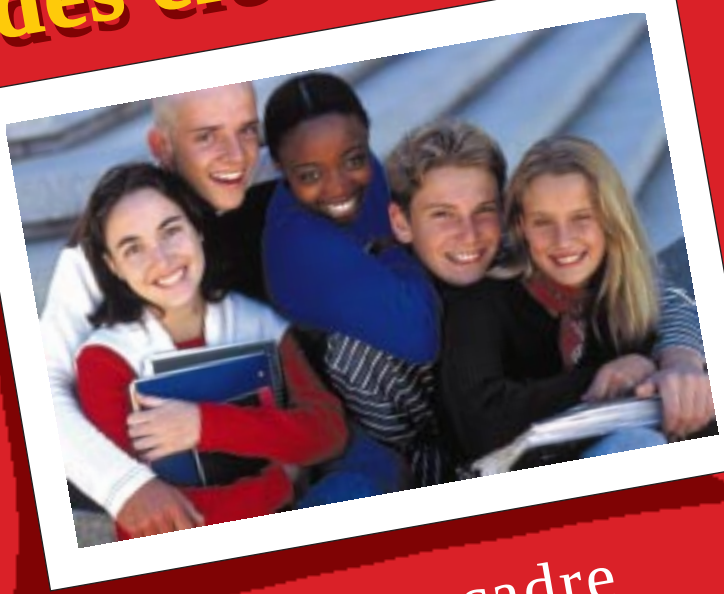
ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

Mesurer les connaissances et compétences des élèves



Un nouveau cadre d'évaluation

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

© OCDE, 1999
© Logiciel, 1987-1996, Acrobat, marque déposée d'ADOBE.
Tous droits du producteur et du propriétaire de ce produit sont réservés. L'OCDE autorise la reproduction d'un seul exemplaire de ce programme pour usage personnel et non commercial uniquement. Sauf autorisation, la duplication, la location, le prêt, l'utilisation de ce produit pour exécution publique sont interdits. Ce programme, les données y afférant et d'autres élémentsdoivent donc être traités comme toute autre documentation sur laquelle s'exerce la protection par le droit d'auteur.

Les demandes sont à adresser au :

Chef du Service des Publications,
Service des Publications de l'OCDE,
2, rue André-Pascal,
75775 Paris Cedex 16, France

MESURER LES CONNAISSANCES ET COMPÉTENCES DES ÉLÈVES

Un nouveau cadre d'évaluation

ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

En vertu de l'article 1er de la Convention signée le 14 décembre 1960, à Paris, et entrée en vigueur le 30 septembre 1961, l'Organisation de Coopération et de Développement économiques (OCDE) a pour objectif de promouvoir des politiques visant :

- à réaliser la plus forte expansion de l'économie et de l'emploi et une progression du niveau de vie dans les pays Membres, tout en maintenant la stabilité financière, et contribuer ainsi au développement de l'économie mondiale ;
- à contribuer à une saine expansion économique dans les pays Membres, ainsi que les pays non membres, en voie de développement économique ;
- à contribuer à l'expansion du commerce mondial sur une base multilatérale et non discriminatoire conformément aux obligations internationales.

Les pays Membres originaires de l'OCDE sont : l'Allemagne, l'Autriche, la Belgique, le Canada, le Danemark, l'Espagne, les États-Unis, la France, la Grèce, l'Irlande, l'Islande, l'Italie, le Luxembourg, la Norvège, les Pays-Bas, le Portugal, le Royaume-Uni, la Suède, la Suisse et la Turquie. Les pays suivants sont ultérieurement devenus Membres par adhésion aux dates indiquées ci-après : le Japon (28 avril 1964), la Finlande (28 janvier 1969), l'Australie (7 juin 1971), la Nouvelle-Zélande (29 mai 1973), le Mexique (18 mai 1994), la République tchèque (21 décembre 1995), la Hongrie (7 mai 1996), la Pologne (22 novembre 1996) et la Corée (12 décembre 1996). La Commission des Communautés européennes participe aux travaux de l'OCDE (article 13 de la Convention de l'OCDE).

Also available in English under the title:
MEASURING STUDENT KNOWLEDGE AND SKILLS
A New Framework for Assessment

Crédit photo : PIX/Denis Boissavy

© OCDE 1999

Les permissions de reproduction partielle à usage non commercial ou destinée à une formation doivent être adressées au Centre français d'exploitation du droit de copie (CFC), 20, rue des Grands-Augustins, 75006 Paris, France, tél. (33-1) 44 07 47 70, Fax (33-1) 46 34 67 19, pour tous les pays à l'exception des États-Unis. Aux États-Unis, l'autorisation doit être obtenue du Copyright Clearance Center, Service Client, (508)750-8400, 222 Rosewood Drive, Danvers, MA 01923 USA, ou CCC Online : http://www.copyright.com/. Toute autre demande d'autorisation de reproduction ou de traduction totale ou partielle de cette publication doit être adressée aux éditions de l'OCDE, 2, rue André-Pascal, 75775 Paris Cedex 16, France.

AVANT-PROPOS

Avec le Programme international de l'OCDE pour le suivi des acquis des élèves (PISA), les gouvernements des pays de l'OCDE s'engagent à suivre, au sein d'un cadre international commun, l'évolution des résultats des systèmes éducatifs en se fondant sur les connaissances des élèves. Le programme PISA est avant tout une initiative concertée regroupant des compétences scientifiques des pays participants et dirigée conjointement par leurs gouvernements dans le but de répondre à des préoccupations communes touchant l'action des pouvoirs publics. Les pays participants sont responsables du projet au niveau de l'action gouvernementale. Des experts des pays participants sont également membres de groupes de travail chargés d'établir un lien entre les objectifs d'action de PISA et les meilleures compétences disponibles, aussi bien au niveau général qu'au niveau technique, dans le domaine des évaluations comparatives internationales. En collaborant aux travaux de ces groupes d'experts, les pays veillent à ce que les instruments d'évaluation utilisés dans le cadre de PISA soient valides au plan international et prennent en compte le contexte culturel et éducatif des pays de l'OCDE, à ce qu'ils se fondent sur des méthodes de mesure rigoureuses et à ce qu'ils mettent l'accent sur la fiabilité des données et leur validité sur le plan éducatif.

Les enquêtes sont conçues et réalisées sous la direction du Secrétariat de l'OCDE, par l'intermédiaire d'un consortium international dirigé par l'Australian Council for Educational Research (ACER). Sont également membres de ce consortium l'Institut national des Pays-Bas pour l'évaluation de l'enseignement, le Service de pédagogie expérimentale de l'Université de Liège et WESTAT.

Cette publication présente le cadre conceptuel sur lequel se fondent les évaluations réalisées aux fins du programme PISA : elle décrit les domaines à évaluer ainsi que l'objet des évaluations et examine la façon et dans quels contextes et limites celles-ci sont effectuées. La publication est publiée sous la responsabilité du Secrétaire général de l'OCDE.

OCDE 1999

REMERCIEMENTS

Les cadres d'évaluation du programme PISA ont été mis au point par des groupes d'experts placés sous la direction de Raymond Adams, de l'ACER. Le groupe d'experts chargé de la lecture était présidé par le Professeur Irwin Kirsch, de l'Educational Testing Service, le groupe chargé des mathématiques était présidé par le Professeur Jan de Lange, de l'Université d'Utrecht, et le groupe d'experts chargé des sciences était présidé par le Professeur Wynne Harlen, du Scottish Council for Research in Education. La liste des membres des groupes fonctionnels d'experts figure en annexe 1. Les cadres d'évaluation ont également été mis au point sur la base de recherches menées par des groupes d'experts dans chacun des pays participants. Ils ont été adoptés en décembre 1998 par les gouvernements des pays de l'OCDE, par l'intermédiaire du Conseil des pays participants. La publication a été élaborée par la Division des statistiques et indicateurs de la Direction de l'éducation, de l'emploi, du travail et des affaires sociales, sous la responsabilité d'Andreas Schleicher.

OCDE 1999

TABLE DES MATIÈRES

Introduction 7

La structure du programme OCDE/PISA 2000 11
Structure générale du programme OCDE/PISA 11
Originalité du programme OCDE/PISA par rapport aux autres évaluations internationales 13
Aspects sur lesquels porte l'évaluation dans chaque domaine 14
Mise en œuvre de l'évaluation et présentation des résultats 17
Les questionnaires contextuels et leur utilisation 18
Le programme OCDE/PISA – Un instrument évolutif 20

Élaboration du programme OCDE/PISA et de ses plans d'évaluation – Une initiative concertée 21

Chapitre 1. **La compréhension de l'écrit** 23
Définition du domaine 23
Organisation du domaine et caractéristiques des épreuves 25
Structure des épreuves d'évaluation 41
Échelles des résultats 44
Autres questions 44

Chapitre 2. **La culture mathématique** 49
Définition du domaine 49
Organisation du domaine 50
Caractéristiques des épreuves 59
Structure des tests d'évaluation 64
Échelles des résultats 65
Autres questions 66

Chapitre 3. **La culture scientifique** 67
Définition du domaine 67
Organisation du domaine 69
Caractéristiques des épreuves 75
Structure des tests d'évaluation 77
Échelles des résultats 79
Autres questions 82

Bibliographie 83

Annexe 1. **Membres du groupe fonctionnel d'experts** 87

Annexe 2. **Questions pour les cycles à venir du programme OCDE/PISA** 89

OCDE 1999

INTRODUCTION

Dans quelle mesure les jeunes adultes sont-ils prêts à relever les défis de demain ? Sont-ils véritablement à même d'analyser des situations, de raisonner et de communiquer leurs idées ? Ont-ils les moyens de continuer à apprendre tout au long de leur vie ? Les parents, les élèves, le public et les responsables des systèmes d'enseignement ont besoin de savoir ce qu'il en est.

De nombreux systèmes d'enseignement suivent l'évolution des savoirs des élèves afin de fournir certaines réponses à ces questions. Les analyses comparatives réalisées à l'échelon international peuvent compléter et enrichir les travaux menés au plan national en déterminant les niveaux de compétence atteints par les élèves de pays différents et en offrant un contexte plus large au sein duquel interpréter les résultats nationaux. Elles permettent de définir des orientations en ce qui concerne l'action des établissements en matière d'enseignement et l'acquisition de connaissances par les élèves ; elles donnent également des indications sur les points forts et les points faibles des programmes d'enseignement. Associées à des mesures d'incitation appropriées, elles peuvent pousser les élèves à mieux apprendre, les enseignants à mieux enseigner et les établissements scolaires à se montrer plus efficaces. Elles peuvent enfin offrir au pouvoir central des instruments lui permettant de suivre l'évolution des niveaux d'instruction, même dans les cas où la gestion des systèmes éducatifs est décentralisée et où la direction des établissements est assurée en coopération avec les collectivités locales.

Les gouvernements et le grand public ont besoin d'éléments d'appréciation fiables et se prêtant à des comparaisons internationales en ce qui concerne les résultats de l'enseignement. Pour répondre à cette attente, l'Organisation de coopération et de développement économiques (OCDE) a lancé le Programme international pour le suivi des acquis des élèves (OCDE/PISA). Celui-ci doit fournir périodiquement des indicateurs des acquis des élèves orientés sur l'action des pouvoirs publics et comparables au niveau international. L'évaluation sera centrée sur les élèves âgés de 15 ans et les indicateurs devront permettre de mieux déterminer dans quelle mesure les systèmes d'enseignement des pays participants préparent leurs élèves à acquérir des savoirs tout au long de leur vie et à jouer un rôle constructif en tant que citoyens au sein de la société.

Le programme OCDE/PISA représente une nouvelle expression de la volonté des gouvernements des pays Membres de l'OCDE d'étudier de façon suivie, à l'intérieur d'un cadre conceptuel commun à l'échelon international, les résultats des systèmes d'enseignement en se fondant sur le niveau des connaissances acquises par les élèves. On peut, certes, s'attendre à ce que les résultats de l'enquête soient utilisés à des fins multiples dans les pays participants, par de nombreux acteurs, tant spécialistes que profanes ; la raison première qui a présidé à la mise au point et à la réalisation d'une évaluation internationale d'une telle ampleur est cependant la volonté de disposer d'informations reposant sur une base empirique pour étayer les décisions des pouvoirs publics.

Les résultats des évaluations réalisées par l'OCDE, qui seront publiés tous les trois ans parallèlement à d'autres indicateurs des systèmes d'enseignement, permettront aux décideurs nationaux de comparer les résultats de leurs systèmes éducatifs à ceux d'autres pays. Ils contribueront également à orienter et à promouvoir les réformes de l'enseignement et l'amélioration du fonctionnement des établissements scolaires, en particulier dans les cas où des établissements ou des systèmes d'enseignement dotés de ressources comparables obtiennent des résultats sensiblement différents. Qui plus est, ils fourniront une base à partir de laquelle il sera possible de mieux évaluer et observer de manière suivie l'efficacité des systèmes éducatifs à l'échelon national.

OCDE 1999

Qu'est-ce que le Programme international de l'OCDE pour le suivi des acquis des élèves (PISA) ? Brève présentation de ses principales caractéristiques

Données fondamentales

- Il s'agit d'une évaluation standardisée à l'échelon international, dont les instruments sont développés conjointement par les pays y prenant part et sont administrés dans les écoles à des groupes d'élèves âgés de 15 ans.
- L'évaluation est réalisée dans 32 pays, dont 28 sont Membres de l'OCDE.
- En règle générale, entre 4 500 et 10 000 élèves seront testés dans chaque pays.

Domaines d'application

- Le PISA recouvre 3 domaines : les compétences en lecture, les compétences en mathématiques et les compétences en sciences.
- Le PISA vise à définir chaque domaine non seulement du point de vue de la maîtrise des savoirs scolaires mais aussi du point de vue de connaissances et d'aptitudes qu'il est important de posséder dans la vie adulte. L'évaluation de compétences transversales fait partie intégrante du PISA.
- L'accent est placé, dans chaque domaine, sur la maîtrise des processus, la compréhension des concepts et la capacité de réagir dans diverses situations.

Méthodes utilisées

- Des tests papier-crayon sont utilisés, l'évaluation durant au total 2 heures pour chaque élève.
- Le matériel de test comprend à la fois des items à choix multiple et des questions pour lesquelles les élèves doivent construire leurs propres réponses. Les épreuves sont constituées de groupes d'items fondés sur des extraits de textes décrivant des situations réelles.
- Des items correspondant au total à 7 heures de test sont prévus, chaque élève ne passant qu'une partie des épreuves.
- Les élèves répondent à un questionnaire contextuel d'une durée de passation de 20 à 30 minutes et dans lequel ils donnent des informations sur eux-mêmes. Les directeurs d'école doivent remplir un questionnaire concernant leur établissement d'une durée de passation de 30 minutes.

Cycle d'évaluations

- La première évaluation aura lieu en 2000, et les premiers résultats seront publiés en 2001. Des évaluations ultérieures auront lieu tous les trois ans.
- Chaque cycle d'évaluations porte sur l'étude en profondeur d'une matière « dominante » à laquelle sont consacrés les deux tiers de la durée des tests ; les deux autres matières font l'objet d'une analyse plus succinte des compétences. Les matières dominantes sont les suivantes : la lecture en 2000, les mathématiques en 2003 et les sciences en 2006.

Résultats

- Un profil général des connaissances et des compétences des élèves à l'issue de la scolarité obligatoire.
- Des indicateurs contextuels établissant des liens entre les compétences des élèves et les caractéristiques de ces derniers et des établissements.
- Des séries temporelles montrant l'évolution des résultats dans le temps.

OCDE 1999

Le programme OCDE/PISA est un processus qui met en synergie les compétences scientifiques des divers pays participants et qui est dirigé conjointement par les gouvernements de ces pays en fonction de préoccupations communes touchant l'action des pouvoirs publics. On trouvera ci-dessous une présentation de la structure générale du projet et un exposé succinct des principales caractéristiques des évaluations ; on décrira ensuite les modalités données à cette collaboration, ainsi que leur mise en œuvre pour l'élaboration des plans d'évaluation utilisés dans le programme OCDE/PISA.

Les autres sections de la publication précisent le cadre conceptuel des évaluations entreprises aux fins de PISA : chacun des domaines devant être évalués est défini et des explications sont fournies quant à l'*objet* des évaluations et quant à *la manière* dont celles-ci seront effectuées. Sont également décrits le contexte dans lequel s'inscrivent les évaluations PISA ainsi que les contraintes auxquelles celles-ci sont soumises.

OCDE 1999

LA STRUCTURE DU PROGRAMME OCDE/PISA 2000

Structure générale du programme OCDE/PISA

Les évaluations que réalisera l'OCDE en l'an 2000 porteront sur les compétences en lecture, en mathématiques et en sciences. Les élèves répondront également à un questionnaire contextuel et des informations complémentaires seront recueillies auprès des autorités scolaires. Les premières évaluations de l'OCDE seront effectuées en 2000 et leurs résultats seront disponibles à partir de 2001. Trente-deux pays, dont 28 pays Membres de l'OCDE, prévoient de prendre part aux évaluations OCDE/PISA. Ensemble, ces pays représentent plus du quart de la population mondiale, soit une population supérieure à celle jamais prise en compte jusqu'ici dans une évaluation internationale touchant l'éducation.

L'objectif du programme OCDE/PISA étant d'évaluer le rendement cumulé des systèmes éducatifs à un âge auquel la scolarisation est quasiment universelle, l'évaluation sera centrée sur les élèves de 15 ans inscrits à la fois dans les filières d'enseignement scolaire et les filières d'enseignement professionnel. Entre 4 500 et 10 000 élèves seront en général testés dans chaque pays, ce qui permettra de disposer d'un échantillon suffisant à partir duquel ventiler les résultats en fonction de diverses caractéristiques des élèves.

Bien que la lecture, les mathématiques et les sciences soient des matières scolaires, les évaluations de l'OCDE n'auront pas pour objet essentiel d'examiner la mesure dans laquelle les élèves ont acquis la maîtrise de contenus spécifiques de l'enseignement. Leur but principal est plutôt d'évaluer la mesure dans laquelle les jeunes ont acquis, dans ces domaines, les connaissances et les aptitudes dont ils auront besoin dans leur vie adulte. C'est pourquoi l'évaluation de compétences transversales fait partie intégrante du programme OCDE/PISA. Les principales raisons pour lesquelles il a été décidé de situer l'évaluation dans cette perspective « large » sont les suivantes :

– Tout d'abord, bien que l'acquisition de connaissances spécifiques soit une composante essentielle de l'apprentissage scolaire, l'application de ces connaissances dans la vie adulte dépend énormément de l'acquisition par l'individu de notions et d'aptitudes plus vastes. S'agissant de la lecture, interpréter des documents écrits et réfléchir sur le contenu et la qualité de textes constituent des aptitudes fondamentales. En mathématiques, être capable de raisonner en termes quantitatifs et de concevoir des relations de dépendance ou des correspondances est plus important que de savoir répondre aux questions figurant habituellement dans les carnets de tests, lorsqu'il s'agit de faire preuve de compétences mathématiques dans la vie courante. En sciences, posséder des connaissances particulières concernant par exemple des noms de plantes et d'animaux présente moins d'intérêt que d'avoir assimilé de grandes notions telles que la consommation d'énergie, la biodiversité et la santé humaine lorsqu'on se penche sur les thèmes scientifiques qui suscitent des débats de société.

– Ensuite, privilégier le contenu des programmes d'enseignement aurait pour effet, dans un contexte international, de polariser uniquement l'attention sur les éléments de ces programmes qui sont communs à tous les pays, ou à la plupart d'entre eux. Il faudrait donc procéder à de nombreux compromis qui déboucheraient sur une évaluation trop étroite pour être utile aux gouvernements souhaitant savoir quels sont les points forts et les innovations de systèmes éducatifs d'autres pays.

OCDE 1999

– Enfin, il existe certaines aptitudes générales qu'il est indispensable que les élèves acquièrent. Il faut qu'ils sachent communiquer, s'adapter, faire preuve de souplesse, résoudre les problèmes et utiliser les technologies de l'information. Ces compétences s'acquièrent dans le cadre des diverses matières des programmes d'études et leur évaluation nécessite que l'on se situe dans une optique transversale.

Le programme OCDE/PISA s'appuie sur un modèle dynamique d'apprentissage à vie dans le cadre duquel de nouvelles connaissances et aptitudes nécessaires pour pouvoir s'adapter à l'évolution des situations sont acquises de manière continue tout au long du cycle de vie. Les élèves ne peuvent apprendre à l'école tout ce qu'ils auront besoin de savoir dans leur vie adulte. Ce qu'ils doivent acquérir sont les bases indispensables à un bon apprentissage futur. Ces bases sont à la fois de nature cognitive et affective (motivation). Les élèves doivent en effet devenir capables d'organiser et d'ordonner leur propre apprentissage, d'apprendre seuls et en groupes et de surmonter les difficultés qu'ils rencontrent au cours du processus d'apprentissage. Cela exige d'eux qu'ils aient conscience de leur mode de réflexion ainsi que de leurs stratégies et méthodes d'apprentissage. En outre, la formation continue et l'acquisition de connaissances complémentaires tendent de plus en plus souvent à prendre place dans des situations où les personnes sont appelées à travailler ensemble et dépendent les unes des autres. Pour évaluer ces aspects, l'évaluation OCDE/PISA 2000 envisage la mise au point d'un instrument permettant de recueillir des informations sur l'organisation personnelle de l'apprentissage.

Le programme OCDE/PISA n'est pas une évaluation internationale isolée des compétences en lecture, en mathématiques et en sciences des élèves de 15 ans. Il s'agit d'un programme d'évaluation durable, dans le cadre duquel seront rassemblées des données concernant chacune de ces matières tous les trois ans. A long terme, il débouchera sur la constitution d'une base de données permettant d'enregistrer les tendances en matière d'acquisition de connaissances et de compétences des élèves dans les divers pays ainsi que dans divers sous-groupes démographiques au sein de chacun d'entre eux. Lors de chaque enquête, l'une des matières sera évaluée en détail, son étude représentant près de deux tiers de la durée totale des tests. La matière « dominante » sera la lecture en 2000, les mathématiques en 2003 et les sciences en 2006. Ce cycle d'enquêtes permettra d'entreprendre, tous les neuf ans, une analyse approfondie sur le niveau des acquis dans chacune des matières, et d'établir un bilan plus restreint tous les trois ans.

Le temps total qui sera consacré aux tests par chaque élève sera de deux heures, mais la collecte d'informations se fondera sur une batterie d'items correspondant à près de sept heures de tests. L'ensemble des questions sera divisé en blocs. Chaque bloc sera administré à un nombre d'élèves suffisant pour que des estimations correctes puissent être réalisées en ce qui concerne les niveaux de compétences pour l'ensemble des items dans chaque pays et à l'intérieur de sous-groupes spécifiques au sein d'un même pays (les garçons et les filles par exemple, ou les élèves se situant dans des contextes sociaux et économiques différents). Les élèves passeront également 20 minutes à répondre au questionnaire contextuel.

Les évaluations fourniront divers types d'indicateurs :

– des indicateurs de base donnant un profil général des connaissances et des compétences des élèves ;

– des indicateurs contextuels montrant comment ces compétences sont liées à d'importantes variables démographiques, sociales, économiques et éducatives ;

– des indicateurs de tendances qui deviendront disponibles en raison du caractère cyclique des collectes de données et qui montreront l'évolution des niveaux de compétence, l'évolution de la ventilation des compétences et l'évolution des rapports entre les caractéristiques des élèves et des écoles, et les compétences au fil du temps.

Bien que les indicateurs constituent un bon moyen d'appeler l'attention sur des questions importantes, ils ne sont en général pas à même d'apporter de réponses aux questions qui se posent aux pouvoirs publics. C'est pourquoi le projet OCDE/PISA a également mis au point un schéma d'analyse des problématiques relatives aux politiques éducatives qui ira au-delà de la simple présentation des indicateurs.

Les pays participant au premier cycle d'enquêtes OCDE/PISA sont les suivants : l'Allemagne, l'Australie, l'Autriche, la Belgique, le Brésil, le Canada, la Chine, la Corée, le Danemark, l'Espagne, les États-Unis, la Fédération de Russie, la Finlande, la France, la Grèce, la Hongrie, l'Irlande, l'Islande, l'Italie, le Japon, la Lettonie, le Luxembourg, le Mexique, la Norvège, la Nouvelle-Zélande, les Pays-Bas, la Pologne, le Portugal, la République tchèque, le Royaume-Uni, la Suède et la Suisse.

Originalité du programme OCDE/PISA par rapport aux autres évaluations internationales

Le PISA ne constitue pas la première enquête internationale comparative sur les acquis des élèves. D'autres études ont été effectuées au cours des 40 dernières années, essentiellement par l'Association internationale pour l'évaluation du rendement scolaire (IEA) et par l'Education Testing Service dans le cadre de l'International Assessment of Educational Progress (IAEP). La qualité et la portée de ces enquêtes ont beaucoup progressé au fil des ans mais elles ne donnent cependant que des images partielles et isolées des connaissances acquises par les élèves, à un moment donné et dans un nombre limité de matières. Les trois enquêtes réalisées en sciences et en mathématiques par l'IEA fournissent certaines indications sur la façon dont la situation a évolué en 30 ans mais le tableau qu'elles dressent n'est que partiel compte tenu du nombre restreint de pays qui ont participé aux premières enquêtes et du fait que les tests ne sont pas entièrement comparables.

Surtout, ces enquêtes se sont concentrées sur les compétences directement liées aux programmes d'enseignement et n'ont en outre porté que sur les parties des programmes qui sont en principe communes à tous les pays participants. Les aspects des programmes spécifiques à un pays ou à un petit nombre d'entre eux n'ont en général pas été pris en compte dans les évaluations, quelle que soit leur importance pour les pays concernés.

L'originalité du programme OCDE/PISA tient notamment à un certain nombre de points importants, à savoir :

– son *origine* : ce sont les gouvernements qui ont pris l'initiative de mettre en œuvre une enquête conçue de manière à répondre à leurs besoins ;

– sa *périodicité* : la décision de traiter plusieurs matières, en procédant à des mises à jour tous les trois ans, permettra aux pays de suivre régulièrement et à des dates fixées d'avance les progrès sur la voie de la réalisation d'objectifs clés en matière d'apprentissage ;

– le *groupe d'âge considéré* : faire porter l'évaluation sur des jeunes gens en fin de scolarité obligatoire fournit une indication précieuse sur la performance des systèmes éducatifs. Alors que la plupart des jeunes gens des pays de l'OCDE poursuivent leur formation au-delà de l'âge de 15 ans, cet âge est normalement proche de la fin de la scolarité de base dans le cadre de laquelle tous les élèves suivent un programme d'enseignement pratiquement commun. Il est utile de déterminer, à ce stade, la mesure dans laquelle ils ont acquis des connaissances et des compétences qui leur serviront dans leur vie ultérieure, y compris dans le cadre des formations complémentaires individualisées qu'ils pourront suivre ;

– les *connaissances et les compétences évaluées* : celles-ci ne sont pas avant tout définies en tant que dénominateur commun des programmes scolaires nationaux, mais en tant qu'aptitudes jugées essentielles dans la vie future. Il s'agit là de l'apport le plus important et le plus ambitieux du programme OCDE/PISA.

  Il serait arbitraire d'établir une distinction trop stricte entre les aptitudes « scolaires » et les aptitudes « à vivre en société », l'école ayant toujours eu pour objectif d'équiper les jeunes pour la vie, mais cette distinction a néanmoins son importance. Traditionnellement, les programmes scolaires sont surtout conçus en tant qu'ensembles d'informations et de techniques à maîtriser, et accordent moins d'importance, dans le cadre de chaque matière, à l'acquisition d'aptitudes susceptibles d'être utilisées en général dans la vie adulte. Ils accordent une importance encore moindre à des compétences d'ordre encore plus général, acquises dans le cadre des diverses matières des programmes d'études, qui permettent à l'individu de résoudre des problèmes et d'appliquer ses idées et ses savoirs aux situations qu'il rencontre dans la vie. Le programme

OCDE 1999

OCDE/PISA ne met pas à l'écart les connaissances et les compétences liées aux programmes d'enseignement, mais il évalue essentiellement celles-ci dans la perspective de l'acquisition de notions et aptitudes à caractère général permettant d'exploiter les connaissances. De plus, l'évaluation OCDE/PISA ne se heurte pas aux limites imposées par le dénominateur commun constitué par le contenu spécifique de l'enseignement dispensé dans les écoles des pays participants.

Cette approche de l'évaluation en terme de maîtrise de grands concepts est particulièrement importante, compte tenu de la volonté exprimée par les nations de mettre en valeur le capital humain, que l'OCDE définit comme suit :

« les connaissances, qualifications, compétences et autres qualités possédées par un individu et intéressant le bien-être personnel, social et économique ».

Jusqu'ici, la tendance a été d'estimer le stock de capital humain ou la base des compétences humaines en les déterminant, dans le meilleur des cas, à partir de variables telles que le niveau d'études atteint. Lorsque l'intérêt porté au capital humain s'étend, en outre, aux qualités qui permettent une participation démocratique pleine et entière à la vie de la collectivité à l'âge adulte et qui aident les individus à devenir des « apprenants permanents », l'inadéquation de ces variables apparaît encore plus clairement.

En évaluant directement les connaissances et les compétences qui ont été acquises à la fin de la scolarité obligatoire, le programme OCDE/PISA examine le degré de préparation des jeunes gens à la vie adulte et, dans une certaine mesure, l'efficacité des systèmes éducatifs. Son ambition est d'évaluer le niveau de formation des élèves par rapport aux objectifs sous-jacents des systèmes éducatifs (tels qu'ils sont définis par la société), et non par rapport à l'enseignement et à l'acquisition d'un ensemble de connaissances. Il est nécessaire de procéder à des évaluations de ce type, qui tiennent véritablement compte de la réalité, si l'on veut inciter les établissements scolaires et les systèmes éducatifs à se préoccuper des défis de l'époque moderne.

Aspects sur lesquels porte l'évaluation dans chaque domaine

Le tableau 1 présente de manière succincte la structure de chacun des trois domaines sur lesquels porte le programme OCDE/PISA, en donnant une définition de chacun d'eux et en décrivant les dimensions qui structurent les items prévus dans les tests.

Les définitions données pour les trois domaines mettent l'accent sur les connaissances et les compétences fonctionnelles qui permettent de participer activement à la vie de la société. Une telle participation exige davantage qu'être simplement à même d'exécuter des tâches imposées de l'extérieur, par un employeur par exemple. Elle exige également que l'on possède les aptitudes nécessaires pour prendre part aux processus de prise de décisions. Les épreuves les plus complexes prévues par le programme OCDE/PISA demanderont aux élèves de réfléchir sur le matériel qui leur est présenté, et non de répondre simplement à des questions pour lesquelles il n'existe qu'une seule réponse « correcte ».

Afin de rendre opérationnelles ces définitions, chaque domaine est décrit en fonction de trois dimensions, à savoir :

– le *contenu* ou la *structure* des connaissances que les élèves doivent acquérir dans chaque domaine ;

– un ensemble de *processus* qu'il faut savoir mettre en oeuvre et qui exigent diverses aptitudes cognitives ;

– la *situation* ou le *contexte* dans lesquels les connaissances et les compétences sont appliquées ou exploitées.

