



Faculté de Psychologie et des Sciences de l'Education
Département Education et Formation
Pédagogie théorique et expérimentale
Professeur Marcel CRAHAY

Université de Liège

PISA 2000 : PROGRAMME INTERNATIONAL DE L'OCDE POUR LE SUIVI DES ACQUIS DES ÉLÈVES

Présentation de l'enquête

Dominique Lafontaine
Ariane Baye
Anne Matoul

PISA 2000 : PROGRAMME INTERNATIONAL DE L'OCDÉ POUR LE SUIVI DES ACQUIS DES ÉLÈVES

Présentation de l'enquête

Dominique Lafontaine
Ariane Baye
Anne Matoul

VUE D'ENSEMBLE DU PRO- GRAMME PISA :

En quoi consiste PISA ?

Une évaluation internationale des connaissances et des savoir-faire des élèves de 15 ans

Développé conjointement par les pays membres de l'OCDÉ, le Programme international pour le suivi des acquis des élèves (PISA) vise à **évaluer dans quelle mesure les élèves qui s'acheminent vers la fin de leur scolarité obligatoire ont acquis certaines des connaissances et des savoir-faire indispensables pour participer pleinement à la vie en société.**

Une collaboration entre gouvernements

PISA est une opération de type coopératif, qui se fonde sur une mise

en commun de l'expertise scientifique des pays participants. Les gouvernements respectifs la dirigent ensemble, sous la houlette de l'OCDÉ et à partir d'une plate-forme d'intérêts communs en matière de politique d'enseignement. Ces pays ont travaillé conjointement à la mise au point d'une méthode d'évaluation des élèves valide pour l'ensemble des pays, susceptible de fournir une mesure robuste des savoir-faire pertinents et fondée sur des situations de la vie réelle.

Un mécanisme de pilotage périodique

La première campagne de tests PISA (1^{er} cycle) s'est déroulée en 2000 dans 32 pays. Douze autres pays non-membres de l'OCDÉ ont rejoint ce 1^{er} cycle et ont testé leurs élèves en 2001 (Pisa +).

Tableau 1 : Pays participant à PISA

	Pisa +
Allemagne	Argentine
Australie	Albanie
Autriche	Bulgarie
Belgique	Chili
Brésil*	Hong Kong
Canada	Indonésie
Corée	Israël
Danemark	Lituanie
Espagne	Pérou
Etats-Unis	Macédoine
Finlande	Roumanie
France	Thaïlande
Grèce	
Hongrie	
Irlande	
Islande	
Italie	
Lettonie	
Japon	
Lichtenstein	
Luxembourg	
Mexique*	
Norvège	
Nouvelle Zélande	
Pays-Bas	
Pologne	
Portugal	
République tchèque	
Royaume Uni	
Russie	
Suède	
Suisse	

* Pays non-membre de l'OCDE

Par la suite, des évaluations prendront place tous les trois ans (en 2003, 2006, 2009...). Trois «domaines» – la compréhension de l'écrit, la culture mathématique et la culture scientifique – seront évalués dans chaque cycle. Toutefois, dans chacun des cycles particuliers, les

deux tiers du temps de test seront consacrés à un domaine «majeur», évalué en profondeur. PISA 2000 a fait la part belle à la compréhension de l'écrit ; la culture mathématique et la culture scientifique deviendront les domaines majeurs en 2003 et 2006 respectivement.

Figure 1 : Visualisation des cycles de PISA

1998	1. Production/Évaluation du matériel de test		
1999	2. Prétest		
2000	3. Test domaine majeur : lecture		
2001	4. Analyse des résultats	1. Production/Évaluation du matériel de test	
2002	5. Publications	2. Prétest	
2003		3. Domaine majeur : mathématique	
2004		4. Analyse des résultats	1. Production/Evaluation du matériel de test
2005		5. Publications	2. Prétest
2006			3. Domaine majeur : sciences
2007			4. Analyse des résultats
2008			5. Publications

Une enquête internationale à grande échelle

Dans chaque pays, les échantillons évalués comptent entre 4.000 et 10.000 élèves; diverses épreuves papier/crayon sont proposées à chacun de ces échantillons. Les élèves remplissent en outre un questionnaire sur leur environnement éducatif (famille et école) et sur leurs attitudes

des par rapport à l'apprentissage. Un large éventail d'informations sur les contextes scolaires est également recueilli via les chefs d'établissement.

En quoi PISA est-il différent des enquêtes internationales antérieures ?

Il est prioritairement ciblé sur les aspects qui sous-tendent les politiques d'enseignement. Les gouvernements veulent pouvoir répondre à des questions telles que : « Nos écoles préparent-elles les enfants à participer pleinement à la vie en société ? », « quelles sont les structures et les pratiques pédagogiques qui optimisent les chances de succès pour les élèves issus de milieux défavorisés ? », « peut-on quantifier l'impact sur les résultats des élèves de la qualité des ressources dont disposent les établissements scolaires ? ». PISA est conçu pour contribuer à une meilleure compréhension de ces questions, notamment en fournissant un ensemble de résultats comparables à travers un grand nombre de pays.

Le fondement de PISA est un modèle dynamique d'apprentissage, dans lequel les nouvelles connaissances et les savoir-faire nécessaires pour réussir son adaptation dans un monde en perpétuelle mutation sont acquis de manière continue par l'individu. Les élèves ne peuvent apprendre à l'école tout ce dont ils auront besoin dans leur vie d'adulte. Pour qu'ils soient des apprenants efficaces dans la vie, il leur faut des bases solides dans des domaines-clés comme la

compréhension de l'écrit, les mathématiques et les sciences. Ils doivent aussi être capables d'organiser et de réguler leur apprentissage, d'apprendre par eux-mêmes et en groupe, et de surmonter les difficultés rencontrées. Cela exige d'eux qu'ils soient conscients de leurs propres processus de réflexion, de leurs stratégies et de leurs méthodes d'apprentissage. Afin de pouvoir évaluer ces aspects, PISA teste non seulement les connaissances et savoir-faire des élèves, mais pose aussi des questions sur leurs stratégies, les intérêts qu'ils privilégient et les différents types de situations d'apprentissage qui leur conviennent le mieux.

PISA offre un instrument de contrôle à long terme dont les échéances sont pré-établies. En effet, l'engagement pris de couvrir chaque domaine d'évaluation tous les neuf ans en profondeur, avec des mises à jour tous les trois ans, donne aux pays la possibilité de contrôler de manière régulière les progrès obtenus dans leur poursuite des objectifs-clés de l'éducation.

Les compétences testées correspondent à des situations les plus proches possible de la vie réelle. Les enquêtes internationales antérieures se concentraient davantage sur les acquis et les connaissances « scolaires ». PISA vise à mesurer des performances des élèves qui vont au-

delà du curriculum enseigné dans les domaines ciblés.

Qui est concerné par l'enquête ?

Dans les études internationales antérieures, le principal critère de sélection était l'année d'étude fréquentée par la majorité des élèves d'un âge donné. Une fois l'année d'étude sélectionnée, tous les élèves de cette année étaient concernés, qu'ils soient à l'heure, en avance ou en retard scolaire. Mais l'objectif de PISA est d'évaluer les acquis des élèves au moment où ils s'apprêtent à quitter la scolarité obligatoire à temps plein. Dès lors, le **critère de sélection de la population testée est l'âge**. L'étude concerne donc **tous les élèves âgés de 15 ans** au moment de la passation des épreuves, quelle que soit leur année d'étude, le type de programme, de filière ou d'établissement dans lequel ils sont inscrits, à temps plein ou à temps partiel. Les élèves intégrés dans l'enseignement ordinaire et certaines formes de l'enseignement spécial sont également concernés. En fonction des politiques nationales - modalités de promotion, de différenciation et d'orientation - ces élèves de 15 ans peuvent être répartis dans un nombre plus ou moins important d'années d'études et dans des programmes d'études plus ou moins différents. Ainsi, dans tel pays, la majorité des

élèves fréquentent la 9^e ou 10^e année d'études et suivent un programme commun; dans tel autre, une partie des élèves sont dans le cycle inférieur et une autre partie dans le cycle supérieur (collège/lycée par exemple); dans tel autre enfin, les élèves se répartissent sur plusieurs années d'études et suivent des parcours fortement différenciés. PISA ne mesure pas l'efficacité des établissements où les élèves sont actuellement inscrits, mais l'impact cumulé des expériences d'apprentissage qu'ils ont connues depuis leur entrée dans le système scolaire.

Qu'évalue PISA ?

Les connaissances et les savoir-faire des élèves

PISA évalue un certain nombre de connaissances et de savoir-faire fondamentaux pour une insertion sociale et économique réussie et pour l'apprentissage tout au long de la vie. Certes, ces aspects sont partiellement inclus dans le curriculum scolaire – par exemple, la maîtrise de concepts scientifiques fondamentaux – mais les tests PISA veulent évaluer davantage que la maîtrise d'un ensemble défini de connaissances acquises récemment par les élèves au cours de leurs deux ou trois dernières années d'étude. Ils examinent la capacité des élèves à

tirer parti d'une réflexion active sur leurs connaissances et leur expérience et à traiter des questions qui seront pertinentes pour leur avenir.

Des facettes multiples du produit de l'éducation

PISA 2000 teste trois «domaines» : la compréhension de l'écrit, la culture mathématique et la culture scientifique. Le terme anglais *literacy*, transposé ici en «compréhension de l'écrit» pour la lecture, en «culture» pour les mathématiques et les sciences, a été choisi pour mieux refléter la vaste gamme des connaissances, savoir-faire et compétences qui sera évaluée. Les domaines d'évaluation sont définis en termes de contenus ou de structure des connaissances que les élèves doivent acquérir dans chaque domaine, de processus à mettre en œuvre et de contextes dans lesquels ces connaissances et savoir-faire sont sollicités. Pour chacun des domaines, on dispose d'une échelle continue dont les scores permettent de représenter les niveaux de performance des individus et la distribution des performances des populations. Il n'y a pas sur ces échelles un seuil unique distinguant ceux qui maîtrisent ou non le domaine : les résultats des élèves sont définis en fonction d'une série de niveaux de compétence successifs.

Des attitudes et des compétences transversales

Les compétences transversales gagneront en importance dans PISA au fur et à mesure que le programme évoluera dans le temps. PISA 2000 analyse la motivation des élèves ainsi que leurs attitudes et leurs stratégies d'apprentissage¹. En 2003, PISA évaluera plus spécifiquement la capacité des élèves à résoudre des problèmes.

Les épreuves de compréhension de l'écrit requièrent des élèves qu'ils exécutent tout un éventail de tâches à partir de différents types d'écrit : trouver une information spécifique, démontrer qu'ils comprennent globalement le texte, interpréter celui-ci ou encore réfléchir sur son contenu ou sa forme. Les textes utilisés ne se cantonnent pas aux traditionnels passages en prose mais comprennent une grande variété de types d'écrit : des listes, des formulaires, des graphiques et des diagrammes.

Les épreuves de culture mathématique font appel à différents niveaux de compétences en mathématiques : l'élève doit effectuer des opérations mathématiques élémentaires, mais aussi raisonner et faire preuve de compréhension mathématique (mathématisation). Elles mettent également en œuvre la connaissance et l'application d'une série de

¹ Cette évaluation se fait via l'option internationale CCC (*Cross Curriculum Competencies*), à laquelle la Communauté française de Belgique a décidé de ne pas participer en 2000.

contenus mathématiques empruntés à des domaines tels que le hasard, les variations et la croissance, l'espace et les formes, le raisonnement quantitatif, l'incertitude, la dépendance et les relations, ce qui inclut des domaines spécifiques du curriculum scolaire comme l'algèbre, les nombres ou la géométrie.

Les épreuves de culture scientifique font appel à l'utilisation de concepts scientifiques clés pour comprendre le monde naturel et contribuer à la prise de décisions à son propos. Cela implique en outre d'être capable d'identifier les questions qui sont d'ordre scientifique, de tirer des conclusions scientifiques et d'être capable de les communiquer. Les concepts scientifiques qui sont utilisés se rapportent à l'environnement actuel et à venir des élèves; ils portent sur les questions scientifiques relatives à la vie et à la santé, à la Terre et l'environnement, et aux technologies.

Qu'apportera PISA ?

Les résultats de PISA seront publiés en anglais et en français dans une série de rapports internationaux, à partir de 2001. Un premier rapport international prévu pour décembre 2001 présentera les résultats globaux, tandis que les gouvernements nationaux mettront au point leurs propres dispositifs de présentation

des résultats, afin de les situer dans le cadre de leur propre système éducatif et dans le contexte local. L'OCDE publiera en outre une série de rapports thématiques examinant les implications des résultats pour les politiques d'enseignement. Enfin, la riche base de données de l'enquête sera mise à la disposition de tout chercheur souhaitant entreprendre des analyses complémentaires. Les résultats seront également résumés sur le site de PISA (www.pisa.oecd.org/).

Les résultats de PISA, comme on le verra ci-après, fournissent un profil de base des compétences et savoir-faire des élèves. Pour chacun des domaines, le niveau des acquis des élèves est situé sur un continuum décrivant leur capacité à effectuer des tâches spécifiques. Les résultats sont également disponibles pour des sous-groupes déterminés, par exemple en fonction du sexe, de l'origine socio-économique ou du statut d'immigré ou d'enfant d'immigré.

Les analyses tentent de dégager les facteurs démographiques, sociaux, économiques et pédagogiques qui sont déterminants pour les résultats des élèves et des écoles. Par exemple, les données PISA permettent :

- de mettre les acquis des élèves en relation avec le contexte d'enseignement;
- d'analyser les relations entre les résultats des élèves et des facteurs

scolaires tels que la qualité des ressources humaines et matérielles, le caractère privé ou officiel de l'instance de contrôle, le financement et les mécanismes de prise de décision;

- d'analyser les différences entre les profils de rendement au sein des pays, en particulier la mesure dans laquelle les variations de rendement entre élèves se situent entre écoles ou au sein des écoles, ainsi que l'importance de l'impact exercé par l'école sur le rapport entre la performance des élèves et le capital socio-économique et culturel de leurs familles;
- de comparer divers aspects de la vie des élèves, tels que leurs attitudes par rapport à l'apprentissage, et leur vie au sein de l'école et dans leur environnement familial.

PISA devrait permettre de mieux comprendre les forces et les faiblesses relatives des différents systèmes éducatifs. Ainsi, si l'on observe dans certains pays une moyenne élevée des performances qui va de pair avec un faible écart entre les élèves les plus et les moins performants, cela tend à prouver que l'élévation de niveau peut s'obtenir sans que ce soit au détriment des plus faibles. Ou encore si l'on constate que le lien entre l'origine socio-économique des élèves et leurs performances est d'une ampleur variable selon les pays, cela montre que certaines modalités

d'organisation du système éducatif contribuent à limiter le poids du déterminisme social.

Tous les trois ans, une nouvelle série de résultats et d'analyses fournira des informations sur la façon dont les caractéristiques des élèves évoluent. En comparant l'orientation et le rythme du changement dans différents pays, les décideurs politiques auront la possibilité de situer les développements entrepris au niveau local dans le contexte des changements globaux.

Ainsi, pour la Communauté française de Belgique, PISA représente ainsi une occasion d'évaluer les effets de réformes fondamentales telles que la mise en œuvre des **Socles de compétences** ou des **compétences terminales et savoirs requis**, les discriminations positives, ou de réformes/dispositifs plus ciblés tels que les efforts entrepris pour améliorer l'enseignement des sciences ou le projet Cyber-Ecoles (une des options internationales porte sur la familiarité avec l'informatique, voir infra, p. 21).

Comment est organisée l'évaluation ?
--

Une batterie d'items d'évaluation

Les évaluations PISA 2000 se présentent sous forme papier/crayon¹. Il est demandé aux élèves de lire un certain nombre d'extraits de textes (parfois en combinaison avec des schémas ou des graphiques) et de répondre aux questions qui les accompagnent. Certaines questions sont à choix multiple, d'autres sont des questions à réponse construite. L'objectif d'une grande partie du test est de déterminer si les élèves sont capables de réfléchir activement au domaine plutôt que de simplement restituer les connaissances acquises.

