissement disposent d'outils pour organiser l'enseignement et mettre en place les dispositifs de remédiation. Comme pour la sixième, elles testent des compétences normalement acquises au début cycle central, et d'autres en cours d'acquisition.

Les programmes de 1985 (Complément intitulé *Savoir lire*) conseillaient, pour évaluer les compétences des élèves en lecture, d'utiliser «des textes cohérents et assez longs, plutôt que des textes trop courts qui obligent à reconstituer un contexte inconnu.» Le protocole de cinquième propose justement un seul texte, une nouvelle de cent-sept lignes, découpé en quatre passages sur lesquels portent les questions des exercices. Les élèves découvrent ainsi le déroulement du récit à mesure qu'ils passent d'un exercice à l'autre, ce qui permet de tester leur activité de lecteur. L'exercice de production écrite, lui, est classique, la composition d'un récit.

Cette évaluation a reçu un accueil contrasté. Certains professeurs ont été heureux de la présence d'un texte littéraire et intégral. Mais d'autres ont trouvé qu'on évaluait des compétences «de lecteur lettré», en cours de construction, sans volet dans le protocole qui puisse dire ce que les élèves avaient appris en sixième. C'est pourquoi certains ont été consternés par les résultats de leurs élèves, et par ce qu'ils semblaient signifier sur l'enseignement dispensé en sixième. Et les résultats comparés des deux années paraissent leur donner raison. Les scores globaux sont plus bas, les pourcentages d'élèves ayant le moins bien réussi plus élevés, et la dispersion des résultats plus grande.

Or il est difficile d'en tirer ces conclusions, parce que les protocoles de sixième et de cinquième ne sont effectivement pas comparables. Cela montre la nécessité, si l'on veut suivre le chemin parcouru par les élèves, de ne pas changer brusquement la logique des protocoles.

Certes, il s'agit toujours de tirer d'un texte des informations ponctuelles, de construire des informations, de comprendre l'organisation logique. Mais, comme si l'on se trouvait dans la logique des protocoles de seconde, les questions du protocole de cinquième visent toutes à voir si l'élève sait, en mettant en œuvre ces opérations mentales, construire le sens du texte qu'il a sous les yeux. Or ce texte, parce qu'il est littéraire et d'une langue soutenue, rend toutes ces opérations plus délicates. Par exemple, les élèves savent presque tous prélever des indices explicites à l'entrée en sixième ; mais, ici, il fallait prélever ceux qui permettent de «caractériser» un personnage, ce qui s'est révélé plus difficile, sans compter le poids du métalangage (caractériser, le terme d'indice qui revient très souvent). D'autre part, on teste en cinquième l'activité du lecteur, la façon dont il participe à cette construction du sens en faisant des hypothèses de lecture, en les vérifiant et en les modifiant, quitte à revenir en arrière au cours de la lecture. Dans le tableau de compétences figurent donc deux compétences nouvelles : élaborer le sens d'un texte, émettre des hypothèses de lecture. Quant aux outils de la langue, ils ne sont plus évalués comme tels. Ils sont censés pouvoir être automatiquement utilisés par l'élève pour élaborer le sens en répondant aux questions. Enfin, les réponses à formuler par des mots ou par des phrases sont plus nombreuses (près de 20%). C'est dire que, contrairement aux protocoles de sixième, on a plus souvent recours à l'écrit comme moyen de rendre compte d'une lecture.

Cet écart entre les deux protocoles rend honorables les résultats des élèves à l'entrée en cinquième (savoir lire : 66,3% / savoir écrire : 59,4% / score global : 63,2%) ; il explique aussi la dispersion des résultats. Dans la perspective de l'évolution des compétences en lecture, on passe légitimement à un autre niveau, celui qui demande de mobiliser ses ressources pour la lecture attentive et récurrente d'un texte long ou intégral; c'est la tâche du collège. Mais dans la perspective d'une observation des progrès des élèves et d'un retour sur l'efficacité de la classe de sixième, celle qu'attendent aussi certains professeurs et chefs d'établissement, cette évaluation apprend peu de choses, sinon qu'elle est prématurée pour un grand nombre d'élèves. Si donc on évalue les acquis des élèves à différents moments de leur cursus, à quel moment un protocole doit-il rompre avec la logique des protocoles

précédents pour prendre en compte les seuils de complexité des compétences pour une même capacité ? Quelle logique préférer : celle qui permet de mesurer les progrès des élèves sur des compétences identiques, ou celle qui mesure les acquis en fonction des niveaux différents d'un même savoir-faire ?

Résultats des élèves à l'évaluation de 2002 à l'entrée en cinquième

1. Plus de neuf élèves sur dix reconnaissent le genre narratif.

Au moins huit élèves sur dix savent toujours opérer sur le texte les opérations habituelles : prélever les informations explicites demandées, construire des informations à partir de quelques indices textuels explicites, synthétiser des observations pour en déduire, par exemple, le point de vue des personnages sur l'objet de ces observations. Environ sept ou huit élèves sur dix savent fournir, pour justifier une réponse, des indices pertinents, même s'ils ne fournissent pas tous les indices demandés ou que se glisse une erreur parmi les indices fournis.

Sept à huit élèves sur dix comprennent la progression des répliques dans un dialogue et rapportent les phrases du dialogue à celui qui les énonce. Plus d'un sur quatre n'y parvient encore pas, mais l'étude du dialogue dans la narration est au programme de la cinquième.

Six élèves sur dix ont réussi à émettre une hypothèse de lecture et à proposer une interprétation en se fondant sur des observations préalables. Mais ils ne sont plus que quatre à cinq sur dix à savoir émettre une hypothèse de lecture à laquelle ils n'ont pas été préparés (item 15), à vérifier une hypothèse de lecture en revenant sur le texte et à la corriger si nécessaire.

2. Les difficultés rencontrées par les élèves.

Les questions qui obtiennent des taux élevés de réponses erronées ou une grande dispersion des réponses donnent des pistes pour interpréter les difficultés des élèves, qu'évidemment seuls des entretiens avec ces élèves ou des évaluations ciblées permettraient d'expliquer plus exactement.

* Plus d'un élève sur deux n'a pas reconnu le genre spécifique du texte à l'intérieur du genre narratif (ils ont cru que c'était un conte). Or ce savoir est maîtrisé dès la sixième, mais à partir d'extraits différents à identifier par QCM, ce qui est une tâche plus facile. Cela signifie peut-être que plus de la moitié des élèves se contente d'une réponse rapide et routinière au lieu de réfléchir et de vérifier sa réponse. Il s'agirait moins d'un manque de savoir, que d'une procédure insuffisante pour construire ce savoir.

* La lecture des consignes est encore difficile. Alors que la majorité des élèves identifie la saison en donnant les indices pertinents, ils sont beaucoup moins nombreux à le faire pour le lieu, et 23,6% donnent une réponse erronée, ce qui est énorme puisqu'il ne s'agit que de prélever des informations explicites. C'est que tous les indices du texte portent sur la même saison. Au contraire, les nombreux indices qui portent sur le lieu indiquent des éléments de nature différente selon qu'ils précisent la distance, la région, des éléments du paysage, des repères dans le parcours des personnages, etc. Pour répondre, les élèves devaient donc interpréter la consigne («préciser le lieu où se passe l'histoire») pour sélectionner des indices parmi la multitude des indices possibles. Près d'un élève sur cinq ne sait pas le faire. Or, c'est là une démarche plutôt qu'un savoir.

* Une stratégie de lecture mal adaptée explique peut-être qu'environ 20% des élèves n'aient pas compris deux éléments essentiels pour le sens du récit, l'identité d'un personnage et la mort de ce personnage. Or, des formules explicites permettaient soit de les déduire, soit de les relever. Mais ces indices étaient éparpillés dans le texte ou rapidement énoncés dans un paragraphe ; et une lecture de balayage, survolant le texte, pouvait ne pas les remarquer.

* Les questions demandant de relever les indices qui permettent d'identifier et surtout de «caractériser» un personnage obtiennent chaque fois 53% de bonnes réponses (77,6% et 81,3% si l'on accepte une omission et une erreur), avec 14,5% et 12,5% de réponses erronées, ce qui étonne puisqu'il s'agit de relever des indices nombreux et explicites. Quand le même exercice est demandé pour vérifier une hypothèse de lecture, avec un aussi grand nombre d'indices explicites, le score est encore plus faible : 40,4% de bonnes réponses seulement (63% avec les codes 3 et 4), 22,3% de réponses erronées et 14,7% de non réponses. Même si la notion de caractérisation est difficile, il semble que les élèves aient du mal à sélectionner et à regrouper en une même classe les indices correspondant à cette classe. Là encore, c'est une démarche de recherche qui est en cause.

Conclusion.

Dans l'ensemble, les résultats en lecture sont plutôt encourageants, dans la mesure où les activités

testées à l'entrée en sixième n'ont pas baissé ou sont en progrès, alors que le texte proposé diffère notablement des extraits courts et relativement plus simples des protocoles de sixième.

Ce qui fait baisser le score global relève de compétences en cours d'acquisition en cinquième et parfois au collège, comme le dialogue et la caractérisation des personnages, outre que le terme technique de «caractérisation» n'est pas nécessairement connu.

Mais il semble que les démarches de construction de la réponse soient en cause dans les scores les plus bas : appliquer au texte les opérations qu'on sait faire et les connaissances dont on dispose pour construire le savoir demandé est encore difficile. En ce qui concerne l'activité du lecteur, nouvelle compétence, si 20% à 25% des élèves n'ont encore aucune activité de lecteur, il est intéressant de constater qu'environ un élève sur deux est déjà un lecteur actif. Enfin, si justifier son opinion en fournissant les indices textuels dont on s'est servi est mieux réussi à l'entrée en cinquième qu'à l'entrée en sixième, ils sont plus d'un quart, à l'entrée en cinquième, à ne pas savoir justifier leur avis par une phrase à compléter. C'est sur ces dernières compétences que l'écart entre les élèves en ZEP et hors ZEP (13 points ou plus) est le plus important.

Résultats des élèves en écriture

1. Les élèves devaient rédiger un récit de voyage à partir d'un libellé et d'une image. L'ensemble de cette consigne précisait les contraintes sur le temps, le lieu et le moyen de transport, un épisode et des personnages. Comme pour la sixième, les items sont classés en trois catégories qui s'inspirent des nouveaux programmes du collège : le discours, le texte, la phrase.

* Selon les résultats, plus de neuf élèves sur dix savent rédiger un récit, en utilisant bien la personne verbale. Parmi eux, huit à neuf élèves sur dix sont capables de rédiger un récit complet d'une page, et sept sur dix le font en tenant compte des éléments essentiels de la consigne. Le pourcentage de ceux qui rédigent autre chose qu'un récit ou n'élaborent presque pas de texte est infime. Déjà, en entrant en sixième, des pourcentages équivalents d'élèves montraient les mêmes compétences.

Cinq à six élèves sur dix (66% en sixième) savent écrire un texte cohérent, avec des réussites diverses selon la source de cette cohérence. La cohérence qui relève de l'identification des personnages est la mieux maîtrisée (sept élèves sur dix, un peu moins à l'entrée en sixième), puis celle qui relève de l'emploi des temps du récit (six élèves sur dix) ; celle qui relève du cadre spatio-temporel et de la progression du texte est la moins maîtrisée (un peu plus d'un élève sur deux, pour six sur dix en sixième).

La variété des termes qui désignent les personnages ou qui y renvoient dépend de l'étendue du répertoire lexical et s'exerce tout au long du collège : elle n'est encore possible que pour un quart des élèves (moins de 25% à l'entrée en sixième), ce qui est normal.

Cinq élèves sur dix savent utiliser la ponctuation pour délimiter les phrases de leur récit, score équivalent à celui des élèves entrant en sixième. Un peu plus d'un quart seulement segmente son récit en paragraphes (moins d'un quart à l'entrée en sixième), ce qui est normal pour une compétence exercée tout au long du collège.

* Comme à l'entrée en sixième, c'est la maîtrise de la langue qui obtient les scores les plus bas à l'entrée en cinquième.

- Environ quatre élèves sur dix construisent des phrases syntaxiquement correctes ; ces scores sont un peu plus élevés à l'entrée en sixième.

- Quatre élèves sur dix également savent accorder le verbe et le sujet (36% à l'entrée en sixième) ; si l'on admet les scores de ceux qui n'ont fait que quelques erreurs, ils sont toujours plus d'un sur deux à ne pas savoir accorder le verbe et le sujet. Le nombre des élèves chez qui la syntaxe des phrases est en cours d'acquisition est à peu près équivalente à l'entrée en sixième (35,9%) et à l'entrée en cinquième (38,5%) ; mais le nombre de ceux chez qui l'accord du verbe et du sujet est mal maîtrisé à l'entrée en sixième (31,3%) s'est nettement réduit à l'entrée en cinquième (12%).

-La maîtrise de l'orthographe lexicale et grammaticale n'est maîtrisée que par un peu plus d'un élève sur dix à l'entrée en cinquième, moins qu'à l'entrée en sixième (14%). Avec le pourcentage d'élèves qui la maîtrisent d'une manière «acceptable», on obtient le score de 48,8%, presque un élève sur deux. A l'entrée en sixième, ce score dépasse légèrement la moyenne (53,3%).

Si donc sur certaines compétences les élèves ont des scores légèrement supérieurs à ceux des élèves qui entrent en sixième, environ 20 à 25% des élèves montrent encore de grandes carences en production écrite à l'entrée en cinquième.

Les élèves de ZEP à l'entrée en 5° : l'écart moyen avec le score global est de 8,3 (il était de 10,9 à l'entrée en sixième de la même année)

En lecture.

Les écarts les plus importants qui signalent les faiblesses particulières des élèves de ZEP portent justement sur les compétences où l'ensemble des élèves a rencontré des difficultés : sélection des indices pour indiquer le lieu, caractérisation des personnages, en particulier quand elle sert à vérifier une hypothèse de lecture, repérage des émetteurs des répliques dans la partie dialoguée du récit, compréhension d'un événement important du récit. Contrairement à l'ensemble des élèves (plus de 80% de réussites), les élèves de ZEP ont eu du mal à synthétiser les observations textuelles pour formuler l'opinion des personnages (les écarts sont, selon les items, de 11,9 et de 13,6 points).

Cependant, les écarts sont faibles ou peu significatifs pour les compétences les plus complexes : formuler un avis et le justifier par une phrase, formuler une hypothèse de lecture, la modifier si nécessaire. Il est vrai que le retour sur le texte pour chercher les signes annonciateurs de cette hypothèse étant mal réussi (11 points d'écart, un tiers seulement des élèves l'ont réussi), il était difficile de comprendre la nécessité de modifier cette mauvaise hypothèse de lecture. Mais les autres hypothèses à émettre ne distinguent pas fortement les élèves de ZEP des élèves hors ZEP. Une seule observation : là où six sur dix des élèves hors ZEP ont formulé une hypothèse de lecture qui se fondait en partie du moins sur des observations textuelles qu'ils avaient trouvées préalablement, les élèves de ZEP sont à peine moins nombreux à formuler cette hypothèse mais avec 11 à 13 points d'écart sur les observations textuelles préalables. Il est possible qu'ils aient l'intuition juste du sens du texte, sans pouvoir se fonder efficacement sur le texte pour construire ce sens.

En production écrite

Comme les autres élèves, la majorité des élèves de ZEP (plus de sept élèves sur dix) sait élaborer un récit complet qui tienne compte de la consigne, qui utilise clairement la personne du récit et les substituts pronominaux. Comme eux, certains ont négligé, dans la consigne, les éléments qui précisent la nature du récit (récit de voyage) et sa construction (nécessité d'une rencontre).

Ils rencontrent la même relative difficulté à employer avec cohérence les temps du récit (un élève sur deux y réussit en ZEP pour six élèves sur dix hors ZEP) et à faire progresser les informations (autour d'un élève sur deux en ZEP et hors ZEP). Ils sont aussi faibles que les autres pour la constitution de paragraphes, moins d'un élève sur deux y réussit en ZEP.

C'est dans la maîtrise de la langue que l'écart se creuse en ZEP : orthographe lexicale et grammaticale (moins de 40% des élèves de ZEP ont une orthographe bonne ou acceptable), accord du verbe et du sujet, syntaxe de la phrase (moins d'un tiers des élèves de ZEP construit des phrases correctes), emploi de la ponctuation pour délimiter les phrases (écart de 13,7 points, le plus important).

Conclusion sur les évaluations du second degré

La logique des protocoles de sixième et de seconde générale et technologique ainsi que la nature de leurs exercices montrent clairement les seuils à franchir en lecture et en écriture du collège au lycée. C'est plus obscur pour les évaluations récentes sur les seuils internes au collège comme au début du cycle central (début cinquième en 2002). C'est sur ce point qu'une échelle des niveaux d'exigence et des seuils de compétence serait utile pour construire des protocoles qui soient différents tout en restant comparables, donc utilisables pour qui veut suivre l'évolution des compétences des élèves d'un niveau d'études à l'autre.

L'observation des résultats des élèves fournit cependant des pistes éventuellement utilisables dans une telle réflexion. De la sixième à la seconde, **que ce soit en lecture ou en connaissance de la langue,**

- reconnaître ou identifier paraît plus facile que nommer et analyser, et plus facile que produire ;
- justifier ce qu'on a identifié en recopiant les mots utilisés pour répondre est plus difficile qu'identifier, mais mieux réussi que rédiger une phrase pour commenter ou expliquer ;
- l'ignorance de la terminologie (littéraire comme grammaticale), c'est-à-dire l'ignorance des notions utiles et des éléments de l'écrit qui les expriment, entrave la compréhension des

consignes, celle des textes et la maîtrise de la langue.

En lecture,

- le relevé des indices explicites devient difficile dès le collège quand il s'agit de les regrouper en une classe (ce qui caractérise un personnage, en seconde ce qui constitue un argument ou une thèse,...), surtout s'ils sont épars dans le texte, qu'il faut les comparer, les classer à plus forte raison les hiérarchiser, ou que le texte se présente comme une situation complexe ;
- quand il s'agit d'utiliser concurremment les savoirs ou les savoir-faire apparemment acquis pour élaborer le sens d'un écrit ou pour interpréter, ces mêmes acquis se trouvent souvent fragilisés ;
- enfin, la lecture guidée est mieux réussie que la lecture autonome dès la sixième.

En production écrite,

- l'écriture d'argumentation et de commentaire semble plus difficile que l'écriture de fiction ou d'invention ;
- comme en connaissance de la langue, la maîtrise de la ponctuation, de la morphologie verbale, des articulations et des structures logiques (cause, conséquence, but), du paragraphe comme unité signifiante et non purement formelle, d'un vocabulaire exact et adapté, surtout abstrait, fait partie des compétences qui se construisent au cours du collège sans être acquises à la fin de la seconde ;
- en revanche, comme le montrent les évaluations-bilans de collège évoquées ci-après, certaines notions ou compétences grammaticales semblent acquises ou proches de l'être à la fin de la troisième, comme l'accord des noms au pluriel, l'orthographe des homonymes grammaticaux (ni / n'y, ça / sa,...), certaines articulations logiques comme l'opposition, les types de phrases, par exemple. Les indices de temps et de lieu sont longtemps délicats à reconnaître : en fin de 5^e, les élèves ne reconnaissent que les groupes nominaux prépositionnels (par exemple *dans la maison* , mais non la maison *dans laquelle*...).

Les évaluations-bilan

Les évaluations-bilan dans le premier degré

Le souci d'évaluer les acquisitions scolaires au niveau de toute une génération d'écoliers est relativement récent.

I. C'est en 1979 que les services statistiques du M.E.N. mettent en place un dispositif national d'évaluation visant à préciser dans quelle mesure et dans quelle proportion les élèves atteignent les objectifs fixés par les I.O. aux différents paliers de la scolarité. Pour l'école élémentaire des évaluations ont été effectuées en 1979 en fin de C.P., en 81 en fin de CE2 et de CM2 et en 83 en fin de CM2.

Dans l'avant propos de la revue "Etudes et Documents" de mars 1980, C. SEIBEL et R. COUANAU exposent les objectifs et les caractéristiques du dispositif.

"Il a pour finalité première de fournir à la Direction des Ecoles des moyens d'apprécier aussi objectivement que possible en vue de l'améliorer, l'efficacité de l'enseignement et de l'action éducative de l'école; c'est en identifiant avec précision les insuffisances qu'on peut déterminer les opérations à conduire pour y remédier, émission d'instructions, actions de formation continue, aménagement de formation initiale."

L'évaluation porte sur :

- les apprentissages instrumentaux (langue française et mathématiques),
- l'éveil,
- les pédagogies pratiquées dans les classes.

Le dispositif a été conçu de manière à permettre:

1. de déterminer dans quelle mesure chacun des objectifs assignés à l'école élémentaire est atteint et par quelle population d'élèves;
2. de discriminer, dans les échecs relatifs constatés, ce qui tient:
 - à des facteurs "internes" (et notamment pédagogiques),
 - à des facteurs externes (environnement des écoles, des classes, des élèves, caractéristiques individuelles de ces derniers...").

La population de l'opération

EN 1979, un panel de 19938 élèves répartis dans 1390 écoles publiques et 3068 élèves de CP répartis dans 227 écoles privées a été constitué; il doit permettre de suivre la scolarité de tous les enfants du Panel jusqu'au CM2. En 1983, le dispositif a été allégé et ce sont 7631 élèves, 424 classes et 330 écoles qui seront étudiés.

Le dispositif

Les objectifs sont déclinés à partir des textes officiels de 1978 et 1980

Un questionnaire à destination des enseignants recueillaient des données sur

- la difficulté des exercices,
- les critères d'évaluation,
- les priorités mises en œuvre dans la classe.

Les résultats

Faute de pouvoir rappeler tous les résultats, nous avons choisi des types d'épreuves qui se retrouvent dans tous les protocoles suivants.

En fin de CE2, lecture silencieuse d'un texte de 140 mots

Un premier exercice, exercice jugé plutôt facile par les maîtres et très utile, voire indispensable par 91% des maîtres n'est réussi que par la moitié des élèves: il s'agit de repérer des informations importantes du texte, données littéralement

Un deuxième exercice sur le même texte, jugé lui aussi indispensable (95%) et moyennement difficile par 75% des maîtres, n'est réussi totalement que par 2% des élèves mais à 30% pour 7 réponses sur 9. Il s'agit de choisir parmi plusieurs propositions un titre, des interprétations simples et des interprétations qui impliquent des relations de cause / conséquence.

Les évaluateurs concluent : " *les questions qui demandaient une prise de renseignements ont été mieux réussies que celles qui demandaient une interprétation et un jugement. La présence effective des éléments de réponse dans le texte constitue un facteur de réussite. Le choix du titre dans la liste proposée montre la difficulté qu'éprouvent 41% des enfants à dégager l'idée essentielle d'un texte.* "

Récit : écrire un texte en racontant une histoire présentée par quatre dessins

Estimé indispensable par 47% des maîtres et plutôt difficile par 90% d'entre eux cet exercice appelle les commentaires suivants des évaluateurs:

- Equilibre du récit: environ 36% des élèves ne font pas référence à tous les dessins;
- Vivacité et cohérence du récit: un enfant sur deux est capable de varier plus ou moins les modes et les temps verbaux. Un sur cinq ne sait pas utiliser correctement les modes et les temps.
- Lisibilité et expression générale: environ un tiers des enfants produit un texte difficile à lire et à comprendre. On peut être inquiet du niveau d'écriture atteint alors que cet apprentissage devrait être achevé à la fin du CE2.

Distinction orale et écrite de phonèmes voisins

Cet exercice est jugé indispensable par plus de 84% des maîtres et facile par plus de 71%.

Les résultats correspondent à leur jugement et à leur attente, commentent les évaluateurs.

"Cependant 64% des enfants seulement réussissent totalement ou presque l'ensemble de l'exercice . Ces résultats montrent que la distinction orale écrite de phonèmes voisins n'est pas complètement maîtrisée.

Les erreurs les plus fréquentes correspondent

- à des difficultés phonétiques habituellement rencontrés
 - dessert/désert
 - gaine /gêne
 - croupe/groupe
 - enfouis/enfuis
- à une mauvaise perception de la position des lettres (reliure-reliure).

En fin de CM2

La réflexion pédagogique a évolué et l'analyse des résultats se présente sous deux formes:

- l'une en référence aux programmes formulés en terme d'objectifs généraux et d'objectifs spécifiques
- l'autre sous forme de compétences .

La compétence 1 regroupe les exercices qui demandent après lecture de différents type de message d'apporter une réponse pertinente, fondée sur un relevé d'information.

Les résultats sont très dispersés:

73% de réussite quand il s'agit d'identifier et de nommer cinq documents divers(lettre, horaires, annonce publicitaire, calendrier...); 43% quand il s'agit de sélectionner et de souligner dans deux textes de nature encyclopédique ("Le lièvre", "La marmotte")des informations sur la "transformation de ces animaux à l'approche de l'hiver".

La compétence II regroupe les exercices qui demandent aux élèves une restitution de connaissances ou de pratiques résultant de l'expérience ou d'un enseignement systématique dont l'acquisition est explicitement prévue dans les programmes.

Cette rubrique compte 7 exercices dont 6 de grammaire.

Les résultats montrent que certains "automatismes sont assez bien maîtrisés, comme la conjugaison (identification et reconnaissance des formes verbales) réussie à 90%, la pronominalisation à 76%."Des difficultés notables apparaissent dans la désignation de la nature ou de la fonction des mots."

La compétence III demande de mettre en œuvre une démarche logique d'organisation et de l'explicitier le cas échéant.

Un des exercices implique la mise en relation de plusieurs documents dont la lecture d'une fiche horaire: il est réussi par 51% des élèves.

Un autre concerne une activité d'expression écrite qui constitue la suite à la lecture sélective sur les transformations des deux animaux à l'approche de l'hiver;

La compétence IV permet de mesurer comment les élèves investissent les acquis dans des situations nouvelles.

La lettre et la rédaction d'un récit continué constituent deux des exercices regroupés sous cette compétence;

On constate dans l'écriture de la lettre, l'information essentielle, la date d'arrivée, n'est apportée que par 56% des élèves alors que 65% d'entre eux l'avaient repérée avec exactitude.

Le récit que l'élève est invité à continuer est extrait de "La gloire de mon père" de M. Pagnol

L'évaluation prend en compte

- la lisibilité générale(54% de réussite),
- la continuité du récit sous l'angle des marques linguistiques (55%) et la signification (80%),
- la cohérence interne du texte (65% pour les marques linguistiques et 52% pour le sens),
- la variété des discours utilisés: 46% pratiquent l'alternance récit-dialogue.

L'orthographe est évaluée dans le récit continué:

8% seulement des élèves ne font aucune erreur alors que 60% font au moins cinq erreurs.

Les erreurs sont rarement lexicales: 78% des élèves ne font pas plus de trois erreurs. La fréquence des erreurs grammaticales est plus élevée 45% ne font pas plus de trois erreurs de ce type, et 25% en font au moins sept.

La Compétence V apprécie la capacité à inventer et/ou manifester différents aspects de leur personnalité.

Ce classement montre que la notion de compétence a été introduite bien avant les évaluations et les programmes récents. Mais la définition et les contenus mêmes de ces compétences, en harmonie certes avec les conceptions dominantes de l'époque, semblent ne pas avoir été reprises dans des évaluations ultérieures.

Conceptions et critères d'évaluation des maîtres de CM2

Les conceptions de l'enseignement du français

Les questions posées aux maîtres ont fait apparaître des écarts que relèvent les analyses publiées (Pratiques pédagogiques Juin 1983):

1. Alors que dans les Instructions officielles de 1980 "parfaire la maîtrise de la lecture devient un impératif de première urgence", les enseignants des CM2 évalués n'ont pas classé la lecture parmi leurs trois objectifs prioritaires; ils ne la situent qu'après l'orthographe et le vocabulaire.

2. Alors qu'au CE2, les enseignants considèrent à plus des deux tiers l'apprentissage de la lecture comme le premier des trois objectifs prioritaires, au CM2 le développement "de comportements appropriés de lecture" n'est plus jugé prioritaire(12% classent cet objectif entête et 9% en fin de liste).

Des hypothèses sont formulées pour expliquer cette distorsion:

- une formulation insuffisamment explicite de cet objectif
- une représentation des enseignants de CM2 sur la maîtrise de la lecture: la majorité d'entre eux la considère comme ayant du être acquise au cycle précédent.

Ces évaluations de 1979 à 1983 ont permis quantité d'analyses de type sociologique sur les corrélations entre réussite et C.S.P., le niveau des redoublants, le choix des langues au collège ...mais on peut regretter qu'elles n'aient pas donné lieu à une diffusion large des résultats qui auraient permis

- aux enseignants de situer leurs pratiques et leurs conceptions par rapport à ce qui était analysé;
- aux cadres intermédiaires que sont les inspecteurs et leurs équipes d'intervenir sur des points ciblés, par exemple la maîtrise de la lecture au cycle moyen;
- aux formateurs, en particulier en formation continue(à l'époque, certains stages duraient trois mois) de travailler sur une meilleure appropriation des Instructions et les points faibles que révélaient ces évaluations.

II Une enquête en 87 vise à actualiser et à compléter ces données, selon un objectif différent: elle mesure non pas les compétences visées par le programme de CM2 mais cherche "à vérifier dans quelle mesure les élèves ont acquis les savoirs et savoir faire correspondant aux exigences du fonctionnement des classes de sixième."

Les écoliers savent-ils lire, écrire, compter?

La commande de 1987 est formulée de "façon rustique" dit J. Vogler qui introduit ainsi les son analyse (Education et Formation 1988 n° 14): *"On peut s'étonner que, en 1987, on ne dispose pas encore d'une réponse claire à cette question C'est que le problème est moins simple qu'il n'y paraît de prime abord. En premier lieu, tout dépend du degré de compétences à partir duquel on dira que tel ou tel élève sait lire et calculer. En second lieu, lire et calculer suppose la mise en œuvre d'un ensemble de compétences en interaction qui se manifestent différemment selon les situations dans lesquelles on les a sollicitées. Cela suffit à expliquer que les recherches effectuées aboutissent à des résultats difficiles à comparer et à généraliser."*

Les résultats en fin d'école élémentaire

Seuls la lecture, l'orthographe et le calcul sont évalués.

La lecture

Les objectifs étaient définis ainsi: "Evaluer la compréhension de textes divers(mais appartenant à l'univers scolaire) par des élèves placés en situation de lecture silencieuse et personnelle(sans rythme imposé)."

L'épreuve de lecture portait sur 8 textes de même longueur (une dizaine de lignes) de nature différentes: 2 textes littéraires, 2 textes scientifiques, 2 textes de caractère utilitaire.

Les questions portant sur la compréhension globale du texte sont réussies en moyenne à 72% et à près de 80% sur un texte narratif ("En classe" de Marcel Pagnol).

Le prélèvement ponctuel d'informations est réussi par 75% des élèves.

Les questions visant une compréhension plus fine, exigeant un traitement logique des informations prélevées sont réussies par 47% des élèves avec des résultats très différents selon les textes (90 à 20%). La complexité syntaxique et la connaissance générale du sujet paraissent être deux éléments déterminants le niveau de compréhension des élèves. Les opérations intellectuelles qui supposent un regroupement ou une reformulation d'informations contenues dans le texte sont réussies par environ la moitié des élèves. Celles qui exigent une interprétation du sens général pour inférer une information non explicite sont nettement moins bien réussies.

Jean Vogler conclut :

" La grande majorité des élèves (trois quarts environ) maîtrisent bien la "lecture de survol" et la lecture "sélective". Mais ils ne sont guère que la moitié à pouvoir répondre à des questions nécessitant une lecture plus approfondie, plus réfléchie."

L'orthographe

Les objectifs sont conformes aux I.O. de 1985:

- "Acquisition, emploi et reproduction sans faute sous la dictée ou de mémoire de mots usuels,
- initiation aux accords simples(pluriel, féminin, terminaison des verbes)"

La dictée choisie (7 phrases, 65 mots) contenait les deux phrases qui avaient constitué la dictée de l'évaluation de 1981. Ce qui permettait la comparaison à six ans d'intervalle.

"Les gamins se sont certainement perdus, dit maman. S'ils n'ont pas retrouvé leur chemin, nous les verrons arriver très fatigués à la maison."

Les résultats sont stables:

- les erreurs lexicales sont très rares,
- l'essentiel des erreurs relève de la non application de règles grammaticales: accords du participe passé, accord des adjectifs (entre 20 et 25% d'erreurs), accord sujet-verbe (10 à 20%) confusion participe passé -infinitif (43 et 21%).

L'analyse des variables

L'influence des variables donne lieu à des conclusions qui sonnent aujourd'hui comme des évidences:

- les filles ont des performances légèrement supérieures à celles des garçons,
- plus les élèves sont jeunes mieux ils réussissent,
- ils réussissent d'autant mieux qu'ils ont fréquenté l'école maternelle plus longtemps,
- le milieu socioprofessionnel d'origine a une influence importante sur la réussite scolaire.

III L'enquête de Janvier 1989

En janvier 1989 la Direction des Ecoles a initié une expérimentation visant à étudier les effets des bibliothèques d'écoles sur les pratiques de lecture. Cette opération concerne 60 écoles et 5768 élèves. Elle comportait entre autres prises d'informations, une évaluation des compétences de lecture des élèves. Cette évaluation des compétences de lecture fournit des indications sur l'évolution des apprentissages durant le cursus scolaire et porte donc sur les quatre niveaux au-delà du C.P.

Le dispositif

- L'évaluation est centrée sur la compréhension en conformité avec les I.O. de 1985: "L'apprentissage de la lecture doit faire l'objet d'une acquisition systématique permettant la compréhension d'écrits de plus en plus complexes."
- Les écrits proposés sont diversifiés: contes et récits, textes scientifiques ou historiques, écrits sociaux (recette, annuaire, calendrier...)

Plutôt que de déterminer une sorte de niveau moyen d'exigence, on a cherché "un échelonnement dans la difficulté des items en fonction de :

- la nature de l'écrit,
- la quantité d'information à traiter,
- l'activité de compréhension sollicitée: ("il est plus aisé de repérer des noms dans un annuaire que d'indiquer le sujet général d'un texte"),

pour permettre de positionner les élèves à différents degrés de compétences.

Le dispositif comporte des épreuves identiques d'un niveau à l'autre, et permet ainsi d'observer l'évolution des compétences réalisée en une année par les élèves.

La population : 1178 élèves de CE1, 1156 de CE2, 1080 de CM1 et 1051 élèves de CM2

Les résultats

- En CE1, on distingue trois niveaux de compréhension
 - compréhension de phrase plus de 70%,
 - supérieure à la phrase 50 à 55%,
 - compréhension de texte 40 à 50%.

Près de deux tiers des élèves sont capables de comprendre le sens général d'un conte;

Le repérage d'informations est bien maîtrisé dès lors qu'il ne nécessite pas de discrimination fine. Quand les activités de repérage portent sur deux mots qu'une seule lettre distingue, ou quand il faut isoler des mots qui ne sont pas familiers, les erreurs apparaissent: ainsi dans ce dernier item, 60% des élèves isolent moins de 9 mots sur 17.

- Du CE1 au CE2 les performances des élèves s'améliorent nettement et en particulier ceux qui supposent la mise en œuvre d'une discrimination visuelle fine.
- Au CE2, mobiliser des informations écrites pour effectuer une tâche met en difficulté un élève sur deux. Près de deux tiers des élèves sont capables d'appréhender le sens global d'un texte et de prélever des informations. En revanche, les taux de réussite baissent nettement dès qu'il s'agit d'interpréter, d'inférer à partir d'éléments du texte, une information qui n'est pas donnée explicitement (9%).
- Du CE2 au CM1, les progrès se font essentiellement dans la compréhension du sens général d'un texte; les questions exigeant une interprétation du sens restent difficiles et ne donnent lieu qu'à une très faible progression d'une année sur l'autre.
- Du CM1 au CM2 les progrès sont significatifs; cette progression concerne tous les élèves même ceux qui réussissent moins bien. Ces progrès varient selon :
 - l'activité de compréhension sollicitée: les réponses demandant une compréhension fine continuent de poser problème. elle demande la mise en œuvre d'une lecture "précise et raisonnée" dont ne semble être capable qu'un élève sur deux. au CM2,
 - le degré de complexité syntaxique du texte.

Jean VOGLER qui rapporte et analyse tous ces résultats dans la revue "Education et Formations" ce décembre 89 conclut: « *Sont généralement bien réussies les activités relevant d'une lecture que l'on peut qualifier de "superficielle": repérage simple d'éléments d'informations uniques, mises en relations terme à terme, relevé ponctuels d'informations. Dès que l'activité de lecture s'affine la*

réussite est moindre mais progresse régulièrement... en revanche, les activités relevant d'une lecture approfondie et plus raisonnée posent encore problème au CM2. La moitié environ des élèves aborde l'enseignement du collège sans avoir une maîtrise suffisante de ce niveau de lecture."