L'objectif poursuivi est d'évaluer un ensemble d'aptitudes nécessaire à l'exécution de diverses tâches que les citoyens sont susceptibles d'avoir à accomplir. Il convient toutefois de souligner que l'importance relative à attribuer à ces trois dimensions doit encore faire l'objet d'études. Les essais de terrain qui auront lieu en 1999 permettront de mettre à l'épreuve un grand nombre de questions portant sur

Tableau 1. **Synthèse des composantes du PISA**

Domaine	Compréhension de l'écrit	Culture mathématique	Culture scientifique
Définition	***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.***	***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.***	***Capacité d'associer des connaissances scientifiques à la formulation de conclusions fondées sur l'observation des faits 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.***
Composantes/ aspects du domaine	Lecture de différents types de *textes :* les textes continus, divisés en sous-catégories (par exemple les textes descriptifs, les textes narratifs), et les documents, classés selon leur structure.	*Contenu* des mathématiques – pour l'essentiel les «idées mathématiques majeures», qui dans le premier cycle sont : les variations et la croissance, et l'espace et les formes. Les cycles futurs comprendront : le hasard, le raisonnement quantitatif, l'incertitude et les relations de dépendance.	*Concepts scientifiques* – par exemple conservation, transformation et décomposition de l'énergie – se rapportant aux domaines majeurs de la science (physique, biologie, chimie, etc.) où ils sont utilisés dans le but de résoudre des problèmes touchant l'utilisation de l'énergie, la conservation des espèces ou l'utilisation des matières premières.
	Exécution de divers types de *tâches* de lecture, tels que la restitution d'informations, le développement d'une interprétation ou une réflexion sur le contenu et la forme du texte.	Application de *compétences* mathématiques, par exemple capacités de modélisation, résolution de problèmes, classées en trois catégories *i)* exécution de procédures, *ii)* mises en relation et *iii)* réflexion dans une optique mathématique (mathématisation) et généralisation.	*Application de savoir-faire* concernant les démarches scientifiques – par exemple identifier des faits et communiquer, évaluer et en tirer des conclusions valides. Ils ne nécessitent pas de connaissances scientifiques préalables mais doivent nécessairement être mis en œuvre en relation avec des contenus scientifiques.
	Lecture de textes écrits, dans des *situations* différentes, par exemple pour satisfaire des intérêts personnels ou à des fins professionnelles.	Utilisation des mathématiques dans diverses situations, par exemple les problèmes qui concernent un individu, une communauté ou l'ensemble du monde.	Utilisation des sciences dans des *situations* ou contextes différents dans le cas de problèmes, par exemple concernant les individus, les collectivités ou le monde entier.

les trois dimensions retenues et leurs diverses sous-catégories, avant qu'une décision soit prise sur les catégories les plus utiles à retenir et sur la façon de traduire en notes globales les résultats obtenus en ce qui concerne les divers items.

Au sein de ce cadre commun en trois dimensions, chaque domaine est structuré de manière spécifique. Il existe en effet une différence importante entre la maîtrise de la lecture d'une part, et celles des sciences et des mathématiques d'autre part. La première est en soi une compétence qui transcende les diverses matières des programmes d'enseignement, particulièrement au niveau secondaire, et qui ne peut être associée à aucun « contenu » particulier. Bien que certaines connaissances scolaires, concernant par exemple la structure de la phrase, ne soient pas dénuées d'importance, ce type de connaissances ne peut se comparer, par exemple, à la maîtrise d'un ensemble de principes ou de concepts scientifiques.

Les principaux aspects des trois domaines, résumés dans le tableau 1, sont décrits plus en détail ci-après.

OCDE 1999

On entend par **compréhension de l'écrit** l'aptitude des individus à utiliser un texte écrit pour atteindre des buts qu'ils se sont fixés. Cet aspect de la compréhension de l'écrit a été bien défini à l'occasion d'enquêtes précédentes telles que l'*Enquête internationale sur la littératie des adultes* (EILA), mais le programme OCDE/PISA le développe en introduisant un élément « actif » – la capacité non seulement de comprendre un texte mais de l'analyser, en se fondant sur ses propres réflexions et expériences. La compréhension de l'écrit est évaluée en fonction des aspects suivants :

– Premièrement, la *forme du matériel de lecture*, à savoir le type de texte à lire. De nombreuses évaluations de la lecture ont été centrées sur des « textes continus » organisés en phrases et paragraphes. Le programme OCDE/PISA fera en outre intervenir des « textes non continus » présentant l'information sous d'autres formes (listes, formulaires, graphiques ou diagrammes). Il distinguera également entre diverses formes de textes continus (textes narratifs, textes informatifs, textes argumentatifs, etc.). Ces distinctions se fondent sur le principe qu'une personne rencontrera une large variété de textes écrits dans sa vie adulte, et qu'il ne suffit pas de savoir lire un nombre limité de types de textes habituellement utilisés dans le cadre scolaire.

– Deuxièmement, le *type de tâche de lecture*. Cet aspect concerne d'une part, les diverses démarches cognitives qu'il faut pouvoir mettre en oeuvre pour être un lecteur efficace et d'autre part, les caractéristiques des questions posées dans l'évaluation. L'évaluation des élèves ne portera pas sur les aptitudes à la lecture les plus fondamentales, car on suppose que la plupart des jeunes de 15 ans ont acquis celles-ci. Il leur sera plutôt demandé de montrer leur capacité de restituer l'information, de comprendre globalement le texte présenté, de l'interpréter et de réfléchir sur son contenu ainsi que sur sa forme.

– Troisièmement, l'utilisation pour laquelle le texte a été conçu – à savoir le contexte ou la situation dans lesquels il doit être utilisé. Ainsi, un roman, une lettre personnelle ou une biographie sont rédigés en vue d'une lecture à usage « privé » ; les documents ou les avis officiels sont destinés à un usage « public » ; un manuel d'entretien ou un rapport sont destinés à une lecture à des fins professionnelles et un manuel scolaire ou une feuille de travail sont destinés à une utilisation « scolaire ». L'une des raisons importantes qui amène à établir ces distinctions est que certains groupes d'élèves peuvent obtenir de meilleurs résultats dans une situation de lecture que dans une autre, auquel cas il est souhaitable de prévoir des items couvrant ces divers types de situations.

La **culture mathématique** est définie en tant que capacité d'un individu de comprendre le rôle des mathématiques et d'utiliser cette discipline suivant des voies adaptées à ses besoins. Cette approche met l'accent sur l'aptitude à poser et à résoudre des problèmes mathématiques plutôt que l'aptitude à faire des opérations mathématiques spécifiées. La culture mathématique est évaluée en fonction des aspects suivants :

– Premièrement, le *contenu* des mathématiques, essentiellement défini en termes « d'idées mathématiques majeures » (le hasard, les variations et la croissance, l'espace et les formes, le raisonnement quantitatif, l'incertitude et les relations de dépendance) et ensuite seulement, par référence aux « domaines enseignés » (par exemple, les nombres, l'algèbre et la géométrie). Un ensemble représentatif (et non la totalité) des principaux concepts qui sous-tendent le raisonnement mathématique a été retenu aux fins du programme OCDE/PISA. Dans le cadre du premier cycle d'évaluation – dans lequel les mathématiques constituent un domaine « mineur » –, le nombre de ces concepts a été encore réduit et seules deux idées majeures ont été retenues : les variations et la croissance, et l'espace et les formes. Ce choix permet de prendre en considération de nombreux aspects des programmes d'enseignement, sans que l'on ait à accorder d'importance excessive aux compétences relatives aux nombres.

– Deuxièmement, la *démarche* mathématique, définie en termes de compétences générales en mathématiques. Parmi celles-ci figurent l'utilisation du langage mathématique, la capacité de modélisation et la capacité de résoudre des problèmes.

Il n'est cependant pas question d'évaluer ces compétences à l'aide d'items différents, dans la mesure où l'on suppose qu'un ensemble de compétences est nécessaire pour accomplir toute tâche mathématique donnée. Les questions sont donc organisées en fonction de trois « catégories de compétences » définissant le type d'opération mentale nécessaire. La première catégorie concerne des calculs simples ou des définitions que l'on rencontre habituellement dans les évaluations mathématiques traditionnelles. La deuxième nécessite des mises en relation visant à résoudre des problèmes simples. La troisième catégorie de compétences porte sur le raisonnement mathématique ou « mathématisation », la généralisation et la compréhension en profondeur, et elle exige des élèves qu'ils soient à même de procéder à une analyse leur permettant d'identifier les éléments mathématiques d'une situation et de poser les problèmes.

– Troisièmement, les *situations* dans lesquelles sont utilisées les mathématiques. Le plan d'évaluation identifie cinq situations : privée, éducative, professionnelle, publique et scientifique. Dans le cas des mathématiques, cependant, ce troisième aspect est considéré comme moins important que ceux se rapportant à la démarche ou au contenu.

La **culture scientifique** est définie en tant que capacité d'utiliser des connaissances et des processus scientifiques, non pas simplement en vue d'appréhender le monde naturel mais pour participer à la prise de décisions concernant celui-ci. La culture scientifique est évaluée en fonction des aspects suivants :

– Premièrement, les *concepts scientifiques*, qui constituent les liens permettant de comprendre des phénomènes connexes. Dans le cadre du programme OCDE/PISA, les concepts pris en compte sont des concepts courants se rapportant aux sciences physiques, chimiques et biologiques ainsi qu'aux sciences de la terre et de l'espace, mais il sera demandé aux élèves de les appliquer au contenu des items plutôt que de se borner à un simple rappel des concepts. Le contenu des items sera principalement lié à trois grands champs d'application : les sciences de la vie et la santé, les sciences de la terre et l'environnement et la science dans la technologie.

– Deuxièmement, les *savoir-faire concernant des démarches scientifiques*, centrés sur la capacité de constater, d'interpréter et d'utiliser des faits. Le programme OCDE/PISA étudie notamment à cet égard : *i)* la reconnaissance de questions scientifiques, *ii)* l'identification de faits probants, *iii)* l'établissement de conclusions, *iv)* la communication de ces conclusions et *v)* la manifestation de la compréhension des concepts scientifiques. A l'exception de la dernière, ces démarches n'impliquent pas que l'on ait acquis un ensemble préalable de connaissances scientifiques, mais aucune démarche scientifique ne peut être « sans contenu », de sorte que les questions relatives aux sciences abordées dans le cadre de l'évaluation OCDE/PISA nécessiteront toujours une connaissance des concepts dans divers champs d'application décrits ci-dessus.

– Troisièmement, *des situations scientifiques*, qui se rapportent davantage à la vie quotidienne plutôt qu'au travail scientifique réalisé dans les salles de classe ou les laboratoires, ou au travail de chercheurs. Dans l'évaluation OCDE/PISA, les situations scientifiques sont conçues comme des problèmes en rapport avec la personne et la famille, avec la collectivité, avec le monde dans son ensemble, ainsi qu'avec l'évolution historique des connaissances scientifiques.

Mise en œuvre de l'évaluation et présentation des résultats

Pour des raisons de faisabilité, l'évaluation OCDE/PISA 2000 sera réalisée à l'aide de tests papier-crayon. La possibilité de recourir à d'autres supports sera activement étudiée lors des cycles ultérieurs.

L'évaluation utilisera divers types d'items. Certains d'entre eux seront de type « fermé » – c'est-à-dire qu'ils réclameront des élèves qu'ils choisissent ou produisent des réponses simples, pouvant être directement comparées à une réponse correcte unique. D'autres auront un caractère plus ouvert, imposant aux élèves de fournir des réponses plus élaborées destinées à évaluer des dimensions plus larges que celles mesurées par d'autres enquêtes à caractère plus traditionnel. L'évaluation de compétences d'ordre supérieur, souvent par le biais de questions à réponse ouverte, constituera une innovation importante

OCDE 1999

du programme PISA. La place exacte à donner à ce type d'items dépendra de l'intérêt qu'ils présenteront lors des essais de terrain ainsi que de la possibilité de mettre au point des grilles de correction fiables. Après avoir tenu une place relativement modeste dans le premier cycle d'enquête, les items à réponse ouverte seront sans doute plus largement utilisés au cours des cycles ultérieurs du PISA.

Dans la plupart des cas, les épreuves auront la forme de groupes d'items se rapportant à un texte, à un matériel d'incitation ou à un thème communs. Il s'agit là d'un aspect important qui permet aux questions d'avoir un caractère plus approfondi que ce ne serait le cas si chacune d'entre elles portait sur un contexte entièrement nouveau. Cette formule donne davantage de temps à l'élève pour assimiler le support qui peut être ensuite utilisé pour évaluer des aspects multiples de ses performances.

Dans l'ensemble, les items utilisés dans le programme OCDE/PISA seront très différents de ceux utilisés, par exemple, dans les études internationales telles que la Troisième étude internationale sur les mathématiques et les sciences (IEA/TIMSS), qui se concentrait sur de brèves questions à choix multiple fondées sur des connaissances acquises en classe. Par exemple, certaines questions touchant les sciences nécessitent des connaissances scientifiques de base (les élèves devaient indiquer, par exemple, combien de pattes et de parties du corps possèdent les insectes) ou une exploitation simple des connaissances (les élèves devaient, par exemple, déterminer laquelle parmi des cuillères en métal, en bois ou en plastique serait la plus chaude après avoir été plongée pendant 15 secondes dans de l'eau chaude). Les items prévus dans l'évaluation OCDE/PISA nécessitent en revanche le plus souvent l'association d'un ensemble de connaissances et de compétences et parfois (comme dans l'exemple n° 4 du plan d'évaluation des compétences scientifiques – voir figure 18), une évaluation active de décisions pour lesquelles il n'existe pas de réponse correcte unique.

Les résultats de l'évaluation OCDE/PISA seront présentés sous forme de niveaux de performances atteints par les élèves sur les diverses échelles de compétence dans chaque domaine. Les échelles obtenues par la calibration des items du test permettront de décrire les résultats en termes des compétences que maîtrisent les élèves se situant aux divers niveaux de l'échelle. En d'autres termes, il sera possible de dire ce que les élèves, à un niveau donné de l'échelle, savent et sont capables de faire par rapport à ceux qui se trouvent à des niveaux inférieurs. L'utilisation d'items nécessitant un niveau de raisonnement supérieur ainsi que d'autres demandant un niveau de connaissances relativement faible, permettra de prendre en compte une large gamme de compétences.

Savoir s'il faut présenter les niveaux de compétence dans chaque domaine sur une échelle ou plusieurs est un point important. Les compétences d'une personne peuvent-elles être facilement regroupées et situées à un niveau spécifique, ou est-ce justifié d'indiquer qu'elles se situent à des niveaux différents dans divers sous-domaines ? La réponse à cette question dépendra de deux facteurs, que les essais de terrain feront apparaître plus clairement. Premièrement, il faut se demander dans quelle mesure les performances d'un individu pour certains types de questions sont en corrélation avec ses performances pour d'autres types de questions, en d'autres termes s'il existe des profils de performance différents liés à des dimensions particulières du domaine évalué. En second lieu, il faudra vérifier s'il est possible de développer plus d'une échelle de résultats pour chaque domaine, compte tenu du nombre d'items pouvant figurer dans l'évaluation. Le scénario le plus probable est que dans le domaine majeur, qui comptera le plus grand nombre de questions (la lecture, dans l'évaluation OCDE/PISA 2000), il sera possible de prévoir plusieurs échelles, mais que dans les domaines mineurs, une seule échelle sera utilisée.

Les questionnaires contextuels et leur utilisation

Afin de rassembler des informations contextuelles, il sera par ailleurs demandé aux élèves et aux directeurs d'établissement de remplir des questionnaires sur leur environnement éducatif pour lesquels un temps de réponse de 20 à 30 minutes est prévu. Ces questionnaires ne constituent pas un élément annexe de l'évaluation OCDE/PISA. Il s'agit d'instruments centraux, devant permettre l'analyse des résultats en fonction des caractéristiques des élèves et des écoles.

Les questionnaires rassembleront des informations sur les aspects suivants :

– les élèves et leur milieu familial, y compris les ressources économiques, sociales et culturelles des élèves et de leurs familles ;

– certains aspects de la vie des élèves tels que leur attitude par rapport à l'apprentissage, leurs habitudes et leur mode de vie à l'école et dans leur milieu familial ;

– certaines caractéristiques des écoles, telles que la qualité des ressources humaines et matérielles dont elles disposent, le caractère public ou privé de leur gestion et de leur financement, les processus de prise de décision et les méthodes de gestion du personnel ;

– le contexte scolaire, à savoir les structures et les types d'établissement, la taille des classes et le niveau de participation des parents.

Le premier cycle de l'évaluation OCDE/PISA comprendra également un instrument demandant aux élèves de décrire la façon dont ils organisent leur apprentissage. Cet instrument se fonde sur les *constructs*[1] suivants :

– les stratégies d'organisation personnelle de l'apprentissage, qui déterminent dans quelle mesure les informations reçues seront traitées de manière approfondie et systématique ;

– les profils de motivation et d'objectifs poursuivis, lesquels influent sur l'investissement en temps et en énergie mentale consenti pour apprendre, ainsi que sur le choix des stratégies d'apprentissage ;

– les mécanismes personnels d'acquisition de connaissances, qui régulent les normes que l'élève se donne, les objectifs poursuivis et les méthodes d'action ;

– les stratégies de contrôle de l'action, en particulier l'effort et la persévérance, qui évitent que des objectifs concurrents viennent entraver l'action et qui contribuent à surmonter les difficultés d'apprentissage ;

– les préférences concernant les situations d'apprentissage, les styles d'apprentissage et les aptitudes sociales nécessaires à un apprentissage en collaboration.

Les questionnaires contextuels prévus dans l'évaluation OCDE/PISA fourniront une base de données détaillée pour analyser les résultats de l'évaluation dans l'optique de l'action des pouvoirs publics. Ils permettront notamment :

– de mettre en relation les différences de résultats scolaires avec les différences entre les systèmes éducatifs et les contextes scolaires ;

– de mettre en relation les différences de résultats scolaires avec les différences entre les contenus des divers programmes d'enseignement et les diverses méthodes pédagogiques ;

– d'examiner le rapport entre les résultats scolaires et certaines caractéristiques des écoles, telles que leur taille et les ressources dont elles disposent, ainsi que les différences entre les pays à cet égard ;

– d'examiner les différences entre les pays en ce qui concerne le rôle joué par l'école pour atténuer ou accentuer l'impact des caractéristiques des élèves sur leur rendement scolaire ;

– d'examiner les différences entre les systèmes éducatifs et les contextes nationaux qui sont liées à des différences en matière de résultats scolaires entre les pays.

Les informations contextuelles rassemblées par le biais des questionnaires destinés aux élèves et aux établissements ne fourniront qu'une partie de l'ensemble total des informations auquel aura accès le programme PISA. Des indicateurs décrivant la structure générale des systèmes éducatifs (leurs contextes démographique et économique définis notamment en termes de coûts, de nombre d'élèves inscrits, de rendement, de caractéristiques des écoles et des enseignants et de méthodes pédagogiques, et concernant les retombées de l'enseignement sur le marché du travail) sont déjà régulièrement élaborés par l'OCDE.

1. NDT : Le terme « construct » a volontairement été laissé en anglais, car il est d'un emploi courant en tant que terme technique relatif aux tests ; il renvoie à la dimension latente que cherche à mesurer une épreuve grâce aux données observables que constituent les réponses des élèves.

OCDE 1999

Le programme OCDE/PISA – Un instrument évolutif

Étant donné l'horizon à long terme du projet et les accents différents qui seront placés sur les divers domaines dans le cadre de chaque cycle d'étude, les plans d'évaluation OCDE/PISA représentent à l'évidence un instrument qui ne manquera pas d'évoluer. Ils sont conçus pour être flexibles afin qu'ils puissent :

– être modifiés en fonction de l'évolution des intérêts des pays participants ;

– servir en même temps de base aux composantes permanentes, qui sont susceptibles de susciter un intérêt constant au cours du temps et qui devraient par conséquent figurer dans tous les cycles d'évaluation.

Les plans d'évaluation seront modifiés à la lumière des essais de terrain de 1999 avant qu'un instrument définitif soit mis au point pour le programme OCDE/PISA 2000. Par ailleurs, l'enquête conservera un caractère évolutif, lui permettant de prendre en compte tant les modifications apportées aux objectifs des systèmes éducatifs que les améliorations touchant les techniques d'évaluation. Il faudra certes veiller à concilier les avantages offerts par ce caractère évolutif et ces améliorations, et la nécessité de procéder à des comparaisons fiables dans le temps, et c'est pourquoi un grand nombre d'éléments fondamentaux de l'évaluation OCDE/PISA seront conservés au fil des ans.

Les objectifs de l'OCDE sont ambitieux. Pour la première fois, une évaluation internationale des résultats scolaires vise à déterminer non seulement si les élèves ont acquis les connaissances prévues dans les programmes, mais encore si les savoirs et les aptitudes qu'ils ont acquis dans leur enfance les ont bien préparés à la vie adulte. Les pays qui souhaitent vérifier de façon continue l'adéquation de leur système éducatif dans un contexte mondial ont besoin de ce type de mesures. On ne parviendra certes pas immédiatement à réaliser des évaluations parfaites ; certains des objectifs du programme PISA se heurteront dans un premier temps à des obstacles d'ordre pratique dus au fait que l'instrument d'évaluation devra à la fois être fiable et permettre des comparaisons entre un grand nombre de cultures différentes. Mais les objectifs poursuivis ont été clairement définis, et les évaluations, tout en évoluant au fil des ans, s'attacheront à les réaliser progressivement.

Dans le plan OCDE/PISA 2000, les capacités des élèves seront évaluées grâce à une série d'épreuves administrées sur papier qui devront être faites dans un temps limité. Le terme de « test » a plusieurs significations dans les pays participants. Il implique parfois que les résultats ont une répercussion particulière sur l'élève. Le but de PISA est d'effectuer une enquête sur les caractéristiques globales des élèves dans chaque pays, plutôt que d'examiner les performances d'individus particuliers. Le terme d'« évaluation » est donc utilisé dans PISA, même si les conditions auxquelles sont soumis les élèves seront les mêmes que celles d'un contrôle scolaire.

OCDE 1999

ÉLABORATION DU PROGRAMME OCDE/PISA ET DE SES PLANS D'ÉVALUATION UNE INITIATIVE CONCERTÉE

Le programme OCDE/PISA représente une initiative concertée des gouvernements des pays Membres de l'OCDE visant à déboucher sur un nouveau type d'évaluation périodique des acquis des élèves. Les évaluations entreprises dans le cadre du programme ont été définies d'un commun accord par les pays participants et sont réalisées par des organismes nationaux.

Un Conseil des pays participants au sein duquel chaque pays est représenté définit, dans le contexte des objectifs de l'OCDE, les priorités d'action concernant le programme OCDE/PISA et veille au respect de ces priorités au cours de la mise en œuvre du programme. Il est chargé de déterminer les priorités en ce qui concerne l'élaboration d'indicateurs, la mise au point des instruments d'évaluation et la présentation des résultats. Des experts des pays participants sont par ailleurs membres de groupes de travail chargés de mettre au service des objectifs d'action du programme OCDE/PISA les meilleures compétences techniques disponibles à l'échelon international dans les différents domaines évalués. En contribuant aux travaux de ces groupes d'experts, les pays veillent à ce que :

– les instruments utilisés soient valides au plan international et prennent en compte les contextes culturels et éducatifs des pays Membres de l'OCDE ;

– le matériel d'évaluation se fonde sur des méthodes de mesure rigoureuses ;

– les instruments d'évaluation mettent l'accent sur l'authenticité et la valeur éducative des tâches proposées aux élèves.

Par l'intermédiaire des chefs de projet nationaux, les pays participants mettent en œuvre le projet OCDE/PISA à l'échelon national dans le cadre des procédures d'exécution convenues. Les directeurs de projet nationaux ont un rôle de premier plan à jouer pour garantir la bonne qualité de l'exécution du projet et pour contrôler et évaluer les résultats de l'enquête, ainsi que les analyses, les rapports et les publications.

La conception et l'exécution des enquêtes, à l'intérieur du cadre défini par le Conseil des pays participants, relève de la responsabilité d'un consortium international dirigé par l'Australian Council for Educational Research (ACER). Parmi les autres membres du Consortium figure l'Institut national néerlandais pour l'évaluation de l'enseignement (CITO), le Service de pédagogie expérimentale de l'Université de Liège (Belgique) et WESTAT, aux États-Unis.

Le Secrétariat de l'OCDE est responsable de la gestion globale du programme, suit la mise en oeuvre de ce dernier au jour le jour, assure le secrétariat du Conseil des pays participants, facilite la recherche de consensus entre les pays participants et sert d'interlocuteur entre le Conseil des pays participants et le consortium international chargé de la mise en œuvre des activités. Le Secrétariat de l'OCDE produira également les indicateurs et les analyses, et préparera les publications et les rapports internationaux conjointement avec les membres du projet PISA et en consultation étroite avec les pays de l'OCDE, tant au niveau des orientations (par l'entremise du Conseil des pays participants) qu'au niveau de la réalisation (par l'intermédiaire des chefs de projet nationaux).

Les plans d'évaluation utilisés dans le programme OCDE/PISA, qui sont décrits ci-après, ont été élaborés par des groupes d'experts sous la direction de l'ACER. Le groupe d'experts sur la lecture a été dirigé par le Professeur Irwin Kirsch, de l'International Testing Service, le groupe chargé des mathématiques

OCDE 1999

a été dirigé par le Professeur Jan de Lange, de l'Université d'Utrecht, et le groupe d'experts chargé des sciences a été dirigé par le Professeur Wynne Harlen, du Scottish Council for Research in Education. La mise au point des plans d'évaluation a bénéficié en outre des consultations effectuées par les groupes d'experts auprès de leurs collègues des pays participants. Les plans ont été adoptés par les gouvernements des pays de l'OCDE par l'intermédiaire du Conseil des pays participants.

L'élaboration des plans d'évaluation de l'OCDE/PISA s'est déroulée selon un processus en six étapes :

– la mise au point d'une définition opérationnelle du domaine visé et la formulation des hypothèses qui sous-tendent cette définition ;

– la détermination de la structure à donner à l'ensemble des tâches proposées aux élèves pour permettre de rendre compte le mieux possible aux décideurs et aux chercheurs des niveaux de compétences acquis dans chaque matière par les élèves de 15 ans des pays participants ;

– l'identification d'un ensemble de caractéristiques clés qu'il convient de prendre en compte lors de la définition de tâches devant être utilisées à l'échelon international ;

– la mise au point opérationnelle de l'ensemble des caractéristiques clés qui seront prises en compte dans l'élaboration des tests, par référence aux définitions fournies par la littérature disponible et à l'expérience acquise au cours d'autres évaluations à large échelle ;

– la validation de ces variables et l'étude de la contribution que chacune d'entre elles apporte à la description de la difficulté des tâches dans les divers pays participants ;

– la mise au point d'un schéma d'interprétation des résultats.

Les quatre premières étapes sont décrites dans cette publication. Les deux dernières étapes seront achevées quand les résultats des plans d'évaluation seront disponibles. Les principes suivants ont guidé le processus d'élaboration de ces plans :

– le point de départ d'un plan d'évaluation doit être une définition générale ou l'énoncé de l'objectif poursuivi – indiquant la raison d'être de l'enquête et l'objet des mesures ;

– la définition d'un plan d'évaluation doit fournir une description des diverses caractéristiques des tâches et indiquer comment ces caractéristiques seront utilisées lors de la mise au point des tâches ;

– les variables associées à chacune des caractéristiques des tâches doivent être spécifiées, et celles qui semblent avoir la plus forte incidence sur la variance (en termes de difficulté des tâches) doivent être utilisées pour mettre sur pied un dispositif d'interprétation.

Élaborer et valider un plan d'évaluation pour chacun des domaines visés a principalement pour objectif d'améliorer la mesure ; la démarche comporte cependant plusieurs autres avantages potentiels :

– le plan d'évaluation fournit un langage et un cadre communs aux discussions sur la finalité de l'évaluation et sur l'objet qu'elle s'efforce de mesurer. Un tel débat encourage l'établissement d'un consensus autour du plan ainsi élaboré et des objectifs de l'évaluation ;

– l'analyse des types de connaissances et de compétences associées à de bons résultats fournit une base à partir de laquelle il devient possible de définir des normes ou des niveaux d'aptitude. A mesure qu'évolue la perception de ce que l'on mesure et la capacité d'interpréter les scores d'une échelle particulière, la base empirique permettant de communiquer un ensemble plus important d'informations à divers groupes intéressés peut être développée ;

– le plan d'évaluation permet d'identifier et d'appréhender les variables particulières qui sous-tendent de bons résultats et il renforce les possibilités d'évaluer l'objet mesuré et d'apporter des modifications à l'évaluation au cours du temps ;

– il établit un lien entre la recherche, l'évaluation et l'action des pouvoirs publics qui, non seulement incite à améliorer l'enquête et à l'utilisation de ses résultats, mais encore contribue à une meilleure appréhension de l'objet sur lequel elle porte.

Chapitre 1

LA COMPRÉHENSION DE L'ÉCRIT

Dans le cadre de l'étude PISA de l'OCDE, l'expression *compréhension de l'écrit* doit être entendue au sens large. Comme, dans nos sociétés, il existe relativement peu de jeunes adultes ne possédant pas de compétences en lecture, le plan d'évaluation n'est pas destiné à évaluer si oui ou non les élèves de 15 ans peuvent lire au sens technique du terme. Il constitue plutôt le reflet des opinions contemporaines sur la lecture : les élèves sur le point de quitter l'enseignement secondaire devraient être capables de construire du sens à partir de ce qu'ils ont lu, d'étendre leur interprétation et de réfléchir sur la signification de ce qu'ils lisent dans les divers types de textes continus et non continus qu'ils peuvent communément rencontrer dans un large éventail de situations, tant à l'école qu'en dehors de l'école.

Définition du domaine

Les définitions de la lecture et de la compréhension de l'écrit ont évolué dans le temps, parallèlement aux changements de société, d'économie et de culture. Le concept d'apprentissage et, particulièrement, le concept d'apprentissage tout au long de la vie, ont fait évoluer les opinions et les attentes vis-à-vis de la lecture. Celle-ci n'est plus seulement considérée comme une aptitude que l'on acquiert dans l'enfance, durant les premières années de scolarité, mais plutôt comme un ensemble de connaissances, de compétences et de stratégies en perpétuelle évolution, que les individus élaborent au fil de leur vie, dans divers contextes et en interaction avec leurs pairs.

Les conceptions de la compréhension de l'écrit fondées sur les sciences cognitives mettent l'accent sur la nature interactive de la lecture et la nature constructive de la compréhension (Bruner, 1990 ; Dole *et al.*, 1991 ; Binkley et Linnakylä, 1997). En réponse au texte, le lecteur construit du sens en utilisant ses connaissances antérieures et une série d'indices liés au texte et à la situation, qui sont souvent partagés par les membres de sa culture ou de sa société. Pour construire du sens, le lecteur utilise divers processus, compétences et stratégies servant à alimenter, réguler et maintenir sa compréhension. On peut s'attendre à ce que ces processus et ces stratégies soient variables en fonction des contextes et des objectifs que poursuivent les lecteurs lorsqu'ils interragissent avec divers types de textes continus et non continus.

Deux enquêtes internationales récentes sur la lecture (l'enquête *Reading Literacy Study* conduite par l'Association internationale pour l'évaluation du rendement scolaire – IEA/RLS, et l'*Enquête internationale sur la littératie des adultes* entreprise conjointement par l'OCDE et Statistique Canada – EILA) ont également souligné la nature fonctionnelle de la lecture. Dans l'enquête de l'IEA, la compréhension de l'écrit est définie comme :

> *« l'aptitude à comprendre et à utiliser les formes écrites de la langue qui sont requises par la société et/ou qui sont valorisées par l'individu ».*

L'*Enquête internationale sur la littératie des adultes* (EILA) a, de même, mis l'accent sur le côté fonctionnel de la lecture en insistant particulièrement sur son rôle dans le développement de l'individu et de la société. Plutôt que de se focaliser sur les formes linguistiques, la définition met en avant la notion d'information. La compréhension de l'écrit y est définie comme :

> *« l'usage de l'information imprimée et écrite pour fonctionner dans la société, pour atteindre ses objectifs et pour développer ses connaissances et son potentiel ».*

OCDE 1999

Ces définitions de la compréhension de l'écrit sont centrées sur l'aptitude du lecteur à utiliser des textes écrits ou imprimés à des fins requises par la société ou valorisées par les individus pour développer leurs connaissances et leur potentiel. Elles vont au-delà du simple décodage et de la compréhension littérale, et impliquent que la compréhension de l'écrit englobe à la fois la compréhension et l'usage d'informations écrites à des fins fonctionnelles. Toutefois, ces définitions ne mettent pas en valeur le rôle actif et les initiatives du lecteur dans le processus de compréhension ou d'utilisation de l'information. Pour l'étude OCDE/PISA, la définition de la compréhension de l'écrit sera donc la suivante :

« Comprendre l'écrit, c'est non seulement comprendre et utiliser des textes écrits, mais aussi réfléchir à leur propos. Cette capacité devrait permettre à chacun(e) de réaliser ses objectifs, de développer ses connaissances et son potentiel, et de prendre une part active dans la société[2]. »

Il est nécessaire de donner quelques explications pour rendre plus claire cette définition du domaine concerné.

