

LE DISPOSITIF D'ÉVALUATION EXTERNE EN COMMUNAUTÉ FRANÇAISE DE BELGIQUE¹

Dominique Lafontaine²

Contexte

La Communauté française de Belgique possède un système éducatif fortement décentralisé. L'ensemble des établissements, qu'ils soient publics ou privés, bénéficient de subventions de l'Etat et doivent se soumettre à son contrôle, via notamment les services d'inspection. Mais nombre de décisions essentielles au fonctionnement de ce système sont de la responsabilité des « réseaux » ou pouvoirs organisateurs.

On distingue ainsi trois grands réseaux organisateurs d'enseignement : le réseau de la Communauté française de Belgique (dit « réseau Etat »), qui est directement organisé par les Ministères de l'Education; le réseau provincial et communal, organisé par les provinces et les communes, et le réseau subventionné libre, qui est organisé par des

instances privées, principalement catholiques.

La définition des programmes d'enseignement et des méthodes pédagogiques sont de la responsabilité des réseaux. Chaque réseau, et, à l'intérieur des réseaux, chacun des pouvoirs organisateurs fédérés dans ce réseau (par exemple une commune) peut rédiger son propre programme, qui doit cependant être approuvé par l'autorité centrale (la Communauté française de Belgique).

Chaque réseau ou pouvoir organisateur est aussi libre d'organiser des évaluations externes à l'intention des écoles et élèves relevant de sa responsabilité. De nombreuses initiatives se sont ainsi multipliées au fil du temps. En revanche, dans un tel contexte, on saisit aisément combien peut apparaître délicate sur le plan institutionnel, la mise en place d'un dispositif d'évaluation

¹ Texte paru en 2001 dans la revue *Le Point sur la Recherche en Education*, 19, 37-56.

² Responsable scientifique de l'évaluation externe dans le domaine de la langue maternelle depuis 1994-95.

externe qui concernerait l'ensemble des élèves de la Communauté française de Belgique tous réseaux confondus.

Initiative

Au début des années 90, le *Rapport sur l'état du système éducatif* rédigé par des experts de l'Ocdé (Ocdé, 1991) met en avant deux caractéristiques inquiétantes du système éducatif en Communauté française de Belgique : l'importance des taux d'échecs et la disparité de réussite en fonction de l'établissement fréquenté.

Ce rapport fait grand bruit à l'époque et serait notamment à l'origine de la mise en place, en 1994-95, du dispositif d'évaluation externe. Par ailleurs, l'instauration de celui-ci ne peut se penser qu'en articulation étroite avec une autre initiative d'importance : la définition et mise en place de « socles de compétences » concernant tous les élèves. Des socles de compétences provisoires, élaborés en commun par des représentants des différents réseaux et composantes du système éducatif, voient le jour en 1994. La version définitive de ces socles sera arrêtée en 1999, en même temps que des documents définissant les compétences terminales (fin de l'enseignement secondaire) pour l'ensemble des disciplines concernées.

Il s'agit là d'un événement sans précédent dans l'histoire du système éducatif de la Belgique : **pour la première fois, la Communauté française de Belgique se donne un ensemble de références communes, transcendant les programmes particuliers**, et il devient donc pleinement légitime, et plus aisé, sur le plan pratique, de mettre en place un dispositif destiné à vérifier dans quelle mesure les élèves atteignent les niveaux de compétences recommandés au moment attendu.

L'initiative de la mise en place d'un dispositif d'évaluations externes revient à un haut responsable (le secrétaire général Jean Magy) de l'Administration générale de l'Enseignement et de la Recherche scientifique du Ministère de la Communauté française de Belgique (Magy, 1998), fortement soutenu dans son initiative par une partie de l'inspection primaire. Même si l'opération a le soutien du Ministère de l'Education de l'époque, elle ne fait cependant l'objet d'aucune disposition législative.

Par la suite, une cellule responsable de ces évaluations est instaurée au sein de l'administration (Service général, des Affaires générales, de la Recherche en Education et du Pilotage interréseaux). Il ne s'agit toutefois pas, loin s'en faut, d'une « direction » à la française (comme celle de l'ancienne DEP ou de l'actuelle DPD) chargée de réaliser

les évaluations externes. La responsabilité de cette cellule est d'abord d'ordre administratif : elle définit le calendrier, signe des conventions avec des organismes de recherche extérieurs, coordonne les activités des différents partenaires, apporte un soutien logistique aux évaluations...

La réalisation des évaluations externes, au delà de cette cellule, associe différents partenaires, dans un dispositif original, qui a sans doute peu d'équivalents dans le monde, dans la mesure où il découle des caractéristiques particulières du système éducatif en Communauté française de Belgique.

Partenaires et modalités de fonctionnement

Pour chaque opération d'évaluation externe, un groupe de travail est mis sur place, dont les différents membres vont, à des titres divers, contribuer à l'élaboration, à la réalisation et au suivi de l'évaluation. Dans ce groupe, on retrouve des responsables de l'administration, des inspecteurs, des représentants des réseaux (qui sont le plus souvent des enseignants en fonction ou détachés pour mission) et des chercheurs¹ recrutés sur des conventions extérieures.

¹ De l'ordre d'un chercheur à mi-temps par an et par discipline concernée (français ou mathématique jusqu'à présent), ceci pour donner une idée de l'ampleur des moyens investis. Le fait que les chercheurs soient

Les directions que prennent les évaluations externes sont la résultante des discussions de ce groupe, qui en général, travaille à la recherche de consensus circonstanciels et provisoires (on reviendra sur cet aspect). La première phase (définition des compétences à évaluer, choix du matériel et des modalités de questionnement) repose sur une véritable collaboration. Il s'agit de se mettre d'accord sur le sens et les grandes lignes de l'évaluation. Dans la seconde phase (réalisation), les chercheurs jouent un rôle technique plus prépondérant : choix et rédaction des items, prétest, sélection de l'échantillon, recueil, encodage et analyse des données, rédaction du document de résultats. L'administration, de son côté, assure le soutien logistique de l'opération : reproduction et envoi des documents aux écoles, contacts téléphoniques avec les établissements, réponses aux nombreuses questions qui ne manquent pas de se poser. Lors de la troisième phase (pistes didactiques et suivi), les différents partenaires interagissent à nouveau de façon étroite autour du document de pistes didactiques. Une fois les documents envoyés aux écoles, les inspecteurs et les réseaux, dans la mesure de

recrutés sur contrat extérieur ne signifie pas que les chercheurs soient chaque année différents. Dans les faits, une certaine continuité s'est installée. L'auteur du présent article a participé depuis 1994 aux évaluations externes dans le domaine de la langue maternelle.

leurs possibilités¹, assurent un suivi sur le terrain. Ce dispositif, on le voit, associe d'une façon souple et singulière, des partenaires « non institutionnels » qui contribuent, pour une période plus ou moins limitée (cela dépend des partenaires) à la réalisation des évaluations externes.