Des items conçus pour avoir une bonne validité de pays à pays

Une des caractéristiques les plus importantes de PISA est que des connaissances et des savoir-faire réellement utiles sont évalués d'une manière valide dans beaucoup de pays différents. En 1999, un grand nombre d'items de test ont été soumis à un essai de terrain dans tous les pays participants. Les résultats de l'essai ont été utilisés pour sélectionner les items qui contribuent le mieux à l'élaboration d'indicateurs de connaissances et savoir-faire pertinents et qui s'avèrent valides pour établir des comparaisons entre élèves des différents pays.

L'équivalent de près de sept heures de test

Le test dure deux heures par élève mais, étant donné que différentes combinaisons de matériel de test sont administrées aux différents élèves, les items de la batterie équivalent au total à sept heures de test, ce qui assure une large couverture des domaines à évaluer.

¹ Il n'est pas exclu, dans des étapes ultérieures, d'évaluer les élèves sur support informatique.

Des données contextuelles sont également recueillies.

Chaque élève consacre environ 20 minutes et chaque directeur d'école environ 30 minutes à répondre à un questionnaire. Ces questionnaires fournissent des renseignements contextuels importants, utiles pour l'interprétation et l'analyse des résultats. Les questions posées aux élèves concernent leurs caractéristiques personnelles (sexe, milieu socio-économique, langue parlée à la maison, statut d'immigré, organisation de la famille...), leurs attitudes envers l'apprentissage et l'école, leurs loisirs, leurs habitudes et leurs attitudes dans le domaine de la lecture. Le questionnaire du chef d'établissement envisage de nombreux aspects de l'organisation de l'établissement : population, formes d'enseignement organisées, ressources humaines et matérielles, climat de l'école, etc.

Les élèves sont testés dans leur école

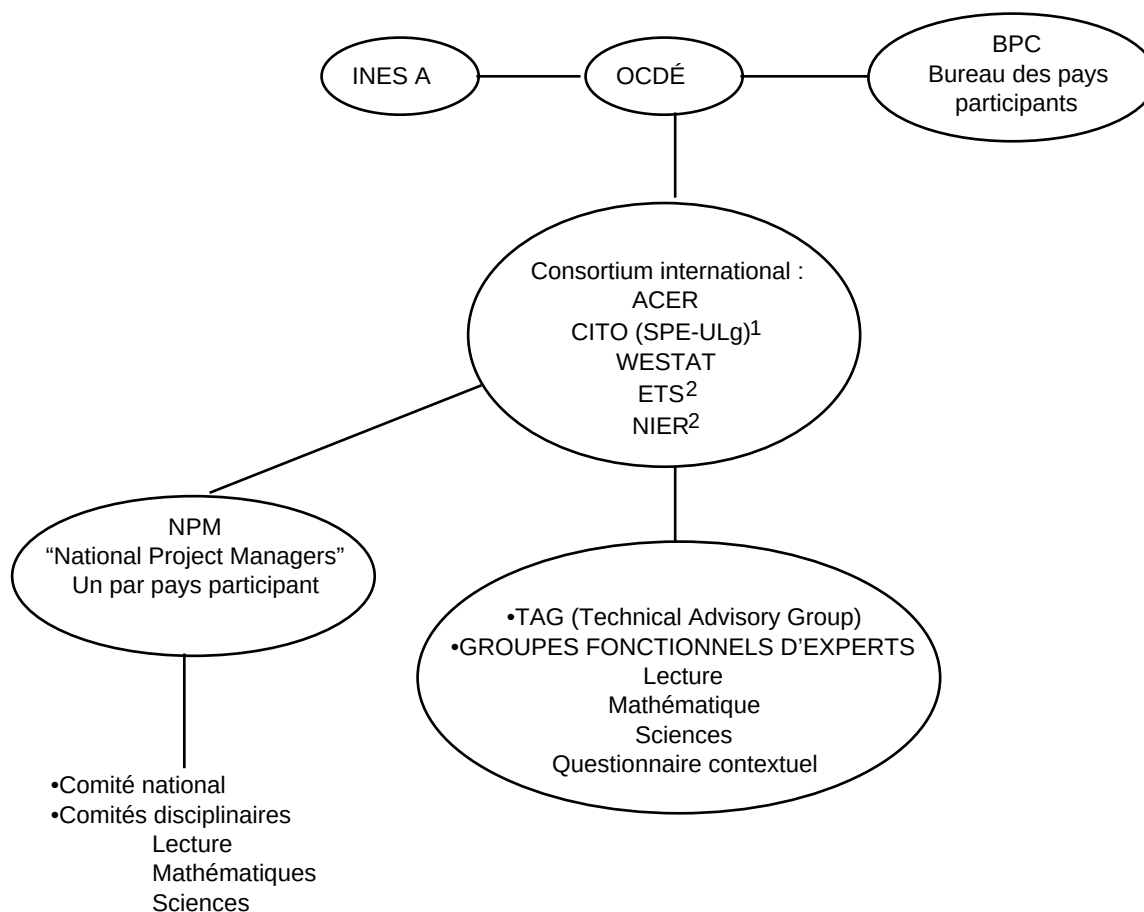
Les résultats ont été communiqués aux écoles de façon confidentielle, en conformité avec les règles en vigueur pour ce genre d'études. Chaque établissement participant ne reçoit que ses propres résultats. Les autres résultats sont traités de manière strictement anonyme.

ASPECTS MÉTHODOLOGIQUES ET DÉROULEMENT DE PISA

Par qui est conduit PISA ?

Figure 2 :

ORGANIGRAMME PISA



¹ De 1998 à juin 1999.

² A partir de juin 1999.

PISA est le fruit d'une collaboration entre les gouvernements des Etats membres de l'OCDÉ. Après avoir, pendant plusieurs années, utilisé, pour les indicateurs de *Regards sur l'éducation*, les informations issues d'autres enquêtes internationales,

l'OCDÉ a décidé d'organiser son propre recueil de données, avec l'enquête relative à la littératie des adultes d'abord (OCDÉ, 1995 ; OCDÉ, 2000), avec PISA ensuite. Le processus de prise de décision dans PISA suit un cheminement

complexe et fortement hiérarchisé. Au sommet de la pyramide décisionnelle, se trouve le **secrétariat de l'OCDE** (Andreas Schleicher); celui-ci préside le **BPC** (*Board of Participating Countries* : Bureau des pays participants), auquel reviennent les décisions finales concernant PISA. Ce bureau est composé de représentants des pays membres de l'OCDE, proches ou issus des instances politico-administratives responsables en matière d'éducation. Le représentant officiel de la Communauté française de Belgique au BPC est Dominique Barthélemy¹.

Le **réseau INES A** de l'OCDE (Acquis des élèves) joue un rôle moteur dans le développement de PISA². C'est en effet au sein du réseau A que sont développés les appels d'offre (on parle à ce propos de « TOR » : *Termes Of Reference*) pour la mise en œuvre des cycles successifs de PISA. Il revient ensuite au BPC d'approuver ces « TOR » et de choisir, parmi les différents candidats, le contractant qui dirigera l'exécution du programme pour un cycle, sous la houlette du BPC. Pour le premier cycle, le contractant, au départ, est un **consortium international** regroupant quatre centres de

recherche : ACER (Australian Council of Educational Research), CITO (Institut national d'évaluation pédagogique, Pays-Bas), WESTAT (Etats-Unis) et le SPE (Service de pédagogie expérimentale de l'Université de Liège). A partir de juillet 1999, le SPE, suite au départ de deux de ses membres pour l'étranger, est écarté du consortium, qui, dans le même temps, s'élargit à ETS (Educational Testing Service, Etats-Unis) et NIER (National Institute for Educational Research, Japon).

Le consortium international, pour la conduite du programme, s'est entouré de différents **groupes d'experts**, qui définissent les axes prioritaires et font des propositions soumises ensuite à la réflexion des gestionnaires nationaux de projet (« NPM » : gestionnaire national de projet, son rôle sera décrit ci-dessous) et à l'approbation du BPC (c'est nous qui soulignons). Cinq groupes d'experts, composés de spécialistes issus d'horizons divers (sans représentation « nationale » à ce niveau) interviennent ainsi dans PISA : le TAG (Technical Advisory Group : groupe technique chargé des aspects logistiques, échantillon, rotation de carnets, traitement de données...), un groupe fonctionnel pour chacun des domaines évalués (lecture, mathématique et sciences) et un groupe responsable des questionnaires contextuels. Dominique Lafontaine, du Service de pédagogie

¹ Ce dernier se fait régulièrement suppléer au BPC par Christiane Blondin, représentante au réseau A, ou par Dominique Lafontaine, gestionnaire nationale du projet PISA pour la Communauté française de Belgique.

² Christiane Blondin, du Service de Pédagogie expérimentale de l'Université de Liège, représente la Communauté française de Belgique au réseau A.

expérimentale de l'Université de Liège, est membre du groupe fonctionnel d'experts en lecture.

Dans chacun des pays participants, un **gestionnaire national de projet** (ci-après **NPM**) est désigné pour la mise en œuvre nationale de PISA¹. La « NPM », pour la Communauté française de Belgique est Dominique Lafontaine (du Service de pédagogie expérimentale de l'Université de Liège), elle a été secondée dans ce rôle pour le 1^{er} cycle par Anne Matoul et Ariane Baye du même service². Le NPM joue un rôle d'interface entre le consortium international et les instances nationales. Il est chargé d'exécuter le programme en conformité avec les règles et standards définis internationalement et de relayer l'avis du pays sur les différentes questions pour lesquelles il est consulté (par exemple : choix des instruments ou modalités d'évaluation). Les NPM assistent, plusieurs fois par an, à des réunions de coordination internationale ou à des sessions de formation. Afin de représenter d'une manière large et significative l'avis du pays, chaque NPM doit s'entourer de **comités nationaux**. En

Communauté française de Belgique, quatre comités ont été mis en place : un comité général, composé de représentants de l'administration, des cabinets des Ministères concernés et de l'Inspection générale de l'enseignement secondaire, ainsi que trois comités « disciplinaires » (lecture, mathématique et sciences)³, composés d'inspecteurs et de représentants des réseaux. La tâche de ces comités disciplinaires, si on voulait se conformer à la lettre aux exigences du consortium, serait considérable : ainsi, depuis le début du programme à la mi-98, ces comités auraient dû examiner et rendre un avis sur les différentes versions successives des plans d'évaluation des trois disciplines, des questionnaires, des textes et items (environ 600 items). Dans certains pays, de véritables comités d'experts nationaux sont constitués pour l'occasion et rémunérés (au Canada, aux Etats-Unis, dans certains pays nordiques notamment). En Communauté française de Belgique, tous les membres des comités assument d'autres fonctions et leur intervention dans PISA est strictement gracieuse. Si leur implication à certains moments clés est incontournable, il ne paraît ni possible ni raisonnable de les consulter sur l'ensemble des questions liées à PISA. C'est ainsi, par exemple, que la révision générale des items exigée

¹ La participation de la Communauté française de Belgique à PISA est financée par le Secrétariat général – Direction des Affaires Internationales du Ministère de la Communauté française de Belgique.

² En 1999, Christian Monseur participait à PISA, avant son départ pour ACER en Australie. Il a ensuite été remplacé dans l'équipe par Ariane Baye et Anne Matoul. Isabelle Demonty et Réginald Burton sont en outre intervenus dans PISA de façon ponctuelle pour les aspects relatifs aux mathématiques et aux sciences.

³ La liste des membres des comités figure à l'annexe 1.

des centres nationaux pendant les congés scolaires de juillet et août a été effectuée uniquement par la NPM et ses collaboratrices, pour des raisons que l'on peut aisément comprendre.

Il faut préciser qu'aux yeux de l'OCDE, la Belgique est considérée comme un seul système éducatif. Les résultats de PISA dans les analyses internationales concernent donc la Belgique dans son ensemble¹; chaque Communauté disposera évidemment par ailleurs de ses propres résultats. À tous les niveaux de décision cités plus haut (BPC, réseau A, NPM, comités), les deux Communautés ont leurs propres représentants et la gestion du survey sur le terrain se fait de façon indépendante. Chaque Communauté a augmenté la taille de son échantillon de façon à obtenir une estimation correcte pour son propre système éducatif (il aurait suffi de tester 150 établissements pour toute la Belgique; la Communauté française et la Communauté flamande en testent respectivement 100 et 150).

Quelles garanties avons-nous que la comparaison est valide ?

Cette question est la clé de voûte de toute étude comparative. Pour s'engager dans une telle étude, un pays a en effet besoin d'un certain nombre de garanties lui assurant que les performances observées ne sont pas dues à des caractéristiques particulières des instruments observés, à la nature des échantillons testés, ou encore à des procédures de recueil des données qui ne seraient pas identiques dans les différents pays, pour ne prendre que quelques exemples. Les grandes enquêtes internationales prennent donc des mesures, définissent des procédures standards, instaurent des mécanismes de contrôle de qualité destinés à se prémunir contre les risques énoncés plus haut. Il ne fait guère de doute que l'étude PISA, est, de toutes les études comparatives jamais entreprises², celle qui a poussé le plus loin les procédures de standardisation et de contrôle de qualité, d'une part parce qu'elle bénéficie de l'expérience accumulée au cours des trente dernières années en matière d'évaluation comparative, d'autre part parce que l'OCDE est hantée par le souci de se prémunir contre toute critique de « partialité » du type de celles qu'a encourues l'étude IALS (International Adult Literacy Study) de la part de la France³. Pour

¹ Les résultats aux tests dans les trois domaines seront toutefois présentés par communauté linguistique dans une annexe du 1^{er} rapport international.

² Pour plus de détails sur cette question, on pourra consulter l'étude historique que Lafontaine (2000) a consacrée à trente ans d'évaluations comparatives de la lecture, ainsi que Baye (2000).

³ Pour rappel, la France, après avoir participé aux différentes phases d'élaboration de IALS, a mis en

PISA, on peut dire que les pays participants sont sous haute surveillance. L'élévation du niveau de contrôle est d'ailleurs l'un des éléments qui conduit à une augmentation très sensible des coûts nationaux d'une telle enquête comparativement à des enquêtes antérieures du même type.

La question de la validité de la comparaison recouvre deux aspects principaux : d'une part, quels sont les contrôles mis en place pour s'assurer que les élèves sont évalués dans des conditions comparables, d'autre part, quelles garanties a-t-on que les épreuves n'avantagent pas certains pays ou certaines catégories de pays (par exemple les pays anglo-saxons) ?

l'échantillonnage (Westat) s'assure que les pays tirent leur échantillon en respectant les règles définies dans le manuel d'échantillonnage; ce centre se propose d'ailleurs de tirer l'échantillon pour les pays qui le demandent (ce qui fut le cas de la Communauté française de Belgique).

Les contrôles de qualité

Echantillon

Chaque pays doit fournir les chiffres de population et les sources précises permettant de contrôler que tous les élèves de 15 ans concernés par l'étude sont effectivement répertoriés. L'exclusion de certains élèves pour des motifs divers (handicaps, écoles situées dans des zones difficilement accessibles...) ne peut dépasser certains seuils fixés (2,5 %) et doit être dûment justifiée au cas par cas. L'organisme responsable de

question plusieurs aspects méthodologiques de l'étude et s'en est retirée, en sorte que les résultats de la France ne figurent pas dans les rapports internationaux.