Comprendre l'écrit…

L'expression *compréhension de l'écrit* a été utilisée de préférence à *lecture* parce qu'elle est susceptible de mieux traduire, pour un public de non-spécialistes, l'idée de ce que l'enquête va mesurer. Par *lecture*, on entend souvent un simple décodage, ou même le fait de lire à haute voix, alors que l'intention de cette enquête est de mesurer quelque chose de bien plus vaste et de bien plus profond. Nous voulons mettre l'accent sur le fait que la lecture intervient dans une série de situations et sert des fins multiples. Historiquement, le terme anglais *literacy* a désigné l'outil dont les lecteurs se servent pour acquérir des connaissances. Le terme français *littératie* ne conviendrait pas, car il est trop souvent associé à la notion d'analphabétisme/illettrisme ou à toute autre façon de désigner un niveau minimum de savoir-faire nécessaire pour fonctionner dans une société donnée. Ceci étant dit, la référence à la lecture en tant qu'outil est proche de la connotation que le terme *compréhension de l'écrit* est supposé prendre dans l'étude OCDE/PISA, qui concerne une population d'élèves dont le spectre de compétences est très large. Certains de ces élèves poursuivront leurs études en s'inscrivant dans une université, d'autres s'inscriront dans une formation secondaire ou tertiaire préparant à une entrée dans la vie professionnelle, d'autres encore se mettront directement à travailler après avoir accompli leur scolarité obligatoire. Quelles que soient les aspirations en termes de cursus scolaire ou de profession future, on attend de la part de ces élèves qu'ils deviennent des membres actifs de leurs communautés respectives.

… c'est non seulement comprendre, mais aussi utiliser et réfléchir…

Les mots « *réfléchir à leur propos* » ont été ajoutés à « *comprendre* » (tiré de IEA/RLS) et « *utiliser* » (tiré de IEA/RLS et de l'EILA) pour insister sur le caractère interactif de la lecture : face à un texte, les lecteurs mobilisent leur pensée et leur expérience. Cela peut impliquer de la part du lecteur une réflexion sur le contenu du texte qui met en jeu des connaissances ou des concepts acquis au préalable, ou une réflexion sur la structure ou la forme du texte.

… à propos de textes écrits…

Les mots *« textes écrits »* désignent des textes, qu'ils soient imprimés, manuscrits ou affichés électroniquement, dans lesquels la langue est utilisée. Sont inclus dans cette définition les supports visuels, tels que des diagrammes, des photos, des cartes géographiques et des tableaux ou des graphiques, mais non les films, images télévisées, dessins animés ou autres types d'images non accompagnées

2. Bien que les hypothèses qui sous-tendent la définition de la compréhension de l'écrit soit généralement bien comprises dans les pays participants, des mots spécifiques n'existent pas dans certaines langues. Pour montrer que le sens donné à cette section peut se traduire dans d'autres langues sans modifier la signification sous-jacente du terme *« compréhension de l'écrit »* ou les hypothèses qui ont été élaborées autour de sa définition, des traductions de cette section ont été faites et sont accessibles sur le site web du programme OCDE/PISA : http://www.pisa.oecd.org.

d'écrits. Ces textes visuels peuvent se présenter seuls ou être inclus dans des textes continus. Par textes électroniques, on entend également les textes ayant un format électronique, bien que certains d'entre eux puissent se différencier des textes imprimés par la structure et le format, et puissent faire appel à des stratégies de lecture différentes. Il est prévu que des textes électroniques seront utilisés dans les prochains cycles de l'enquête PISA mais ils ne seront pas inclus dans ce premier cycle pour des raisons de temps et d'accessibilité. Le terme « *textes* » a été choisi de préférence au mot *« informations* », utilisé dans la définition de l'*Enquête internationale sur la littératie des adultes* (EILA), car on a estimé que cette expression ne convenait pas pour un ensemble incluant des textes littéraires.

… pour parvenir à ses objectifs, développer ses connaissances et son potentiel, et prendre une part active dans la société

Par cette phrase, on entend désigner l'ensemble des situations où la compréhension de l'écrit joue un rôle : de la vie privée à la vie publique, de l'école au travail, au cours de toute une vie d'apprentissage et de citoyenneté active. « *Parvenir à ses objectifs, développer ses connaissances et son potentiel* » contient l'idée que la compréhension de l'écrit permet l'accomplissement des aspirations individuelles, qu'elles soient bien définies (obtention d'un diplôme ou d'un travail), ou qu'elles le soient de manière moins précise et moins immédiate (enrichissement et développement de la vie personnelle, poursuite de l'apprentissage au cours de la vie). « *Prendre une part active* » a été utilisé de préférence à « *fonctionner* » (Enquête EILA), parce que cette expression implique que la compréhension de l'écrit permet aux gens d'apporter une contribution à la société, autant que de répondre à leurs propres besoins. Le terme « *fonctionner* » comporte une connotation limitative et pragmatique, alors que « *prendre une part active* » suppose un engagement social, culturel et politique. Prendre une part active peut correspondre à l'adoption d'une position critique, l'accomplissement d'un pas vers la libération personnelle, à l'émancipation et à la prise de responsabilités. Le terme « *société* » désigne tant la vie économique et politique que la vie sociale et culturelle (Linnakylä, 1992 ; Lundberg, 1991, 1997 ; MacCarthey et Raphael, 1989).

Organisation du domaine et caractéristiques des épreuves

Après avoir défini la compréhension de l'écrit et avoir fait état de l'ensemble des principes sur lesquels nous nous sommes fondés pour élaborer cette définition, il importe de fixer un cadre conceptuel afin d'organiser ce domaine. Cette organisation doit tenir compte de la façon de présenter les scores qui résulteront de l'administration de la batterie d'épreuves de lecture. C'est une question importante, car la façon dont est organisé le domaine peut affecter la conception des tests. Des travaux de recherche suggèrent que la lecture n'est pas une compétence isolée, unidimensionnelle, et que par conséquent la compréhension de l'écrit ne peut être représentée de façon appropriée par une échelle unique ou par un seul résultat sur cette échelle. Déterminer le nombre et les caractéristiques des échelles qu'il y aura lieu d'utiliser pour présenter les scores de compréhension de l'écrit constitue un point crucial : il faut s'assurer qu'un nombre suffisant d'items sont développés pour définir et interpréter ces échelles de manière adéquate.

On peut envisager différentes perspectives pour organiser les échelles. Le plus facile consisterait à se fonder sur les travaux antérieurs de même nature, menés aux niveaux national et international. Le programme américain d'évaluation en lecture *National Assessment of Educational Progress* (NAEP) et l'étude de l'IEA sur la lecture ont tous deux traduit les scores sur trois échelles, liées essentiellement au format des textes. L'étude de l'IEA traduit ses résultats pour les élèves des 4e et 9e années d'études sur des échelles portant respectivement sur des textes *narratifs*, des textes *informatifs* et des *documents*. L'enquête NAEP, avec la même approche, a présenté les résultats sur trois échelles : une échelle pour les textes *littéraires* (lire pour l'expérience littéraire), une échelle pour les textes *informatifs* (lire pour être informé), et une échelle pour les *documents* (lire pour accomplir une tâche). Ces échelles sont également voisines de celles utilisées lors de l'*Enquête internationale sur la littératie des adultes* (EILA). En plus d'une échelle quantitative, l'EILA a développé des échelles de compétences relatives à la *prose* et aux *documents*. Dans

OCDE 1999

cette étude, la *prose* concernait surtout des textes informatifs, alors que dans les enquêtes de la NAEP portant sur des populations scolaires, un meilleur équilibre a été recherché entre textes narratifs et textes informatifs.

La deuxième façon d'envisager l'organisation des épreuves de lecture consisterait à considérer les situations à propos desquelles les tâches sont construites. L'un des objectifs du programme OCDE/PISA est de mesurer la compréhension de l'écrit non seulement dans un contexte scolaire, mais aussi dans une variété d'autres situations, dans la mesure où l'évaluation vise à établir si les élèves de l'âge ciblé sont préparés à entrer dans le monde du travail et à jouer un rôle actif en tant que membres de leur communauté.

Une autre manière encore d'organiser et de situer internationalement les niveaux de compréhension de l'écrit des élèves consiste à utiliser un schéma basé sur le type de contenus des épreuves – une des distinctions qui a été proposée est « *textes à caractère technique* » par opposition à *« textes relatifs aux sciences humaines* ».

Une autre modalité de structuration pourrait se fonder sur des aspects de la lecture tels que comprendre globalement un texte ; trouver l'information ; développer une interprétation ; réfléchir sur le contenu d'un texte ; réfléchir sur la structure et la forme d'un texte. La compréhension approfondie d'un texte implique que les lecteurs sont engagés dans tous les aspects de la lecture (Langer, 1995) et cela constitue un schéma d'organisation majeur dans le cadre du développement des tâches destinées à cette enquête.

Il y a tout lieu de penser que le programme OCDE/PISA d'évaluation de la compréhension de l'écrit va fournir une riche moisson de données, à partir desquelles une décision sera prise sur la façon la plus efficace de rendre compte des résultats aux décideurs politiques, aux éducateurs et aux chercheurs. Cette décision sera prise en fonction de la combinaison de trois critères, d'ordre conceptuel, empirique et politique. Les décisions définitives sur les échelles qui seront utilisées pour présenter les résultats seront prises lorsque les données de l'essai de terrain auront été recueillies et analysées.

En plus d'organiser le domaine de la compréhension de l'écrit pour pouvoir rendre compte des résultats, il faut également être attentif à l'identification et à un début d'opérationalisation des caractéristiques des épreuves, afin de pouvoir commencer à élaborer et à sélectionner des tâches. On ne peut manipuler qu'un nombre fini de caractéristiques dans l'élaboration des tâches, et seul un petit nombre de variables associées à ces caractéristiques est susceptible d'influencer de manière sensible les résultats des élèves.

D'après Almond et Mislevy (1998), les variables décrivant les caractéristiques des épreuves peuvent jouer un des cinq rôles suivants dans un test :

– limiter le champ de l'évaluation ;

– caractériser les aspects à utiliser lors de la construction des épreuves ;

– contrôler le regroupement des épreuves dans les carnets ou les sous-sections du test ;

– caractériser les performances des sujets ou les réponses qu'ils doivent apporter aux items ;

– contribuer à caractériser certains aspects des compétences ou des capacités évaluées.

Certaines de ces variables peuvent à la fois contribuer à l'élaboration des épreuves et à la compréhension des compétences évaluées, ainsi qu'à la caractérisation des performances. Il y a un nombre fini de caractéristiques susceptibles d'influencer les performances des élèves pour un ensemble d'épreuves de lecture, et il faut en tenir compte lors de l'élaboration de ces tâches et de la correction des réponses. Ces caractéristiques sont des composantes du processus de lecture qui feront l'objet de manipulations lors du survey PISA sur la compréhension de l'écrit, afin de stimuler et d'évaluer la nature interactive du processus de lecture. Ces caractéristiques sont les suivantes :

– ***La situation –*** Puisque les adultes qui lisent des textes le font toujours dans une situation particulière, il est important d'identifier une série de situations à partir desquelles on peut échantillonner du matériel destiné à l'évaluation en lecture. Il faut garder à l'esprit qu'un des objectifs

poursuivis par le projet OCDE/PISA est de ne pas se contenter de textes scolaires mais d'inclure une série de documents que les élèves rencontreront aussi bien dans les salles de classe qu'en dehors de l'école.

– ***Les textes –*** Il va de soi qu'une épreuve de compréhension de l'écrit doit comporter une série de matériaux de lecture ; mais ce qui est crucial pour la conception du test et l'interprétation des scores obtenus, c'est l'éventail et les caractéristiques spécifiques des textes choisis pour l'élaboration des épreuves. Une palette diversifiée de textes tant continus que non continus sera donc utilisée dans ce survey et il faudra veiller à proposer des épreuves où l'élève aura à lire tantôt un texte isolé, tantôt une combinaison de textes (par exemple, deux textes continus traitant d'un même sujet, ou bien un texte continu et un graphique ou un tableau qui l'illustrent).

– ***Les modalités des questions-réponses –*** Cette expression désigne les caractéristiques des questions et des consignes des tâches qui seront données aux élèves, les formats utilisés pour obtenir les réponses ainsi que les grilles de correction qui seront appliquées aux réponses. En général, les questions et les consignes font référence à un but ou un objectif dont le lecteur est censé tenir compte au cours de sa lecture et de son interaction avec le texte. L'enquête sur la lecture ne doit pas uniquement reposer sur l'utilisation de questions à choix multiple, mais doit également comporter des questions à réponses ouvertes, dont le but est d'amener les élèves à mettre en oeuvre des stratégies et des processus plus variés et de niveau taxonomique plus haut.

Pour pouvoir utiliser ces trois caractéristiques principales des épreuves lors de la conception de l'instrument et, plus tard, lors de l'interprétation des résultats, il faut les rendre opérationnelles. Cela signifie qu'il faut spécifier les diverses valeurs que peut prendre chacune de ces caractéristiques. Cela permettra aux concepteurs d'items de classer par catégories les supports écrits avec lesquels ils travaillent et les questions qu'ils élaborent, afin qu'on puisse ensuite les utiliser pour organiser la présentation des données et l'interprétation des résultats. Ces variables peuvent également servir à spécifier quel poids doit être donné dans le test à chacune des catégories. Le rôle exact des variables lors de l'interprétation est bien évidemment une question empirique, mais elles ne peuvent jouer aucun rôle si elles ne sont pas intégrées dès le départ dans la conception du test.

Situations

La notion de *situation* se réfère davantage à l'usage auquel l'auteur destine son texte qu'aux circonstances dans lesquelles le texte est effectivement lu. Certes, il est dans notre intention d'évaluer les types de lecture associés à l'école, comme ceux qui prennent place en dehors de l'école ; cependant, la façon dont est spécifiée la situation ne peut se borner à indiquer l'endroit où se déroule l'activité de lecture. Par exemple, les manuels sont lus aussi bien à l'école qu'à la maison, et il est probable que les processus et les objectifs de la lecture se différencient très peu d'un endroit à l'autre. De même, ainsi que l'a démontré Hubbard (1989), certains types de lectures destinées aux enfants et habituellement associées à des cadres non scolaires (règlements d'un club, compte-rendu de matches) prennent souvent place aussi, de manière non officielle, à l'école.

Bien que, dans de nombreux cas, la lecture constitue une activité solitaire, elle n'en comporte pas moins une facette sociale. D'autres personnes sont impliquées : les auteurs, les personnages, et les personnes qui imposent la tâche de lire (par exemple, les enseignants). La situation fait donc aussi référence aux personnes (et, pour la lecture effectuée dans un cadre professionnel, aux objets) mis en jeu par l'activité de lecture.

Dans le cadre de l'évaluation OCDE/PISA des élèves de 15 ans, la dimension « situation » peut être considérée comme une classification des tâches de lecture qui est fondée sur l'usage auquel elles sont destinées, sur la relation à d'autres qui est explicite ou implicite dans la tâche, et sur le contenu en général (voir le tableau 2). Ainsi, lire un manuel est un exemple de situation scolaire, puisque son usage premier est l'acquisition d'informations dans le cadre d'une tâche scolaire (usage), cette activité est associée aux consignes émises par l'enseignant ou d'autres instructeurs (les autres), et son contenu est généralement orienté vers l'instruction et l'apprentissage (contenu).

OCDE 1999

Tableau 2. **Variables de situation pour la lecture**

	Lecture à usage privé	Lecture à usage public	Lecture à des fins professionnelles	Lecture à des fins scolaires
Les autres	Soi-même Proches Amis	Anonyme	Objets Collègues Supérieurs hiérarchiques	Instructeurs
Usage	Curiosité Contact	Information	Pour agir	Pour apprendre
Contenus	Lettres Fiction Biographies Modes d'emploi Livres et magazines Cartes géographiques	Notices Règlements Programmes Brochures Formulaires	Instructions Manuels Horaires Mémos Rapports Tableaux/graphiques	Textes Cartes Schémas Tableaux Graphiques

Dans le cadre de cette évaluation, le contenu n'est pas une variable qui fait l'objet de manipulations spécifiques. Cependant, en échantillonnant des textes relatifs à des situations très variées, notre intention est bien de diversifier autant que possible les contenus inclus dans l'enquête OCDE/PISA sur la compréhension de l'écrit. On accordera une attention toute particulière tant à l'origine et au contenu des textes sélectionnés pour cette enquête qu'aux types de questions et de consignes utilisés pour obtenir des informations sur la compréhension de l'écrit chez les élèves. Le but est de parvenir à un équilibre entre le développement de tâches qui reflètent le mieux possible la définition globale de la compréhension de l'écrit utilisée dans le cadre du programme OCDE/PISA, et la nécessité que ce matériel soit représentatif de la diversité linguistique et culturelle des pays participants. Cette diversité permettra d'avoir la garantie qu'aucun groupe ne sera avantagé ou désavantagé par le contenu de l'évaluation.

Les travaux sur le langage réalisés par le Conseil de l'Europe (1996) fournissent une structure utile pour l'opérationalisation des variables de situation :

– *La lecture à usage privé* (personnel) : on effectue ce type de lecture pour satisfaire ses intérêts personnels, qu'ils soient d'ordre pratique ou intellectuel. Ceci inclut également la lecture effectuée dans le but de maintenir, ou de développer, des liens personnels avec d'autres personnes. Parmi les contenus typiques de ce genre de lecture figurent les lettres personnelles, les textes de fiction, les biographies ainsi que les textes informatifs lus par curiosité, dans le cadre d'activités récréatives et de loisirs.

– *La lecture à usage public :* ce type de lecture est pratiqué pour prendre part à des activités sociales plus larges. Cela comprend l'usage de documents officiels ainsi que des informations concernant des événements publics. Généralement, ces tâches sont associées à des contacts avec les autres qui demeurent relativement anonymes.

– *La lecture à des fins professionnelles :* bien que seuls quelques jeunes de 15 ans aient réellement à lire dans un cadre professionnel, il est important d'inclure des tâches représentant des exemples typiques d'activités de lecture sur les lieux de travail : celles-ci ont souvent un lien étroit avec telle ou telle tâche à effectuer immédiatement et leur contenu est particulièrement pertinent dans le cadre de ce projet. Il importe également d'évaluer si les jeunes de 15 ans sont prêts à entrer dans le monde du travail puisque beaucoup d'entre eux rejoindront la population active un ou deux ans après le moment où les élèves sont testés. On fait souvent référence à ce genre de tâches en les qualifiant de « *lecture pour agir* » (Sticht, 1975 ; Stiggins, 1982).

– *La lecture à des fins éducatives :* ce type de lecture, qui a pour finalité l'acquisition d'informations, s'inscrit normalement dans le cadre plus vaste d'une tâche d'apprentissage. Le support écrit n'a généralement pas été choisi par le lecteur, mais donné à lire par un enseignant. Son contenu est habituellement conçu à des fins spécifiques d'enseignement. On fait souvent référence à ce genre de tâches en les qualifiant de « *lecture pour apprendre* » (Sticht, 1975 ; Stiggins, 1982).

Les textes

L'acte de lecture nécessite un support (quelque chose que le lecteur doit lire). Dans un test, ce support doit avoir une certaine autonomie, c'est-à-dire que le texte doit se suffire à lui-même, sans nécessiter du matériel imprimé supplémentaire[3]. Il existe un grand nombre de textes différents et il va de soi que toute évaluation doit s'efforcer d'inclure autant de genres différents que possible ; cependant, il n'existe pas forcément une classification idéale des types de textes. Beaucoup de catégorisations ont été proposées ; bon nombre d'entre elles ont été créées avec une visée pratique plutôt que théorique. Toutes présentent la même caractéristique : aucun texte réel ne semble s'inscrire vraiment dans une seule catégorie. Par exemple, un chapitre de manuel peut comprendre des définitions (souvent identifiées comme un type de texte en soi), des consignes sur la façon de résoudre un problème particulier (un autre genre de texte), un bref historique relatant la découverte de la solution (encore un autre type de texte) et des descriptions des objets particuliers impliqués dans la solution (quatrième type).

On pourrait estimer qu'une définition, par exemple, pourrait être isolée et traitée comme un texte unique à des fins d'évaluation. Mais cela supprimerait le contexte de la définition et créerait un type de texte artificiel (il est rare que les définitions soient fournies de manière isolée, si ce n'est dans les dictionnaires) ; et cela ne permettrait pas aux développeurs de tests de créer des tâches de lecture demandant à l'élève d'intégrer des informations provenant d'une définition avec des informations provenant de consignes, par exemple.

Certains textes se présentent comme reflétant le monde tel qu'il est (ou était), et sont donc considérés comme factuels ou non fictionnels. Les récits fictionnels supposent une relation plus métaphorique au monde tel qu'il est, et se présentent sous la forme de récits décrivant comment il pourrait être ou comment il paraît être. Cette distinction s'estompe lorsque les auteurs utilisent des formats et structures propres aux textes factuels en créant leur fiction. L'évaluation OCDE/PISA proposera des textes factuels et fictionnels (ainsi que des textes dont il sera sans doute malaisé de dire s'ils sont du premier ou du second type), mais ne tentera pas de mesurer des différences de compétences en lecture basées sur l'un ou l'autre de ces types de textes.

Une classification des textes bien plus importante et qui est l'une des dimensions sur lesquelles se fonde l'organisation de l'évaluation OCDE/PISA, fait une distinction entre textes continus et non continus. Les textes continus sont spécifiquement constitués de phrases, elles-mêmes organisées en paragraphes. Ceux-ci peuvent à leur tour s'inscrire dans des structures plus vastes (des sections, des chapitres, des livres). Les textes non continus s'organisent le plus souvent en format matriciel, basé sur des combinaisons de listes.

Les textes continus sont conventionnellement constitués de phrases, elles-mêmes organisées en paragraphes. Dans ces textes, l'organisation est indiquée au moyen de paragraphes, de retraits et de la décomposition du texte selon une hiérarchie signalée par des titres qui aident les lecteurs à la reconnaître. Ces divers indices renseignent aussi sur les limites du texte (en signalant la fin d'une section, par exemple). Le repérage des informations est souvent facilité par l'usage de caractères de différentes tailles ou polices (italique, gras), d'encadrements et de mises en page particulières. Pouvoir utiliser ce type d'indices constitue chez le lecteur une des sous-compétences essentielles pour lire efficacement.

L'information sur l'organisation du texte peut également être véhiculée par les marques du discours. Des connecteurs séquentiels (*premièrement, deuxièmement, troisièmement,* etc.), par exemple, signalent le rapport de chacune des unités qu'ils introduisent et indiquent comment les unités sont reliées à l'ensemble du texte.

La classification principale des textes continus se fonde sur leur objectif rhétorique, c'est-à-dire sur le type de textes.

3. Cela n'exclut pas l'usage de plusieurs textes pour une seule tâche, mais chacun d'eux doit être cohérent par lui-même.

OCDE 1999

Les textes non continus, ou documents, ainsi qu'on les désigne parfois, peuvent être classés de deux manières. La première est l'approche par la structure formelle qu'ont utilisée Kirsch et Mosenthal dans leurs travaux[4] ; ces auteurs catégorisent les textes selon la manière dont sont agencées les listes sous-jacentes qui composent les différents types de textes non continus. L'autre approche utilise des descriptions ordinaires de ces textes. La méthode de Kirsch et Mosenthal est systématique et propose une façon d'organiser l'ensemble des types de textes non continus, qui est indépendante de leur usage.

Types de textes continus

Les types de texte correspondent à des façons standardisées de classer les contenus et les objectifs de l'auteur pour des textes continus. Chaque type se présente sous des formats spécifiques. Ceux-ci sont décrits ci-dessous[5].

1. Le *texte descriptif* est un type de texte où l'information porte sur les propriétés physiques des objets *dans l'espace* ou sur les caractéristiques des gens. La question typique à laquelle les textes descriptifs apportent une réponse est « *quoi ?* ».
   - Les *descriptions « impressionnistes »* présentent l'information du point de vue d'un sujet (relations, qualités et orientations dans l'espace).
   - Les *descriptions techniques* présentent l'information du point de vue d'une observation objective dans l'espace. Elles utilisent fréquemment des formats de texte non continus, comme les diagrammes et les illustrations.
2. Le *texte narratif* est un type de texte où l'information concerne les propriétés des objets *dans le temps*. La narration répond typiquement à la question « *quand ?* », ou à la question « *dans quel ordre* ? »
   - Les *récits* présentent les changements survenus sous un angle subjectif (sélection et emphase), relatant des actions et des événements d'un point de vue subjectif.
   - Les *rapports* présentent les changements survenus d'un point de vue objectif, relatant des actions et des événements qui peuvent être vérifiés par d'autres personnes.
   - Les *articles d'actualités* sont censés permettre aux lecteurs de se forger leurs propres opinions sur les faits et les événements relatés sans être influencés par les opinions personnelles du journaliste.
3. Le *texte informatif* est un type de texte dans lequel l'information se présente sous la forme d'un ensemble de concepts ou de *constructs* (voir note 1, p. 19) mentaux, ou sous la forme d'éléments constitutifs de ces concepts et *constructs*. Le texte fournit une explication sur la façon dont ces composantes s'associent pour former un tout significatif et répondent souvent à la question « *comment* ? »
   - Les *essais informatifs* fournissent une explication simple des concepts, constructions mentales ou conceptions d'un point de vue subjectif.
   - Les *définitions* expliquent de quelle façon les termes ou les noms sont reliés aux concepts mentaux. En montrant ces interrelations, la définition explique la signification des « mots ».
   - Les *explications* sont une forme d'exposé analytique utilisée pour expliquer comment un concept mental peut être relié à des mots ou des termes. Le concept est traité comme un tout composite qui peut être compris en étant décomposé en éléments constituants et leurs interrelations, chacun d'entre eux recevant un nom.

4. Le modèle de Kirsch et Mosenthal a été présenté en détail dans une série d'articles mensuels intitulés « Understanding Documents », parus entre 1990 et 1992 dans le *Journal of Reading*.
5. Ce paragraphe se base sur le travail de Werlich (1976). A noter que la catégorie « hypertexte » ne relève pas du schéma de Werlich.

– Les *résumés* sont une forme d'exposé synthétique utilisée pour expliquer et communiquer des informations à propos de « textes » sous une forme plus courte que le texte original.

– Les *procès verbaux* sont des rapports plus ou moins officiels des résultats de réunions ou de communications.

– Les *interprétations de texte* sont une forme d'exposés à la fois analytique et synthétique utilisés pour expliquer les concepts abstraits qui sont contenus dans un texte particulier (fictionnel et non fictionnel) ou dans un groupe de textes.

4. Le *texte argumentatif* est un type de texte qui présente des propositions relatives aux rapports entre des concepts ou d'autres propositions. Les textes argumentatifs répondent souvent à la question « *pourquoi ?* ». Une autre sous-catégorie importante des textes argumentatifs est celle des textes *persuasifs*.

   – Le *commentaire* établit un lien entre les concepts relatifs à des événements, à des objets et des idées et un système personnel de pensée, de valeurs et de croyances.

   – L'*argumentation scientifique* établit un lien entre les concepts relatifs à des événements, à des objets et des idées, et des systèmes de pensées et de connaissances, tel que les propositions qui en résultent peuvent être vérifiées ou infirmées.

5. Le *texte donnant des instructions* (appelé parfois texte *injonctif*) est un type de texte qui fournit des directives sur ce qu'il faut faire.

   – Les *consignes* sont des directives destinées à guider le comportement pour accomplir une tâche.

   – Les *règles, règlements* et *statuts* spécifient les exigences concernant certains comportements. Ces textes fondent leur légitimité sur une autorité impersonnelle (contraintes pratiques ou autorité publique).

6. L'*hypertexte* est un ensemble de fragments de textes reliés les uns aux autres en sorte que les unités peuvent être lues dans des ordres séquentiels différents, ce qui permet aux lecteurs de suivre différents chemins véhiculant l'information. Ces textes contiennent fréquemment des supports visuels et peuvent solliciter chez le lecteur des stratégies d'approche non linéaires.

Textes non continus (structure et format)

Les textes non continus sont organisés différemment des textes continus et font donc appel à des démarches de lecture de types différents. Il y a deux manières pratiques de considérer ces textes. La première classification tient compte des principes selon lesquels les éléments du texte sont organisés. Cette variable *structure du texte* identifie dans les textes non continus des caractéristiques dont la fonction est semblable à celle que jouent les caractéristiques des phrases et des paragraphes dans les textes continus. La seconde approche identifie un certain nombre de *formats* habituels de ce type de texte.

• Textes non continus : classement par structures

On peut considérer que tous les textes non continus sont composés d'un certain nombre de listes. Certains d'entre eux sont seulement des listes simples, mais la plupart d'entre eux combinent plusieurs listes. L'analyse ne fait pas référence à l'usage fait de ces listes ni n'utilise les classifications habituelles souvent appliquées aux textes non continus. Elle identifie plutôt les principales caractéristiques structurelles communes à de nombreux textes différents. Pour décrire de manière complète un quelconque texte non continu, il est nécessaire de le catégoriser à la fois selon sa structure et selon son format. Les lecteurs qui comprennent la structure des textes sont mieux à même d'identifier les rapports entre leurs divers éléments et de comprendre quels textes sont semblables et quels textes sont différents.

1. Les *listes simples* contiennent une seule collection d'éléments. La liste des livres à lire pour le cours de littérature constitue un exemple de liste simple, tout comme la liste des élèves au tableau d'honneur. Les éléments d'une liste peuvent être ordonnés, comme lorsque les élèves d'une classe sont rangés selon l'ordre alphabétique de leur nom de famille, ou non

OCDE 1999

ordonnés, comme la liste des fournitures à acheter pour un cours artistique. La recherche à effectuer est plus facile dans le premier cas. Si la liste non ordonnée est longue, il peut s'avérer difficile de déterminer si un article figure sur la liste ou non. Si la liste est ordonnée, on peut déterminer facilement si un élément y figure ou non, à condition de connaître le principe de classement.

2. Les *listes combinées* sont constituées de deux listes simples, ou davantage, sur lesquelles chaque élément d'une liste va de pair avec un élément d'une autre liste. Une des listes peut être considérée comme la liste primaire (liste d'indexation) ; la liste primaire est ordonnée pour faciliter la localisation des éléments qui y figurent, ce qui permet de localiser aisément l'information parallèle située sur les autres listes. Un exemple de liste élémentaire combinée pourrait être une liste d'élèves accompagnée de la liste correspondante de leurs scores à un test d'évaluation. Les éléments peuvent apparaître plus d'une fois sur une des listes, bien que cela soit rarement le cas de la liste primaire. Par exemple, sur une liste d'élèves et de scores, il est possible qu'un même score apparaisse plusieurs fois. Une liste combinée peut comporter plusieurs listes composantes : c'est le cas, par exemple, pour une liste de chansons populaires sur laquelle figurent le titre de la chanson, le nom du chanteur, la maison de disques et le nombre de semaines durant lesquelles la chanson a figuré au hit-parade. Les recherches sur une liste non indexée peuvent s'avérer plus difficiles et il peut être malaisé de savoir si on a trouvé toutes les informations que l'on cherche. Ainsi, il sera facile d'utiliser la liste des élèves et des scores pour savoir la note obtenue par un élève donné, surtout si les noms des élèves sont classés par ordre alphabétique. Il peut s'avérer plus difficile de trouver tous les élèves qui n'ont pas obtenu une note suffisante au test d'évaluation.

3. Une *liste croisée* est constituée de trois listes, qui ne sont pas parallèles, mais croisées, et qui forment une matrice de rangées et de colonnes. Un exemple typique de liste croisée est le programme de télévision, qui est constitué d'une liste d'heures, d'une liste de chaînes et d'une liste de programmes. Les programmes apparaissent dans les cellules formées par l'intersection d'une heure (figurant habituellement dans une des colonnes) et d'une chaîne (figurant habituellement sur une rangée). Un département universitaire présentera un tableau des cours proposés sous forme de matrice, où les colonnes représentent les jours, les rangées les heures, et où figurent dans les cellules les cours qui sont proposés à une certaine heure, un certain jour ; cela permet aux étudiants de repérer aisément quels sont les cours qui ne se chevauchent pas dans le temps. Sur une liste croisée, les cellules sont toutes d'un type simple (intitulés de cours, programmes télévisés, etc.). De nombreux tableaux « statistiques » sont des listes croisées. Par exemple, un tableau où figurent les taux de chômage des grandes villes présentera probablement les villes dans les rangées, les diverses dates dans les colonnes et, dans les cellules, le taux de chômage relevé pour une ville au cours d'une période déterminée. Le tableau peut être destiné à comparer les différences de taux de chômage selon les dates ; dans ce cas, il aura plusieurs colonnes, chacune représentant une période différente (mois, années, etc.).