Population et niveaux scolaires concernés

L'évaluation externe touche chaque année tous les établissements et tous les élèves d'un niveau scolaire déterminé (d'une année d'étude). Compte tenu de la structure du système éducatif (voir ci-dessus), il n'est pas envisageable de contraindre les établissements de se soumettre à l'évaluation externe. Quelques établissements refusent donc parfois de participer, surtout au niveau de l'enseignement secondaire². Les chercheurs universitaires, pour des raisons de coût et de temps, ne traitent que les résultats d'un échantillon représentatif.

Dans un premier temps, les évaluations externes ont porté en priorité sur les niveaux scolaires de l'enseignement de base (6-14 ans) couvert par le document de référence « Socles de compétence ». Des évaluations ont ainsi été réalisées pour les élèves de 8 ans, 10 et 12 ans. Le dispositif s'est ensuite étendu aux élèves de 14 et de 16 ans. Les références en termes d'âge ne sont bien sûr que des indications destinées à montrer que l'on procède par cycles de deux ans. Tous les élèves de l'année sont testés, qu'ils soient à l'heure, en avance ou en retard scolaire. On procède à une évaluation par an, par rotation des niveaux concernés.

Dans l'enseignement primaire, tous les élèves de l'enseignement ordinaire sont concernés; l'enseignement spécial n'est pas envisagé. Dans l'enseignement secondaire, le souci d'évaluer tous les élèves sur une base commune trouve vite ses limites, dans la mesure où, dès la première secondaire, les élèves se trouvent répartis dans des filières de niveau extrêmement différents.

La question « faut-il ou non évaluer tous les élèves ? » a régulièrement alimenté les discussions du groupe de travail. Pour le niveau 1^{re} et 3^e secondaires, les socles de compétences s'appliquent en principe à tous les élèves (avec des nuances subtiles toutefois, qu'il serait trop long de détailler) et il serait donc normal de les évaluer tous. En faisant cela, on sait que l'on place

¹ Etant donné leur faible nombre, il est beaucoup plus difficile aux inspecteurs du secondaire d'assurer un tel suivi. Certains inspecteurs de l'enseignement primaire ont en revanche consacré beaucoup de temps et d'énergie à cette tâche.

² Ces rares établissements réfractaires ne relèvent pas d'un réseau particulier (on en trouve dans les différents réseaux). Il s'agit en général d'établissements « élitistes », qui, soucieux de défendre leur réputation d'exigence, mettent en avant le fait qu'ils ont mieux à faire que de participer à une évaluation.

nécessairement les élèves fréquentant les filières les plus faibles (1^{re} B ou 3^e P) dans une situation d'échec difficilement défendable sur le plan pédagogique et humain. Pour ces niveaux, on a donc tranché la question par un compromis : deux versions différentes du test ont été réalisées. Au niveau de la 5^e secondaire, le choix a été fait d'évaluer uniquement les élèves de la filière transition (générale et technique) qui donne accès à l'enseignement supérieur sans année complémentaire. Les référentiels pour les filières de transition et de qualification ne sont pas identiques (deux tests différents auraient donc été nécessaires) et les moyens étaient insuffisants pour réaliser deux évaluations distinctes.

Moment et disciplines concernées

Les évaluations sont toujours réalisées en début d'année scolaire, en conformité avec les objectifs qui leur sont assignés (voir ci-après). Le but est de permettre à l'enseignant de réaliser un diagnostic des compétences acquises par ses (nouveaux) élèves et d'orienter son action éducative en fonction notamment de ces résultats. De plus, en effectuant l'évaluation en début d'année, on évite que l'enseignant se sente lui-même évalué.

De 1994 à 1998, deux disciplines ont été concernées : le français et les mathématiques. En français,

l'évaluation s'est essentiellement portée sur les compétences en lecture. En 1999-2000, l'évaluation n'envisage que la langue maternelle, mais s'ouvre à la production d'écrit. A cet égard, il faut préciser que jusqu'en 1999, les tests ont toujours été corrigés par les enseignants titulaires des classes évaluées, en fonction de critères standardisés. Motivée par des raisons de coût et de temps, cette disposition retentit évidemment sur le choix des compétences et des modalités de questionnement. Une tâche supplémentaire trop lourde ne peut être demandée aux enseignants et, pour des productions ouvertes relativement complexes (questions ouvertes en lecture, production d'écrits, résolution de problèmes), se poserait la question de la fidélité entre les correcteurs. C'est pourquoi jusqu'à présent, la plupart des questions posées ont été des questions à choix multiple ou des questions ouvertes à réponse « objective » (peu sujettes à interprétation); c'est aussi pourquoi on n'avait guère jusqu'ici évalué l'expression écrite¹, si ce n'est d'une façon limitée.

Objectifs : le sens de la mesure

Au départ, lors de sa création en 1994, ce dispositif répond, dans la

¹ De nouvelles modalités de correction et d'échantillonnage des copies seront en vigueur pour l'évaluation de l'écrit en 2000-2001.

tête de son initiateur, à un double souci :

- a) sortir les enseignants de leur isolement, en leur fournissant un référentiel (les résultats d'un échantillon représentatif) auquel comparer les performances de leur classe. L'outil se veut formatif et diagnostique. Il est passé en début d'année;
- b) réduire les disparités de rendement entre écoles, dont l'importance a été soulignée par les recherches internationales (IEA, OCDE) et par le *Rapport sur l'état du système éducatif* des experts de l'Ocdé (Ocdé, 1991).

Dans les faits, le dispositif, fortement investi par les services d'inspection du primaire, est surtout utilisé, lors des deux premières années, comme un « levier didactique », i.e. comme un instrument permettant aux inspecteurs de faire passer leur message pédagogique (test « modèle » montrant aux enseignants comment « bien évaluer », en se centrant sur les compétences essentielles, accompagnement de l'évaluation par un document de pistes didactiques exploité sur le terrain par les inspecteurs). Ainsi, si l'évaluation se veut officiellement « au service des enseignants », elle est tout autant au service de l'inspection et mise implicitement sur l'effet en retour (*backwash effect*) de l'évaluation sur les pratiques (Anderson & Wall, 1996).