L'analyse des résultats de cette enquête permet de repérer ce qui est "facile" pour la plupart des élèves d'un niveau, ce qui est en cours d'acquisition et qui reste difficile en fin de scolarité primaire; elle aurait pu être complétée pour permettre de construire des échelles de compétences comme le font certaines évaluations internationales.

Conclusion.

Entre 1979 et 1989, ont été conduites des évaluations de qualité qui visaient toutes à apprécier la mise en œuvre de nouveaux programmes, à mesurer l'efficacité du système.

1. Sur les compétences des élèves, elles n'ont pas été conçues pour être comparées mais elles font un certain nombre de constats qui sont encore, on le verra, pour certains vrais et concluent dans le mêmes sens:

- En lecture,
 - les correspondances phonies/graphies ne sont pas encore maîtrisées en fin de CE1, ni même en fin de CE2,
 - la compréhension de phrases, de textes simples est acquise par la très grande majorité des élèves de CE2 mais elle dépend du type de questions posées, du type de raisonnement demandé, de la complexité syntaxique du texte, du thème et du type de texte. La moitié des élèves abordent la sixième sans avoir la maîtrise d'une lecture "raisonnée".
 - La connaissance de la langue est encore incertaine au CM2 ;
 - 60 à 70% des élèves sont capables de rédiger à peu près correctement un texte répondant à des consignes simples, en utilisant les temps du récit, en restant cohérent par rapport au sens,
 - En orthographe les erreurs sont plutôt grammaticales que lexicales.

2. L'importance des panels observés a donné lieu à des analyses de type sociologique, abondamment reprises et confirmées depuis.

3. Grâce aux questions posées aux enseignants, on disposait de données importantes sur les conceptions et les pratiques des maîtres sur la lecture en particulier; certes, elles n'étaient pas suffisamment précises pour dégager des types de pédagogie ; mais à les relire, en 2003, on regrette qu'elles n'aient pas été diffusées largement pour permettre que l'ensemble de la communauté éducative s'en saisisse.

On disposait aussi de données sur les seuils de compétences qui elles non plus n'ont pas été utilisées.

4. La question des élèves en échec n'est posée, ni dans les questionnaires aux maîtres, ni dans les analyses pour l'institution. Seuls les redoublants sont observés quant à leur niveau, à leur orientation ou à leur origine sociale.

On a l'impression que beaucoup d'énergie, d'intelligence et de travail ont été mises en œuvre et que peu d'enseignements pour améliorer les résultats en ont été tirés.

Ces conclusions rejoignent celles du Haut Conseil de l'évaluation de l'école (Avis n°3 octobre 2001: "La variété des dispositifs et la richesse de leurs résultats contrastent avec la faiblesse de leurs usages."

Les évaluations-bilan dans le second degré

Evaluations de fin de cinquième, en 1982/1991-1992

Un dispositif d'évaluation a été mis en place en 1980 pour évaluer les élèves à l'entrée en sixième (1980), deux ans après à la fin de la cinquième c'est-à-dire du cycle d'observation d'alors (1982), et en fin de troisième (1984).

Le protocole de 1982 de fin de cinquième comprenait vingt exercices (cent soixante-deux items). Les domaines évalués portaient largement sur le maniement de la langue (formation des mots, syntaxe, orthographe, invention, ponctuation, analyse et repérage des fonctions dans la phrase), sur l'appréhension du sens dans l'écrit (sens des mots et des expressions, sens global du texte), du langage poétique, de l'expression écrite (une rédaction de dix lignes). Il reprenait certains items du protocole de 1980 à l'entrée en sixième.

Dans le maniement de la langue, les réussites (de 60% à 90%) portaient sur la maîtrise de la conjugaison (malgré une difficulté à orthographier le participe passé avec avoir), les transformations de phrases (négative et, à un moindre degré, passive), l'emploi des conjonctions adversatives (*mais*, *pourtant*), celui des comparatifs et superlatifs. Mais moins d'un élève sur deux maîtrisait la composition des mots (terme de base, préfixe et suffixe), l'analyse et l'identification des fonctions dans la phrase, l'emploi des conjonctions de cause et conséquence. Moins de 30% maîtrisaient la ponctuation et le langage poétique (versification, rythmes...).

En lecture, les résultats en compréhension de mots ou en compréhension globale (titre d'un texte, 33%) étaient moyens ou médiocres.

En écriture, la grande majorité des élèves avaient produit le récit demandé, en respectant la cohérence des temps. Près de 50% utilisaient un vocabulaire pertinent. Les taux les plus bas (les deux-tiers des élèves) concernaient l'orthographe grammaticale. Les items repris du protocole de sixième montraient une progression de la sixième à la cinquième.

Un tel protocole permettait de faire un état assez large des compétences des élèves à la fin du cycle d'observation d'alors, et sans ruptures avec le protocole de sixième deux ans avant. On ne peut en comparer les résultats avec ceux du protocole de 2002 d'entrée en cinquième. En effet l'évaluation de 2002 en cinquième, se situant à l'entrée du cycle central actuel, est construit sur une tout autre logique.

Un second dispositif d'évaluation sur trois ans a été conduit de 1989 à 1992. Deux études ont conduit à faire passer le même test à un même échantillon d'élèves, à l'entrée en sixième puis, deux ans plus tard, en fin de cinquième en prenant en compte les redoublants (octobre 1989 / mai 1991, octobre 1990 / mai 1992). Les items communs portaient sur l'orthographe, la grammaire et le vocabulaire. Les résultats ont montré des progrès sensibles dans les trois domaines abordés. La proportion moyenne de bonnes réponses est passée de 37% à 51%. «La proportion des élèves réussissant les items proposés s'accroît de 19 à 34 points» pour les items de la première étude, de 13 points en moyenne pour les items de la deuxième (NI n° 93-37 de septembre 1993). En sixième, les élèves avaient été répartis en quatre groupes : faibles, moyens-faibles, moyens-forts, forts. Quelle que soit la catégorie de départ, 40 % des élèves ne changent pas de catégorie à la fin de la cinquième. Les plus faibles ont progressés plus vite, les plus forts se sont maintenus (DT n° 409). Faute d'items communs au protocole de fin de cinquième de 1988 et de 1991, l'évolution n'a pas pu être mesurée. La sixième et la cinquième constituaient alors le cycle d'observation, réduit aujourd'hui à la classe de sixième. Mais l'évaluation de 1992 permet de mesurer les progrès des élèves au début du collège, ce que celle de 2002 n'a généralement pas permis.

Les évaluations nationales en fin de troisième.

3°G = troisième générale, 3°T= troisième technologique. LEP= lycée d'enseignement professionnel.

Les dispositifs mis en place.

En mai-juin 1995, les élèves de troisième des établissements publics, troisième générale et troisième technologique de collège ou de lycée professionnel, ont été évalués. 9,1% des élèves de troisième étaient alors en troisième technologique, classes anciennes en lycée professionnel mais ouvertes

également en collège une dizaine d'années auparavant. Quelques temps après leur ouverture en collège, en mai-juin 1988, une évaluation des élèves de troisième technologique (88 établissements, lycées professionnels et collèges), avait déjà été mise en place, où les résultats étaient confrontés, quand c'était possible, avec les scores des élèves de troisième générale de 1984. D'autre part, en 2003, a eu lieu une évaluation des *compétences générales* de tous les élèves en fin de troisième générale qui fonctionne comme un bilan du collège mais dont les résultats ne sont pas encore analysés.

Chaque fois, le protocole d'évaluation a été élaboré par des groupes de travail spécifiques, avec des phases d'expérimentation. Mais les protocoles sont très différents. En 1988, on évaluait la compréhension d'un court texte informatif, la richesse du vocabulaire, les registres de langue, la maîtrise de quelques règles de grammaire, soit douze compétences toutes jugées importantes par les professeurs consultés ; mais, au grand regret de ces mêmes professeurs, on n'évaluait ni la production écrite ni l'aptitude à l'argumentation. En 1995, le protocole comprenait 22 items d'orthographe, 22 items de grammaire, 41 items de compréhension sur trois textes de type différent (un extrait de roman contemporain, une fable de La Fontaine, un ensemble de petites annonces), et deux exercices d'écriture.

Les résultats des élèves.

En 1995, pour tous les élèves de fin de collège, les résultats sont meilleurs en orthographe qu'en lecture, et meilleurs en orthographe et en lecture qu'en grammaire. Dans ce dernier domaine, reconnaître et identifier les éléments et les fonctions dans la phrase est mieux maîtrisé qu'expliquer et interpréter les phénomènes grammaticaux (l'emploi d'un temps ou d'un mode, la valeur d'une transformation, par exemple).

	orthographe	compréhension	grammaire
3° générale	77%	72%	62%
3° tech. collège	56%	53%	34%
3° tech. LEP	49%	46%	29%

En lecture, tous les élèves savent (à hauteur d'environ 79% en 3° générale, 61% en 3°T de collège, 57% en 3°T de lycée professionnel) :

- repérer des indices concernant le temps, le lieu, des événements et des personnages,
- identifier le narrateur et les personnages qui parlent (récit littéraire),
- rétablir l'intégralité d'un message informatif (petite annonce),
- rechercher et traiter l'information (ce qui n'est souvent pas acquis à l'entrée en 6°).

Les écarts de réussite sont importants entre les différents types de classe (de 24 à 38 points selon la compétence), surtout pour la lecture d'un texte littéraire, qui demande d'interpréter.

Les élèves de 3° technologique ont des difficultés, avec des écarts parfois de 30 à 50 points entre leurs résultats et ceux de la 3° générale, à :

- identifier le personnage désigné par un mot ou un groupe de mots (plus de sept élèves sur dix en 3° générale, mais moins d'un élève sur deux en 3° technologique),
- rétablir l'ordre chronologique d'un texte (un peu plus de six élèves sur dix en 3°G, mais moins de 20% en 3°T.),
- reconnaître le genre d'un texte (sept élèves sur dix en 3°G., mais de 40 à moins de 30% en 3°T.),
- analyser des discours descriptifs ou injonctifs (sept à huit élèves sur 10 en 3°G., mais de quatre à six en 3°T de collège et de trois à quatre en 3°T de LEP),
- trouver le sens de mots dans le texte (moins d'un sur deux en 3°G., environ un tiers en 3°T de collège, moins de 25% en 3°T de LEP).

Parmi ces dernières compétences, les deux premières sont déjà difficiles pour les élèves entrant en sixième sur des textes à leur niveau ; quant à la reconnaissance du genre et du sens d'un mot dans son contexte, bien réussie à l'entrée en sixième, elle devient difficile en troisième dans une situation plus complexe.

NB. Notons que, sur les 21 items de lecture communs aux évaluations de 1990 et de 1995 pour les troisièmes générales, les scores baissent très légèrement en 1995 (76% de réussite en 1995 pour 78% en 1990).

En 1988, les résultats des élèves de troisième technologique étaient satisfaisants en lecture (de

70 à 80% de réussite en comptant les codes 1 et 2) sur un texte dont il fallait identifier l'idée essentielle, une cause et un effet, questions courantes à l'entrée en sixième. Ils étaient encore plus d'un sur deux à savoir reconstituer une phrase (non un texte) dont les éléments étaient proposés en désordre.

Pour tous les élèves de troisième, en revanche, les résultats étaient modestes (moins de 50% en troisième générale, moins de 40% en troisième T de collège, moins de 30% en troisième T de LEP) sur un exercice demandant d'utiliser la ponctuation pour faire apparaître le sens différent de deux phrases identiques ; ils étaient très faibles (de 30% à 25%) pour le même exercice quand une différence orthographique distingue les deux phrases *forts /fort*). Ces deux derniers exercices montrent à quel point, en 3^e, les élèves ont encore du mal à être attentifs au détail de la langue pour aller au sens.

En orthographe, pour tous les élèves, les difficultés sont plus importantes en orthographe grammaticale qu'en orthographe lexicale, ce que confirme leur faiblesse générale en grammaire.

	orth. grammaticale	orth. lexicale
3 ^e générale	68%	86%
3 ^e tech. Collège	44%	71%
3 ^e tech. LEP	36%	66%

Comme permettait de le voir l'exercice proposé, sept à huit élèves de troisième générale peuvent repérer les erreurs de syntaxe et de langue dans un texte, ce qui suppose qu'ils ont les moyens d'orthographier correctement leur propre texte, savoir faire essentiel en production écrite ; mais ce n'est pas le cas pour les élèves de troisième technologique.

Les résultats montrent que tous les élèves de troisième savent recopier un texte sans erreurs et accorder les noms au pluriel. En troisième générale, de six à huit élèves sur dix savent également accorder l'adjectif en genre et distinguer les homonymes grammaticaux (*ça / sa ; ni / n'y ; ...*). Ils ont donc acquis des compétences par rapport aux résultats habituels à l'entrée en sixième. Mais ce n'est pas le cas en troisième technologique : de 35% à 45% y réussissent en troisième T de collège, de 29% à 37% en troisième T de LEP.

Ce qui est le moins réussi, pour tous les élèves de troisième, concerne la conjugaison (moins de 50% en troisième G., 20% en troisième T de collège, 14% en troisième T de LEP). C'était déjà le cas en 1988, où l'on notait en particulier, chez les élèves de troisième technologique, une méconnaissance de l'emploi des temps de l'indicatif qui semblent n'être reconnus que par une idée imprécise de leurs terminaisons (d'où 40% d'élèves qui confondent l'imparfait et le futur), et la méconnaissance du subjonctif. Notons qu'aux évaluations de 1984, les élèves de troisième générale avaient des résultats globalement honorables pour la connaissance des formes verbales.

En grammaire, 65% des élèves de troisième générale reconnaissent les différentes sortes de mots dans la phrase, identifient les formes verbales, les modes et les fonctions. Ils sont capables également d'effectuer avec succès des transformations qui portent sur le texte. Ils ont donc acquis les connaissances grammaticales élémentaires. Mais ils ont des difficultés à commenter, analyser, justifier les identifications ou les transformations qu'ils effectuent portant avec succès (82% des élèves identifient par exemple un subjonctif sans pouvoir expliquer son emploi). La réflexion sur la langue est encore en cours d'acquisition en fin de collège.

En troisième technologique, de collège comme de lycée professionnel, les fondamentaux de la grammaire ne sont pas acquis (les substantifs ne sont pas toujours repérés). Pour la reconnaissance et l'identification des formes, les écarts avec la troisième générale sont de 28 à 34 points.

Le vocabulaire n'a pas été évalué en 1995, sinon par l'orthographe lexicale, qui montre une bonne réussite pour l'ensemble des élèves. En 1988, l'ensemble des élèves (de 70% à 90%) savait reconnaître si un mot relève d'une langue soutenue, courante ou familière. La connaissance du sens des mots et de leur emploi dans un texte était variable, avec une difficulté plus sensible sur les mots polysémiques.

En production écrite, les scores de réussite rejoignent globalement ceux de la lecture. Ils montrent que les élèves de troisième technologique sont plus habitués à produire des écrits d'ordre

pratique (rédiger une petite annonce) que des écrits de type littéraire (texte à partir d'une fable de La Fontaine).

	3° générale	3° tech. Collège	3° tech. LEP
expr. écrite	72%	49%	40%
petites annonces	74% *(1990:80%)	58%	53%
score global de production	73%	53%	46%
score général	71%	50%	43%

* Cet exercice figurait à l'évaluation des troisièmes générales en 1990, avec un score nettement meilleur. En 1995, 13% des élèves n'ont produit aucun texte, contre 2% en 1990.

Les élèves de troisième générale sont plus de sept sur dix à produire un texte de plus de quinze lignes, cohérent et montrant une bonne maîtrise des temps verbaux. Les difficultés portent sur la structuration du texte en paragraphes (45%) et sur l'utilisation, pour le développement du sujet, de toutes les propositions de la consigne (37%).

Les élèves de troisième technologique sont nombreux (31% en collège) à n'avoir pu écrire de texte de plus de cinq lignes ; leurs difficultés portent surtout sur la cohérence interne du texte qu'ils élaborent et sur les règles qui régissent, à l'intérieur de ce texte, les types de discours (narration, description).

Conclusion.

Par l'ampleur des domaines de compétences explorés par le protocole, l'évaluation de 1995 montre que les élèves de troisième générale, malgré une faiblesse relative en conjugaison et en grammaire, maîtrisaient globalement les savoirs et savoir-faire de base, ce qui les rendait capables d'être vigilants sur la correction linguistique de leur propres productions écrites. Ils avaient de bons résultats en lecture, surtout quand elle est guidée et que la situation n'est pas trop complexe, et en production écrite. La réflexion sur la langue par le commentaire ou l'analyse, moins bien réussie que la reconnaissance ou la simple transformation, est une compétence en cours d'acquisition en fin de collège, comme la structuration du texte en paragraphes. On note une faible baisse des résultats en 1995 par rapport aux évaluations de 1984 ou de 1990.

La principale difficulté des élèves de troisième technologique portait sur la maîtrise des éléments fondamentaux de la langue, surtout en grammaire. S'ils réussissaient à répondre convenablement à des questions sur un texte, ils étaient démunis pour en saisir le sens global ou la structure logique. En écriture, s'ils réussissaient convenablement à produire les textes d'ordre pratique demandés, ils avaient du mal à élaborer un écrit de type littéraire d'une certaine longueur et à se représenter les règles qui le régissent.

Notons qu'une étude, fondée sur l'évaluation de 1984, a observé en juin 1987 les données sur lesquelles étaient orientés les élèves de fin de collège dans les différentes sections de seconde existant alors. On y relève le rôle que jouent l'acquisition de connaissances linguistiques et culturelles, en particulier grammaticales. Mais cette étude énumère surtout, dans ce dernier domaine, les compétences qui n'ont pas de poids sur l'orientation soit parce qu'elles sont acquises par tous, soit parce qu'elles ne sont généralement pas maîtrisées. Elles correspondent globalement, pour ce qui est comparable, aux résultats de l'évaluation de 1995.

Les évaluations spécifiques et comparatives

Les évaluations spécifiques

L'évaluation des compétences de élèves au cours préparatoire

L'étude des compétences des élèves à l'entrée au CP a été réalisée à partir d'un panel d'écoliers recruté en 1997. La DPD a mis ce panel en place pour mieux connaître le déroulement des parcours scolaires. L'échantillon était constitué de 10 000 élèves scolarisés en 1997 dans 1570 écoles publiques et privées de France métropolitaine. Il était prévu que ces écoliers soient suivis jusqu'à la fin de la cinquième.

L'étude des compétences et des comportements de ces élèves doit fournir une description initiale pour mieux mesurer l'action de l'école et pour déterminer l'influence de ces compétences testées sur la réussite. Des équipes d'universitaires avaient élaboré les épreuves, analysé et commenté les résultats. Ce sont des données techniques, parfois éloignées des pratiques pédagogiques et de la culture des maîtres et même sans doute de la plupart de leurs formateurs, qui fournissent un inventaire étalonné de certains aspects de certaines compétences : compréhension orale, compétences verbales et familiarité avec l'écrit ; en , lecture : conscience phonologique, conscience syntaxique, attention et mémorisation. Ces compétences ont été mesurées à deux moments différents de l'année en grande section et au CP.

Ces épreuves ont permis de construire les outils d'évaluation pour la grande section et le CP qu'il est recommandé de faire passer en début d'année depuis 2001; diffusées sous la forme d'un CD-ROM , leurs résultats ne sont pas collectés nationalement mais permettent aux enseignants d'analyser les performances de leurs élèves et de comparer leurs scores avec les résultats moyens collectés auprès des enfants de la cohorte. Elles sont difficiles à mettre en œuvre car elles supposent des moments de passation individuelle, ce qui est difficile dans des classes de parfois 30 élèves en maternelle, et qui obligent donc à une organisation spécifique au niveau de l'école. Elles ne sont pas faciles non plus à analyser, mais permettent justement d'alerter les maîtres sur des aspects qu'ils n'ont pas l'habitude d'évaluer comme l'attention, le vocabulaire, la bonne discrimination phonologique.

Cette étude a aussi aidé à concevoir les cahiers de "prévention de l'illettrisme" pour les CP allégés(2002), avec plusieurs prises d'informations (dont une en septembre 2003) et qui pour l'instant n'ont pas encore été analysées complètement.

La mise en relation des observations faites en CP avec les résultats de l'évaluation CE2 permet de dégager des enseignements sur les compétences qui sont prédictives des réussites en lecture; ainsi, les capacités d'attention, d'analyse phonologique apparaissent comme fondamentales à exercer pour préparer les élèves de l'école maternelle à l'apprentissage de la lecture.

Etude spécifique relative aux difficultés de lecture à l'entrée en sixième

A la demande de l'Observatoire National de la Lecture, en 1997, une étude spécifique a été conduite par la DEP. Elle visait à mieux comprendre la nature des difficultés des élèves qui ne maîtrisaient pas les compétences de base à l'évaluation nationale d'entrée en sixième. Ont été retenus, les élèves n'ayant pas répondu de manière satisfaisante à au moins 9 des 12 items répertoriés parmi ces compétences.

Ce protocole est essentiellement axé sur l'évaluation du système d'identification des mots. Il est composé de huit exercices testant successivement : les habiletés attentionnelles des élèves, la voie indirecte d'identification des mots, la sensibilité phonologique et la maîtrise orthographique, la voie directe d'identification des mots, la compréhension d'énoncés écrits, la capacité à utiliser la morphologie pour reconnaître des familles de mots, et enfin, l'aptitude à utiliser le contexte pour pallier les difficultés d'identification des mots. La plupart de ces exercices devaient être réalisés dans un temps limité.

Analyse des résultats de la population soumise à l'étude

Au total, la population représentait 14,9 % des élèves de sixième en 1997. L'étude a mis en évidence l'existence de trois groupes qui se différencient assez nettement par le score obtenu à l'évaluation nationale. Les pourcentages qui seront exprimés par la suite le seront par rapport à la population générale des élèves de sixième ayant participé à l'évaluation nationale de sixième.

Les élèves en très grande difficulté. Ils sont 4,3 % dans ce cas et leur situation est particulièrement préoccupante, car ils cumulent des retards dans des domaines élémentaires de l'apprentissage de la lecture. A l'évaluation nationale, ces élèves ont réussi significativement moins d'items que les autres élèves en difficulté. Deux sous groupes homogènes ont été identifiés :

- Le groupe A est constitué de 2,2% de l'ensemble des élèves de sixième. Ces élèves ont des lacunes graves dans tous les domaines indispensables à une bonne maîtrise de la lecture. Dans ce groupe, les garçons sont sur-représentés (80% contre 64,2% de l'ensemble des élèves en difficultés), la proportion d'élèves accusent un retard scolaire est plus importante que dans les autres groupes.
- Le groupe C représente 2,1 % des élèves. Leurs performances sont nettement moins bonnes que celles des élèves en difficultés pour : l'utilisation de la voie indirecte d'identification des mots, les capacités phonologiques, l'orthographe et l'utilisation de la morphologie et la délimitation de la frontière des mots.

Le groupe B. Il est composé de 7,8 % des élèves qui se caractérisent par une extrême lenteur. Ils sont, à la fois, plus lents que l'ensemble des élèves de sixième, mais aussi que les élèves présentant des difficultés de lecture. Leur lenteur est compensée par leurs bonnes performances d'identification des mots (d'ailleurs très proche de celles de l'ensemble de la population), par leur capacité à utiliser la morphologie et à délimiter la frontière des mots. En revanche, leurs performances sont équivalentes à celles des autres élèves en difficultés de lecture en ce qui concerne l'orthographe, les capacités phonologiques et la capacité de compréhension des énoncés.

Le groupe D. Il regroupe les 2,8 % dont la situation est moins préoccupante. Ils paraissent maîtriser les apprentissages fondamentaux nécessaires à la mise en œuvre de l'acte de lire, mais leurs problèmes se situent sur des compétences de niveau un peu plus élevé. Ils sont dotés d'une vitesse de traitement supérieure à l'ensemble des élèves de sixième. S'ils ont de meilleures compétences que les trois autres groupes en phonologie, en orthographe et pour identifier les mots, ils présentent des difficultés analogues pour comprendre les énoncés, utiliser la morphologie et pour délimiter la frontière des mots.

Une analyse des liaisons entre les réussites à ce protocole spécifique et le protocole national de français sixième a été conduite. Les élèves des Groupes A et C les plus en difficulté présentent un profil identique de résultats aux trois types de compétences classiquement étudiées dans les évaluations nationales jusqu'à une date récente. Ils réussissent environ 55% des items relevant des Compétences de base, 37% pour les Compétences approfondies, et 23 % pour les Compétences remarquables. Le groupe B fait preuve de davantage de maîtrise qu'eux des Compétences approfondies (44%), et des Compétences remarquables (28 %). Enfin le groupe D identifié comme celui qui est le moins en difficulté a un profil différent des trois autres groupes : il réussit environ 62% des items relevant des Compétences de base, 50 % pour les Compétences approfondies, et 12 % pour les Compétences remarquables.

En résumé, cette étude a permis de mettre à jour les spécificités des élèves présentant les plus grandes difficultés en lecture à l'orée de la classe de sixième. Cependant, aucune orientation pédagogique n'est proposée pour prendre ces élèves en charge, pour tenter de combler leurs faiblesses.

Les évaluations comparatives

La lecture en CM2

Comparaison des résultats en lecture des élèves en fin de CM2 à dix ans d'intervalle (1987 – 1997)

En juin 1987, la DEP avait mis en place un dispositif pour évaluer les acquis en français et en mathématiques des élèves de fin de CM2 afin de vérifier dans quelle mesure les élèves avaient acquis les savoirs et savoir-faire correspondant aux exigences du programme de sixième. Des épreuves standardisées portant sur la lecture, l'orthographe et la grammaire avaient été élaborées pour évaluer le français. Cette évaluation a été reprise en juin 1997 afin d'apprécier l'évolution à dix ans d'intervalle de compétences en lecture d'élèves de fin de CM2 et d'estimer les effets sur la lecture de diverses politiques introduites au cours de ces dix années : mesures en faveur du livre et de la lecture, plans « Maîtrise de la langue », introduction des cycles à l'école, limitation des redoublements, etc.

La population : 1 994 élèves en 1987 et 3108 en 1997 ont passé l'épreuve de lecture, constituée de huit textes presque tous de « type explicatif » chacun étant suivi de cinq questions à réponses construites, destinées à en estimer la compréhension. En 1997, les élèves ont passé une épreuve supplémentaire composée d'exercices provenant de protocoles nationaux de CE2 et de sixième.

Comparaisons globales

A dix ans d'intervalle, les élèves de fin d'école primaire atteignent des performances de compréhension très proches 65, 6 % vs. 66, 1 %, leurs dispersions étant respectivement de 14,8 et de 15,3. Sauf indication particulière, nous donnons d'abord les résultats obtenus en 1987, puis ceux de 1997. Le pourcentage de non réponses tend à s'accroître (6, 5% vs 8, 5 %) ; de manière corollaire, les réponses erronées diminuent (27,9 vs 25,4%).

Comparaisons partielles

Une discussion figure dans le rapport dédié à cette étude (page 40 et suivantes). Elle fait état de la grande difficulté à reprendre les terminologies de classification « des types de textes » établies en 1987 et le « regroupement des questions selon le type de compétences sollicitées » (p. 42). Par exemple, une question est classée comme estimant la compréhension globale alors qu'elle ne nécessite que d'effectuer un prélèvement d'information dans le texte. Une première analyse des résultats a cependant été menée en conservant le regroupement de 1987 ; elle donne des résultats identiques à dix ans d'intervalle pour les trois types de compréhension (globale, ponctuelle et fine). Face aux problèmes évoqués et comme cette évaluation conjugue la comparaison des résultats avec l'appréciation des capacités des élèves au moment où ils entrent en classe de sixième, le contenu de chaque item a été soumis au tableau de compétences et aux critères utilisés pour construire l'évaluation nationale en français à l'entrée en sixième. Cependant, certaines questions se classent difficilement dans cette typologie car elles combinent plusieurs « types de compétences ». Des solutions ont été trouvées pour rendre compte, au mieux, des niveaux de compréhension effectivement évalués en décidant de la « dominante » de la question. Les difficultés de classement dont nous faisons état mettent en lumière combien les dispositifs d'évaluation sont contraints par les choix qui ont été faits lors de leur construction.

Au final, trois niveaux de compréhension ont été retenus :

- La compréhension immédiate relève du simple relevé d'informations.
- La compréhension logique nécessite de comprendre l'enchaînement des phrases, l'organisation des textes et le rôle des pronoms ou des déterminants.
- La construction d'informations et de significations sur un texte nécessite que l'élève traite l'implicite du texte et élabore des inférences.
-

On remarque entre les pages 49 et 50, le glissement de « *type de compréhension* » à « *type de compétences* », sans qu'une justification en soit vraiment apportée, nous adoptons cette terminologie pour présenter les résultats suivants.

Le tableau ci dessous présente de manière synthétique les pourcentages moyens de réussite des élèves de CM2 en 1987 et en 1997, selon « la compétence dominante ».

« Type de compétences »	1987	1997
Compréhension immédiate – 22 questions	75, 7	76, 2
Compréhension immédiate seule – 14 questions	79, 1	80, 3
Compréhension immédiate et compréhension logique - 6 questions	70,3	69, 2
Compréhension immédiate et construction d'informations et de significations sur un texte – 2 questions	67, 4	68, 6
Construction d'informations et de significations – 14 questions	49, 9	50, 3
Construction d'informations et de significations seule – 9 questions	52, 8	53, 7
Construction d'informations et de significations et compréhension logique – 5 questions	44, 5	44, 1
Pourcentages moyens – 40 questions	65, 6	66, 1

On peut remarquer que le niveau de compréhension appelé « *Compréhension logique* » n'est jamais évalué seul. Ce tableau entre la grande stabilité des résultats au deux moments étudiés ; cette stabilité se retrouve aussi bien dans les résultats des garçons (65, 6 vs 65, 5 %) que dans ceux des filles (65, 7 vs 66, 8 %). Au total, en fin de CM2, les trois quarts des questions faisant appel à *La compréhension immédiate* sont réussies contre la moitié des questions mettant en jeu *La construction d'informations* à partir du texte et, en conséquence, des opérations cognitives plus complexes.

L'âge est un facteur déterminant de réussite aux épreuves de compréhension proposées, comme le montre le tableau ci dessous.

Pourcentages moyens de réussite des élèves des deux cohortes, selon leur cursus scolaire.

	1987	1997
En avance	73, 8	71, 9
A l'heure	69, 5	68, 5
En retard	57, 4	55, 2

Pour les deux cohortes, les résultats des élèves en avance sont meilleurs que ceux qui sont « à l'heure », eux – mêmes supérieurs à ceux des élèves présentant au moins un an de retard à l'entrée en CM2. Dans l'intervalle de cette étude, des mesures ont été prises pour limiter le nombre des redoublements. Aussi, on constate une diminution importante du pourcentage d'élèves en retard en CM2 : passage de 33 % dans l'échantillon de 1987 à 19 % dans celui de 1997. Les résultats des redoublants doivent être observés de près. En effet, les élèves présentant un retard d'un an en 1997 ont des performances très proches de celles des redoublants en retard de deux ans en 1987, et significativement plus faibles que la population 87 ayant le même retard scolaire qu'eux en termes d'années. Autrement dit, les populations de redoublants de CM2 se sont modifiées et l'élève redoublant en 1997 en classe de CM2 présente probablement plus de difficultés que celui de 1987 n'ayant, comme lui, redoublé qu'une fois. L'augmentation du nombre d'élèves « à l'heure » et la réduction des élèves en retard se traduisent par un maintien des scores de compréhension, aussi peut – on conclure que la politique de limitation des redoublements n'a pas eu d'effets négatifs sur les performances en lecture comparées à dix ans d'intervalle.

Une analyse a été menée sur les résultats des élèves selon qu'ils se situent dans le groupe des 10% les meilleurs ou dans celui des 10 % les plus faibles.

Comparaison des pourcentages moyens de réussites des élèves les plus faibles et les plus forts des deux cohortes.

Cohortes	1987		1997	
	faibles	fort	faibles	fort
Compréhension immédiate CI	47,7	94,5	48,3	93,8
Compréhension immédiate seule	54,6	95	55,8	95,2
C.I. et compréhension logique	37,3	93,2	35,9	93,6
C.I. et construction d'informations et de significations	30,7	94,8	33	93,4
Construction d'informations et de significations	23,1	80,5	22,6	82
Construction d'informations et de significations seule	29,3	52,8	28,7	80,7
Construction d'informations et de significations et compréhension logique	12	81,8	11,6	84,4
Pourcentages moyens	38,1	89,1	38,3	89,2

Les performances moyennes des 10% les plus faibles et des 10% les plus forts gardent le caractère de stabilité déjà mentionné. Les meilleurs réussissent en moyenne environ 90% des items, pourcentage qui frôle les 95% quand les questions mettent en jeu la *compréhension immédiate* et se réduit à environ 80% pour les questions nécessitant de *construire des informations*.

Les élèves les plus faibles réussissent environ 38% des items (48 % pour les questions mettant en jeu la *compréhension immédiate* et 23 % pour les questions nécessitant de *construire des informations*). Autrement dit, ces élèves parviennent dans un cas sur deux à prélever de l'information et à l'utiliser pour formuler leur réponse par écrit, mais trois fois sur quatre, ils échouent quand la réponse à la question exige la construction d'informations. Quand cette construction se combine à la *compréhension logique* du texte, seul un élève sur dix peut résoudre la tâche. Toutes les réponses devant être données par écrit, ces élèves ont pu cumuler des difficultés liées à la lecture et à l'expression écrite.

Nous remarquons, et nous y reviendrons, un travail très détaillé sur l'utilisation des modèles de réponses à l'item (MRI), qui semble être l'un des premiers menés à la DEP. Il a conduit à éliminer certains items de mauvaise qualité. Des items de mauvaise qualité avaient été repérés lors d'une analyse du matériel (Rémond, 1997). Par exemple, la rédaction de certains items posait problème ; la réponse attendue pouvait relever du prélèvement d'une information explicite du texte, mais la complication de la consigne pouvait nécessiter une inférence pour prendre conscience de ce qu'il fallait rechercher. Cet exemple pourrait se retrouver dans d'autres évaluations.

Usages attendus et usages réels des évaluations nationales

On observe globalement deux types d'usage. L'un sert aux responsables, évaluateurs et décideurs, et révèle des attitudes et des besoins éventuellement identiques quelle que soit la discipline évaluée. L'autre, surtout pédagogique, peut révéler des comportements différents selon la discipline évaluée. En particulier la nature globale de la maîtrise du français, l'histoire de cette discipline et de ses traditions scolaires peuvent influencer sur la façon dont sont reçus et utilisés les résultats des évaluations.

Un indicateur à la disposition des responsables

Référent national et annuel, elles constituent un indicateur utilisé, mais dans des proportions variables, par les évaluateurs et les décideurs du système éducatif.

Aux niveaux national et académique.

Ce sont surtout les évaluations diagnostiques qui jouent ce rôle d'indicateur. En effet, les élèves passent tous les mêmes tests aux mêmes moments de leur cursus : ce caractère national fournit donc une sorte de référent commun qui distingue les évaluations diagnostiques nationales des autres types d'évaluation. Leurs résultats sont des indices à la disposition des cadres institutionnels, qui les croisent avec d'autres données pour apprécier l'état de l'enseignement dans telle ou telle unité territoriale (académie, département, circonscription, ...), ou pour s'interroger sur les décisions à prendre. Ce qui est éclairant dans ce cas, si l'on tient compte de la situation spécifique des établissements dans l'unité considérée, c'est l'écart entre les scores locaux et les scores d'unités plus larges (moyennes nationale, académiques, départementales) ainsi que l'évolution de cet écart dans le temps. Pourquoi par exemple, dans tel département, les élèves ont-ils durablement des résultats inférieurs ou supérieurs à ceux des élèves de tel autre département comparable d'une même académie ?

Mais la comparaison des scores peut aussi s'effectuer entre différents secteurs d'enseignement. Le rapport Moisan-Simon (1997) a utilisé les évaluations pour analyser la réussite des ZEP. Même si les scores des élèves de ZEP sont globalement inférieurs aux scores moyens des autres élèves, la situation est différente entre les établissements situés en ZEP. Ainsi, ce rapport a pu constater que certaines ZEP avaient des résultats supérieurs à la moyenne nationale (par exemple dans l'enseignement primaire à Gennevilliers), alors que d'autres, à population scolaire comparable, obtenaient des résultats inférieurs de vingt points à cette moyenne. Cette observation a permis de dénoncer la croyance courante au déterminisme socio-culturel, et a mis en lumière les nombreux facteurs qui constituent des critères de réussite : taille de la zone et de l'établissement, cohérence des équipes, type de pilotage, stabilité des personnels, entre autres. Autant de données qui devaient permettre d'améliorer le dispositif des ZEP (voir aussi le dossier *Education et formations* réalisé par la DEP sur ce sujet en 1995).