4. Les *listes emboîtées* consistent en un ensemble de listes combinées. Ainsi, dans certaines listes croisées, les catégories en colonnes (par exemple, les jours de la semaine), ne se trouvent pas seulement en intersection avec les rangées (par exemple, les heures) mais aussi avec une quatrième liste (par exemple, celle des divers départements d'une université). Dans une véritable liste emboîtée, le même type de catégories doit être utilisé pour chacune des listes en intersection. La liste croisée des taux de chômage peut comporter des entrées séparées pour chaque mois pour les hommes et pour les femmes ; dans ce cas, le sexe est emboîté dans le mois.

5. Les *listes en combinaison* sont celles où plusieurs types de listes, ou plusieurs listes du même type, peuvent être rassemblées en une seule. Par exemple, la liste croisée créée pour présenter un tableau statistique des taux de chômage pour différents mois dans les grandes villes peut être combinée avec une autre liste croisée des variations saisonnières des taux de chômage pour ces mêmes villes.

• Textes non continus : classement par formats

Classer les textes non continus en catégories selon leur format ouvre une autre perspective sur ces textes. Rappelons que tout texte non continu peut être classé à la fois selon sa structure et selon son format. Par exemple, le formulaire est une des catégories de format, mais chaque formulaire a également une structure ; c'est le plus souvent une liste combinée dans laquelle une liste d'entrées va de pair avec une liste de blancs à compléter, destinés à contenir l'information qui correspond à l'entrée. Le format d'un horaire d'autobus, de train ou d'avion est celui d'un tableau dont la structure est normalement celle d'une liste croisée ou emboîtée. Reconnaître le format est important car des textes ayant une même structure peuvent avoir des mises en pages très différentes. Par exemple, la table des matières d'un livre et d'un formulaire sont généralement des listes combinées. Sur le formulaire, les deux listes sont celle des entrées et celle des blancs, comme indiqué plus haut. Dans une table des matières, les deux listes sont les titres des chapitres et les pages auxquelles commencent ces chapitres ; ces deux listes présentent le même parallélisme que les entrées et les blancs sur le formulaire. Mais personne ne confondrait un formulaire avec une table des matières.

1. Les *formulaires* sont des textes structurés et formatés qui demandent au lecteur de répondre à des questions spécifiques d'une manière spécifique. Un grand nombre d'institutions utilisent des formulaires pour recueillir des données. Ceux-ci comportent souvent des formats de réponse structurés ou pré-codés. Les déclarations d'impôts, les formulaires d'immigration, de visa, de demande d'emploi, les questionnaires utilisés lors d'enquêtes statistiques, etc., constituent des exemples typiques de formulaires.

2. Les *feuilles d'information*, contrairement aux formulaires, proposent une information au lieu de la demander. Elles résument l'information d'une façon structurée et dans un format tel que le lecteur peut facilement et rapidement localiser des informations particulières. Les feuilles d'information peuvent contenir diverses formes de textes, ainsi que des listes, des tableaux, des schémas, et une typographie sophistiquée (titres, polices, retraits, encadrements, etc.) pour résumer ou mettre en évidence l'information. Les horaires, les listes de prix, les catalogues, les programmes, constituent des exemples de ce genre de document.

3. Les *coupons* servent à certifier que leur détenteur est autorisé à utiliser tel ou tel service. L'information que contient le coupon permettra de savoir s'il est valide ou non. Des exemples typiques de coupons sont les tickets, les factures, etc.

4. Les *certificats* sont des écrits reconnaissant la validité d'un accord ou d'un contrat. C'est leur contenu qui est formalisé plutôt que leur format. Ils requièrent en général la signature d'une ou plusieurs personnes compétentes pour (ou habilitées à) garantir la véracité de la déclaration contenue dans le document. Les garanties, les certificats scolaires, les diplômes, les contrats, etc., sont des documents qui présentent ces caractéristiques.

5. Les *annonces et les réclames* sont des documents destinés à inviter le lecteur à faire quelque chose, par exemple, acheter des marchandises ou des services, participer à des réunions, des rassemblements, élire une personne pour une fonction officielle, etc. Le but de ces documents est de persuader le lecteur. Ils offrent quelque chose et, en même temps, requièrent de l'attention et appellent à une action. Les publicités, les invitations, les assignations, les avertissements, les notices, constituent des exemples de ce type de format.

6. Les *graphiques* constituent des représentations iconiques de données. Ils sont utilisés dans les exposés scientifiques et sont également présents dans des revues et des journaux pour diffuser au public des informations chiffrées dans un format visuel.

7. Les *diagrammes* accompagnent souvent les descriptions techniques (par exemple, pour illustrer les éléments d'un appareil électroménager), les textes informatifs et injonctifs (par exemple, pour présenter la façon de monter un appareil électroménager). Il est souvent utile de faire la distinction entre diagrammes de procédure (comment faire) et diagrammes de processus (comment cela fonctionne).

OCDE 1999

8. *Tableaux et matrices.* Les tableaux sont des matrices de colonnes et de rangées. Typiquement, toutes les entrées de chaque colonne et de chaque rangée ont des propriétés communes, ce qui fait que les titres des colonnes et des rangées font partie intégrante de la structure informative du texte. Parmi les tableaux les plus courants, citons les horaires, les tableurs, les formulaires de commande et les index.
9. Les *listes* sont les textes non continus les plus élémentaires. Elles consistent en un certain nombre d'entrées qui ont en commun la (les) même(s) propriété(s). Cette propriété commune peut servir d'étiquette ou de titre à la liste. Les entrées de la liste peuvent être ordonnées (par exemple, les noms des élèves de la classe rangés par ordre alphabétique), ou non ordonnées (par exemple, la liste d'articles à acheter dans un magasin).
10. Les *cartes* sont des textes non continus qui indiquent les relations géographiques entre des lieux. Il existe de nombreuses sortes de cartes. Les cartes routières indiquent les distances et les itinéraires entre des lieux donnés. Les cartes thématiques indiquent les rapports entre des endroits et des caractéristiques sociales ou physiques.

Modalités de questions/réponses et caractéristiques des tâches

Trois ensembles de variables sont concernés ici : les questions ou consignes, qui présentent au sujet la tâche à effectuer ; les formats de réponse, qui indiquent la façon dont les sujets sont appelés à montrer leur compétence à accomplir la tâche ; enfin, les règles de correction, qui spécifient comment les réponses de l'élève doivent être évaluées. Chacune de ces variables sera discutée ci-dessous, mais c'est la première d'entre elles qui va, de loin, requérir le plus d'attention.

Questions et consignes

On peut considérer l'ensemble des tâches proposées aux sujets testés d'un point de vue « macro » ou « micro ». Au niveau « macro », les tâches peuvent être identifiées selon qu'elles font appel à un ou plus des cinq grands aspects de la lecture. Au niveau « micro », il a été démontré que plusieurs variables peuvent influencer la difficulté des tâches figurant sur une échelle donnée.

• Aspects au niveau « macro »

Puisqu'on part du principe que la plupart des élèves de 15 ans qui participeront aux tests auront maîtrisé les principes de base du décodage de textes, il n'est pas nécessaire de faire une place explicite aux compétences techniques qui sous-tendent la lecture. L'évaluation sera plutôt centrée sur des stratégies de lecture plus complexes (Dole *et al.*, 1991 ; Paris, Wasik et Turner, 1991).

Pour tenter de simuler des situations authentiques de lecture, l'évaluation OCDE/PISA mesurera les cinq aspects suivants, associés à une pleine compréhension d'un texte, continu ou non continu. Les sujets auront à démontrer leurs compétences pour ces cinq aspects :

– comprendre globalement un texte ;

– trouver l'information ;

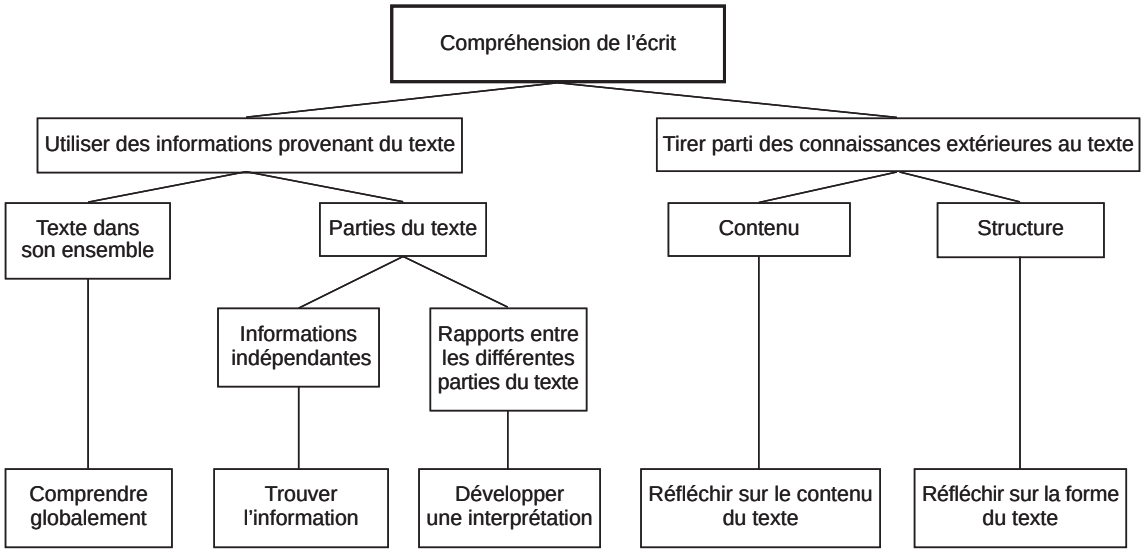
– développer une interprétation ;

– réfléchir sur le contenu d'un texte ;

– réfléchir sur les qualités formelles d'un texte.

La bonne compréhension d'un texte implique la mise en oeuvre de tous ces aspects. Tous les lecteurs, quel que soit leur niveau global, doivent être capables de faire preuve d'un certain niveau de compétence pour chacun d'entre eux (Langer, 1995). Bien qu'il existe un lien entre les divers aspects – chacun d'eux requiert plusieurs des mêmes compétences sous-jacentes –, la maîtrise de l'un n'est pas nécessairement dépendante de la maîtrise de l'autre. On considère souvent ces aspects comme faisant partie du répertoire de chaque lecteur à chaque étape de son développement, plutôt que comme un

Figure 1. **Caractéristiques permettant de distinguer les cinq aspects de la lecture**



ensemble hiérarchisé et séquentiel de compétences. Bien sûr, il sera possible et nécessaire d'explorer cette hypothèse une fois que les items auront été développés et que les données de l'essai de terrain auront été recueillies.

Ces cinq aspects de la lecture seront opérationnalisés par le biais d'une série de questions/consignes proposées aux élèves participant à l'évaluation OCDE/PISA. Par leurs réponses, les élèves pourront prouver qu'ils sont à même de comprendre et d'utiliser une série de textes continus et non continus et réfléchir à leur propos.

La figure 1 présente les caractéristiques clé qui permettent de distinguer les cinq aspects de la lecture que OCDE/PISA va mesurer. Chaque aspect y apparaît, certes, de manière simplifiée ; cependant, ce schéma fournit un cadre utile pour l'organisation et le rappel des rapports existant entre ces différents aspects.

On peut distinguer ces cinq aspects selon quatre dimensions :

– La première oppose les questions où il est demandé au lecteur d'utiliser des informations provenant essentiellement du texte à celles où il doit davantage tirer parti de connaissances extérieures au texte.

– La deuxième oppose les questions où l'on demande au lecteur de considérer le texte dans son ensemble à celles où il doit se concentrer sur des informations spécifiques contenues dans ce texte.

– Parfois, il est demandé au lecteur de retrouver des informations spécifiques ou indépendantes ; dans d'autres cas, on lui demande de montrer qu'il a compris les rapports existant entre des parties différentes du texte. La distinction entre informations isolées ou indépendantes *vs* rapports entre les différentes parties du texte constitue la troisième dimension.

– La quatrième caractéristique oppose les questions où l'on demande au lecteur de traiter le contenu ou la substance du texte à celles portant sur sa forme ou sa structure.

Les cinq aspects de la lecture sont représentés à la dernière ligne de la figure 1, c'est-à-dire l'extrémité des différentes branches. En partant du haut du schéma et en suivant chaque branche, on peut voir les dimensions qui sont associées avec chacun des aspects.

OCDE 1999

Bien que la figure 1 simplifie forcément le contenu et la complexité de chacun des aspects, nous nous efforçons de présenter ici une première ébauche de définition opérationnelle de ces aspects de la compréhension d'un texte et de les associer avec des types particuliers de questions et de consignes. Même si chaque aspect est discuté ci-après dans le cadre d'un texte unique, il faut comprendre qu'ils peuvent également s'appliquer à plusieurs textes lorsque ceux-ci constituent un ensemble[6]. La description de chaque aspect comprend deux parties. La première partie consiste en une approche globale de l'aspect, tandis que la seconde décrit des manières spécifiques d'évaluer celui-ci.

a) Comprendre globalement un texte

Souvent, le lecteur souhaite se limiter à une compréhension globale du texte avant de se décider s'il va le lire de manière plus approfondie. Au travers de cette première lecture, le lecteur habile peut déterminer si le texte, qu'il soit continu ou non continu, convient au but recherché.

Pour saisir globalement le texte qu'il a lu, le lecteur doit le considérer comme un tout, en avoir une vision d'ensemble. Cela ressemble à ce qui se produit lors de la première rencontre avec une personne ou un lieu. En un premier coup d'œil et une première réflexion, le lecteur avance des hypothèses ou fait des prédictions sur ce dont le texte traite. Il s'agit souvent d'une impression assez générale, mais elle est très importante pour la sélection du matériel de lecture le plus pertinent et le plus intéressant.

Les tâches relatives à cet aspect se fondent sur le texte lui-même, et sont de ce fait voisines de celles qui demandent au lecteur de trouver une information ou de développer une interprétation. Mais, contrairement à ces dernières catégories de tâches, celles qui demandent au lecteur de dégager une compréhension globale requièrent de lui qu'il saisisse l'essence du texte dans son ensemble – il doit expliquer de quoi il s'agit, identifier le sujet discuté, etc. Plusieurs composantes sont importantes, comme trouver l'idée principale ou le thème et identifier l'usage auquel est généralement destiné le document ou le texte non continu.

Divers types de tâches demandent une compréhension globale de la part du lecteur. Les élèves testés peuvent faire preuve de leur compréhension initiale en identifiant le sujet traité ou le message principal, en cernant l'objectif ou l'usage général du texte. Demander au lecteur d'imaginer le titre ou le thème du texte, ou d'expliquer l'ordre d'une série d'instructions simples, ou de dire quelles sont les dimensions principales d'un graphique ou d'un tableau, constituent des exemples de ce type de tâches. Autres exemples : décrire le personnage principal, le décor ou le milieu où se passe une histoire ; identifier le thème ou le message d'un texte littéraire ; expliquer l'objectif ou l'usage d'une carte ou d'un schéma ; identifier le sujet principal ou le destinataire d'un message électronique ; déterminer quel type de livre pourrait contenir un certain type d'information ; saisir globalement l'objectif d'une page de site Internet ou déterminer à qui elle s'adresse.

Certaines des tâches relatives à cet aspect peuvent requérir d'établir une correspondance entre un extrait du texte et la question. Cela peut être le cas, par exemple, lorsqu'un thème ou une idée maîtresse sont formulés explicitement dans le texte. D'autres tâches peuvent requérir de l'élève testé qu'il se focalise sur plus d'une référence spécifique du texte : c'est le cas, par exemple, lorsque le lecteur doit déterminer le sujet en se fondant sur la répétition d'une catégorie particulière d'informations. Sélectionner l'idée maîtresse implique d'établir une hiérarchie parmi les idées et de choisir les plus générales et les plus globales. Une tâche de ce type permet de vérifier si les élève testés sont capables de faire la distinction entre les idées essentielles et les détails mineurs, ou de reconnaître le résumé du thème principal dans une phrase ou dans un titre.

6. Notre intention est d'illustrer chacun des aspects par des exemples. Cependant, leur inclusion dans ce document destiné au public risque de compromettre la confidentialité des instruments du test PISA. Il est prévu que des exemples servant à définir et à distinguer chacun des aspects seront introduits après que les données de l'essai de terrain ont été recueillies et analysées et que la sélection finale des items a eu lieu.

b) Trouver l'information

Dans leur vie quotidienne, les lecteurs ont souvent besoin de telle ou telle information particulière. Il peut s'agir de trouver un numéro de téléphone, ou de vérifier l'heure de départ d'un autobus ou d'un train. Ils peuvent souhaiter trouver un fait particulier pour étayer ou réfuter une affirmation avancée par quelqu'un. Dans de telles situations, les lecteurs s'intéressent à la recherche d'une information isolée. Pour ce faire, ils doivent parcourir des documents, chercher, localiser et sélectionner l'information qui leur sera utile. Le processus impliqué dans cet aspect de la lecture se situe le plus souvent au niveau de la phrase, bien que, dans certains cas, l'information puisse se situer dans deux ou plusieurs phrases, ou être répartie dans différents paragraphes.

Maîtriser efficacement la recherche d'informations requiert une compréhension immédiate. Pour trouver l'information nécessaire, il se peut que le lecteur doive traiter plus d'une information. Par exemple, pour trouver quel autobus va lui permettre de quitter un endroit le plus tard possible tout en arrivant quand même à temps à destination, un lecteur aura à consulter l'horaire des autobus et comparer les heures de départ et d'arrivée de plusieurs autobus effectuant le même trajet. Dans ce cas, le lecteur devra forcément examiner plus d'une information.

Dans les épreuves d'évaluation qui demandent de trouver une information, les élève testés doivent établir entre l'information fournie dans la question et celle qui figure dans le texte une correspondance qui peut être littérale ou synonymique, et l'utiliser pour trouver la nouvelle information demandée. A ce titre, la localisation d'informations se fonde sur le texte lui-même et sur l'information explicite qu'il contient. Les tâches de localisation demandent à l'élève testé de trouver des informations en se fondant sur les conditions ou les éléments définis dans les questions ou les consignes. L'élève testé doit repérer ou identifier les éléments essentiels du message (personnages, rythme/temps, cadre, etc.), puis il doit rechercher un équivalent qui soit littéral ou synonyme.

Les tâches requérant de trouver une information peuvent également impliquer différents niveaux d'ambiguïté. On peut demander à l'élève testé de sélectionner, dans le texte ou dans un tableau, une information explicite – par exemple une indication de temps ou de lieu. Une version plus difficile du même type de tâche peut consister à trouver une information synonyme. Cela repose parfois sur une catégorisation, ou une discrimination entre deux informations semblables. En faisant varier systématiquement les paramètres qui contribuent à la difficulté, on peut mesurer les différents niveaux de compétence associés à cet aspect de la compréhension.

c) Développer une interprétation

Développer une interprétation demande aux lecteurs d'étendre leurs impressions initiales de façon à développer une compréhension plus spécifique ou plus complète de ce qu'ils ont lu. Cela implique de naviguer dans le texte, de relier les informations présentées dans ses différentes parties et aussi de se concentrer sur des détails spécifiques en tant que parties intégrantes d'un tout.

Les tâches de cette catégorie font appel à une compréhension logique : le lecteur doit traiter l'organisation de l'information dans le texte. Pour ce faire, il doit comprendre l'interaction existant entre la cohésion locale et globale dans le texte. Dans certains cas, développer une interprétation revient pour le lecteur à traiter une séquence de deux phrases seulement, unies par une relation qui relève de la cohésion locale, éventuellement facilitée par la présence de connecteurs textuels. Dans des cas plus difficiles (par exemple, dans le cas d'une relation de cause à effet), il se peut qu'aucun indicateur explicite ne soit présent.

Un texte contient plus d'informations qu'il n'en exprime explicitement. Inférer constitue une opération mentale importante, car les inférences remplissent un grand nombre de fonctions dans la compréhension d'un texte. Les inférences se fondent sur les informations et les idées activées au cours de la lecture du texte, même lorsque celles-ci ne sont pas formulées explicitement. Elles dépendent (plus ou moins) des connaissances du monde que possède le lecteur. Certaines sont considérées comme indispensables à la compréhension et sont reliées au traitement de mécanismes linguistiques (par exemple, les relations anaphoriques) ; elles jouent un rôle important dans la cohérence de l'interprétation du

texte qui se développe au cours de la lecture. D'autres inférences aboutissent à construire des informations nouvelles basées à la fois sur des données contenues dans le texte et sur les connaissances du lecteur.

Comme exemples de tâches utilisées pour évaluer cet aspect, on peut citer : comparer et confronter des informations, inférer, identifier et énumérer des éléments étayant l'information. Les tâches de comparaison et de confrontation d'informations requièrent de la part de l'élève testé qu'il intègre deux informations, ou plus, présentes dans le texte. Pour traiter des informations explicites ou implicites à partir d'une ou de plusieurs sources, le lecteur doit souvent inférer les relations ou la catégorisation que l'auteur a voulu suggérer. Les tâches qui demandent à l'élève testé de faire des inférences sur les intentions de l'auteur, et d'identifier les éléments lui permettant d'inférer ces intentions illustrent également cet aspect de la compréhension.

D'autres exemples de questions caractéristiques relevant de cet aspect sont : inférer la signification à partir du contexte, identifier les motivations ou les intentions d'un personnage précis, ou identifier une cause et son effet.

d) Réfléchir sur le contenu d'un texte

Réfléchir sur le contenu d'un texte requiert que le lecteur relie des informations qu'il a trouvées dans le texte aux connaissances qu'il détient en provenance d'autres sources ou qu'il évalue les affirmations énoncées dans le texte par rapport à sa propre connaissance du monde. Dans de nombreuses situations, les lecteurs sont appelés à justifier et défendre un point de vue. Pour ce faire, ils doivent être capables de développer une compréhension de ce qui est dit et sous-entendu dans un texte, et confronter cette représentation mentale à leurs propres connaissances et convictions, basées sur des informations antérieures ou des informations provenant d'autres textes. Ils doivent faire appel à des éléments de preuve contenus dans le texte et les confronter à d'autres sources d'informations, en utilisant des connaissances générales et spécifiques, tout comme leur aptitude à raisonner de manière abstraite.

Cet aspect de la compréhension exige un haut niveau d'aptitude métacognitive. Les lecteurs doivent réguler leur raisonnement et leurs réactions par rapport au texte, tout en mettant à l'épreuve des modèles mentaux potentiels. Pour répondre aux exigences de ce type de tâches, il faut faire appel à des informations pertinentes et les organiser d'une façon cohérente.

Parmi les tâches représentatives de cette catégorie de traitement, citons : fournir des éléments probants ou des justifications ne provenant pas du texte, évaluer la pertinence de certaines informations ou données ; établir des comparaisons avec des normes morales ou esthétiques (standards). On peut demander à l'élève testé de proposer ou d'identifier une autre information susceptible de confirmer le point de vue de l'auteur, ou bien d'évaluer si l'information ou le fait fourni dans le texte sont des preuves suffisantes.

Les informations externes à mettre en rapport avec l'information textuelle peuvent provenir des connaissances du lecteur, ou bien d'autres textes fournis dans l'évaluation, ou encore d'idées explicitement fournies dans la question.

e) Réfléchir sur la forme et la structure d'un texte

Les tâches de cette catégorie demandent au lecteur de se tenir à distance du texte et de le considérer avec objectivité, d'évaluer ses qualités et sa pertinence. Ces tâches incluent l'évaluation critique, ainsi que l'appréciation de l'impact suscité par certaines caractéristiques du texte, comme l'ironie, l'humour et l'organisation logique. Les questions demandant au lecteur de déceler les cas où le texte donne un point de vue partial, et ceux où l'auteur met en œuvre des techniques subtiles de persuasion relèvent aussi de cet aspect.

La connaissance d'éléments comme la structure, le genre et le registre du texte revêt une grande importance dans ces tâches. Ces aspects, qui sont une partie essentielle du « métier » de l'auteur, jouent un rôle considérable dans la compréhension des normes, qui est inhérente aux tâches de cette nature. Pour évaluer le talent d'un auteur à décrire certaines caractéristiques ou à persuader, ce dernier doit non seulement connaître le contenu, mais aussi pouvoir détecter les nuances de langage –

comprendre, par exemple, comment le choix d'un adjectif peut colorer l'interprétation. Un tel traitement en profondeur fait appel à des activités comme le raisonnement, l'analyse critique, la capacité d'évaluer si l'auteur a convenablement transmis son message, la capacité de distinguer un fait d'une opinion, etc. Le lecteur doit pouvoir sélectionner les unités du texte pertinentes, intégrer des unités secondaires et être à même de justifier un point de vue.

Parmi les exemples de tâches d'évaluation demandant de réfléchir sur la forme et la structure d'un texte, on peut citer : déterminer l'utilité d'un texte donné pour atteindre un objectif spécifique, évaluer l'usage fait par l'auteur de certaines techniques textuelles pour parvenir à un but précis. On peut aussi demander à l'élève testé d'identifier le style de l'auteur, de commenter l'usage qu'il en fait et d'identifier l'objectif qu'il poursuit ou l'attitude qu'il a prise.

• Aspects au niveau « micro »

Lors de la mise en œuvre des cinq aspects ci-dessus, pour lesquels les élèves testés auront à faire la preuve de leur compétence, trois variables de processus peuvent être prises en considération ; elles s'inspirent des travaux sur la lecture menés à l'occasion d'autres enquêtes internationales (IEA/RLS et EILA). Ces trois variables sont : le type d'information requise ; le type de correspondance à établir entre l'information donnée et l'information requise ; le caractère plus ou moins plausible d'éventuelles informations parasites. Les paragraphes suivants décrivent les propriétés générales de chacune de ces caractéristiques et proposent une série de considérations sur le format à donner aux réponses sollicitées, ainsi que sur le codage des réponses obtenues.

a) Type d'information requise

Il s'agit ici des divers types d'informations que le lecteur doit identifier pour répondre correctement à une question du test. Plus l'information requise est concrète, plus la tâche est jugée facile. Dans des travaux précédents basés portant sur des enquêtes à grande échelle sur la compréhension de l'écrit chez les adultes et les enfants (Kirsch, 1995 ; Kirsch et Mosenthal, 1994 ; Kirsch, Jungeblut et Mosenthal, 1998), la variable « type d'information » a été codée sur une échelle à cinq points. Un score de un point correspond à l'information la plus concrète, et donc la plus facile à traiter, alors qu'un score de 5 points correspond à l'information la plus abstraite et donc la plus difficile à traiter. Par exemple, les questions demandant d'identifier une personne, un animal ou une chose (c'est-à-dire les noms d'entités que l'on peut se représenter mentalement) sont considérées comme requérant une information éminemment concrète et on leur attribue une valeur de un point. Les questions demandant d'identifier des buts, des conditions ou des objectifs sont considérées comme requérant des informations plus abstraites. Ces tâches sont jugées plus difficiles et se voient donc attribuer une valeur de 3 points. Les questions qui demandent aux élèves testés d'identifier « un équivalent » sont considérées comme les plus abstraites et se voient attribuer une valeur de 5 points. Souvent, « l'équivalent » est une expression ou un terme peu familier pour lequel il est demandé aux élèves testés d'inférer une définition ou une interprétation à partir du texte.

b) Type de correspondance

Le « type de correspondance à effectuer » fait référence à la façon dont les élèves testés traitent le texte pour répondre correctement à une question. Il s'agit des processus utilisés pour relier l'information fournie dans la question (information donnée) à l'information qu'il y a lieu de trouver dans le texte (nouvelle information) ainsi que les processus requis pour identifier ou pour construire une réponse correcte à partir des informations disponibles.

On a identifié quatre stratégies de mise en correspondance : localiser, passer en revue, intégrer et générer. Les tâches demandant de « localiser » requièrent de l'élève testé qu'il compare une ou plusieurs caractéristiques de l'information contenue dans la question à une information identique ou synonyme située dans le texte. Les tâches « passer en revue » exigent aussi qu'on compare une ou plusieurs caractéristiques de l'information mais, contrairement aux tâches de localisation, elles demandent aux élèves testés d'établir une série de correspondances entre caractéristiques pour rencontrer les conditions énoncées dans la question. Les tâches demandant d'« intégrer » nécessitent de la part du lecteur

OCDE 1999

qu'il rassemble deux informations provenant du texte, ou davantage, en fonction d'un type donné de relation. Par exemple, il peut être demandé aux élèves testés d'établir une relation de similitude (comparaison), de différence (contraste), de degré (plus petit ou plus grand), ou de cause à effet. L'information peut être localisée dans un seul paragraphe ou peut apparaître dans différents paragraphes ou différentes sections du texte. En intégrant ces informations, les élèves testés se réfèrent à des catégories d'information fournies dans la question pour localiser l'information correspondante dans le texte. Ils relient ensuite les informations à ces diverses catégories en s'appuyant sur la relation spécifiée dans la question. Dans certains cas, les élèves testés doivent « générer » eux-mêmes ces catégories et/ou ces relations avant d'intégrer l'information fournie par le texte.

Le type de correspondance à établir entre la question et le texte n'est pas seulement déterminé par celle de ces quatre stratégies qu'il est demandé à l'élève d'appliquer. Il est, en outre, influencé par plusieurs autres conditions de traitement, qui contribuent à la difficulté générale de la tâche. La première de ces conditions est le nombre de phrases qui doivent être utilisées lors de la recherche. Par ailleurs, la difficulté de la tâche augmente avec la quantité d'informations contenues dans la question pour lesquelles l'élève doit effectuer des recherches dans le texte. Par exemple, les questions qui consistent en une seule proposition principale ont tendance à être plus faciles, en moyenne, que celles qui contiennent plusieurs propositions, principales ou subordonnées. La difficulté augmente encore avec le nombre de réponses que l'élève testé est censé donner. Les questions qui requièrent une seule réponse sont plus faciles que celles qui en requièrent trois ou plus. De plus, les questions qui indiquent le nombre de réponses à donner tendent à être plus faciles que celles qui ne le précisent pas. Par exemple, la question « *Énumérez les trois raisons…* » sera plus facile que la question « *Énumérez les raisons…* ». La difficulté des tâches est également influencée par le nombre d'inférences à effectuer pour confronter l'information donnée dans une question à l'information correspondante dans le texte et identifier l'information requise.

c) Degré de plausibilité des distracteurs

Il s'agit ici de la présence éventuelle dans le texte d'informations qui possèdent une ou plusieurs caractéristiques communes avec l'information requise dans la question, mais qui ne répondent pas précisément à ce qui est demandé. Une tâche est considérée comme très facile lorsqu'il n'y a aucune information distractrice dans le texte. Elle tend à devenir plus difficile lorsque le nombre de distracteurs augmente, lorsque les distracteurs ont un plus grand nombre de caractéristiques communes avec la réponse correcte, et lorsque les distracteurs se trouvent à proximité immédiate de la réponse correcte. Par exemple, les tâches tendent à être jugées plus difficiles lorsqu'un ou plusieurs distracteurs répondent à certaines des conditions spécifiées dans la question, mais non à toutes, et lorsqu'ils apparaissent dans un paragraphe ou une section du texte différent de celui qui contient la réponse correcte. On considère comme plus difficiles les tâches où deux distracteurs, voire davantage, présentent un très grand nombre de points communs avec la réponse correcte et qu'ils se situent dans le même paragraphe, ou nucléus d'information, que la réponse correcte.

d) Formats de réponse

Pour évaluer la compréhension de l'écrit, on a utilisé des questions à choix multiple aussi bien que des questions à réponse construite, mais la littérature sur les tests nous donne peu d'informations permettant de savoir quels sont les formats qui permettent de mesurer le plus correctement les stratégies ou les processus mis en oeuvre. Comme le remarque Bennett (1993), *« En dépit de ce qu'affirment avec force les théoriciens cognitivistes, la recherche empirique n'a fourni que des preuves peu convaincantes sur la question de savoir si les tâches à réponse construite mesurent des compétences fondamentalement différentes de celles visées par les questions à choix multiple* » (p. 8). En particulier, les travaux de Traub sur les différences entre les deux formats de réponse dans les tests de compréhension en lecture aboutissent à la conclusion qu'il n'y a pas de preuve d'un effet important induit par le format (Traub, 1993).