Administration des épreuves

Comme dans toutes les études internationales, des procédures standards pour l'administration des tests aux élèves sont définies; celles-ci sont réunies dans le Manuel de l'administrateur : les mêmes consignes et explications doivent être fournies aux étudiants, ceux-ci doivent passer les épreuves dans les mêmes limites de temps. Pour garantir le respect des consignes, trois procédures de contrôle sont mises en place :

- les épreuves doivent être dans tous les cas administrées par du personnel extérieur à l'établissement, dûment formé pour cette tâche;
- ces personnes doivent rédiger un rapport qui détaille le déroulement de la séance et les éventuels problèmes rencontrés (problèmes de discipline, problèmes décelés dans les carnets, ...);
- des visites de contrôle de la qualité du déroulement des séances sont effectuées à l'improviste par des « SQM » (*School quality monitors*). Ces SQM, proposés par les pays, sont recrutés et formés pour cette tâche par le consortium international. Ils n'ont pas de lien avec le centre national. En

Communauté française de Belgique, cette tâche est assumée par des inspecteurs de l'enseignement secondaire.

Lors de l'étude définitive, cinq inspecteurs ont ainsi observé une vingtaine de séances sur les 100 et ont fait directement rapport au consortium sans en référer au centre national. Il s'agit d'un véritable contrôle indépendant.

Correction des épreuves

Les tests PISA comportent une proportion considérable de questions ouvertes complexes (55 % des items en lecture), dont la correction nécessite le recours à des guides de correction standardisés. En dépit de tous les efforts mis dans l'élaboration de tels guides, le risque d'une certaine subjectivité dans la correction subsiste. Pour limiter le plus possible ce risque, différentes procédures ont également été définies :

- Les NPM et les responsables des corrections suivent une formation d'une semaine (pour l'essai de terrain et pour l'étude définitive). De longues séances d'entraînement sont ainsi prévues pour essayer d'harmoniser au maximum les procédures de correction. A leur retour, les NPM forment de la même façon tous les

correcteurs qui ont été recrutés pour corriger les carnets de tests.

- La moitié des carnets ont dû être corrigés par quatre correcteurs différents, afin de contrôler la fidélité entre correcteurs.
- Pour l'étude définitive, des corrections croisées entre pays ont été effectuées. Un certain nombre de carnets sont ainsi corrigés en dehors du pays, sous la responsabilité du consortium. En effet, la fidélité entre correcteurs d'un même pays peut être bonne, même si ceux-ci s'écartent des standards définis; il suffit pour cela d'être cohérents dans l'erreur. C'est pourquoi de tels contrôles croisés sont nécessaires.

Evaluation de la fiabilité internationale des corrections

En décembre 2000, une évaluation de la fiabilité internationale du codage multiple a été effectuée (« *Intercountry reliability report* »). Pour tous les pays, une partie des carnets¹ soumis à la procédure de codage multiple a été réexaminée par des correcteurs internationaux. Il s'agissait de confronter le codage

de ces codeurs internationaux avec les codes attribués par les codeurs nationaux afin d'évaluer d'une part le niveau de fiabilité de ces derniers et d'autre part, le niveau de cohérence entre les procédures de codage multiple mises en place dans les différents pays participants. Voici brièvement comment s'est déroulée cette procédure internationale. Dans un premier temps, les 48 carnets ont été soumis à un correcteur international qui a codé les questions ouvertes indépendamment, c'est-à-dire sans connaître les codes attribués par les codeurs nationaux. Les codes nationaux ont ensuite été comparés aux codes attribués par le correcteur international. En cas de discordance, le correcteur international réexaminait ses corrections et décidait alors soit de se conformer à l'avis majoritaire des codeurs nationaux, soit de maintenir son code initial (en fournissant ses commentaires et justifications). Dans ce dernier cas, la réponse litigieuse fournie par l'élève et les codes discordants étaient soumis à deux adjudicateurs chargés de commenter, justifier et, finalement, de donner raison à l'une ou l'autre partie.

Dans la grande majorité des pays, les résultats de cette étude de fiabilité internationale sont satisfaisants. L'indice moyen de « concordance » est de 92 %. On estime qu'il y a concordance lorsque le correcteur international donne le même code

¹ Ce sont 1248 cas qui ont ainsi fait l'objet de cette vérification internationale : 26 items à réponse ouverte du carnet 7 x 48 élèves.

que 3 des correcteurs nationaux ou lorsque celui-ci est en désaccord avec les correcteurs nationaux mais à tort (selon l'adjudicateur). Les taux de concordance ne sont inférieurs à 90 % que dans 8 pays. En moyenne, 1 % des cas sont corrigés trop sévèrement (avec un maximum de 2,5 % en Lettonie) et 2,5 % des cas sont corrigés avec trop peu de sévérité (avec un maximum de 9,4 % en Lettonie). Ces contrôles sont essentiels car ils permettent d'écarter toute suspicion de fraude ou de laxisme dans la correction des carnets à l'intérieur des pays. Ils permettent aussi d'éliminer les items mal corrigés dans certains pays des données nationales et internationales.

En Communauté française de Belgique, la fiabilité des codeurs multiples a été évaluée à 92,2 %.

Visites de site

Peu avant la période définie pour le test PISA, les centres nationaux et le gestionnaire responsable font l'objet d'une inspection fouillée par un membre du consortium. Celui-ci envisage, au cours d'un entretien enregistré avec le NPM et son équipe, tous les aspects stratégiques de l'étude, afin de vérifier si aucun « dérapage » n'est intervenu. Toutes les procédures sont systématiquement passées en revue.

L'équivalence linguistique et culturelle

Origine du matériel de test

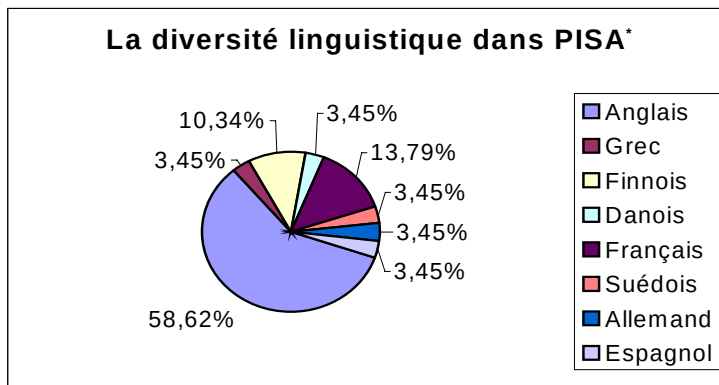
Comme d'autres études antérieures, PISA a fait appel aux pays participants pour qu'ils fournissent du matériel d'évaluation (textes, documents, items) qu'ils estiment adapté pour un public d'élèves de 15 ans. Un matériel nombreux et diversifié a ainsi été recueilli. Malgré cela, il faut constater que les textes d'origine anglo-saxonne restent majoritaires : seules 12 unités sur les 54 que comporte l'étude (soit 22 %) ne proviennent pas d'un pays de

langue anglaise. Vingt-cinq unités au moins sont en anglais (mais peuvent venir de pays différents), auxquelles il faut ajouter une partie des unités IALS reprises dans PISA (17 unités)¹. Les 12 unités non anglaises à l'origine dans PISA appartiennent à 7 langues différentes².

A notre sens, la dominante anglo-saxonne n'est pas problématique en soi dès lors que des contrôles en vue d'éviter les biais culturels et linguistiques sont effectués avec un maximum de rigueur, ce qui est le cas dans PISA (voir ci-dessous).

¹ Une partie des items de l'étude IALS a en effet été reprise dans PISA, afin de comparer les performances des élèves de 15 ans avec celles des adultes. Parmi les unités IALS, certaines étaient à l'origine en anglais et d'autres dans d'autres langues, mais nous ignorons exactement lesquelles.

² Deux des textes proposés par la Communauté française de Belgique ont été retenus dans PISA pour l'évaluation de la compréhension de l'écrit.

Figure 3 :

D'après Baye (2000)

* Ce graphique a été réalisé sans tenir compte des unités IALS, dont on ignore l'origine linguistique exacte.

Adéquation du matériel de test aux différents contextes nationaux

PISA, conforme en cela aux évaluations comparatives de la lecture qui l'ont précédé (Lafontaine, 2000), n'a pas effectué une étude du curriculum de l'enseignement de la lecture dans les différents pays participants, qui aurait permis de déterminer quels y sont les objectifs ou les compétences à atteindre, les types de textes généralement étudiés, ou encore les modalités d'évaluation les plus répandues. Tout au long du processus de choix du matériel et de construction des items, les pays, via leur NPM et les comités nationaux, ont cependant été à plusieurs reprises consultés sur l'adéquation du contenu des épreuves aux réalités des pays. A cette occasion, les NPM devaient indiquer, pour chaque texte et pour chacun des items proposés :

- le degré d'exposition à l'école,

- le degré d'exposition en dehors de l'école,
- une estimation de la difficulté,
- les problèmes culturels éventuels,
- les autres problèmes possibles,
- les problèmes de traduction,
- la familiarité du contenu,
- la familiarité du type de texte,
- l'intérêt pour les 15-16 ans,
- le degré de priorité que le pays donnait à l'item pour son inclusion dans le test (de 1=faible à 4=très forte).

Les valeurs moyennes obtenues pour les différents critères de la grille sont satisfaisantes (des valeurs supérieures à 2 sur des échelles qui comptent 4 niveaux). Les valeurs les plus élevées sont observées pour les mathématiques qui, rappelons-le, ne sont pas la discipline majeure dans cette première phase de PISA.

Les questionnaires contextuels ont fait l'objet du même processus de révision; chaque pays a eu de

multiples occasions de commenter et de solliciter des modifications des questions de l'instrument. Pour de nombreuses questions, des possibilités d'adaptation nationale existent (sous contrôle du consortium) : chaque pays peut choisir les termes les mieux adaptés aux réalités du contexte national (par exemple, les noms des programmes d'études ou les noms des diplômes).

Equivalence des traductions

Les élèves impliqués dans une enquête internationale, classiquement, passent une même épreuve traduite dans les différentes langues d'enseignement des pays concernés, au départ d'une langue source : en général l'anglais. Même si, à l'origine, les textes sont en espagnol, en russe, ou en français, ils doivent tous, à un moment, être traduits en anglais. Dans PISA, la procédure est un peu différente. L'OCDÉ est en effet une organisation bilingue français/anglais et toutes ses publications officielles doivent être fournies dans les deux langues. Les textes proposés par les pays pour inclusion dans les épreuves pouvaient donc être soumis en français ou en anglais; l'ensemble du matériel de test disponible pour les pays est ensuite fourni aux pays dans deux langues sources. La comparabilité des deux versions sources française et anglaise fait l'objet de nombreux contrôles (traduction par deux traducteurs

indépendants, conciliation des deux traductions par un expert, vérification de l'équivalence psycholinguistique des versions anglaise et française par un expert français, relecture par un professionnel de tous les aspects formels).

Un véritable centre de vérification de l'équivalence des traductions a été institué pour PISA, sous la direction d'Aletta Grisay. Celui-ci s'assure les services de traducteurs compétents dans les différentes langues, qui reçoivent en outre une formation spécifique pour l'occasion. Chaque pays doit se soumettre aux procédures suivantes :

- Le matériel de test doit être traduit deux fois d'une langue source vers la langue cible par deux traducteurs indépendants. Il est recommandé aux pays d'effectuer une traduction au départ de l'anglais et l'autre au départ du français. Si ce n'est pas possible, le pays peut faire traduire le test deux fois au départ de l'anglais ou du français.
- Ces deux traductions indépendantes sont ensuite comparées par un traducteur professionnel du centre de vérification, qui les concilie en vue d'arriver au meilleur résultat, en conformité avec un ensemble de règles définies dans un document de référence. Un expert ressource peut être consulté en cas de difficulté.

- Chaque pays dispose d'une marge de manœuvre limitée et définie dans le même document, pour réaliser les adaptations nationales qui s'imposent : par exemple formuler les prix dans la monnaie du pays ou adapter les noms ou prénoms de certains personnages. Toutes ces adaptations souhaitées par les pays doivent être répertoriées et soumises au centre de vérification des traductions et au consortium international pour approbation. En ce qui concerne la Communauté française de Belgique, les adaptations « nationales » pour les épreuves, ont été peu nombreuses et consistaient, pour l'essentiel, en des corrections d'erreurs ou de coquilles dans la version source française qui avaient échappé aux nombreuses relectures précédentes. En revanche, les adaptations nationales dans les questionnaires contextuels sont nombreuses, car il s'agit d'adapter les termes aux réalités du système éducatif concerné : les filières, les autorités de tutelle, les diplômes, les sections... portent des noms différents dans tous les pays.

Parallèlement au centre de vérification, un panel de révision culturelle réunissant différents experts arbitre enfin l'ensemble des questions d'équivalence. Ce panel n'a identifié aucun problème majeur pour le 1^{er} cycle.

Lors de l'essai de terrain, la manière dont les items se comportent dans les différents pays et les différentes langues a fait l'objet de toutes les attentions. Pour chaque item, le programme d'analyse des items¹ vérifie, outre les contrôles relatifs à la qualité psychométrique habituels, que celui-ci ne se comporte pas d'une façon « suspecte » dans l'un ou l'autre pays. Si un item se révélait, dans un pays, exceptionnellement bien réussi, ou l'inverse, comparativement à d'autres items de difficulté équivalente, ce pourrait être pour des raisons de non-équivalence dans la traduction du texte, de l'item ou du guide de correction de cet item. Un rapport intitulé *Dodgy items report* (Rapport sur les items suspects) a été consacré à ces questions; tous les biais y ont été systématiquement traqués (biais liés au sexe, au milieu socio-culturel des élèves, au pays). Très peu d'items présentaient des problèmes d'interaction avec le pays (aucun en Communauté française de Belgique). Ce rapport a servi de base à l'élaboration du test définitif; les items suspects de non-équivalence d'un pays à l'autre ont été systématiquement écartés. Le même type de contrôle a été répété lors de l'étude définitive et a conduit à l'élimination de 8 items litigieux, qui n'entrent dès lors pas en ligne de

¹ Il s'agit du logiciel CONQUEST, qui effectue des analyses selon le modèle de la réponse à l'item (analyses IRT).

compte pour la construction des
échelles de compétences.

La marge de manœuvre des pays : options internationales et nationales

On vient de le voir, un maximum de garanties ont été prises, dans PISA, pour que les pays utilisent finalement un matériel de test le plus équivalent possible et que, si divergence il y a, celle-ci soit intentionnelle et contrôlée. Les pays ont une marge de manœuvre limitée pour quelques adaptations nationales; leur seule véritable marge de manœuvre réside dans le choix de participer ou non aux options internationales et d'ajouter des options nationales à PISA, moyennant l'accord du consortium.

Qu'entend-on par options internationales ? La possibilité est offerte aux pays participants qui le souhaitent, de tester les élèves sur d'autres aspects que ceux de l'étude principale obligatoire, en se servant d'instruments communs. Pour le premier cycle de PISA, deux options internationales sont ainsi proposées :

- Un questionnaire bref (une page recto-verso) portant sur la familiarité avec l'informatique. La Communauté française de Belgique, après consultation du comité national, a décidé de prendre cette option, peu coûteuse en temps. A l'heure où le projet Cyber-écoles est en voie de réalisation, l'occasion d'effectuer à peu

de frais une mesure d'entrée semblait devoir être saisie.

- Un questionnaire portant sur des compétences dites transversales (instrument CCC : Cross Curricular Competencies). Cet instrument a été développé par un groupe d'experts en marge des travaux du réseau A. Il porte sur différentes composantes socio-affectives telles que les méthodes de travail, l'auto-évaluation, la motivation, les attitudes envers l'apprentissage, la préférence pour un apprentissage compétitif ou coopératif, ce qui ne correspond pas vraiment à ce que l'on entend habituellement par « compétences transversales » en Communauté française de Belgique. Lors de l'essai de terrain, la Communauté française de Belgique, par le biais de son comité national, avait, à l'unanimité, renoncé à cette option. Les principales raisons de ce refus étaient :

- une qualité jugée peu satisfaisante de certains items (ce que l'essai de terrain a largement confirmé);
- la longueur et la redondance de l'instrument, qui devait être administré au terme de deux heures et demie de testing;
- des valeurs implicites aisément perceptibles : à certains indices, on sentait que l'apprentissage compétitif avait plus de valeur que l'apprentissage coopératif (du moins les items relatifs à

l'apprentissage compétitifs étaient plus attractifs);

- un problème d'ordre éthique dû au fait que tous les items des échelles étaient orientés dans le même sens (l'élève cochant systématiquement la case 4 représenterait en quelque sorte l'élève «modèle»). Le comité déplorait que l'élève non conforme à ce modèle devait marquer son désaccord ou sa divergence avec le portrait « idéal » proposé.