Pour les cadres institutionnels qui doivent élaborer une politique éducative (recteurs, inspecteurs d'académie), cet indicateur joue souvent aussi un rôle d'alerte. Grâce au recueil des résultats, l'évolution des moyennes peut rendre visibles des phénomènes jusqu'alors mal perçus parce que dispersés, comme des difficultés persistantes dans un domaine donné, en mathématiques par exemple dans une académie ; dans une autre, la «régression, faible mais régulière, des résultats aux évaluations nationales» est prise en compte dans le projet académique. C'est pourquoi les tableaux de moyennes aident souvent aussi à définir la politique des bassins dans les académies.

Il est enfin utilisé pour orienter l'action des cadres intermédiaires chargés de la pédagogie. Ainsi, notent les auteurs d'un rapport sur une académie, «dans le sillage du projet académique, ... les IA-IPR s'engagent dans une réflexion concertée sur les disciplines, sur l'aide individualisée, sur l'évaluation.» On voit ailleurs apparaître des initiatives de pilotage pédagogique à partir des évaluations nationales, comme dans un département où «l'inspection académique forme des professeurs référents pour l'exploitation des évaluations de sixième et la mise en place des remédiations individuelles.»

Les études faites sur le terrain pour les rapports d'évaluation de l'enseignement dans les

académies ou dans le cadre normal du travail des inspecteurs font état de l'utilisation des évaluations par les inspecteurs de l'éducation nationale (IEN) pour piloter leur circonscription. Parmi les exemples possibles, on peut citer le cas d'une circonscription du département du Rhône située en ZEP et comprenant vingt-deux écoles. Sous l'impulsion de l'IEN, le projet élaboré en 1995 se fixait entre autres pour objectif «de réduire les écarts à la moyenne nationale des évaluations CE2». Les rapporteurs notent que l'objectif a été atteint : les résultats de cette évaluation en 1993-1994 «situaient les écoles de la circonscription dans les 10% des écoles en grande difficulté par rapport à l'ensemble national des ZEP» ; mais, en 2000, la circonscription a «atteint ou dépassé la moyenne nationale des réseaux d'éducation prioritaire (REP). ... L'écart par rapport à la moyenne nationale générale a été aussi réduit de manière significative.» Ce résultat est dû à un ensemble d'actions mises en place : dans l'ensemble de la circonscription, un conseil de maîtres de cycle situé en novembre «porte sur la remédiation à partir des résultats» de l'évaluation. Comme le protocole et les exercices présents dans les banques de données «sont peu utilisés tels quels dans la pratique des classes» les enseignants estimant que les évaluations CE2 sont «loin du niveau des élèves de REP», tous ces exercices «ont été retravaillés dans la circonscription, où l'on a également produit des outils d'évaluation de l'oral.» L'un des effets de la politique de l'IEN a donc été «l'amélioration nette des résultats des élèves aux évaluations» pour développer chez les élèves les compétences correspondantes.

Le pilotage pédagogique conduit par les IA-IPR prend aussi pour levier les évaluations nationales, grâce essentiellement à des opérations de diffusion, d'information et formation, de recherche. Dans une académie par exemple, l'IA-IPR de lettres conduit des séances de formation avec ses collègues de sciences humaines et de sciences de la vie et de la terre et avec leurs formateurs. Il s'agit, à partir des protocoles de sixième, d'explicitier les compétences évaluées pour la maîtrise de cet outil transversal qu'est la langue, la nature des tâches demandées aux élèves (prélever, reformuler synthétiquement, etc.), les difficultés rencontrées, en particulier dans la lecture du texte informatif fréquent dans ces disciplines, les sources éventuelles de ces difficultés dans des pratiques pédagogiques, enfin les moyens d'y remédier. Par le biais du centre régional de documentation pédagogique (CRDP) ou de la messagerie, une information générale ou ciblée est donnée au public, aux chefs d'établissement et aux professeurs pour présenter ou clarifier les évaluations et leurs résultats. Des groupes de travail produisent des exercices d'évaluation en couplant cette tâche avec la pratique des devoirs communs. Enfin, un document de travail est en projet, pour expliciter les relations entre les compétences des protocoles et les programmes.

Dans les établissements.

Cet usage des évaluations comme outil de pilotage est plus rare dans les établissements, et de grandes disparités existent tant dans le premier degré que dans le second. Ainsi, tout en signalant quelques exemples d'écoles et de collèges «où l'évaluation constitue le point de départ d'une prise en charge différenciée par groupes de besoin», un rapport de l'Inspection générale sur une académie exprime un fait fréquemment observé ailleurs en constatant qu'«à l'échelle de l'établissement, les évaluations servent peu au pilotage et fournissent rarement le socle d'un projet pédagogique».

Dans les écoles, les évaluations gardent leur fonction diagnostique. Elles servent d'abord à objectiver les résultats des élèves. En effet, la fin du cycle des apprentissages fondamentaux (cours élémentaire 1^o année) est perçue comme une étape décisive. Même si les résultats de l'évaluation CE2 confirment souvent à grands traits ce que pensent les instituteurs et le conseil de cycle 2, ils permettent d'objectiver les appréciations des maîtres. Elles répondent aussi au besoin de se situer par rapport à un niveau moyen national, repère particulièrement apprécié dans les ZEP. Dans les écoles où le niveau est excellent (plus de 80% de réussite à l'ensemble des épreuves), l'échec voire les décisions de redoublement peuvent être relativisés. Elles donnent enfin un contenu à la notion de niveau, surtout quand on a défini des compétences de base, ce qui est la question même du passage d'un cycle à l'autre.

Mais les évaluations peuvent également jouer un rôle de pilotage quand elles permettent :

- d'évaluer le cycle 2 pour, en retour, modifier éventuellement les pratiques et rééquilibrer les temps d'apprentissage,

- de mettre en place des actions, en particulier des PPAP, pour aider les plus faibles à acquérir les compétences indispensables à la poursuite du travail au cycle 3,
- de programmer au moins le début du cycle 3.

Ainsi, une école qui a constaté des difficultés en production d'écrit en 1998 a mis en œuvre des ateliers d'écriture au CE1 et a constaté de progrès. Dans le meilleur des cas, qui n'est pas rare mais qui n'est pas chiffré, les résultats aux évaluations constituent un outil de pilotage en aval : les évaluations CE2 pour piloter le cycle 2, les évaluations de sixième pour piloter le cycle 3. Pour ces dernières évaluations cependant, les maîtres disent eux-mêmes que l'engagement de l'équipe de circonscription, décisif pour l'analyse des résultats et la réflexion sur les pratiques, l'est encore très peu pour l'organisation d'améliorations pédagogiques.

La mise en place des PPAP (définis dans une circulaire de 1988) a obligé les écoles et les réseaux d'aides spécialisées à se saisir des résultats et des analyses des évaluations pour construire des programmes de travail avec les élèves repérés comme ne possédant pas les compétences de base. Des exemples de dispositifs de circonscription et de dispositifs départementaux ont été décrits lors d'un séminaire national organisé par la DESCO en octobre 2000. Aucune statistique ne permet de dire combien de ces projets ont réellement été conduits ni quels en sont les résultats. Mais cette décision ministérielle montre que, quand l'institution s'appuie elle-même sur les évaluations, elle permet aux acteurs de terrain de se mobiliser pour faire de ces évaluations des outils destinés à améliorer les résultats des élèves.

Dans les collèges et les lycées, l'utilisation des évaluations est, comme dans les écoles, inégale, et elle reste parfois formelle ; mais elle a fait naître des besoins.

L'incidence des évaluations nationales sur la politique de l'établissement demeure plutôt ponctuelle. Bien que la majorité des collèges fasse maintenant passer les évaluations, elles y sont généralement peu exploitées, et le sont encore moins dans les lycées. Présentés au conseil d'administration et figurant à titre de données dans le projet d'établissement, les résultats ne suscitent pas toujours ensuite de réel débat pédagogique ni de politique de formation. Inversement on voit par exemple, dans un collège classé en ZEP, que les résultats peu satisfaisants des élèves à l'évaluation de sixième, en particulier en expression écrite et en vocabulaire, ont conduit à inscrire dans le projet d'établissement, à côté de groupes de consolidation, des actions destinées à améliorer la maîtrise de la langue « en créant des situations dans lesquelles les élèves auront à lire et à s'exprimer devant un vrai public » : défi-lecture, élaboration d'un recueil de contes lus aux grandes sections de maternelle, par exemple.

Pour ce qui est de l'évaluation de seconde, les auteurs du rapport sur les modules au lycée (1998) notent que « les résultats de cette évaluation ne sont guère pris en compte pour constituer les groupes de travail. » « L'évaluation à l'entrée en seconde ne provoque plus les vives résistances ou réticences constatées lors de sa mise en place, mais elle est rarement prise en compte tant pour la composition des groupes de modules que pour la différenciation des thèmes de travail à proposer aux élèves. Les groupes sont souvent constitués avant de connaître les résultats de l'évaluation. Le passages des tests se fait trop tard, et cela leur enlève une bonne part de leur efficacité. »

Mais les collèges soucieux des résultats de leurs élèves aux évaluations ne se contentent pas de les utiliser pour construire le projet pédagogique. Et la façon dont il les utilisent met bien en lumière d'une part le besoin d'un référent national pour piloter un établissement, d'autre part ce qui limite actuellement un usage de ces évaluations qui réponde véritablement à leurs attentes.

Les résultats individuels des élèves aux évaluations de sixième sont une base de dialogue avec les parents d'élèves et avec l'équipe enseignante quand il s'agit, surtout au moment des conseils de classe, de réfléchir sur le bilan, sur l'avenir ou l'orientation d'un élève qui connaît des difficultés. Ils permettent de renvoyer aux enseignants un autre regard sur les possibilités d'un élève s'il est considéré par eux comme uniformément faible, d'éviter de faire porter la décision sur son seul comportement en classe, d'expliquer aux parents la nature de quelques-unes des difficultés importantes de leur enfant et le sens de la décision qui est envisagée pour lui. Quand la politique du collège repose sur une équipe pluriprofessionnelle, le conseiller d'orientation psychologue ou le conseiller d'éducation peuvent, selon la qualité de leurs interlocuteurs, utiliser également ces résultats.

Du coup, la relative opacité de l'outil est une entrave à la fois pour les parents et pour certains acteurs qui ne sont pas enseignants dans la discipline considérée. Il est certain qu'une brochure explicative à l'intention des parents et des non enseignants, donc distincte des *Commentaires* actuellement élaborés par la DEP, serait ici bienvenue. D'autre part, cet usage des évaluations diagnostiques révèle le besoin de bilans situés aux moments-clefs du cursus d'un élève pour éclairer (et non déterminer) les décisions d'orientation. Cela demanderait évidemment de repenser dans leur ensemble la finalité, la périodicité et les moments de passation des évaluations. Il faudrait également que les compétences évaluées d'une étape du cursus scolaire à une autre soient assez proches pour rendre visible le chemin parcouru par l'élève. De ce point de vue, l'harmonisation effectuée en 2002 entre les protocoles de CE2 et de sixième répond à ce besoin. Mais la nature de la récente évaluation d'entrée en cinquième, vécue par certains professeurs comme un angoissant échec, n'a pas permis, aux dires de certains principaux, de suivre en cinquième l'évolution des compétences en lecture et en écriture évaluées en début de sixième, parce que l'écart entre les deux protocoles est trop grand. Il faudrait donc que les compétences évaluées le soient en fonction de niveaux de compétence et de niveaux d'exigences clairement déterminés et affichés dans les programmes.

Quelques facteurs de cette utilisation disparate des évaluations nationales comme outil de pilotage pédagogique.

Plus courante et plus efficace dans le premier degré que dans le second, la prise en compte par les acteurs de l'enseignement des résultats aux évaluations nationales demeure pourtant trop dépendante des initiatives individuelles. Elle souffre également d'une circulation imparfaite de l'information et d'une attention encore trop peu partagée aux besoins des élèves.

L'exploitation des évaluations dépend encore trop du seul engagement individuel.

Dans le premier degré, si l'un des effets attendus des évaluations était d'aider directeurs et maîtres à identifier les difficultés des élèves pour les traiter, cet effet ne va pas de soi. Quand les projets d'école placent réellement au centre de leur action l'analyse des besoins des élèves, cela reflète surtout l'engagement, et même l'action durable de l'inspecteur de la circonscription.

Ce dernier peut alors avoir recours aux possibilités qui existent dans le premier degré pour accompagner les évaluations. L'équipe de circonscription (IEN, conseillers pédagogiques, maîtres formateurs parfois) dispose de moyens pour s'assurer que les évaluations sont passées, que les résultats sont communiqués aux parents ; les résultats peuvent être collectés au niveau de la commune, analysés, comparés aux moyennes nationales. D'autre part, douze heures d'animation pédagogique sont incluses dans les obligations de service annuelles des enseignants ; elles peuvent être utilisées en partie ou en totalité pour analyser les scores, chercher des pistes de remédiation au cycle 3 et prendre des décisions pour le cycle 2. L'IEN a une certaine souplesse d'organisation : avec l'aval de l'inspecteur d'académie, de nombreuses écoles ont bénéficié d'une demi-journée pour corriger en équipe les évaluations. Enfin, des stages de formation (de plusieurs semaines selon les académies) existent, jusqu'ici en grand nombre puisque les classes ainsi libérées sont confiées aux stagiaires, en particulier des stages spécifiques en ZEP.

Dans le second degré, l'exploitation des évaluations est le plus souvent l'affaire d'individus motivés et actifs, de personnalités comme celle du chef d'établissement ou de son adjoint, et rarement d'équipes ; le départ de ces personnes sonne alors le glas des initiatives qu'ils avaient lancées et soutenues.

Quant à l'engagement des inspecteurs du second degré chargés de l'enseignement général ou professionnel (IA-IPR ou IEN) pour soutenir ces initiatives, il dépend largement des moyens dont ils disposent. Or, ils ne disposent généralement pas, à la différence de leurs collègues du premier degré, de la disponibilité et des moyens (conseillers pédagogiques, journées de formation avec remplacement dans les classes, par exemple) qui leur permettraient de travailler durablement et efficacement, dans les établissements, avec principaux et professeurs. Quand ils sont nombreux de la même discipline dans une académie, un inspecteur au moins peut s'occuper plus particulièrement des évaluations, suivre les groupes académiques qui élaborent des exercices pour les groupes nationaux, analyser les résultats académiques, les utiliser lors de réunions pédagogiques et avec les formateurs, mais sans pouvoir faire l'information et l'accompagnement «de proximité» qui seraient nécessaires auprès des professeurs des collèges et des lycées qui dépendent de lui. Quand un inspecteur est seul dans son académie ou que l'équipe de ses collègues est réduite, cette action même est difficile voire impossible.

L'insuffisance des processus de diffusion et de communication souligne cette disparité.

Signe peut-être du désintérêt des établissements pour les évaluations, on observe fréquemment que les cahiers d'exercices et les résultats des élèves aux évaluations ne sont pas systématiquement communiqués aux enseignants des classes ou des cycles précédents.

Dans le second degré, les professeurs de troisième sont loin d'être toujours informés des résultats de l'évaluation de seconde ni destinataires des cahiers d'épreuves, alors même qu'ils reçoivent souvent des lycées les résultats des conseils de classe du premier trimestre de seconde. Le collègue ne peut donc pas croiser les données des évaluations de sixième et de seconde pour apprécier et réguler les effets de l'enseignement dispensé. Cela a été observé par exemple dans un collège de ZEP, où le Principal communiquait aux professeurs les résultats comparés des quatre collèges du bassin, dont le sien, aux évaluations de sixième. Les professeurs analysaient de près, en réunion de ZEP et entre eux, les scores des élèves «item par item» : mais, privés de toute autre information, comme les scores de leurs élèves aux évaluations de seconde, ils exploitaient moins qu'ils ne l'auraient pu les résultats de leur réflexion.

D'ailleurs, les résultats aux évaluations de seconde, très peu exploités dans les lycées, n'étaient même pas toujours transmis aux services des statistiques des rectorats, qui ne pouvaient donc en fournir une analyse générale. Il arrive même que des IA-IPR doivent réclamer aux rectorats les résultats des évaluations qui, autrement, ne leur sont pas communiqués.

Dans le premier degré au contraire, s'il est plus habituel que les écoles reçoivent des collèges de leur secteur les résultats des évaluations de sixième, cela n'est pas systématique. En outre, même quand ces résultats sont communiqués, ils sont le plus souvent globaux : ils ne permettent pas aux instituteurs, qui le regrettent, d'analyser, item par item, les résultats de chacun de leurs élèves. Or, comme le dit une IEN de l'Essonne qui note que «l'école a une demande forte vis-à-vis des résultats de sixième», «les maîtres de CM2... veulent des résultats item par item, parce qu'ils veulent avoir une analyse de ce qui se passe au cycle 3». Enfin, les instituteurs auraient souvent besoin, surtout en ZEP où ils les ressentent parfois «douloureusement», que la publication des résultats de l'évaluation soit «accompagnée, comme pour les lycées,» de l'indication des résultats attendus.

Enfin, la diffusion par la DEP des analyses et commentaires des groupes nationaux sur les évaluations est considérée comme trop tardive et faite en trop peu d'exemplaires (trois ou quatre par circonscription dans le premier degré).

Conclusion.

Les évaluations, surtout diagnostiques, constituent donc un référent national précieux dans l'éventail des indicateurs qui permettent aux cadres et responsables institutionnels d'apprécier les effets de l'enseignement, et elles fonctionnent parfois comme un système d'alerte sur des phénomènes masqués jusque-là par leur dispersion. Dans les collèges qui les utilisent, elles sont senties comme nécessaires dans le dialogue qui s'établit, à l'interne et à l'externe, sur les élèves. Cependant, elles ne jouent pas encore systématiquement, dans les établissements, le rôle de révélateur et de régulateur qui pourrait être le leur.

Ce qui est en jeu, c'est d'une part la diffusion et la communication à tous les intéressés des outils d'évaluation (cahiers, scores, analyses rectorales), d'autre part la mobilisation des relais intermédiaires comme les inspecteurs territoriaux et pédagogiques, pour développer, surtout dans les collèges, une mentalité qui donne du prix aux indicateurs afin de «mesurer les effets des actions menées». Cette culture de l'évaluation, comme on l'a appelée, est en cours, et grâce aux évaluations nationales. Mais elle n'est que naissante, et n'a pas encore entraîné partout une organisation administrative qui permette une exploitation efficace des protocoles et des résultats. Surtout, les évaluations nationales ne sont pas encore pensées et accompagnées de manière à permettre, surtout dans le second degré, ce que certains collèges en attendent : un suivi des élèves suffisamment régulier et sûr pour réguler les cursus et favoriser le travail en partenariat.

Les évaluations nationales comme outil de formation.

En formation initiale.

Les protocoles constituent un support de travail dans les stages ou les modules de formation. La théorie de l'évaluation est éclairée par les processus concrets d'évaluation disciplinaire que proposent les protocoles : interprétation des types de codage et des erreurs qu'ils induisent, analyse des scores des élèves, suivi d'une même compétence d'un protocole à l'autre, d'une année sur l'autre, d'une discipline à l'autre (lecture en français et en mathématique dans les protocoles de CE2 et de sixième, en histoire et en français dans ceux de seconde, par exemple).

On retrouve la fonction de visibilité et d'alerte qu'elles jouent parfois auprès des décideurs ; mais ici la visibilité porte sur les apprentissages à conduire. Telle composante de la compétence de lecture présente dans les protocoles, comme le relevé des indices qui permettent d'identifier les personnages d'un récit, avec ses scores de réussite et d'échec, produisent, pour prendre l'image d'un formateur, «un effet d'éclairage» sur la nécessité de prendre en compte cette compétence dans les apprentissages. Le protocole joue alors un rôle prescriptif.

Cependant, le peu de temps accordé aux Instituts universitaires de formation des maîtres (IUFM) pour la formation didactique des professeurs stagiaires, une année, fait que cette formation est limitée et n'a pas de rôle central. Il s'agit plus d'une sensibilisation, et qui se fait souvent dans des ateliers optionnels et hors de la formation générale commune. Cette formation modeste explique en partie la difficulté souvent observée des jeunes enseignants «à prendre en compte les acquis des élèves et à s'adapter au public» à cause de leur «carence en matière d'évaluation...».

Ce fait est souligné également dans le rapport sur *Les modules au lycée* : «La formation à l'évaluation, quand elle est dispensée, est trop généraliste. ... Il importe ... de donner à l'évaluation des élèves (sous toutes ses formes : sommative, formative, diagnostique) et à son utilisation dans les pratiques d'enseignement une place plus importante dans la formation initiale et continue des enseignants. Cela suppose que les IUFM aient les ressources nécessaires au sein de leurs établissements pour assurer ces formations.»

En formation continue.

Dans les plans académiques de formation pour le second degré, il est rare que les évaluations nationales soient présentes en tant que telles. Alors même que, par exemple, l'aide individualisée en seconde (diagnostic des compétences et des difficultés, exploitation des erreurs, projets de remédiation, modalités d'accompagnement des élèves, etc.) faisait partie des nouveautés du lycée, la partie du PAF consacrée aux Lettres dans une académie en 2000-2001 ne proposait aucune action sur l'analyse et l'exploitation des évaluations nationales, ni pour la sixième ni pour la seconde. Et ce cas n'est pas unique. Si l'on s'attend à voir l'évaluation trouver naturellement sa place dans les stages sur des problèmes didactiques demandés par les programmes, comme la construction de séquences d'apprentissage, on voit dans les PAF que le terme d'évaluation ne figure pas nécessairement dans le descriptif des stages ou qu'il y figure sans que soient évoquées les évaluations nationales.

Par ailleurs, la plupart des propositions des plans de formation continue répondent aux orientations des programmes ou des réformes. Il s'agit en général d'une activité globale de lecture ou d'écriture à partir de techniques (comme l'informatique), de supports (comme la littérature de jeunesse) ou de démarches nouveaux. La lecture ni l'écriture n'y sont présentées sous la forme de compétences hiérarchisées à atteindre, donc à évaluer, ce que se proposent les évaluations nationales. Il est donc logique que ces dernières ne figurent pour ainsi dire pas dans les PAF comme un outil spécifique à connaître et à utiliser dans l'enseignement.

Dans le premier degré, de la même manière, les programmes de formation continue offrent peu de stages qui partent des difficultés précises des élèves. Seuls les secteurs de l'éducation spécialisée en proposent, mais dans une perspective plus souvent psychologique que pédagogique. Ainsi des stages passionnants et très fréquentés ont été organisés dans les Hauts de Seine sur «les enfants hors du lire.»

Si donc les évaluations sont exploitées, mais encore trop rapidement, en formation initiale, elles n'occupent qu'une place marginale dans les plans de formation continue.

Cependant, comme on l'a dit plus haut, de nombreux stages sont organisés, par les circonscriptions pour les écoles et par les académies pour les collèges, sur les résultats des évaluations

en cours, et sur la liaison école-collège, c'est-à-dire CM2-6°, ces derniers étant parfois inscrits dans les plans de formation continue.

Ces stages CM2-6° prennent les résultats des évaluations comme support des échanges entre les enseignants. Pour les instituteurs et professeurs des écoles, elles constituent une évaluation du travail réalisé au cycle 3 ; pour les professeurs de collège, elles indiquent le niveau de leurs nouveaux élèves. Ces stages sont difficiles à organiser et à conduire : les professeurs de collège sont moins disponibles car ils ne sont pas, comme les maîtres du premier degré, remplacés dans leur classe ; d'autre part, les discussions peuvent être perçues comme des remises en cause. Mais, une fois ces obstacles dépassés, un travail fructueux a pu parfois s'engager sur les pratiques et les contenus d'enseignement de part et d'autre.

Les évaluations nationales comme outil pédagogique pour les enseignants

Dans le premier degré, les évaluations nationales permettent de constituer des groupes de besoin à partir des faiblesses constatées. Il est difficile de savoir exactement combien de classes fonctionnent ainsi; c'est souvent dans les secteurs où de nombreux élèves sont en difficultés que ces groupes sont constitués. Ainsi dans une école de ZEP des Hauts de Seine suite à de mauvais résultats aux évaluations en production d'écrit, les maîtres ont mis en place en CE1 pour les évaluations à venir et au cycle 3 pour les élèves déjà évalués, des ateliers d'écriture selon une méthodologie rigoureuse. Le suivi de cohorte a montré que les élèves évalués en CE2 avaient progressé aux évaluations de sixième. Les élèves de CE1 se sont situés l'année suivante au-dessus de la moyenne nationale en production d'écrit.

Même si ces pratiques sont plus développées dans le premier degré, elles sont encore loin d'être majoritaires.

Dans le second degré, les évaluations nationales n'ont pas encore suffisamment développé le rôle de l'évaluation dans les pratiques pédagogiques.

«Les évaluations nationales en CE2 lorsqu'on en dispose, à l'entrée en 6°, à l'entrée en 2° sont sous-utilisées par les enseignants.» Cette constatation d'Inspecteurs généraux pour une académie est formulée dans des termes proches pour beaucoup d'autres académies. Un récent rapport sur la classe de sixième fait le constat suivant : dans les collèges visités, «l'évaluation nationale a certes eu lieu aux dates indiquées, mais les résultats qu'elle fournit ne sont utiles, au niveau statistique, que pour l'équipe de direction et pour les autorités académiques, ce qui n'est pas leur destination première... Cette opération ne développe pratiquement pas chez les enseignants une culture de l'évaluation.»

Sous-utilisation des évaluations...

La plupart des professeurs de collège considèrent désormais comme normal de faire passer et de corriger les évaluations de sixième. Cela ne signifie pas qu'ensuite ils exploitent pédagogiquement ces résultats, qu'ils considèrent d'ailleurs souvent comme un «bilan de l'école primaire» plus que comme un état des compétences et des besoins de leurs élèves utile pour l'année à venir. Il a été observé par exemple que les résultats aux évaluations aient conduit un collège à décider d'organiser des groupes de besoin ou de soutien, sans que ces résultats, ensuite, servent de critères aux professeurs de français pour constituer les groupes. Les problèmes d'organisation matérielle, le souci de maintenir la cohérence du groupe-classe ou le fait que beaucoup d'élèves partagent finalement les mêmes erreurs prévalent alors sur une exploitation plus ciblée des évaluations. Déjà, lors de la mise en œuvre des dispositifs de consolidation en sixième (1996), une enquête conduite dans l'académie de Nantes en 1998 montrait que les élèves étaient affectés dans les groupes «à partir des évaluations réalisées par les professeurs» ou «en prenant appui sur les dossiers de l'école primaire» ; les rédacteurs ajoutaient que «dans tous les cas, l'évaluation de début d'année en sixième semble jouer un rôle minime.»

...due en partie à la faiblesse de la culture de l'évaluation,

Cette difficulté à apprécier les besoins des élèves et à leur donner une réponse pédagogique adaptée prend «racine dans la faiblesse de la culture d'évaluation, notamment formative : sous-utilisation des évaluations nationales,... absence de rétroaction de l'évaluation sur les pratiques professionnelles.»

...»La difficulté à évaluer les compétences rejaillit sur la mise en œuvre des démarches pluridisciplinaires transversales.» Même dans l'enseignement professionnel, «l'évaluation nationale de la mise en œuvre des PPCP (Projet Pluridisciplinaire à Caractère Professionnel) ... a fait apparaître «l'absence fréquente, dans la conception du projet, de la dimension d'évaluation des acquis.» Ces jugements portés par les évaluateurs de l'enseignement dans une académie sont vrais, à quelques nuances près, pour beaucoup d'autres académies. De même, après avoir constaté chez les professeurs de lycée un accueil progressivement moins hostile à l'évaluation de seconde, les inspecteurs généraux notent que, malgré tout, l'exploitation des cahiers d'évaluation «reste limitée, quand elle n'est pas purement et simplement ignorée. Dans leur grande majorité, les professeurs de Français n'ont pas encore intégré à leurs pratiques ces nouvelles formes d'évaluation.»

...à la difficulté de concevoir une authentique remédiation,

Les professeurs n'ont pas les moyens de mettre en place une authentique remédiation à partir des résultats des évaluations.

Ce qui est en jeu, c'est l'insuffisance de l'accompagnement pédagogique et didactique des professeurs : à telle erreur révélée par les évaluations ne correspond pas tel remède, tout prêt dans les manuels ou élaboré par l'enseignant. La «remédiation» demandée dans les textes n'est pas un remède au sens médical du terme, comme le mot est compris généralement, mais une nouvelle médiation, ce qui est tout autre chose. L'erreur révélée par les tests d'évaluation ne porte que sur une composante déterminée et évaluable de la compétence de lecture ou d'écriture. Or, dans la réalité de l'activité globale de l'élève, cette erreur est liée à d'autres, mais aussi à des comportements psychologiques ou cognitifs, aux univers sociaux et culturels où se meuvent les élèves, toutes choses qui font que traiter l'erreur ponctuelle est souvent insuffisant, voire inopérant. Les textes de présentation publiés par la DEP avec l'analyse des résultats le disent : la remédiation n'est pas un entraînement répétitif, mais «une combinatoire». Les composantes des compétences ne s'additionnent pas, elles se combinent (livret du professeur 1993). Si donc on multiplie les remèdes à des items ponctuels (sens d'un mot dans le contexte, par exemple), on n'améliore pas mécaniquement la compétence globale. Seulement, cette combinatoire n'apparaît pas comme telle dans les protocoles. Les notes d'information de la DEP sur les évaluations annuelles le rappellent explicitement ; mais cela ne suffit pas à l'inscrire dans la réalité. Il reste très difficile de savoir comment traiter efficacement chez les élèves des carences complexes dont un seul des éléments se signale dans les évaluations.

...de lier les évaluations aux programmes,

Dans la représentation des professeurs, les évaluations n'ont pas de lien avec les programmes et restent extérieures à leur enseignement. Cette difficulté est spécifique au second degré.

Les protocoles sont élaborés à partir des capacités affichées par les programmes. Pourtant, les professeurs ne voient pas le lien entre les évaluations et les programmes, ce qui explique en partie le rejet ou du moins l'indifférence de nombre d'entre eux. C'est que les protocoles sont conçus selon une logique analytique, les programmes selon une logique globale.

Les protocoles de sixième évaluent, sur des supports diversifiés et limités, des composantes précises des capacités : pour la lecture, par exemple, saisir l'essentiel, reconstituer l'organisation logique ou chronologique, identifier les locuteurs à partir de leurs substituts, repérer ou construire des informations, etc. Pour les professeurs, lire c'est une activité globale, c'est lire une pièce de Molière ou une fable de La Fontaine : c'est-à-dire mobiliser simultanément des savoir et savoir-faire relevant de domaines différents pour construire le sens d'objets complexes constitués avec la langue française. La maîtrise de composantes limitées de cette compétence ne conduit pas automatiquement à la maîtrise de la compétence globale, qui est d'une autre nature. Aussi les professeurs peuvent-ils constater les scores flatteurs de leurs élèves aux évaluations, tout en observant dans leurs classes à quel point la compétence globale reste mal assimilée, peu maîtrisée.

Même chose en production écrite. Les protocoles évaluent des composantes déterminées de la capacité : hiérarchiser des informations par la mise en page, tenir compte du libellé et de ses données, élaborer un texte cohérent, respecter les contraintes de la langue, écrire lisiblement, etc. Pour les professeurs, la

production écrite est surtout un entraînement à une compétence globale qui s'exerce sur un répertoire de lieux rhétoriques, sur des types de sujets propres à chaque niveau (fiches de lecture, fictions ou récits autobiographiques, lettres, essai littéraire, commentaire composé, ...). Cette compétence globale demande, elle aussi, de traiter simultanément des contraintes multiples et de natures diverses en fonction d'une vision claire du type d'écrit et de l'effet à obtenir. Là encore, les professeurs peuvent constater une grande divergence entre des scores honorables aux exercices limités des évaluations et l'absence de maîtrise de ces mêmes composantes dans les tâches globales auxquelles ils exercent leurs élèves. L'orthographe en est un bon exemple : les protocoles de sixième montrent que les élèves la maîtrisent mal en production écrite, alors qu'ils répondent mieux, et parfois bien, aux exercices ponctuels où ils doivent orthographier correctement une dizaine de termes dictés. Les résultats aux évaluations de seconde en 1998 montrent que plus de six élèves sur dix respectent l'orthographe grammaticale en production écrite, et plus d'un sur deux la syntaxe des phrases ; ces scores étaient si différents de ce que l'on pouvait constater dans les classes, que les évaluateurs essayaient de les expliquer en les attribuant par exemple à un codage trop peu contraignant ou à des écrits trop brefs pour une évaluation rigoureuse de ces items.

Ces divergences sont normales et légitimes. Même quand elles portent sur des tâches globales et complexes (la lecture d'un texte long et littéraire en cinquième et en seconde, la production d'un écrit complet dans tous les protocoles), les évaluations indiquent quelle composante de la compétence globale est ou n'est pas maîtrisée. C'est une mesure qui peut servir à celui qui conduit un élève sur le chemin d'une compétence globale qui, elle, est de nature différente : elle suppose que l'emploi harmonieux et convergent de ces multiples composantes soit assez assimilé pour s'exercer presque automatiquement, et donc aussi longtemps que nécessaire, sur des objets complexes. Tant que l'élève n'y peut réussir qu'un court laps de temps et parce qu'on le lui demande expressément (situation des évaluations), la compétence globale n'est pas nécessairement assimilée, elle ne va pas de soi pour l'élève dans toute autre situation. La présentation générale des protocoles rappelle sans cesse cette différence, et donc vainement, comme le montre une page des commentaires des résultats nationaux de l'évaluation de seconde de septembre 1997. «Le reproche de parcellisation qui revient fréquemment repose sur un malentendu : lors de l'évaluation de septembre, il ne s'agit pas de faire étudier un texte par les élèves mais d'évaluer les compétences de lecture de chaque élève. Ces compétences mettent en jeu des savoirs notionnels de lecture et d'écriture qui sont en interaction et que souvent il n'est pas possible d'isoler. Cette spécificité rend le questionnement délicat et le codage difficile, d'où des réticences vis à vis de l'évaluation dont pourtant l'enseignement de la discipline ne peut se passer... Il n'est pas inutile de rappeler ici qu'une situation d'évaluation n'est pas une situation d'apprentissage.»

Pourtant, les programmes de 1985 pour le collège et de 1987 pour le lycée énumèrent les objectifs à atteindre et les compétences que les élèves doivent avoir acquises à la fin de leurs études. On peut d'ailleurs mettre en parallèle les compétences assignées dans les programmes à chaque cycle et saisir ainsi leur continuité et leurs seuils.

«A la fin du cycle d'observation (fin 5°) l'élève doit savoir :	«A la fin de la scolarité du collège, l'élève doit savoir :
1. consulter les usuels : distinguer les types de dictionnaires, maîtriser l'ordre alphabétique à deux initiales au moins, exploiter les outils de référence contenus dans les manuels(...), utiliser index et tables des matières. ;	1. consulter les usuels, distinguer les types de dictionnaires et tirer parti des manuels, trouver un article ou un document, utiliser un index et une table des matières, prendre quelques notes pour un travail simple, écrit ou oral ;
2. reconnaître et utiliser les principaux outils linguistiques : paradigmes verbaux, verbes irréguliers d'emploi fréquent, structures syntaxiques de la phrase simple ;	2. reconnaître et utiliser les principaux outils linguistiques, paradigmes verbaux, verbes irréguliers d'emploi fréquent, structures syntaxiques (fonctions, constructions des verbes), pronoms et mots de reprise ;
3. utiliser le vocabulaire usuel défini par les tables de fréquence ;	3. employer exactement le vocabulaire usuel, distinguer les synonymes et les différents sens d'un mot ;
4. rédiger un texte d'intérêt pratique ou d'un ton plus personnel, c'est-à-dire tout à la fois : composer des phrases simples, les enchaîner; respecter l'orthographe (mots outils, vocabulaire courant, accords fondamentaux, conjugaison des verbes d'emploi fréquent, ponctuation); viser un destinataire ;	4. rédiger un texte d'intérêt pratique ou d'un ton plus personnel, c'est-à-dire tout à la fois : composer des phrases simples, les enchaîner; respecter l'orthographe, la ponctuation; organiser un développement en plusieurs paragraphes; viser un destinataire ;
5. comprendre le sens littéral d'un texte simple ;	5. comprendre un texte lu, saisir le sens littéral, dégager les lignes directrices, percevoir les articulations, logiques ou chronologiques, reconnaître l'absence de liaison ;
6. définir la spécificité d'un texte: reconnaître les principaux traits qui caractérisent certains types (narratif, théâtral, poétique, documentaire); distinguer entre auteur, narrateur et personnage; relever un champ lexical ;	6. définir la spécificité d'un texte; replacer un texte dans ses contextes et définir son statut; distinguer entre discours et récit, style direct et style indirect; observer des champs de significations (lexicaux et sémantiques) et préciser la portée de certaines figures (métaphore et métonymie); définir le statut du narrateur, celui des personnages, l'enjeu de l'action ou de l'argumentation, reconnaître les particularités du texte poétique; utiliser un vocabulaire élémentaire de l'analyse des textes ;
7. résumer oralement un texte narratif ;	7. résumer oralement et par écrit un texte lu, un texte entendu ;
8. rendre compte oralement d'un livre (le sujet, l'action, les personnages);	8. rendre compte oralement de la lecture d'un livre, formuler et justifier une appréciation, choisir quelques courts extraits, les lire à voix haute ;
9. observer l'image fixe : analyser la composition d'une image fixe; dégager d'une image une signification ;	9. observer l'image, fixe et mobile, analyser la composition d'une image fixe, dégager d'une image une signification, examiner la trame narrative d'un film ;
10. situer dans le temps œuvres, textes et images étudiés.»	10. situer historiquement œuvres, textes et images étudiés. «
Ref. <i>Complément aux Instructions et Programmes de 1985.</i>	Ref. <i>Complément aux Instructions et Programmes de 1985.</i>

«On peut se proposer, comme objectifs essentiels à atteindre à la fin de la classe de seconde, de conduire l'élève à :
1. savoir prendre des notes, les classer, les utiliser ;
2. savoir définir une situation de communication, une question à traiter ;
3. savoir rendre compte d'une rencontre, d'une visite, d'une expérience ;
4. savoir établir ou découvrir la composition d'un message en relation avec une situation donnée et un objectif à atteindre ;
5. savoir comprendre et manier les articulations logiques explicites ou implicites ;
6. savoir résumer oralement ou par écrit un débat, un texte qu'on a lu, ou entendu, ou composé soi-même ;
7. savoir repérer un type de texte et définir son statut ;
8. savoir poser et exploiter concrètement les principales questions qui peuvent guider une lecture méthodique et aider à construire l'essentiel d'un commentaire ;
9. savoir utiliser un vocabulaire minimum d'analyse littéraire ;
10. savoir situer exactement dans l'histoire les auteurs et les textes abordés pendant l'année.»
Ref. <i>Instructions</i> de Français pour la classe de 2°, BO spécial n°1 du 5 février 1987.