Il existe assez peu de travaux empiriques sur les effets du format. Lors de son enquête bibliographique, Traub n'a trouvé que deux études, l'une effectuée auprès d'étudiants de l'enseignement supérieur et l'autre auprès d'élèves de 3e année du primaire. A noter que celle menée auprès d'étudiants de

l'enseignement supérieur (Ward, Dupree et Carlson, 1987) a mesuré les aspects les plus complexes de la compréhension en lecture. Toutefois, dans le discours qu'il a prononcé lors de son investiture comme président de *l'American Psychological Association*, Frederickson (1984) a souligné que, dans les tests, les vrais biais proviennent des limites imposées par l'usage exclusif d'items à choix multiple. En outre, il se peut que les élèves de certains pays de l'OCDE ne soient pas familiarisés avec le format standard d'items à choix multiple. C'est pourquoi l'inclusion d'un ensemble d'items à réponse ouverte permettra à coup sûr d'obtenir un meilleur équilibre entre les types de tâches qui sont familières aux élèves. La recherche de cet équilibre peut également aider à élargir les *constructs* mesurés.

Il existe une grande variété de tâches à réponse construite. Certaines d'entre elles requièrent peu de jugement de la part du correcteur ; par exemple, celles qui demandent au lecteur de simplement cocher des parties du texte pour indiquer une réponse, ou citer quelques mots. D'autres tâches requièrent du correcteur une part considérable de jugement subjectif, comme celles où le lecteur doit résumer un texte avec ses propres mots.

En raison de l'absence d'éléments de preuve solides pouvant confirmer l'existence d'effets dus à la méthode, et en tenant compte des conseils des responsables du développement des tests, la solution la plus sage est sans doute d'inclure les deux types d'items, ceux à choix multiple et ceux à réponse construite, dans l'évaluation de la compréhension de l'écrit.

e) Notation

La notation est relativement simple pour des items à choix multiple corrigés de façon dichotomique : soit l'élève testé a choisi la bonne réponse, soit il ne l'a pas choisie. Des modèles de crédit partiel permettent une notation plus complexe pour les items à choix multiple. Dans ce système, certaines mauvaises réponses sont considérées comme plus « correctes » que d'autres ; les élèves testés qui choisissent ces « réponses presque bonnes » reçoivent un crédit partiel. Les modèles psychométriques pour ce genre de notation à valeurs multiples sont maintenant bien établis et, de fait, sont préférables aux barèmes dichotomiques, parce qu'ils prennent davantage en compte l'information contenue dans les réponses. L'interprétation des barèmes à valeurs multiples est cependant plus complexe, puisque chaque tâche a plusieurs localisations sur l'échelle de difficulté : une pour la réponse complète et d'autres pour chacune des réponses partiellement correctes.

La notation est relativement simple aussi pour des items dichotomiques à réponse construite, mais la spécification des bonnes réponses est plus difficile à établir. Plus on attend des élèves testés qu'ils génèrent des idées plutôt que de se borner à identifier simplement des informations du texte, plus les variations sont importantes parmi les bonnes réponses. Une formation approfondie des correcteurs et un processus de régulation important lors des corrections seront indispensables si l'on veut garantir la comparabilité d'un correcteur à l'autre, même pour un seul pays. Il faut trouver le juste équilibre entre spécificité et ouverture. Des consignes de notation trop spécifiques peuvent conduire à noter comme incorrectes certaines réponses correctes, mais formulées de manière originale. Si les consignes de notation sont trop ouvertes, des réponses qui ne correspondent pas vraiment à la tâche peuvent être notées comme correctes.

Les items à réponse construite se prêtent particulièrement bien à la notation à crédit partiel, bien que cela ajoute quelque peu à la complexité du processus (et du développement des consignes de correction). La notation à crédit partiel permet également d'utiliser de nombreux types de tâches où une catégorie donnée de réponses indique une compréhension plus fine du texte qu'un autre type de réponse, même dans les cas où toutes deux sont « correctes ». L'usage de la notation à crédit partiel est recommandé, du moins pour les items à réponse construite les plus complexes.

Structure des épreuves d'évaluation

Dans cette partie, nous décrivons la répartition des tâches proposées pour l'évaluation de la compréhension de l'écrit selon la situation, le format de texte, l'aspect et le type d'item.

OCDE 1999

Une manière évidente de distribuer les tâches de compréhension de l'écrit serait de les répartir uniformément sur les quatre situations (tableau 3). Toutefois, il convient de donner moins de poids aux situations relative à la vie professionnelle, pour deux raisons. D'une part, la sélection de textes à caractère professionnel peut rendre l'épreuve dépendante de connaissances professionnelles trop spécifiques, ce qui est à éviter. D'autre part, nous estimons qu'il est possible de construire le même type de questions et de consignes à partir d'autres situations – au contenu desquelles des élèves de 15 ans auront accès plus facilement.

Tableau 3. **Distribution recommandée des épreuves de lecture selon la situation**

Situation	% du total des épreuves
Personnel	28
Pédagogique	28
Professionnel	16
Public	28
Total	**100**

La distribution et la diversité des textes proposés aux élèves dans le cadre du programme OCDE/PISA constituent une caractéristique essentielle du test. Les tableaux 4 et 5 montrent la distribution recommandée entre textes continus et non continus. On peut aisément constater que les textes continus représenteront environ les deux tiers des textes inclus dans l'épreuve. Au sein de cette catégorie, le plus grand pourcentage sera constitué de textes informatifs (33 pour cent), tandis que les textes injonctifs représenteront le pourcentage le plus petit (7 pour cent). Les autres types de textes continus devraient être distribués de façon égale, représentant environ 20 pour cent chacun. Les textes non continus devraient représenter environ un tiers des textes de l'épreuve de compréhension de l'écrit. La grande majorité de ceux-ci (66 pour cent) sera constituée de tableaux, de diagrammes et de graphiques. Les autres textes non continus seront des cartes géographiques, des messages publicitaires ou des formulaires que des élèves âgés de 15 ans sont censés pouvoir lire et utiliser. Il est important de garder à l'esprit que ces pourcentages sont des objectifs fixés pour la campagne de test définitive et non pour l'essai de terrain. La sélection des textes entrant dans l'essai de terrain et ensuite dans la campagne de test définitive ne se fondera pas exclusivement sur des caractéristiques structurelles telles que le format et le type de texte. La diversité culturelle, la gamme des degrés de difficulté d'un texte à l'autre, leur intérêt potentiel pour les élèves et leur authenticité seront également pris en considération.

Tableau 4. **Distribution recommandée des épreuves de lecture selon le type de texte continu**

Type de texte	% des textes continus	% du test
Narratif	20	13
Informatif	33	22
Descriptif	20	13
Argumentatif/Persuasif	20	13
Injonctif	7	5
Total	**100**	**66**

le tableau 6 montre la distribution recommandée des tâches de compréhension de l'écrit en fonction de chacun des cinq aspects décrits plus haut. Le pourcentage le plus élevé concerne le développement d'une interprétation ; un peu plus des deux tiers des tâches couvrent les trois premiers aspects (70 pour cent). Ces trois aspects – comprendre globalement un texte, trouver l'information et développer une interprétation – mettent l'accent sur la capacité du lecteur à comprendre et utiliser l'information

Tableau 5. **Distribution recommandée des épreuves de lecture selon le type de texte non continu**

Type de texte	% des textes non continus	% du test
Diagrammes/Graphiques	33	11
Tableaux	33	11
Schémas	10	3
Cartes	10	3
Formulaires	8	3
Publicité	6	2
Total	**100**	**33**

Tableau 6. **Distribution recommandée des épreuves de lecture selon les aspects**

Aspect	% du test
Trouver l'information	20
Comprendre globalement un texte	20
Développer une interprétation	30
Réfléchir sur le contenu	15
Réfléchir sur la forme	15
Total	**100**

contenue dans le texte. Dans les 30 pour cent de tâches restants, on demandera à l'élève de réfléchir, soit sur le contenu de l'information fournie dans le texte, soit sur la structure et la forme du texte lui-même. Le tableau 7 présente la distribution des tâches en fonction du format et de l'aspect du texte.

Tableau 7. **Distribution recommandée des épreuves de lecture selon les aspects**

Aspect	% du test
Trouver l'information	20
Comprendre globalement un texte	20
Développer une interprétation	30
Réfléchir sur le contenu	15
Réfléchir sur la forme	15
Total	**100**

Pour déterminer avec précision la place à donner aux items à réponse construite, il faut quelques principes de distribution des tâches établis sur une base à la fois pratique et conceptuelle. Le tableau 8 présente la distribution des tâches proposée pour les items à réponse construite et à choix multiple, en fonction des cinq aspects de la lecture.

Tableau 8. **Distribution recommandée 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**

OCDE 1999

Ce tableau montre qu'environ 45 pour cent de l'épreuve de compréhension de l'écrit seront constitués d'items à réponse construite qui requièrent un jugement de la part du correcteur. Les 55 pour cent restants consisteront en épreuves à choix multiple et à réponse construite nécessitant peu de jugement subjectif de la part du correcteur. On voit également dans le tableau 8 que, bien que les items à choix multiple et à réponse construite soient présents dans les cinq aspects, ils ne seront pas distribués uniformément. Un pourcentage plus important d'items à choix multiple sera associé aux trois premiers aspects de la lecture.

Échelles des résultats

Pour arriver aux buts que s'est fixé le programme OCDE/PISA, il est essentiel d'élaborer des échelles décrivant les résultats des élèves. Le développement de ces échelles fera l'objet d'un processus itératif : on partira de propositions initiales basées sur l'expérience antérieure et les recherches dans le domaine de l'apprentissage et du développement cognitif qui seront ensuite révisées grâce aux données empiriques recueillies durant l'essai de terrain de l'enquête PISA.

Les deux variables d'organisation du plan d'évaluation de la lecture qui sont actuellement considérées comme pouvant donner lieu à des échelles de résultats sont le type de texte (continu ou non continu) et les aspects au niveau « macro » (comprendre globalement un texte ; trouver l'information ; développer une interprétation ; réfléchir sur le contenu d'un texte ; réfléchir sur les qualités formelles d'un texte). Cela signifie qu'il y aura soit deux, soit cinq échelles de résultats en lecture, en plus de l'échelle globale de compréhension de l'écrit.

Autres questions

Plusieurs autres questions relatives à la manière dont la compréhension de l'écrit sera définie et mesurée dans le cadre de ce test doivent être abordées. Trois d'entre elles ont trait au questionnaire contextuel auquel les élèves doivent répondre, deux autres concernent les rapports avec d'autres tests. Ces questions sont abordées ci-dessous.

Problématique du questionnaire

Plusieurs des volets de l'étude envisagée se prêtent bien mieux à une collecte de données par questionnaire que par une épreuve par test. L'un de ces volets est le recueil d'informations sur les pratiques et les centres d'intérêt en lecture, le second porte sur certains aspects métacognitifs, le troisième a trait au rôle de la technologie dans la vie des élèves participant à l'étude OCDE/PISA.

Pratiques de lecture et intérêt pour la lecture

A l'instar d'autres enquêtes à grande échelle sur la lecture (Conseil des ministres de l'Éducation du Canada, 1994 ; Elley, 1992 ; Jones, 1995 ; Smith, 1996 ; Taube et Mejding, 1997), le questionnaire contextuel OCDE/PISA comprendra un ensemble de questions destinées à évaluer les pratiques de lecture des élèves et à connaître le contexte de leurs activités de lecture en général, que ce soit à l'école ou en dehors de l'école. L'intérêt de ces questions sera à la fois descriptif et explicatif. D'une part, elles serviront à décrire la population des élèves de 15 ans en termes d'accès à l'écrit, d'intérêt et d'attitudes par rapport à diverses activités de lecture, ainsi que des activités de lecture effectivement mises en œuvre. Les éléments d'information ainsi recueillis devraient, d'autre part, permettre d'expliquer en partie les variations de compétences en compréhension de l'écrit observées parmi les élèves de 15 ans des différents pays participant au programme OCDE/PISA.

Deux considérations principales ont guidé le choix des aspects à inclure dans l'enquête :

– les aspects évalués doivent être pertinents en termes de politique éducative ;

– l'évaluation de la compréhension de l'écrit doit avoir une contrepartie au niveau des questionnaires, en cherchant à recueillir des informations sur l'attitude des élèves quant à leur pratique de la lecture et leur intérêt pour la lecture.

Cette section du questionnaire contextuel inclura des items portant sur les aspects suivants :

– *Le degré d'exposition à différents types d'écrit à la maison, à l'école ou dans un environnement public :* cette rubrique comprendra les questions suivantes : nombre de livres à la maison ; l'élève possède-t-il ses propres livres ; achat régulier d'un quotidien ou d'un hebdomadaire par la famille ; fréquentation d'une bibliothèque publique ou scolaire, etc.

– *Pratiques et habitudes de lecture :* il semble important de faire en sorte que les items couvrent les divers types d'imprimés et la fréquence de leur usage, pour les différents types de textes et de formats de textes distingués par le plan d'évaluation, et en relation avec le contexte où ils sont évalués. Vu les limitations de temps et les problèmes méthodologiques posés par les inventaires d'« activités de lecture », il faut parvenir à un équilibre raisonnable entre, d'une part, la nécessité d'inventorier les différents types d'imprimés dans différents types de situations (afin d'évaluer la diversité) et, d'autre part, les diverses contraintes.

– *Les attitudes et les intérêts par rapport à la lecture :* les attitudes à l'égard de la lecture et les motivations sont susceptibles d'influencer les activités de lecture et le rendement. Il s'agit, par ailleurs, d'aspects sur lesquels il est possible d'agir, en créant un climat favorable à la compréhension de l'écrit, à l'école et en dehors de l'école. Dans le cadre du projet OCDE/PISA, cet aspect est évalué au moyen d'un certain nombre de questions ciblées, ne nécessitant qu'un temps de réponse très court (par exemple : Aimez-vous recevoir un livre comme cadeau ? Aimez-vous aller à la bibliothèque ? etc.). On demande par ailleurs à l'élève d'indiquer ses préférences en matière d'activités de loisirs, en comparant la lecture à d'autres activités (télévision, musique, sorties, jeux vidéo, etc.). Les réponses à ce genre de questions pourraient s'avérer moins sujettes aux effets de désirabilité que l'on observe souvent dans les évaluations des pratiques de lecture.

Métacognition

Plusieurs études ont montré une corrélation entre les connaissances métacognitives et les performances en lecture chez les jeunes lecteurs (Ehrlich, Kurtz-Costes et Loridant, 1993 ; Ehrlich, 1996). Elles ont mis en évidence chez les lecteurs des variations dans la compréhension du concept de lecture, dans les buts et objectifs de la lecture, dans les stratégies destinées à la compréhension et dans la détection des incohérences. Ces variations concernent les deux composantes de base de la métacognition : les connaissances métacognitives des élèves et la régulation de la métacognition. La première composante a trait, chez le lecteur, à la capacité de réfléchir à ses propres processus cognitifs, et comprend le fait de savoir quand, comment et pourquoi accomplir différentes activités cognitives. La seconde composante, la régulation, concerne l'utilisation de stratégies qui permettent au lecteur de contrôler ses efforts cognitifs (Baker, 1991).

L'intérêt de la mesure de la métacognition dans le cadre du projet OCDE/PISA tient au fait que les résultats sont jugés pertinents pour les décideurs politiques et susceptibles d'influencer les pratiques de lecture et d'apprentissage, dans la mesure où l'on considère qu'il est possible d'enseigner ces savoir-faire, et que ceux-ci sont transférables à d'autres domaines que la lecture. La difficulté à surmonter est de développer un outil de mesure de la métacognition utilisable dans des enquêtes à large échelle comme OCDE/PISA. La plupart des travaux disponibles se fondent sur des expériences menées auprès d'élèves plus jeunes que ceux qui vont participer aux tests OCDE/PISA (Myers et Paris, 1978). En conséquence, il est apparu qu'il n'existe aucun instrument susceptible de fournir des résultats fiables et valides pour le groupe d'âge visé.

Étant donné qu'il n'existe aucun instrument fiable pour mesurer la métacognition chez les élèves âgés de 15 ans, et que les ressources adéquates permettant de concevoir et de développer un tel instrument ne sont pas disponibles, il a été décidé que la métacognition ne ferait pas partie de l'essai de

OCDE 1999

terrain ni de la campagne de test définitive sur la compréhension de l'écrit lors de ce premier cycle. Cependant, compte tenu de l'intérêt que ce thème suscite, l'élaboration d'un instrument de mesure de la métacognition sera prise en considération pour les cycles à venir.

Technologie

Dans notre monde en constante évolution, une enquête portant sur les habitudes de lecture doit s'ouvrir à des questions en rapport avec la technologie de l'information, particulièrement l'ordinateur. Il est évident que l'accès aux textes sur support électronique et l'utilisation de ceux-ci pour obtenir et échanger de l'information gagneront en importance dans la vie des élèves au cours des années à venir.

Pour jeter les bases d'une présence accrue de la technologie dans les cycles à venir, l'enquête OCDE/PISA inclura un bref questionnaire destiné à une collecte d'information concernant l'accès des élèves à l'ordinateur à la maison, à l'école, sur le lieu de travail ou dans leur communauté ; leurs attitudes par rapport à l'utilisation d'un ordinateur ; la fréquence d'utilisation de l'ordinateur dans différents contextes ; le type d'activités auquel l'élève s'adonne sur un ordinateur.

Liens avec d'autres tests d'évaluation

Il est important d'utiliser toutes les sources d'informations disponibles pour interpréter pleinement les données provenant de cette enquête. Une partie des informations complémentaires proviendront des données des questionnaires, mais les liens établis avec d'autres enquêtes peuvent également fournir des données qui enrichissent l'interprétation du test.

OECD/PISA n'est pas la seule évaluation des compétences en lecture. Comme on l'a mentionné plus haut, l'*Enquête internationale sur la littératie des adultes* (EILA) a évalué les compétences des adultes et il serait utile que les résultats d'OCDE/PISA puissent également être analysés à la lumière de cette étude, malgré les différences significatives dans la nature et la conception des instruments PISA et EILA. Les liens avec PISA permettraient de comparer également les résultats des élèves à ceux des adultes. Cela est partiellement réalisable, en incorporant des épreuves provenant de l'EILA en vue de créer un lien direct. Une réflexion a été menée sur le nombre d'items de l'enquête EILA qui devraient être intégrés à l'enquête OCDE/PISA pour que l'ancrage ainsi créé soit valide tant du point de vue conceptuel que statistique. La discussion a également porté sur l'équilibre à trouver entre la nécessité d'inclure un nombre suffisant d'items d'ancrage EILA et le désir de préserver la nature d'une évaluation dont la palette se veut plus large que celle utilisée dans les enquêtes antérieures.

Il sera possible d'établir un ancrage avec une seule des trois échelles utilisées dans l'*Enquête internationale sur la littératie des adultes*, car trop peu de temps est disponible dans l'évaluation pour inclure dans les tests une longue série d'items de l'enquête EILA. Puisqu'il est probable que l'évaluation OCDE/PISA comprendra moins d'épreuves basées sur des textes non continus que sur des textes continus, il ne serait pas raisonnable d'utiliser l'échelle informative de l'enquête sur la littératie des adultes : le nombre important d'items EILA nécessaire pour réaliser cet ancrage aurait pour effet de les rendre dominants dans cette section du test OCDE/PISA. La même objection vaut pour l'échelle quantitative EILA, car celle-ci se base essentiellement sur des textes non continus. Ce sont donc les épreuves de l'échelle de textes en prose d'EILA qui paraissent les mieux appropriées pour PISA, car les items à inclure ne seraient pas trop nombreux par rapport aux autres textes continus de PISA.

Afin de déterminer si les paramètres des items s'ajustent aux populations ciblées par OCDE/PISA, deux blocs d'items en provenance de l'échelle de prose EILA seront inclus dans l'essai de terrain OCDE/PISA. En supposant que la plupart des paramètres conviendront, un nombre approprié de tâches provenant de l'échelle de prose EILA sera inclus dans la campagne de test définitive de PISA. Ces items auront un double objectif : tout d'abord, les paramètres des items EILA serviront à estimer la compétence en lecture de textes de prose dans les différents pays où sera menée la campagne PISA. Deuxièmement, ces items seront codés comme faisant partie du plan d'évaluation de la compréhension de l'écrit de PISA et serviront à estimer les compétences en compréhension de l'écrit des élèves de 15 ans des pays participants.

L'enquête OCDE/PISA évalue également deux autres domaines de rendement – les mathématiques et les sciences –, quoique de manière moins approfondie. Il y a là une occasion de tenter d'estimer les liens entre ces différents domaines. L'approche la plus simple consiste à utiliser uniquement le fait que des populations semblables sont testées en même temps, dans les mêmes pays, sans tenter de modéliser ces relations. Une approche plus intéressante, mais aussi plus complexe, est de prévoir un certain nombre d'items qui se situent à l'intersection des différents domaines (variables intégrées), en sorte que la relation entre les rendements en lecture et en sciences ou en mathématiques pourrait être modélisée directement à partir des résultats de ces items communs. Les modèles psychométriques pour ce type d'évaluations croisées existent, mais les items nécessitent une mise au point très rigoureuse, et l'interprétation de tels résultats conditionnels est moins directe qu'avec des modèles traditionnels, dans lesquels chaque item se voit attribuer une position unique sur une seule échelle. Étant donné l'importance des informations susceptibles d'être obtenues grâce aux items croisés, le premier cycle de la campagne OCDE/PISA comprendra un bloc d'items intégrés (qui recevront un codage pour plus d'un domaine d'évaluation), ainsi que plusieurs blocs incluant des items distincts de lecture, de mathématiques et de sciences, mais basés sur des textes ou des situations communs.

OCDE 1999

Chapitre 2

LA CULTURE MATHÉMATIQUE

Définition du domaine

Le domaine de la « culture mathématique » a trait à la capacité des élèves de tirer parti de leurs compétences mathématiques pour affronter les défis de l'avenir. Il renvoie à la capacité des élèves d'analyser, raisonner et communiquer des idées de façon efficace en sachant poser, formuler et résoudre des problèmes mathématiques dans une grande diversité de domaines et situations.

Dans le cadre de l'étude OCDE/PISA, la définition de la culture mathématique est la suivante :

« La culture mathématique est l'aptitude d'un individu à identifier et à comprendre les divers rôles joués par les mathématiques dans le monde, à porter des jugements fondés à leur propos, et à s'y engager, en fonction des exigences de sa vie présente et future en tant que citoyen constructif, impliqué et réfléchi. »

Quelques explications s'imposent pour clarifier la définition de ce domaine.

Culture mathématique…

Le terme de *culture* a été choisi pour souligner que les connaissances et compétences mathématiques, telles que définies dans le cadre d'un curriculum traditionnel de mathématiques, ne sont pas, à nos yeux, la cible principale du projet OCDE/PISA. Nous mettons plutôt l'accent sur des connaissances mathématiques utilisées de manière fonctionnelle dans toute une série de contextes différents et sur les façons de susciter la réflexion et la compréhension. Pour que de telles démarches soient réellement possibles et efficaces, quantité de connaissances fondamentales et de savoir-faire mathématiques (tels qu'ils sont souvent enseignés dans les écoles) sont indispensables. Tout comme la maîtrise de la lecture présuppose – mais ne peut se réduire à – un riche vocabulaire et une connaissance approfondie des règles grammaticales, de la phonétique et de l'orthographe, la « culture mathématique » ne peut se réduire à une connaissance de la terminologie mathématique, de propriétés et de procédures, ni à une série de savoir-faire permettant d'effectuer certaines opérations ou d'appliquer certaines méthodes. Cependant, elle en présuppose bien sûr l'existence.

… traiter la réalité…

Le terme *réalité* signifie l'environnement physique, social et culturel dans lequel vit l'individu. Comme Freudenthal (1983) l'affirmait, « *les concepts, les structures et les idées mathématiques ont été créés en tant qu'outils d'organisation des phénomènes issus de réalités de nature physique, sociale ou mentale* ».

… s'engager…

Le terme s'*engager* n'a pas été choisi pour qualifier des actes physiques ou sociaux dans le sens strict du terme. Ce terme englobe aussi le fait de *communiquer*, de *prendre position par rapport à*, d'*établir des liens*, d'*évaluer* ou même d'*apprécier* les mathématiques. La définition ne devrait donc pas être perçue comme étant limitée au seul usage fonctionnel des mathématiques. En fait, les dimensions esthétique et récréative des mathématiques sont intégrées dans la définition de la culture mathématique.

OCDE 1999

… vie présente et future

Les termes *vie présente et future* de l'individu comprennent sa vie privée, professionnelle et sociale avec son entourage et ses proches, ainsi que sa vie en tant que citoyen et membre d'une collectivité.

Au cœur même de la notion de culture mathématique, se trouve la capacité de poser, de formuler et de résoudre des problèmes mathématiques dans des domaines et des contextes très divers, depuis des situations purement mathématiques jusqu'à celles qui ne présentent, au départ, aucune structure mathématique – celle-ci devant être introduite par la personne qui pose ou résout le problème.

Il faut également insister sur le fait que cette définition n'inclut pas uniquement un niveau de connaissances mathématiques minimal, mais qu'elle s'applique aussi à l'utilisation des mathématiques dans toute une série de situations.

Des attitudes et des affects tels que la confiance en soi, la curiosité, l'intérêt, la sagacité, l'envie d'agir et de comprendre le monde (pour n'en citer que quelques-uns) ne sont pas des composantes proprement dites de la définition de la culture mathématique. Néanmoins, ce sont des prérequis importants. Il est possible, en principe, d'avoir une culture mathématique sans pour autant posséder les attitudes et affects mentionnés ci-dessus. En pratique, il est cependant peu probable qu'une personne fasse véritablement usage de sa culture mathématique, telle qu'elle est définie ici, si elle n'a pas un certain degré de confiance en elle, un intérêt et un désir d'accomplir ou de comprendre des choses possédant une dimension mathématique.

Organisation du domaine

Dans le cadre du projet OCDE/PISA, il est utile d'identifier un certain nombre d'aspects de la culture mathématique.

Le projet OCDE/PISA a utilisé deux aspects majeurs et deux aspects mineurs pour organiser le domaine. Les aspects majeurs sont :

– les compétences mathématiques ; et

– les idées mathématiques majeures.

Les aspects mineurs sont :

– les domaines mathématiques enseignés ; et

– les situations et les contextes.

Les aspects majeurs servent à déterminer le champ d'application de l'évaluation et à décrire les compétences. La fonction des aspects mineurs est d'assurer une bonne couverture du domaine et d'équilibrer l'éventail d'épreuves d'évaluation sélectionnées.

Il est important de noter que ces quatre aspects ne doivent pas être combinés de sorte à ne former qu'un seul schème de classification. Deux des aspects, à savoir les « idées mathématiques majeures » et les « domaines mathématiques enseignés », constituent des alternatives de description des contenus mathématiques.

Les *compétences mathématiques* sont les compétences et savoir-faire généraux tels que la résolution de problèmes, l'utilisation du langage mathématique et les savoir-faire ayant trait à la modélisation mathématique.

Les *idées mathématiques majeures* représentent des assemblages de concepts mathématiques interdépendants qui apparaissent de façon pertinente en contexte ou dans des situations de la vie réelle. Certaines de ces idées majeures sont désormais classiques, par exemple le *hasard*, les *variations* et la *croissance*, les notions de *dépendance* et de *relation*, et la notion de *forme*. La notion d'« idée majeure » a été retenue parce qu'elle ne débouche pas sur une façon artificielle de compartimenter les mathématiques par sujet.

L'aspect *domaines mathématiques enseignés* renvoie aux contenus mathématiques scolaires tels qu'ils figurent dans de nombreux curricula. Dans le cadre de l'étude OCDE/PISA, nous proposons la liste suivante : nombres, mesures, estimations, algèbre, fonctions, géométrie, probabilités, statistiques et

mathématiques discrètes. Les *domaines mathématiques enseignés* ont été retenus comme un aspect mineur de cette étude afin que les contenus traditionnels soient couverts par cette évaluation. Toutefois, la sélection effective des contenus à inclure dans PISA sera préférablement fondée sur les *idées mathématiques majeures*, qui est un concept englobant et plus important.

Le deuxième aspect mineur est celui des *situations*, c'est-à-dire le contexte dans lequel sont présentés les problèmes mathématiques, comme par exemple un contexte éducatif, un contexte professionnel, public ou privé.

Les quatre aspects organisationnels sont analysés en détail ci-après.

Les compétences mathématiques

Le premier aspect majeur du plan d'évaluation OCDE/PISA de la culture mathématique est celui des compétences mathématiques. Cet aspect consiste en une liste non hiérarchisée de compétences générales mathématiques censées être pertinentes et applicables à tous les niveaux d'enseignement. Cette liste comprend les éléments suivants :

1. *Capacité de pensée mathématique*. Savoir poser des questions caractéristiques des mathématiques (« Existe-t-il… ? », « Si oui, combien ? », « Comment trouve-t-on… ? ») ; connaître les types de réponses que les mathématiques réservent à de telles questions ; savoir distinguer différents types d'énoncés (définitions, théorèmes, conjectures, hypothèses, exemples, assertions conditionnelles) ; comprendre et manier la portée et les limites de concepts mathématiques donnés.
2. *Capacité d'argumentation mathématique.* Appréhender ce qu'est une démonstration mathématique et en quoi elle diffère d'autres formes de raisonnements mathématiques ; comprendre et évaluer différents types d'enchaînements d'arguments mathématiques ; posséder un certain sens de l'heuristique (« que [ne] peut-il [pas] se produire, et pourquoi ? ») ; savoir développer une argumentation mathématique.
3. *Capacité de modélisation mathématique*. Savoir structurer le domaine ou la situation qui doit être modélisé ; « mathématiser » (c'est-à-dire opérer une traduction de la « réalité » vers la structure mathématique) et « démathématiser » (c'est-à-dire interpréter des modèles mathématiques en termes de « réalité ») ; travailler avec un modèle mathématique ; savoir valider le modèle ; réfléchir, analyser et se montrer critique à l'égard des modèles et de leurs résultats ; savoir communiquer à propos du modèle et de ses résultats (y compris au sujet des limites de ces derniers) ; savoir maîtriser le suivi et le contrôle du processus de modélisation.
4. *Capacité de poser et résoudre des problèmes.* Savoir poser, formuler et définir différents types de problèmes mathématiques (« purs », « appliqués », « à réponses ouvertes », « à réponses fermées ») ; savoir résoudre différentes sortes de problèmes.
5. *Capacité de représentation.* Savoir décoder, interpréter et distinguer différentes formes de représentation d'objets et de situations mathématiques ainsi que les relations entre les diverses représentations ; savoir choisir entre différentes formes de représentations et passer de l'une à l'autre en fonction de la situation et du but recherché.
6. *Capacité symbolique, formelle et technique.* Savoir décoder et interpréter le langage symbolique et formel, et saisir les relations qu'il entretient avec la langue naturelle ; opérer la conversion de la langue naturelle vers le langage symbolique et formel ; manier des énoncés et des expressions contenant des symboles et des formules ; utiliser des variables, résoudre des équations et effectuer des calculs.
7. *Capacité de communiquer.* Savoir s'exprimer de diverses façons sur des sujets à contenu mathématique, tant oralement que par écrit ; savoir comprendre une information orale et écrite sur les mêmes sujets, formulée par autrui.
8. *Capacité de manier les outils et les instruments.* Connaître et pouvoir utiliser différents outils et instruments (y compris les nouvelles technologies d'information) pouvant contribuer à l'activité mathématique ; être informé des limites effectives de tels outils et instruments.

OCDE 1999

Les classes de compétences

Le but du projet OCDE/PISA n'est pas d'élaborer des items évaluant séparément les capacités ci-dessus. En effet, lorsqu'on fait de *vraies* mathématiques, il est habituellement nécessaire de recourir en même temps à de nombreux savoir-faire (si pas à tous), si bien que tout effort visant à les évaluer individuellement aboutit immanquablement à des épreuves artificielles et une compartimentation du domaine de la culture mathématique.