Pour l'étude définitive, le comité national a une nouvelle fois été consulté quant au choix de cette option. Le groupe a été informé que des développements futurs de PISA, à caractère obligatoire, viendraient immanquablement se greffer sur l'instrument CCC et que la pression de la hiérarchie de l'OCDÉ en vue de convaincre la Communauté française de Belgique de participer (une majorité de systèmes éducatifs, dont la Communauté flamande, participent à cette option) était importante. Au terme d'un débat nourri, le comité national a maintenu, à l'unanimité, sa position de refus : en dépit d'une amélioration sensible de la qualité de l'instrument après l'essai de terrain, il est apparu que l'instrument restait redondant, et que les problèmes de valeurs¹ et

d'éthique restaient entiers. La Communauté française de Belgique, clairement, ne souhaite pas donner la priorité à une comparaison internationale pour l'évaluation de ces aspects à l'aide de l'instrument proposé, sans toutefois nier l'intérêt d'une évaluation de compétences autres que cognitives.

Pour conclure, il reste à mentionner la possibilité pour les pays de développer des options nationales autour de PISA, en profitant du fait qu'un échantillon représentatif d'élèves de 15 ans est testé pour recueillir de l'information sur des questions qui intéressent le pays. Certains pays, tels l'Allemagne, l'Autriche ou le Canada, exploitent au maximum cette possibilité. D'autres pays, comme le nôtre, ne l'exploitent que de façon marginale pour ce premier cycle. Seuls quelques items qui paraissaient indispensables pour combler des manques évidents, ont été ajoutés au questionnaire Elève ou au questionnaire Etablissement ; par exemple, à une liste relative au manque de matériel dans les écoles

d'une motivation intrinsèque (j'étudie parce que les matières m'intéressent ou pour enrichir mes connaissances ou ma culture, pour pouvoir réaliser mon projet, pour pouvoir faire le métier ou les études qui m'intéressent), ou d'une motivation extrinsèque plus circonstanciée (j'étudie pour réussir mon année ou parce que mes parents m'y obligent). Pas de trace non plus d'une motivation plus pragmatique souvent évoquée par les jeunes (« j'étudie pour échapper au chômage »). De l'avis général, l'ensemble du questionnaire CCC est peu « parlant » pour des adolescents de 15 ans, en tout cas en Communauté française de Belgique et ne donne qu'un éclairage partiel, voire partiel, sur l'éventail des valeurs possibles.

¹ Ainsi, par exemple, les seuls motifs pour lesquels on peut étudier proposés au choix de l'étudiant sont d'« accroître ses chances sur le plan professionnel » de « trouver un bon emploi » ou de « s'assurer un avenir stable sur le plan financier ». Pas de trace

qui ne mentionnait que les laboratoires de sciences et les ateliers d'art plastique, a été ajouté un item relatif aux ateliers destinés aux formations professionnelles et techniques. Une question relative aux activités de loisir dans lesquelles s'investissent les jeunes a également été ajoutée.

Caractéristiques du matériel d'évaluation

La compréhension de l'écrit dans PISA

Dans PISA, la compréhension de l'écrit est définie comme « *la capacité de comprendre, d'utiliser et d'analyser des textes écrits, afin de pouvoir réaliser des objectifs personnels, développer des connaissances et des capacités et prendre une part active dans la société.* » (OCDÉ, 1999, p. 15).

PISA évalue la compréhension de l'écrit en fonction de trois dimensions :

- Tout d'abord, le *type d'écrit* ou de texte. De nombreuses enquêtes sur les compétences en lecture des élèves sont principalement fondées sur des textes en prose organisés en phrases et en paragraphes, c'est-à-dire des textes «continus». PISA y ajoute des textes «non continus», qui présentent l'information sous

d'autres formes : des listes, des formulaires, des graphiques ou des schémas. 33 % des unités dans PISA sont des textes non continus, 66 % des textes en prose. Une autre distinction est établie entre les différentes formes de prose, comme les textes narratifs, informatifs ou argumentatifs. Ces distinctions partent du principe que les individus seront confrontés à des formes d'écrit diverses au cours de leur vie d'adulte et qu'il ne leur suffira pas d'être capables de lire un nombre restreint de types de textes, tels qu'on les rencontre habituellement à l'école.

- En second lieu, le *type de tâche de lecture*. Cette classification correspond d'une part aux différentes démarches nécessaires pour être un lecteur efficace, d'autre part aux caractéristiques des questions posées dans le test. Les élèves ne sont pas testés sur les compétences en lecture les plus rudimentaires, car on peut supposer que la majorité des élèves de quinze ans les ont déjà acquises. Il leur est plutôt demandé de démontrer leur aptitude à trouver une information, à comprendre globalement le texte, à l'interpréter et à réfléchir sur son contenu ou sur sa forme en relation avec leurs propres connaissances du monde, et à justifier leur point de vue.

- En troisième lieu, *l'usage pour lequel le texte a été conçu* – son contexte ou sa situation. Un roman, une lettre personnelle ou une biographie sont des exemples de lecture «à usage privé». Les documents officiels ou les annonces sont «à usage public». Un mode d'emploi ou un rapport peuvent être assimilés à la lecture

«à des fins professionnelles». Un manuel scolaire ou un tableau représentent la lecture «à des fins scolaires».

Le test de lecture est composé de 45 % de Q.C.M. et de 55 % de questions ouvertes. Le tableau suivant présente la répartition des épreuves en fonction du type de tâche (aspect) et du type de questions.

Tableau 2 : Distribution des épreuves à réponse construite et à choix multiple en fonction des cinq aspects de la lecture

Aspect	% du test	% des épreuves requérant une réponse construite	% des items du test requérant une réponse construite	% des items du test à choix multiple
Trouver l'information	20	35	7	13
Comprendre globalement	20	35	7	13
Développer une interprétation	30	35	11	19
Réfléchir sur le contenu	15	65	10	5
Réfléchir sur la forme	15	65	10	5
Total	100		45	55

La culture mathématique dans PISA

Dans PISA, la culture mathématique se définit comme « *la capacité d'identifier et de comprendre les rôles joués par les mathématiques et de porter des jugements fondés à leur propos, ainsi que d'utiliser les mathématiques, en fonction des exigences de la vie actuelle et future, en tant que citoyen constructif, responsable et intelligent.* » (OCDE, 1999, p. 15). PISA évalue la culture mathématique en fonction de trois dimensions :

- Premièrement, *les contenus mathématiques*, définis principalement en termes d'idées mathématiques majeures qui sous-tendent la pensée mathématique (par exemple, le hasard, les variations et la croissance, l'espace et les formes, l'incertitude, la dépendance et les relations), et de manière secondaire, par rapport aux «domaines mathématiques enseignés» (par exemple, les nombres, l'algèbre et la géométrie). L'évaluation PISA 2000, où les mathématiques sont un domaine mineur, ne s'articule qu'autour de deux aspects : les variations et la croissance d'une part, l'espace et les formes d'autre part. Ces deux domaines permettent une représentation assez large des différents aspects du curriculum, sans attribuer une

importance excessive aux savoir-faire numériques.

- En second lieu, *les processus mathématiques*, définis en termes de compétences mathématiques générales, parmi lesquelles le maniement du langage mathématique, la capacité de modélisation et de résolution de problèmes. Cependant, il ne s'agit pas de proposer des items qui évaluent ces savoir-faire séparément, étant donné que chaque épreuve mathématique sollicite plusieurs savoir-faire en même temps. Les questions s'organisent plutôt autour de trois «classes de compétences» qui définissent le type de raisonnement auquel l'item fait appel :

- ◇ La première classe de compétences mathématiques comprend les calculs simples ou les définitions que l'on rencontre le plus souvent dans les tests d'évaluation mathématique.
- ◇ La deuxième classe exige une mise en relation pour résoudre des problèmes élémentaires.
- ◇ La troisième classe de compétences implique une mathématisation, une généralisation et une compréhension en profondeur; l'élève doit analyser, identifier les éléments mathématiques d'une situation et

poser ses propres problèmes.

- En troisième lieu, les situations dans lesquelles les mathématiques sont utilisées, qui vont du contexte privé au cadre plus vaste des questions scientifiques ou publiques.

La culture scientifique dans PISA

Dans PISA, la culture scientifique se définit comme « *la capacité d'associer des connaissances scientifiques à la formulation de conclusions fondées sur l'observation et d'élaborer des hypothèses en vue de comprendre le monde naturel et les transformations qui y sont apportées par l'activité humaine, et de contribuer à la prise de décisions à cet égard.* » (OCDE, 1999, p. 15).

PISA évalue la culture scientifique en fonction de trois dimensions :

- Premièrement, *les concepts scientifiques* qui sont nécessaires pour comprendre certains phénomènes du monde naturel et les changements que l'activité humaine y apporte. Les concepts mis en œuvre sont ceux, familiers, qui se rapportent à la physique, à la chimie, aux sciences biologiques, à la Terre et à l'espace; les élèves doivent les appliquer à des problèmes scientifiques existant dans la réalité, et non simplement

les restituer. Pour l'essentiel, le contenu du test est emprunté à trois champs d'application: les questions scientifiques relatives à la vie et à la santé, celles relatives à la Terre et à l'environnement, et celles relatives à la technologie.

- En second lieu, *les démarches scientifiques*. L'accent est mis sur la capacité de recueillir des éléments de preuve, de les interpréter et d'agir en fonction des conclusions tirées. Cinq de ces démarches sont représentées dans PISA :

- ◇ reconnaître les questions auxquelles on peut répondre par une investigation scientifique;
- ◇ identifier les éléments probants;
- ◇ tirer des conclusions;
- ◇ communiquer ces conclusions;
- ◇ faire preuve de sa compréhension des concepts scientifiques.

- Aucun de ces processus – excepté le dernier – ne requiert un corpus préétabli de connaissances scientifiques. Cependant, dans la mesure où aucune démarche scientifique ne peut être «vide de contenu», les questions de PISA portant sur les sciences feront toutes appel à la maîtrise de

notions scientifiques fondamentales.

- En troisième lieu, *les situations scientifiques*, tirées de la vie quotidienne plutôt que de la science telle qu'elle est généralement pratiquée dans une salle de classe ou dans les travaux de professionnels de la science. Comme pour les mathématiques, la science se manifeste dans la vie des individus, depuis la sphère privée jusqu'aux questions d'intérêt public.

En mathématique et en sciences, 35 % des items sont des questions ouvertes.

Exemples d'items

L'ensemble du matériel d'évaluation PISA 2000 ne peut être rendu public, car une partie non négligeable de celui-ci sera réutilisée lors des cycles suivants, pour suivre l'évolution des compétences. Un certain nombre d'entre eux ont toutefois été rendus accessibles, afin d'illustrer la manière dont les trois domaines sont évalués dans PISA. Les exemples d'items, accompagnés de leurs guides de correction, figurent en annexe 2.

Questionnaires de contexte

Comme on l'a vu, les élèves et la direction des écoles participant à PISA doivent répondre à un questionnaire visant à recueillir des informations sur les contextes dans lesquels se déroulent les apprentissages. Le but est de lire les performances des différents pays, mais aussi, à l'intérieur des pays, des élèves et des établissements, à la lumière de toutes les informations recueillies au travers des questionnaires. Celles-ci permettent en effet de relativiser les résultats cognitifs ou de les interpréter d'une manière juste et réfléchie.

Quels sont les différents aspects abordés par ces questionnaires ?

Le milieu social

On sait l'influence que le milieu social d'origine exerce sur l'acquisition des compétences. Vu cette importance, il a été décidé, pour PISA, de tenter de saisir d'une façon complète et précise les caractéristiques de ce milieu. Le concept de « milieu social » recouvre à la fois la position socio-économique des parents et des notions plus larges touchant à l'environnement familial.

Les variables relatives à l'environnement familial sont la taille de la famille, sa structure (famille traditionnelle, mono-

parentale ou recomposée), et les styles parentaux.

Les variables relatives à la position socio-économique englobent :

- l'emploi occupé par les parents : la profession citée par l'élève est codée et convertie en un indice socio-économique international du statut professionnel ;
- le capital culturel : celui-ci est mesuré par le degré de familiarité de l'étudiant avec la culture dominante (littérature classique, poésie, possession d'œuvres d'art ou d'instruments de musique) et par le niveau d'éducation des parents ;
- le capital social : ressources sociales dans la famille (rôle que les parents et les amis peuvent jouer dans le travail scolaire) ;
- le bien-être économique, mesuré au travers de la possession d'une série de biens et d'équipements (ordinateur, téléphones, automobiles, lave-vaisselle...).

L'investissement des étudiants

L'investissement des étudiants par rapport à l'école exerce un rôle important dans la réussite. Il est mesuré dans PISA au travers de questions portant sur la régularité de la fréquentation scolaire, le sentiment d'appartenance à la communauté éducative, l'investissement dans des activités à l'école en

dehors des cours et les relations avec les enseignants.

Les habitudes et les attitudes dans le domaine de la lecture

Comme la lecture est la discipline majeure, des informations ont été recueillies sur le degré d'accès des élèves à l'écrit dans sa famille et en dehors de la famille, sur ses habitudes de lecture (à quelle fréquence lisent-ils et que lisent-ils ?) et sur leurs attitudes par rapport à la lecture (*aiment-ils lire et pourquoi ?*).

Culture et langue d'origine

Des questions ont été posées sur le pays d'origine des parents et sur la (les) langues parlées à la maison.

Ressources humaines et matérielles

Des questions ont été posées sur les ressources matérielles auxquelles les étudiants peuvent avoir accès :

- à la maison : Internet, logiciels éducatifs, pièce tranquille où étudier, manuels et calculatrices,
- à l'école : bibliothèques, ordinateurs, calculatrices et laboratoires de sciences.

Pour ce qui est des ressources humaines, des informations relatives aux enseignants ont été collectées:

existe-t-il suffisamment d'enseignants qualifiés dans les disciplines concernées par PISA, quelles sont les possibilités de développement professionnel ?

Climat de l'école

Le climat qui règne au sein de l'établissement peut aussi affecter d'une façon non négligeable les acquis des élèves. A cet égard, des informations ont été recueillies concernant :

- le degré de pression mis sur les étudiants : attentes et attitudes des professeurs en matière de réussite, rythme de travail, encouragements des professeurs...
- la discipline au sein de l'établissement et des classes,
- les relations élèves-professeurs : les étudiants se sentent-ils soutenus dans leurs efforts, les professeurs sont-ils respectés par les élèves ?
- l'investissement des professeurs dans leur métier : ceux-ci sont-ils dynamiques, fiers d'appartenir à leur établissement ?

Caractéristiques de l'école

Dans une enquête comparative comme PISA, un élément crucial – car il diffère fortement selon les systèmes éducatifs – est constitué par l'organisation de l'école et les parcours de formation offerts par le système et les établissements.