En outre, on trouve ponctuellement dans les programmes des définitions d'activités présentées sous forme de compétences . Par exemple, le début du Complément aux *Programmes et Instructions de 1985* intitulé *Lire au collège* , définit ainsi la lecture :

«Savoir lire, c'est savoir accéder, à travers l'écrit, de manière autonome, à la pensée d'autrui : savoir lire et savoir écrire sont complémentaires.»

Dans la même brochure, on lit plus loin, sous la rubrique *Objectifs* :

« A l'issue des classes de collège, tout élève doit pouvoir :

- pratiquer une lecture autonome, silencieuse et, s'il est nécessaire, rapide,
- accéder à une lecture consciente qui permet de saisir les informations, les idées et les significations en les replaçant dans leurs contextes définis par l'histoire, la langue, l'appartenance à un type de texte, en exerçant sur elles une réflexion critique,
- acquérir le goût de la lecture qui naît d'une compétence maîtrisée et qui répond aux appels de l'intelligence, de la sensibilité et de l'imagination.»

On peut relever également que la nécessité de l'évaluation est rappelée dans les programmes. « Cet effort permanent de formation et de culture doit permettre aux professeurs d'assurer une évaluation plus sûre des élèves ... Cette évaluation vise à identifier, à côté des réussites, les difficultés rencontrées par certains élèves.» (*Collèges ,Programmes et Instructions,, 1985, Orientations et objectifs*). Le *Complément* intitulé *Lire au collège* demande d'«évaluer la compétence des enfants à leur entrée en sixième».

Enfin, au moins au début, les groupes d'élaboration ont défini les compétences à évaluer en exprimant les objectifs des programmes en termes d'activités observables ; ils ont traduit les objectifs des programmes en objectifs mesurables par des instruments à application collective.

On ne peut donc pas dire que les compétences à atteindre soient absentes des programmes, ni que la nécessité d'évaluer les acquis des élèves le soit également. Si cependant peu de professeurs ont pris l'habitude de lier leur enseignement à ces compétences, et ces compétences aux protocoles d'évaluation qui en reprennent l'essentiel, c'est peut-être que leur formation ne les a pas rendus sensibles à cet aspect de leur pratique, qu'ils ne l'ont donc pas reconnu et retenu dans les programmes. Il est possible aussi qu'il s'agisse d'une question de présentation : peut-être faudrait-il que les activités elles-mêmes soient présentées sous forme de compétences à acquérir, hiérarchisées et caractérisées par des niveaux d'exigence progressifs.

...à la réticence à l'égard d'un outil imposé,

Les protocoles d'évaluation ont été envoyés tout faits aux professeurs du second degré sans que la grande majorité de ces professeurs ait participé à leur élaboration : cela n'a pas été sans influence sur le façon dont ils ont été reçus. Ils ont été considérés comme une tâche imposée «d'en haut», un «pensum» dont on subissait clairement la lourdeur sans en voir d'emblée le profit. Si ce rejet s'est

atténué avec le temps, le caractère imposé, étranger, de l'outil est demeuré, et cela n'a pas suscité chez les professeurs, comme on pouvait l'espérer, le besoin d'évaluer les effets de l'enseignement dispensé. Ce n'est pas l'outil qui crée le besoin, c'est l'inverse.

...à l'influence des traditions surtout dans le second degré.

Enfin, dans la mentalité des professeurs, la «variable élève n'est pas le centre du système», ce qui ne les rend pas attentifs à l'évaluation.

N'étant pas habitués, contrairement à leurs collègues du premier degré, à considérer les élèves d'une manière «différenciée», comme pouvant être forts sur une compétence, faibles sur une autre, en cours d'acquisition ailleurs, les professeurs ont du mal à accepter les résultats de protocoles qui évaluent les acquis autant que les déficits, qui ne classent pas les élèves, qui font de l'erreur un processus normal en cours d'apprentissage. Là encore, l'appréhension de l'élève est le plus souvent globale chez les professeurs.

C'est pourquoi beaucoup de professeurs pensent que les résultats surévaluent les élèves, ce qui discrédite à leurs yeux les évaluations elles-mêmes, d'autant plus que ces résultats viennent alors contredire leur propre jugement sur les capacités de leurs élèves. Ou bien, au contraire, ils estiment que les résultats des évaluations ne leur apportent aucune connaissance particulière de leurs élèves, comme semblent le penser les professeurs de lycée : «Loin de favoriser la connaissance par l'enseignant de ses nouveaux élèves, elle (l'évaluation de 2^o) la retarde par la mise en place de tâches supplémentaires dont l'aspect scientifique ne nous a pas convaincus qu'il puisse remplacer le jugement humain et nuancé d'un professeur. ... Les résultats aux premiers devoirs contredisent même parfois les codages.» (rapport sur *Les modules au lycée*). Même constat dans le rapport récent sur la sixième : «L'évaluation de début de 6^o est la plupart du temps considérée par les professeurs comme un travail supplémentaire qui ne leur apporte guère plus d'informations que l'observation directe des élèves.»

Conclusion.

Les évaluations nationales sont paradoxalement de moins en moins utilisées à mesure qu'on arrive au niveau de l'établissement et de la classe. Même quand leurs résultats fondent ou influencent la politique de l'établissement, ce qui n'est pas le cas général, ils n'influencent pas nécessairement et en profondeur les pratiques pédagogiques de la majorité des enseignants, bien que cette observation soit plus fréquente dans le second degré que dans le premier.

Certes, on voit se développer grâce à elles une «culture de l'évaluation» qui touche surtout l'encadrement mais qui commence à atteindre aussi le corps enseignant : après le temps de passation des épreuves nationales, on a pu voir dans les cahiers de textes des classes, par exemple en seconde, des tests «maison» élaborés par les professeurs pour évaluer les acquis des élèves, ce qui était rarement le cas avant la naissance des évaluations nationales. L'attention à l'élaboration des consignes et à la nature des erreurs s'est développée, et les évaluations commencent à influencer les modes habituels d'évaluation en classe, ce qui est déjà un résultat notable. Mais les pratiques pédagogiques ne se sont pas réellement ni profondément modifiées.

La formation initiale et continuée, trop généraliste ou trop modeste, ne permet pas aux enseignants de s'approprier les cahiers et les critères d'évaluation. Ils ne voient pas comment relier les programmes qu'ils doivent appliquer et les tests que passent leurs élèves. Ils n'ont même pas l'impression que les résultats des évaluations leur donnent une connaissance des élèves plus vraie et plus nuancée que celle qu'ils peuvent avoir par leurs propres moyens, c'est-à-dire par les exercices habituels de leur discipline. Et souvent les résultats leur arrivent quand les groupes de besoin ou les modules sont déjà constitués. Enfin, ils disposent rarement des cahiers, des tests et des résultats des élèves dont ils ont ou dont ils ont eu la responsabilité. Tout reste largement affaire d'initiatives et d'intérêts individuels. Si donc les évaluations fournissent aux décideurs du système éducatif un indicateur utile, elles n'ont pas encore atteint, chez les professeurs, l'objectif pédagogique qui était le leur.

Dans le premier degré, les évaluations sont utilisées plus largement. Elles donnent un contenu aux compétences essentielles du cycle 2. Elles constituent souvent l'objet d'un conseil de cycle qui permet parfois de programmer et de réguler les actions pédagogiques. Elles objectivent les difficultés des

élèves et favorisent la mise en place de groupes différenciés et la prise en charge des élèves les plus en difficulté grâce aux PPAP. Enfin, elles constituent pour les ZEP un indicateur essentiel du niveau à atteindre impérativement.

Evaluation hors système éducatif : la Journée d'Appel de Préparation à la Défense (JAPD)

Depuis avril 2000, tous les jeunes français et françaises, environ 650 000, doivent être convoqués à une Journée d'Appel de Préparation à la Défense lorsqu'ils ont environ 17 ans. Dans le cadre de cette journée, leurs compétences en compréhension de l'écrit sont évaluées au moyen d'épreuves conçues par la direction de l'évaluation et de la prospective (MJENR), le ministère de la défense et l'Université de Paris V.

Première demi – journée, détermination du niveau général

L'évaluation se déroule en deux temps : une épreuve d'orientation a lieu le matin pour toute la population convoquée. Elle a pour but de départager cette population en deux « filières » :

- filière A : les jeunes en difficulté de lecture.
- filière B : les jeunes « bons lecteurs ».

Ce partage de la population se fait à l'issue de deux épreuves qui sollicitent des compétences de recherche et de sélection d'informations dans un document de la vie quotidienne (un programme de télévision organisé en un tableau à double entrée) d'une part, et la compréhension d'un texte narratif court, d'autre part.

Les résultats auxquels nous allons faire référence, concernent l'année 2001 – 2002, la seconde année d'existence de la JAPD. Ces résultats ne sont pas représentatifs de l'ensemble d'une génération puisque tous les jeunes ayant eu 17 ans en 2001 ne se sont pas présentés à la JAPD et que par ailleurs, d'autres plus âgés qui ne s'étaient pas encore présentés ont été intégrés à la population 2001 – 2002, ce qui fausse l'indicateur « compétences en lecture des jeunes de 17 ans », qui ne pourra être pris comme référence pour les années à venir du fait de ce biais d'échantillon.

La filière B est composée de 88 % de jeunes qui n'éprouvent pas de difficultés de lecture à l'épreuve d'orientation. La filière A comporte donc 12 % des jeunes (garçons : 14, 5 %, filles : 9, 4 %) rencontrant de réelles difficultés de compréhension qui pourraient les placer dans une situation d'illettrisme. Notons que si l'on considère les élèves de la classe d'âge (les 17 ans), ce pourcentage de jeunes en difficulté se réduit à 6 %, chiffre comparable à 2000 - 2001.

Deuxième demi – journée : caractérisation des compétences

A l'issue des épreuves, les jeunes en difficulté de lecture sont répartis en trois groupes :

- groupe 1 : les jeunes maîtrisant insuffisamment les mécanismes de base de la lecture.
- groupe 2 : les jeunes qui ne possèdent que partiellement les mécanismes de base de la lecture et qui éprouvent de graves difficultés à lire tous les types d'écrits.
- groupe 3 : les lecteurs « rigides » qui ne peuvent s'adapter efficacement aux différents écrits proposés, mais qui n'ont toutefois pas de troubles apparents des mécanismes fondamentaux de la lecture.

Les Groupes 1 et 2 rassemblent les jeunes en grave difficulté de lecture, soit 6% de la population totale, dont 8 % des garçons et 3, 9 % des filles. Ces jeunes se voient proposer un entretien individualisé destiné à les orienter vers un réseau qui pourra les prendre en charge pour leur éviter l'illettrisme. Comme on pourrait s'y attendre, leur niveau scolaire est bas (dominance de niveau 1, puis de niveau 2). Cependant, on trouve aussi des élèves appartenant à des niveaux plus élevés, ce qui pose le problème de la limite de cette évaluation : manque de motivation, volonté de parasiter le dispositif ?

Deuxième temps pour la filière B : les élèves déclarés « bons lecteurs » passent une épreuve destinée à évaluer de manière plus fine leur compréhension. Trois supports la composent : un support de la vie courante – dossier de candidature à un concours administratif -, un texte plus long – nouvelle d'un auteur contemporain – et un extrait d'un texte classique servant de base à une épreuve d'orthographe.

Les jeunes de la filière B réussissent en moyenne près de 80 % des items de lecture. A toutes les épreuves, sauf au programme de T.V., les filles obtiennent de meilleurs scores de lecture, et l'écart s'accroît en leur faveur pour les supports littéraires les plus longs. Ceci se vérifie aussi pour l'épreuve d'orthographe réussie par près de 80 % de la population totale.

Des analyses plus fines des résultats ont été réalisées. Les pourcentages de réussites en fonction du classement des items en trois niveaux de compréhension sont les suivants : 90 % des items testant la compréhension immédiate, 70 % de ceux qui estiment la compréhension logique et environ

60 % des items demandant une compréhension fine. La compréhension immédiate ne départage pas les garçons des filles, mais pour la compréhension logique et la compréhension fine, qui requièrent des traitements plus approfondis, les filles surpassent significativement les garçons.

En résumé, près de neuf jeunes sur dix de la JAPD 2001 – 2002 se révèlent « bons lecteurs » du protocole qui leur a été proposé, et maîtrisent les compétences orthographiques visées par une épreuve de complètement orthographique.

Le nouveau dispositif

Cependant, comme nous l'avons dit, le biais d'échantillon ne permet pas que ce résultat soit pris comme référentiel dans les années à venir. Les épreuves sont en cours de révision, notamment pour pouvoir mieux caractériser les bas niveaux de lecture.

Depuis septembre 2002, dans la majorité des centres JAPD, seule l'épreuve d'orientation est passée par les jeunes et elle a même été utilisée en 2003 pour repérer les jeunes en grande difficulté alors que ce n'était pas son objectif initial. Pour 2004, une nouvelle épreuve d'orientation a été construite afin de repérer directement les jeunes illettrés. Il est prévu dans ce nouveau cadre d'établir des indicateurs au plan régional.

Chapitre 2

Les évaluations internationales

Les enquêtes internationales donnent une appréciation de l'efficacité relative des politiques éducatives et aboutissent souvent à un **palmarès des pays**, parfois jugé comme une forme de Jeux Olympiques. Les performances des pays sont classées les unes par rapport aux autres sur une échelle de compétence, selon les notes que ces pays ont obtenues. Ces évaluations sommatives ne s'intéressent pas à la manière dont le résultat est obtenu, ni aux processus qui le sous-tendent. Elles ne sont pas conçues pour repérer les élèves en difficultés et il ne faut pas chercher à le faire avec les outils qu'elles comportent.

Ces enquêtes se sont développées depuis la fin des années cinquante dans le but de comparer les compétences des élèves de plusieurs pays dans des domaines tels que les mathématiques, la lecture ou les sciences. L'IEA (*International Association for the Evaluation of Education*) en a mené la plupart, rejointe récemment par l'OCDE (Organisation pour la Coopération et le Développement Economique) soucieuse de connaître l'efficacité des systèmes éducatifs.

La dimension internationale contraint le cadre de ces évaluations qu'il faut rendre comparable d'un pays à un autre. Leur conception repose bien sûr sur des aspects de la lecture et de la compréhension, mais il ne sera pas possible d'entrer dans le détail du traitement de telle ou telle marque linguistique qui, peut-être, n'existe pas dans d'autres langues, alors qu'elle est tellement importante pour notre propre langue. Aussi ces évaluations ont un **caractère très général** : l'appui sur les programmes scolaires devient illusoire dès l'instant où l'on essaie de préparer un référent commun.

Actuellement, des évaluations sommatives de la « *reading literacy* »¹ sont menées : PIRLS (Progress in International Reading Literacy Study) par l'IEA et PISA (*Programme for International Student Assessment*) par l'OCDE.

¹ PIRLS et PISA se réfèrent au concept de « literacy », employé pour désigner la « culture », la notion historique d'aptitude à lire et à écrire (capacités de lecture et d'écriture) étant dépassée pour adopter une dimension beaucoup plus vaste défini comme le rapport qu'on développe avec l'écrit. Nous avons volontairement conservé ici le mot anglais « literacy », car le néologisme « *littératie - litéracie* », concept mal défini en France, ne permet pas d'envisager l'ampleur de l'objet traité dans PISA. La participation effective à la société actuelle fait appel à une « culture de l'écrit », ce qui suppose de disposer de connaissances et de savoir-faire de base et de les maîtriser.

Le programme PIRLS – Progress in International Reading Literacy Study

www.pirls.org

Le programme PIRLS de l'IEA a été développé par une équipe de « *l'International Study Center de Boston College* » (Etats - Unis) pour mesurer les performances en lecture des élèves de 9 à 10 ans. Trente cinq pays ont participé à cette enquête qui, comme PISA, a un caractère cyclique (première application en 2001, puis en 2005 et 2009). Dans les faits, les élèves testés ont en moyenne 10 ans, et à quelques exceptions près, ils achèvent leur quatrième année de scolarité obligatoire comme prévu par les concepteurs de l'enquête. L'échantillon français était composé de 3500 élèves scolarisés dans 230 classes de CM1, leur moyenne d'âge était de dix ans et un mois.

Le processus d'élaboration

Le processus d'élaboration de PIRLS ressemble aussi à celui de PISA puisqu'il repose sur un **cadre conceptuel** incluant une **définition** de la « *reading literacy* » : « *l'aptitude à comprendre et à utiliser les formes du langage écrit que requiert la société ou qui sont importants pour l'individu. Les jeunes lecteurs peuvent construire du sens à partir de textes très variés. Ils lisent pour apprendre, pour s'intégrer dans une société où la lecture joue un rôle essentiel et pour leur plaisir.* ». L'étude pose comme hypothèse que la compétence en lecture est liée aux raisons qui enclenchent la lecture, et le test dans son ensemble est bâti sur le croisement de deux dimensions : *les processus de lecture* et *les objectifs de la lecture*. Une troisième dimension constituée par « *les habitudes de lecture et les attitudes vis à vis de l'écrit* » est prise en compte dans des questionnaires dits de contextes proposés aux enfants sur leurs habitudes de lecture, à leurs parents et à leurs enseignants. Une encyclopédie « *PIRLS 2001 Encyclopedia. A reference guide to Reading Education in the countries participating in IAE's PIRLS* » a été élaborée au cours de ce programme. Les quatre *processus de compréhension* évalués sont les suivants : *Prélever des informations explicites, Faire des inférences directes, Interpréter et assimiler idées et interprétation, Examiner et évaluer le contenu, la langue et les éléments textuels*. Les pourcentages d'items que chaque catégorie doit couvrir se répartissent de manière équivalente pour les deux objectifs de lecture appelés : *Lire pour accéder aux textes littéraires* et *Lire pour acquérir des informations*.

Le matériel

Le matériel d'évaluation est composé de huit textes : quatre « littéraires » et quatre « informatifs » qui servent à constituer dix « *cahiers tournants* ». Nous mettons entre guillemets les termes « littéraires » et « informatifs » en reprenant les termes génériques utilisés dans PIRLS. En effet, si nous cherchions à caractériser chaque texte en usage dans ce dispositif, nous n'arriverions pas à deux catégories vu la variété des textes proposés. Chaque cahier comporte un texte « littéraire » et un texte « informatif », l'enfant dispose de 60 minutes fractionnées en deux temps pour accomplir sa tâche d'évaluation.

Les textes « littéraires », textes complets, sont des fictions narratives dans lesquelles « *le lecteur s'engage pour s'investir dans des événements, des actions, des conséquences, .. des sentiments, des idées imagées et pour apprécier la langue elle – même* ». Les textes « informatifs » se présentent sous des formes variées (prospectus, texte documentaire portant sur des comportements animaux, ...) et l'information y est présentée, ou non, sous forme chronologique. Des illustrations sont insérées dans les cahiers, et l'un d'eux se présentant comme un petit livre est illustré en couleur, les questions étant proposées sur un autre support, ceci afin de proposer à l'élève un support de lecture motivant.

Les questions

Comme on l'a indiqué précédemment, les questions visent à vérifier que les élèves maîtrisent la mise en œuvre de quatre processus de lecture dans deux « types de textes ». Ces processus sont dénommés *compétences* dans la publication de la DEP. La moitié des questions sont présentées sous forme de

QCM à quatre modalités dont une seule est correcte. L'autre moitié des questions nécessite de construire des réponses qui sont corrigées en fonction d'un barème en points, elles valent 1, 2 ou 3 points selon le niveau de compréhension requis pour répondre, ou selon l'étendue des connaissances textuelles sollicitées. Il s'agit du même système de correction aboutissant à un score que dans les évaluations anglaises présentées dans ce rapport ; il faut noter que l'organisme NFER a construit les épreuves de PIRLS. Ces deux formats de questions sont utilisés de manière inégale selon le processus évalué, car les QCM par exemple, ne se prêtent pas bien à la formulation de questions qui ciblent les capacités interprétatives ou des évaluations complexes du contenu du texte.

Le contenu des quatre processus évalués par les deux formats d'items s'exemplifie par des tâches représentatives de chacun d'eux :

- Prélever des informations explicites (*Prélever* – 20 % des items) : repérer les informations directement liées à l'objectif de la lecture, chercher des informations précises, chercher la définition de mots ou d'expressions, repérer le contexte de l'histoire, retrouver l'idée principale, etc..
- Faire des inférences directes (*Inférer* – 40 % des items) : déduire que tel événement a entraîné tel autre, déduire l'élément principal d'une série d'arguments, décrire la relation entre deux personnages, etc..
- Interpréter et assimiler idées et interprétation (*Interpréter* – 25 % des items) : déduire le message global ou le thème d'un texte, envisager une alternative aux actions des personnages, saisir l'atmosphère ou le ton du récit, etc..
- Examiner et évaluer le contenu, la langue et les éléments textuels (*Examiner* – 15 % des items) : évaluer la probabilité que les événements décrits se passent réellement, décrire la manière dont l'auteur a amené la chute, juger de l'intégralité ou de la clarté des informations fournies dans le texte, décrire comment le choix des adjectifs modifie le sens.

La construction des scores de PIRLS

Une échelle de scores a été élaborée à partir des modèles de réponse à l'item (M.R.I.). Par cette approche, il est possible de situer les performances des élèves sur une échelle, même si ils ont répondu individuellement à des cahiers différents (cahiers tournants). Comme dans PISA, la moyenne internationale est fixée à 500 et l'écart –type à 100 ; les pays sont classés sur une dimension commune. En termes statistiques, cela signifie qu'environ les deux tiers de la population testée ont un score compris entre 400 et 600 points. Cette méthodologie met l'accent sur la comparaison des performances entre les pays.

Les résultats de la France à PIRLS

Résultats généraux

Si la France a obtenu un score de 525, significativement supérieur à la moyenne internationale, elle occupe un rang médian au classement des pays par scores. Les pays qui obtiennent les meilleurs scores sont : la Suède (561), les Pays – Bas (554), l'Angleterre (553) et la Bulgarie (550). La notion de palmarès est toujours présente dans ce genre d'enquête, et celle ci n'y échappe pas. Douze pays ont des scores significativement supérieurs à la France, sept ont des performances comparables et quinze font moins bien qu'elle. Quand on reconstruit les tableaux de résultats en prenant des critères de proximité géographique ou des critères de nature économique, la position française n'est pas plus enviable : le score se trouve en deçà des nouvelles moyennes de références (par exemple, Union européenne 2001 – 7 pays sur 15 - ou 2004 – 14 pays sur 25, respectivement 541 et 534, ou encore référence aux pays de l'OCDE participant à PIRLS 16 pays sur 30, score français = 529). Si l'on s'intéresse aux dispersions des résultats, l'écart –type français est de 70 (écart moyen de la population générale fixé à 100), les résultats sont assez peu dispersés. Comme dans tous les pays, les filles obtiennent toujours des résultats supérieurs à ceux des garçons, il faut noter que c'est en France et en Italie que cette différence est la plus faible.

Nous avons reporté dans le tableau ci dessous la ventilation des résultats de la France et des trois pays qui sont en tête du classement des pays en quartiles, médiane et dernier décile.

Comparaison des caractéristiques de la distribution des résultats français et de trois autres pays.

Scores	1 ^{er} Quartile 435 et plus	Médiane 510 et plus	3 ^{ème} Quartile 570 et plus	9 ^{ème} Décile 615 et plus
France	90 %	60 %	26 %	9 %
Suède	96 %	80 %	41 %	20 %
Pays Bas	98 %	79 %	40 %	14 %
Angleterre	90 %	72 %	45 %	24 %

Si 90 % d'élèves de CM1 atteignent le score de 435, représentant le premier quartile de la population totale, 10 % de nos élèves sont en dessous de ce niveau. Le score médian (510) est atteint par 60 % de nos élèves, alors que 26% parviennent au troisième quartile (score 570) par 26 %. Enfin, pour le score limitant le neuvième décile (615), la population française est proche des 10%, comme dans la population générale de l'enquête, mais beaucoup moins performante que celles des trois pays portés dans le tableau.

En conclusion, la population française ne se démarque pas trop des pays les plus performants à PIRLS pour la distribution des plus faibles niveaux de performances, mais au delà, elle est en total décalage négatif.

Quelle est la signification des scores ?

Scores de 435 et plus : Pour les textes « littéraires », les élèves manifestent leur compréhension d'histoires courtes comprenant un ou deux épisodes et essentiellement deux personnages principaux. Ils peuvent retrouver et citer des informations explicitement exposées dans le texte qu'il s'agisse d'une narration, d'une description ou d'un dialogue, ces informations sont relatives aux actions ou aux sentiments des personnages. A partir de matériaux informatifs divers, mais courts et organisés dans un ordre chronologique ou thématique, ils peuvent retrouver et citer des informations explicites concernant les personnes, les animaux et les lieux. Ils sont capables de faire des inférences très simples et fortement suggérées par le texte.

Scores de 510 et plus : A ce niveau, les élèves sont capables de faire des inférences et des interprétations simples relatives aux actions et objectifs des personnages, à partir d'éléments puisés dans différentes parties du texte. Ils peuvent aussi identifier des éléments portant sur l'organisation textuelle. Pour les textes « informatifs », ils sont capables d'inférences simples, d'extraire des informations pertinentes à partir d'une brochure contenant divers matériaux (textes, carte, illustrations, ...), et de donner une appréciation générale sur l'ensemble du texte.

Scores de 570 et plus : Pour les textes « littéraires », les élèves sont capables de faire des inférences basées sur différentes caractéristiques des personnages et des événements, et de les justifier en s'appuyant sur le texte. En particulier, ils peuvent faire des inférences pour décrire les personnages en les contrastant à partir de leurs actions, de leur caractère ou de leurs sentiments. A partir de matériaux informatifs divers, mais courts et organisés dans un ordre chronologique, ils peuvent retrouver des informations enchevêtrées, réaliser des inférences fondées sur plusieurs phrases, faire des interprétations en faisant appel à leurs connaissances et expériences personnelles. Ils comprennent les métaphores simples.

Scores de 615 et plus : Pour les textes « littéraires », les élèves manifestent leur compréhension d'histoires courtes. Ils sont capables d'interpréter les sentiments et les comportements des personnages en s'appuyant sur le texte. Ils sont aussi capables d'intégrer des idées puisées au long du texte pour en dégager le thème ou le sens général. Ils sont capables d'intégrer des informations provenant de différentes parties d'un texte « informatif », et après les avoir combinées, de les appliquer à des situations réelles.

Résultats analytiques

La variable Processus de lecture.

Les trois premiers processus portent sur le traitement d'informations contenues explicitement ou non dans le texte. Dans tous les pays, les items **Prélever** des informations explicites sont en moyenne mieux réussis que ceux qui nécessitent d'**Inférer**, eux mêmes mieux réussis que ceux qui requièrent d'**Interpréter**. La France est toujours au dessus de la moyenne internationale, et obtient le score le plus élevé pour « Inférer ». Le processus **Apprécier** qui met en jeu des aspects métalinguistiques de la compréhension impose de prendre de la distance par rapport au texte. Il aboutit en moyenne à de plus fortes réussites qu'« Interpréter », mais les scores français dérogent à cette règle et se montrent équivalents. Parmi les différences internes aux pays, voir aux langues, il convient de noter le cas des pays anglophones qui se démarquent des autres pays en ayant un bien meilleur classement pour « Interpréter » et pour « Apprécier ».

Scores et rangs des élèves Français aux quatre processus de lecture.

	Prélever	Inférer	Interpréter	Apprécier
Score français	521	524	514	515
Rang	15	16	20	14

La variable Objectifs de lecture

Nos élèves réussissent légèrement mieux les items portant sur les textes « littéraires » que ceux qui portent sur des textes « informatifs », ce qui apparaît cohérent avec les données recueillies par les questionnaires que les enseignants ont remplis concernant leurs pratiques. La France se place au troisième rang international pour une épreuve basée sur un support informatif : il s'agit d'un dépliant publicitaire présentant les prestations d'un loueur de bicyclettes et proposant un plan de piste cyclable près de laquelle sont localisés un certain nombre d'attractions et de lieux culturels. Pour le fabliau tiré de la culture indienne, la France se place en dixième position. Autrement dit, et malgré les précautions prises dans les procédures de pré-test, il existe une certaine variabilité des performances selon le support. A l'inverse, pour l'un des textes « littéraires », la France se place au vingtième rang. Pour ce texte, il est probable que les élèves ont éprouvé à des difficultés liées à un alourdissement du texte dans sa version française, ce qui a pu leur imposer une surcharge cognitive dans le traitement qu'ils devaient en faire.

Le format des questions

Les élèves français paraissent influencés par le format des questions puisqu'ils réussissent nettement mieux les QCM que les items leur demandant de rédiger une réponse plus ou moins longue. Ceci vaut pour tous les processus évalués, même pour ceux qui semblent de plus bas niveau comme Prélever. Ceci se retrouve dans les autres pays. Cette sensibilité au format de l'item est relativement classique. Une analyse de variance intégrant les variables « Processus » et « Format des questions » met en évidence, en France, le rôle de ces deux variables qui expliquent à elles deux 35% de la variance. Cependant, la variable Processus est responsable de 30% de la variance expliquée, ce qui signifie qu'elle joue un rôle plus important sur les performances que le Format de questions.

Les non – réponses

Nos élèves ont une tendance plus forte que dans les autres pays à s'abstenir de répondre aux questions à réponses construites. Pour les QCM, le taux de non réponses est de l'ordre de 1%, ce qui place la France au septième rang pour cet indicateur alors que ce taux atteint 10% pour les items nécessitant de produire une réponse écrite, nous plaçant au vingt troisième rang. Ceci veut dire que nos élèves, plus que ceux d'autres pays, se réfugient en ne répondant pas à un certain nombre d'items pour lesquels ils devraient rédiger une réponse. Le même constat a été fait pour PISA. Faut-il en voir les racines chez nos élèves de CMI, et surtout quelles en sont les raisons ? Par ailleurs, un des questionnaires d'attitudes face à la lecture révèle que nos élèves ont tendance à sous-estimer leurs capacités de lecture. Faut-il chercher un lien avec ce qui pourrait être interprété comme une peur de l'erreur ?

En résumé

L'enquête PIRLS place les élèves dans des situations de lecture longue de textes complets (environ 1800 mots). Ils sont ensuite soumis à une séquence de douze à treize questions pour la moitié desquelles, en moyenne, ils doivent produire une réponse par écrit, la durée de l'épreuve étant d'une heure. Les élèves Français scolarisés en CM1 obtiennent des résultats un peu supérieurs à la moyenne internationale et réussissent légèrement mieux les items portant sur les textes « littéraires », tels qu'ils sont classés dans l'enquête. Ces résultats sont modulés par le Processus de lecture et le Format de la question. Si l'âge moyen de la population internationale au moment de la passation de l'épreuve était de 10, 3 ans (entre 9, 7 ans et 11, 2 ans, selon les pays), une répartition des élèves en trois groupes d'âges montre que les Français appartiennent au groupe des plus jeunes. Il n'est pas exclus qu'une variable liée à l'âge ait pu jouer sur leurs performances. Pour expliquer les résultats, il faudrait mettre à l'épreuve des hypothèses et se livrer à des études complémentaires, notamment analyser finement les réponses des élèves pour tenter de cerner les obstacles qu'ils ont rencontrés.

Nous en venons maintenant à l'enquête *Reading literacy* menée dans le cadre de l'année internationale de l'alphabétisation par l'IEA, qui s'était lancé en 1988 le défi suivant : créer un instrument international de mesure de la « reading literacy » et évaluer le niveau des élèves de 9 ans et de 14 ans dans une trentaine de pays.

L'enquête Reading literacy de l'IEA (1991)

Dans cette enquête la « reading literacy » a été définie comme « *la capacité à comprendre et à utiliser les formes du langage écrit requises pour la société et / ou valorisées sur le plan individuel* ». Les épreuves de compréhension de l'écrit ont pour support : des textes informatifs, narratifs ou des documents. Presque tous les items sont de format QCM.

Population A - élèves de 9 ans

Le matériel était composé d'un cahier regroupant diverses épreuves dont un test de reconnaissance de mots qui a été particulièrement bien réussi par les élèves français (près de 100 % de bonnes réponses). Pour les épreuves de compréhension de textes, seules quatre questions demandaient des réponses construites. Aucune analyse n'a été faite par l'IEA pour caractériser quel processus ou quelle compétence est visé par chaque item. Comme pour PIRLS, les scores ont été normalisés, leur moyenne étant fixée à 500 et leur écart –type à 100. Le score global de la France est de 531 (écart –type 74), ce qui la place au quatrième rang des 27 pays participants, après la Finlande, les Etats – Unis et la Suède. Les élèves français scolarisés en CM1 comme dans PIRLS obtiennent un score moyen de 532 (93) pour les textes narratifs, de 533 (84) pour les textes informatifs et de 527 (81) pour les documents.

Des comparaisons entre Reading literacy 91 et PIRLS ?

Les deux dispositifs ont été conçus dans les deux cas par l'IEA avec une grande rigueur et des méthodologies proches. Les modèles I.R.T. sont utilisés dans les deux études.

Le critère de même niveau de scolarisation : La population ciblée dans cette partie de l'enquête a sensiblement le même âge que dans PIRLS, elle est scolarisée en CM1. On a vu dans l'enquête DEP de suivi des populations 87 – 97 que la limitation des redoublements imposée en 1990 modifie sensiblement la composition des populations scolaires.

Les épreuves sont différentes. Les scores normalisés n'ont pas été obtenus à partir d'épreuves dont l'équivalence a été vérifiée. L'analyse rapide, que nous avons menée, montre que les processus sollicités ne sont pas identiques.

Les pays participants ne sont pas tout à fait les mêmes ; si seize pays ont participé aux deux programmes, sept seulement ont pratiqué les deux évaluations aux mêmes âges moyens.

En conséquence, les comparaisons entre *Reading Literacy* et PIRLS paraîtraient particulièrement fragiles.

Population B de Reading Literacy – 14 ans.

Le matériel était composé des mêmes types de supports que pour la population A, à l'exception du test de reconnaissance de mots non utilisé avec les élèves les plus âgés. La France se place au deuxième rang après la Finlande et la Suède, où les élèves légèrement plus jeunes, sont scolarisés en septième année alors qu'en France, ils sont en huitième année (en termes de définition de l'enquête). Le score global de la France est de 549 (écart –type 68), le score moyen est de 556 (86) pour les textes narratifs, de 546 (84) pour les textes informatifs et de 544 (77) pour les documents.

Au total, pour cette enquête, la France a obtenu pour ses deux populations des résultats de bon niveau. On remarquera que les exercices étaient courts et que la modalité de réponse quasi exclusive était le QCM.