Afin de rendre opérationel l'aspect « *compétences mathématiques* » à travers l'élaboration d'items et de tests, il est utile d'organiser les savoir-faire ou capacités en trois grandes classes de compétences non hiérarchisées. Les trois classes de compétences sont :

– Classe 1 : reproduction, définitions, calculs ;

– Classe 2 : mise en relation et intégration pour résoudre des problèmes ;

– Classe 3 : mathématisation, généralisation et compréhension en profondeur.

Chacun des savoir-faire ci-dessus sera vraisemblablement amené à jouer un rôle dans chacune des classes de compétences ; c'est-à-dire que les savoir-faire ne peuvent pas être rattachés à une seule classe de compétences. Les classes forment un continuum conceptuel qui va de la simple reproduction de faits et des capacités en calcul à la faculté de mettre en relation divers domaines en vue de résoudre des problèmes simples de la réalité, jusqu'à la troisième classe, qui comprend la mathématisation (ce terme est expliqué en détail plus loin) de problèmes réels en vue de leur résolution, ainsi qu'une réflexion sur les solutions du problème en situation, en recourant à la pensée mathématique, au raisonnement et à la généralisation.

Ce qui précède laisse entendre que les classes forment une hiérarchie, au sens où un ensemble d'épreuves faisant appel à des compétences de classe 3 sera généralement plus difficile qu'un ensemble d'épreuves faisant appel à des compétences de classe 2. Toutefois, cela n'implique pas que les compétences de classe 2 soient un prérequis pour les compétences de classe 3. En effet, certaines études (de Lange, 1987 ; Shafer et Romberg, à paraître) ont montré qu'il n'est pas indispensable d'exceller dans les compétences de classe 1 pour obtenir de bons résultats dans les classes 2 ou 3 et qu'a contrario les élèves réalisant de bonnes performances dans la classe 3 n'excellent pas nécessairement dans les savoir-faire de classe 1.

La définition de la culture mathématique dans le projet OCDE/PISA accorde une grande importance aux élèves faisant preuve d'aptitudes à réaliser des tâches requérant des aptitudes dans les trois classes de compétences. Les épreuves d'évaluation proviendront des trois classes afin que les décideurs politiques aient loisir de repérer la place donnée par les écoles et les curricula de leur pays aux savoir-faire requis dans chacune des classes de compétences.

Les compétences de classe 1 : reproduction, définitions, calculs

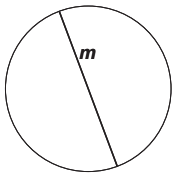
Dans cette classe, nous avons principalement affaire à des aspects que l'on rencontre fréquemment dans les tests d'évaluation standardisés, ainsi que dans les études comparatives internationales. Cette classe inclut les connaissances factuelles, les représentations, l'identification d'équivalences, la mémorisation d'objets et de propriétés mathématiques, l'exécution de procédures routinières, l'application d'algorithmes simples et la mise en œuvre de savoir-faire techniques. Le maniement d'énoncés utilisant des expressions symboliques et des formules « standard » et l'exécution de calculs font également partie de cette classe. Les types d'items utilisés pour tester les compétences de cette classe sont habituellement des questions à choix multiple ou des questions ouvertes à réponse courte.

La classe 1 renvoie en particulier aux capacités symboliques, formelles et techniques décrites plus haut. Quelques exemples de cette classe sont présentés dans la figure 2.

Figure 2. **Exemples tirés des compétences de classe 1**

Résoudre l'équation 7x – 3 = 13x + 15
Quelle est la moyenne de : 7, 12, 8, 14, 15, 9 ?
Écrire 69 % en écriture fractionnaire.

Le segment ***m*** est appelé ..... du cercle.



Les compétences de classe 2 : mise en relation et intégration pour résoudre des problèmes

Dans cette classe, prime l'établissement de liens entre les différents éléments et domaines des mathématiques, et l'intégration d'informations diverses dans le but de résoudre des problèmes simples. Pour cela, les élèves doivent choisir entre plusieurs stratégies et plusieurs outils mathématiques. Quoique ces problèmes soient présumés non routiniers pour l'élève, ils n'exigent qu'une mathématisation relativement élémentaire.

Dans la classe 2, on attend aussi des élèves qu'ils soient capables de manier diverses méthodes de représentation en fonction de la situation et de l'objectif visé. La composante « mise en relation » demande encore que les élèves soient à même de distinguer et de relier différents énoncés, tels des définitions, affirmations, exemples, assertions conditionnelles et démonstrations.

Cette classe renvoie à plusieurs des capacités mathématiques mentionnées plus haut. Il est clair que la résolution des problèmes donnés en exemple nécessite de pouvoir raisonner et argumenter. Elle nécessite ainsi la mise en œuvre de *capacités d'argumentation mathématique.* Les élèves ont de plus à « modéliser » le problème pour le résoudre. Les *capacités de modélisation* sont ainsi requises. La résolution même du problème fait appel à la capacité de poser et résoudre des problèmes. Dans le processus de résolution de problèmes, les élèves ont à utiliser différents types de représentation – un tableau, un graphique, un dessin – pour lesquels les *capacités de représentation* sont sollicitées.

En ce qui concerne le langage mathématique, le fait de pouvoir décoder et interpréter le langage symbolique et formel, ainsi que saisir ses relations avec la langue naturelle est aussi une compétence cruciale de cette classe. Les items de la classe 2 sont souvent contextualisés et demandent une prise de décision mathématique de la part de l'élève.

Deux problèmes de cette classe sont donnés à titre d'exemple dans la figure 3. Contrairement aux exemples de la classe 1, le domaine mathématique dont relèvent les questions n'est pas immédiatement clair, pas plus que la méthode, la stratégie ou l'algorithme le mieux à même de résoudre le problème. En fait, dans certains cas, le domaine mathématique dépend de la (ou des) stratégie(s) choisie(s) par l'élève.

Figure 3. **Exemples tirés des compétences de classe 2**

Vous avez parcouru en voiture deux tiers de la distance totale (en partant avec un réservoir plein) et le réservoir est encore plein au quart. Votre voiture a-t-elle un problème ?

Marie habite à deux kilomètres de l'école et Martin à 5 kilomètres. A combien Marie et Martin vivent-ils l'un de l'autre ?

OCDE 1999

Les compétences de classe 3 : mathématisation, généralisation et compréhension en profondeur

Dans les items de cette classe, on demande aux élèves de « mathématiser » des situations. Ils doivent pouvoir identifier et extraire les mathématiques partie prenante d'une situation donnée et se servir des mathématiques pour résoudre le problème, analyser, interpréter, élaborer leurs propres modèles et stratégies, ainsi que pour développer une argumentation mathématique, y compris des démonstrations et des généralisations.

Ces compétences comprennent une analyse du modèle et une réflexion sur le processus mathématique. Pour cette classe de compétences, les élèves devraient non seulement être à même de résoudre des problèmes, mais aussi de les poser.

Toutes ces compétences ne peuvent réellement bien fonctionner que si les élèves sont capables de communiquer correctement, et de différentes manières : oralement, par écrit, visuellement, etc. La communication s'entend ici comme un processus bi-directionnel : les élèves doivent aussi bien pouvoir communiquer leurs points de vue mathématiques que comprendre ceux d'autrui.

Il est enfin important de souligner que les élèves ont également besoin de *comprendre en profondeur* la nature scientifique des mathématiques, sans en oublier les dimensions culturelles et historiques, et qu'ils doivent percevoir l'intérêt de l'usage des mathématiques dans d'autres contextes et d'autres domaines du programme qui sont sujet à des modélisations mathématiques.

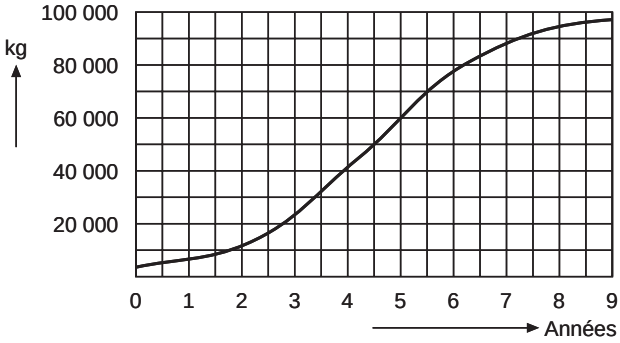
Les compétences de classe 3 incluent souvent des savoir-faire et compétences issus des deux autres.

Cette classe est une composante centrale de la culture mathématique. Elle est malheureusement la classe la plus difficile à évaluer, particulièrement dans les enquêtes à grande échelle comme PISA. Les items à choix multiple, par exemple, ne sont souvent pas adaptés pour évaluer ces compétences. Des questions à réponses multiples et longues seraient de format plus adéquat, mais la conception de tels items et l'évaluation des réponses des élèves posent de nombreuses difficultés. Cependant, comme cette classe constitue une partie fondamentale de la culture mathématique telle qu'elle est définie dans le projet OCDE/PISA, il nous a paru important de l'inclure dans l'évaluation, même s'il ne lui est accordé qu'une place restreinte.

A titre d'exemple, un problème est proposé dans la figure 4.

Figure 4. **Exemples tirés des compétences de classe 3**

Quelques poissons ont été introduits dans un plan d'eau. Le graphique représente le modèle de croissance du poids de poisson dans le plan d'eau.



On suppose qu'un pêcheur décide d'attendre quelques années pour pêcher du poisson dans ce plan d'eau. Combien d'années le pêcheur devrait-il attendre s'il désire maximiser le nombre annuel de poissons qu'il pêchera par la suite ? Argumentez votre réponse.

La « mathématisation »

La mathématisation – telle qu'elle est utilisée dans le cadre de PISA – est le processus qui permet d'organiser la réalité à l'aide d'idées et de concepts mathématiques. C'est une activité organisatrice dans laquelle sont mises en œuvre des connaissances acquises et des capacités pour découvrir des régularités inconnues, des relations et des structures (Treffers et Goffree, 1985). Ce processus est parfois qualifié de mathématisation horizontale (Treffers, 1986). Il fait appel à des activités telles que :

– identifier, dans un contexte général, les éléments spécifiquement mathématiques ;

– schématiser ;

– formuler et visualiser le problème ;

– découvrir des relations et des régularités ;

– reconnaître des similitudes dans des problèmes différents (de Lange, 1987).

A partir du moment où le problème a été transformé en un problème mathématique, il peut être résolu au moyen d'outils mathématiques. En d'autres termes, on peut utiliser des outils mathématiques pour manipuler et affiner le modèle mathématique du problème réel. On parle alors d'un processus de *mathématisation verticale*, qui se distingue par les activités suivantes :

– représenter une relation par une formule ;

– démontrer des régularités ;

– affiner et ajuster des modèles ;

– combiner et intégrer des modèles ; ainsi que

– généraliser.

Le processus de mathématisation se joue ainsi en deux phases différentes : la mathématisation horizontale, processus qui traduit le monde réel en monde mathématique ; la mathématisation verticale, processus de traitement du problème au sein des mathématiques à l'aide d'outils mathématiques, en vue de résoudre le problème. Réfléchir sur la solution trouvée au regard du problème original est une étape essentielle dans le processus de mathématisation qui bien souvent ne bénéficie pas de l'attention nécessaire.

On pourrait évidemment soutenir que le processus de mathématisation intervient dans toutes les classes de compétences puisque, dans tout problème contextualisé, on doit forcément identifier les mathématiques pertinentes. Toutefois, dans le cadre du projet OCDE/PISA, on donne une importance particulière au type de mathématisation sollicité dans la classe 3. C'est cette forme de mathématisation qui dépasse la simple *reconnaissance* de problèmes connus.

Les différents niveaux de complexité de la mathématisation sont illustrés par les deux exemples suivants (voir figures 5 et 6). Tous deux sont conçus pour des élèves âgés de 13 à 15 ans, et tous deux font appel à des concepts mathématiques analogues. Le premier demande une mathématisation simple, et le second, une mathématisation plus complexe.

L'item de la figure 5 est un exemple d'item de compétences de classe 2 car il ne fait appel qu'à une mathématisation relativement mineure.

Figure 5. **Exemple d'item sollicitant une « mathématisation » simple**

Dans une classe, il y a 28 élèves. Le rapport filles-garçons est de 4:3. Combien de filles y a-t-il dans cette classe ?

Source : TIMSS Mathematics Achievement in the Middle Years, p. 98.

OCDE 1999

Figure 6. **Exemple d'item sollicitant une « mathématisation » plus complexe**

Dans un pays donné, en 1980, le budget de la défense nationale était de 30 millions dans la monnaie du pays. Cette année-là, le budget total s'élevait à 500 millions. L'année suivante, le budget de la défense était passé à 35 millions, tandis que le budget total passait à 605 millions. Le taux d'inflation pendant la période couvrant les deux exercices budgétaires s'élevait à 10 %.

a) Vous êtes invité à donner une conférence pour une association pacifiste. Vous souhaitez expliquer que le budget de la défense a diminué cette année. Expliquez comment vous vous y prenez.

b) Vous êtes invité à donner une conférence dans une école militaire. Vous souhaitez expliquer que le budget de la défense a augmenté cette année. Expliquez comment vous vous y prenez.

Source : de Lange (1987). Voir aussi MSEB (1991). Reproduction autorisée.

La mathématisation sollicitée dans l'exemple de la figure 6 relève clairement des compétences de classe 3, en ce que cet exemple nécessite la reconnaissance par l'élève de mathématiques pertinentes, l'élaboration puis la communication d'une argumentation mathématique.

Les « idées mathématiques majeures »

Comme indiqué précédemment, les objectifs de l'étude OCDE/PISA sur l'évaluation des mathématiques sont assez différents de ceux d'études comparatives antérieures sur les mathématiques, en particulier de l'étude TIMSS de l'IEA. Les différences les plus importantes sont discutées plus loin.

Le processus d'élaboration des tests TIMSS de l'IEA attachait beaucoup d'importance à une bonne couverture des programmes des pays participants et recourrait à un schéma détaillé, fondé sur les domaines traditionnellement enseignés pour décrire le programme national. Cependant, les mathématiques scolaires sont souvent présentées aux élèves comme une science strictement compartimentée, et elles surestiment le calcul et les formules. Il est typique que les élèves quittant l'enseignement secondaire ne soient pas conscients que les mathématiques se développent continûment, en s'étendant à de nouveaux champs et de nouvelles *situations*. En conséquence, les instruments de TIMSS s'en tenaient pour la plupart à des connaissances mathématiques factuelles qui étaient testées isolément, le plus souvent au moyen d'items très courts.

En revanche, l'enquête OCDE/PISA porte davantage sur la culture mathématique telle que définie plus haut. C'est pourquoi il est important de souligner que l'objectif du projet OCDE/PISA est d'évaluer, dans toute son étendue, l'éventail des acquis des élèves, d'une manière cohérente et intégrée, plutôt que de tester des fragments de connaissances factuelles se situant parmi les compétences de classe 1. Pour le projet PISA, les interrelations et les idées communes à plusieurs domaines sont centrales. Les mathématiques sont le langage qui décrit des régularités – tant les régularités présentes dans la nature que celles qui sont inventées par l'esprit humain. Pour être « cultivés » sur le plan mathématique, les élèves doivent pouvoir reconnaître ces régularités et en discerner la diversité, les invariants et les interrelations.

C'est pour cette raison que le découpage traditionnel des contenus ne devrait pas être une dimension essentielle de l'étude OCDE/PISA sur le plan d'évaluation de la culture mathématique. L'alternative proposée est de structurer les contenus de l'évaluation autour des *idées mathématiques majeures.*

Le concept *d'idée majeure* n'est pas nouveau. En 1990, Senechal publiait *On the Shoulders of Giants : New Approaches to Numeracy* (« Sur les épaules des géants : nouvelles approches de la culture numérique ») qui est un plaidoyer énergique pour aider les élèves à approfondir et trouver les concepts sous-jacents à toutes les mathématiques, et à mieux appréhender ainsi leur signification dans la réalité. A cette fin, il

est nécessaire d'explorer les idées ayant des racines profondes dans les mathématiques sans considération des programmes scolaires actuels. D'autres mathématiciens défendent cette idée, les publications les plus fameuses étant celles de Devlin (1994, 1997), *Mathematics : The Science of Patterns* (« Mathématiques : la science des régularités »).

On peut identifier et décrire de nombreuses idées majeures. En fait, le domaine des mathématiques est si riche et varié qu'il ne serait pas possible de constituer une liste exhaustive d'idées majeures. Cependant, il est important pour l'objectif que poursuit le projet OCDE/PISA sur la culture mathématique, de proposer un choix d'idées majeures qui puissent suffisamment révéler, dans leur diversité et leur profondeur, l'essence même des mathématiques.

La liste suivante d'idées majeures utilisées par PISA satisfait à ces conditions :

– le hasard ;

– les variations et la croissance ;

– l'espace et les formes ;

– le raisonnement quantitatif ;

– l'incertitude ;

– la dépendance et les relations.

Dans le premier cycle d'évaluation de l'étude OCDE/PISA, le temps limité laissé aux mathématiques oblige à restreindre l'étendue du domaine évalué. Le premier cycle se centrera donc sur les deux idées majeures suivantes :

– les variations et la croissance ; et

– l'espace et les formes.

La suggestion de restreindre le premier cycle d'enquête à deux idées majeures est motivée pour deux raisons principales :

– premièrement, ces deux domaines couvrent un vaste éventail de sujets parmi les contenus présentés ci-dessus ;

– deuxièmement, ces domaines offrent une couverture satisfaisante des programmes scolaires existants.

Le raisonnement quantitatif a délibérément été omis du premier cycle de l'enquête, car il aurait pu entraîner une sur-représentation des savoir-faire typiquement numériques.

Ces deux idées majeures sont précisées plus bas.

Variations et croissance

Tout phénomène naturel subit des variations : les organismes qui changent en grandissant, le cycle des saisons, le flux et le reflux des marées, les cycles de fluctuation des taux de chômage, les variations météorologiques et l'indice Dow-Jones. Certains de ces processus de croissance peuvent être décrits ou modélisés par des fonctions mathématiques simples : les fonctions linéaires, exponentielles, périodiques, logistiques – soit discrètes, soit continues. Mais de nombreux processus relèvent de catégories différentes, et l'analyse des données y est souvent essentielle. L'utilisation des technologies informatiques a rendu les techniques d'approximation plus puissantes, et a permis une visualisation de données plus sophistiquée. Dans la nature comme dans les mathématiques, les régularités auxquelles obéissent les variations ne suivent en rien les découpages traditionnels des contenus scolaires.

Comme le dit Stuart (1990), pour être sensibles aux régularités des variations nous devons :

– représenter les variations sous une forme compréhensible ;

– comprendre les types fondamentaux de variations ;

– reconnaître des types de variations particuliers lorsqu'ils se manifestent ;

– appliquer ces techniques au monde extérieur ; et

– contrôler à notre avantage un univers en changement.

OCDE 1999

Ces compétences se rapportent étroitement à notre définition de la culture mathématique, et aux compétences telles que définies plus haut dans ce plan d'évaluation.

De nombreuses subdivisions du découpage traditionnel des contenus scolaires apparaissent dans cette idée mathématique majeure des variations et de la croissance. Les plus évidentes sont les relations, les fonctions et leurs représentations graphiques. Les séries et les gradients entretiennent des rapports étroits avec les fonctions. Observer les taux de croissance de différents phénomènes de croissance conduit à des courbes de croissance linéaires, exponentielles, logarithmiques, périodiques ou logistiques ainsi qu'à leurs propriétés et relations. A leur tour, celles-ci mènent à des aspects de la théorie des nombres, comme la suite de Fibonacci et le nombre d'or. Les liens qui relient ces idées aux représentations géométriques peuvent également jouer un rôle.

La géométrie peut aussi permettre d'observer des régularités présentes dans la nature, dans l'art ou l'architecture. L'analogie et la congruence peuvent avoir un rôle à jouer ici, tout comme ce serait le cas pour le rapport entre l'augmentation d'une aire et celle du périmètre ou de la circonférence.

Les modèles de croissance peuvent trouver leurs expressions algébriques, et ces dernières peuvent à leur tour être représentées par des graphiques.

La croissance peut aussi être mesurée de façon empirique, ce qui provoque un questionnement sur ce qu'il est possible d'inférer à partir des données et sur la meilleure manière de les représenter. Des aspects d'analyse des données et de statistiques émergent alors tout naturellement.

Espace et formes

Les processus de variations et de croissance ne sont pas les seuls à présenter des régularités de structure. Celles-ci sont omniprésentes autour de nous : dans le langage, la musique, la vidéo, la circulation, les constructions, l'art. Les formes sont des structures : maisons, églises, ponts, étoiles de mer, flocons de neige, plans de villes, feuilles de trèfle, cristaux, ombres. Les structures géométriques peuvent servir de modèles relativement simples pour quantité de phénomènes, et leur étude est possible et souhaitable à tous les niveaux (Grünbaum, 1985). En mathématiques, les formes constituent un thème fondamental en plein développement et tout à fait fascinant, un thème profondément enraciné dans la géométrie traditionnelle (quoique le lien soit plus ténu avec la géométrie scolaire), mais qui va bien au-delà par son contenu, son sens et sa méthode (Senechal, 1990).

Dans l'étude des formes et des constructions, nous nous intéressons à leurs similitudes et à leurs différences en analysant leurs composantes formelles et en nous efforçant de reconnaître des formes sous des représentations et dans des dimensions différentes. L'étude des formes est étroitement liée à *l'appréhension de l'espace* (Freudenthal, 1973). Cela signifie apprendre à connaître, à explorer, à conquérir l'espace afin de mieux y vivre, y respirer et s'y déplacer.

Cela implique d'être capable de comprendre les positions relatives des objets, de savoir comment nous percevons les choses et pourquoi nous les percevons de cette manière. Nous devons apprendre à voyager dans l'espace, au sein de constructions et de formes. Cela implique que les élèves soient capables de comprendre la relation entre la forme et son image, ou sa représentation visuelle – par exemple, entre une ville réelle et ses plans ou ses photographies. Ils doivent aussi comprendre comment des objets en trois dimensions peuvent être représentés en deux dimensions, comment les ombres se forment et s'interprètent, ce qu'est la « perspective » et comment elle fonctionne.

Décrite sous cet angle, l'étude de l'espace et des formes est ouverte et dynamique, et elle s'accorde bien, à la fois à la notion de culture mathématique et à celle de compétences mathématiques, telles que définies dans PISA.

Les domaines mathématiques enseignés

Il est évident que les domaines traditionnels d'enseignement des mathématiques ne peuvent, ni ne doivent être ignorés. C'est pourquoi nous les avons intégrés dans PISA de manière explicite, sous forme d'un aspect mineur de l'organisation du domaine de la culture mathématique. L'aspect *domaines*

mathématiques enseignés peut être utile pour trouver un bon équilibre entre les items et pour assurer une couverture raisonnable de l'éventail des contenus définis par les programmes scolaires. Les catégories de contenus définies par domaine pour l'étude OCDE/PISA seront les suivantes :

– nombres ;
– mesures ;
– estimation ;
– algèbre ;
– fonctions ;
– géométrie ;
– probabilités ;
– statistiques ; et
– mathématiques discrètes.

Cette liste des domaines enseignés a été élaborée en collaboration avec les pays participants à PISA. Des items couvrant chacun des domaines mentionnés ci-dessus seront inclus dans les tests d'évaluation de PISA.

Les situations

Un aspect important de la définition de la culture mathématique est lié à la grande diversité des situations dans lesquelles sont utilisées les mathématiques. En effet, on a remarqué que le choix des méthodes mathématiques et la présentation des résultats dépendent souvent des contextes dans lesquels les problèmes sont présentés au départ. Chaque situation devrait permettre aux élèves de participer au processus social de mathématisation, en reconnaissant comment les pratiques apprises dans une situation peuvent aisément être appliquées dans d'autres situations similaires.

On peut se représenter les *situations* comme se trouvant à des « distances » variables par rapport à l'élève. Parmi les situations identifiées pour l'enquête PISA, la plus proche de l'élève est d'abord la vie privée, puis viennent la vie scolaire, le travail et le sport (ou les loisirs en général), ensuite la collectivité locale et la société dans les diverses formes qu'elles prennent au quotidien ; viennent enfin les contextes scientifiques. Il est à noter que les contextes scientifiques incluent la démonstration de conjectures de caractère abstrait, la généralisation de régularités numériques ou spatiales et d'autres activités de type semblable.

A ainsi été définie une échelle plus ou moins continue que l'on considérera comme une autre dimension ou un autre axe du plan d'évaluation de PISA dans lequel l'accent sera mis principalement sur les cinq situations suivantes : les situations privée, scolaire, professionnelle, publique et scientifique.

Un aspect lié aux *situations* est l'authenticité des contextes utilisés dans les problèmes. Cet aspect sera examiné plus en détail ci-dessous.

Caractéristiques des épreuves

Le domaine lié à la culture mathématique a été défini dans les parties précédentes du plan OCDE/PISA et la structure en a été définie. Nous considérons maintenant les épreuves qui serviront à évaluer les élèves. Dans cette partie, la nature des épreuves est décrite : leur contexte mathématique, leur format et leur processus d'élaboration.

Contexte mathématique des épreuves

Une question à prendre en considération lors de l'élaboration des épreuves de test porte sur les contextes mathématiques dans lesquels les items sont construits. Le terme de *contexte* est utilisé conformément à l'acception en usage dans l'enseignement des mathématiques. Un contexte est un cadre

OCDE 1999

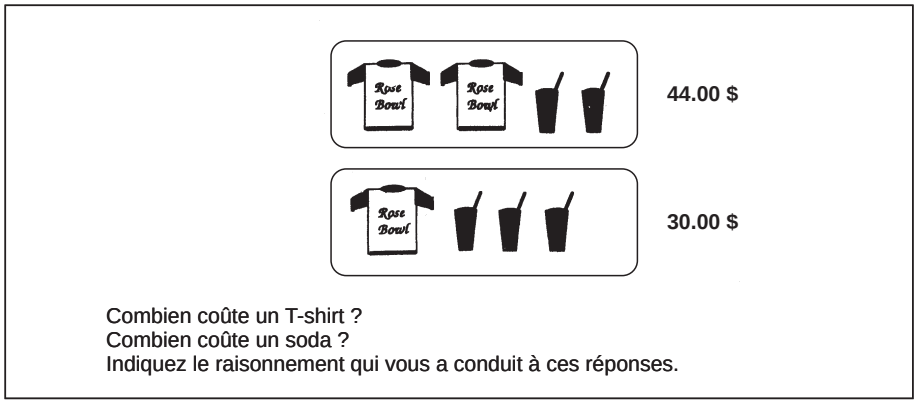
*extra*mathématique ou *intra*mathématique dans lequel les éléments d'un *complexe mathématique* (à savoir un problème, une tâche ou un ensemble d'objets, de relations, de phénomènes mathématiques, etc.) doivent être interprétés. En somme, un contexte est soit un cadre dans lequel un complexe mathématique est déjà inséré (cadre intramathématique), soit un cadre qui se prête à l'activation d'un tel complexe, qui s'insère alors dans ce contexte (cadre extramathématique). L'insertion d'un complexe mathématique dans un contexte extramathématique implique toujours la présence explicite ou implicite (tacite) d'un modèle mathématique qui représente le cadre (ou certains aspects de celui-ci) en opérant une traduction dans le complexe en question.

La définition du contexte proposée ci-dessus permet le recours à une grande variété de contextes. On pourra par exemple emprunter des contextes à d'autres domaines, comme ceux de la pratique professionnelle, de la vie quotidienne, de la vie de la collectivité ou de la société. Cette définition englobe également le contexte des loisirs comme le sport et les jeux. Les *situations* définies précédemment en tant qu'aspect mineur, constituent une forme de contexte. L'ensemble des épreuves d'évaluation devra puiser dans un large éventail de contextes : leur variété est indispensable pour assurer la diversité culturelle et représenter le spectre des rôles que les mathématiques peuvent jouer.

Les épreuves d'évaluation du plan OCDE/PISA seront limitées aux contextes *authentiques*. Un contexte est considéré comme authentique s'il se situe dans le domaine du vécu et des pratiques effectives des participants, dans un cadre réel. Il est à noter que cette définition ne demande pas que les élèves évalués fassent eux-mêmes partie de ce cadre. Nous pouvons prendre comme exemple des questions ayant trait à de l'argent placé dans une banque, à un taux d'intérêt réaliste, même si elles sortent de la sphère d'expérience de l'élève évalué.

Il est important de reconnaître que l'usage de composantes de la réalité ne suffit pas à rendre un contexte authentique. Considérons par exemple les épreuves des figures 7 et 8. Deux de ces épreuves comportent des éléments réels mais elles ne sont pas authentiques parce que personne dans un cadre extra-scolaire ne pourrait être conduit à aborder ce genre de problème.

Figure 7. **Une épreuve réelle mais non authentique**



Les contextes de ces problèmes ont été choisis pour leur donner (superficiellement) une allure réelle. Le projet OCDE/PISA devrait autant que faire se peut éviter le recours à cette sorte de contextes.

L'accent mis par PISA sur les contextes authentiques n'exclut pas l'insertion de contextes mathématiques importants et/ou intéressants (ce peut être parfois des contextes virtuels). Considérons par exemple l'épreuve de la figure 9. Ici le contexte est stylisé ou généralisé. Il peut être ou ne pas être

Figure 8. **Exemple d'item dans un contexte créé**

Parmi les expressions numériques suivantes, laquelle pourrait être utilisée pour résoudre le problème ? Bill pesait 48 kilos l'été dernier. Il a perdu 2 kilos, puis en a repris 5. Combien pèse-t-il à présent ?

a) 48 - (2 + 5) = A
b) (48 - 2) + 5 = A
c) (48 + 5) + 2 = A
d) -2 + 5 = 48 + A
e) (48 - 5) + 2 = A

authentique. De tels contextes auront leur place dans l'évaluation OCDE/PISA s'ils sont mathématiquement intéressants et pertinents. L'usage des mathématiques pour expliquer des scénarios hypothétiques et explorer des systèmes et situations potentiels, même non réalisables dans la réalité, est l'une de leurs plus importantes caractéristiques.

Figure 9. **Exemples d'item dans un contexte «virtuel»**

Serait-il possible d'instaurer un système de monnaie (ou un système de timbres) qui utiliserait exclusivement les valeurs 3 et 5 ? Plus spécifiquement, quels montants pourrait-on obtenir ainsi ? S'il s'avérait possible, un tel système serait-il souhaitable ?

Formats des épreuves

La création d'instruments d'évaluation doit soigneusement prêter attention à l'impact du format des épreuves sur les performances des élèves et à la définition du *construct* (voir note 1, p. 19) qui est évalué. Ce point est particulièrement important dans un projet du type de PISA dans lequel les questions d'échelle et le caractère transnational créent de sérieuses contraintes sur le spectre de formats d'items réalisables.

Comme pour le domaine de compréhension de l'écrit, le programme OCDE/PISA évaluera la culture mathématique à travers plusieurs items à choix multiple, des réponses fermées et des réponses ouvertes. L'annexe 2 examine un spectre plus large de formats susceptibles d'être utilisés quand les mathématiques constitueront un domaine majeur dans le deuxième cycle de l'enquête.

Commentant la deuxième étude IEA sur les mathématiques, Travers et Westbury (1989) relèvent que « la construction et la sélection des items à choix multiple n'ont pas été difficiles pour les niveaux de comportement cognitif les plus bas – le calcul et la compréhension ». Ils ajoutent cependant que les difficultés se sont manifestées pour des niveaux plus élevés. Il y a une place pour les items à choix multiple (voir la figure 10), mais seulement dans un nombre restreint de cas et pour les objectifs (ou les comportements) les moins élevés. Pour les objectifs d'ordre supérieur et les processus plus complexes, on devrait privilégier d'autres formats de test, dont le plus simple est la *question ouverte*.

Les items fermés à réponse construite posent des questions analogues à ceux des items à choix multiple mais on demande aux élèves de produire une réponse qui peut être aisément jugée correcte ou incorrecte. Quand les réponses ne sont pas notées par machine, c'est un format de prédilection pour évaluer les compétences de classe 1, parce qu'il est improbable que la bonne réponse puisse être

OCDE 1999

Figure 10. **Exemple d'item proposant un nombre limité de réponses possibles**

Un phoque doit remonter à la surface pour respirer, même quand il dort. Martin a observé un phoque pendant une heure. Au début de l'observation, le phoque a plongé au fond de l'eau et s'est endormi. Au bout de 8 minutes, il s'est lentement laissé remonter à la surface et a respiré.