Progressivement, avec le passage de l'évaluation dans le secondaire en 1996-97, et le départ à la retraite de l'initiateur du dispositif, on assiste à un recul du rôle de l'inspection primaire et le côté « levier didactique » est moins à l'avant-plan. Le débat et les tensions entre les partisans d'une évaluation « modèle » (une évaluation en conformité avec la tonalité des recommandations pédagogiques du moment) et les exigences techniques d'une évaluation à large échelle à caractère diagnostique restent cependant d'actualité et resurgissent d'une manière plus ou moins vive dès qu'un nouveau groupe de travail se met en place.

La permanence de ce débat illustre bien le statut particulier de ces évaluations. **Un dispositif institutionnel existe, mais c'est bien aux acteurs qui le mettent en oeuvre qu'est dévolu le rôle de lui donner son sens et ses orientations précises**¹. Jusqu'à présent, ni les ministères de l'éducation ni l'administration n'ont défini d'attentes exactes en la matière (ni écrit officiel, ni cahier des charges n'existe à ce propos); ceux-ci s'en remettent aux groupes de travail pour opérationnaliser l'épineuse question de « comment savoir si les

¹ Aux membres du groupe de travail qui interrogent les responsables de l'Administration sur ce que veut le Ministère de l'Éducation, il est classiquement répondu que « le Ministère ne veut rien d'autre que ce que veut ce groupe de travail ».

élèves atteignent à temps les compétences attendues ? ».

Un débat récurrent est de déterminer dans quelle mesure il faut s'en tenir à évaluer la mobilisation des compétences face à des tâches relativement complexes, ambitieuses, intégrées, fonctionnelles (le vocabulaire peut varier, mais on saisit la tendance générale) ou s'il faut coupler ceci d'une évaluation plus ciblée, portant sur des compétences plus spécifiques. En général, les chercheurs universitaires sont plutôt en faveur de cette seconde option, car ils l'estiment nécessaire pour rencontrer la fonction diagnostique assignée aux évaluations externes. Il est clair que l'évaluation de micro-compétences de façon ponctuelle (hors d'une situation ou d'un contexte jugé « fonctionnel ») heurte la sensibilité de certains partenaires, car elle va à l'encontre du message pédagogique dominant et leur fait redouter que l'inclusion de telles questions dans les évaluations externes leur confère une légitimité qui pourrait miner les efforts de réforme et d'innovation entrepris dans les classes. Ces craintes sont sans doute légitimes. Il n'empêche qu'il se révèle pratiquement impossible de couler dans un même moule les formes multiples que peut ou doit prendre l'évaluation en classe par l'enseignant, qui peut être formative, sommative, certificative, continue, quantitative et/ou qualitative et une évaluation externe où toute

l'information souhaitée doit être recueillie d'un seul coup auprès de 50 000 élèves environ, avec une nécessité de standardisation et un droit à l'erreur limité.

Matériel d'évaluation et documents d'accompagnement

Typiquement, une opération d'évaluation externe annuelle suppose l'élaboration des documents suivants :

- une **épreuve cognitive**;
- un **carnet d'informations et de consignes à l'intention de l'enseignant** où sont notamment exposés les objectifs de l'évaluation, la composition de l'épreuve (types de texte, grands domaines évalués), les compétences évaluées par épreuve et par item, les consignes de passation, les consignes et les critères de correction;
- un **questionnaire à l'élève** où sont brièvement recueillies des informations relatives à l'élève (sexe, âge, langue parlée à la maison), à son parcours scolaire (redoublement au cours de la scolarité antérieure, options suivies, changement d'école, projets d'étude...), à des pratiques ou habitudes dans le domaine de la lecture (attitudes envers la lecture, nombre de livres à la maison, emprunt de livres en bibliothèque...);

- un **questionnaire à l'enseignant titulaire** où sont recueillies des informations relatives aux caractéristiques de l'établissement (situation géographique, taille de l'établissement, école en zone d'éducation prioritaire ou en discrimination positive), aux caractéristiques de la classe (nombre d'élèves, pourcentage d'élèves en retard scolaire, pourcentage d'élèves ne parlant pas habituellement le français à la maison, pourcentage d'élèves issus de milieu favorisé, pourcentage d'élèves vivant en situation de précarité). L'avis de l'enseignant sur le degré d'adaptation des épreuves et des items à l'année d'étude concernée est également sollicité. Certaines informations sur les pratiques pédagogiques des enseignants sont parfois demandées;
- une **grille de correction** et une liste de **critères de correction** standardisés;
- un document présentant et commentant **les résultats**, qui suit le plus rapidement possible la passation des épreuves;
- un document de **pistes didactiques**, qui suggère des possibilités d'action didactique en relation avec les compétences évaluées. Sur ce plan, l'exercice relève de la haute virtuosité institutionnelle, car il faut savoir que les « méthodes pédagogiques » relèvent de la prérogative des réseaux. Il faut donc se montrer concret, tout en restant suffisam-

ment vague pour ne pas empiéter sur un territoire appartenant au réseau;

- un questionnaire qui fait le **bilan** de l'ensemble de l'opération, à destination de l'enseignant.

Textes et compétences évaluées en français

Comme cela a été dit précédemment, l'évaluation, s'est, à l'une ou l'autre exception près, centrée sur la lecture. Il n'y a jamais eu, jusqu'à présent, d'évaluation de l'oral, ni de connaissances formelles (grammaire, conjugaison, orthographe, lexique...) qui ne soient pas couplées à un support textuel. On a parfois évalué (en 3e et en 5e primaires notamment) l'expression écrite d'une façon circonscrite (une seule production : écrire une carte postale et continuer un récit amorcé).

Les compétences en lecture ont toujours été évaluées au départ de textes entiers (jamais de phrases isolées), de genres diversifiés (narratifs, informatifs, argumentatifs, injonctifs). Un large éventail de compétences, différent selon les niveaux, est envisagé, défini par rapport aux référentiels interréseaux que constituent les socles de compétence ou les compétences terminales.

Ce sur quoi va porter l'évaluation et ses modalités sont ainsi déterminées en référence aux instructions

officielles et aux recommandations des différentes autorités pédagogiques que représentent l'inspection, d'une part, les réseaux, d'autre part. L'évaluation est clairement centrée sur les grands objectifs prioritaires ou sur les compétences essentielles (lire, écrire...) et ne laisse aucune place à l'évaluation de savoirs décontextualisés à l'acquisition desquels nombre d'enseignants consacrent pourtant l'essentiel de leurs énergies. **Les évaluations externes véhiculent donc un message implicite sur ce qui constitue l'essentiel. Il n'est toutefois pas évident que le divorce entre ces évaluations et les pratiques habituelles de certaines classes soit clairement interprété dans ce sens par les enseignants.**

Traitement des données; types de résultats fournis aux enseignants

Les textes et items sont toujours prétestés avant leur utilisation pour le test. De même, après le test, des contrôles sont effectués pour vérifier la qualité psychométrique des items.