Le programme PISA

Programme International d'Évaluation des Acquis des élèves de 15 ans

Ce programme évalue trois domaines dénommés « *reading literacy* », « *mathematical literacy* », « *scientific literacy* ». Il est organisé en trois cycles de trois ans, chacun comportant une « dominante » (« *reading literacy* » en 2000, mathématiques en 2003, sciences en 2006, (« *reading literacy* » en 2009). Ce programme cherche à évaluer « *l'aptitude des jeunes à exploiter leurs connaissances et leurs compétences pour faire face aux défis de la vie réelle plutôt que d'examiner dans quelle mesure les élèves ont assimilé une matière spécifique du programme d'enseignement* ». Cette évaluation est axée sur leur capacité de résolution de problèmes dans des contextes proches de la vie quotidienne. Afin de mieux comprendre les résultats de l'évaluation directe des élèves, des données contextuelles ont été recueillies par des questionnaires aux parents, aux établissements scolaires et aux élèves eux mêmes.

Le contenu de PISA

PISA étant mis en œuvre dans plus de trente pays, la perspective de fonder directement l'évaluation à partir des programmes scolaires a été rapidement abandonnée. Ce n'est donc pas un outil qui permet d'apprécier l'atteinte des objectifs que les programmes fixent au système scolaire. L'accent est placé sur la maîtrise des processus, la compréhension de concepts et la capacité d'être efficace dans des situations variées.

Le cadre de référence : définition, domaines évalués

Un cadre de référence élaboré en langue anglaise sert de base à la construction des outils d'évaluation qui a été confiée à un consortium dirigé par A.C.E.R. (*Australian Council for Educational Research*). La France avait répondu à l'appel d'offres préalable dans le cadre d'un Consortium européen, son influence n'a pas été négligeable dans la mesure où des aspects théoriques de son projet ont été intégrés dans PISA et qu'un expert français a été nommé pour participer au programme. De plus, à l'initiative de la France, il y a eu des efforts de révision des traductions du matériel initialement conçu en anglais (version source) vers le français (autre version source).

La définition de la compréhension de l'écrit, spécifique, à PISA est la suivante : « *La capacité de comprendre, d'utiliser et d'analyser des textes écrits afin de pouvoir réaliser ses objectifs, développer ses connaissances et son potentiel et jouer un rôle actif dans la société* ». Il fixe les trois principales dimensions à prendre en compte dans l'organisation de ce dispositif d'évaluation : *les tâches de lecture*², *les types de textes* et *le contexte*. Les aspects de la compréhension à évaluer ont été fixés en croisant ces dimensions et en les faisant varier.

Les trois « types de tâche de lecture » sont prévus pour que les élèves montrent leurs capacités à *Retrouver des informations dans des matériaux variés*, *Interpréter ce que l'on lit* et *Réfléchir à ce qu'on vient de lire* (*Retrieve, Interpret, Reflect on*). Pour les deux premières catégories appelées désormais **S'informer** et **Interpréter**, seul le recours au texte (partie ou tout) est sollicité alors que pour **Réagir**, où il faut réfléchir sur le contenu et la forme des écrits, les lecteurs doivent faire intervenir leurs connaissances du monde, évaluer ces écrits en sachant argumenter leur point de vue.

Les « types de textes » donnés à lire sont soit des textes continus (narratifs, descriptifs ou argumentatifs) ou des textes non continus (listes, formulaires, diagrammes, graphiques, etc.), respectivement 66% vs 33 %.

L'utilisation pour laquelle le texte a été écrit est intégrée sous forme de la variable « **contexte** » qui se

² Dans le cadre de référence établi à l'origine en anglais, cette dimension est indiquée par les mots anglais « *questions* » ou « *aspects* ». Elle a été traduite en français, par l'OCDE, alternativement par « processus » (employé avec guillemets) ou par tâches de lecture, ou type de tâches de lecture. Dans les publications propres à la France, c'est le terme « compétences » qui a été employé.

décline sous usage privé, publique, professionnel et scolaire.

L'évaluation du niveau des élèves est fondée sur leurs réponses à des tâches (questions) de compréhension de natures diverses. Pour chaque tâche, on peut distinguer les caractéristiques des textes sur lesquels porte la tâche et celles qui sont liées à la question elle-même. Les textes peuvent être de niveau de difficulté variable. Par construction, les questions sont de degrés de difficultés variables, mais il faut remarquer que les trois types de tâches de lecture, objets centraux de PISA, par leur nature se prêtent avec plus ou moins de facilité à un questionnement par QCM (Questionnaire à choix Multiple) ou par questions ouvertes. Ainsi, le QCM est peu approprié à la tâche « Réfléchir sur un texte » qui est majoritairement évaluée par des questions suscitant une réponse écrite, parfois longue. Les pourcentages de questions fermées et de questions ouvertes sont au total respectivement de 55 % et 45 %. Un total de 141 items constitue le pool d'items répartis en cahiers tournants dont 129 pour PISA, les autres items provenant de l'enquête I.A.L.S. (*International Adult Literacy Survey*), nous n'en tiendrons pas compte. Chaque élève de l'échantillon a été soumis à une épreuve d'une durée de 2h divisée en deux séquences.

La population.

PISA vise une population scolarisée née en 1984, ayant entre 15 ans 3 mois et 16 ans 2 mois au moment de l'évaluation. Trente deux pays ont ainsi soumis un échantillon de leur population de 15 ans à PISA en 2000. En France, les élèves de 15 ans sont dispersés entre le collège et les lycées dans plusieurs niveaux scolaires et différents cursus. Un peu plus de la moitié de notre échantillon est à l'heure ou en avance dans sa scolarité (43% en seconde Générale et technologique) et 36 % en classe de troisième, 10 % en quatrième.

Les échelles de compréhension

Pour chaque « type de tâche de lecture », la difficulté des items est contrôlée de manière à pouvoir situer le niveau réel de la population et ses acquis sur une échelle de difficulté dite « échelle de compréhension ». Selon l'OCDE, l'échelle « *Retrouver de l'information* » renvoie à la « *capacité* » de l'élève à localiser des informations dans un texte. Pour la compréhension de l'écrit, une **échelle combinée** regroupe les résultats des **trois échelles spécifiques** (S'informer, Interpréter, Réagir).

Pour faciliter l'interprétation des résultats, la moyenne de chaque échelle a été fixée à 500 points et l'écart – type qui exprime la dispersion des résultats à 100. Les scores des 2/3 de la population participante sont compris entre 400 et 600. Chacune des trois échelles de compréhension de l'écrit comporte cinq niveaux dits de « *connaissance et de compétence* » ou « *niveau de compétence* », qui traduisent la variation des tâches en terme de difficultés. Chaque élève de l'échantillon obtient un score pour chaque échelle. Cependant, différents profils de réponses peuvent conduire au même score (Rocher, sous presse).

Contenu des « niveaux de compétences »

Les résultats en lecture sont présentés sur des échelles de compétences ; sur chaque échelle, cinq niveaux de compétences hiérarchisés ont été définis. A chaque niveau correspondent des tâches de lecture que les élèves sont capables d'accomplir avec une certaine réussite. Il n'y a pas de maximum ni de minimum sur les échelles ; les scores ne représentent ni des points, ni des pourcentages de réussite. Le principe est d'exprimer les scores en écarts par rapport à une moyenne fixée arbitrairement à 500, afin de faciliter les comparaisons d'une échelle et d'une étude à l'autre. Les modèles I.R.T. gouvernement PISA, comme PIRLS.

Les résultats généraux – échelle combinée

Les élèves français obtiennent le score 505, proche de la moyenne internationale fixée à 500. La Finlande se classe première avec un score de 546, les pays anglo-saxons, à l'exception des Etats Unis (score de 504), réalisent de bonnes performances à cette évaluation. En revanche, l'Allemagne (485) et les pays de l'Europe de l'Est et du Sud réussissent moins bien.

Les niveaux de compétence et les tâches leur correspondant sont listés ci - dessous.

- **Niveau 5 plus de 626** : Les élèves sont capables de mener à terme des tâches de lecture élaborées telles que traiter des informations difficiles à trouver dans des textes non familiers, faire preuve d'une compréhension fine et déduire les informations pertinentes pour la tâche à accomplir, être capable d'évaluer de manière critique et d'élaborer des hypothèses, faire appel à des connaissances spécifiques et développer des concepts contraires aux attentes.
- **Niveau 4 de 552 à 626** : Les élèves sont capables de réaliser des tâches de lecture complexes comme extraire des informations enchevêtrées , utiliser les nuances de la langue pour interpréter, évaluer de manière critique un texte.
- **Niveau 3 de 480 à 552** : Les élèves sont capables d'effectuer des tâches de lecture de complexité modérée telles que repérer plusieurs éléments d'information, établir des liens entre différentes parties du texte et les relier à des connaissances familières de la vie courante.
- **Niveau 2 de 408 à 480** : Les élèves sont capables d'effectuer des tâches de lecture de base, notamment localiser des informations explicites, établir des inférences de niveau élémentaire dans des textes variés, comprendre le sens d'un passage limité du texte et le relier à des connaissances familières.
- **Niveau 1 de 335 à 407** : Les élèves sont capables de réaliser les tâches les plus simples de PISA : repérer un élément simple, identifier le thème principal d'un texte ou d'établir une relation simple avec les connaissances de la vie courante.
- **Au dessous du niveau 1, moins de 335** : Les élèves ne sont pas capables de mettre couramment en œuvre les connaissances et les compétences les plus élémentaires requises par PISA. Cela ne peut être interprété comme une absence totale de compétences en lecture, mais il faut s'interroger sur leurs possibilités d'utiliser la lecture comme un outil au service de la vie courante.

Répartition en pourcentages des élèves selon les Niveaux de performances en compréhension pour la France et l'OCDE

	France	OCDE
Niveau 5	8	9
Niveau 4	24	22
Niveau 3	31	29
Niveau 2	22	22
Niveau 1	11	12
Au dessous	4	6

Les résultats français sont peu dispersés si on les compare à d'autres pays : 4,2 % des élèves sont classés dans le plus bas niveau de performances et 8, 5 % dans le plus haut niveau (Belgique, respectivement 7,7 et 12 %) Notre distribution est très proche de la distribution internationale. Ces résultats masquent une forte variabilité des résultats puisque la France se classe parmi les cinq premiers pays pour 29 items et pour 29 autres items dans les dix derniers. Ce constat très grossier incite à la prudence pour tirer des conclusions sur les résultats. En effet, PISA est présenté comme une mesure unidimensionnelle robuste d'une compétence de portée générale, ce dont ces résultats disparates font douter (pour une discussion, Rocher, sous presse).

Les trois échelles de compréhension de l'écrit sont les composantes de l'échelle générale de compréhension que nous avons présentée au point précédent. Nous allons maintenant nous intéresser à chacune des dimensions. Le tableau ci – dessous présente les résultats en % pour chacune de ces dimensions, pour la France et l'OCDE (entre parenthèses).

	S'informer	Interpréter	Réagir
Niveau 5	13, 2 (11,6)	9 (9, 9)	8,6 (10, 9)
Niveau 4	25, 2 (21, 2)	23, 4 (21, 7)	21 (22, 5)
Niveau 3	27 (26, 1)	30, 3 (28, 4)	28, 7 (27, 6)
Niveau 2	19,2 (20, 7)	21, 8 (22, 3)	23, 4 (20, 7)

Niveau 1	10, 5 (12, 3)	11, 5 (12, 2)	12, 5 (11, 4)
En dessous Niv. 1	4, 9 (4, 9)	4 (5, 5)	5, 9 (6, 8)
Moyennes	67, 9 (64, 4)	64, 3 (63, 3)	51 (52)

Les élèves français retrouvent un peu plus facilement l'information (67, 9 % de bonnes réponses) qu'ils ne réalisent des inférences ou interprétation du texte (64, 3 % de bonnes réponses). Les performances chutent pour *Réagir*, là où le lecteur doit prendre de la distance par rapport au document proposé. Ces performances aux trois tâches de lecture sont proches des moyennes internationales. Cependant, selon les pays, les configurations peuvent être très différentes de cette apparence donnée par les moyennes ; ainsi, la Grande Bretagne et les Etats Unis réalisent une bien meilleure performance à *Réagir* qu'à *S'informer*, où vérification faite auprès de nos collègues des deux pays, la sollicitation n'est pas nouvelle pour les élèves.

Le pourcentage d'élèves français parvenant au niveau 5 est supérieur à la moyenne internationale pour *S'informer*, et nettement inférieur à celle - ci pour *Réagir*. Typiquement, certains pays atteignent des proportions d'élèves plus élevées au niveau 5 pour la démarche *Retrouver de l'information*, tandis que dans d'autres on observe ce phénomène pour la démarche *Réagir* qui demande de *Réfléchir sur le texte* -.

Effet des supports et des questions

Les élèves français obtiennent les meilleurs résultats aux deux exercices de PISA portant sur des supports proposés par la France, même si les questions ont été modifiées pour les besoins du dispositif (texte d'Anouilh et texte documentaire sur le Pôle Sud). De tels effets ont été retrouvés pour d'autres pays.

Pour les QCM, le taux de non réponses est faible et identique à la moyenne OCDE. Cependant, un effet du niveau de performances des élèves existe en France ; plus ils sont faibles et plus ils ont recours à la non réponse, à la fois, aux QCM et aux questions ouvertes. De manière générale, pour les items qui nécessitent de construire une réponse écrite, l'ensemble de nos élèves a recours à des non réponses beaucoup plus souvent que dans les autres pays. La dimension *Réagir* est très concernée puisque ce format de réponse est le plus employé. Les élèves sont confrontés à un format de réponse exigeant et à une tâche peu pratiquée dans notre système scolaire, et c'est encore plus vrai pour les collégiens, qui constituent une part importante de la population évaluée. Le plus souvent, on leur demandait d'exprimer une opinion personnelle, ce qui a dû les désarçonner en situation scolaire, comme le confirment des entretiens (Rémond, 2003). Quand *Réagir* est évalué par des QCM, nos élèves réussissent comme ceux des autres pays.

Nous retrouvons donc comme dans PIRLS, la manifestation d'un taux de non réponses aux questions ouvertes plus élevé en France que dans la moyenne des autres pays. Avons nous une particularité culturelle ou pédagogique qui favoriserait ce comportement et de manière précoce ? Nos élèves ont – ils peur de se tromper ?

Effets du niveau scolaire sur les performances à PISA

Les élèves qui sont en première obtiennent un score moyen de 609, en seconde Générale et Technologique 561, en seconde Professionnelle 476, en troisième 449 et en quatrième 397. Les élèves scolarisés en première ou en seconde G T ont des scores supérieurs à la moyenne française 505, les autres classes ici mentionnées basculent en deçà de cette moyenne et cela d'autant plus fortement que leur retard scolaire s'accroît. Niveau de compétences et retard scolaire sont fortement liés (Murat & Rocher, 2002).

Quelques commentaires sur les enquêtes internationales et sur PISA, en particulier.

Les grandes enquêtes internationales sont souvent riches d'enseignement pour les pays qui y participent. Nous avons montré que si la manière de définir les populations n'est pas la même dans les deux enquêtes (critère d'année de scolarité vs critère d'âge), l'objectif des concepteurs réside dans la garantie de la qualité de la mesure, qui sous - tendra le palmarès final. Des centaines de pages

justifient toutes les précautions qui ont été prises aussi bien pour assurer l'équivalence des traductions que pour mettre en œuvre l'arsenal psychométrique destiné à rendre l'enquête inattaquable. Malgré les efforts faits, des biais subsistent dans PISA (Rocher, sous presse). Tout d'abord, les concepteurs de l'enquête ont fait l'hypothèse d'unidimensionnalité ; pour eux, tous les jeunes de 15 ans vont être placés dans des situations similaires, indépendantes de leur contexte culturel. Ils font aussi l'hypothèse du maintien de la difficulté de la tâche à l'issue de la traduction. Pourtant malgré toutes les précautions prises, l'équivalence de la difficulté des versions ne peut être totalement garantie. Ainsi, Adams et Wu (2002) ont montré que les procédures de traduction allongent, en moyenne, les textes passant de l'anglais au français de 12 % de mots et de 19% de caractères. Aucune interrogation sur les processus en jeu dans les tâches ne figure dans PISA. Nous nous interrogeons sur leur nature et sur la charge cognitive du traitement selon que le texte est rédigé en anglais, en français et bien sûr dans d'autres langues. Par ailleurs, l'adaptabilité de chaque épreuve à chaque contexte culturel va agir sur la robustesse de l'échelle de mesure. Nous avons mentionné quelques biais au cours de la présentation de PISA ; nous renvoyons le lecteur aux analyses détaillées de Rocher (sous presse) pour un panorama des difficultés de cette enquête. Une question forte subsiste : l'hypothèse d'unidimensionnalité se justifiant par une échelle unique de compétence est –elle finalement valide ? Ces analyses évoquent plutôt des différences culturelles entre les différents pays d'une part, entre les objectifs de leur systèmes éducatifs d'autre part, que des dysfonctionnements techniques de l'enquête.

L'enquête PISA s'intéresse davantage à la performance des élèves qu'à leur compétence. En effet, il s'agit avant tout de comparer les performances des élèves de la trentaine de pays participants lorsqu'on leur donne à lire une épreuve supposée identique à l'issue de la traduction.

Peut –on évaluer à l'échelle de plusieurs pays sans partir d'un corpus commun traduit ?

Deux exemples : étude de faisabilité « Socrates » et C- BAR

Ce n'est pas le principe des enquêtes internationales qui pose problème, mais les présupposés sur lesquels celles-ci se fondent. Pour tenter d'éviter les procédures de traduction, une méthodologie alternative a été inventée dans le cadre d'une recherche pilote menée par des chercheurs en sciences cognitives dans le cadre d'un projet « SOCRATES ». Cette recherche avait pour objectif de spécifier les conditions dans lesquelles des données réellement comparatives peuvent être obtenues à partir des épreuves nationales qui utilisent des documents originaux et non traduits (Bonnet et al., 2001). L'étude a porté sur les compétences de jeunes, âgés de 15 ans, scolarisés en Finlande, France, Grande-Bretagne et Italie. Ses trois points forts sont :

- l'utilisation d'un modèle psycholinguistique de la lecture,
- une approche interculturelle qui tient compte des caractéristiques nationales : les épreuves sont construites au niveau national après établissement de règles communes pour les pays concernés
- une méthodologie qui assure la comparabilité des résultats obtenus dans les différentes épreuves nationales : le protocole comporte ainsi la comparaison des performances entre des groupes contrôle de sujets bilingues – anglais/finlandais , italien/français, ... dans les épreuves des deux pays dont ils parlent la langue.

Les résultats sont convaincants : il est possible de mener des comparaisons internationales sans avoir recours à un matériel commun et traduit.

Le projet C-BAR “Culturally balanced assessment of reading”

Une seconde phase associant cette fois huit pays européens, a eu lieu et le rapport vient d'être publié. Il s'agit du projet C - BAR qui a cherché à dépasser les faiblesses révélées par la mise à l'épreuve du projet précédent, tout en intégrant quatre pays supplémentaires de langues différentes. Ces huit pays participants ont élaboré un nouveau tableau de compétences à partir du précédent. Ils ont aussi mis au point une procédure de définition de l'équivalence des textes qui seraient utilisés par les différents pays pour en assurer la meilleure comparabilité. Cette procédure a été extrêmement détaillée à la suite de la première phase de l'enquête. Des critères d'élaboration des questions ont aussi été rédigés. Comme dans la première phase, des procédures d'ancrage ont été prévues tout comme les comparaisons des résultats (NIPALS algorithm, MRI, etc.). Rappelons qu'un modèle de lecture est à la base de « Socrates » et de C - BAR. A partir des procédures évoquées ci dessus, les huit pays ont mis au point des épreuves d'évaluation à partir de deux textes sélectionnés en commun (un « littéraire » et un « scientifique »), textes existant dans les huit pays dans des « traductions dites naturelles ». Chaque pays a construit des items portant sur ces textes à partir des règles définies, en respectant les compétences à évaluer. Ces épreuves serviraient d'ancrage dans la phase d'expérimentation. Il a donc été possible d'élaborer un matériel équivalent sans passer par la traduction. Reste à le soumettre aux élèves des huit pays concernés, sous réserve de financement de l'opération, et du développement d'autres épreuves.

Chapitre 3

Les évaluations nationales de la lecture dans d'autres pays

La variété des pratiques d'évaluation de la lecture est forte dans les pays étrangers (PIRLS Encyclopedia) et même dans les pays proches en Europe (Nyssen, Terwagne, Godevir, 2001). Certains n'ont aucun système d'évaluation en dehors des évaluations internationales, seules sources de données de leur propre système, d'autres ont une culture forte de l'évaluation. Nous avons choisi d'étudier deux systèmes assez différents : celui de l'Angleterre et celui des Etats Unis. Ce choix a été motivé par la forte culture de l'évaluation de la Grande Bretagne dans de nombreux domaines qui dépassent le système éducatif. La NAEP, évaluation nationale américaine, fait partie des outils qui conditionnent fortement les enquêtes internationales et il n'est pas inutile de souligner cette filiation.

L'évaluation britannique : National Curriculum Reading Tests in England

L'évaluation du niveau des élèves fait partir du dispositif d'évaluation du système éducatif anglais. La première moitié des années 1990 a vu la mise en place par une agence indépendante (actuellement, QCA - *Qualification and Curriculum Authority*) et à la demande du ministère anglais, d'un dispositif national de tests standardisés pour la langue anglaise, les mathématiques et les sciences, destiné à évaluer les résultats des élèves par rapport aux programmes nationaux (*National Curriculum*).

Les résultats sont publiés par le ministère, établissement par établissement, et sont repris par la presse sous forme d'un palmarès censé permettre d'identifier les « bons » établissements (Bonnet, 2002). Nous n'examinerons pas ici l'évaluation des élèves par les examens et nous nous limiterons aux dispositifs d'évaluation de la langue anglaise : *National Curriculum Reading Tests in England*. Ces tests seront mentionnés par NCRT dans la suite du texte.

Les programmes sont divisés en trois cycles (K.S. – *Key Stage* – étapes clés) : 5 – 7 ans, 7 – 11 ans, 11 – 14 ans. Une première catégorie d'instructions détaille ce qui doit être enseigné (compréhension de textes, stratégies de lecture, structure du langage et ses variations, littérature, etc.) L'autre partie spécifie les écrits sur lesquels les élèves doivent apprendre à lire. On notera que depuis 1998, l'enseignement de la lecture et de la production d'écrit est régi par la *National Literacy Strategy* qui fournit un large éventail de matériel, et propose une aide pour mettre en œuvre le programme national.

Les programmes précisent les compétences attendues sous forme d'une échelle en huit niveaux ; les enfants doivent normalement parcourir chaque niveau en une moyenne de deux ans. Au Royaume Uni, les classes sont organisées par âge, mais dans les faits, elles sont multi-niveaux. A chaque niveau de compétence attendue correspond une description de ce que les élèves doivent savoir faire.

Les tests NCRT sont obligatoires et passés chaque année, au mois de mai, par les élèves de 7, 11, 14 ans, à la fin de chaque cycle (K.S. – *Key Stage*, respectivement K.S. 1, K.S. 2, K.S. 3). Les performances des élèves de KS. 1 doivent s'inscrire au niveau 3 visé par le test KS1. Des épreuves supplémentaires et facultatives sont cependant proposées pour tous les niveaux intermédiaires, dans ce cas Niveaux 1 et 2 (niveaux 3 à 5 pour les 11 ans, etc.).

Toutes ces évaluations sont développées par des équipes de recherches de NFER (*National Foundation for Educational Research in England and Wales*), à l'exception des évaluations pour les élèves de 14 ans.

Principes généraux de NCRT

La construction des épreuves NCRT repose sur un certain nombre de principes dérivés du programme d'enseignement de la langue maternelle.

Les évaluations doivent refléter le contenu des programmes d'enseignement et les compétences attendues, mais aucune évaluation annuelle ne peut couvrir l'ensemble. Aussi elle doit porter sur la variété la plus large possible de compétences et de types de textes.

Les programmes incluent aussi des aspects liés aux attitudes et il y est recommandé de choisir des textes bien écrits, intéressants, distrayants pour les élèves.

Les réponses données par les élèves doivent s'appuyer sur leurs propres expériences et connaissances. C'est pourquoi la plupart des questions doivent être ouvertes pour que les élèves puissent exprimer leur point de vue personnel, leur explication ou leur opinion en réponse au texte. Cependant, une petite proportion de questions fermées (QCM, appariement, vrai /faux, numérotation, etc.) doit être proposée car elle aide le lecteur peu expérimenté à accéder à une compréhension élémentaire des points essentiels du texte, que l'enfant ne serait pas lui – même capable d'exprimer avec une clarté suffisante par une réponse construite.

Elaboration des tests

Choix des textes

Les programmes imposent de travailler sur une grande variété de textes, littéraires ou non. Aussi, n'importe quelle évaluation annuelle comporte entre 2 et 5 textes de différents types. En général, elle est organisée autour d'un thème qui conditionne le choix des textes (Les feux de forêt, traités de différents points de vue, par exemple). Les textes dits non littéraires sont généralement réunis sous forme d'un magazine ou d'un recueil thématique d'une quinzaine de pages, comportant un sommaire et parfois un éditorial et quelques articles. Les questions s'y référant sont réunies dans un autre livret ; ce principe est valable pour l'ensemble du matériel. Les supports de lecture sont toujours rassemblés dans un livret de présentation attractive et illustrée. Pour le K.S. 1 (7ans), le maître doit évaluer la lecture à haute voix d'un passage choisi parmi une liste d'ouvrages.

La construction des questions

A partir des programmes et des compétences attendues (voir plus haut), l'organisme Q.C.A établit d'autres listes détaillées de compétences, une pour la lecture et une autre pour la production d'écrits, dégageant des niveaux de compétences ; ces listes précisent les opérations mise en œuvre pour lire et produire des textes à chaque niveau. Les constructeurs de tests doivent s'y référer et chaque question élaborée doit pouvoir être mise en correspondance avec un de ces niveaux. Ceci est aussi prévu pour que les maîtres puissent interpréter facilement les résultats des tests. Il est clair que les élèves doivent dépasser la compréhension littérale et faire appel aussi bien à leurs expériences qu'à leurs connaissances personnelles.

Le codage des réponses

Pour les questions ouvertes, il est construit pour pouvoir attribuer des points en fonction de la qualité de la réponse. Chaque élève obtient un score qui sera rapporté à un niveau de la liste de Compétences attendues. A partir des scores obtenus aux évaluations, un niveau est dégagé : « *National Curriculum Level* ».

La correction des tests

Les maîtres n'assurent le codage que pour le K.S. 1. Un organisme extérieur spécialisé code, en six semaines environ, les protocoles des KS 2 et KS 3, qui sont alors retournés aux établissements, vers fin juin. Les enseignants peuvent, s'ils le souhaitent, se livrer à une analyse des épreuves de leurs élèves, mais la fin de l'année scolaire est proche pour en tirer un bénéfice direct.

La publication des résultats

Tous les ans vers le mois de novembre, QCA publie un rapport appelé « *Standards report* », dans lequel les enseignants peuvent trouver des conseils pédagogiques établis sur la base de l'analyse des erreurs et des réussites aux questions, et sur celle de profils de réponses. Cette Agence produit aussi un CD – ROM appelé « *Textbase* » qui permet aux enseignants d'entrer les résultats de leurs élèves et de les comparer aux standards nationaux. Cependant tout ceci a lieu plusieurs mois après le passation des tests.

Déroulement des épreuves

Les consignes pour les élèves sont toujours très détaillées et données à la fois oralement et par écrit sur les cahiers. En règle générale (sauf pour les 7 ans), le temps alloué est précisé en distinguant la lecture du livret de celle de la réponse aux questions vs la préparation du plan pour la production écrite de celle de la rédaction. Des questions d'entraînements sont proposées pour s'assurer de la compréhension de ces consignes. Les différents types de questions qui seront rencontrés dans l'épreuve sont aussi insérés au début du livret de lecture et commentés. Enfin, le nombre de points maximum accordé à chaque question est indiqué en regard de celle –ci.

Quelques exemples d'épreuves

Key Stage 1 –7 ans – exemple de l'épreuve de 2003

Une partie du test consiste en une épreuve d'orthographe. Pour évaluer la lecture, deux textes sont proposés, l'un porte sur une fiction « *Grandfather's pencil* » et l'autre sur un dépliant décrivant les activités d'un musée. Parmi les compétences évaluées pour le premier texte, on note « *Juger l'atmosphère d'une histoire* » ou encore « *Identifier l'idée principale* ». Les questions fermées (QCM ou autre procédé) représentent moins de la moitié de l'épreuve de lecture, ce pourcentage sera bien plus faible pour les autres K.S.. Le score total de lecture peut atteindre 26 points ; on considère que les élèves ayant obtenu entre 15 et 26 sont au niveau requis et que pour les autres, ce n'est pas le cas (le niveau fixé pour le K.S. 1 étant *Level 3*). Afin d'évaluer ces derniers au mieux on peut leur faire passer des tests « additionnels » et facultatifs envoyés aux écoles en même temps que les autres tests.

Les maîtres peuvent aussi transformer les scores de leurs élèves (total lecture – orthographe) en « *âge standardisé* », et situer chacun de leurs élèves dans un tableau croisant ces deux éléments. Le score moyen standardisé est de 100. On rapporte la performance à l'âge réel, comme dans le test « *Neale Analysis of Reading Ability* » (où l'on obtient un Age de lecture) et l'on peut apprécier comment se situe l'élève par rapport à sa classe d'âge.

Key Stage 2 –11 ans – exemple de l'épreuve de 2001

Trois champs sont évalués : la lecture, l'orthographe et l'écriture, la production d'écrits. Au total, 2 h 15 sont consacrées à l'évaluation de la langue maternelle. Pour la lecture, un magazine en couleur appelé « *Ocean Voices* » a été constitué. Les questions portent sur chacun des trois textes et sur la technique éditoriale générale ; dans deux questions, la compétence ciblée est la suivante : « *Comprendre une technique éditoriale. Donner son jugement personnel.* ». Les enfants peuvent choisir la production d'écrit qu'ils vont réaliser parmi quatre propositions. A la base, il y a toujours un texte sur lequel se greffe une situation de production. Deux domaines sont évalués : le contenu et son organisation, et la « *grammaire* » - style et ponctuation.

Key Stage 3 – 14 ans – exemple de l'épreuve de 2003

La lecture – compréhension de l'écrit est évaluée comme précédemment grâce à un support du type magazine documentaire illustré donnant uniquement lieu à des questions ouvertes, dont certaines requièrent des réponses très longues. Une heure est attribuée pour les questions après les 15 mn exclusivement dévolues à la lecture du « magazine ». Pour certaines questions, le codage extrêmement raffiné peut aller jusqu'à un score de 5 points. Des questions portent sur les procédés utilisés par les auteurs. La compréhension est aussi testée sur trois extraits de Shakespeare, qui servent aussi à des tâches de production de textes (temps total : 1h 15 par texte). Enfin à partir d'un message électronique portant sur un sauvetage, il faut s'improviser journaliste et réaliser un reportage écrit pour lequel on donne des recommandations sur les personnes à interroger et sur le ton à adopter. Sur la page en regard de cette situation initiale, on donne un guidage : « *des mots et des phrases à intégrer dans le titre et dans la première phrase de votre texte, lister les événements de ce sauvetage, les commentaires des témoins, les points sur lesquels insister dans ce reportage* ». On laisse 15' pour cette préparation, puis 30' pour se consacrer à la rédaction elle – même.

Les productions écrites sont jugées sur trois registres s'articulant avec les compétences listées par QCA : la structure des phrases et la ponctuation, la structure du texte et son organisation, le contenu et les effets produits.

Un temps total de 5 h 45 est dévolu à l'évaluation de la langue des élèves de 14 ans. Comme pour le K.S. 2, les résultats doivent être consultés sur le site Web de Q.C.A après une étude centralisée des résultats permettant de fixer des seuils de réussite.

L'évaluation américaine de la lecture NAEP - National Assessment of Educational Progress.

Un des premiers dispositifs d'évaluation de grande ampleur est, certainement, l'enquête américaine NAEP dont la première phase eut lieu en 1969. Cette enquête de type bilan, reproduite périodiquement, a pris en compte différentes disciplines et différents niveaux scolaires. Le choix du Ministère américain n'a pas été de créer un service scientifique de haut niveau, mais de faire financer, par le *NCES (National Center for Education Statistics du département américain d'Education)*, lui-même sous le contrôle du NAGB (National Assessment Governing Board), des organismes scientifiques chargés de piloter l'ensemble de l'enquête jusqu'à la publication des résultats. Depuis 1983, cette mission est confiée à ETS (*Educational Testing Service*), le plus important organisme mondial dans le domaine de l'évaluation et de la recherche (signalons qu'ETS est un organisme privé). Cependant, la NAEP apparaît sous le label NCES. Elle est aussi connue comme « *The Nation's Report Card* ».

La NAEP apparaît par son ancienneté et les travaux scientifiques et méthodologiques qui ont accompagné son développement, comme un dispositif prototypique. La NAEP repose une méthodologie basée sur les Modèles de Réponses aux Items (MRI) pour le traitement et la présentation des résultats. Un soin particulier a été apporté à l'organisation des plans d'enquête et des questionnaires en vue de permettre un suivi des performances dans le temps ainsi qu'un suivi du début de la scolarité à l'âge adulte. De plus, tant les résultats de la NAEP que ses méthodologies ont fait l'objet de nombreuses publications. Par ailleurs, ETS a tenté d'en faire une référence pour les enquêtes internationales, et il est facile d'en identifier l'empreinte tant dans PISA que dans PIRLS, leurs cadres conceptuels présentant un certain nombre d'analogies. Il faut dire que l'on retrouve aux commandes des grandes enquêtes actuelles des personnes connaissant particulièrement le sujet. Si le président des experts internationaux en lecture pour PISA travaille toujours à ETS, d'autres personnels ont migré vers des organismes comme ACER et Boston College.

Depuis 1990, NAEP est menée à la fois au niveau national et au niveau des états, alors qu'auparavant, elle ne portait que sur un échantillon national. Pour la lecture, elle est appliquée aux *grades 4, 8 et 12*.

Le cadre conceptuel de NAEP, développé par le N.A.G.B., prévoit d'évaluer **trois « contextes » de lecture : lire pour l'expérience littéraire** (les lecteurs explorent les éléments textuels tels que les événements, les personnages, les actions et le langage des travaux littéraires lors de la lecture de nouvelles, histoires courtes, poèmes, biographies, mythes, histoires du folklore, etc.), **lire pour s'informer** (les lecteurs s'informent et comprennent mieux le monde qui les entoure en lisant des journaux, magazines, manuels, discours, etc.), **lire pour réaliser une tâche** (cf. utilisation d'horaires, d'imprimés, de directives, etc.). Ce dernier contexte n'est pas proposé au *grade 4*.

L'autre dimension prévue dans le cadre conceptuel s'appelle « **aspects** » et se répartit en : *Elaborer une compréhension générale, Développer une interprétation, Etablir des relations entre le lecteur et le texte* (le lecteur doit relier des informations présentes dans le texte avec ses connaissances et son expérience), *Etudier le contenu et la structure du texte* (y compris les effets de l'ironie, de l'humour et de l'organisation du texte).

Les performances obtenues sont standardisées et ventilées en fonction de trois catégories préalablement établies par une commission ministérielle. Pour chacune et à chaque niveau scolaire, il est défini ce que les élèves doivent savoir et savoir-faire. Au niveau dit « **fondamental** » (ou de base), les connaissances et les performances sont insuffisamment maîtrisées pour travailler efficacement à ce niveau scolaire. Au niveau suivant, « **maîtrise** », les performances académiques évaluées sont acquises, de manière solide, et les élèves l'ont démontré au travers d'activités variées. Enfin, le dernier groupe concerne les élèves dont les performances atteintes sont au dessus de la moyenne et jugées comme « **remarquables** ». Cependant, une quatrième catégorie apparaît car un pourcentage notable d'élèves n'atteint pas le niveau fondamental.

Répartition par grade des pourcentages d'élèves en fonction du niveau de performances en lecture.

	En dessous du fondamental	Fondamental	Maîtrise	Remarquable
Grade 4	36 %	32 %	24 %	7 %
Grade 8	25 %	43 %	30 %	3 %
Grade 12	26 %	38 %	31 %	5 %

Les résultats

Niveau national – écoles publiques et privées. Il s'agit d'un échantillon représentatif. En 2002, et selon les standards établis, 31 % des élèves de Grade 4 font preuve de « maîtrise » et au -delà, ils sont 33 % en Grade 8 et 36 % en Grade 12. Des données sont disponibles, à des fins de comparaisons pour les années 1992, 1994, 1998 et 2002. Sans entrer dans le détail, l'évolution des performances entre ces différentes années n'est pas notable au Grade 4 ; une certaine baisse semble s'installer au Grade 12, alors qu'une évolution lente et positive s'installe au Grade 8. Des analyses plus circonstanciées utilisant les décilages sont également réalisées, ce qui permet d'étudier, par exemple, l'évolution des performances des élèves les plus faibles au cours des ans ou encore celle des moyens ou des meilleurs.