En 3 minutes, il avait regagné le fond de la mer et le même cycle a recommencé depuis le début, selon un rythme très régulier.

Au bout d'une heure, le phoque était :

a) au fond
b) en train de remonter à la surface
c) en train de respirer
d) en train de redescendre vers le fond

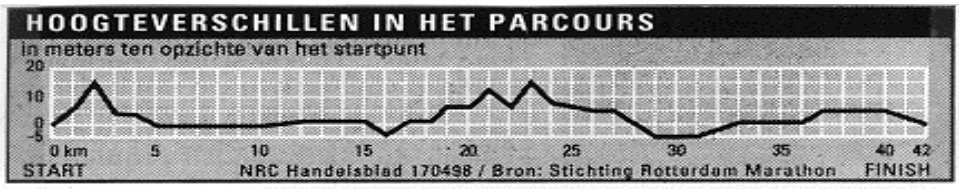
devinée et qu'il n'est pas nécessaire de prévoir des distracteurs qui pourraient influencer le *construct* à évaluer. Par exemple, dans le problème de la figure 11, il n'y a qu'une réponse correcte et de multiples réponses incorrectes possibles.

Figure 11. **Exemple d'item avec une réponse correcte et de nombreuses réponses incorrectes**

Tepla Loroupe a remporté le marathon de Rotterdam en 1998. « C'était facile », dit-elle, « le parcours était plutôt plat ».

Ci-dessous, vous voyez un graphique des dénivellations dans le parcours du marathon de Rotterdam :

[***DÉNIVELLATIONS DANS LE PARCOURS — en mètres par rapport au point de départ***]



Quelle était la différence d'altitude entre le point le plus élevé et le point le plus bas de la course ?

Les items ouverts à réponse construite

Ces items sollicitent une réponse plus longue de la part de l'élève évalué et le processus d'élaboration de la réponse est susceptible de mettre en jeu des activités d'ordre plus élevé. Ces questions requièrent souvent de l'élève non seulement qu'il fournisse une réponse, mais qu'il montre aussi les étapes de son raisonnement ou explique comment il est arrivé à cette réponse. La caractéristique principale des items ouverts à réponse construite est qu'ils donnent l'occasion aux élèves de faire preuve de leurs compétences en fournissant des solutions se situant sur tout un spectre de différents niveaux de complexité mathématique. L'item présenté à la figure 12 en est un exemple.

Figure 12. **Exemple d'item nécessitant une réponse construite**

L'Indonésie se situe entre la Malaisie et l'Australie. Quelques données sur la population de l'Indonésie et sa répartition sur les îles sont représentées dans le tableau suivant :

Région	Superficie (km^2)	Pourcentage de la superficie totale	Population en 1980 (millions)	Pourcentage de la population totale
Java/Madura	132 187	6.95	91 281	61.87
Sumatra	473 606	24.86	27 981	18.99
Kalimantan (Bornéo)	539 460	28.32	6 721	4.56
Sulawesi (Célèbes)	189 216	9.93	10 377	7.04
Bali	5 561	0.30	2 470	1.68
Irian Jaya	421 981	22.16	1 145	5.02
TOTAL	1 905 569	100.00	147 384	100.00

L'un des problèmes importants de l'Indonésie est la répartition inégale de sa population sur les îles. A l'aide du tableau, nous pouvons voir que Java, qui a moins de 7 % de la superficie totale, possède presque 62 % de la population.

Question : Faire un graphique qui montre la répartition inégale de la population indonésienne.

Source : de Lange et Verhage (1992). Reproduction autorisée.

Pour le programme OCDE/PISA, 25 à 35 pour cent de l'évaluation des mathématiques seront consacrés à des items ouverts à réponse construite. Ces items nécessitent une notation par un correcteur expérimenté qui met en place un barème qui peut exiger des éléments de jugement professionnel. En raison de désaccords éventuels entre correcteurs, le plan OCDE/PISA mettra en œuvre des études pour réguler l'étendue des désaccords entre correcteurs. La conclusion de ces études montre que des barèmes clairs peuvent être élaborés et qu'on peut obtenir une notation fiable.

Le premier cycle de l'enquête OCDE/PISA fera usage d'un format d'épreuve dans lequel plusieurs items dépendent d'un même matériau source. Les épreuves de ce format offrent aux élèves l'occasion de s'impliquer dans le contexte ou le problème, en posant une série de questions ouvertes avec une gradation croissante du niveau de complexité. Les premières questions sont à choix multiple ou fermées à réponse construite alors que les items suivants sont ouverts à réponse construite. Ce format est approprié pour toutes les classes de compétences.

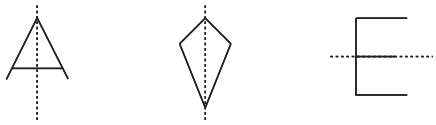
Une des raisons à l'origine de tels formats d'épreuve est de rendre les tâches réalistes et de permettre qu'elles reflètent la complexité de la vie réelle. Une autre raison tient à l'efficacité d'utilisation du temps de test. La nécessité de rendre, dans l'épreuve, chaque point du barème indépendant des autres est reconnue et prise en compte dans la conception des tâches assignées au programme PISA. Il est aussi reconnu qu'il est important de minimiser les biais dus au nombre restreint de situations.

La figure 13 donne un exemple d'une épreuve à plusieurs items.

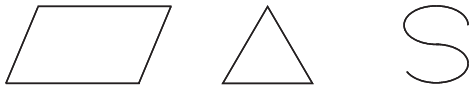
OCDE 1999

Figure 13. **Une épreuve à plusieurs items**

Si une figure peut être pliée de telle manière que les deux moitiés soient exactement l'une sur l'autre, la droite de pliage est dite de symétrie.



Question A
Quelles sont les figures ci-dessous qui admettent une droite de pliage de symétrie ?



(1) (2) (3)

Question B
Tracer toutes les droites de symétrie du carré.



Question C
Lesquelles des huit premières lettres majuscules de l'alphabet admettent exactement deux droites de pliage de symétrie ?

Question D
Jean a dit : « Je connais une règle qui permet de dire quand une figure à 4 côtés a une droite de pliage de symétrie. Si les triangles de chaque côté de la droite ont la même taille et la même forme, la figure a une droite de pliage de symétrie. »
Expliquez pourquoi vous êtes d'accord ou en désaccord avec Jean.

Structure des tests d'évaluation

Cette partie décrit la structure des tests d'évaluation de la « culture mathématique » du premier cycle de PISA dans lesquels un temps total de 60 minutes est consacré à évaluer la culture mathématique des élèves.

Dans le premier cycle du plan OCDE/PISA, la durée prévue pour les tests sera répartie à part égale entre les deux *idées mathématiques majeures* : variations et croissance ; espace et formes. La répartition approximative entre les trois *classes de compétences* sera de 1-2-1. Ces informations sont résumées au tableau 9, dans lequel est indiqué le nombre d'items pour chaque idée majeure et classe de compétences. Les items sont ventilés selon qu'ils nécessitent un ou plusieurs correcteurs.

Les aspects mineurs ne sont pas présentés dans le tableau 9. La répartition des items du test entre les aspects mineurs des domaines enseignés et les situations est à peu près uniforme. C'est dire que le test OCDE/PISA accordera une égale importance à chacun des neuf domaines enseignés et à chacune des cinq situations mentionnées plus haut.

Le tableau 9 montre qu'il est prévu que le test comprenne :

– 15 items pour chacune des deux idées majeures ;

– 21 points pour chacune des deux idées majeures ;

Tableau 9. **Nombre d'items recommandés et de points du barème par idée majeure et par classe de compétences**

Type d'item	Variations et croissance			Espace et formes		
	Compétences			Compétences		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
Un seul correcteur	6(6)	5(5)		6(6)	5(5)	
Plusieurs correcteurs		2(5)	2(5)		2(5)	2(5)

Note : Les nombres entre parenthèses correspondent au nombre de points présumés.

– 8 items à plusieurs correcteurs et 22 items à correcteur unique ;

– 12 points dans la classe de compétences 1, 20 points dans la classe de compétences 2 et 10 points dans la classe de compétences 3.

A long terme, le plan OCDE/PISA a l'intention de donner une plus grande importance aux classes 2 et 3 que ce n'est le cas dans le premier cyle de l'enquête où le temps imparti au test de mathématiques est très limité.

Dans le premier cycle, la composition d'un module type d'une demi-heure pour le test sur la culture mathématique est la suivante :

– un petit nombre (2-4) d'items à choix multiple ou d'items fermés à réponse construite, destinés à tester les compétences des classes 1 ou 2 ;

– un petit nombre (1-2) de problèmes contenant chacun deux ou trois items dans un même contexte, destinés à tester les compétences des classes 1 ou 2 ;

– un *bloc d'items* consistant en plusieurs items situés dans un même contexte. Les items commenceraient par des tests relativement directs évaluant les compétences de classe 1 pour aller vers des tests de plus grande complexité évaluant les compétences de classe 3.

Échelles des résultats

L'élaboration d'échelles des acquis des élèves est essentielle pour atteindre les objectifs du plan OCDE/PISA. Le processus d'obtention d'une échelle doit être itératif en ce que les propositions initiales basées sur l'expérience passée et la recherche sur l'apprentissage ainsi que sur le développement cognitif en mathématiques évoluent grâce aux résultats des essais pilotes réalisés durant le projet PISA.

Le choix d'échelles pour l'étude mineure de la culture mathématique doit encore être fait. Les choix les plus évidents consistent à présenter les scores sur *i)* une échelle unique pour la culture mathématique, *ii)* une échelle séparée pour chacune des idées majeures ou *iii)* une échelle séparée pour chacune des trois classes de compétences. Le choix entre ces trois alternatives se fera après l'analyse des données de l'essai de terrain.

Les résultats de l'étude OCDE/PISA seront plus informatifs si, au moins pour quelques items, des notes sont données non seulement pour les bonnes réponses mais aussi selon les différentes stratégies utilisées par les élèves. On ne devrait pas être surpris si, par exemple, les élèves de deux pays différents avaient le même score sur l'échelle de culture mathématique mais différaient grandement dans la façon dont ils avaient atteint ce niveau de culture mathématique : dans un pays, les élèves peuvent se fier davantage aux stratégies formelles et aux routines alors que dans un autre des stratégies plus informelles et de bon sens sont peut-être plus employées.

OCDE 1999

Autres questions

Liens avec d'autres tests d'évaluation

Le centre d'intérêt du plan OCDE/PISA est différent de celui des enquêtes comparatives en mathématiques précédentes, telles que la Troisième étude internationale sur les mathématiques et les sciences (TIMSS). Si l'étude TIMSS de l'IEA est un instrument construit autour du dénominateur commun du programme scolaire national des pays participants, le but du programme OCDE/PISA est d'évaluer la culture mathématique telle que nous l'avons définie plus haut. Les personnes familières avec TIMSS reconnaîtront des liens entre les « compétences mathématiques » du plan OCDE/PISA et les « performances attendues » de TIMSS, ainsi que la similitude des « domaines mathématiques enseignés ». Cependant, dans le plan OCDE/PISA, les domaines enseignés sont sous-jacents aux idées majeures abordées par la culture mathématique. De la même façon, si la plupart des items dans les catégories de performance de TIMSS correspondent aux compétences de classe 1, le but du plan OCDE/PISA est de couvrir également les compétences de classes 2 et 3.

Pour cette raison, on n'établira pas de liens psychométriques entre les échelles de l'étude PISA et celles de TIMSS.

Outils et instruments

Trois réglementations possibles concernant l'usage des calculatrices et autres instruments peuvent être prises dans l'enquête OCDE/PISA :

– les élèves n'ont pas le droit d'utiliser de calculatrice ;

– les élèves ne peuvent utiliser qu'une calculatrice fournie dans le cadre de l'enquête OCDE/PISA ;

– les élèves sont libres d'utiliser leur propre calculatrice et leurs instruments.

C'est la troisième des possibilités mentionnées ci-dessus qui sera retenue dans le cadre de PISA. Cette option a été choisie car elle constitue l'évaluation la plus authentique des apprentissages des élèves, et fournira la meilleure possibilité de comparaison entre systèmes éducatifs. La décision de permettre aux élèves d'utiliser des calculatrices s'ils le souhaitent n'est pas différente en principe de toute autre décision de réglementation prise par des systèmes éducatifs et non contrôlée par le plan OCDE/PISA.

En outre, l'argument selon lequel les deux premières options conduiraient à une évaluation plus juste en raison de l'égalité apparente des conditions de test est sans valeur. Les élèves habitués à répondre à des questions en s'aidant d'une calculatrice seront désavantagés si on la leur enlève. Les élèves à qui l'on fournit une calculatrice (avec laquelle ils ne sont pas familiers) peuvent ne pas l'utiliser de façon efficace, ou le simple fait de la fournir peut les encourager à y avoir recours de façon inutile ou inappropriée. Par exemple, pour de nombreux élèves, le simple problème $6 + 4 \times 3 = ?$ est plus difficile avec une calculatrice que sans – particulièrement pour les élèves non initiés aux calculatrices.

Le plan OCDE/PISA propose donc que les élèves soient autorisés à utiliser comme ils l'entendent calculatrices et autres instruments qu'ils utilisent généralement à l'école. Cependant, les items du test OCDE/PISA seront sélectionnés de telle manière que l'emploi de calculatrices lors de l'évaluation ne pourra pas améliorer les performances mathématiques d'un élève.

Chapitre 3

LA CULTURE SCIENTIFIQUE

Un atout pour les jeunes d'aujourd'hui est d'acquérir des compétences leur permettant de tirer des conclusions adéquates et avisées à partir des preuves et des informations dont ils disposent, de critiquer des affirmations émises par d'autres en mettant en cause les éléments sur lesquels elles se fondent, et de faire la distinction entre opinions et conclusions étayées par des faits. La science a ici un rôle bien particulier à jouer, puisque son propos est d'établir un principe de rationalité dans la confrontation des idées et des théories avec les faits observables dans le monde environnant. Cela ne veut pas dire que la science exclut la créativité et l'imagination : celles-ci ont toujours joué un rôle essentiel dans les progrès de la compréhension par l'homme du monde qui l'entoure. Certaines idées, qui semblent parfois être « tombées du ciel », ont, en fait, été saisies au vol selon un mécanisme qu'Einstein décrivait comme *« le chemin de l'intuition, qui est favorisé par une certaine sensibilité pour l'ordre caché derrière les apparences »* (Einstein, 1933). Le fait que ces idées aient été « saisies au vol » à une époque bien précise est déterminé, historiquement, par leur acceptabilité sociale à cette époque. Les progrès du savoir scientifique ne sont donc pas uniquement le fruit de la créativité des individus, mais également de la culture dans laquelle ils se produisent. Quand le « sursaut » créatif s'est produit et qu'un nouveau schéma de compréhension théorique a été mis en place, l'étape suivante consiste à confronter l'idée à la réalité des faits. Comme l'écrit Hawking :

> *« une théorie est bonne si elle satisfait à deux exigences : elle doit décrire fidèlement une large catégorie d'observations sur la base d'un modèle qui ne contient que peu d'éléments arbitraires, et elle doit fournir des prédictions précises sur les résultats de futures observations » (Hawking, 1988, p. 9).*

Les théories qui ne répondent pas à ces exigences – ou qui ne peuvent pas être mises à l'épreuve – ne sont pas des théories scientifiques, et il importe qu'un citoyen instruit soit capable de faire la distinction entre les questions auxquelles la science peut répondre et celles auxquelles elle ne le peut pas, comme entre ce qui est scientifique et ce qui relève du pseudo-scientifique.

Définition du domaine

La conception contemporaine des objectifs visés par l'éducation scientifique met surtout l'accent sur le développement, chez tous les citoyens, d'une compréhension globale d'aspects scientifiques essentiels et des structures fondamentales des sciences, des méthodes par lesquelles elles recueillent les éléments permettant de fonder les connaissances, et donc aussi de ce que sont les potentialités et les limites de la science dans le monde réel. On considère comme essentiel que l'individu soit capable d'appliquer ces acquis dans des situations du monde réel où la science est impliquée, où des affirmations doivent être vérifiées et des décisions prises. Ainsi, Millar et Osborne (1998) estiment qu'un enseignement moderne en sciences doit avoir avoir pour cible « la capacité de lire et d'assimiler des informations scientifiques et techniques et d'évaluer leur signification ». Ils ajoutent :

> *« Dans cette optique, l'essentiel n'est pas de « faire de la science ». Ce n'est pas non plus la façon dont on parvient à la connaissance scientifique, ou dont on se la remémore brièvement lors d'un examen final. […] Ainsi, en sciences, on demande aux élèves de se montrer capables d'évaluer des faits, de faire la distinction entre théories et observations, et d'estimer le degré de confiance qu'on peut avoir à propos des explications avancées » (Millar et Osborne, 1998).*

OCDE 1999

Ces objectifs sont ceux que l'enseignement des sciences devrait poursuivre pour l'ensemble de la population. Pour la minorité d'élèves qui deviendront les chercheurs de demain, le champ sera étendu à l'étude approfondie de notions scientifiques et au développement de la capacité de « faire de la science ».

Dans cette perspective, on estime que l'objet essentiel de l'enseignement des sciences, qui doit être au centre de l'évaluation conduite par PISA, est de produire des élèves *scientifiquement cultivés*. Cette expression a été utilisée dans différents contextes. Par exemple, la définition proposée par l'*International Forum on Scientific and Technological Literacy for All* en donne plusieurs versions, dont la suivante :

« la capacité d'agir avec compréhension et assurance, aux niveaux adéquats, selon des modalités qui conduisent à une meilleure maîtrise du monde matériel et du monde des idées scientifiques et technologiques » (UNESCO, 1993).

Ces diverses conceptions de la culture scientifique (voir la synthèse de Shamos, 1995 ; et aussi Grabe et Bolte, 1997) font appel à la notion de *niveau de culture scientifique*. Par exemple, Bybee (1997) propose quatre niveaux, dont les deux inférieurs désignent « la culture scientifique nominale » (connaissance des noms et de la terminologie), et la « culture scientifique fonctionnelle », qui s'applique aux élèves capables d'utiliser un vocabulaire scientifique dans des contextes limités. Nous considérons ces deux niveaux comme trop bas pour être pris en compte dans le plan d'évaluation OCDE/PISA. Le niveau le plus haut identifié par Bybee, celui de la « culture scientifique multidimensionnelle », requiert une bonne compréhension de la nature de la science et de son histoire, ainsi que de son rôle dans notre culture, à un niveau s'appliquant plutôt à l'élite scientifique qu'à l'ensemble des citoyens. C'est peut-être l'idée que la culture scientifique implique une réflexion aussi spécialisée qui freine la diffusion d'une conception plus « accessible » de cette culture. La définition la plus pertinente et la plus proche des objectifs du plan d'évaluation scientifique de PISA paraît être celle relative au troisième niveau de culture scientifique, désigné par Bybee comme la « culture scientifique conceptuelle et procédurale ».

Après avoir examiné un certain nombre de descriptions existantes, le programme OCDE/PISA définit la culture scientifique comme suit :

« La culture scientifique est le fait de pouvoir utiliser des connaissances scientifiques, d'identifier les questions et de tirer des conclusions fondées sur des faits, en vue de comprendre le monde naturel et de prendre des décisions à son propos, ainsi que de comprendre les changements qui y sont apportés par l'activité humaine. »

Il est nécessaire de détailler la signification des éléments condensés dans cette définition.

Culture scientifique…

Il est important de noter ici que d'une part, les connaissances scientifiques (au sens des connaissances sur la science) ainsi que les démarches selon lesquelles elles sont produites, sont des aspects essentiels de la culture scientifique et qu'elles sont étroitement liées dans l'acception donnée ici à cette expression. Comme on le verra plus loin de manière détaillée, les savoir-faire concernant les démarches scientifiques ne sont réellement des savoir-faire scientifiques que lorsqu'ils sont appliqués à des matières scientifiques. L'utilisation de ces savoir-faire implique donc nécessairement une certaine compréhension du domaine concerné. La conception de la culture scientifique adoptée ici met en avant la combinaison entre raisonnement et compréhension à propos des aspects scientifiques du monde.

… pouvoir utiliser des connaissances scientifiques, identifier les questions et tirer des conclusions fondées sur des faits…

Dans la définition proposée ci-dessus, l'expression « connaissances scientifiques » désigne bien plus que la connaissance de faits, de noms et de la terminologie. Elle inclut la compréhension de concepts scientifiques fondamentaux, la perception des limites de la connaissance scientifique ainsi que l'appréciation de l'aspect culturel de la science. Les questions qui doivent être identifiées sont celles auxquelles on peut répondre par la méthode scientifique, impliquant une connaissance des sciences ainsi que des aspects scientifiques des sujets spécifiques. Par capacité de « tirer des conclusions fondées sur des faits », on entend connaître et être capable d'appliquer les procédures nécessaires

pour la sélection et l'évaluation d'informations ou de données. Cependant, les informations fournies ne suffisent souvent pas à tirer des conclusions définitives et il faut être capable de spéculer, prudemment et en toute conscience, à partir des informations disponibles.

… comprendre le monde naturel et prendre des décisions à son propos…

Comprendre le monde naturel et prendre des décisions à son propos indique tout d'abord que la compréhension du monde naturel est valorisée à la fois comme un objectif en soi et comme un élément nécessaire à la prise de décisions ; d'autre part, s'il est vrai qu'elle peut contribuer à la prise de décisions, cette compréhension n'en sera pas nécessairement un facteur déterminant. Dans la vie pratique de tous les jours, les décisions à prendre reposent toujours sur des composantes sociales, politiques ou économiques et les connaissances scientifiques s'intègrent au niveau des valeurs humaines relatives à celles-ci. Quand, dans une situation donnée, apparaît un consensus de valeurs, l'apport de faits scientifiques sert souvent d'appui à ce qui a été accepté. Par contre, en cas de désaccord, l'apport de faits scientifiques nécessaires à la prise de décision sera plus enclin à la controverse.

… comprendre les changements qui y sont apportés par l'activité humaine

L'expression « le monde naturel » désigne de façon abrégée l'environnement physique, les êtres vivants et les rapports entre eux. Les « décisions à propos du monde naturel » se réfèrent à des décisions ayant un rapport à la science et qui concernent des problèmes de l'individu, de sa famille, de la collectivité ou du monde en général. L'expression « les changements apportés par l'activité humaine » se rapporte aux modifications introduites dans le monde naturel, qu'elles soient ou non intentionnelles, en vue de satisfaire les besoins de l'homme (technologies simples et complexes), ainsi qu'à leurs conséquences.

Il convient de signaler ici, et nous y reviendrons de façon plus explicite, que la culture scientifique n'est pas polarisée, c'est-à-dire que nous ne voulons pas laisser entendre que les gens peuvent être répartis en deux catégories : les scientifiquement cultivés et les scientifiquement incultes. Nous dirions plutôt qu'il existe un continuum entre culture scientifique peu développée et culture scientifique de haut niveau. Par exemple, un élève doté d'une culture scientifique peu développée pourra néanmoins se montrer capable d'identifier certains des faits objectifs nécessaires à l'évaluation d'une hypothèse ou à la défense d'un point de vue, ou même de fournir une analyse plus complète concernant des situations simples et familières. Un autre fera preuve d'une culture scientifique de plus haut niveau en fournissant des réponses plus complètes et en étant capable d'utiliser des savoirs scientifiques et d'évaluer des hypothèses relatives à des faits survenant dans des situations moins familières et plus complexes.

Organisation du domaine

La définition de la culture scientifique adopté par PISA se fonde sur trois composantes clé :

- Les *savoir-faire scientifiques* qui, parce qu'ils sont scientifiques feront intervenir les connaissances des sciences, bien que dans l'évaluation scientifique, celles-ci ne doivent pas constituer une barrière majeure à la réussite.
- Les *concepts scientifiques* dont la compréhension sera évaluée par l'application dans certains domaines de contenu.
- Les *situations* dans lesquelles les tâches d'évaluation sont présentées (il est d'usage de faire référence à cet aspect en utilisant les termes « contexte » ou « environnement »).

Bien que ces aspects de la culture scientifique soient étudiés séparément, il importe de noter que, lors de l'évaluation proprement dite de celle-ci, ils seront toujours combinés.

Les deux premiers aspects seront utilisés aussi bien pour construire les tâches que pour caractériser le résultat des élèves. La dimension « situations » servira à garantir, lors du développement des évaluations, que l'on fait intervenir un éventail suffisamment large de situations pertinentes en termes de culture scientifique.

OCDE 1999

Ces trois aspects organisateurs du domaine sont présentés de manière plus détaillée dans les sections suivantes. En se structurant autour de ces dimensions, le plan d'évaluation OCDE/PISA met l'accent sur le produit de l'éducation scientifique conçue comme un tout.

Savoir-faire scientifiques

Les savoir-faire scientifiques sont des actions mentales (et quelquefois physiques) mises en œuvre lors de la conception, du recueil, de l'interprétation et de l'utilisation de faits ou de données, dans le but de mieux connaître ou comprendre des phénomènes. Les savoir-faire scientifiques sont nécessairement liés à un thème ; l'idée qu'un savoir-faire scientifique puisse être utilisé sans contenu est dépourvue de sens. Ils peuvent être appliqués à toutes sortes de thèmes ; on parlera de *savoir-faire scientifiques* lorsque le thème abordé relève d'aspects scientifiques du monde et que la mise en œuvre de la démarche a pour but de déboucher sur une avancée dans la compréhension scientifique.

Les démarches scientifiques, telles qu'on les décrit habituellement, recouvrent un large ensemble de savoirs et de savoir-faire nécessaires pour recueillir et interpréter des faits relatifs au monde environnant et d'en tirer des conclusions. Les savoir-faire relatifs au recueil des faits sont ceux qui ont trait aux pratiques d'investigation – concevoir et mettre en œuvre des situations expérimentales, prendre des mesures et relever des observations à l'aide d'instruments appropriés, etc. Le développement de ces savoir-faire figure parmi les objectifs des programmes scientifiques enseignés à l'école ; ceux-ci visent à permettre aux élèves d'expérimenter et de comprendre la façon dont se construit la connaissance scientifique, et de se familiariser par là, du moins dans l'idéal, avec la nature de la recherche et du savoir scientifiques. En fait, peu d'élèves auront effectivement besoin de ces savoir-faire pratiques après l'école, mais ils auront besoin de maîtriser les concepts dont la compréhension s'acquiert grâce à ces pratiques de recherche et à ces manipulations. Ajoutons qu'il existe de solides raisons de penser que la vision traditionnelle de la « démarche scientifique », selon laquelle des conclusions sont tirées inductivement à partir d'observations – vision qui est encore largement présente dans bon nombre de programmes scientifiques enseignés dans les écoles – est, en réalité, contraire à la manière dont se développe effectivement le savoir scientifique (voir par exemple, Ziman, 1980).

La culture scientifique, telle qu'elle est définie ici, donne la priorité à la capacité de *tirer des conclusions fondées sur des faits*, plutôt qu'à la capacité d'accumuler des faits pour eux-mêmes. La capacité d'établir un lien entre les faits ou les données recueillies et les affirmations et conclusions est considérée comme l'élément central, dont tous les citoyens ont besoin pour former des jugements à propos des aspects de leur vie influencés par la science. Dès lors, tout citoyen doit savoir à quel moment le recours à la connaissance scientifique est pertinent, quelles sont les questions auxquelles la science peut répondre et celles auxquelles elle ne peut pas répondre. Tout citoyen doit être à même de juger si les preuves apportées sont valides, tant en termes de pertinence que par la méthode utilisée pour les recueillir. Et surtout, tout citoyen doit être capable de faire le lien entre les faits et les conclusions qui en découlent, et d'évaluer le poids respectif des éléments qui s'inscrivent en faveur ou en défaveur de tel ou tel programme d'action affectant sa vie à un niveau personnel, social ou global.

Les distinctions qui viennent d'être proposées peuvent se résumer en quelques mots : la priorité doit être donnée aux démarches s'appliquant *aux sciences* par rapport à celles s'appliquant *dans le cadre des sciences*. Il est important que les savoir-faire concernant les démarches cités dans la figure 14 soient d'abord conçus comme concernant les sciences et non comme s'appliquant dans le cadre des sciences. Toutes les démarches décrites dans la figure 14 impliquent une connaissance des concepts scientifiques. Pour les quatre premières démarches, cette connaissance est nécessaire mais pas suffisante puisque ici, c'est la connaissance du recueillement des données et de l'application des faits scientifiques qui est essentielle. Pour la cinquième démarche, la compréhension des concepts scientifiques est la composante la plus importante.

Une description détaillée des démarches scientifiques est proposée ci-dessous. Ces démarches sont explicitées en termes opérationnels à la figure 19.

Figure 14. **Démarches scientifiques sélectionnées**

1. Reconnaître les questions auxquelles on peut répondre par une investigation scientifique.
2. Identifier les faits nécessaires à une investigation scientifique.
3. Tirer ou évaluer des conclusions.
4. Communiquer des conclusions valides.
5. Faire preuve de maîtrise des concepts scientifiques.

Reconnaître les questions auxquelles on peut répondre par une investigation scientifique

Reconnaître les questions auxquelles on peut répondre par une investigation scientifique peut impliquer de pouvoir reconnaître ou formuler une hypothèse qui est (ou qui pourrait être) vérifiée dans une situation donnée ou en utilisant certaines données scientifiques. Il peut aussi s'agir de distinguer les questions ou les idées auxquelles on pourra répondre par des méthodes scientifiques de celles qui ne pourront pas être traitées de cette façon, ou encore de suggérer ouvertement une question qu'on pourra traiter scientifiquement dans une situation donnée.

Identifier les faits nécessaires à une investigation scientifique

Pouvoir identifier les faits/données nécessaires à une investigation scientifique sous-entend de pouvoir identifier ce qui est nécessaire à une vérification valide d'une hypothèse donnée. Cela peut exiger, par exemple, d'identifier ou de reconnaître quelles données devraient être comparées, quelles variables devraient être changées ou gardées, ou quelle décision devrait être prise pour le recueillement de données pertinentes.

Tirer ou évaluer des conclusions

Tirer ou évaluer d'une manière critique des conclusions qui relèvent des données fournies peut impliquer d'avancer une conclusion dérivant de faits ou de données scientifiques ou résultant de procédés alternatifs et choisie parce qu'elle s'applique aux données. Cela peut aussi impliquer de fournir des raisons pour ou contre une conclusion se rapportant aux données, ou souligner les présomptions faites pour aboutir à telle conclusion.

Communiquer des conclusions valides

Pouvoir communiquer à un public déterminé des conclusions valides à partir de faits ou données disponibles implique la production de points pertinents à la situation et aux données fournies ou d'informations supplémentaires, d'une manière appropriée et compréhensible pour l'auditoire en question.

Faire preuve de maîtrise des concepts scientifiques

Montrer une compréhension des concepts scientifiques en appliquant des concepts pertinents à la situation donnée peut impliquer d'expliquer les rapports et les causes possibles aux changements, d'émettre des prédictions en ce qui concerne les conséquences dues à certains changements, ou d'identifier les facteurs influençant un résultat donné en utilisant des idées scientifiques ou des informations venant de l'extérieur.

Dans chacun des cinq savoir-faire, certaines connaissances scientifiques sont requises. Cependant, dans le cas des quatre premiers, les connaissances ne constituent pas le seul critère, puisque le but est d'évaluer les démarches intellectuelles impliquées lors du recueillement, de l'évaluation et de la communication de faits scientifiques valides. Quant au cinquième savoir-faire, c'est la compréhension du concept scientifique impliqué qui est évaluée et celle-ci représente la difficulté principale de l'exercice.

OCDE 1999

Il y a lieu de préciser que, pour chacune des démarches citées ci-dessus, la gamme des difficultés proposées aux élèves est très variable, selon les concepts scientifiques et les situations mis en œuvre. Grâce aux *feedbacks* des divers pays et à l'essai de terrain, la campagne de test définitive OCDE/PISA permettra de sélectionner des items dont le niveau de difficulté est approprié pour des jeunes de 15 ans.

Concepts et contenus

Les concepts sont les larges représentations qui nous rendent capables de donner un sens à de nouvelles expériences en les reliant à ce que nous savons déjà. Les *concepts scientifiques* sont ceux qui nous aident à donner un sens à des aspects du monde naturel et à l'univers de la technologie. Les expressions qui les désignent sont d'ordre très divers, depuis les étiquettes très générales qui leur sont attribuées lorsqu'on les propose comme disciplines scolaires (biologie, physique, sciences naturelles, etc.), jusqu'aux longues listes d'objectifs génériques qu'on rencontre souvent dans les programmes définissant des standards éducatifs ou des curriculums.