Chaque enseignant corrige et conserve les carnets de sa propre classe. Les enseignants de l'échantillon renvoient les grilles de correction et les questionnaires aux universités qui traitent les données.

Des pourcentages moyens de réussite par item et pour l'ensemble de l'épreuve sont calculés. Dans

certain cas, des sous-scores peuvent aussi être calculés, si les conditions techniques de cohérence interne sont remplies (calcul de l'alpha de Cronbach). On a ainsi calculé, pour certaines des évaluations, un sous-score pour les questions littérales et un sous-score pour les questions inférentielles.

Lorsque l'enseignant reçoit ces scores moyens, il peut comparer la moyenne obtenue dans sa classe avec celle obtenue par l'échantillon représentatif. Cette information est intéressante, car elle permet à celui-ci de relativiser les résultats de sa classe. Des recherches (Grisay, 1984) ont en effet montré qu'en raison de leur isolement, les enseignants ont tendance à adapter leur enseignement et leur degré d'exigence en fonction du niveau de leur classe. Ainsi certains enseignants travaillant en milieu très défavorisé pourraient avoir tendance à revoir leurs exigences à la baisse. A l'inverse, d'autres enseignants qui se trouvent quotidiennement face à un public d'élèves plus « sélectionné » ou plus privilégié, pourrait avoir tendance à placer la barre très haut et à déclarer en échec des élèves dont le seul problème est qu'ils sont en dessous d'une moyenne très élevée. Confronter les enseignants avec les résultats d'une évaluation standardisée reflétant le niveau d'exigence « officiel » peut, dans cette perspective, se révéler intéressant. **En entamant l'isolement du maître face à sa classe, on**

peut espérer « modérer » des « débordements » (dépassements de programme, surestimation de l'hétérogénéité de la classe, exigences trop ou trop peu élevées...) bien compréhensibles sur le plan individuel, mais aux effets combien néfastes à l'échelle du système.

Afin de nuancer cette comparaison, on s'est aussi efforcé, au-delà de ce pourcentage moyen, de fournir aux enseignants un tableau un peu plus nuancé. En fonction du pourcentage obtenu par sa classe, l'enseignant peut déterminer quelle position il occupe en termes de percentiles (voir annexe 1).

Pour compléter le tableau, on fournit par ailleurs des pourcentages en fonction des grandes caractéristiques de la composition des classes (pourcentage d'élèves en retard, d'allophones, d'élèves issus de milieu privilégié ou d'élèves provenant de milieux très défavorisés...). L'enseignant peut ainsi comparer les résultats de sa classe à ceux de classes présentant, dans leur composition socio-pédagogique, des caractéristiques approchantes (voir annexe 2).

Au travers de ces différentes possibilités de comparaison, on espère que l'enseignant se trouve en situation de relativiser la valeur des performances observées dans sa

classe, à la lumière d'informations contextuelles dont il est le seul à disposer. Si les grandes caractéristiques de la composition de la classe sont en effet connues, les nuances du contexte particulier à chaque classe échappent au champ d'investigation des évaluations externes. Seul l'enseignant ou le chef d'établissement connaît la pédagogie mise en oeuvre par l'école ou par les enseignants des années antérieures à celle évaluée, ou sait qu'un tiers de la classe est composé d'élèves en échec venant d'un autre établissement ou sait encore le degré de soutien qui a été apporté lors de la passation de l'épreuve externe...(nul n'est censé aider, mais donner un petit coup de pouce paraît parfois si naturel...). **Les résultats au niveau d'un établissement ou d'une classe doivent toujours être envisagés avec toute la prudence et le recul nécessaire; bien des facteurs dont on ne soupçonnerait même pas l'existence peuvent en effet expliquer les résultats d'une classe particulière.** Si les menus aléas qui peuvent affecter l'exactitude de la mesure sont sans effets majeurs et sans conséquences pour les valeurs moyennes observées, il n'en va pas de même pour les mesures individuelles (d'élèves, de classes, d'établissements), qu'il faudra toujours envisager avec circonspection, en tentant de recouper les résultats par d'autres sources d'information.

Publicité des résultats

Il découle des considérations qui précèdent que le degré de publicité des résultats est strictement limité. Au niveau central, les résultats sont traités par les universités dans le respect des règles de confidentialité en vigueur pour ce type d'étude (traitement anonyme). Aucun palmarès, aucun classement n'est publié, aucune décision d'octroi de moyens supplémentaires n'est prise sur la base des résultats aux évaluations externes. Les résultats par établissement ne sont communiqués ni aux Ministères de l'Education, ni à leur administration. La garantie d'anonymat est totale à ce niveau. Comme on l'a dit plus haut, l'absence de toute garantie quant aux conditions de passation de ces tests (c'est l'enseignant titulaire qui les administre à sa propre classe) rendrait d'ailleurs, au-delà des principes, toute conclusion relative à un établissement ou une classe particulière fort peu valide. A l'inverse, la garantie offerte de non-publicité des résultats est sans doute le meilleur rempart contre la tentation d'améliorer un peu les résultats, en donnant le cas échéant un petit coup de pouce aux élèves.

A un niveau plus local, on l'a dit, les enseignants conservent les résultats de leur classe et rien n'empêche des autorités telles que les directeurs ou les inspecteurs de s'arranger pour obtenir les résultats des classes relevant de leur pouvoir d'intervention, en vue d'actions de

micro-pilotage dont on espère qu'elles s'appliquent avec la prudence et le doigté nécessaires.

Des indicateurs destinés au pilotage du système éducatif

Même si les évaluations externes se sont vu assigner comme destinataire principal l'enseignant dans sa classe (en misant ainsi implicitement sur leurs capacités d'auto-régulation au départ d'informations écrites), on ne peut toutefois mettre de côté le fait que celles-ci sont aussi destinées à informer un certain nombre de décideurs de l'état du système éducatif. On soulignera cependant que cette mission apparaît en Communauté française de Belgique comme relativement secondaire tant dans le chef de l'Etat que dans celui des citoyens. Aucune politique d'évaluation claire et structurée n'est en effet impulsée par les décideurs¹.

¹ En 1997, voit le jour un texte législatif d'importance pour le système éducatif de la Communauté française de Belgique : le « Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre ». Ce décret, qui coule dans un texte de loi la charpente du système éducatif, institue, dans un chapitre intitulé "du pilotage", trois types de commissions : une commission chargée de définir les socles de compétences, une commission des programmes et une commission responsable des outils d'évaluation. La composition des commissions est décrite avec une abondance de détails (il faut respecter tous les savants équilibres politico-administratifs desquels notre système éducatif se nourrit et dans lesquels il perd son sens).