Niveau des états – écoles publiques seules.. La participation des écoles se fait sur la base du volontariat. Les comparaisons sont établies entre les années 1998 et 2002, car auparavant, les élèves handicapés n'étaient pas inclus dans le processus d'évaluation, aussi les échantillons ne sont pas comparables. En 2002, les élèves de Grade 4 et 8 ont participé à NAEP dans 45 états au lieu de 40 en 1998. En Grade 4, la moitié des états ont des résultats en progression et un seul en régression, les autres étant stables. Des comparaisons sont établies entre ces résultats et ceux de l'échantillon national. Elles font apparaître qu'environ un quart des états ont des résultats inférieurs à la moyenne nationale au Grade 4 et au Grade 8. Une cartographie montre qu'un bon nombre d'états du nord dépasse les performances nationales alors que les performances significativement plus faibles proviennent essentiellement d'états du sud. Des analyses plus raffinées permettent de connaître, par exemple, le pourcentage d'élèves qui a atteint ou dépassé le niveau de « maîtrise ». Au Grade 4, 19 états ont de meilleurs pourcentages qu'au niveau national et 15 états de plus mauvais.

Les résultats sont étudiés en fonction des variables classiques : sexe, niveaux d'études des parents, types d'écoles fréquentées, implantation de l'école. Ils sont aussi exprimés, par exemple, en fonction de la provenance ethnique et des aides accordées aux familles.

Comparaison des évaluations nationales britanniques et américaines avec les protocoles nationaux français CE2 – sixième

Comme en France, on retrouve l'idée d'évaluer aux grandes articulations des systèmes éducatifs. En Angleterre et en France, la définition de la lecture repose sur les Programmes et sur les Compétences attendues. Les tests anglais NCRT et la NAEP n'ont pas de fonction diagnostique, mais une fonction bilan afin de vérifier la qualité de l'enseignement et de piloter le système éducatif. Aussi, le codage se fait par attribution de points sans entrer dans une analyse des réponses.

En Angleterre, les maîtres pourraient analyser eux mêmes les réponses de leurs élèves au moyen des instructions de codage et de la compétence que chaque item est censé véhiculer, mais cela ne va pas de soi. Ce n'est qu'en temps très différé du moment de l'évaluation qu'ils peuvent consulter des documents d'analyse et de conseils pédagogiques.

Alors qu'en Angleterre et en France, tous les élèves d'un niveau doivent être soumis aux protocoles d'évaluation, aux Etats - Unis, seuls des échantillons représentatifs ou des états volontaires participent aux dispositifs.

On note de grandes différences dans les supports employés et pour les questions posées pour apprécier les compétences de lecture – compréhension : textes longs (environ 1800 mots en Angleterre, 1000 en France, à 11 ans) et souvent intégraux, questions essentiellement ouvertes en Angleterre (l'inverse en France), certaines compétences (cf. la liste en annexe) ne le sont jamais en France, sauf lors des évaluations internationales, par exemple « *Faire preuve d'une certaine capacité de réflexion sur le texte, son contenu, sa forme, sa structure, et faire appel à des connaissances extérieures au texte* ». En outre, nous remarquons que les deux pays étrangers s'efforcent de donner aux élèves du matériel ayant de fortes similarités avec l'écrit tel qu'il est présenté dans un ouvrage : illustrations conservées (ou créées si le matériel correspond à une commande spécifique). En France, la présentation est beaucoup moins soignée et il faut déplorer les problèmes de lisibilité créés en « tassant » des items sur une seule page (observation toujours valable pour des protocoles récents).

Les productions écrites en Angleterre sont jugées sous l'angle d'une liste de 8 compétences (voir KS2 et KS3), de manière beaucoup plus détaillée qu'en France et parfois sur des aspects que nous ne prenons pas en compte, incluant l'opinion personnelle (comme dans les évaluations internationales).

On remarquera le grand volume d'écrit produit par les élèves anglais dans les tâches estimées « évaluer la lecture / compréhension », alors que les élèves français sont majoritairement soumis à des questions fermées, et que les réponses construites demandées sont brèves, voire très brèves.

Les élèves anglais, et dans une moindre mesure les américains, se voient imposer une contrainte de lecture, puisque les tâches d'évaluation de la lecture sont, dans un premier temps, dissociées de celles de prise de connaissance des questions. Il en est de même pour les tâches de production écrite qui sont, elles aussi, séquencées. Pour la lecture, cela peut éviter comme on le voit encore trop souvent, que les élèves s'engagent dans les réponses aux questions alors qu'ils n'ont pas lu le texte qui s'y réfère.

Comme nous l'avons mentionné, les évaluations internationales de la lecture en cours (PIRLS et PISA) trouvent leurs racines dans d'autres outils d'évaluation. Pour PISA, il s'agit davantage de NAEP, IALS. Pour PIRLS, il y a de fortes similitudes avec le type d'outils utilisés dans NCRT, les concepteurs de tests de NFER ayant intégré le cadre conceptuel imposé par l'IEA et Boston College. Dans les deux cas, les dispositifs sont gouvernés par les méthodes IRT.

Chapitre 4

Fiabilité des protocoles

Nous avons effectué une présentation assez détaillée de divers dispositifs d'évaluations. Nous allons maintenant nous intéresser à la fiabilité des protocoles.

Que mesurent les protocoles d'évaluation de la DEP ?

Les protocoles sont des **épreuves standardisées** regroupant des exercices. Chaque exercice comprend en général plusieurs questions. Il est d'usage dans le champ de la mesure de considérer ces questions comme des items. Les questions se présentent sous différents formats : à choix multiples (QCM), ouvertes (à réponse construite, plus ou moins complexe). Des situations de production de textes existent dans plusieurs protocoles. Ces dernières situations se distinguent des productions scolaires habituelles par la mise en place d'une standardisation de la situation de production, ainsi que du codage utilisé par les correcteurs. Les protocoles sont administrés selon un calendrier identique pour tout l'échantillon.

Objectifs pris en compte par les évaluations françaises

Les évaluations de type bilan visent à déterminer dans quelle mesure les objectifs fixés par les programmes sont atteints. Les évaluations de type diagnostique sont conçues pour servir aux enseignants dans la connaissance de leurs élèves : forces et faiblesses. Les informations collectées sur les performances des élèves s'inscrivent dans une conception des connaissances et des compétences de type scolaire.

Les exercices et les items sont élaborés à partir d'une nomenclature d'objectifs identifiés à partir des programmes. Cette conception est, de plus, sous-tendue par les principes de construction des épreuves. Les protocoles sont construits par des groupes réunissant les représentants des principaux acteurs de l'enseignement dans chaque discipline (Inspection générale, enseignants, chercheurs, parfois des représentants de directions pédagogiques ministérielles) et des personnels de la DEP.

Les objectifs sont hiérarchisés selon une taxonomie en niveaux, depuis les objectifs les plus généraux jusqu'aux items. Un des principaux problèmes vient du caractère trop descriptif des objectifs les plus fins. Les objectifs du niveau hiérarchique le plus élevé semblent davantage accessibles à un public de non-spécialistes. La terminologie utilisée supporte mal, selon nous, le transfert vers des utilisations plus générales : pilotage du système éducatif dans son ensemble. Par ailleurs, **cette dépendance à l'égard des programmes peut poser problème pour les comparaisons temporelles**. La terminologie a subi au fil des évaluations des modifications. Les objectifs et leur dénomination ont parfois changé, comme nous l'avons déjà indiqué.

Il ne faut donc pas confondre les **évaluations-bilan** avec les **évaluations à finalité diagnostique** et formative qui sont proposées à l'ensemble des élèves aux niveaux CE2, sixième, seconde (arrêtées en 2002). Les évaluations dites diagnostiques ont pour but de donner aux enseignants et aux équipes éducatives des informations sur les acquis des élèves, afin de mieux gérer les apprentissages dans la classe. Une certaine confusion peut s'établir entre ces types d'évaluation dans la mesure où la DEP publie des résultats nationaux en utilisant les résultats obtenus à ces évaluations diagnostiques pour un échantillon tiré *a posteriori*. Or, l'examen de l'évolution du dispositif et de son mode de construction nous a montré que s'étaient sans doute produites des contaminations entre les représentations des différents dispositifs (Bilan vs. Diagnostique). Ces contaminations ont eu des effets préjudiciables sur la fiabilité du dispositif des évaluations nationales et de la lecture de leurs résultats. Peut-on réellement considérer qu'il y a un diagnostique dans les évaluations sans se référer à un modèle du fonctionnement du lecteur et donc à des théories de la lecture – compréhension ? Des évaluations diagnostiques peuvent – elles être analysées au niveau d'une population ? On ne peut pourtant pas tout attendre des évaluations diagnostiques. Elles ne permettent pas de repérer toutes les difficultés en raison, notamment, du choix limité des compétences évaluées (lié aux objectifs). Les modalités de passation (collectives) peuvent restreindre les observables, etc.. D'une année à l'autre, les résultats ne

sont pas comparables, car ils ne proviennent pas d'épreuves strictement comparables.

Etude de la fiabilité des protocoles

Importance des analyses psychométriques.

L'étude de nos savoirs sur les connaissances des élèves conduit à se poser deux questions : ce que nous savons et comment nous le savons. La seconde question est préalable à la première car la manière dont nous recueillons les informations conditionne leur nature et leur fiabilité. Nous avons souvent trouvé dans les publications de la DEP des annexes, des encarts méthodologiques assez pertinents et didactiques, qui témoignent du souci apporté à la fiabilité statistique du dispositif et à la communication de réserves éventuelles des statisticiens. Cependant, les références aux concepts et aux méthodes de la mesure en éducation, de la psychométrie sont rares.

Schématiquement, **l'objet de la psychométrie** est la construction d'une mesure fiable (voir, Dickes et al., 1994.. En psychométrie, on considère que le score observé est composé d'au moins deux parts : le score vrai et l'erreur de mesure. Le concept d'erreur en psychométrie intervient à tous les stades de la démarche : du choix des questions jusqu'à l'interprétation des résultats. Les études de validation ont pour but d'identifier et de réduire autant que faire se peut les différentes sources de biais pour garantir la fiabilité des inférences finales sur l'interprétation des résultats.

La distinction entre **performance et compétence** est une autre manière de poser le problème de la mesure en éducation (sur cette distinction voir en particulier Flieller, 2001). Les compétences ne sont pas directement observées : on va inférer sur les compétences à partir des performances des élèves. Les performances sont dépendantes de nombreux autres facteurs que de la seule compétence évaluée, facteurs qu'il faudra essayer de prendre en compte.

Analyse des protocoles de la DEP

L'échantillonnage des items

La première étape est l'échantillonnage de l'univers des items. Ce point est en général moins bien connu que l'échantillonnage des sujets. Il est spécifique des problèmes de la mesure en sciences humaines en général et en éducation en particulier. En effet, le protocole doit être représentatif de l'univers à évaluer. De même que la représentativité de l'échantillon des sujets permet de généraliser sur la population parente (non observée), de même la représentativité de l'échantillon des items quant au domaine évalué permet de généraliser sur l'ensemble du programme (non observé).

La méthode mise en œuvre pour la construction des protocoles de la DEP remplit en principe cette fonction. Le programme est découpé en objectifs, mais certains objectifs ne sont parfois pas représentés du tout. Contrairement aux évaluations internationales, le nombre de questions, leur répartition par « compétence » et par format sont largement dépendants des possibilités offertes par le matériel, ce qui peut nuire à la qualité de la mesure.

La standardisation

Les conditions de standardisation des passations sont bien prises en considération. On peut néanmoins s'interroger sur un certain nombre d'éléments implicites susceptibles de compromettre cette standardisation. Signalons par exemple la présence éventuelle de différences de conditions de passation entre les classes, et entre les établissements. Ainsi, dans les cas où les élèves passent deux cahiers d'épreuves, le délai entre les deux passations peut varier : soit que les élèves passent les deux cahiers dans une matinée, soit qu'ils en passaient un le matin, l'autre l'après-midi (Derouet, 2003). On peut penser que dans la première condition, les effets de fatigue et de lassitude étaient plus importants que dans la seconde.

Nous n'avons pas trouvé d'études sur les conditions de passation sur le terrain, ni sur le respect de la standardisation du codage par les correcteurs (fidélité inter-juges). Ces vérifications sont indispensables. Elles paraissent d'autant plus nécessaires que les protocoles comportent des questions ouvertes ainsi que de la production de textes. Dans ce dernier cas, le dispositif devrait comporter une

formation des correcteurs pour garantir une harmonisation maximale (Vrignaud, sous presse).

Analyse interne des épreuves

En général, les analyses internes se basent sur trois types d'indicateurs. Deux se situent au niveau de l'item : sa difficulté (la fréquence de réussite à l'item), sa discrimination (la liaison entre l'item et le score total) ; et un autre au niveau du score ou des sous-scores : un indicateur permettant d'apprécier la consistance de l'ensemble des items, c'est-à-dire le fait que l'ensemble des items appartiennent à un même domaine, homogène, et que le score reflète bien la compétence dans le domaine évalué.

La difficulté des items

Dans l'approche psychométrique classique, on désigne sous le terme de difficulté des items la fréquence de réussite (pourcentage de bonnes réponses). Les difficultés des items sont utilisées pour piloter le niveau de difficulté global de l'épreuve. La difficulté d'un item n'est pas une mesure absolue, elle est dépendante de l'échantillon. La difficulté d'un protocole est liée aux intentions des constructeurs. Nous n'avons pas trouvé d'indications sur les intentions des constructeurs quant aux niveaux de difficulté recherchés.

En dehors de l'image qu'il donne de la maîtrise des objectifs, le niveau global de difficulté d'un protocole a plusieurs implications méthodologiques importantes. En effet, le niveau de difficulté moyen dépend de la distribution des scores. En psychométrie, on cherche souvent à éliminer les items dont la fréquence de réussite est trop élevée (e.g. > 90%, effet plafond) ou trop faible (e.g. < 10%, effet plancher). Un protocole facile (réussi à plus de 80% en moyenne) induit une distribution asymétrique, c'est à dire que les résultats des sujets les plus performants sont moins dispersés que les résultats des sujets faibles et moyens. Un tel protocole est, de ce fait, peu sensible pour distinguer les sujets à performance élevée des sujets à performance très élevée. Un tel protocole, par contre, distinguera bien les sujets faibles des sujets moyens. La distribution des scores a d'autres conséquences, car elle joue sur les liaisons entre le score et d'autres indicateurs qui seront étudiés : par exemple les relations entre les disciplines. De plus, dans la perspective d'étude des évolutions temporelles, la marge d'évolution de protocoles faciles, est plus restreinte, car le tassement observé à la première passation aura tendance à s'accroître à la seconde.

En conclusion, nous dirons que si le choix du niveau de difficulté appartient aux constructeurs du protocole, il est par contre indispensable d'explicitier les raisons de ce choix.

Discrimination et consistance interne

Les indicateurs de discrimination³, et les indicateurs de consistance sont publiés depuis peu. Il faut remarquer que certains items de faible qualité sont cependant souvent exploités et conservés dans les analyses.

Problèmes posés par la généralisation sur un objectif à partir d'un nombre restreint d'items

La généralisation sur un objectif à partir d'un item pose problème du fait de l'influence, sur la difficulté intrinsèque d'un item, de différents facteurs indépendants de la compétence évaluées. La performance à un item isolé ne permet pas de distinguer entre performance et compétence, donc de

³ Le concept de discrimination peut apparaître moins habituel que celui de difficulté. De manière générale, il répond à la question « les sujets qui réussissent l'item ont-ils une compétence plus élevée que les élèves qui échouent ? ». Cette question est opérationnalisée en comparant les performances au test des sujets qui réussissent et des sujets qui échouent. On s'attend à ce que le score des sujets qui réussissent soit en moyenne plus élevé que celui des sujets qui échouent. Si c'est le cas, alors on peut dire que l'item discrimine bien les sujets forts (du point de vue de la compétence) des sujets faibles. On voit l'intérêt de cette notion : un item présentant une bonne discrimination apporte une information importante sur les compétences des sujets ; un item de discrimination faible apporte peu d'information, il n'est pas très utile du point de vue du protocole ; un item de discrimination nulle, n'apporte pas d'information voire introduit du bruit dans la mesure ; un item de discrimination inverse (les sujets qui le réussissent posséderaient moins la compétence que les sujets qui échouent !) fonctionne à l'envers.

généraliser sur la maîtrise de l'objectif.

Pour pallier cet inconvénient, il est recommandé d'utiliser un nombre d'items « variés » suffisamment important pour neutraliser les biais liés aux contenus spécifiques d'un item. Un item isolé ne renseigne, à la limite, que sur lui-même ; les résultats à un ensemble d'items permettent d'inférer sur ce qui leur est commun : en principe la maîtrise d'un objectif. Nous retrouvons ici les idées de base de la psychométrie : un seul item ne permet pas d'apprécier l'erreur de mesure, plusieurs items évaluant la même compétence permettent de construire une mesure de la compétence du sujet (variable non observée). Dans les protocoles récents, les mises en garde faites, par exemple, dans l'Avis n°2 du HCcé, ont été prises en compte et les items isolés ont disparu. Conduire des comparaisons temporelles entre items ou entre objectifs évalués par un faible nombre d'items, ne paraît pas offrir les garanties nécessaires pour construire des inférences valides.

La publication des résultats

Nous notons des différences dans la publication des résultats.

Il nous paraît important de réfléchir à d'autres types de traitement et de présentation des résultats. Les documents publiés par différents organismes étrangers ou ceux des évaluations internationales pourraient fournir dans ce domaine une base de réflexion intéressante. Il s'agit de considérer les objectifs et les items qui servent à les évaluer dans une perspective globale. Les items sont hiérarchisés selon leur difficulté. On peut donc les ordonner sur une échelle. On s'appuie alors sur cette propriété des protocoles pour présenter les items selon leurs niveaux de difficulté.

De nombreux dispositifs (NAEP, PIRLS, PISA,...) utilisent les modèles de réponse à l'item (MRI) pour le traitement et la présentation des résultats. Les MRI sont des modèles puissants pour assurer la mesure en éducation (pour une présentation, voir Dickes et al., 1994 ; Vrignaud, 1996). Leur avantage est de placer conjointement sur une même échelle compétence des sujets et difficulté des items. La mesure de la compétence d'un sujet permet d'estimer dans quelle mesure il peut réussir un item d'une difficulté donnée. La mesure de la difficulté d'un item permet d'estimer le niveau de compétence requis pour pouvoir le réussir.

Prise en compte des caractéristiques des élèves

Les analyses des résultats cherchent à estimer l'effet de différentes variables sur la performance des élèves. Elles sont intéressantes car elles mettent bien en évidence l'effet de ces variables et attirent l'attention des éducateurs sur des faits auxquels il conviendrait, dans un souci d'équité, de remédier. Il serait utile d'aller au-delà des constats en essayant de mieux comprendre en quoi ces variables agissent sur les performances scolaires afin d'y remédier plus directement. Il n'est bien sûr pas question de transformer la DEP en laboratoire de recherches, mais plutôt d'attirer l'attention sur la présentation un peu « mécanique » et statique, des mises en relation des caractéristiques des élèves avec leurs performances.

Identifier les biais

L'étude des caractéristiques revient, de fait, à comparer les performances de groupes de sujets sur une caractéristique commune (par exemple le genre, ZEP – non ZEP). Les comparaisons ne sont fiables que si les instruments utilisés pour conduire cette comparaison ne sont pas biaisés au détriment de l'une des populations comparées.

Nous insistons particulièrement sur un type de biais que nous avons observé à plusieurs reprises dans les protocoles : des biais d'items. Les biais d'items sont des nuisances pour la qualité de la mesure psychométrique. En effet, on peut montrer qu'un item biaisé mesure autre chose que la variable qu'il est censé mesurer et que cette variable parasite favorise - ou défavorise - un des groupes étudiés. Si nous nous situons dans le cadre de l'utilisation de tests ou de questionnaires pour des comparaisons entre groupes, l'équivalence entre deux épreuves doit être démontrée, en premier lieu, par l'absence de biais au niveau des items. Ces études devraient être systématiquement effectuées et leurs résultats publiés.

Les comparaisons temporelles et internationales

Comparaisons temporelles et comparaisons internationales apportent des éléments d'informations essentiels pour évaluer les acquis des élèves. La comparaison du niveau de nos élèves avec ceux d'autres pays est l'un des moyens disponibles pour se faire une idée de l'efficacité de notre système éducatif (Bonora & Bacher, 1986 ; OCDE, 2003). En effet, nous avons montré que du fait de la dépendance de la mesure par rapport au choix des objectifs et à la difficulté des items, il est virtuellement impossible d'avoir une évaluation absolue des acquis de nos élèves. Les comparaisons temporelles et internationales peuvent, par contre, nous fournir des informations sur le niveau relatif de nos élèves, et permettre d'étudier les effets de certains facteurs d'efficacité du système, surtout dans des systèmes assez fortement centralisés comme le système français, où l'effet de nombreux aspects ne peut être mis en évidence du fait de l'absence de variabilité.

Les comparaisons temporelles

Le nombre de publications sur les comparaisons temporelles est limité (enquête 1987 – 1997), mais les reprises d'items ou d'exercices dans les protocoles nationaux sont fréquents. Malheureusement, du fait de modifications dans les exercices appelés communs, (conditions de passation, aménagement des codages), ces données ne permettent qu'imparfaitement d'effectuer des comparaisons fiables du fait de l'influence de différents biais. Il faut noter l'habitude de comparer des items isolés dans des protocoles différents (items repris), ce qui conduit à des comparaisons peu robustes (ex. protocoles CE2, sixième). Les références aux indicateurs psychométriques restent discrètes à l'exception du numéro 102 des Dossiers (comparaison 1987 – 1997) étude qui aurait marqué à la DEP la mise en œuvre des méthodes psychométriques. Il faut aussi souligner dans cette étude l'élimination des items biaisés pour l'élaboration des résultats, ce qui n'est pas toujours le cas, comme nous l'avons déjà évoqué.

- **biais d'administration** : La comparaison entre les consignes des différents protocoles montre des différences dans les conditions de passation qui peuvent produire des biais difficiles à apprécier. La durée des épreuves varie, Ces détails peuvent modifier les représentations du dispositif par les élèves et avoir un effet sur leur motivation.
- **biais d'items** : Les changements de contextes (programmes, environnement culturel, informations) peuvent engendrer des biais d'items.

Nous pensons que la difficulté de conduire des comparaisons temporelles fiables est réduite du fait de trois facteurs : (1) des évolutions du contexte éducatif ; (2) des changements de personnels parmi les responsables de la DEP et les constructeurs de tests, (3) d'une connaissance limitée de la psychométrie. Le premier point (changements de programmes, modification de la composition des populations scolarisées) échappe à la DEP, tout au plus peut-on essayer de mettre en place des mécanismes de rattrapage, mais dont l'efficacité est souvent limitée. Le second point peut être considéré comme un aléa. Ceci étant, des procédures par écrit devraient assurer la conservation de la mémoire et de la transmission des savoir-faire. Nous développerons davantage le troisième point dans la partie « propositions » car les méthodologies des comparaisons temporelles ont été l'objet de nombreux et importants développements dans le domaine de l'évaluation des systèmes éducatifs. Les méthodes, souvent très sophistiquées, ont été intégrées récemment à la DEP à la suite de travaux d'expertise (Bonora et Vriгдаud, 1997 ; Flieller et al., 1994).

Les comparaisons internationales

Dans les évaluations internationales, des informations de nature différente ou plus larges que les connaissances scolaires définies à partir des programmes sont prises en compte. Les résultats des comparaisons internationales sont une source d'information incontournable. L'intérêt de situer les acquis de nos élèves et le rendement de notre système éducatif comparé à celui d'autres pays, ces

enquêtes permettent une réflexion enrichissante sur nos pratiques d'évaluation. Cependant, nous ne savons pas bien définir ce que PISA évalue vraiment. La nécessité de rechercher un cadre commun à un ensemble de pays aux conceptions pédagogiques parfois assez hétérogènes conduit à réfléchir sur une conception des objectifs différente de celle développée à partir de la seule analyse de nos programmes.

La DEP a développé une réflexion sur les méthodologies employées dans les enquêtes internationales. Cette réflexion menée en liaison avec différents laboratoires universitaires (Dickes & Flieller, 1999 ; Guérin-Pace & Blum, 1999 ; Vrignaud & Bonora, 1998 ; Vrignaud, 2001) et d'autres institutions lui a permis d'acquérir une expertise dans ce domaine. Dans ce cadre, la DEP a piloté une enquête européenne sur les compétences en lecture des élèves de 16 ans dans quatre pays européens (projet SOCRATES, cf. Bonnet et al., 2001, puis C - bar). Elle réinvestit ces apports dans le cadre de nouvelles études : « *Evaluation bilan fin d'école* » et « *Evaluation bilan fin de collège* » pour lesquelles elle a adopté la méthodologie des cahiers tournants (cf. PISA et PIRLS) et mis en œuvre une méthodologie qui devrait permettre des comparaisons dans le temps (système d'ancrage des items, comme dans étude SOCRATES et C - BAR). Dans ces deux évaluations (en cours), le souci psychométrique est clairement affiché. Il faut noter que les préconisations du HCéé ont eu des effets au niveau du recrutement de personnels formés à la psychométrie (Avis 2) ; une évolution certaine a eu lieu au niveau des préoccupations psychométriques, et un effort a été fait pour installer une culture de la psychométrie. C'est chose faite chez les statisticiens, les autres personnels y deviennent sensibles, mais cette préoccupation a du mal à gagner les enseignants qui composent les groupes de pilotage et qui ont une vision beaucoup trop centrée sur la docimologie. Une réelle opposition existe dans certains de ces groupes et les enseignants refusent de prendre en compte les apports de ces méthodes au moment de l'élaboration des résultats ; ils continuent à garder des items sans pouvoir discriminant sans pouvoir vraiment justifier leur refus.

Chapitre 5

Conclusion générale

Des évaluations portant entre autres sur les compétences en lecture et écriture existent depuis 1979 dans le cadre de l'évaluation permanente du système éducatif. A partir de 1989, elles sont plus particulièrement destinées à fournir aux enseignants un état diagnostique des compétences de leurs élèves. La finalité des évaluations a changé mais aussi la logique des protocoles, la nature des exercices, l'équilibre des domaines disciplinaires, les formes de codage. Simultanément, les programmes ont évolué, ainsi que l'organisation de la scolarité et donc l'âge des élèves évalués à un niveau donné. Les résultats sont donc difficilement comparables. En outre, ils ne font plus de différences entre les compétences requises à telle étape du cursus scolaire et celles qui les dépassent. Enfin, les études comparatives permettant de suivre l'évolution des compétences dans le temps sont rares.

Cependant, les résultats, l'analyse des protocoles et la comparaison avec d'autres observations permettent de dégager des tendances générales sur les compétences des élèves. Dans l'ensemble, le niveau de compétence des élèves est stable dans le temps.

- En lecture, les élèves ont de bons résultats surtout quand il s'agit de prélever des informations explicites ; les difficultés apparaissent dans le traitement de l'information avec la complexité des tâches et des supports, et également pour passer à l'interprétation.
- En production écrite, les élèves sont capables de rédiger un texte cohérent qui réponde aux consignes, les cadres formels sont à peu près acquis. Ce sont les contraintes linguistiques qui posent le plus de difficultés aux élèves. Le récit est bien maîtrisé ; le discours argumentatif à la fin du collège ne l'est pas encore.
- Le fonctionnement et l'utilisation de la langue sont les domaines les moins maîtrisés par les élèves. On le comprend pour le premier degré puisqu'en fin de cycle 2, l'apprentissage systématique de ces compétences en est à son début. Cependant, les résultats en orthographe se sont dégradés au cours de ces 25 dernières années. Au collège, quand un grand nombre d'éléments de la langue sont évalués, on constate jusqu'en seconde les difficultés importantes des élèves particulièrement dans le vocabulaire et l'orthographe grammaticale.

En l'absence d'un indicateur qui précise les compétences requises et le pourcentage d'élèves qui ne maîtrisent pas ces compétences, on est obligé d'utiliser les écarts qui séparent les scores des élèves faibles des scores moyens ou des scores des élèves les plus forts. Les chiffres avancés sont donc à prendre avec prudence.

Au CE2, plus de 15% des élèves ne maîtrisent pas les compétences requises à ce niveau. Ces élèves sont particulièrement en difficulté sur le déchiffrage et la compréhension.

Au collège, environ 15% des élèves ont des difficultés sur toutes les tâches les plus complexes de lecture ; et la combinatoire n'est pas automatisée chez les 10% des élèves des plus faibles. Ce qui les distingue fortement des autres élèves, c'est la connaissance du fonctionnement de la langue et la mobilisation des ressources de la langue pour écrire.

Cependant les analyses fines des évaluations montrent que ces élèves, dont les scores moyens sont très faibles, peuvent avoir de bons voire de très bons résultats à certains exercices. On ne peut donc pas savoir si ce sont des illettrés au sens rigoureux de ce terme. En effet, ils ne sont pas démunis de compétences et forment une population hétérogène à laquelle les items des protocoles ne sont peut-être pas adaptés.

Les évaluations nationales, même si elles sont sous-utilisées, ont fait entrer dans la culture de l'institution la notion et l'importance de l'évaluation, avec une certaine représentation de ses méthodes et de son vocabulaire. Peut-être cette familiarité s'est-elle progressivement muée en un rituel, qui demande donc une réflexion et une orientation nouvelles, attentive en outre aux apports des évaluations internationales. Mais cet effet des évaluations nationales est certain. D'autre part, les évaluations ont confirmé des constats importants en pédagogie : le poids des facteurs externes sur la

réussite des élèves ; le fait que la grande difficulté n'est pas due à un simple retard qu'un redoublement permettrait de combler ; le fait qu'à une erreur ne correspondent ni une seule origine, ni une seule solution ; l'acquisition des compétences par une refondation permanente de leurs composantes. En effet, les résultats montrent que les élèves n'acquièrent pas une compétence en additionnant des savoirs et savoir faire isolés, acquis une fois pour toutes, et dont la somme constituerait la compétence ; au contraire, les savoirs et les savoir faire dont les élèves disposent pour une tâche donnée sont remis en cause, à reconstruire et à consolider avec la complexité croissante des situations de travail qui leur sont proposées. Rien n'est jamais définitivement acquis, vérité d'un poids décisif en pédagogie.

Dans le premier degré en particulier, les évaluations permettent aux maîtres de situer leur classe à partir d'un référent national, de réguler et de mieux programmer l'action pédagogique dans la classe et dans le meilleur des cas dans les cycles. Elles donnent aussi un contenu à la définition du savoir-lire et du savoir-écrire. Elles ont objectivé la nécessité de construire et de mettre en œuvre des dispositifs précis pour les élèves en grande difficulté (PPAP programmes personnalisés d'aide et de progrès). Elles ont contribué à diffuser de nouvelles pratiques didactiques et pédagogiques favorables à des apprentissages différenciés. Enfin les études engagées à partir des analyses des résultats ont permis de construire de nouvelles évaluations, mises à la disposition des maîtres, pour l'apprentissage de la lecture et de l'écriture en grande section et au cours préparatoire. Ces effets très positifs des évaluations de masse ne sont peut-être pas ceux qu'attendaient les responsables institutionnels, qui espéraient un effet plus rapide et direct sur la réussite scolaire. Pour être moins visibles, ils n'en sont pas moins importants pour améliorer, à plus long terme, l'efficacité du système.

La situation actuelle est cependant paradoxale. Les protocoles, semblables ou proches d'une année sur l'autre surtout depuis 1997, semblent pouvoir permettre aux décideurs de suivre l'évolution du niveau des élèves. Mais comme, pour satisfaire aux besoins des enseignants, ils ne sont pas absolument identiques, ils permettent malaisément cette comparaison et ce suivi, sans pour autant être massivement utilisés par les enseignants. Les décideurs régionaux ou locaux comparent surtout les moyennes nationales avec les scores de leurs propres élèves, sans s'intéresser réellement au contenu des items eux-mêmes qui sont, dans leur esprit, faits pour les enseignants. Mais les enseignants, pour qui sont prioritairement élaborés les items, s'y intéressent également peu parce qu'ils se sont rarement appropriés les protocoles. Le diagnostic sur l'état des compétences des élèves et l'évolution de ces compétences, fondé sur un nombre suffisant de comparaisons, et qui puisse également donner la raison des constats, n'est pas véritablement fait. Il serait pourtant nécessaire à la fois aux responsables institutionnels pour piloter le système éducatif, aux enseignants pour situer leurs élèves, et aux citoyens.

A cela, on l'a vu, s'ajoute le tournant qu'on observe à partir de 1989-90. Avant cette date, les protocoles cherchaient à mesurer, parfois sur une même cohorte d'élèves, l'évolution des résultats d'une classe ou d'un cycle à l'autre, avec des items en partie communs qui permettaient ce suivi, et avec une consultation fréquente des enseignants. Ces protocoles étaient fondés sur un éventail assez large de compétences. La finalité dominante semble avoir été alors de suivre l'évolution des compétences des élèves au fil de leur cursus. Après cette date, l'éventail des compétences évaluées s'est réduit, les protocoles ont été construits selon une logique différente, les avis préalables ou postérieurs des enseignants se sont raréfiés, et tous les résultats n'ont pas toujours été analysés. La finalité dominante semble être désormais de fournir aux professeurs et aux chefs d'établissements un outil professionnel ponctuel pour organiser l'enseignement. Du coup, toute analyse rigoureuse de l'évolution des acquis des élèves depuis 1979 est impossible. Et on a l'impression que rien n'est attendu de cette profusion d'évaluations à part une exploitation locale qui n'est pas assurée, que les évaluations se répètent sans planification ni projet visibles.

Les études conjointes à l'évaluation des compétences ont permis de mettre en évidence le lien de l'environnement (familial, socio-culturel, etc.) et des résultats des élèves ; il est important de montrer ainsi que des «facteurs externes» jouent dans la réussite des élèves et lesquels. Mais le rôle, parmi les «facteurs internes», des pratiques pédagogiques, est à peu près absent des protocoles et des analyses. Certes, les premiers protocoles ont été élaborés avec une consultation des professeurs, qui ont donné leur avis sur leur choix et leur pronostic sur les réponses éventuelles des élèves ; mais ces avis ne sont pas des pratiques. Ce serait pourtant un apport précieux des évaluations, si elles pouvaient

permettre de mesurer la part, dans les résultats, des représentations et des pratiques des enseignants. Si par exemple les élèves semblent ne pas avoir une meilleure maîtrise de la conjugaison de la fin de la troisième à la fin de la seconde (ce qu'on ne peut dire pour l'instant qu'en généralisant des résultats limités épars dans le temps et non systématiquement suivis ni largement évalués), est-ce parce que la maturité que demande cette maîtrise n'est pas atteinte par tous les élèves en fin de seconde, ou parce que ce domaine de la langue n'est plus enseigné après le collège? Quand on observe d'une évaluation à l'autre les mêmes carences, faut-il les attribuer à l'identité des protocoles, à un stade identique dans la maturité des élèves, à un type déterminé de pratique pédagogique, ou à la négligence des compétences correspondantes dans l'enseignement? Cette exploitation des évaluations n'est pas faite. Il en est de même des démarches, des processus utilisés par les élèves pour répondre, et qui sont aussi importants que les savoirs eux-mêmes dans l'acquisition d'une compétence. Certains items cherchent à les atteindre, quand ils demandent de justifier une réponse ou de vérifier une hypothèse. L'analyse des résultats suggère parfois une mauvaise lecture des consignes, une exploration trop rapide des textes, un besoin insuffisant de vérifier sa réponse. Mais l'évaluation des démarches, même si elle est difficile, paraît n'être qu'un objet secondaire des protocoles, mis à part ceux de seconde professionnelle.

Les évaluations internationales comme PIRLS et PISA nous interrogent sur notre système éducatif et sur nos habitudes culturelles : la France se classe dans les deux cas à un rang moyen par rapport à la trentaine de pays participants. Elles révèlent toutes les deux des types de questionnement mettant nos élèves en difficulté, questionnement existant par ailleurs dans les protocoles anglais. Nos évaluations proposent beaucoup de réponses fermées contrairement à ces deux évaluations internationales pour lesquelles nos élèves ont un comportement qui intrigue. En effet, ils ont tendance à ne pas répondre aux questions ouvertes beaucoup plus souvent que les élèves des autres pays. Quelle est l'origine de ce comportement?

Nous avons largement souligné les limites de ces évaluations et les biais qui subsistent même si la psychométrie fonde ces enquêtes. Pour limiter ces biais culturels et même les éviter, le développement et la mise en œuvre de méthodologies novatrices comme l'étude Socrates et le projet C - BAR doivent être favorisés.