Des informations sont recueillies à propos :

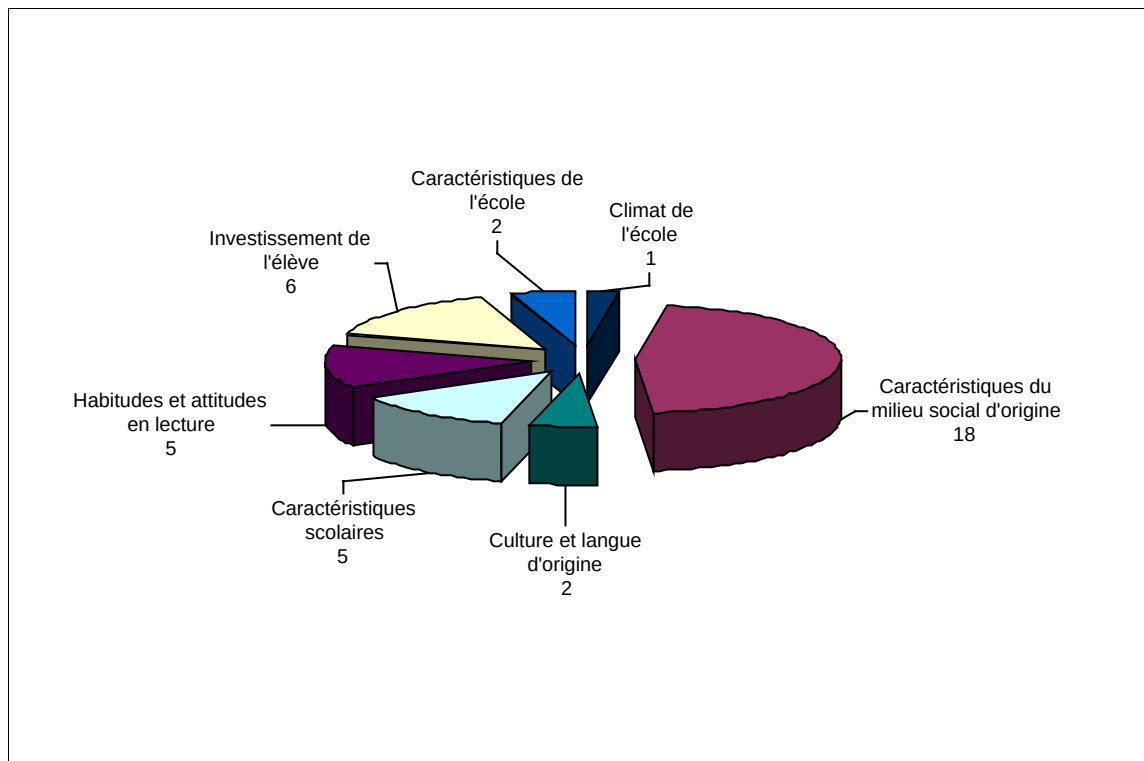
- des filières, formes ou programmes d'études fréquentés par les élèves de 15 ans,
- des sources de financement des établissements,
- de la taille des établissements,
- du temps d'instruction
- du degré d'autonomie dont disposent les établissements et les enseignants.

Politique d'évaluation au sein de l'école

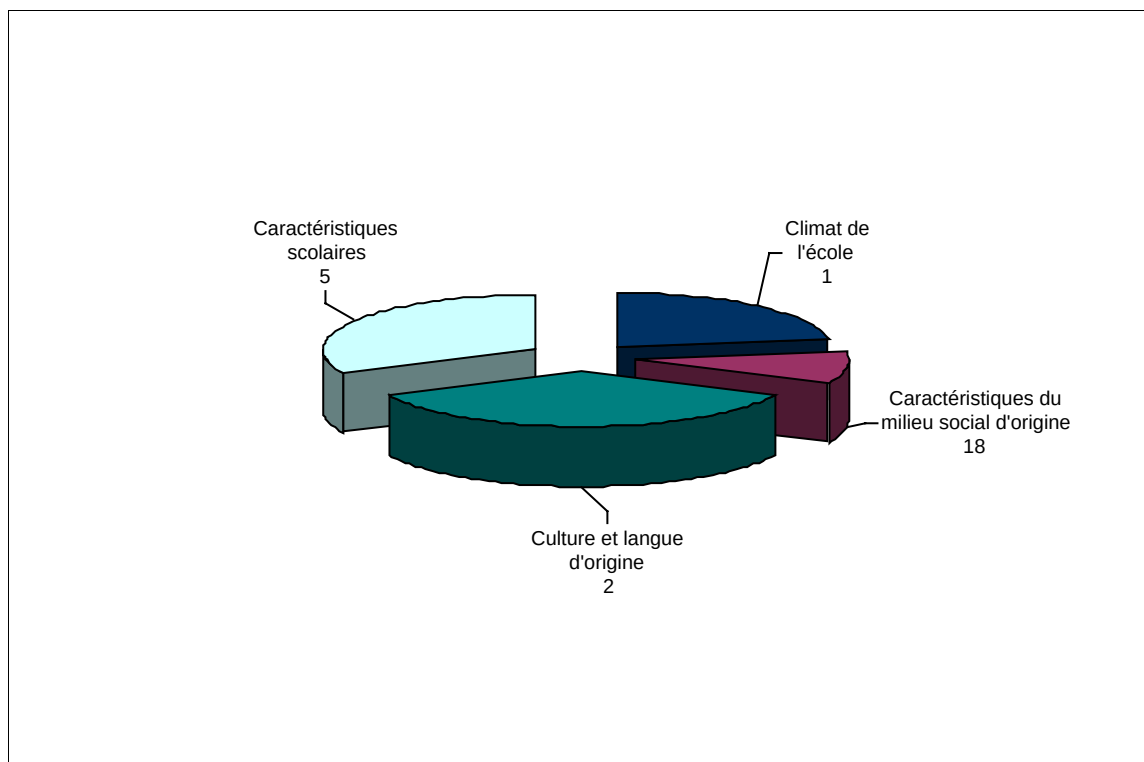
Des informations sont collectées sur la fréquence et la nature des évaluations, le caractère plus ou moins sélectif de la politique des établissements, le niveau d'exigence en matière de réussite en lecture, mathématiques et sciences.

Les deux graphiques ci-dessous illustrent la manière dont les différents domaines d'intérêt sont investis dans les questionnaires PISA.

**Figure 4 : Domaines couverts par le questionnaire Elève
(en nombre de questions)**



**Figure 5 : Domaines couverts par le questionnaire école
(en nombre de questions)**



Design de PISA

Pour rappel, le premier cycle de PISA a pour domaine principal la lecture, les domaines « mineurs » étant les mathématiques et les sciences. Dans les carnets de tests destinés aux élèves, cette répartition se traduit par un volume important de tests consacrés à la lecture, et par une quantité moindre consacrée aux mathématiques et aux sciences. Les carnets comportent également, outre la partie cognitive, des questionnaires contextuels, certains communs à tous les pays et d'autres optionnels.

Chacune des unités de test (texte/support + items) a été attribuée à un bloc d'items. Il existe ainsi 9 blocs de Lecture (R1 – R9), 4 blocs de Mathématiques (M1 – M4) et 4 blocs de Sciences (S1 – S4). Les blocs d'items de lecture ont une durée de passation de 30 minutes chacun, tandis que les blocs d'items de sciences et de mathématiques ont une durée de passation de 15 minutes

chacun. L'ensemble de l'évaluation PISA est riche de quelque 210 items et équivaut à une durée de passation de 390 minutes. Chaque bloc est présenté dans plusieurs carnets et est ainsi administré à un nombre suffisant d'élèves.

Chacun des élèves ne passe bien entendu qu'une partie de l'ensemble de l'évaluation. Les blocs d'items sont répartis en 9 carnets de test selon le plan de rotation présenté ci-dessous. Dans ce tableau, R1 fait référence au bloc de lecture n° 1, et ainsi de suite. Même si chacun des élèves ne passe qu'une partie des items, des méthodes d'analyse sophistiquées (analyse IRT) permettent de déterminer le score qu'il obtiendrait s'il avait passé l'ensemble des items de lecture et rendent ainsi les scores comparables d'un élève à l'autre quel que soit le carnet qu'il a reçu. Tous les carnets comportent un noyau commun d'items (appelés items d'ancrage) au départ desquels cette équivalence est rendue possible.

Tableau 3 : Plan de rotation des carnets

Carnet	30 min.	30 min.	30 min.	30 min.
1	R1	R2	R4	M1/M2
2	R2	R3	R5	S1/S2
3	R3	R4	R6	M3/M4
4	R4	R5	R7	S3/S4
5	R5	R6	R1	M2/M3
6	R6	R7	R2	S2/S3
7	R7	R1	R3	R8

8	M4/M2	S1/S3	R8	R9
9	S4/S2	M1/M3	R9	R8

Dans les six premiers carnets, les épreuves de lecture constituent les trois-quarts des épreuves cognitives et un seul domaine mineur (mathématiques ou sciences) y figure. Le carnet 7 ne comporte que des épreuves de lecture. Les carnets 8 et 9, quant à eux, commencent par des épreuves dans les deux domaines mineurs (mathématiques et sciences), et celles-ci ont une durée de passation égale à celles de lecture. A la fin de chaque carnet, figure le questionnaire Elève de 30 minutes auquel s'ajoute, pour la Communauté française de Belgique, le questionnaire optionnel sur la familiarité avec l'informatique.

La passation complète d'un carnet (partie cognitive et questionnaires) dure trois heures (pauses comprises). L'opération d'évaluation est donc lourde à appliquer dans les établissements – pour lesquels il s'agit de réunir des élèves provenant souvent de classes différentes – et surtout, assez éprouvante pour les élèves.

Pour les élèves de l'enseignement spécial prenant part à l'évaluation, un carnet spécial « allégé » (carnet 10) a été conçu. Bien que plus léger (une heure au lieu de deux), ce carnet, qui comporte des items de lecture, de mathématiques et de sciences, constitue néanmoins une épreuve très difficile pour des élèves inscrits dans ce type d'enseignement.

Echantillonnage

Sélection de l'échantillon d'écoles

L'un des enjeux majeurs en Communauté française de Belgique était d'assurer une participation maximale des établissements scolaires à la campagne de tests PISA. Les contraintes internationales concernant la participation des établissements sont en effet très strictes : si le taux de refus de participation des implantations¹ sélectionnées est trop élevé, le pays concerné risque de ne pas voir ses résultats figurer dans les tableaux internationaux². De plus, il est capital que les établissements scolaires de l'échantillon de base acceptent massivement de participer à l'évaluation car moins la participation de l'échantillon de base est élevée, plus la participation des échantillons de remplacement doit être élevée³.

¹ Pour la Communauté française de Belgique, chaque fois qu'il est fait mention d'une « école » ou d'un « établissement scolaire » ayant participé à l'étude, c'est en fait d'une implantation dont il s'agit.

² C'est ce qui s'est d'ailleurs produit dans le cas des Pays-Bas, dont les résultats ont été écartés des tableaux internationaux en raison du trop faible taux de participation de ses établissements.

³ Une première contrainte veut qu'un minimum de 65 % des implantations de l'échantillon de base acceptent de participer pour que les résultats nationaux soient valides. Dans tous les cas, moins la participation de l'échantillon de base est élevée (si elle n'atteint par exemple que les 65 %), plus la participation des implantations figurant dans les échantillons de remplacement doit être élevée. Pour 65 % de participation de l'échantillon de base, il faut obtenir 95 % d'acceptation pour les échantillons de remplacement ; pour 70 % de participation de l'échantillon de base, il faut 92,5 % d'acceptation des échantillons de remplacement. Le taux à atteindre s'obtient en soustrayant le taux de participation avant

L'expérience de l'essai de terrain nous avait appris que le taux de participation des écoles pouvait être amélioré grâce au soutien direct des différents ministres impliqués ainsi qu'à celui des responsables des différents réseaux d'enseignement. En janvier 2000, une réunion des membres du comité « national / communautaire », des représentants des cabinets ministériels concernés (relations internationales et enseignement secondaire), ainsi que des responsables des différents réseaux de l'enseignement en Communauté française a permis de définir les stratégies propres à encourager la participation des établissements sélectionnés. Il a donc été décidé qu'une lettre de soutien signée par les deux ministres concernés viendrait renforcer l'impact du courrier envoyé par le Service de Pédagogie expérimentale. Comme l'attestent les chiffres présentés ci-dessous, cette stratégie bien orchestrée a largement porté ses fruits, puisqu'elle nous a même permis d'obtenir un taux record de participation des établissements en Communauté française.

Selon la procédure d'échantillonnage mise en place par le centre responsable (Westat), le nombre d'écoles reprises dans l'échantillon principal s'élevait à 103

remplacement de 255 et en divisant le résultat par deux (par exemple pour un taux de 70 % avant remplacement :
 $(255 - 70)/2 = 92,5 \%$).

établissements en Communauté française. Chaque pays participant à PISA doit en effet tester un minimum de 150 écoles. Aux yeux de l'OCDÉ, la Belgique est considérée comme un pays et cette exigence de 150 établissements s'applique donc à l'ensemble du pays. Chacune des communautés linguistiques du pays a augmenté le nombre minimal requis pour obtenir une bonne estimation des résultats en son sein. En Communauté française de Belgique, il a été décidé de tester 100 écoles (plus 3 écoles de l'enseignement spécial). La Communauté flamande a visé un échantillon de 150 établissements. La Communauté germanophone a effectué un recensement (elle a testé tous les établissements secondaires).

L'échantillonnage de PISA s'effectue en deux étapes. Lors d'une première étape, des établissements accueillant des élèves de 15 ans sont sélectionnés. Pour chaque école sélectionnée dans l'échantillon de base, deux écoles de remplacement sont prévues. Les écoles ont une probabilité d'être sélectionnées proportionnelles à leur taille (*i.e.* au nombre d'élèves de 15 ans qui la fréquentent). Dans un deuxième temps, des élèves sont sélectionnés à l'intérieur des écoles qui ont accepté de participer.

Une seule variable de stratification explicite a été utilisée en Communauté française de Belgique : la

taille de l'établissement. Pour refléter correctement la réalité de notre système éducatif, il a été décidé de tester un nombre déterminé de petites écoles (nombre d'élèves de 15 ans inscrits inférieur à 17), d'écoles de taille moyenne (comptant entre 17 et 35 élèves de 15 ans) et d'écoles de grande taille (avec un nombre d'élèves de 15 ans supérieur à 35). L'échantillon définitif comporte, 3 écoles de petite taille, 9 écoles de taille moyenne et 86 écoles de grande taille. Etant donné le mode de tirage de l'échantillon adopté, si un tel critère n'avait pas été défini, les chances des petites et moyennes écoles d'être tirées au sort auraient été trop faibles.

Un autre critère de stratification – implicite celui-là – a été utilisé pour garantir la bonne représentativité de l'échantillon. La liste des écoles dans laquelle on tire l'échantillon a été triée par ordre croissant de taux d'élèves en retard scolaire. Dès lors que l'on tire l'échantillon en procédant dans la liste par « pas » réguliers, cette façon de faire garantit que des écoles bien diversifiées sont tirées au sort. Le pourcentage d'élèves en retard est en effet un critère qui en « exprime » beaucoup d'autres : il est fortement lié à la forme d'enseignement ou au milieu socio-culturel et ethnique d'origine des élèves.

Vers la mi-janvier, ces 103 établissements échantillonnés ont donc reçu simultanément un courrier émanant du Service de Pédagogie expérimentale, qui décrivait brièvement en quoi consiste l'étude PISA, la procédure à suivre en cas d'acceptation, ainsi que la lettre signée par le Ministre-Président, H. Hasquin, et le ministre de l'Enseignement secondaire, P. Hazette.

Toujours grâce à l'expérience de l'essai de terrain, nous savions qu'il était bien plus efficace de recontacter par téléphone les écoles encore hésitantes plutôt que de leur envoyer un rappel par courrier. A la fin du mois de février, nous avons obtenu l'acceptation de 84 écoles sur 103, et un second courrier était envoyé à 18 écoles du 1^{er} échantillon de remplacement (même procédure). Huit écoles supplémentaires ont donné leur accord et, enfin, ce sont 9 écoles sur 12, issues du second échantillon de remplacement, qui ont accepté de participer à la campagne de tests PISA. Ainsi, courriers et échanges téléphoniques réguliers avec les écoles ont permis d'atteindre un total de 101 établissements participant à l'étude, ce qui donne un taux de participation de 98 %. Ceci constitue un record en Communauté française de Belgique, qui avait obtenu des taux de participation faibles (de l'ordre de 55 à 60 % pour l'échantillon de base) pour les deux dernières études

internationales en date (IEA Reading Literacy et TIMSS)¹.