En revanche, aucune relation n'est établie entre le dispositif de pilotage préexistant et ces nouvelles dispositions législatives. Le dispositif

Parallèlement, il n'existe pas de demande citoyenne forte pour que « l'école rende des comptes ».

Le traitement des données de l'évaluation externe pour en extraire des informations pertinentes sur l'état du système éducatif est donc laissé à l'appréciation des chercheurs responsables. Pour ce qui concerne l'évaluation de la langue maternelle, la chercheuse responsable a notamment insisté sur les aspects suivants :

- recueillir des données via les élèves et les enseignants sur certaines caractéristiques du milieu familial, de l'établissement, des pratiques pédagogiques, afin de replacer les résultats dans leur contexte;
- calculer les disparités de réussite en fonction des établissements, ou de la filière fréquentée;
- reprendre une partie des items d'une évaluation dans la suivante (mesure d'ancrage), afin de mesurer les progrès accomplis en deux ans;
- reprendre certains items des enquêtes internationales (IEA

de pilotage se trouve ainsi en quelque sorte « oublié » dans le texte légal, mais continue néanmoins à fonctionner et à recevoir des crédits du même Ministre de l'Education. Le texte de loi, qui institue par ailleurs plusieurs réformes d'importance, ne prévoit ni évaluation des performances des élèves, ni évaluation des effets des réformes. Il envisage donc un pilotage sans instruments, « à vue », s'en remettant sans doute implicitement à des actions de micro-pilotage orchestrées par les services d'inspection. Mais le texte de loi n'en dit rien (Lafontaine, 1999).

Reading Literacy, Timss), afin de déterminer si la situation a évolué dans le temps;

- utiliser les modèles de la réponse à l'item (IRT) pour déterminer, au-delà des pourcentages moyens, quelles proportions d'élèves atteignent tel ou tel palier ou niveau de compétences.

Le principe de ces modèles IRT (Laveault et Grégoire, 1997; Demeuse, 2000) est qu'ils permettent d'exprimer, sur une même échelle, le niveau de difficulté d'un item et le niveau de compétence d'un individu. Un individu qui obtient par exemple un score d'une valeur de 2 sur une telle échelle a une probabilité supérieure à un seuil fixé (on a utilisé 80 % pour les évaluations externes) de réussir les items d'une difficulté inférieure à 2 et une probabilité inférieure au seuil fixé (80 %) de réussir les items d'une difficulté supérieure à 2.

Connaissant le score d'un individu sur l'échelle, on peut estimer et décrire son niveau de compétences en mettant son score en rapport avec le groupe d'items qui ont des niveaux de difficulté se situant dans la même zone que celui-ci. Pour un individu qui obtient un score de 2, on pourra par exemple dire qu'il est capable, à propos d'un texte narratif bref, de réaliser des inférences assez simples à propos des intentions des personnages, mais qu'il n'est pas capable, n'ayant pas obtenu un score supérieur, de le faire pour un texte narratif long ou encore pour un texte

narratif bref, si les inférences à effectuer sont complexes. A l'échelle du système, on peut aussi caractériser chacun des paliers de compétences (rudimentaire, élémentaire, suffisant, satisfaisant, excellent ou avancé...) et déterminer quelles proportions d'élèves se situent à chacun des paliers ainsi défini.

On trouvera en annexe 3 un exemple de ce qu'a donné ce type d'analyse appliqué aux données de 3e année primaire, recueillies en 1995-96. Les indications qui se dégagent de ce type d'analyse sont précieuses pour tenter de déterminer d'une façon autre que purement intuitive dans quelle mesure les élèves maîtrisent les compétences attendues à l'issue d'un cycle d'enseignement.

Comment sont reçues les évaluations externes dans le monde enseignant ?

Au terme des trois premières opérations d'évaluation externe, un questionnaire bilan a été envoyé à tous les enseignants concernés, afin de recueillir leur avis sur l'opération, de mieux savoir quels aspects les ont davantage intéressés et quels usages ils ont pu faire des différents documents mis à leur disposition. Environ la moitié des enseignants ont renvoyé leur questionnaire complété.

Les résultats qui se dégagent de ces bilans sont plus qu'encourageants; ils sont convergents quelle que soit

l'année concernée. Voici, à titre d'illustration, les résultats enregistrés après la première évaluation menée en cinquième année primaire. Ce sont les documents les plus pratiques (l'épreuve elle-même, les pistes didactiques) qui recueillent le plus franc succès. Les carnets de tests sont les plus appréciés : plus de 90 % des enseignants les ont trouvés intéressants, 33 % en ont fait repasser des parties en cours d'année, 68 % s'en sont inspirés pour construire d'autres épreuves, et 65 % déclaraient avoir l'intention de la faire passer à leurs élèves à la rentrée scolaire suivante dans l'hypothèse où ils auraient encore une classe de cinquième.

Les pistes didactiques ont également rencontré un intérêt massif : 89 % des enseignants les ont trouvées intéressantes. Le document présentant les résultats de l'échantillon est également apprécié, puisque 75 % des enseignants le jugent intéressant. Ils ne sont toutefois que la moitié à avoir reporté les résultats de leur classe dans les tableaux prévus à cet effet, pour pouvoir comparer en vis-à-vis les résultats de leur classe à ceux de l'échantillon. Et 61 % reconnaissent que c'est important de pouvoir ainsi relativiser ses résultats, alors que 90 % se montrent sensibles à la fonction diagnostique de l'épreuve (pouvoir situer les points forts et les difficultés des élèves).

Les tests ont fait l'objet de nombreux échanges entre

enseignants et avec le directeur (environ la moitié des répondants), et, dans une moindre mesure, avec les inspecteurs (19 %). Les élèves apprécient les tests. La réaction a été positive chez 70 % d'entre eux. Enfin, 71 % des enseignants souhaitent voir des opérations de ce genre se répéter.

Même s'il faut se garder d'un optimisme béat, il semble donc qu'une stratégie à large échelle et à diffusion pyramidale, s'appuyant sur des relais solidement établis (inspection, réseaux) se solde par **un impact non négligeable. Celle-ci suscite en tout cas l'intérêt et le débat et semble jouer en partie le rôle de levier pédagogique que certains lui assignent.**

Si l'on prend en compte le fait que l'action s'inscrit dans la durée, on peut espérer créer chez les enseignants un questionnement ou un déséquilibre (au sens piagétien) qui rendrait ceux-ci plus réceptifs à des actions de formation continuée organisées sur le terrain.