Chapitre 6

Propositions

Propositions méthodologiques

Mieux identifier et distinguer les objectifs des évaluations

La première question qui se pose en psychométrie est de définir les univers à mesurer. La distinction des objectifs des évaluations est fondamentale car, d'une part les principes de construction, et par suite la qualité de chacune des épreuves, sont dépendants de leur finalité ; car d'autre part l'utilisation d'épreuves à tout faire peut induire, lors de la communication des résultats aux utilisateurs, aux décideurs, au grand public, des représentations erronées sur ce que mesure chacun de ces types d'évaluation. Il serait utile de réfléchir à l'intérêt de situer les objectifs pris en compte par nos différentes évaluations nationales sur un continuum de la maternelle à la troisième, en passant par le primaire.

Quels apprentissages évaluer ?

La procédure classique de construction des items d'une épreuve consiste à élaborer un référentiel, extrapolé des textes officiels, qui croise le plus souvent les contenus disciplinaires et les démarches intellectuelles applicables sur ces contenus. Or, cette procédure pourra parfois apparaître insuffisante pour rendre compte exhaustivement des acquis de l'élève. Il semble que la conception d'un instrument d'évaluation puisse tirer profit d'une réflexion préalable sur les acquis les plus significatifs de la formation considérée.

Il serait également intéressant de dégager des pistes d'élaboration des items à partir de la prise en compte des "critères d'intelligibilité" de la discipline, qui évoluent dans le temps, et modifient les caractéristiques générales des compétences de la population.

Comment évaluer les productions écrites ?

Des productions d'écrits sont demandées dans certains protocoles de la direction de l'évaluation et de la prospective. Ce point est relativement original. Ces exercices sont utiles car ils renforcent la validité de contenu des protocoles. Le travail de cotation de telles productions est bien sûr plus lourd que celui des QCM. Il doit faire l'objet d'une élaboration de grilles de dépouillement à partir d'indicateurs factuels à défaut de quoi il n'échapperait pas aux biais habituels liés à la subjectivité des correcteurs. Signalons qu'actuellement des recherches sur la correction automatisée de textes produits, simulant la démarche d'enseignants, sont conduites en particulier dans des organismes tels qu'ETS aux Etats-Unis. Ces exercices de production écrite pourraient être enrichis en associant des spécialistes de sciences du langage à l'élaboration de nouveaux protocoles, spécialistes insuffisamment sollicités par la DEP. Les évaluations anglaises, particulièrement celles concernant la production de textes, mériteraient d'être étudiées de près pour leur richesse et leur méthodologie.

Évaluer les productions orales ?

L'oral n'est pas l'objet privilégié des évaluations. Les nouvelles évaluations bilans (fin de CM2 et de 3ème) comportent des tâches de compréhension de l'oral. Mais les productions orales sont absentes de la plupart des protocoles. Ceci apparaît d'autant plus dommageable que les nouveaux programmes de l'école mettent l'accent sur l'oral. De plus, les compétences de communication orale sont fondamentales dans la vie professionnelle, que ce soit en langue nationale ou en langues vivantes étrangères. Une évaluation standardisée de l'oral requiert des dispositifs encore plus complexes que l'évaluation des productions écrites. En effet, ces épreuves sont encore plus sensibles aux contextes psychologiques et psycho-sociaux. Il est évident que la connaissance préalable - en bien ou en mal - d'un élève ne peut qu'avoir un effet parasite sur l'appréciation de sa prestation par l'enseignant. Les situations doivent non seulement être standardisées mais aussi ne pas induire de prises de position trop marquées chez l'élève. Ce type d'instruments mériterait également des études de fiabilité avant qu'ils

ne soient introduits et systématisés dans les protocoles.

Les protocoles d'évaluation : renforcer la fiabilité de la mesure

L'IEA a rédigé un ouvrage, à la fois méthodologique et bibliographique, pour les évaluations internationales, qui n'a pas son équivalent en français (Gregory et Martin, 2001). Cet ouvrage pourrait être utilisé avec profit.

Augmenter le nombre d'items par objectif

Cette recommandation est récurrente : seule l'utilisation de plusieurs items ou plutôt exercices peut permettre de généraliser sur un objectif donné. Cette exigence peut apparaître contraignante par rapport à l'étendue des objectifs à évaluer et au temps de passation disponible. Les cahiers tournants peuvent être une solution ; cela consiste à mettre en place un système de blocs d'items ou plutôt d'exercices. Chaque sujet ne passe qu'une partie des blocs. Un dispositif d'ancrage permet de placer tous les blocs sur une même échelle, soit à partir d'un bloc commun, soit en utilisant une combinatoire permettant d'établir des relations entre tous les blocs deux à deux (système dit des plans expérimentaux incomplets). Ces dispositifs permettent de recueillir des informations sur un grand nombre d'items sans augmenter trop la charge de travail des sujets. L'introduction de cette procédure dans les nouvelles évaluations bilan a été remarquée. Il s'agit de l'intégration d'une méthodologie en œuvre dans PIRLS et PISA, par exemple.

Réduire les sources de biais

- Minimiser les liaisons artificielles entre items
Les protocoles de la DEP comportent des exercices comprenant plusieurs items. A l'intérieur d'un même exercice, les réponses des élèves ne sont, le plus souvent, pas indépendantes les unes des autres. Ce fonctionnement induit une cohérence artificielle dont il faudrait tenir compte dans l'appréciation des indicateurs psychométriques.
Il serait judicieux d'éviter ce type de situations débouchant sur des items dont la forme crée une dépendance locale forte : par exemple, les réponses en " cascade " où le sujet doit, dans un premier temps, fournir une réponse et, dans un second temps, la justifier, ainsi que les questions sous forme de classifications qui demandent au sujet de mettre en relations plusieurs termes avec des cibles.
- Clarifier le codage
La distinction entre bonne et quasi-bonne réponse (codes 1 et 2) est sans doute intéressante pédagogiquement mais crée un flou certain du fait d'un écart parfois important dans les fréquences de réussite selon que l'on inclut les codes 2 ou non. Or, rappelons le, les items sont censés informer sur l'acquisition de l'objectif, il faut donc décider clairement si les quasi-bonnes réponses indiquent ou non cette maîtrise.

Vérifier la qualité du protocole par une réelle pré-expérimentation

Une pré-expérimentation est la seule procédure fiable pour vérifier les qualités de l'épreuve. Ce n'est pas une simple formalité, les résultats des analyses conduiront à supprimer, éventuellement à modifier, les exercices, les items et leur codage. L'étude des indicateurs de difficulté permettra de fixer le niveau de difficulté global souhaité pour l'épreuve. Les jugements d'experts ne sont en aucun cas fiables, comme l'ont montré toutes les comparaisons antérieures effectuées entre jugements et difficulté observée. De même, discrimination et consistance sont les seuls indicateurs objectifs du rattachement ou non des exercices et des items aux objectifs. Enfin, on pourra effectuer certaines vérifications telles que l'identification des items biaisés selon les caractéristiques des élèves.

Dans les cas où l'on souhaite employer des modèles particuliers pour le traitement des résultats (par exemple les MRI), il sera nécessaire de s'intéresser lors de l'analyse des résultats de la pré-expérimentation, à l'adéquation du modèle aux données, et d'éliminer, si nécessaire, les items témoignant d'une insuffisante adéquation.

Cette pré-expérimentation ne dispensera pas, bien entendu, de conduire les mêmes analyses sur le protocole définitif.

Mieux standardiser les conditions de passation

Nous avons constaté, en particulier, dans le déroulement des passations (gestion du temps, lecture des

consignes), une variabilité entre les classes, et entre les établissements ; les entretiens que nous avons menés sont clairs à cet égard. Cette variabilité devrait être éliminée autant que faire se peut.

En ce qui concerne les évaluations de type bilan, nous sommes favorables à une solution qui éviterait l'administration des épreuves par les professeurs habituels. Une telle solution permettrait de réduire d'éventuels effets induits par la relation développée au cours de l'année scolaire, et de mieux respecter l'anonymat signalé dans les consignes lues aux élèves. Elle limiterait l'éventualité d'une divulgation de tout ou partie des épreuves, point sur lequel nous reviendrons à propos de l'étude des évolutions temporelles. Certains organismes étrangers font administrer les protocoles des évaluations de type bilan par des examinateurs spécialisés.

La communication des résultats : systématiser et renouveler la présentation

Les exercices

La publication du texte intégral des exercices n'est pas forcément indispensable, un choix d'exercices typiques d'un objectif ou d'un niveau peut être suffisant. Signalons que, dans un souci de réduction des biais, de nombreux organismes ne divulguent pas l'ensemble des exercices susceptibles d'être utilisés pour des comparaisons temporelles ultérieures.

Les indicateurs psychométriques

Nous insistons fortement sur les raisons scientifiques et déontologiques de publier l'intégralité des indicateurs psychométriques (difficulté, discrimination, consistance) pour les items et l'ensemble du protocole. Cette publication est indispensable pour permettre un regard scientifique sur la fiabilité de la mesure. Elle fait partie des procédures recommandées par les organismes garants de la qualité des tests et instruments d'évaluation (par exemple la Commission Internationale des Tests, voir les Recommandations parues en 2002). Ces indicateurs doivent bien sûr être accompagnés des commentaires nécessaires, sans oublier de signaler les choix qui ont pu être faits, leurs raisons et leurs incidences : suppression d'un ou de plusieurs items, fusion de codages, maintien d'items de qualité médiocre, etc. Ces informations peuvent, bien sûr, ne faire l'objet que d'un rapport technique diffusé aux seuls spécialistes. Il est, évidemment, nécessaire de signaler dans les documents pour un public plus large, les biais éventuels limitant la généralisation des résultats des items et des exercices problématiques.

Les résultats

La DEP produit plusieurs publications (Dossiers d'Education et Formation, Notes d'Information). Le principal problème nous semble concerner les types d'indicateurs publiés, en particulier la fréquence de réussite des items. La présentation sous forme de fréquences a le mérite de la simplicité. Elle a l'inconvénient d'induire une réification et surtout d'inciter à des comparaisons qui ne sont pas toujours fiables. En effet, nous avons rappelé que le problème principal de la mesure en éducation est la difficulté de distinguer la réussite de la tâche et la compétence des sujets.

D'autres modes de présentation pourraient être envisagés. Le recours aux modèles de réponse à l'item a été souvent adopté par des organismes étrangers et internationaux d'évaluation en éducation : PISA et PIRLS en sont des exemples. L'avantage est de placer la difficulté des items et la compétence des sujets sur une même échelle. On peut donc lire sur un même graphique la difficulté des items et la fréquence des sujets ayant une compétence correspondant à cette maîtrise. De plus on peut indiquer les niveaux de compétence de différents groupes (selon le genre, la PCS, etc.). Signalons enfin, que ces modèles permettent, à condition d'avoir mis en place un système d'ancrage approprié, d'établir des comparaisons entre les niveaux des sujets des différentes cohortes. On voit l'avantage et l'élégance de cette présentation synthétique. L'inconvénient de ces modèles tient aux contraintes qu'ils imposent (exigences d'unidimensionnalité, d'indépendance locale, etc.). Une réflexion devrait être conduite sur ce point.

Un contrôle scientifique des publications ?

Il nous semblerait souhaitable que soit entreprise une réflexion sur le contrôle de la qualité scientifique des publications de la DEP. Dans les publications scientifiques, le contrôle par les pairs est de règle. Les articles dans des revues à comité de lecture sont révisés par plusieurs experts qui peuvent autoriser ou non la publication et surtout demander des modifications et/ou des précisions. Nous avons rencontré dans les publications de la DEP (revue " Education et Formations ") des articles qui auraient pu être nettement améliorés par des expertises extérieures et des Notes d'Information dont la présentation des résultats, assez maladroitement, induisait des problèmes de compréhension des données.

Les comparaisons temporelles : mieux gérer le dispositif

Un observatoire du système éducatif doit nécessairement être en mesure d'informer sur les changements dans les acquis de nos élèves au cours du temps. Il est indispensable de mettre en place toutes les mesures pour assurer la fiabilité de ces comparaisons. Satisfaire cet objectif introduira des contraintes fortes. Il existe une abondante littérature psychométrique sur les procédures et les méthodes à mettre en œuvre pour construire et gérer un dispositif permettant la comparaisons des acquis des élèves dans le temps. Leur mise en place est une entreprise exigeante demandant de respecter de nombreuses contraintes au niveau du dispositif de production des données, en particulier les protocoles. Il est évidemment indispensable d'utiliser les mêmes exercices, les mêmes items sans les modifier, même d'une virgule, et, sans modifier les conditions de passation (consignes, durée de travail, place de l'exercice dans le protocole). Nous avons déjà signalé que les nouvelles évaluations - bilan nous semblent prometteuses.

Comparer dans le temps, veut dire pouvoir placer sur une même échelle de mesure, des résultats obtenus par des générations différentes, dans un contexte sujet aux changements. Ces changements impliquent qu'il est rarement possible d'administrer des protocoles identiques car certains exercices " vieillissent " plus mal que d'autres. La solution la plus largement préconisée est d'utiliser une partie des exercices pour assurer ce qu'on appelle en termes techniques un ancrage. On peut ainsi ne conserver dans les protocoles qu'une partie des exercices acceptables dans le nouveau contexte.

Construire un tel dispositif rend nécessaire de disposer d'un ensemble important d'exercices et d'items pour assurer cet ancrage. Précisons que ces items d'ancrage ne doivent pas forcément être utilisés indéfiniment : il suffit d'assurer un ancrage entre deux évaluations consécutives. Ces exercices doivent évidemment être représentatifs des contenus et aussi des niveaux de difficulté, et bien sûr, être en nombre suffisant.

Il sera également utile d'envisager des formations spécifiques à ces méthodes.

Conclusion

En dépit des critiques formulées, il importe de saluer l'intérêt et l'importance de l'outil que sont les évaluations de type bilan et les travaux de la DEP. Nous avons remarqué des avancées en matière de psychométrie, qui font suite à des recommandations provenant d'avis du Hcéé. Il faudra poursuivre cet effort et *contaminer* de manière décisive les groupes de pilotage responsables de la construction des protocoles.

Propositions structurelles

Arrêter la définition des concepts utilisés en évaluation, lecture et écriture

La construction des protocoles successifs montre un flottement entre les termes d'objectif, de capacité et de compétence, de domaine et de champ, etc. , avec des absences comme celle d'indicateur ou de critère. Des choix ont finalement été faits. Mais, si les protocoles devaient être plus largement utilisés, une définition claire de ces termes et l'explication des choix effectués seraient profitables et éviteraient peut-être la dérive qui consiste à confondre l'item avec la compétence évaluée .

D'autre part, peut-on établir l'éventail le plus complet possible des compétences et des processus mentaux qui sont à l'œuvre dans les activités de lecture et d'écriture ? Ces compétences et ces processus peuvent-ils être décrits avec leurs niveaux différents de développement selon les différents cycles de l'enseignement ? Peut-on, à partir de ces descriptions, déterminer des niveaux d'exigence à tel ou tel moment du cursus d'un élève ? Quelle est la part de la maîtrise de la langue dans la maîtrise de la lecture et de l'écriture ?

Une telle réflexion, si elle était possible, guiderait et éventuellement harmoniserait la construction des protocoles à différents moments du cursus scolaire. Elle permettrait aussi d'établir explicitement un socle commun aux programmes d'enseignement et aux protocoles d'évaluation, qui rendrait peut-être ces derniers plus accessibles et plus « parlants » aux enseignants. Enfin, certains problèmes en seraient peut-être clarifiés comme entre autres la frontière entre compréhension et interprétation qui inclut le statut de ce qu'on appelle parfois la « compréhension littérale » : pour les uns

la compréhension littérale consiste à saisir toutes les informations explicites d'un écrit, elle est préalable à toute interprétation. Pour d'autres, elle est déjà une interprétation et comprend l'accès à l'implicite. Or on peut penser qu'interpréter, au primaire, est à la fois proche et différent d'interpréter en seconde, mais que la compréhension littérale par exemple est une opération préliminaire nécessaire dans tous les cas. La même étude serait indispensable pour la production écrite. Certains la réduisent à un discours cohérent, répondant à une consigne ou à une situation imposée, et respectueux des codes linguistiques et typographiques de l'écrit. D'autres y voient d'abord un processus constant (et peu évaluable par des tests standardisés) de régulations diverses survenant au moment-même de son élaboration, et qui ne se réduisent pas à la correction du «brouillon», qui est d'abord une activité de lecture. Certains y voient un entraînement à des types de sujets selon les niveaux d'enseignement, d'autres la capacité à transcrire pour soi ou pour d'autres ce qu'on veut mémoriser ou retenir, etc.

Cette vision demande sans doute la participation à la réflexion de tous les spécialistes de la question, universitaires, chercheurs ou enseignants, spécialistes des sciences du langage, de la classe, etc.

Précautions souhaitables si l'on conserve la situation actuelle

Conserver la situation actuelle, c'est-à-dire conserver des protocoles élaborés pour les enseignants et passés par tous les élèves d'un niveau, avec analyse précise des résultats d'un échantillon pour les responsables institutionnels.

Pour éviter de figer l'évaluation, de la ritualiser :

- éventuellement composer les groupes selon la commande,
- repenser la composition des groupes d'élaboration des protocoles et prévoir leur renouvellement partiel régulier ; ménager des relations entre les groupes qui travaillent sur des niveaux proches ou dont les travaux seront mis en perspective (cycle 3-6°, 6°-5°, 6°-3° par exemple) ;
- alléger les protocoles en cherchant non l'exhaustivité, mais un choix hiérarchisé de compétences à évaluer chaque fois en fonction d'une planification précise.

Explications.

Les groupes d'élaboration peuvent fonctionner en circuit fermé et développer une sorte de culture propre qui peut conduire à des protocoles de plus en plus pointus ou à des protocoles sur la logique desquels on ne s'interroge plus. Leur renouvellement partiel pourrait assurer la mémoire nécessaire au groupe à condition que celui qui dirige le groupe ne soit justement pas un «ancien», et leur composition devrait éviter le poids prépondérant de certaines parties du groupe sur d'autres. Par ailleurs, le rôle du représentant de la DEP devrait y être précisé : est-il là pour rappeler la commande et assurer le bon déroulement des opérations, ou doit-il travailler avec le groupe pour l'aider à élaborer des items fiables ? Enfin, quand un groupe travaille dans l'urgence, il réduit le temps nécessaire à l'expérimentation, à l'étalonnage dans les académies qui assure également la fiabilité des items. On doit veiller à maintenir un lien étroit avec «le terrain», professeurs en exercice et groupes académiques.

Si l'on veut suivre l'évolution des acquis des élèves à un même niveau du cursus scolaire, il faudrait reconduire toujours le même protocole : mais ce serait sans doute mal reçu des enseignants ; et on évaluerait seulement alors les acquis qui résultent du choix initialement fait parmi les composantes possibles d'une même compétence. On n'échapperait pas au soupçon selon lequel les scores sont équivalents parce que les items sont identiques. Si l'on veut mieux juger des acquis en élargissant l'éventail des savoir faire évalués ou en modifiant la complexité des tâches demandées aux élèves (non plus restituer seulement un savoir, mais l'appliquer à une situation nouvelle, ou à une résolution de problème, par exemple), on risque d'alourdir des protocoles jugés déjà trop lourds, et on ne pourra plus comparer les acquis. Peut-être pourrait-on par conséquent maintenir chaque année des items communs sur lesquels asseoir un suivi des acquis sur des compétences essentielles, et introduire à la place des autres l'évaluation de nouveaux savoirs et savoir-faire, sans alourdir les protocoles. C'est selon ce principe que les protocoles ont évolué, mais timidement. Ce «roulement» d'items régulièrement repris ou renouvelés exige une planification. Les groupes devraient donc disposer d'une «feuille de route» qui leur montre la finalité du protocole auquel ils travaillent et sa place dans la planification des évaluations. Ce dispositif de variation contrôlée n'empêche pas qu'à un moment, tous les quatre ans par exemple, la totalité d'un protocole soit repris à l'identique.

Pour accroître l'efficacité des évaluations :

- renouveler la réflexion sur la nature des items et des tâches,
- se donner les moyens d'accompagner les enseignants en exercice pour les aider à comprendre et à exploiter les évaluations, surtout dans le second degré ;
- faire des protocoles des outils de communication à l'interne et à l'externe, grâce à des brochures de présentation claires et simples destinées à ceux qui ne sont ni des enseignants ni des spécialistes disciplinaires (parents, membres des équipes de direction, partenaires extérieurs,...).

Explications.

La parcellisation des tâches, dans les protocoles, permet-elle d'atteindre la tâche globale qui donne en réalité leur sens aux savoir-faire limités qu'on évalue ? Les savoir-faire manifestés permettent-ils de juger de la capacité ? En quoi une épreuve donne-t-elle un information sur un savoir réel ? Ces questions sont présentes à l'esprit des utilisateurs des évaluations nationales et il faudrait y répondre.

La sous-utilisation des évaluations par les enseignants est généralement attribuée à une carence de leur formation sur ce sujet. Un développement de la formation continue, qui prendrait appui sur des protocoles ancrés dans les disciplines, et formerait à l'évaluation sans être rejetée comme trop théorique, paraît donc nécessaire. Mais ce qu'a mis en évidence l'expérience passée, c'est qu'un dispositif d'accompagnement pédagogique des évaluations, effectué entre autres par les corps d'inspection et des équipes de formateurs, serait indispensable pour en assurer l'exploitation dans les classes. Les anciens membres des groupes d'élaboration des protocoles pourraient d'ailleurs, dans leurs académies respectives, faire partie de ces formateurs. Ce dispositif aurait déjà le mérite de rendre effectives les conséquences pédagogiques que la DEP tire des résultats mais que les professeurs ne comprennent pas toujours : par exemple que «les compétences ne s'acquièrent pas isolément les unes des autres, ni en une seule fois», que «les aides ponctuelles sont insuffisantes», que «toute acquisition doit être consolidée et enrichie dans des contextes nouveaux.» Il devrait montrer aux enseignants l'enjeu des évaluations : quelle connaissance particulière des élèves fournissent les évaluations, comment nourrir de ces informations les choix et les stratégies pédagogiques.

Les pages de présentation des dossiers *Education et formations* qui analysent les scores des évaluations, ou les notes d'information sont d'un langage relativement technique et ne sont pas largement diffusés. Le moment de leur parution (pour les dossiers, trois mois après l'évaluation analysée en 1992, mais ensuite de six mois à un an après) fait qu'ils sont rarement lus par d'autres que les spécialistes des évaluations. Les pages de présentation des cahiers destinés aux professeurs sont peu exploitées. Il faudrait donc réfléchir aux modes de diffusion et de présentation des évaluations à destination d'un public non enseignant.

Pour afficher la cohérence du pilotage pédagogique : organiser et rendre visible le partenariat entre la DESCO et la DEP.

La logique voudrait que les enseignants (et les chefs d'établissement) puissent situer leurs élèves par rapport aux compétences jugées essentielles à des niveaux clefs du cursus scolaire. La DESCO aurait comme rôle d'identifier ces compétences, la DEP d'élaborer les items et les protocoles, qui seraient ensuite validés par la DESCO. Il y aurait adéquation entre l'objet et la finalité des évaluations, donc cohérence aux yeux des acteurs de l'institution. A la DESCO de définir chaque fois la finalité : alerter sur les compétences de base non acquises, ou mécanisées ; mesurer la maîtrise de parties des programmes ou de compétences négligées dans les protocoles précédents ; etc.

A la DESCO également de se saisir des analyses effectuées par la DEP sur les résultats des évaluations, et de faire les commentaires pédagogiques appropriés. Certains commentaires pédagogiques inspirés par les résultats des évaluations dépendent en effet directement de la logique des protocoles, et ne peuvent sans précaution être généralisés par un lecteur inattentif à l'ensemble de l'enseignement. Par exemple, ces dernières années, les élèves ont leurs meilleurs scores dans le champ de la lecture (compréhension) ; pourtant, c'est aussi dans ce champ que se trouvent quelques exercices qui obtiennent les scores les plus bas ; les notes d'information de la DEP en concluent logiquement que c'est donc en lecture qu'il faut surtout travailler. Mais si le protocole analysé avait, comme d'autres autrefois, largement évalué la maîtrise de la langue, ce serait peut-être là qu'on aurait trouvé les scores les plus bas, entraînant une autre conclusion sur les priorités de l'enseignement. D'autre part, on peut rappeler certains éléments du discours institutionnel accompagnant actuellement les résultats des évaluations qui se prêtent à des malentendus et devraient être révisés. Ainsi, le lien établi entre erreur et remédiation est mal compris. Surtout on insiste sur la connaissance des élèves donnée

par les évaluations sans indiquer la finalité de cette connaissance : il ne s'agit pas de connaître pour connaître, mais pour permettre au professeur de réguler son action pédagogique. Du coup, les résultats des évaluations ne remplacent pas, mais complètent, la connaissance que les professeurs ont de leurs élèves par les travaux habituels de la classe.

Une autre possibilité construire : des outils différents pour les enseignants et pour les décideurs

C'est-à-dire ne pas conserver un même protocole pour deux types de destinataire.

On peut récapituler les faits qui suggèrent cette solution. D'abord, il semble qu'actuellement les objectifs des évaluations soient brouillés dans l'esprit des différents utilisateurs, qui sont conduits, selon les cas, à s'en arranger ou à s'en détourner, peu en étant pleinement satisfaits. Il n'est pas satisfaisant non plus que les évaluations aient été utilisées comme des bilans ; mais ce fait montre le besoin, à côté d'évaluations diagnostiques, d'évaluations - bilan.

D'autre part, comme on l'a vu, les protocoles actuels, ni totalement identiques ni vraiment différents, portent les marques contradictoires de leur double finalité. Et l'absence d'un indicateur de niveau, même empirique, ne permet pas non plus de fournir aux décideurs des informations précises et comparables sur les compétences et les déficiences des élèves dits les plus faibles.

Enfin, les évaluations internationale PIRLS et PISA ont jeté, sur les compétences des élèves français, un regard nouveau. On pourrait dire, en forçant le trait, qu'elles ont apporté une authentique information. En effet, les résultats annuels des évaluations nationales montrent en gros les mêmes pourcentages d'élèves maîtrisant ou non les compétences évaluées, avec des différences ponctuelles imputables à divers facteurs implicites dont on laisse aux professeurs le soin de les clarifier avec leurs élèves. Quoi qu'on pense de l'évaluation PISA, l'exploitation approfondie d'une définition de la lecture pour en induire le choix des compétences, la définition des trois compétences et de leurs cinq niveaux respectifs de complexité, les divers formats de réponse qui ne recouvrent pas toujours les nôtres, la répartition des élèves en cinq groupes constitués à partir de compétences communes, ont donné des résultats intéressants sur les forces (s'informer, interpréter) et sur les carences (communiquer sa réponse ou son avis, en particulier par écrit) de nos élèves. Parfois ils recouvrent nos propres constatations, parfois ils en apportent de nouvelles. En outre, la répartition des élèves les plus faibles dans les groupes en fonction des items permet d'avoir une idée des carences (cognitives mais aussi procédurales) de ces élèves. Certes, l'évaluation PISA n'est pas faite pour évaluer le niveau scolaire de nos élèves. Mais elle pose au moins une question : si l'on veut avoir plus de connaissances sur le niveau acquis par les élèves à un moment donné de leur scolarité, faut-il nécessairement faire toujours passer, aux mêmes ou à tous, un protocole identique ? Les scores assez proches de nos élèves, d'une année sur l'autre, seraient-ils différents sur d'autres compétences de la même capacité ou sur d'autres niveaux de la même compétence ? Et cela nous donnerait-il des informations nouvelles donc éventuellement utiles ?

Si l'outil doit différer selon les destinataires, de quoi auraient besoin des décideurs ? Quelles seraient la nature et la finalité de l'outil ?

Il semble que les décideurs aient besoin d'évaluations bilans ou d'évaluations ciblées.

De telles évaluations peuvent difficilement se fonder sur les résultats d'un échantillon représentatif d'élèves, comme on le fait aujourd'hui. Car le fait de constituer un échantillon ne change rien à la logique du protocole qu'on utilise, actuellement destiné d'abord à éclairer les professeurs pour leur enseignement. Or les finalités sont différentes. Et l'une des dérives tient précisément au fait que des évaluations de type diagnostique ou pronostique ancrés dans les programmes sont utilisées comme des évaluations bilan indiquant les effets d'une politique éducative. Si donc on utilise un échantillon, ce choix sera dû à la finalité de l'évaluation correspondante, et il ne dispensera pas d'élaborer des protocoles spécifiques selon les questions habituelles : que veut-on savoir, de quoi a-t-on besoin, que veut-on faire ? Quel public cible-t-on ? A quel niveau du cursus scolaire ? Quelle périodicité choisit-on ? Quelles modalités ?

Certains de nos interlocuteurs ont émis des attentes, d'autres se lisent dans leurs rapports, par exemple :

- des évaluations-bilan permettant d'apprécier l'efficacité de l'enseignement et de mesurer le niveau d'atteinte des objectifs sur un cycle entier (école, collège, lycée, en évitant les examens de fin de cycle) ou sur une partie de cycle (fin du cycle 3, fin du cycle central, fin de 2^o) afin d'asseoir un

diagnostic complet et fiable. La pratique des évaluateurs du système éducatif, qui cherchent plus à saisir une tendance qu'à comparer les résultats annuellement, montre qu'une périodicité de trois ou quatre ans entre chaque évaluation de ce type pourrait être suffisamment éclairante. Mais il faut noter que le caractère national des évaluations, leur rôle de référent unique, semblent importants pour les décideurs, les pilotes et les évaluateurs du système éducatif, qui voudraient donc le conserver. Enfin, ces évaluations n'étant pas destinées aux pédagogues, ces derniers ne seraient pas nécessairement majoritaires dans les groupes qui décideraient du choix des compétences et des modalités de l'évaluation ; la fiabilité des items et la précision de la mesure pourraient s'en trouver renforcées, puisqu'elles n'auraient pas à s'adapter aux attentes et aux usages des enseignants.

- Des évaluations-bilans permettant de vérifier, sur des compétences de base à retenir et à définir, comment évolue le niveau global des élèves à des seuils déterminés de leur cursus : soit en faisant passer le même protocole à un échantillon d'élèves à X années de distance (type évaluation CM2 en 1987 et en 1997), soit en soumettant une même cohorte d'élèves à des évaluations correspondant aux différents niveaux de leur cursus. Cela suppose la reprise des mêmes protocoles, sur des échantillons équivalents, selon une périodicité définie (cinq ans, dix ans,...) dans le premier cas, une observation très fine des réussites et des erreurs dans le second.

- Des évaluations ciblées sur des problèmes identifiés (éventuellement à la suite des évaluations bilans) qu'on veut comprendre pour pouvoir y remédier, des études sur échantillons pour «comprendre ce qui se passe» et qui n'ont lieu qu'aux moments où elles sont nécessaires :

* les enseignants français sont ceux qui font le moins confiance à l'école et qui accordent le plus d'importance aux facteurs socio-culturels, dit-on.

Donc, *quel regard les maîtres portent-ils sur leurs élèves ?*

* Quand les résultats sont bas au cycle 2, ils s'effondrent au cycle 3 : que se passe-t-il chez les maîtres ?

*En référence à l'étude faite sur les résultats des élèves en ZEP et hors ZEP, «trouver des types de situations où on fait la même étude» pour comprendre les effets des pratiques pédagogiques et didactiques sur les compétences des élèves de ZEP.

* Qu'en est-il des handicaps acquis vers 10-11 ans ?

* Au lieu de vérifier toutes les compétences tous les ans, hiérarchiser les compétences et ne vérifier que «ce qui met globalement les élèves en difficulté» en essayant d'identifier la source des erreurs.

* Qui sont les élèves les plus faibles et où sont-ils ? Recouvrent-ils les élèves de ZEP, ou les illettrés de la JAPD ? A quel seuil dans la non-maîtrise des compétences s'arrêter pour les déterminer ? Que savent ou ne savent-ils pas faire ? Pourquoi? etc.

De quoi auraient besoin les enseignants, et les responsables du pilotage pédagogique, cadres intermédiaires et formateurs ?

Selon les formateurs interrogés, il est logique que les protocoles destinés aux enseignants soient, comme aujourd'hui, à finalité diagnostique et/ou pronostique donc ancrés dans les programmes.

Les exercices pourraient adopter des formes et utiliser un métalangage dont la complexité ou l'approximation seraient moins gênants pour la validité de l'évaluation. En effet, pour prendre l'exemple des protocoles de seconde, la complexité des textes conduit parfois à des types de réponse (les tableaux) qui ont pu déconcerter les élèves et donc affaiblir la fiabilité de l'évaluation. Il est arrivé aussi qu'un libellé, scientifiquement approximatif et peut-être source d'erreur lors de l'évaluation, ait été choisi parce que, paradoxalement, les élèves testés au moment des étalonnages l'avaient mieux compris que d'autres libellés. Cette tension entre ce qui est valide pour les techniciens de l'évaluation et ce que peuvent reconnaître, accepter et donc utiliser les praticiens de l'enseignement serait moins vive et moins grave dans le cas d'un protocole destiné d'abord aux élèves et aux enseignants.

Le caractère prescriptif que les formateurs reconnaissent aux protocoles demeurerait, même pour des exercices d'évaluation complexes et peu utilisés dans les séquences d'apprentissage. Et rien n'oblige alors à utiliser des protocoles identiques : ils pourraient s'adapter plus étroitement aux évolutions des programmes.

La question se pose de savoir comment conserver aux protocoles destinés aux enseignants leur valeur diagnostique ou pronostique si ces protocoles ne sont pas annuels. Mais un protocole élaboré

une année pourrait servir, si l'on peut dire, l'année ou les deux années suivantes, puisque ce ne sont normalement pas les mêmes élèves qui les passent. Il est possible également qu'un protocole propose toujours au moins deux exercices pour un même item, de manière à rendre possible l'utilisation d'un même protocole plusieurs années de suite. Ce n'est pas nouveau : il était déjà prévu que des professeurs de seconde utilisent le même protocole au début et à la fin de l'année (ou en cours d'année) pour mesurer les progrès des élèves.

Quelles seraient les limites ou les dangers à ne pas perdre de vue pour que les évaluations soient réalisables et effectives sur le terrain ?

Compte tenu du temps que demande l'élaboration d'un protocole d'évaluation, il semble utopique de confier ce soin aux équipes enseignantes des établissements pour évaluer leurs élèves à des moments et à des niveaux scolaires définis par l'institution.

D'autre part, l'utilisation des banques d'outils suppose chez les enseignants le besoin d'évaluer pour enseigner : ce besoin, les évaluations nationales à leur naissance avaient mission de le satisfaire ou devaient le susciter. Si les évaluations nationales n'y sont pas encore parvenues, comment les banques d'outils le feraient-elles? Et si elles ne le font pas, quelle est leur utilité? Comment éviterait-on en outre la dérive qui s'observe actuellement, l'utilisation des exercices de la banque d'outils (ou des protocoles) comme des exercices d'apprentissage ? Si donc on voulait rapprocher les évaluations nationales des enseignants afin que ces derniers en voient la nécessité, peut-être le niveau académique serait-il préférable pour élaborer, faire passer et exploiter des protocoles en fonction d'une commande, d'un cadrage et d'un suivi définis et organisés nationalement. L'autre limite à laquelle il faut penser porte sur la multiplication des évaluations nationales, surtout quand elles sont couplées à un examen de fin d'année. Les écueils sont connus : la ritualisation, qui fait des évaluations un événement banal et sans lendemain, une saturation qui deviendrait insupportable, si les évaluations se multipliaient dans les établissements. La conscience de cette limite conduit à penser que le calendrier et la périodicité des évaluations, liés à la densité ou légèreté des protocoles selon leur finalité respective, sont des modalités essentielles à toute réflexion sur les évaluations.

Enfin, il serait dangereux que des évaluations-bilan situées en fin de cycle ne soient pas une aide à l'orientation mais la cause invoquée pour orienter.