Sélection des élèves à l'intérieur des implantations

La procédure décrite dans le *Manuel du Directeur National de Projet* (manuel NPM) prévoit que les coordinateurs scolaires ou les chefs d'établissement fournissent *la liste complète des élèves nés en 1984 et inscrits dans l'implantation concernée*. Après avoir vérifié ces listes, le logiciel Key Quest² est utilisé pour tirer aléatoirement les échantillons d'élèves. Pour chaque implantation concernée, il faut tirer un échantillon de 35 élèves. Si l'implantation comporte moins de 35 élèves, tous les élèves sont sélectionnés. Dans le manuel NPM figurent des instructions concernant les conditions d'exclusion des élèves échantillonnés, qui ont été insérées dans le manuel du coordinateur scolaire. Une fois que ceux-ci avaient reçu les listes des 35 élèves échantillonnés dans leur implantation, ils étaient priés de vérifier si aucun d'eux ne correspondait à ces

critères d'exclusion³. Le cas échéant, ils étaient priés de nous en faire part aussi tôt que possible, et l'(es) élève(s) concerné(s) étai(en)t purement et simplement exclu(s) de la liste des élèves à tester, sans être remplacé(s). Le taux global

³ Consignes pour l'exclusion des élèves (Extrait du *Manuel NPM*, du *Manuel du Coordinateur scolaire* et du *Formulaire de suivi des Elèves*) :

« Les consignes suivantes déterminent les catégories générales pour l'exclusion des élèves au sein des écoles sélectionnées. Ces consignes doivent être mises en œuvre avec circonspection, en tenant compte du contexte spécifique de chaque système éducatif. Les chiffres figurant à gauche sont des codes qu'il faut inscrire dans la colonne 7 du Formulaire de Suivi des Elèves afin d'identifier les élèves exclus.

1= Elèves handicapés fonctionnels. Il s'agit d'élèves dont le handicap physique permanent est tel qu'ils ne peuvent passer le test PISA dans les conditions fixées. Les élèves handicapés fonctionnels qui peuvent passer le test doivent être inclus.

2 = Elèves souffrant d'un retard mental, mais aptes à l'apprentissage. Il s'agit d'élèves qui sont considérés par le psychologue de l'école, le proviseur, le chef d'établissement ou tout autre agent qualifié comme des élèves éducatibles souffrant d'un retard mental, ou qui ont subi des tests psychologiques et ont été déclarés tels. On trouve parmi ceux-ci les élèves qui sont émotionnellement ou mentalement incapables de suivre même les instructions générales du test. Ne doivent pas être exclus les élèves qui ont simplement des résultats scolaires médiocres ou des problèmes de discipline.

3 = Elèves ayant une connaissance limitée de la langue du test. Il s'agit d'élèves qui sont incapables de lire ou de parler la langue dans laquelle le test a été rédigé et qui seraient incapables de surmonter la barrière de la langue dans les conditions du test. A titre indicatif, un élève ayant reçu une formation de moins d'un an dans la langue du test devrait être exclu ; [...]

4 = Autres.

Il est important que ces critères soient respectés de manière rigoureuse afin que les résultats de l'étude soient comparables à l'intérieur d'un même pays et entre pays. En cas de doute, acceptez l'élève. »

¹ Malheureusement, la Flandre a obtenu de moins bons résultats pour PISA. Les taux de participation étant calculés pour l'ensemble du pays (avec une pondération de 58 % pour la Flandre, 41 % pour la Communauté française et 1 % pour la Communauté germanophone), le taux de participation national ne reflétera pas les records obtenus en Communauté française.

² Nous avons reçu une formation pour utiliser ce logiciel lors du meeting qui s'est tenu à Bruxelles en février 2000.

d'exclusion des élèves est relativement bas, comme l'atteste le tableau à la page suivante. Le taux imposé

de 85 % de participation à l'intérieur des écoles est largement atteint.

Tableau 4 : Nombre d'élèves sélectionnés, exclus, absents et testés

Elèves sélectionnés	Elèves exclus	Elèves absents (y compris pour des raisons de non-applicabilité ou pour un refus des élèves ou de leurs parents)	Elèves testés : nombre total pour les séances de base et les séances de rattrapage (9)
3.200	71 (0,22 %)	311 (9,7 %)	2.818 (90 %)

Déroulement de l'enquête

Principaux acteurs de l'administration des tests dans les écoles

A. Les Coordinateurs scolaires

Le Coordinateur scolaire est un membre de l'équipe enseignante ou administrative d'une école, désigné par le chef d'établissement, qui est chargé de toute une série de tâches logistiques relatives à l'administration des tests dans l'établissement. Dans la majorité des cas, le coordinateur scolaire est le chef d'établissement lui-même ; dans les autres cas, il peut s'agir du sous-directeur, d'un enseignant ou d'un surveillant qu'il a désigné pour remplir cette tâche. Après réception du formulaire officialisant l'acceptation de l'école et la désignation du Coordinateur scolaire, le

Centre national faisait parvenir à ce dernier le *Manuel du Coordinateur scolaire*, qui énumère et décrit en détails les différentes tâches qu'il doit effectuer, le questionnaire-école, ainsi qu'un courrier lui expliquant la première tâche qu'on attend de lui, à savoir fournir au Centre national dans les délais les plus brefs la liste complète de *tous* les élèves de l'implantation nés en 1984. Parfois, ce critère était accueilli avec réticence par certains coordinateurs, qui auraient préféré ne fournir que la liste des élèves d'une même année d'étude. En effet, il est bien plus aisé d'organiser une séance de tests pour des élèves issus d'une même année d'étude, plutôt que pour des élèves de 3^e, de 4^e ou de 5^e année par exemple.

Le rôle du coordinateur scolaire est essentiel au bon déroulement des opérations dans les établissements :

- coordination avec le Centre national pour mettre au point les détails concrets de l'administration de tests dans l'école concernée ;
- choix d'une date pour la passation des tests dans l'école ;
- assistance à l'administrateur de test pendant la passation des tests notamment pour aider à maintenir la discipline pendant la séance.

B. Les Administrateurs de tests

Selon les règles fixées par le consortium international, les personnes chargées de l'administration des tests auprès des élèves doivent impérativement être extérieures aux établissements, et présenter les compétences nécessaires pour exécuter au mieux une telle tâche. Comme pour le prétest, nous avons fait appel à des étudiants en licence (sciences de l'éducation, psychologie et philologie romane). En plus des dix étudiants recrutés, nous avons également formé deux ou trois membres du personnel auxiliaire du Service de Pédagogie expérimentale, afin d'assurer d'éventuels remplacements de dernière minute.

Les futurs administrateurs de tests ont reçu une formation sur les différents aspects de l'évaluation PISA : contexte et objectifs de

l'évaluation, consignes à respecter pour assurer la passation d'une évaluation internationale standardisée, présentation des instruments de test et explication détaillée du contenu du *Manuel de l'administrateur de tests*, mise au point concernant tous les aspects logistiques relatifs à l'administration de tests. Lors de cette formation, on a également passé en revue les questions ou situations fréquemment rencontrées lors du prétest afin de préparer au mieux les administrateurs de tests à réagir aux situations concrètes qu'ils pourraient rencontrer dans les écoles. Outre le *Manuel de l'administrateur de tests*, on leur a également fourni une version du *Questionnaire élève* (repris à la fin de chaque carnet de tests) spécialement rédigée à leur intention, où ils pouvaient trouver des indications précises destinées à répondre aux questions que les élèves ne manqueraient pas de poser.

Une autre tâche incombant aux Administrateurs de tests était d'organiser une séance de rattrapage conjointement avec le Coordinateur scolaire, en cas d'absence de plus de cinq élèves sélectionnés. Cette séance de rattrapage ne concernait que les élèves absents, en aucun cas il ne pouvait être envisagé de remplacer les absents par d'autres élèves, qu'il s'agisse de la première séance ou de celle de rattrapage.

Selon la procédure établie, l'Administrateur de tests, avant de se rendre dans une école, avait déjà pris contact avec le Coordinateur scolaire afin de vérifier avec lui que tous les élèves concernés (c'est-à-dire ceux figurant sur le *Formulaire de Suivi des Elèves*) avaient été prévenus pour la date fixée, qu'un local approprié était retenu, etc.

Durant toute la période d'évaluation, un membre du Service de Pédagogie expérimentale a encadré l'équipe d'Administrateurs de tests. Il a fallu établir un calendrier en fonction des dates d'évaluation choisies par les écoles et les disponibilités des administrateurs et s'adapter aux modifications survenant en cours de campagne. Cette même personne était chargée de vérifier que les administrateurs étaient munis de tout le matériel nécessaire avant de se rendre dans les écoles (nombre de carnets adéquats et correctement étiquetés, *Formulaires de suivi des Elèves* correctement remplis, et ainsi de suite). De même, au retour des administrateurs, il fallait vérifier que l'ensemble du matériel était bien rentré au Centre national (carnets, questionnaires-école, *Formulaires de suivi des Elèves* complétés, rapports de séance, etc...).

C. Les Contrôleurs de qualité (SQM)

Les Contrôleurs de qualité ne peuvent appartenir ni au Centre

national, ni au Centre de coordination international, ni à un des établissements participant à l'évaluation, mais doivent avoir l'autorisation d'entrer dans une école pour observer le bon déroulement d'une séance de passation de tests PISA. En Communauté française de Belgique, les inspecteurs des trois domaines évalués par le programme PISA semblaient les personnes tout indiquées pour remplir cette mission de contrôle de qualité. Il s'agissait en effet de se rendre dans une série d'établissements scolaires sur le point d'accueillir une administration de tests PISA, et d'observer les séances de passation des tests d'un bout à l'autre, en étant particulièrement attentif aux points suivants (respect absolu de la part de l'Administrateur de tests de la procédure imposée par le programme) :

- lecture fidèle des consignes et de la marche à suivre aux élèves,
- respect des temps impartis à chaque étape de la passation,
- non-intervention des membres du personnel de l'école (chef d'établissement, coordinateur scolaire ou tout autre personne) pour aider les élèves à répondre aux tests cognitifs,
- attitude ferme et responsable de la part de l'administrateur de test concernant les consi-

gnes de sécurité qui reposent
sur le matériel cognitif de
PISA.

Le formulaire de suivi des élèves

C'est ainsi que l'on désigne la liste des élèves échantillonnés pour une implantation donnée. Sur ce document, figurent toutes les données essentielles concernant les élèves participant aux tests PISA : nom et code de l'implantation, code attribué à chaque élève, date de naissance et année d'étude. C'est également sur ce formulaire que l'Administrateur de tests va indiquer toutes les données concernant la passation de tests proprement dite pour chaque élève mentionné : a-t-il participé ou non à chaque étape de la passation, en cas d'absence à la première séance, était-il présent ou non à la séance de rattrapage, etc. ... Ce document est un outil de référence particulièrement important pour chaque étape des opérations de terrain PISA jusqu'au nettoyage des données. Pour cette raison, il est essentiel qu'il soit parfaitement complété et préservé jusqu'à la fin de la procédure.

Administration des tests

Tous les questionnaires-école ont été envoyés aux chefs d'établissement avant le jour de la passation des tests et la tâche incombait aux administrateurs de tests de les ramener au Centre national en même temps que les carnets, le jour même de la passation. Dans certains cas, les chefs d'établissement ont omis

de les remplir pour la date prévue, et il a été nécessaire de les rappeler pour leur demander de nous les renvoyer par courrier, une fois complétés.

Plutôt que d'envoyer les carnets par courrier, nous avons préféré les confier directement aux administrateurs de tests. Cela a permis d'économiser du temps, de l'énergie et des moyens, tout en respectant les consignes strictes de confidentialité qui pèsent sur l'ensemble du matériel de test PISA. Les administrateurs de tests étaient tenus de ne laisser derrière eux aucun document PISA, ni d'en laisser faire aucune copie (du moins en ce qui concerne les tests cognitifs), bien que quelques chefs d'établissement en aient manifesté le désir.

Comme nous l'avons déjà mentionné, les tests cognitifs et les questionnaires contextuels (*Questionnaire-élève* et *Questionnaire sur la familiarité avec l'informatique*) étaient administrés en même temps, avec une courte pause entre les deux (durée moyenne de la pause : 7 minutes).

Les *rapports de séance* rédigés par les Administrateurs de tests (pour chaque administration) nous ont permis de nous faire une idée précise des durées moyennes de chaque partie d'une administration, ainsi que du climat général dans lequel s'est déroulé la séance : bonne

communication entre l'école et le Centre national, bonne compréhension des consignes fournies par le Centre et organisation logistique correcte de la séance dans l'établissement (tous les élèves concernés sont présents à l'heure fixée, le local prévu pour la passation est approprié, etc...), ainsi que les réactions des élèves face aux tests (types de questions posées, problèmes de discipline, bonne intelligence avec le coordinateur scolaire, etc...).

Ainsi que nous l'avons déjà signalé, la durée moyenne de la passation totale de l'épreuve était d'environ 3 heures :

- lecture faite par l'administrateur de tests des consignes à appliquer pour répondre au test cognitif (10 minutes)
- première partie du test cognitif (60 minutes)
- pause (5 minutes)
- deuxième partie du test cognitif (60 minutes)
- pause (10 minutes)
- questionnaire-école et questionnaire sur la familiarité avec l'informatique (entre 15 et 45 minutes)

Dans la pratique, le temps imparti au test cognitif fut parfaitement respecté. Les rapports de séance rédigés par les administrateurs de tests nous ont permis de constater que, dans l'ensemble, les indications fournies

par le courrier du Service de Pédagogie expérimentale, le manuel du Coordinateur scolaire et les échanges téléphoniques avec le Service de Pédagogie expérimentale avaient été parfaitement entendues par les personnes responsables de l'organisation de l'épreuve PISA au sein des établissements (chefs d'établissement, coordinateurs scolaires). Dans la plupart des écoles évaluées, les opérations de testing se sont bien déroulées. Les problèmes de discipline et/ou d'absentéisme qui se sont présentés n'étaient pas spécifiquement dus à l'organisation de l'épreuve PISA, mais étaient plutôt caractéristiques de(s) (l') établissement(s) concerné(s).

La grande majorité des séances ont été observées par le Coordinateur scolaire ou un autre membre de l'équipe enseignante de l'école.

Codage et encodage

1. Codage simple et codage multiple

Cette double procédure de correction appliquée aux tests cognitifs fait l'objet de consignes strictes de la part du consortium. Les carnets complétés par les élèves sont répartis en deux parties inégales : trois-quarts d'entre eux font l'objet d'un codage simple, tandis que le quart restant fait l'objet d'un codage multiple.

Codage simple

Cette procédure implique un seul correcteur par domaine évalué et par carnet. Ainsi, un carnet où sont présents les trois domaines évalués est examiné par trois correcteurs différents : un pour la lecture, un pour les mathématiques et un pour les sciences. Le codage simple doit être entièrement terminé avant que ne débute la procédure de codage multiple.

Codage multiple

Cette procédure est nécessaire vu la proportion importante de questions ouvertes dans PISA. Même s'il existe des guides de correction visant à assurer la standardisation, il s'impose de vérifier s'il existe une cohérence (fidélité) entre les différents correcteurs. Les carnets réservés au codage multiple vont donc être examinés par 4 correcteurs différents par discipline, chacun d'eux ignorant bien sûr de quelle façon son prédécesseur a codé tel ou tel item. Pour être valide, cette procédure implique une organisation stricte. Pour chaque carnet, chacun des trois premiers correcteurs pour une discipline donnée indique les codes qu'il attribue à chaque réponse d'élève sur une grille individuelle (par carnet et par correcteur), le quatrième correcteur indiquant ses codes directement dans le carnet.

Comme expliqué précédemment, une partie des carnets du codage multiple fait ensuite l'objet d'une vérification internationale (voir p. 16).