En conclusion

Le dispositif d'évaluation externe mis en place en Communauté française de Belgique à partir de

1994 présente plusieurs caractéristiques qui sont le reflet des singularités structurelles et organisationnelles du système éducatif :

- il n'est pas explicitement, et encore moins légalement, investi d'une mission de pilotage. Sa principale cible est l'enseignant dans sa classe, et non les décideurs en matière d'éducation;
- il est faiblement « institutionnalisé », dès lors qu'il est mis en oeuvre par un groupe de travail rassemblant, pour un temps limité, des partenaires issus de différents milieux (administration, inspection, réseaux, universités).

A notre connaissance, aucun système éducatif ne s'est doté d'un dispositif aux caractéristiques semblables et il est souvent malaisé, lors de contacts internationaux, d'en faire saisir le sens et reconnaître la portée. Il est encore plus malaisé de convaincre les partenaires qu'en dépit de ses contours flous et de ses bases mal assurées (tout repose en quelque sorte sur des personnes), ce dispositif est parvenu à dégager des analyses intéressantes (Lafontaine, 1997; Lafontaine, 2000), même si - c'est notre regret personnel - la base de données constituée à l'occasion des évaluations externes reste, faute de moyens de recherche, considérablement sous-exploitée.

Pour en savoir plus...

Où se procurer les documents (épreuves, résultats, pistes didactiques, socles, compétences terminales...)?

Tous les documents peuvent être obtenus en s'adressant à :

Fanny CONSTANT
Administration générale de
l'Enseignement et de la Recherche
scientifique du Ministère de la
Communauté française, Service
général, des Affaires générales, de la
Recherche en Education et du
Pilotage interréseaux
rue Belliard, 9-13
1040 Bruxelles
Tél : 02/213 59 32
Fax : 02/213 59 91

Articles et livres publiés à propos des évaluations externes

Herphelin, M. (1996). Bilan de la première opération d'évaluation externe menée en Communauté française. *Pilotinfo*, 12, 1-6.

Lafontaine, D. (1997). Le niveau en lecture à l'entrée du secondaire en Communauté française de Belgique: des compétences fragiles et si diverses... *Enjeux*, 41-42, 129-161.

Lafontaine, D. (1999). Les évaluations externes en Communauté française de Belgique : quel sens donner à la mesure ? *Actes du XXIIIe colloque de l'Admee*. Dijon.

Lafontaine, D. & Schillings, P. (1999). Mesurer les progrès accomplis grâce au modèle de la réponse à l'item : l'évolution des compétences en lecture à 14-15 ans en Communauté française de Belgique entre 1991 et 1998. *Scientia Paedagogica Experimentalis*, vol 36, 2, 267-287.

Lafontaine, D. (2000). La lecture, un ouvrage féminin... Enquête sur les pratiques de lecture auprès des jeunes de 14-15 ans en Communauté française de Belgique. *Enjeux*, 47/48, 161-176.

Magy, J. (1998). *Oser la qualité dans l'enseignement en Communauté française de Belgique*. Namur : Erasme.

Références citées dans l'article

Anderson, J.-C. & Wall, D. (1993). Does Washback exist ? *Applied linguistics*, vol. 14, 2, 115-129.

Demeuse, M. (2000). *Les échelles de mesure : Thurstone, Likert, Guttman et le modèle de Rasch*. Liège : Service de Pédagogie expérimentale de l'Université, Série "Notes techniques" - 2000/2.

Grisay, A. (1984). Les mirages de l'évaluation scolaire (I). Rendement en français, notes et échecs à l'école primaire. *Revue de la Direction générale de l'Organisation des Etudes*, 19, 5, mai, 29-42.

Grisay, A. (1984). Les mirages de l'évaluation scolaire (II). Rendement en français, notes et échecs à l'école primaire. *Revue de la Direction générale de l'Organisation des Etudes*, 19, 6, mai, 9-23.

Laveault, D. et Gregoire, J. (1997). Les modèles de la réponse à l'item. In *Introduction aux théories des tests en sciences humaines*. Bruxelles : De Boeck.

Ministère de la Communauté française (1999). *Socles de compétences*. Bruxelles : Administration générale de l'Enseignement et de la Recherche scientifique, Direction de la Recherche en Education et du Pilotage interréseaux, Commission commune de Pilotage de l'Enseignement obligatoire.

Ministère de la Communauté française (2000). *Compétences terminales et savoirs requis en français. Humanités générales et technologiques*. Bruxelles : Administration générale de l'Enseignement et de la Recherche scientifique. Direction de la Recherche en éducation et du Pilotage interréseaux.

Ocdé (1991). *Les systèmes éducatifs en Belgique : similitudes et divergences*. Bruxelles : Ministère de l'Education, de la Recherche et de la Formation.

Annexes

Annexe 1

Votre classe a obtenu un score de 47,1 au test. A la lecture du tableau précédent vous pouvez constater que 43,7 % des classes se situent dans cet intervalle de scores. La lecture du

tableau ci-dessous vous apprend plus précisément que ce résultat situe votre classe entre la 40e et la 50e position.

Si votre classe a obtenu un score moyen...	Elle se situe...
inférieur à 41,8	entre la 1re et la 10e position
entre 41,8 et 42,8	entre la 10e et la 20e position
entre 42,9 et 44,7	entre la 20e et la 30e position
entre 44,8 et 46,6	entre la 30e et la 40e position
entre 46,7 et 47,9	entre la 40e et la 50e position
entre 48 et 48,9	entre la 50e et la 60e position
entre 49 et 49,9	entre la 60e et la 70e position
entre 50 et 51	entre la 70e et la 80e position
supérieur à 51	entre la 90e et la 100e position

Annexe 2

Comment situer votre classe par rapport à des classes qui lui "ressemblent" ?

Le score moyen de réussite des élèves, on l'a dit, est de 46,9/60. Votre classe obtient, selon les cas, un score qui est proche de cette moyenne, ou qui lui est inférieur ou supérieur, de peu ou de beaucoup. C'est là une première façon de se situer.

Vous savez cependant, même si cela n'obéit pas à un déterminisme strict, qu'il est plus facile ou plus courant d'obtenir de bons résultats quand la classe compte davantage d'élèves de milieu favorisé, parlant le français à la maison ... Pour chacune des classes, nous avons recueilli des informations via le questionnaire "Enseignant" sur la proportion d'élèves qui :

- ont redoublé au moins un an au cours de leur scolarité antérieure;

- ne parlent pas le français à la maison;
- vivent dans des conditions de vie précaire (parents au chômage, familles très nombreuses, problèmes sociaux, ...);
- sont issus de milieux socio-culturellement favorisés (enseignants, cadres, professions libérales, ...).