Entretiens

- Elisabeth Bautier, professeure des universités, Paris-VIII
- Corinne Bonassieux, PRAG de lettres, université, Paris-XII, Val de Marne
- Viviane Bouysse, chef du bureau des Ecoles, DESCO
- Anne-Marie Chartier, maître de conférence, INRP, membre du groupe national de pilotage Evaluation-CE2
- Marc Colmant, chargé d'études, DEP
- Marielle Court, journaliste, *Le Figaro*
- Emmanuel Davidenkoff, journaliste, *Libération*
- Jean-Luc Derouet, professeur des universités, INRP
- Daniel Dubois, IGEN de lettres honoraire
- Jean -Claude Emin, sous-directeur de l'évaluation, DEP
- Gaëlle Fouéré, journaliste, *La lettre de L'Education*
- Claudine Garcia-Debanc, professeure des universités, IUFM Toulouse
- Jean-Emile Gombert, professeur des universités, Rennes II
- Jacqueline Levasseur, chef du bureau de l'évaluation des élèves
- Bernard Maccario, sous-directeur DESCO
- M. Nembrini, IGEN, cabinet du ministre de l'Education nationale
- Claudine Peretti, directrice de la DEP
- Bruno Petey-Girard, maître de conférences, Paris XII Val de Marne
- Dominique Pierrel, IA-IPR de lettres, académie de Nancy-Metz
- Sylvie Plane, professeure des universités, IUFM Paris
- Jean-Pierre Ruffier, IA-IPR adjoint à l'IA de l'académie de Savoie
- Thierry Rocher, chargé d'études, DEP
- Martine Saffra, IGEN, doyenne du groupe enseignement primaire
- Claude Thélot, conseiller à la Cour des comptes, ancien directeur de la DEP
- Henry Théodet, principal du collège Fabien (ZEP), 93100 Montreuil
- Francis Thourigny, professeur de lettres, CRDP de l'académie de Créteil, membre du groupe national de pilotage Evaluation- 6°
- Bruno Troselle, chargé d'études DEP
- Gérard Vignier, IA-IPR de lettres, académie de Versailles
- Pire Vrignaud, maître de conférences, Université de Paris X Nanterre
- Des professeurs des écoles et de collèges de l'Essonne, entretiens mis à disposition par Martine Rémond,
- Des professeurs des écoles et des IEN des Hauts de Seine et de l'académie de Nantes.

Annexe 2.

Liste de documents utilisés dans le rapport.

Evaluation de l'enseignement dans l'académie de Limoges.

Rapport de l'Inspection générale de l'Education nationale et de l'Inspection générale de l'administration de l'Education nationale et de la recherche. Septembre 2000. n° 00-0070

Evaluation de l'enseignement dans l'académie de Rennes.

Rapport de l'Inspection générale de l'éducation nationale et de l'Inspection générale de l'administration de l'éducation nationale et de la recherche. Septembre 2000. n° 00-0071.

Evaluation de l'enseignement dans l'académie de Lyon.

Rapport de l'Inspection générale de l'éducation nationale et de l'Inspection générale de l'administration de l'éducation nationale et de la recherche. Septembre 2001. n°

Evaluation de l'enseignement dans l'académie de Créteil.

Rapport de l'Inspection générale de l'éducation nationale et de l'Inspection générale de l'administration de l'éducation nationale et de la recherche. Janvier 2003 (n° 03 001).

Les modules au lycée.

Rapport de l'Inspection générale de l'Education nationale. Octobre 1998. n° 98-001

Suivi de la mise en place de la réforme des lycées.

Rapport de l'Inspection générale de l'Education nationale. Année 1999-2000.

La classe de sixième, état des lieux et réformes en cours.

Rapport de l'Inspection générale de l'Education nationale
Mars 2002, n° 2002-018.

Evaluation pédagogique dans les collèges

Fin du cycle d'observation juin 1982.

MEN, Direction des collèges, Division de l'évaluation du système éducatif.

Evaluation pédagogique en classe de seconde .

Les dossiers éducation et formations, n° 1 avril 1989.

Evaluation en fin de troisième technologique. Les dossiers, mai 1989.

Evaluation pédagogique en fin de cinquième , 1988.

Dossier Education et formations n°6, juillet 1990.

Evaluation en cycle d'observation, 1° phase : l'entrée au collège

Rapport d'étape de septembre 1991.

DEP 10, document de travail n° 398.

Evaluation en cycle d'observation. 2° phase : fin de cinquième 1991

Document de travail n° 409, DEP 10, décembre 1992.

Evaluation français sixième.

Collection Réponses pédagogiques , CRDP Poitou-Charentes, 1992.

Evaluation CE2-6°, résultats nationaux, septembre 1992.

n°21 de décembre 1992.

- Evaluation CE2-6° de septembre 1992. Résultats complémentaires et analyse.*
Les dossiers, n° 31, août 1993.
- Evaluation au cycle d'observation (1989-1992). Cycle en 3 ans, redoublement et autres mesures d'aide : quelle efficacité ?* Les dossiers, n° 40, mars 1994.
- Comparaison des performances des élèves scolarisés en ZEP et hors ZEP*
dans le dossier Education et formations n° 41 de juin 1995
- Evaluation en fin de troisième générale et technologique 1995*
Dossier éducation et formations n° 86 de mai 1997
- Evaluation pédagogique et orientation à la fin du cycle d'observation* SDEVA (Sous-Direction de l'Evaluation du système éducatif) n° 61, mars 1987.
- Evaluation pédagogique et orientation à la fin du cycle d'orientation* SDEVA n° 62, juin 1987.
- Evaluation CE2-6°, résultats nationaux septembre 1995.*
Les dossiers d'éducation et formations, n° 65 de mars 1996
- Evaluation CE2-6°, Repères nationaux*
Les dossiers , de 1993 à 2001.
- Evaluations CE2- sixième-cinquième, Repères nationaux septembre 2002*
Les dossiers, n° 141, avril 2003.
- Evolution des acquis cognitifs et socio-affectifs des élèves au cours des années de collège .*
Les dossiers, n° 88, août 1997.
- Grand thème : évaluation et notation des élèves*
Education et formations, n° 53, mars 1998.
- Enquête sur les dispositifs pour les élèves en difficulté en 6° et en 5°*
Académie de Nantes, 26.05.1998.
Synthèse rédigée par Bernard Marquis.
- Grand thème : les sortants du système scolaire sans qualification*
Education et formations, n° 57, juillet-septembre 2000.
- Etude spécifique relative aux élèves en difficulté en lecture à l'entrée en sixième (septembre 1997).*
Dossier DEP n°112 d'octobre 1999 (Ravi Baktavatsalou et Corinne Pons).
- Les compétences des élèves français à l'épreuve d'une évaluation internationale , Premiers résultats de l'enquête PISA 2000 .*
Les dossiers , n° 137, de novembre 2002.
- Goffard Serge, *Entrer dans l'écrit, les genres du discours*
collection Argos Références, CRDP de l'académie de Créteil, 1997
- Jouve Vincent , *La lecture* , Hachette (collection Contours littéraires) 1993.

Références

- ADAMS, R. & WU, M. (Eds.). (2002). PISA 2000 Technical Report. Paris : O.E.C.D.
- AUBRET, J. (1995). Psychologie de l'évaluation. In D. Gaonac'h et C. Golder, Manuel de psychologie pour l'enseignant. Paris : Hachette Education.
- AUBRET, J. & BLANCHARD, S. (1991). *L'évaluation des compétences d'un lecteur*. Issy les Moulineaux: EAP.
- AUBRET, J., DEMANGEON, M., LARCEBEAU, S. & PELNARD-CONSIDERE, J. (1981). Révision et étude de quelques tests, p. 99-107. *Document du Service de Recherche de l'INOP, 1981/1*.
- AUBRET, J. & DOSNON, O. (1993). Les illettrés et le bilan de compétences personnelles et professionnelles. Service de Recherches de l'INETOP.
- BAIN, D. et SCHNEUWLY, B. (1993). Pour une évaluation formative intégrée dans la pédagogie du français : de la nécessité et de l'utilité de modèles de références. In L. Allal, D. Bain et P. Perrenoud (Eds.), *Evaluation formative et didactique du français*. Neuchâtel : Delachaux et Niestlé.
- BONNET, G. (2002). Vers une rémunération au mérite des enseignants anglais. Note pour le HCéé.
- BONNET, G., BRAXMEYER, N., HORNER, S., LAPPALAINEN, H.P., LEVASSEUR, J., NARDI, E., REMOND, M., VRIGNAUD, P. & WHITE, J. (2001). *The use of national reading tests for international comparisons : Ways of overcoming cultural bias*. A European Project: The European Network of Policy Makers for the Evaluation of Educational Systems. Paris : D.P.D. Edition.
- BONNET, G., DAEMS, F., de GLOPPER, C., HORNER, S., LAPPALAINEN, H.P., NARDI, E., REMOND, M., ROBIN, I., VRIGNAUD, P. & WHITE, J. (2003). *Cultural balanced assessment of reading. C- BAR*. A European Project: The European Network of Policy Makers for the Evaluation of Educational Systems. Paris : D.P.D. Edition.
- BONORA, D. & BACHER, F. (1986). Intérêt et difficultés des comparaisons dans les enquêtes internationales, l'exemple de l'IEA. *Revue Tunisienne des Sciences de l'Education*, 11, 14, 129-138.
- BONORA, D. & VRIGNAUD, P. (1997). Evolution des connaissances scolaires et Modèles de Réponse à l'Item. Rapport pour le Ministère de l'Education Nationale. DEP.
- CAMPBELL, J. et al. (2000). Framework and Specifications for PIRLS Assessment 2001 – *Progress in International Reading Literacy Study; I.E.A., (2nd edition) Chestnut Hill, MA : Boston: College*.
- DEROUET, J.L., NORMAND, R. & REMOND, M. (2003). *Le développement d'une culture de l'évaluation dans l'Education nationale : comment les enseignants utilisent –ils les évaluations nationales ?* Note d'étape établie pour le contrat DEP – Commissariat au Plan.
- DICKES, P., TOURNOIS, J., FLIELLER, A. & KOP, J.L. (1994). *Psychométrie*, Paris, P.U.F..
English : The National Curriculum for England, Key Stages 1 – 4. London : Department for Education and Employment and Qualifications and Curriculum Authority, 1999.
- FAYOL, M., DAVID, J. DUBOIS, D. & REMOND, M. (2000). *Maîtriser la lecture. Observatoire National de la Lecture*. Paris : C.N.D.P. et Odile Jacob.
- FLIELLER, A. (2001). Les compétences et les performances cognitives dans l'évaluation scolaire. In J. Bourdon & C. Thélot (Eds.). *Education et formation : l'apport de la recherche aux politiques éducatives*. Paris : Editions du CNRS.

- GREGORY, K.D. & MARTIN, M.O. (2001). *Technical standards for I.E.A. studies : An annotated bibliography. I.E.A. (Eds.). Wellington : Freestyle Artworks.*
- GUERIN – PACE, F. & BLUM, A. (1999). *L'illusion comparative. Les logiques d'élaboration et d'utilisation d'une enquête internationale sur l'illettrisme. Population, 54, 271 – 302.*
- LAFONTAINE, D. (1996). *Performances en lecture et contexte éducatif. Enquête internationale menée auprès d'élèves de 9 et 14 ans. Bruxelles : De Boeck.*
- Les compétences en littéracie.* (2003). Rapport thématique de l'enquête PISA 2000. Office fédéral de la statistique – Conférence suisse des directeurs cantonaux de l'instruction publique. Neuchâtel.
- Measuring student knowledge and skills. The PISA 2000 assessment of reading, mathematical and scientific literacy. *Paris : O.C.D.E..*
- Framework and specifications for PIRLS Assessment 2001 – Progress in International Reading Literacy Study. I.E.A., Boston: PIRLS ISC.*
- MULLIS, I., MARTIN, M., GONZALEZ, E. & KENNEDY, A.. (2003). *PIRLS 2001 International report. IEA's study of reading literacy achievement in primary schools in 35 countries. Boston college : I.S.C..*
- MULLIS, I., MARTIN, M.O., KENNEDY, A. & FLAHERTY, C. (2002). *PIRLS 2001 Encyclopedia. A reference guide to reading education in the countries participating in IAE's Progress in International Reading Literacy Study (PIRLS). Boston college : I.S.C..*
- NEALE, M.D. (1966). *The Neale Analysis of Reading Ability.* London: Macmillon Education.
- Europe.* Toulouse : P.U. du Mirail, Collection Questions d'Education.
- O.C.D.E (1999). *Measuring student knowledge and skills. The PISA 2000 assessment of reading, mathematical and scientific literacy. Paris : O.C.D.E.*
- O.C.D.E (2001). *Knowledge and skills for life. First results from PISA 2000. Paris O.C.D.E. (2003). Regards sur l'éducation. Les indicateurs de l'OCDE. Paris : O.C.D.E.*
- SALINES, M. & VRIGNAUD, P. (2001). *Apprécier et certifier les acquis des élèves en fin de collège : diplômés et évaluations – bilans. Les rapports, HCéé, 2.*
- REMOND, M. (1993). L'exploration des stratégies de lecture, in *Lecture-Ecriture : Approches de recherche*, Paris, INRP. (p.p. 115 - 146).
- REMOND, M. (1996). Pour une pédagogie de la compréhension au cycle 3. In *Regards sur la lecture et ses apprentissages*, (pp. 165 - 179). Observatoire National de la Lecture. Paris : M.E.N. E. S. R.
- REMOND, M. (1997). La construction de la compréhension du CE1 au CM2. Rapport d'expertise établi à la demande de la D.E.P..
- REMOND, M. (2001a). *Adapter n'est pas traduire : Adaptation dans différents contextes culturels d'épreuves d'évaluation de la littéracie. In C. Sabatier et P. Dassen (Eds.), Cultures, développement et éducation. Autres enfants, autres écoles. Paris : L'Harmattan.*
- REMOND, M. (2001b). *Evaluer l'activité de lecture. Les Journées de l'Observatoire : La lecture de 8 à 11 ans. Paris : Observatoire National de la Lecture.*
- RUFFIER, J.P. (2001). *Evaluer l'activité de lecture. Les Journées de l'Observatoire : La lecture de 8 à 11 ans. Paris : Observatoire National de la Lecture.*

- ROCHER, T.. (sous presse). La méthodologie des évaluations internationales des compétences. *Psychologie et psychométrie*.
- The National Literacy Strategy : Framework for teaching, <http://standards.dfes.gov.uk/literacy/>
- The Nation's Report Card. N.A.E.P.. Reading Highlights 2002. U.S. Department of Education, Institute of Education Sciences. N.C.E.S. 2003 – 524. <http://nces.ed.gov/nationsreportcard>
- VRIGNAUD, P. (2001). Evaluation sans frontières, in M. Huteau (Ed.), *Les figures de l'intelligence*. E. A. P.
- VRIGNAUD, P. (2001). *Evaluer l'activité de lecture. Les Journées de l'Observatoire : La lecture de 8 à 11 ans*. Paris : Observatoire National de la Lecture.
- VRIGNAUD, P. (sous presse). *Objectivité et authenticité dans l'évaluation. Avantages et inconvénients des Questions à Choix Multiples et des Questions à Réponses Complexes pour l'évaluation des compétences verbales*. Psychologie et psychométrie.
- VRIGNAUD, P. & BONORA, D. (1998). *Literacy assessment and international comparisons, in Dan Wagner, Literacy assessment for out-of-school youth and adults*. Philadelphia: UNESCO/International Literacy Institute.
- VRIGNAUD, P., CASTRO, D. & MOGENET, J.L. (2002). *Recommandations internationales sur l'utilisation des tests. Version 2000. Version élaborée pour la S.F.P.*. Pratiques Psychologiques, numéro spécial H.S..

Evolution des tableaux de compétences dans les protocoles de 6°

I. Désignation et organisation des compétences en lecture.

Les notions d'objectif et de compétence.

En 1993 et jusqu'en 1999, le tableau a pour titre *Tableau des objectifs* ; ce dernier terme désigne aussi explicitement les deux divisions verticales du tableau, *savoir lire* et *savoir écrire*. Il reprend les tableaux antérieurs, où *maîtriser des mécanismes de base*, *savoir lire* et *savoir écrire*, étaient intitulés *Objectifs de l'évaluation et compétences*. En 1997-1998, *savoir lire* et *savoir écrire* sont appelés *objectifs généraux*, ce qui montre le désir de distinguer ce qui est de l'ordre d'un savoir faire global indépendant des situations, et ce qui en dépend et qui s'évalue dans des situations d'activité déterminées (les exercices du protocole). C'est pourquoi l'organisation horizontale du tableau fait apparaître, en face de chacun de ces grands «objectifs», les *compétences* qui en dépendent puis les *composantes évaluées*, qui sont les activités précises demandées dans les exercices du protocole. C'est en 1999 qu'apparaissent le titre et le tableau qui demeureront ensuite : *Tableau des compétences évaluées* à la rentrée en 6°.

Cette évolution traduit une réflexion sur la nature de l'évaluation. En effet le terme d'objectif désigne habituellement ce qui est visé dans une situation d'apprentissage, donc en classe. Ce qu'on cherche à mesurer dans l'évaluation n'est pas l'objectif visé par l'enseignement, mais la compétence qui en résulte si l'objectif est atteint à la fin de l'apprentissage. La différence entre *compétences* et *composantes évaluées* vise également à indiquer que les situations d'activité demandées dans les exercices n'épuisent pas les compétences dont elles sont une mise en œuvre parmi d'autres.

La notion de champ.

Dans le tableau de 1993, le *champ* désigne horizontalement les divers domaines de la langue ou de l'écrit auxquels seront rattachés les exercices d'évaluation : approche des textes, vocabulaire, grammaire, orthographe. On reprend ici les tableaux antérieurs où ces domaines étaient appelés *champs d'application*. Chaque activité demandée dans les exercices est donc, dans le tableau, à l'intersection d'un objectif (savoir lire ou savoir écrire) et d'un champ. Ainsi «composer un récit» est une forme du savoir écrire qui est rattachée à l'approche des textes puisque cette activité aboutit à produire un texte. On a donc un tableau statique qui présente un classement des activités par relations internes.

La notion de champ disparaît des protocoles suivants pour réapparaître à partir de 1999 avec un autre sens. Elle définit alors ce en quoi consistent spécifiquement savoir lire et savoir écrire : la *compréhension* et la *production de texte*, activités qui sont redéfinies, dans la colonne des *compétences*, en *compétences en réception* et *compétences en production*. Ainsi savoir lire est défini comme une activité de compréhension, qui se manifeste entre autres, quand on «reçoit» un texte tout fait au lieu de le produire, par les opérations suivantes : comprendre un texte dans son ensemble, construire des informations à partir d'un texte, comprendre l'organisation logique d'un texte, prélever des informations ponctuelles dans un texte.

Il ne s'agit plus seulement là d'une réflexion sur l'évaluation, mais d'une réflexion simultanée sur la lecture et l'écriture. On ne cherche plus à énumérer les domaines de la langue concernés par les activités demandées à l'élève ; on détermine les opérations mentales exigées spécifiquement de lui quand il lit ou quand il écrit. La langue devient alors un *champ* intermédiaire, le matériau commun dont la maîtrise est nécessaire soit pour lire, soit pour écrire. La lecture devient une activité de construction du sens lorsqu'on se trouve en situation de recevoir un texte lu ou entendu mais élaboré par un autre, l'écriture une activité de construction de sens quand on se trouve en situation de produire soi-même un texte écrit.

Cette redéfinition et cette réorganisation des compétences ont été préparées, dans le *Tableau d'objectifs* de 1997, par l'apparition de deux rubriques complémentaires, «connaissance du code» pour le savoir lire et «maîtrise du code» pour le savoir écrire. Les regroupements d'items opérés, en 1998, par le logiciel Casimir (présentation générale, page 5) pour le calcul des scores partiels les

reprennent et proposent donc les rubriques suivantes : compréhension, connaissance et maîtrise du code, production de textes. On les retrouvera dans le tableau de compétences de 1999, la seconde rubrique étant redéfinie en «maîtrise des outils de la langue», pour lire ou pour écrire.

La notion de capacité.

A partir des protocoles de 2001, le tableau est simplifié et clarifié. Toutes les compétences sont traduites en termes d'activité, c'est-à-dire par des verbes, ce qui fait disparaître *en réception* et *en production*. D'autre part, on rend manifestes les rapports de dépendance entre les compétences en introduisant la notion de *capacité*. L'introduction de cette notion permet de distinguer des aptitudes générales acquises dans un domaine et non mesurables (les capacités), et les compétences qui en dépendent, développées dans l'enseignement et donc évaluables dans le protocole. Ainsi en 2002, comprendre un texte, maîtriser les outils de la langue et produire un texte sont désignées par le terme de *capacités*, dont dépendent les *compétences* évaluées et leurs *composantes*, les activités précises choisies dans les exercices du protocole pour évaluer ces compétences. La notion de champ ne désigne donc plus que ces domaines généraux où s'exercent les capacités, savoir lire et savoir écrire, rubriques dont la désignation a été longtemps soit occultée soit incertaine.

L'aboutissement, sans doute provisoire, de cette réflexion ininterrompue désigne donc désormais toutes les rubriques des protocoles comme des activités de l'élève lecteur et scripteur, appelé à mobiliser selon les cas des savoirs et des savoir-faire spécifiques. La langue, au début décrite selon les divers domaines qui la composent et où elle s'observe (les textes), devient le matériau dont la maîtrise conditionne la mise en œuvre de ces activités, qui visent toutes à construire du sens. Il est regrettable que les professeurs qui ne s'approprient pas les protocoles ne puissent pas partager la réflexion sur l'évaluation, sur la lecture et l'écriture que les groupes d'élaboration des protocoles ont manifestement conduite.

Effets de ces changements sur la redistribution des activités de lecture demandées aux élèves dans les tableaux de compétences.

Quand on suit d'un protocole à l'autre le libellé des exercices et la formulation des savoir-faire visés par ces libellés, on voit que la réflexion sur la nature et l'organisation des compétences a conduit parfois mais rarement à proposer de nouveaux savoir-faire à évaluer, plus souvent à définir plus rigoureusement les compétences visées et donc à les déplacer dans les tableaux, à les joindre à d'autres compétences dépendant de la même capacité, ou à les abandonner.

*»Identifier les référents des pronoms personnels pour comprendre un texte», c'est-à-dire saisir la chaîne référentielle, est le onzième exercice du protocole de 1989, et se retrouve dans tous les protocoles, compte tenu de son importance dans l'acte de lecture, à peu près sous les mêmes formes : croix dans un tableau à colonnes, flèches ou couleurs pour relier les personnages aux pronoms personnels qui les représentent. On adjoint plus tard aux pronoms personnels les groupes lexicaux qui désignent les différents personnages.

Placé pendant un temps dans le champ «grammaire» de l'objectif «savoir lire», donc plutôt considéré comme un savoir-faire linguistique, il dépend, en 1997, de «l'objectif» « construction d'informations et de significations sur un texte » ; il devient donc plutôt une activité de compréhension. A partir de 1999, et jusqu'en 2002, il rejoint les autres savoir-faire de la compétence «comprendre l'organisation logique d'un texte». L'activité de l'élève demeure la même lors de l'exercice. Mais le savoir-faire n'est plus présenté comme évaluant la façon dont l'élève lecteur utilise la langue pour construire une information sur le texte : il est présenté comme une opération mentale indispensable pour comprendre la cohérence interne (sens de «logique») du texte, en se repérant dans les désignations et identifications successives des personnages.

* Il prend alors la place d'un savoir-faire intitulé «reconstituer un texte en utilisant des déterminants et des pronoms», appelé en 1993 «mettre le bon déterminant» et repris de l'évaluation expérimentale des élèves de 6^e de 1980. Ce savoir-faire était, en 1993, classé sous l'objectif «savoir écrire», sans doute parce que l'élève devait inscrire lui-même dans les blancs d'un texte le déterminant adapté au sens (article, démonstratif, possessif, indéfini, etc.). En 1997, il fait partie du savoir lire, et précisément de «la compréhension logique» d'un texte, puisqu'il s'agit effectivement d'une activité qui montre si l'élève saisit ou non la cohésion grammaticale du texte. Mais se repérer dans la chaîne référentielle a été légitimement considéré comme plus fondamental pour comprendre la cohérence interne du texte que la capacité, plus locale, d'utiliser les déterminants adaptés. Ce dernier savoir faire a donc été placé en 1999 dans «la maîtrise des outils de la langue pour lire», en quinzième position

dans ce protocole.

La reformulation des intitulés des exercices, leur déplacement dans les rubriques des tableaux de compétences ne sont donc pas gratuits. Ils signifient que les compétences visées sont de mieux en mieux définies, ainsi que leur rôle dans les opérations de compréhension en lecture. Cette réflexion a conduit à ajouter à la compréhension d'un texte lu celle d'un texte entendu (où la langue écrite ne fait même plus écran à l'évaluation de la seule compréhension) et celle d'un message oral. Pour ne pas alourdir les protocoles, l'évaluation de la maîtrise de l'ordre alphabétique, présent dès 1989, a disparu en 1999. Et l'évaluation du rapport graphie / phonie, présent en 1993, ne se retrouve, sous une autre forme, qu'en 1999.

II. Les tableaux de compétences en production écrite.

Ils ont suivi le cheminement des tableaux de compétences en lecture, mais les composantes évaluées sont toujours à peu près les mêmes d'un protocole à l'autre. Simplement, la réflexion sur l'acte d'écrire a fait passer certaines compétences du savoir écrire au savoir lire.

Ainsi, la reconstitution d'un «texte puzzle» est classée, en 1993, dans le champ intitulé approche des textes pour «l'objectif» savoir écrire : la réorganisation de phrases présentées en désordre (sens de texte puzzle) en un ensemble cohérent y est ressentie en effet comme la production d'un écrit. Plus tard, cet exercice passera dans le savoir lire : en effet, la production d'un écrit est dès lors considérée comme demandant l'élaboration totale et autonome des phrases qui le composent.

Même chose pour l'exercice qui demande d'insérer dans les blancs d'un texte les déterminants et les pronoms adéquats : placé en 1993 dans le savoir écrire / champ de la grammaire parce qu'il permet de constituer des phrases complètes, il passera finalement dans «la maîtrise des outils de la langue pour lire.»

Les compétences relevant de la maîtrise de la langue sont réparties, à partir de 1999, entre leur utilisation pour la lecture et pour l'écriture selon un parallélisme strict : reconnaître / produire les types et formes de phrases, construire le sens en utilisant les accords / faire les accords, avec , en plus, les outils de la langue spécifiques à la production écrite, l'orthographe et la ponctuation. En 2002, savoir écrire est un champ, dont dépendent deux capacités, maîtriser les outils de la langue et produire un texte. La première capacité donne la compétence «maîtriser les outils de la langue pour écrire», avec les quatre composantes évoquées ci-dessus. La seconde capacité donne deux compétences : «maîtriser les contraintes matérielles» et «créer et construire un texte».

L'évolution des tableaux de compétences montre, comme pour la lecture, une réflexion progressivement plus approfondie sur ce qui constitue spécifiquement la production d'un écrit. La constitution, à l'intérieur du champ de l'écriture, de deux capacités, dont l'une porte sur la maîtrise de savoirs et savoir-faire linguistiques tandis que l'autre porte sur l'élaboration d'un texte, présente un grand intérêt : elle permet de mesurer la différence entre la maîtrise de savoirs ponctuels et leur mobilisation dans une tâche globale et complexe comme la rédaction d'un récit.

Les protocoles d'entrée en sixième permettent-ils de suivre l'évolution des compétences des élèves ?

Les modifications introduites dans les exercices ne permettent pas toujours de mesurer rigoureusement l'évolution du niveau des élèves sur plusieurs années. Soit la compétence est momentanément ou durablement absente des protocoles, soit les supports changent pour un même type d'exercice, soit les codages et les items sont modifiés pour un exercice identique. Il arrive qu'un exercice apparemment semblable d'un protocole à l'autre adopte chaque fois une logique en réalité différente qui rend les scores difficilement comparables. Pour les professeurs et les formateurs, cette variété est un avantage, qui répond à la finalité affichée de l'évaluation diagnostique. Pour ceux qui voudraient suivre l'évolution des acquis des élèves, c'est un handicap ; on ne peut noter, dans le meilleur des cas, que des tendances ou éventuellement des constantes.

1° exemple. Faire les accords (maîtrise de la langue).

L'exercice demande aux élèves de récrire une phrase en faisant les accords qui s'imposent quand on modifie le nombre ou le genre de certains de ses termes. Il porte sur des savoirs considérés comme pouvant être acquis à la fin du cycle 3.

Supports. (les termes soulignés sont ceux qui servi à l'évaluation).

*1992, 1998 : *Chassons sous la pluie!*

«Le chat, trempe jusqu'aux os, s'est glissé dans la cuisine. Il est plus noir que jamais et paraît indifférent à tout. Majestueux, il s'approche de la cheminée.»

Récrire en commençant par Les chattes, ...

*1995,1997 : «L'été, l'écran (la toile) était dressé dans un champ (dans une prairie) offert par la commune. Les gens (les spectateurs) s'installaient sur des bancs (sur une chaise) anciens ou sur des sièges (sur un siège) improvisés. Alors, le ronronnement du projecteur forçait le silence et la magie (les sortilèges) commençait.»

Récrire en remplaçant certains mots par le terme entre parenthèses qui les suit.

*1999, 2000 : «La foule enthousiaste est impressionnée par les résultats inattendus du jeune patineur. Elle manifeste sa joie et applaudit à tout rompre.»

Récrire en commençant par Les spectateurs...

*2001 : «Ton fils est immortel, il ne mourra jamais. De plus, il deviendra un grand guerrier, plein d'audace et de bravoure.»

Récrire en commençant par Tes fils ...

*2002 : Le poulain était maintenant tout seul, perdu au milieu des marais, sans sa mère (...). Déjà le magnifique poulain ne pouvait plus entendre l'appel désespéré du garçon.

Récrire en commençant par Les chevaux...

Scores de réussite.

La première colonne indique les accords tels qu'ils sont désignés dans les items des protocoles (un item, ou question, peut porter sur une ou plusieurs formes du texte, ce qui est indiqué dans la dernière ligne).

Le score global n'est indiqué par les documents de la DEP qu'en 2002. A titre de comparaison, les scores globaux pour les années précédentes ont été calculés de la même manière : somme des résultats des codes de réussite (codes 1 et 2), divisée par le nombre d'items. Ils sont indiqués entre parenthèses.

Accord de...	1992 <u>Le chat</u>	1995 <u>l'écran</u>	1997 <u>l'écran</u>	1998 <u>le chat</u>	1999 <u>la foule</u>	2000 <u>la foule</u>	2001 <u>ton fils</u>	2002 <u>le poulain</u>
adj. en genre <u>ou</u> en nombre							72,8% immortels	
adj. en genre <u>et</u> en nombre	53,3% majestueuses 76,3% trempées	51,3% <u>+article</u> offerte, ancienne, improvisé	50,7% <u>+article</u> offerte, ancienne, improvisé	24,3% majestueuses, noires indifférentes	50,3% enthousiastes	50% enthousiastes		
adj. épithète								71,5% magnifiques
adj. Attribut (du sujet)	51% noires 59,5% indifférentes							43,2% seuls perdus
groupe nominal attribut.							30,5% grands guerriers, pleins	
participe passé en genre et en nombre.				70,8% trempées				
p.passé avec être	73,5% glissées	68,2% dressée	66,4% dressée		56,3% impressionnés	56% impressionnés		
du verbe et du sujet	78,% se sont 86 sont 65,6 70%	41,1% s'installaient commentaient	42,8% s'installaient commentaient	39,5% ensemble des 4 formes verbales	31,7% sont, manifestent, applaudissent	28,6% sont, manifestent, applaudissent	66% sont, mourront, deviendront	73,8% étaient pouvaient
Déterminants					49,1% leur	461 % leur	32,6% de (62,2 des)	65,9% leur
pronoms personnels					68,2% ils	65,8% ils	85,1% ils	
Score global	(78,1) 9formes 9 items	(53,5) 6formes 3 items	(53,3) 6formes 3 items	(44,8) 8formes 3 items	(51,1) 7formes 5 items	(49,2) 7formes 5 items	(69,8) 9formes 5 items	63,6% 6formes 4 items

Cet exercice semble rechercher des informations différentes selon les années. On peut évaluer l'aptitude à faire les accords nécessaires sur des éléments distincts de la phrase (adjectifs ou formes verbales) qu'on évalue séparément (1992). Mais, comme il s'agit de transformer une phrase, on peut aussi évaluer l'aptitude à traiter en même temps ces divers changements : d'où l'intérêt de regrouper pour un score unique toutes les formes de même nature (formes verbales en 1998) ou de même statut dans la phrase (groupe nominal attribut en 2001) ; d'où également l'importance de la transformation des déterminants et des pronoms personnels par exemple.

On traite également les mots selon leur nature (verbes, adjectifs ou participes passés, participes passés avec être) ou leur fonction (épithètes et attributs) mais pas toujours. Parfois, on réunit en un même score le pluriel de mots qui sont de même nature (des adjectifs) mais qui ont des fonctions différentes qu'on a distinguées dans d'autres protocoles. Ainsi, *noires* et *indifférentes*, traitées

comme adjectifs attribués du sujet en 1992, sont évaluées en 1998, en même temps que *majestueuses* comme adjectifs accordés en genre et en nombre ; du coup le score de 24,3% (1998) ne peut être interprété. De même, on sait que 86% des élèves accordent correctement le verbe être au pluriel comme *sont* ; il est difficile de le savoir ensuite, puisque cette forme est incluse dans l'évaluation de l'ensemble des verbes. Cette diversité, utile au professeur, enlève beaucoup de visibilité aux scores.

On peut évidemment, pour suivre une compétence, modifier soi-même la nature de l'information donnée par les scores. Par exemple, l'accord de l'adjectif épithète n'est ainsi désigné qu'en 2002 (*magnifiques*) ; mais les formes évaluées auparavant sous la rubrique «adjectif en genre et en nombre» sont des épithètes, parfois détachées (*trempees, majestueuses, offerte, ancienne, improvisé, enthousiastes*). On aurait donc la succession de scores suivante, qu'on pourrait chercher à interpréter : 76,3%, 53,3%, 51,3%, 50,7%, 50,3% et 50%, 71,5%. Mais ce n'est pas ainsi qu'est construite la succession des protocoles.

2° exemple. La reconstitution d'un texte puzzle (compréhension).

L'exercice demande aux élèves de remettre dans l'ordre logique et chronologique les phrases d'un récit qu'on leur propose en désordre (sens du terme de puzzle). Il y a en général six phrases.

Supports.

1989,1995,1997 : *Les deux chats*.

1992 ; *Cédric et le cadeau* (énoncé de problème).

1998, 2001 : *Les nains*.

1999-2000 : *Le faucon déniché*.

2002 : *Le petit indien et le lama*.

Scores de réussite et codages.

Années	code 1 réponse atten- due	code 2 accep- table	code 3 une omission	code 4 une in- version	code 8 erreur intéres- sante.	code 9 réponse erronée	code 0 non réponse	un en- chaîne- ment juste	deux en- chaîne- ments justes
1989	46,2%	21,2%	1,8%	12,3%		16,7%	1,8%		
1992	50,3%	9,5%				37,7%	2,5%		
1995	42%		3,6%	18,8%	13%	20,9%	1,7%		
1997	36,9%		4,9%	16,7%	16,4%	22,3%	2,8%		
1998	48,3%		2,3%	13,7%	5,1%	28,9%	1,7%	/	
2001	54,1%		1,6%	20%	7,0%	15,9%	1,4%	93,9%	
1999	52,4%		3,4%	8,3%		35,2%	0,7%	8,7%	73,3%
2000	49,5%		3,7%	7,4%		37,9%	1,5%	7,7%	75,5%
2002	26,6%		4,4%			67,9%	1,1%	14,4%	68,4%

Cet exercice demande toujours la même tâche, mais sur des supports qui peuvent se révéler de difficulté différente (comme en 2002 sur un texte nouveau, ce qui fait presque doubler le taux des réponses erronées). Avec des résultats très variables, mais proches quand l'exercice est repris d'une année sur l'autre (99-00), environ cinq élèves sur dix savent généralement reconstituer le texte, sans progrès ou baisse notables depuis 1989. Par ailleurs, les réponses erronées (codes 9) évoluent à l'intérieur d'une fourchette de 16% à presque 40% des réponses ; elles atteignent le même score de 1992 à 2000, sans que le taux de non réponses augmente parallèlement. Ce qui est significatif, en revanche, c'est le taux des réponses contenant une inversion, et le fait que plus de sept élèves sur dix sont capables de retrouver deux enchaînements exacts entre des phrases du texte, sous-compétence évaluée à partir de 1999.

Cependant la diversité des scores, utile pour le professeur, rend les résultats délicats à interpréter pour qui veut suivre la compétence. Les codes 2, qui signalent un ordre différent de l'ordre attendu mais possible, font monter les scores des premières années à 67,4% et 59,8% : de six à sept élèves sur dix réussiraient l'exercice. Il aurait été intéressant de proposer toujours aux élèves un texte qui se prête à deux ordres également acceptables, pour confirmer ou infirmer ce taux de réussite.

La difficulté relative des textes supports, les informations différentes recherchées par les questions

et ciblées par le codage fournissent donc des indications très utiles sur les réussites des élèves ; mais elles incitent à la prudence celui qui veut suivre dans la durée l'évolution du niveau des élèves pour la compétence visée.

Conclusion.

Même quand les exercices semblent proches, leurs modifications sont donc cependant assez sensibles pour que les scores des élèves ne permettent pas une réelle comparaison. Dans le chapitre consacré à l'évaluation diagnostique de cinquième, on verra que c'est parfois la logique même des protocoles qui diffère trop pour qu'on puisse envisager une comparaison entre les résultats sur une longue période.