2. Recrutement des correcteurs

Il appartient au NPM de recruter les personnes qui prendront en charge les corrections des questions ouvertes. Là encore, le manuel NPM fournit des consignes strictes concernant les compétences des correcteurs (ils doivent être des « experts » dans un des trois domaines évalués), leur nombre, leur formation et leur rotation sur les différents carnets.

Expertise des correcteurs

Etant donné le niveau de compétences requis par le consortium, nous nous sommes adressés à de jeunes enseignants fraîchement diplômés et en attente d'un poste dans une école ou à de futurs enseignants dans la branche concernée (dernière année de licence).

Nombre de correcteurs

Pour la lecture, il était requis au départ d'engager 8 correcteurs, voire un nombre supérieur, multiple de 8. Cette consigne était particulièrement difficile à respecter en Communauté française, puisque les étudiants en licence se trouvaient à cette époque de l'année (mai-juin) en pleine période d'examens. Avec l'accord du Consortium, nous nous sommes donc contentés de 8 correcteurs en lecture.

Pour les mathématiques et les sciences, nous avons recruté 8 personnes par domaine (comme cela était imposé par le consortium), dont 5 étaient compétentes dans les deux domaines. Ce sont donc 11 personnes qui ont été engagées pour les mathématiques et les sciences.

Formation des correcteurs

La formation des correcteurs s'est déroulée en deux temps :

- a) une formation globale et semblable pour tous, détaillant le contexte, les objectifs et les priorités du programme PISA, ainsi que le type de questions posées dans les carnets (questions fermées, ouvertes, etc. ...), les formats de réponses requis, et la façon d'attribuer les codes prévus.
- b) une formation spécifique par domaine concerné.

Chaque correcteur recevait un manuel de correction reprenant les consignes de correction minutieusement décrites, spécifiques à chaque domaine, illustrées de nombreux exemples, ainsi que la copie de réponses types fournies par des élèves. Au terme d'une première partie, chaque candidat correcteur était prié de corriger seul ces réponses-types. Les résultats de ces corrections étaient ensuite comparés et débattus avec le formateur. Ensuite, au cours de la procédure du

codage simple, les formateurs vérifiaient régulièrement la qualité du codage, en prenant au hasard quelques carnets déjà codés. Lorsqu'un problème se présentait, il était débattu avec tous les codeurs afin qu'une attitude commune et cohérente soit adoptée.

En phase de codage multiple, plus aucun échange ne pouvait avoir lieu entre les différents codeurs.

En ce qui concerne les questionnaires

Le format des réponses aux questionnaires ne nécessitait aucun travail de codage préalable à la saisie des données, si ce n'est pour trois questions du questionnaire-élève : on demandait en effet aux élèves de mentionner la profession de leur père, de leur mère, ainsi que celle qu'ils souhaiteraient exercer plus tard. Dans ces cas précis, il s'agissait bien de réponses ouvertes, l'élève décrivant avec ses mots la/les profession(s) concerné(es). Nous disposions d'une liste de toutes les professions possibles et imaginables, chacune d'elles étant accompagnée d'un code. Nous avons donc recruté une personne pour l'attribution des codes de profession. Ce poste exigeait une bonne connaissance du monde du travail, afin de mettre en relation les réponses des élèves avec la classification internationale CITE (OCDE, 1999).

Saisie et nettoyage des données

Les données ont été encodées via le logiciel *Key Quest*, fourni par le consortium à chaque pays participant pour l'échantillonnage, l'élaboration des formulaires de suivi des élèves, et la saisie des données. Deux membres du Service de Pédagogie expérimentale ont suivi une formation à l'utilisation du logiciel et se sont chargées de la formation des encodeurs pour la Communauté française. Elles ont ensuite procédé aux différentes vérifications prévues par le consortium avant d'envoyer l'ensemble des données à ACER.

Bilan du déroulement de l'enquête sur le terrain

Que pouvons-nous retenir à propos du déroulement de l'enquête PISA ?

- La stratégie bien orchestrée de communication avec les écoles, dont, notamment, le soutien officiel des deux Ministres concernés, Messieurs Hasquin et Hazette, et des fédérations de pouvoirs organisateurs ont largement porté leurs fruits, avec une participation des écoles jamais égalée en Communauté française de Belgique pour une évaluation internationale.
- Un réel esprit de coopération et d'implication dans les écoles lors

de la préparation de l'évaluation et au cours des séances de passation des tests a permis aux différents acteurs de l'évaluation (chefs d'établissement, coordinateurs scolaires, administrateurs de tests, contrôleurs de qualité et élèves) de réaliser l'opération dans d'excellentes conditions et témoigne, entre autres, de l'adhésion progressive du terrain éducatif à l'idée que des évaluations autres que purement internes sont nécessaires et utiles pour un pilotage raisonné du système éducatif¹.

¹ Pour une revue de la littérature sur le pilotage, on consultera *Les cahiers du SPE* n° 5-6, 2001.

PROFILS DE COMPÉTENCES EN LECTURE, MATHÉMATIQUES ET SCIENCES

Echelles et niveaux de compétences

Comme dit précédemment, la présentation des résultats dans PISA ne se limitera pas à l'examen de scores moyens et d'indices de dispersion. Les résultats seront présentés sur des échelles de compétences, établies au départ de « modèles de la réponse à l'item »¹.

Nombre d'échelles

Pour chacun des trois domaines investigués, on disposera d'une échelle générale regroupant l'ensemble des items administrés aux élèves. En outre, pour le domaine majeur (lecture), les résultats seront présentés par sous-domaine, en fonction de l'aspect évalué. On trouvera ainsi, à côté de l'échelle générale, trois sous-échelles spécifiques regroupant les items qui évaluent respectivement la capacité à :

- retrouver de l'information dans un texte²,
- interpréter un texte,

- réfléchir sur un texte.

Ces trois grandes capacités sont définies de la façon suivante :

- *retrouver de l'information dans un texte* : localiser un ou plusieurs éléments d'information dans un texte ;
- *interpréter un texte* : construire le sens et tirer des inférences d'une ou plusieurs parties du texte ;
- *réfléchir sur un texte* : relier le texte à son expérience personnelle et à ses connaissances antérieures relatives au contenu ou à la forme du texte.

On voit que ces trois aspects correspondent assez largement aux notions de *sens littéral*, *sens inférentiel* et *sens personnel*. L'intérêt de ces trois sous-échelles est qu'elles permettent de distinguer les performances des élèves en fonction de grandes catégories de compétences qui sont inégalement investies dans les curriculums visés et implantés des pays participants. Pour la Communauté française de Belgique, on pourrait ainsi s'attendre à ce que les performances des jeunes de 15 ans soient relativement meilleures pour la première échelle (sens littéral) et relativement moins bonnes pour la deuxième et surtout la troisième échelles, compte tenu de l'inégal investissement de ces trois pôles dans les activités d'apprentissage et d'évaluation de la lecture. La décomposition de la compétence

¹ Pour en savoir plus sur ce type d'analyse, on consultera Demeuse (mars 2000).

² Par « texte », il faut entendre aussi bien texte en prose continu que document ou texte « non continu », tel que schéma, graphique, tableau, formulaire etc.

lecture en trois sous-échelles permet donc d'affiner le diagnostic.

Comment sont construites les échelles

L'explication qui suit, au risque d'être simplificatrice, se veut exempte de toute considération technique et mathématique.

Le principe de base des échelles construites sur le modèle de la réponse à l'item est qu'elles permettent d'exprimer dans la même unité (un score sur l'échelle) le niveau de difficulté des items et le niveau de compétence des sujets. Le niveau de compétence des sujets n'est pas un simple pourcentage ou une moyenne de réussite à certains items. Il se définit par la *probabilité* qu'a l'élève de ce niveau d'en réussir chacun des items. Le seuil que l'on se donne pour cette probabilité (50 %, 80 %...) est arbitraire et peut être fixé dans le programme. On verra que dans le présent cas, il a été fixé à 50 % pour des raisons techniques que nous ne développerons pas ici¹.

Dès lors qu'il y a correspondance entre le niveau de difficulté des items et le niveau de compétence des sujets, on peut appréhender le niveau de compétence des sujets de la façon suivante : un sujet se situant par exemple sur l'échelle avec un

score de 500 a une probabilité égale à 50 %² de réussir les items d'une difficulté égale à 500, et une probabilité supérieure à 50 % de réussir les items d'un niveau de difficulté inférieur à 500. Il a en revanche une probabilité inférieure à 50 % de réussir les items d'une difficulté supérieure à 500. Cette probabilité de réussite est établie en tenant compte de la manière dont le sujet a effectivement répondu aux items.

Sur ce type d'échelle, les items sont donc ordonnés sur un continuum, du plus simple au plus complexe, et il en va de même des sujets, ordonnés du moins compétent au plus compétent, comme l'illustre la figure suivante.

¹ La signification pratique à donner à ce seuil de 50 % sera explicitée un peu plus loin, lorsque l'on évoquera les niveaux de compétences.

² Prenons le seuil utilisé dans PISA, mais ce seuil – rappelons-le – peut varier.

Figure 6 : Exemple d'échelle de compétences

SCORES	SUJETS	N° des items				
		48				
6.0						
Plus compétents						Plus complexe
5.0		65				
4.0		59				
	X	39	58			
	X	63				
3.0		28				
	XX	44	50			
	XX	61				
	XXXXX	17	35	60	62	
	XXXXX	38				
	XXXXX	20	51			
2.0	XXXXXXXXXXXXX	16	33	47	49	
	XXXXXXX	10	37	53	54	
	XXXXXXXXXXXXX	19	43	52	57	
	XXXXXXXXXX	3	9	18	30	36
	XXXXXXXXXXXXXXXXXXXX	21	32	56		
1.0	XXXXXXXXXXXXXXXXXXXX	40	41			
	XXXXXXXXXXXXX	15	42			
	XXXXXXXXXXXXX	23	25	34	46	64
	XXXXXXXXXXXXX	11	31	45		
	XXXXXXX	8	24	55		
	XXXXXXXXXXXXX	7	14	29		
0.0						
	XXXX	22				
	XXXX	2				
	XXXX	5	6	27		
	XX	12	26			
-1.0		1				
		4				
		13				
-2.0						
Moins compétents						Moins complexe
-3.0						

Comment sont définis les niveaux de compétence

Dans un premier temps, le modèle de la réponse à l'item débouche sur le type de formalisation des résultats illustré par la figure précédente. Intervient alors un difficile travail d'interprétation de l'échelle généralement confié à des experts du

domaine évalué¹. Pour rendre la lecture de l'échelle intelligible, il importe en effet de constituer des blocs d'items (proches sur l'échelle) qui permettront de définir des niveaux de compétences atteints par les sujets. Tout ce travail repose

¹ Pour PISA, ce travail a été confié au groupe fonctionnel d'experts en lecture et a ensuite été soumis pour avis aux gestionnaires nationaux de projet et pour approbation au BPC.

certain sur un certain arbitraire, éclairé toutefois par les connaissances théoriques des experts. C'est ainsi que le nombre de niveaux que l'on choisit de définir et les endroits de coupure sont arbitraires : on peut définir 3, 4, ou 5 niveaux. Dans le cas de PISA, le choix a été fait de définir, pour les échelles de lecture, 5 niveaux. Pour les échelles en mathématiques et en sciences – domaines mineurs –, aucun niveau n'est défini, compte tenu notamment du trop faible nombre d'items.

Une fois définis le nombre de niveaux et les endroits de coupure, les experts examinent les items caractéristiques de chacun des niveaux et s'attachent à identifier les compétences évaluées par cet ensemble d'items et, en particulier, ce qui fait que les items du niveau 1 sont moins difficiles que ceux du niveau 2 par exemple. Ce minutieux travail d'interprétation consensuelle entre les experts suppose certes une démarche empirique, mais fait aussi appel aux considérations théoriques qui ont présidé à l'élaboration du test. Pour PISA, un cadre théorique de l'évaluation a en effet été élaboré, dans lequel certaines hypothèses sur ce qui fait la complexité du matériel de lecture sont posées au départ : types de texte (continu/non continu), démarches ou processus cognitifs, degré d'abstraction de l'information à repérer, type de correspondance entre la question et le texte (locali-

ser, intégrer, générer de l'information), présence ou non d'éléments distracteurs dans le texte, modalités de réponse... On retrouve évidemment ces principes directeurs de l'élaboration de l'outil au fondement de la hiérarchie des items.

Au terme de ce long processus, on aboutit donc à des échelles où des proportions données de sujets sont situées à un niveau de compétence (de 1 à 5) décrit en termes qualitatifs. Les échelles pour la lecture sont présentées ci-dessous : on n'y trouvera bien entendu que la description des niveaux, puisque les résultats des pays sont encore sous embargo.

Sur ces échelles, les scores sont standardisés : la moyenne est égale à 500 et l'écart-type à 100.

Les niveaux de compétence sont cumulatifs : un étudiant situé au niveau de compétence 3 possède - à un plus haut niveau - les compétences de niveaux 1 et 2. Il n'est pas non plus absolument incapable de réussir les items de niveau 4, mais sa probabilité de réussir les items de niveau 4 est trop faible par rapport au seuil fixé (inférieure à 50 %). Ainsi, un élève particulier se situant au niveau 3 aura par exemple réussi 65 à 70 % des items de niveau 3, 80 % des items de niveau 2 et 95 % des items de niveau 1. En revanche, il n'aura réussi que 40 % des items du niveau 4. C'est la

raison pour laquelle on a considéré qu'il fallait le situer au niveau 3.

Au delà de l'exemple et plus précisément en termes statistiques :

- Tous les étudiants d'un niveau donné sont supposés répondre correctement à au moins la moitié des items de ce niveau.
- Les étudiants situés dans la partie inférieure d'un niveau ont une probabilité de 62 % de réussir les items les plus faciles du niveau et une probabilité égale à 42 % de réussir les items les plus difficiles du niveau.
- Les étudiants situés dans la partie supérieure sont capables de répondre correctement à environ 70 % de l'ensemble des items du niveau. Ils ont une probabilité égale à 62 % de réussir les items les plus difficiles du niveau et une probabilité égale à 78 % de réussir les items les plus faciles du niveau.
- Les étudiants tout en haut d'un niveau (à la limite du niveau supérieur) n'ont pu réussir à au moins 50 % les items du niveau supérieur.

Les étudiants dont les scores sont inférieurs au niveau 1 (niveau '0') ne se sont pas montrés capables de réussir environ un item sur deux des items les plus faciles de PISA. Cela ne signifie pas qu'ils sont totalement incompétents en lecture, mais cela laisse supposer qu'ils ne possèdent

pas les compétences de littératie suffisantes pour pouvoir se servir de la lecture comme d'un outil pour apprendre et acquérir des connaissances dans d'autres domaines.

Que mesurent les trois sous-échelles en lecture par niveau ?

Retrouver de l'information	Interpréter le texte	Réfléchir sur le texte
----------------------------	----------------------	------------------------

Ce qui est mesuré par chacune des trois échelles

Capacité à localiser un ou plusieurs éléments d'information dans un texte.

Capacité à construire le sens et à tirer des inférences d'une ou plusieurs parties d'un texte.

Capacité à relier le texte à son expérience personnelle et à ses connaissances antérieures relatives au contenu ou à la forme du texte

Caractéristiques de la tâche associées à une augmentation de la difficulté pour chacune des trois échelles

La difficulté des tâches dépend :

- du nombre d'éléments d'information à localiser ;
- du nombre de conditions à satisfaire pour localiser l'information requise ;
- du fait que les opérations doivent être ou non effectuées dans un certain ordre ;
- du caractère saillant ou non de l'information ;
- de la familiarité du contexte ;
- de la complexité du texte ;
- de la présence et de la force d'informations distractrices dans le texte.