Nous avons demandé, en outre, si l'école fait ou non partie d'un projet ZEP, ZAP ou SOUTIEN.

A partir de ces renseignements sur la classe, nous sommes en mesure de fournir des résultats par "types" de classes (classes ZEP ou pas ZEP, classes comptant plus de 20 % d'élèves ne parlant pas le français à la maison, ...). Ces résultats moyens par "sous-groupes" vous permettront de vous situer plus précisément par rapport à des classes qui "ressemblent" à la vôtre. Pour rappel, le score **moyen général** au test est de 46,9/60.

	Score total moyen dans ce type de classe	% de ce type de classe dans l'échantillon
Classes ZEP, ZAP ou soutien	44,4	12%
Autres classes	47,5	88%
Classes comprenant 25 % d'élèves ou plus ayant redoublé au moins une fois	45,2	28%
Autres classes	47,7	72%
Classes comprenant 50 % d'élèves ou plus issus de milieu favorisé	49,9	24%
Autres classes	46,1	76%
Classes comprenant 20 % d'élèves ou plus ne parlant pas le français à la maison	43,6	15%
Autres classes	47,7	85%
Classes comprenant 25 % d'élèves ou issus d'une famille en situation précaire	43,4	33%
Autres classes	48,7	66%

Annexe 3

Niveaux de compétence des élèves

Les 60 items du test ont été soumis à une analyse statistique (analyse de RASCH) qui permet :

- de hiérarchiser les items sur une échelle, du plus simple au plus difficile;
- d'exprimer sur la même échelle le niveau de compétences des élèves.

Parmi les items ainsi hiérarchisés, on peut, en regroupant des "blocs" d'items proches, déterminer des "niveaux de compétence" et voir combien d'élèves se situent à chacun des niveaux.

Sur l'échelle, la moyenne de la difficulté des items est par convention fixée à 0 et *l'écart type* à 1. Les items sont rangés du plus facile (indice = -2.588) au plus difficile (2.986). L'indice de

difficulté exprime aussi la distance entre les items : deux items consécutifs peuvent être pratiquement de même difficulté, ou l'un beaucoup plus difficile que l'autre (c'est par exemple le cas du dernier et de l'avant-dernier item).

Les résultats doivent se comprendre comme suit : un élève qui a une compétence correspondant à la difficulté de tel item (par exemple l'item 2.10.1 : lecture de l'heure - 8 h ou 20 h) a une probabilité de 80 % de réussir cet item et tous les items plus faciles de l'échelle (de 1.2.1 à 3.17), et une probabilité inférieure à 80 % de réussir les items plus difficiles sur l'échelle (de 1.3 à 4.7). Le seuil a été fixé à 80 % parce que c'est habituellement à ce niveau que l'on fixe le seuil de maîtrise.

La hiérarchie des items obtenue est la suivante :

Des items les plus faciles...

	Indice de difficulté
Niveau 1 : Élémentaire	
1.2.1. Prix : 70 F	- 2.588
1.2.2. Prix : 200 F	- 2.245
1.2.3. Prix : 560 F	- 2.214
3.9. Calcul : 60	- 1.917
3.7. Calcul : 35	- 1.885
3.14. Calcul : 30	- 1.847
4.2.2. Pion Laurent, case 10	- 1.509
4.2.1. Pion Julie, case 9	- 1.473
3.6. Calcul : 39	- 1.455
3.11. Calcul : 40	- 1.415
3.8. Calcul : 31	- 1.275
3.10. Calcul : 30	- 1.260
3.12. Calcul : 35	- 1.105
Niveau 2 : Intermédiaire	
2.6. Heure d'entrée : 9 heures	- 1.016
3.13. Calcul : 12	- 0.960
3.1. Argent de Julie : 48 F	- 0.724
1.4.2. Adresse Laurent	- 0.677
3.17. Opération : 20 - 5	- 0.597
2.10.1. Lecture d'heure : 8 h ou 20 h	- 0.590
1.3. Ouverture piscine : 9 h	- 0.583
4.1. Laurent va commencer	- 0.500
3.16. Opération : 20 x 5	- 0.436
2.9.1. Relier Lavabo - 20 cm	- 0.405
3.3. Colorier les pièces : 1 p. de 20 F et 2 p de 5 F	- 0.296
4.8. Case 10	- 0.238
1.4.3. Adresse n° 3 : Piscine	- 0.210
1.4.1. Adresse n° 1 : Julie	- 0.179
Niveau 3 : Compétent	
1.1.1. Copie du prénom, du nom et de l'adresse de Laurent.	- 0.132
1.1.5. L'élève a compris la tâche.	- 0.124
3.18. Opération : 20 + 5	- 0.113
4.9. Retourner à la case 4	- 0.094
3.2. Prix glace de Julie : 30 F	- 0.055
2.4. Mesure de la longueur : 25 m	- 0.041
2.10.3. Lecture d'heure : 3 h 30 ou 15 h 30	0.056
1.1.2. Emploi correct des majuscules	0.121
2.9.2. Relier Bassin : 2m50	0.229
2.10.2. Lecture d'heure : 10 h 15 ou 22 h 15	0.303
2.9.3. Relier baignoire : 60 cm	0.324
4.6. Réponse correcte : non	0.342
4.10. Case 8 ou en prison	0.507
2.5. Nombre de mètres : 50 m	0.592
3.4. Reste de Julie : 18 F	0.594
1.1.3. Orthographe : copie correcte et complète	0.657
2.3. Phrases correctes : les deux phrases correctes (3e et 4e) sont entourées	0.776
2.7. Heure de sortie : 10 h 30	0.860
3.15. Calcul : 15	0.870
1.1.4. Contenu de la carte : Présence d'une formule d'adresse à Laurent et d'une signature	0.992

Niveau 4 : Avancé		
2.1.	Profondeur de l'eau : 1 mètre	1.043
2.2.	Nombre de longueurs : 4	1.126
4.3.	Le pion de Julie est sur la case 21	1.216
1.1.6.	Nom de la rue où se trouve la piscine correct : "Grand-rue"	1.275
2.11.2.	Durée du film : 2 heures 30	1.347
4.5.	Réponse correcte : 4	1.392
2.11.1.	Durée du film : 2 heures 15	1.537
2.8.	Durée : 1 heure 30	1.615
3.5.	La monnaie de retour : 55 F	1.615
4.4.	Le pion de Laurent est sur la case 4	1.628
1.1.7.	Heure d'ouverture le samedi correcte : 9 heures, 9-13 heures ou de 9 à 13 heures	1.635
2.11.3.	Durée du film : 30 minutes	2.454
4.7.	Justification précise du NON	2.986

...aux items les plus difficiles

D'un point de vue technique, cette analyse est très solide : la hiérarchie des items qui en résulte est en effet **indépendante** de la population d'élèves qui a passé le test. Le même test, administré à des élèves plus jeunes ou plus âgés donnerait bien sûr des pourcentages de réussite plus ou moins élevés, mais la hiérarchie des items devrait rester la même. Quand sur une telle échelle, un item apparaît plus facile qu'un autre, ce n'est dû ni au hasard, ni à la force des élèves; il y a à cela de "bonnes" raisons sur lesquelles il convient de s'interroger.