La difficulté des tâches dépend :

- du type d'interprétation requise (les tâches les plus simples demandent d'identifier l'idée principale d'un texte, les tâches un peu plus difficiles supposent la compréhension des relations entre les parties d'un texte et les plus complexes requièrent la compréhension d'éléments linguistiques en contexte, ou un raisonnement analogique) ;
- du degré avec lequel le texte explicite les informations requises pour effectuer la tâche (plus le texte est implicite sur ce point, plus le niveau d'inférence requis est élevé) ;
- du caractère saillant ou non de l'information ;
- de la présence et de la force d'informations distractrices ;
- de la longueur et de la complexité du texte,
- de la familiarité du contenu du texte.

La difficulté des tâches dépend :

- du type de réflexion requis (les tâches les plus aisées demandent simplement de mettre en relation le texte et l'expérience ou les connaissances extérieures au texte, et les tâches les plus complexes requièrent que soient émises des hypothèses ou qu'une évaluation critique soit effectuée) ;
- de la familiarité des connaissances extérieures auxquelles il faut faire appel ;
- de la complexité du texte ;
- du niveau de compréhension du texte demandé ;
- du degré de clarté des consignes orientant le lecteur vers les facteurs à prendre en compte dans le texte et dans la tâche.

Niveau

	RETROUVER DE L'INFORMATION	INTERPRETER	REFLECHIR SUR LE TEXTE
1	Ne doit prendre en compte qu'un seul critère pour localiser un ou plusieurs éléments d'information explicitement cités dans le texte.	Identifie le thème principal ou l'intention de l'auteur dans un texte portant sur un contenu familier, quand l'information est saillante.	Etablit un simple lien entre l'information du texte et des connaissances communes, de la vie de tous les jours.
2	Localise un ou plusieurs éléments d'information, chacun d'entre eux pouvant requérir que soient rencontrés plusieurs critères. Réussit à gérer de l'information distractive.	Identifie l'idée principale d'un texte, comprend les relations, construit ou applique des catégories simples, ou construit la signification d'une portion limitée du texte quand l'information n'est pas saillante et qu'un faible niveau d'inférence est requis.	Etablit des liens ou une comparaison entre le texte et des connaissances extérieures, ou explique un trait du texte en se fondant sur son expérience personnelle et ses attitudes.
3	Localise et, dans certains cas, identifie le lien entre des éléments d'information, chacun d'entre eux pouvant requérir que soient rencontrés plusieurs critères. Réussit à gérer de l'information fortement distractive.	Etablit des liens ou des comparaisons, fournit des explications ou évalue un trait d'un texte. Fait montre d'une compréhension détaillée d'un texte en lien avec des connaissances familières, de la vie de tous les jours, ou s'appuie sur des connaissances moins communes.	Etablit des liens ou des comparaisons, fournit des explications ou évalue un trait d'un texte. Fait montre d'une compréhension détaillée d'un texte en lien avec des connaissances familières, de la vie de tous les jours, ou s'appuie sur des connaissances moins communes.
4	Localise, agence et combine de nombreux éléments d'information, chacun d'entre eux pouvant requérir que soient rencontrés plusieurs critères, dans un texte à la forme ou au contenu peu familier. Infère quelle information du texte est pertinente pour la tâche.	Met en œuvre un haut niveau d'inférences fondées sur le texte afin de comprendre et d'appliquer des catégories dans un contexte non familier, et pour construire la signification d'une section du texte en tenant compte du texte dans son ensemble. Traite les ambiguïtés, les idées contraires aux attentes ou exprimées en termes négatifs.	Utilise des connaissances formelles ou courantes pour émettre des hypothèses à propos d'un texte ou l'évaluer de façon critique. Fait montre d'une bonne compréhension d'un texte long et complexe.
5	Localise, agence et combine de nombreux éléments d'information profondément enchassés, certains pouvant se trouver en	Elabore avec nuance la signification nuancée d'éléments linguistiques ou fait montre d'une compréhension approfondie	Evalue de façon critique ou émet des hypothèses en se fondant sur des connaissances extérieures au texte

dehors du corps principal du texte. Infère quelle information du texte est pertinente pour la tâche. Réussit à gérer des informations distrayantes hautement plausibles.	du texte dans ses détails.	spécialisées. Traite des concepts contraires aux attentes et fait montre d'une compréhension approfondie d'un texte long et complexe.
--	----------------------------	---

SYNTHÈSE

Le programme pour le suivi des acquis des élèves (PISA), par son ampleur et plusieurs de ses caractéristiques, marque une rupture par rapport aux enquêtes internationales qui l'ont précédé.

En guise de conclusion, nous rappellerons quelques-uns de ces aspects distinctifs.

1. En termes d'orientation et de caractéristiques du programme

- PISA couvre trois domaines, un majeur (la compréhension de l'écrit pour 2000) et deux mineurs (la culture mathématique et scientifique).

Cette particularité présente deux avantages :

- elle permettra de mieux comprendre les liens complexes qui unissent les compétences en lecture et les compétences en mathématiques et en sciences. On évoque souvent, face aux difficultés que rencontrent les étudiants en résolution de problèmes notamment, l'insuffisance de leurs compétences en lecture. Qu'en est-il réellement ? Le design de l'étude PISA a été conçu pour permettre une meilleure

compréhension de cette problématique cruciale.

- elle permettra de suivre de 3 en 3 ans l'évolution des compétences des élèves de 15 ans dans ces trois domaines d'une manière particulièrement rigoureuse.

À l'heure où des réformes importantes (instauration des référentiels de compétences, révision des programmes) achèvent de se mettre en place en Communauté française de Belgique, le programme PISA peut apporter des éléments d'information pour le pilotage du système éducatif qui ne sont pas accessibles par d'autres sources¹.

- Grâce à un schéma de rotation de carnets, PISA permet de tester un nombre considérable d'items sans alourdir à l'excès la tâche que cela représente pour un élève. La mesure obtenue est particulièrement solide : elle permet d'évaluer un éventail de compétences diversifié, en multipliant les angles d'approche – du moins en ce qui concerne le domaine majeur.
- Contrairement aux enquêtes comparatives qui l'ont précédé, PISA utilise, à côté des Q.C.M.,

¹ Les évaluations externes interréseaux apportent des éléments d'information complémentaires.

une proportion importante de questions ouvertes.

- Des efforts considérables ont été déployés pour augmenter la qualité et la comparabilité des données entre pays. Des contrôles de qualité ont été mis en place visant à garantir
 - que les traductions soient aussi équivalentes que possible d'une langue à l'autre ;
 - que les tests soient administrés d'une façon semblable dans les différents pays ;
 - que les questions ouvertes soient corrigées d'une façon uniforme et équitable entre pays et à l'intérieur des pays.

Par rapport aux études comparatives du début des années 90, le rehaussement des standards de qualité est indéniable. Ce bond en avant apparaissait en effet nécessaire pour répondre aux critiques dont ce type d'enquête a fait l'objet par le passé. Il a cependant un prix : les exigences qualitatives nécessitent la mise en place de procédures coûteuses en temps et en moyens humains dont l'augmentation des cotisations internationales et le budget national nécessaire pour mener à bien l'enquête dans chaque pays sont le reflet le plus immédiat.

- PISA, à la différence des études menées sous les auspices de l'I.E.A., n'est pas une étude de

rendement scolaire classique. Son regard est davantage tourné vers l'avenir des élèves que vers leurs acquisitions proprement scolaires. Ceci explique pourquoi PISA, à la différence des études I.E.A., ne se préoccupe guère de l'adéquation des tests aux curriculums et pourquoi il envisage les mathématiques et les sciences sous un angle plus « culturel » (en anglais on parle de « *literacy* »). L'accent est délibérément placé sur le « bagage » que possèdent les élèves de 15 ans à l'heure où ils s'apprêtent à quitter la scolarité obligatoire – définitivement dans certains pays, la scolarité obligatoire à temps plein pour ce qui nous concerne.

Ceci explique aussi les caractéristiques particulières de la population de référence – les élèves de 15 ans où qu'ils en soient dans leur parcours scolaire –. A partir de cet âge, les élèves peuvent en effet, dans la plupart des pays participants, sortir du système d'instruction obligatoire à temps plein et c'est le bagage possédé par les élèves à ce moment que PISA entend précisément évaluer.

2. En termes de résultats attendus

PISA ne se contentera pas de produire un classement des différents pays pour les trois

domaines. À côté du palmarès, il est prévu :

- de présenter les résultats des élèves sur des échelles où différents niveaux de compétences seront définis d'une manière qualitative et concrète. Chaque pays pourra ainsi appréhender quelles proportions d'élèves de 15 ans atteignent les différents paliers de compétences définis et tirer de l'évaluation des informations non seulement comparatives, mais à caractère diagnostique pour l'ensemble des élèves ou pour certains groupes d'élèves particuliers.
- de publier une série de rapports thématiques approfondis qui développeront des analyses internationales traitant de questions essentielles de politique éducative. Grâce aux informations recueillies via le questionnaire Elève et le questionnaire au chef d'établissement, on dispose d'informations riches sur le contexte scolaire et familial dans lequel évoluent les élèves. La mise en relation de ces différentes variables de contexte avec les performances des élèves devraient notamment permettre de mieux répondre aux questions suivantes :
 - La relation entre milieu socio-culturel d'origine et les compétences est-elle d'ampleur variable selon les pays ? Certains pays réussissent-ils mieux que d'autres à modérer le poids de l'origine sociale ? Si oui, pourquoi et comment ?
 - Comment le milieu socio-culturel d'origine, les pratiques culturelles dans la famille et la motivation s'agencent-elles pour produire la réussite – ou l'échec ?
 - Quelle est l'ampleur des différences garçons-filles dans les trois domaines et sur le plan des attitudes, de la motivation et des aspirations pour l'avenir ?
 - Quelle est l'ampleur des différences de compétences en fonction de l'établissement fréquenté ? Certains systèmes éducatifs sont-ils plus équitables que d'autres ?
 - Les systèmes éducatifs à filières sont-ils plus efficaces et/ou plus équitables que ceux organisés en tronc commun ?

RÉFÉRENCES

- Baye, A. (2000). De Consensus en Blocus. *La gestion des spécificités nationales, linguistiques et culturelles dans les évaluations internationales de la lecture*. Mémoire dans le cadre du D.E.A. en Relations internationales et intégration européenne. Ulg : Faculté de Droit.
- Service de Pédagogie expérimentale (2001). Les Cahiers du SPE, n° 5-6.
- Demeuse, M. (2000). *Les échelles de mesure : Thurstone, Likert, Guttman et le modèle de Rasch*. Série « Notes techniques » du Service de Pédagogie expérimentale.
- Lafontaine, D. (2001). Quoi de neuf en littératie ? Regard sur trente ans d'évaluation de la lecture. *Didactique des langues romanes. Le développement des compétences chez l'apprenant*. Bruxelles : De Boeck, 67-83.
- OCDE et Statistique Canada (2000). *La littératie à l'ère de l'information*. Rapport final de l'Enquête Internationale sur la Littératie des Adultes. Paris : OCDE et Ministère de l'Industrie au Canada.
- OCDE (1995). *Littératie, Economie et Société*. Paris : OCDE et Canada : Ministère de l'Industrie.
- OCDE (1999). *Nomenclature des systèmes d'éducation. Guide d'utilisation de la CITE-97 dans les pays de l'OCDE*. Paris : OCDE.
- OCDE (2000). *Mesurer les connaissances et compétences des élèves – Lecture, mathématiques et science : l'évaluation de PISA 2000*. Paris : OCDE.
- OCDE (1999). *Mesurer les connaissances et compétences des élèves – Un nouveau cadre d'évaluation*. Paris : OCDE.
-

TABLE DES MATIERES

VUE D'ENSEMBLE DU PROGRAMME PISA :	1
EN QUOI CONSISTE PISA ?	1
EN QUOI PISA EST-IL DIFFÉRENT DES ENQUÊTES INTERNATIONALES ANTÉRIEURES ?	4
QUI EST CONCERNÉ PAR L'ENQUÊTE ?	5
QU'ÉVALUE PISA ?	5
QU'APPORTERA PISA ?	7
COMMENT EST ORGANISÉE L'ÉVALUATION ?	8
 ASPECTS MÉTHODOLOGIQUES ET DÉROULEMENT DE PISA	 11
PAR QUI EST CONDUIT PISA ?	11
QUELLES GARANTIES AVONS-NOUS QUE LA COMPARAISON EST VALIDE ?	14
<i>Les contrôles de qualité</i>	15
<i>L'équivalence linguistique et culturelle</i>	19
<i>Adéquation du matériel de test aux différents contextes nationaux</i>	20
<i>Equivalence des traductions</i>	19
<i>La marge de manoeuvre des pays : options internationales et nationales</i>	22
CARACTÉRISTIQUES DU MATÉRIEL D'ÉVALUATION	24
<i>La compréhension de l'écrit dans PISA</i>	24
<i>La culture mathématique dans PISA</i>	25
<i>La culture scientifique dans PISA</i>	26
<i>Exemples d'items</i>	27
<i>Questionnaires de contexte</i>	27
DESIGN DE PISA	31
ÉCHANTILLONNAGE	33
<i>Sélection de l'échantillon d'écoles</i>	33
<i>Sélection des élèves à l'intérieur des implantations</i>	36
DÉROULEMENT DE L'ENQUÊTE	37
<i>Principaux acteurs de l'administration des tests dans les écoles</i>	37
<i>Le formulaire de suivi des élèves</i>	41
<i>Administration des tests</i>	41
<i>Codage et encodage</i>	42
<i>Saisie et nettoyage des données</i>	46

<i>Bilan du déroulement de l'enquête sur le terrain</i>	<i>46</i>
---	-----------

PROFILS DE COMPÉTENCES EN LECTURE, MATHÉMATIQUES ET SCIENCES... 47

ECHELLES ET NIVEAUX DE COMPÉTENCES	47
--	----

<i>Nombre d'échelles</i>	<i>47</i>
--------------------------------	-----------

<i>Comment sont construites les échelles</i>	<i>48</i>
--	-----------

SYNTHÈSE..... 52

RÉFÉRENCES 55

ANNEXES

LISTE DES MEMBRES DES COMITÉS PISA EN COMMUNAUTÉ FRANÇAISE DE BELGIQUE

REPRÉSENTANTS DES MEMBRES DES TROIS COMITÉS « DISCIPLINAIRES »

LISTE DES MEMBRES DU COMITÉ DE MATHÉMATIQUES

Enseignement de la Communauté française :

Monsieur René MARQUET

Monsieur Robert HEINE

Enseignement catholique :

Madame F. VAN DIEREN-THOMAS

Monsieur Philippe TILLEUL

Enseignement subventionné officiel :

Madame Brigitte DELCOURT

Monsieur Marc VERDEBOUT

Enseignement subventionné libre non confessionnel :

Madame Françoise DAUBANTON

Madame Christine KINT

Inspection :

Monsieur Carlo BENEDETTI

LISTE DES MEMBRES DU COMITÉ DE SCIENCES

Enseignement de la Communauté française :

Monsieur Jean-Paul CHAPELLE

Monsieur Serge SEMAL

Enseignement catholique :

Madame Simone BERTRAND

Monsieur Michel CASTELBLUM

Enseignement subventionné officiel :

Monsieur J.-P. RAVASI

Monsieur M. VERDEBOUT

Enseignement subventionné libre non confessionnel :

Monsieur Christophe LEYS

Madame Anne PIRON

Inspection :

Monsieur Jacques FURNEMONT

Monsieur Philippe DELFOSSE

LISTE DES MEMBRES DU COMITÉ « LANGUE D'ENSEIGNEMENT »

Expert :

Monsieur Georges LEGROS

Enseignement de la Communauté française :

Monsieur Michel LIEMANS

Monsieur Claude DELVAUX

Enseignement catholique :

Madame Françoise JOIRET-DARVILLE

Monsieur Thierry HULHOVEN

Enseignement subventionné officiel :

Madame N. COLEN

Madame N. DE MAYER

Enseignement subventionné libre non confessionnel :

Madame M.F. STRENS

Madame Béatrice FRANCE

Inspection :

Monsieur Michel BAAR

Monsieur Jean-Paul HOGENBOOM