Au départ de cette échelle robuste, il est donc légitime de tenter de regrouper des items en "blocs" ou "niveaux" de compétence.

Le nombre de niveaux et l'endroit précis où l'on place les barres entre niveaux dépendent certes de la personne ou du groupe de personnes qui effectuent ce travail. **Le groupe**

de concepteurs de l'épreuve d'évaluation externe, composé de chercheurs et d'inspecteurs, après examen attentif de la hiérarchie des items, s'est mis d'accord pour déterminer quatre niveaux de compétence dans l'épreuve. Ce choix d'un découpage en quatre niveaux résulte donc du consensus d'un groupe; il a nécessairement un caractère relatif et en partie arbitraire.

Les quatre niveaux de compétences se définissent comme suit :

- Niveau 1 : Élémentaire

Les élèves qui se situent à ce niveau ne maîtrisent vraiment (à 80 %) que les compétences isolées les plus simples (celles dont plusieurs enseignants dans le questionnaire ont dit qu'elles étaient **trop** simples) : les calculs les plus simples de la situation 3, relier les prix à ce à quoi ils correspondent dans la situation 1. Selon les membres du groupe d'inspecteurs et de chercheurs qui

a conçu l'épreuve, tous les élèves de fin de 2e devraient maîtriser ces compétences.

- **Niveau 2 : Intermédiaire**

Les élèves qui se situent à ce niveau maîtrisent les compétences du niveau 1 et des compétences déjà plus complexes. Parmi celles-ci, on retrouve des compétences isolées :

- . certains calculs ou opérations ($20 - 5$) et (20×5), l'identification des adresses, l'heure d'ouverture de la piscine (QCM), la lecture des heures les plus simples (heure juste);
- . certaines questions peu complexes posées en situation : combien Julie a-t-elle d'argent ?, colorier les pièces, qui va commencer le jeu ? ...
Toutes ces questions sont bien ciblées et demandent un traitement de l'information simple (une seule opération à effectuer, pas de compréhension complexe). Selon les membres du groupe concepteur de l'épreuve, ces compétences sont toutes à la portée d'élèves de fin de 2e et doivent donc être maîtrisées. Aucune notion qui n'aurait été abordée dans le programme ne figure parmi ce groupe.

- **Niveau 3 : Compétent**

Les élèves qui se situent à ce niveau sont capables de réussir à 80 % des questions qui correspondent :

- . à des compétences testées isolément telles que : la lecture de l'heure (demi-heure et quart d'heure), des opérations ($20 + 5$), la lecture isolée des règles du jeu de la situation 4, l'évaluation de la profondeur de la baignoire et du bassin et le calcul $50 - 35$;
- . à des compétences testées en situation, mais restant d'un ordre de complexité accessible : copie de la carte (sans rien oublier, majuscules et orthographe), compréhension globale de la tâche (Sit. 1), heure de sortie de la piscine, prix de la glace de Julie, reste de Julie (18 F), Laurent ne peut pas gagner, longueur de la piscine, ...
Il apparaît souhaitable aux concepteurs de l'épreuve que tous les élèves maîtrisent ce niveau de compétences.

- **Niveau 4 : Avancé**

Les compétences figurant dans ce groupe d'items sont en effet des compétences complexes dont on peut penser :

- . soit qu'elles excèdent quelque peu, sur le plan des notions, ce qui est requis en fin de 2e : on retrouve ici tous les items relatifs à la durée, poser et résoudre la soustraction $100 - 45 = 55$, l'estimation des profondeurs, et les questions relatives à la longueur;
- . soit qu'elles sont évaluées dans un contexte qui n'est pas totalement adéquat (le contexte de jeu simulé de la situation 4 a

posé certains problèmes) ou est trop chargé sur le plan cognitif (l'élève a trop de choses à faire à la fois).

Par ailleurs, certains des items de ce groupe ne sont pas pleinement satisfaisants sur le plan technique : ils ont posé des problèmes de compréhension ou de correction.

Pour ces différentes raisons, il apparaît légitime aux concepteurs de l'épreuve, de considérer que les élèves de fin de 2e ne doivent pas réussir (à 80 %) ce type d'items. Maîtriser ces

compétences telles qu'elles ont été évaluées ici représente en quelque sorte "un plus", mais ne fait pas partie du bagage que le groupe juge indispensable à ce niveau.

Pour chacun des paliers ainsi définis, on peut déterminer combien d'élèves se situent à ce niveau. Rappelons que ces niveaux n'ont pas de valeur absolue; ils sont déterminés par rapport à l'épreuve et aux compétences qui y sont évaluées, sous la forme où elles y sont évaluées.

	% d'élèves qui ont une probabilité de 80 % de réussir les items de ce groupe
Niveau 1 : Élémentaire	11 %
Niveau 2 : Intermédiaire	20 %
Niveau 3 : Compétent	37 %
Niveau 4 : Avancé	32 %

Selon le groupe concepteur de l'épreuve, le niveau 3 peut être considéré comme celui que devrait atteindre les élèves à l'issue du 1er cycle. On voit que 69 % des élèves ont atteint ce niveau de compétences. Parmi ces 69 %, 32 % atteignent, en outre, un niveau de compétences approfondies ou complexes. Vingt pour cent des élèves (ceux du niveau 2) sont en outre sur le chemin de cette maîtrise. Leur probabilité de réussir les items du niveau 3 est en effet inférieure à 80 %, sans être négligeable. Pour les

élèves se situant au niveau 1 (11 % des élèves), le chemin vers la maîtrise reste encore long à parcourir; ces élèves ne maîtrisent en effet que le niveau des compétences les plus simples évaluées isolément et sont en difficulté dès que les compétences se complexifient ou sont insérées dans un contexte.

Même si le cas de ces 11 % d'élèves est inquiétant, il n'empêche que le bilan, pour les compétences évaluées, s'avère globalement satisfaisant.