Il y a plusieurs manières de grouper les concepts scientifiques pour nous aider à comprendre les aspects scientifiques du monde qui nous entoure. Dans certains cas, les concepts sont des étiquettes précisant la série de caractéristiques qui définissent un groupe particulier d'objets ou d'événements, pouvant en compter des milliers d'exemplaires (« mammifères », « accélération », « solvant »). Les concepts peuvent aussi être formulés sous forme de principes s'appliquant à des phénomènes donnés (les lois ou les théorèmes en physique ou en chimie), qui existent par centaines. Enfin, ils peuvent être formulés sous forme de thèmes scientifiques majeurs (notions larges), qui sont applicables de manière plus générale et plus facile à mettre en œuvre dans le cadre d'une évaluation et de la communication des résultats observés.

Le programme OCDE/PISA utilise quatre critères pour déterminer le choix des concepts scientifiques à évaluer :

– Le premier d'entre eux est la pertinence par rapport à des situations de tous les jours. Les concepts se distinguent par leur degré d'utilité dans la vie quotidienne. Par exemple, bien que la théorie de la relativité fournisse une description plus précise des rapports entre longueur, masse, temps et vitesse que les lois de Newton, ces dernières sont plus utiles dans des aspects de la vie quotidienne touchant aux notions de forces et de mouvements.

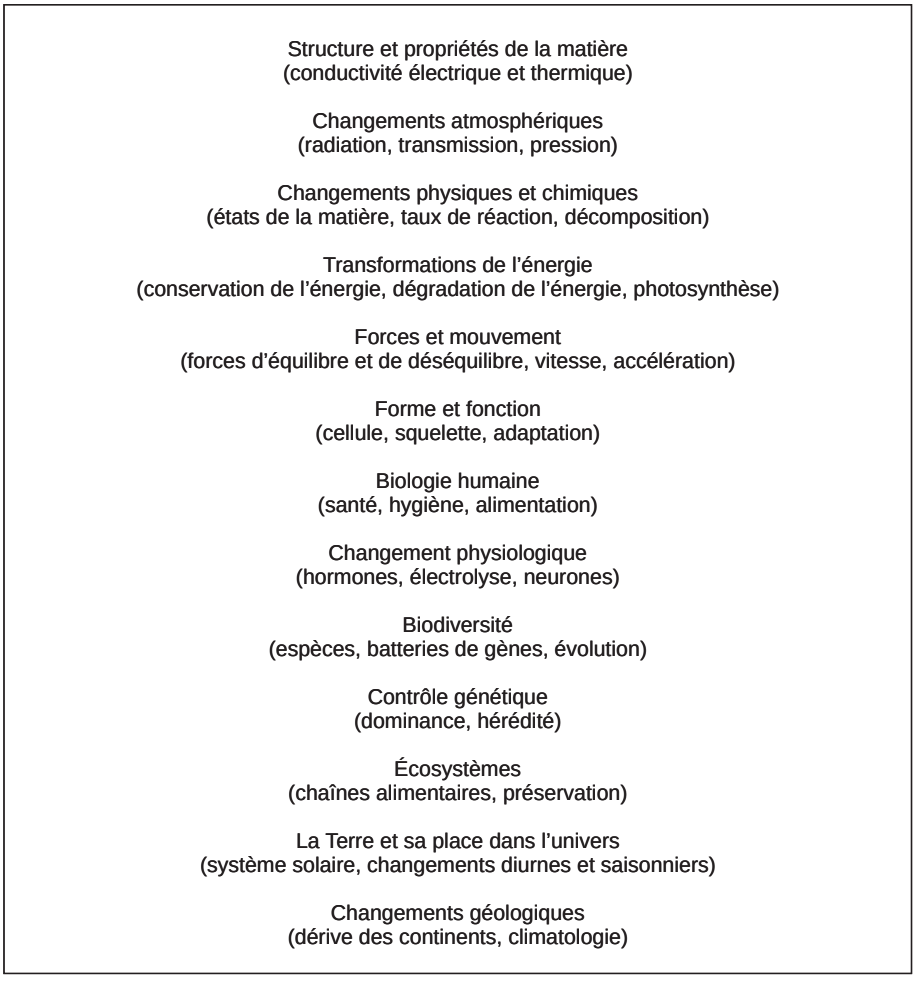
– Le deuxième critère relève de la durée : les concepts et les contenus choisis doivent demeurer pertinents au cours de la prochaine décennie et au-delà. En tenant compte du fait que la campagne de test où l'accent « majeur » sera mis sur les sciences est planifiée pour l'an 2006, le premier cycle de PISA s'intéressera essentiellement aux concepts susceptibles de conserver durablement leur importance dans les sciences et dans les politiques publiques.

– Le troisième critère de choix est la pertinence par rapport à des contextes identifiés comme étant ceux où l'élève doit pouvoir faire preuve de culture scientifique.

– Le quatrième critère est que les concepts retenus puissent être utilisés en combinaison avec les démarches scientifiques qui ont été sélectionnées. Cela ne serait pas le cas dans des épreuves impliquant la simple restitution d'un terme ou d'une définition.

Les figures 15 et 16 présentent les résultats provenant de l'application de ces critères aux concepts et aux contenus scientifiques. La figure 15 comprend la liste des thèmes scientifiques majeurs, avec quelques exemples des concepts s'y rapportant. Ces vastes concepts sont nécessaires pour être à même de comprendre le milieu naturel, pour donner un sens à des nouvelles expériences et pour prendre des décisions les concernant. Leur acquisition dérive de l'étude des phénomènes spécifiques et de faits, mais va au-delà des connaissances détaillées provenant des objets étudiés. La liste des concepts présentés dans la figure 15 renforce la signification des thèmes ; aucune tentative n'est faite de présenter d'une manière complète tous les concepts qui pourraient se rapporter à chacun d'eux.

Figure 15. **Thèmes scientifiques majeurs (et exemples des concepts qui y sont associés) pour l'évaluation de la culture scientifique**



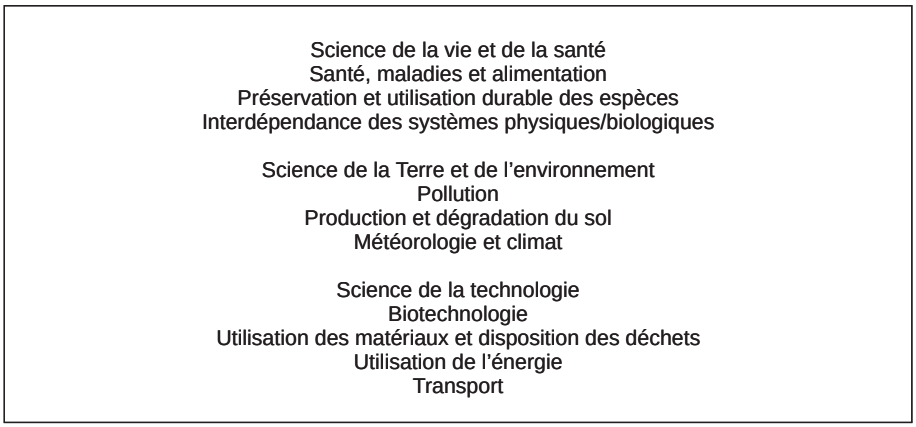
Les concepts donnés en exemples dans la figure 15 indiquent que les connaissances qui seront évaluées se rapportent aux domaines majeurs de la science : physique, chimie, biologie, sciences de la terre et de l'espace. Les items d'épreuves sont classés selon les domaines majeurs de la science ainsi que suivant le thème, le champ d'application et la démarche qu'ils tentent de vérifier.

La figure 16 inventorie les champs d'application de la science susceptibles de soulever des problématiques que le citoyen d'aujourd'hui et de demain aura à comprendre, et à propos desquelles il devra prendre des décisions. Ce sont ces applications qui déterminent le choix des contenus pour les tâches proposées aux élèves et les items qui les accompagnent. La figure 16 présente donc les champs d'application pour lesquels sera évaluée la compréhension des concepts présentés dans la figure 15.

Comme nous l'avons indiqué précédemment, le programme OCDE/PISA évaluera des concepts importants qui se rapportent aux contenus des curriculums de la science des pays participants sans être obligé de prendre pour dénominateur commun les curriculums nationaux. En conformité avec l'accent mis sur la culture scientifique, le programme veillera à ce que les concepts scientifiques sélectionnés soient appliqués et à ce que les démarches scientifiques soient utilisées dans des situations importantes représentant le monde réel et impliquant des idées scientifiques.

OCDE 1999

Figure 16. **Champs d'application de la science pour l'évaluation de la culture scientifique**



Situations

En dehors de la nature des démarches et des concepts évalués, il y a une troisième caractéristique des tâches d'évaluation qui affecte la performance des élèves et qu'on désigne souvent par des expressions comme *contexte* ou *mise en œuvre* des questions. Nous préférons utiliser le mot *situation*, afin d'éviter les confusions que peut entraîner l'emploi de ces termes. On sait que les performances de l'élève peuvent être influencées par les situations particulières dans lesquelles il est interrogé ; il importe donc de déterminer et de contrôler les types de situations utilisées dans les épreuves. Il n'est pas dans nos intentions de présenter les résultats de l'étude en termes de performances liées à des situations particulières ; il y a lieu cependant d'identifier les situations jugées importantes afin de pouvoir contrôler la répartition des tâches proposées aux élèves. Cela permettra, si les résultats des essais de terrain en montrent la nécessité, de garantir que la distribution des tâches à travers les situations reste semblable d'un champ d'étude à l'autre, et, par là, d'assurer une meilleure comparabilité internationale.

En sélectionnant les situations, on gardera à l'esprit que l'objectif de l'évaluation en sciences est de mesurer la capacité des élèves à appliquer dans leur vie de citoyens les savoir-faire et les connaissances qu'ils auront acquis au terme de la scolarité obligatoire. Le programme OCDE/PISA requiert des tâches qu'elles relèvent de la vie en général, et non de la seule vie à l'école. En situation scolaire, il arrive que la mise en œuvre des démarches et des concepts scientifiques soit confinée au laboratoire ou à la salle de classe ; mais, de plus en plus souvent, des efforts sont aussi faits lors de l'élaboration des curriculums scientifiques nationaux, pour en étendre l'application au monde extérieur à l'école.

Les situations du monde réel mettent en jeu des problèmes qui peuvent nous affecter en tant qu'individus (par exemple, l'alimentation, l'utilisation de l'énergie), en tant que membres d'une collectivité locale (par exemple, le traitement des eaux, le choix de l'emplacement d'une centrale électrique), et en tant que citoyens du monde (par exemple, le réchauffement du globe, la diminution de la biodiversité). Tous ces aspects devront figurer dans la série d'épreuves utilisées dans les enquêtes OCDE/PISA sur les sciences. Il y a lieu d'y ajouter une dimension supplémentaire, particulièrement indiquée pour certains contenus : le contexte historique, qui offre la possibilité d'évaluer les représentations de l'élève à propos du progrès des connaissances scientifiques. Dans le plan d'évaluation OCDE/PISA, l'accent sera donc mis sur le développement d'items portant sur des aspects qui concernent l'individu et sa famille (situation personnelle), la collectivité (situation publique), la vie dans le monde en général (situation globale) et sur des situations qui illustrent comment évolue le savoir scientifique et l'impact qu'ont pour la société les décisions impliquant des questions scientifiques (pertinence historique).

S'agissant d'une étude internationale, il importe de sélectionner les situations utilisées dans les épreuves à la lumière de la pertinence par rapport aux préoccupations et aux modes de vie des élèves de tous les pays. Les situations retenues doivent également se prêter à l'évaluation des démarches et des concepts scientifiques. La sensibilité aux différences culturelles est indispensable dans le développement et la sélection des épreuves, non seulement en termes de validité de l'évaluation, mais également pour respecter les différentes valeurs et traditions des pays participants. Les *feedbacks* receuillis lors de l'essai de terrain seront utilisés pour s'assurer que les situations retenues dans les épreuves sont pertinentes et appropriées à l'ensemble des pays participants, tout en présentant la combinaison souhaitée entre connaissances scientifiques et mise en application des démarches scientifiques.

En retenant ce type de situations pour les items proposés aux élèves, PISA s'efforce d'évaluer leur capacité à appliquer des connaissances qu'ils ont très probablement acquises grâce au curriculum de sciences auquel ils ont été exposés (bien qu'une partie de ces connaissances puisse être acquise à travers d'autres matières scolaires ou de sources non scolaires). Toutefois, bien que les connaissances mises en œuvre dans les épreuves relèvent du programme scolaire, il y a lieu de vérifier si celui-ci a offert à l'élève davantage que la connaissance de faits isolés, et s'il a contribué au développement d'une réelle culture scientifique ; c'est ce qui conduit le programme OCDE/PISA à fonder son évaluation sur l'application de ces connaissances dans des situations de la vie réelle. Quelques-uns des exemples d'items présentés ci-dessous servent à illustrer ce point précis.

Caractéristiques des épreuves

Conformément à la définition que PISA donne de la culture scientifique, chaque épreuve, ou « tâche », fera appel à un ou plusieurs savoir-faire représentés dans la figure 14 et à des connaissances scientifiques. Ces tâches sont conçues comme une série de questions (items) portant sur un matériel qui sert de stimulus et qui présente la situation. Certaines des tâches proposées incluront, en plus des items concernant la culture scientifique, des questions qui évaluent la lecture et/ou la culture mathématique.

Exemples d'items mesurant des savoir-faire scientifiques

Quelques exemples d'items envisagés pour l'évaluation de certains de ces savoir-faire aideront à mieux faire comprendre leur signification opérationnelle. Les deux premières démarches sont évaluées dans deux des items d'une épreuve intitulée « Arrêtez ce germe ! ». On demande aux élèves de lire un texte court à propos de l'histoire de l'immunisation, dont est extrait le passage cité dans la figure 17.

Figure 17. **Science : Exemple 1**

Dès le XIe siècle, les médecins chinois manipulaient le système immunitaire. En soufflant dans les narines de leurs patients des croûtes desséchées et pulvérisées prélevées sur des victimes de la variole, ils arrivaient souvent à provoquer une légère poussée de la maladie, qui empêchait une attaque plus virulente. Au XVIIIe, les gens se frottaient la peau avec les croûtes desséchées des boutons pour se protéger de cette maladie. Ces pratiques primitives furent introduites en Angleterre et dans les colonies d'Amérique. En 1771 et 1772, durant une épidémie de variole, un médecin de Boston nommé Zabdiel Boylston griffa la peau de son fils âgé de six ans, ainsi que celle de 285 autres personnes, et frotta ces égratignures avec du pus provenant de boutons de variole. A l'exception de six d'entre eux, tous ses patients survécurent.

Exemple d'item 1 : Quelle hypothèse Zabdiel Boylston était-il ainsi en train de vérifier ?

Exemple d'item 2 : Citez deux autres informations dont vous auriez besoin pour savoir dans quelle mesure la démarche de Boylston a été couronnée de succès.

OCDE 1999

Figure 18. **Science : Exemple 2**

... Une autre façon par laquelle Peter rassemble ses informations est au moyen d'une caméra de télévision placée sur un pylône de 13 mètres de haut pour filmer la circulation routière sur les routes secondaires. Les photographies obtenues informent les chercheurs sur la vitesse à laquelle les véhicules circulent, la distance qui sépare chaque voiture, et les portions de routes que celles-ci utilisent. Plus tard, des lignes sont peintes sur la voie. Les chercheurs peuvent alors utiliser la caméra de télévision pour noter si la circulation est différente une fois que les lignes sont présentes. Est-ce que la circulation va plus vite ou plus lentement ? Est-ce que les voitures sont plus rapprochées ou plus éloignées les unes des autres qu'avant ? Est-ce que les automobilistes conduisent plus près du bord de la chaussée ou plus près du centre maintenant que les lignes sont présentes ? Une fois que Peter possède ces données, il peut recommander de peindre ou de ne pas peindre des lignes sur les routes secondaires.

Exemple d'item 3 : Pour que Peter soit certain qu'il donne de bons conseils, il doit aussi regrouper d'autres informations en plus du filmage des routes secondaires. Parmi ces propositions, lesquelles l'aideraient à prodiguer des conseils sur les effets des lignes peintes sur la circulation sur les voies secondaires ?

a) Répéter ses expériences sur d'autres routes secondaires	Oui/Non
b) Répéter ses expériences sur des routes plus importantes	Oui/Non
c) Estimer les taux d'accidents avant et après que les lignes soient peintes sur les routes	Oui/Non
d) Estimer le nombre de véhicules empruntant la route avant et après la peinture des lignes	Oui/Non

Résultats : Oui *a)* et *c),* Non *b)* et *d)* (2 points)
Oui *a),* Non *b), c)* et *d)* (1 point)
Autres séquences (0 point)

Exemple d'item 4 : Supposez que sur une section de route secondaire Peter observe qu'une fois que les lignes sont peintes, des changements concernant la circulation ont lieu :

Vitesse	La circulation va plus vite
Position	La circulation se concentre vers la chaussée
Distance entre les véhicules	Pas de changement

Sur la base de ces résultats, il a été décidé de peindre des lignes sur les voies de toutes les routes secondaires.

Pensez-vous que ce soit la meilleure décision ?

Donnez vos raisons expliquant votre accord ou désaccord.

D'accord

Pas d'accord

Raisons : _____

Exemple d'item 5 : Les automobilistes sont avisés de laisser plus d'espace entre leur véhicule et celui de devant quand ils conduisent plus vite parce que, à plus grande vitesse, les voitures prennent plus de temps pour s'arrêter.

Expliquer pourquoi une voiture qui va plus vite prend plus de temps pour s'arrêter qu'une voiture qui roule plus lentement.

Raison : _____

Exemple d'item 6 : En examinant son programme télévisé, Peter observe qu'une voiture (A) roulant à 45 km/h est dépassée par une voiture (B) se déplaçant à 60 km/h. A qu'elle vitesse la voiture B semble se déplacer pour le conducteur de la voiture A ?

a) 0 km/h
b) 15 km/h
c) 45 km/h
d) 60 km/h
e) 105 km/h

L'item n° 1 demande une réponse construite, valant 2, 1 ou 0 point(s), selon les détails fournis dans la réponse. Un score de 2 points sera accordé à une hypothèse mentionnant, en gros, que « percer la peau et appliquer du pus directement dans le flux sanguin va accroître les chances de développer une immunité contre la variole ».) Cet item évalue le savoir-faire 1 – *Identifier les questions auxquelles répond une recherche scientifique* – en utilisant des connaissances sur la *biologie* appliquées au domaine des *sciences de la vie et à la santé*.

L'item n° 2 vaut également 2, 1 ou 0 point(s), selon que l'élève mentionne ou non les deux informations requises (taux de survie sans le traitement de Boylston ; ses patients ont-ils été exposés à la maladie en dehors de son traitement ?). Cet item évalue le savoir-faire 2 – *Identifier des faits nécessaires à une recherche scientifique* – en utilisant des connaissances sur la *biologie* appliquées au domaine des *sciences de la vie et à la santé*.

Les quatre items suivants font partie d'une épreuve où le matériel de stimulus est un article à propos de Peter Cairney qui travaille à la Commission australienne pour la recherche des transports routiers (voir figure 18).

L'item n° 3 illustre le savoir-faire 2 – *Identifier les faits nécessaires à une recherche scientifique* – en utilisant la connaissance des *forces et du mouvement* dans le domaine de la *technologie scientifique*.

L'item n° 4 illustre le savoir-faire 3 – *Tirer ou évaluer des conclusions* – en utilisant la connaissance des *forces et du mouvement* dans le domaine de la *technologie scientifique*. Il ne s'agit pas d'approuver ou de désapprouver la décision prise mais plutôt de montrer le raisonnement suivi, en accord avec le choix fait et en se rapportant aux informations procurées (par exemple, être d'accord parce qu'il y a moins de chance de collision si la circulation se concentre près du bord de la chaussée même si le traffic est rapide ; s'il est plus rapide, on est moins encouragé à vouloir doubler ; ou désapprouver, parce que si la circulation est plus rapide et que la distance entre les voitures reste la même, cela peut signifier qu'il n'y a pas assez d'espace pour que celles-ci s'arrêtent en cas d'urgence).

L'item n° 5 illustre le savoir-faire 5 – *La compréhension des concepts scientifiques* – en utilisant la connaissances des *forces et du mouvement* dans le domaine de la *technologie scientifique*. Des réponses structurées sont requises, notées 2, 1 et 0 selon qu'un ou deux des aspects importants à prendre en compte sont mentionnés – *a)* la quantité de mouvement du véhicule qui se déplace plus vite est plus grande, et plus grande est la force requise pour l'arrêter, *b)* à une plus grande vitesse, un véhicule mettra plus de temps à ralentir qu'un véhicule plus lent, dans un même intervalle de temps.

Item n° 6 illustre le savoir-faire 5 – La *compréhension des concepts scientifiques* – en utilisant la connaissances des *forces et du mouvement* dans le domaine de la *technologie scientifique*. C'est un item à choix multiple avec seulement une réponse correcte *b)* à laquelle est attribuée un point.

Pour répondre à ces questions, on demande à l'élève d'utiliser des connaissances acquises grâce à l'enseignement scientifique qu'il a reçu et de les appliquer dans une situation nouvelle. Si l'évaluation et la compréhension de certains concepts n'est pas le but principal de l'item, les connaissances exigées ne constituent pas un obstacle, et la réussite de l'épreuve dépendra donc essentiellement de la capacité particulière d'appliquer la démarche. Si l'évaluation de la compréhension de certains concepts est le but principal, comme dans les exemples d'items 5 et 6, la démarche est de démontrer cette compréhension.

Structure des tests d'évaluation

Comme le montrent ces exemples, chaque épreuve ou « tâche » particulière se présentera sous la forme de plusieurs items accompagnant un matériel de départ servant de stimulus. L'ensemble d'une épreuve mesure plus d'une démarche et plus d'un champ d'application, mais chacun des items pris isolément évalue une des démarches scientifiques présentées dans la figure 14.

Cette structure vise à rendre les tâches proposées aux élèves aussi réalistes que possibles et de représenter, dans une certaine mesure, la complexité des situations de la vie réelle. Une autre raison tient à la nécessité d'utiliser de manière efficiente le temps de passation disponible : vu le temps nécessaire à un élève pour « se mettre » dans une situation, mieux vaut proposer un peu moins de situations permettant de poser plusieurs questions, plutôt que de poser moins de questions concernant un plus grand nombre de situations. Cette approche impose des contraintes techniques qui ont été identifiées et dont il sera tenu compte : il faut faire en sorte que chaque point attribué lors de la correction soit indépendant des autres ; par ailleurs, il importe d'autant plus de minimiser les biais dus à la situation que le nombre de situations proposées est moins élevé.

OCDE 1999

Les épreuves seront relativement longues, certaines d'entre elles comprenant jusqu'à huit items, dont chacun sera corrigé indépendamment des autres. La majeure partie des épreuves (sinon toutes) incluront à la fois des questions faisant appel à la connaissance et à la compréhension des concepts mis en jeu (comme dans les exemples 5 et 6) et d'autres questions, distinctes, nécessitant la mise en œuvre d'un ou de plusieurs savoir-faire de collecte et d'utilisation des faits et données relatifs aux démarches scientifiques, comme dans les exemples d'items 1 à 4. Comme cela a été signalé plus haut, le programme OCDE/PISA n'incluera pas dans les enquêtes des tâches pratiques impliquant des manipulations, du moins pour les cycles des années 2000 et 2003, où la science sera un domaine « mineur ».

D'un point de vue global, l'équilibre souhaité entre les démarches est exprimé dans le tableau 10 en terme de pourcentages de nombres de points. Ces chiffres seront peut-être révisés lors de l'évaluation de 2006 où la science constituera un domaine majeur de l'enquête PISA.

Tableau 10. **Distribution recommandée des points pour chaque démarche scientifique**

Démarches scientifiques	% de points
Reconnaître les questions auxquelles on peut répondre par une investigation scientifique	10-15
Identifier les faits nécessaires à une investigation scientifique	11-20
Tirer ou évaluer des conclusions	12-20
Communiquer des conclusions valides	13-15
Faire preuve de maîtrise des concepts scientifiques	14-50

Il peut arriver qu'en raison du sujet traité, certaines des épreuves soient davantage orientées vers l'évaluation de la compréhension de concepts (savoir-faire 5), et que, dans d'autres cas, ce soit le contraire. Il importe que là où c'est possible, les types d'items évaluant les démarches 1-4 et la démarche 5 soient présents dans chaque tâche, d'une part, pour assurer la « couverture » d'un certain nombre de concepts scientifiques importants que les élèves sont censés avoir acquis grâce au programme scientifique dont ils ont bénéficié à l'école et en dehors et, d'autre part, parce que la capacité d'appliquer des démarches scientifiques dépend fortement de la situation dans laquelle ces savoir-faire sont utilisés (d'où la nécessité d'évaluer les savoir-faire en relation avec une série de concepts). En considérant les objectifs de PISA, il semble clair que la compréhension des concepts, tout comme la combinaison de connaissances scientifiques avec la capacité de tirer des conclusions basées sur des faits, constituent toutes deux des acquis de l'enseignement jugés importants. La répartition recommandée (attribuant un nombre égal de points aux acquis résultant de ces deux apprentissages fondamentaux) s'inscrit dans cette perspective.

Comme on l'a déjà vu, tous les types d'items retenus seront destinés à appliquer les concepts scientifiques qui, selon toute probabilité, auront été acquis par les élèves à l'école, à travers leur programme scientifique. Là ou les items de sciences de PISA se distinguent de ceux utilisés dans d'autres évaluations de l'enseignement des sciences (quoique ce soit loin d'être le cas de toutes), c'est quand ils font appel à l'application de ces concepts dans des situations de la vie réelle. De même, la capacité de tirer des conclusions basées sur des faits en appliquant les savoir-faire décrits ci-dessus fait partie des objectifs visés par bon nombre de programmes de sciences. L'évaluation OCDE/PISA fera appel à l'application d'informations ou de données dans des situations qui dépassent le cadre du laboratoire scolaire ou de la salle de classe. Savoir si cela constituera un élément nouveau pour les élèves dépendra de l'importance accordée par le programme scolaire auquel ils ont été exposés à l'application des concepts et démarches scientifiques dans le monde réel.

Par rapport aux domaines d'application des sciences, le tableau 11 montre que parmi les trois principaux groupes, l'amplitude et l'éventail de pourcentage de points attribués sont les mêmes.

Par rapport à la sélection des situations, PISA aspirera à répartir les items de manière homogène entre les quatre situations suivantes : personnelles, publiques, globales et historiques.

Tableau 11. **Distribution recommandée des points en fonction du champ d'application**

Champs d'application de la science	% de points
Sciences de la vie et de la santé	30-40
Sciences de la terre et de l'environnement	31-40
Sciences de la technologie	32-40

Les situations représentées dans les tâches seront simulées au moyen de matériel de stimulus qui pourrait être un court passage rédigé, accompagné d'un tableau, d'un graphique ou d'un diagramme. Les items seront constitués d'un ensemble de questions corrigées indépendamment les unes des autres et faisant appel aux comportements décrits dans la figure 19. A noter que chacun de ces comportements implique la capacité de comprendre les concepts scientifiques en les appliquant aux divers champs d'évaluation. Les comportements requis se rapporteront aux situations et aux champs d'application présentés dans le matériel servant de stimulus.

Avant qu'aient eu lieu l'essai sur le terrain et l'analyse des résultats, il est difficile d'avoir des certitudes sur certains aspects – par exemple l'équilibre du nombre d'items d'une épreuve à l'autre, le fait que les épreuves portent sur un ou plusieurs champs d'application, ou la répartition des divers formats d'items. Nous pouvons cependant résumer ce qui est envisagé actuellement :

– *Toutes les tâches* seront relativement étendues ; elles incluront des items évaluant la compréhension d'un ou plusieurs concepts scientifiques (figure 14), et la combinaison de savoir-faire (figure 15) et de connaissances relatives à un ou plusieurs champs d'application scientifique (figure 16) ; elles feront appel à des réponses sur papier (écrit ou dessin).

– *La plupart des tâches* seront présentées sur support écrit. Ce sera certainement le cas pour l'enquête qui aura lieu en l'an 2000 ; mais l'usage d'autres formes de stimuli sera mis à l'étude en vue de l'enquête de 2006 où la science sera le domaine majeur.

– *Quelques-unes des tâches* intégreront des items de lecture et/ou de mathématiques et contribueront à l'évaluation de ces domaines. Toutefois, aucun des items proposés concernant la culture scientifique ne se limitera à la répétition d'informations provenant du matériel de stimulus, pas plus qu'il n'y en aura se limitant à la mémorisation d'informations factuelles isolées.

Afin de couvrir la série de savoir-faire et de connaissances décrits dans ce plan d'évaluation, il faut avoir recours à divers types de formats de réponses aux items. Par exemple, on peut proposer des items à choix multiple pour évaluer de manière valide les démarches impliquant une reconnaissance ou une sélection. Toutefois, pour mesurer la capacité d'évaluer ou de communiquer, un format à réponse ouverte conviendra mieux, en termes de validité et d'authenticité. Cela dit, dans de nombreux cas, le format le plus approprié dépendra du contenu particulier de l'item.

Échelles des résultats

Échelles et sous-échelles

Pour répondre aux objectifs du programme OCDE/PISA, le développement d'échelles de compétence des élèves est un élément essentiel. L'élaboration d'une échelle est un processus itératif : les propositions initiales, qui se fondent sur l'expérience acquise au cours des enquêtes de sciences antérieures et sur les résultats des recherches menées dans le domaine de l'apprentissage et du développement cognitif en sciences, sont révisées ensuite en fonction des résultats empiriques fournis par les essais de terrain OCDE/PISA.

Les travaux de recherche disponibles à l'heure actuelle, ainsi que les expériences des enquêtes précédentes suggèrent qu'il sera possible d'établir une échelle de la culture scientifique où le niveau d'acquis des élèves pourra être décrit comme :

OCDE 1999

Figure 19. **Types d'items pour évaluer la culture scientifique**

Ce qui est donné	Comportements requis
Reconnaître les questions auxquelles on peut répondre par une investigation scientifique	
Un exposé d'une investigation ou d'une procédure pour lesquelles des données ont été rassemblées ou des comparaisons établies	Choisir ou formuler une question ou une idée qui peut être (ou pourrait être) vérifiée d'une manière scientifique
Une description d'une situation pour laquelle des questions pourraient être traitées de manière scientifique	Formuler une question qui pourrait être traitée de manière scientifique
Plusieurs questions/hypothèses dérivant ou se rapportant à la situation présentée	Choisir celle(s) qui peut être traitée par une investigation scientifique
Identifier les épreuves nécessaires à une évaluation scientifique	
Une théorie ou une hypothèse émise dans la question ou le matériel de stimulus qui doit être vérifiée	Choisir ou formuler les informations requises pour vérifier l'hypothèse ou pour émettre une prédiction basée sur celle-ci. Les informations peuvent concerner : *a)* Ce qui doit être comparé *b)* Quelles variables devraient être changées ou utilisées en tant que témoin *c)* Quelles informations supplémentaires sont requises *d)* Quelle décision devrait être prise pour permettre de rassembler les données
Tirer ou évaluer des conclusions	
Les données (résultats d'expériences, observations) à partir desquelles des conclusions peuvent être tirées	Formuler des conclusions qui s'appliquent aux données
Les données (résultats d'expériences, observations) et les conclusions déduites des données	Choisir la conclusion qui s'applique aux données et fournir une explication
Les données (résultats d'expériences, observations) et une conclusion déduite des données	Formuler les raisons expliquant pourquoi accepter ou refuser la conclusion ou définir le degré de confiance dans lequel elle pourrait s'insérer
Communiquer des conclusions valides	
Une situation à partir de laquelle des conclusions peuvent être tirées ou qui nécessite de collecter des informations pour soutenir une conclusion ou une recommandation, et un auditoire	Formuler un argument qui est expliqué clairement à l'auditoire présent et qui est soutenu par des faits/données pertinents trouvés dans le matériel de stimulus
Faire preuve de maîtrise des concepts scientifiques	
Une situation dans laquelle une prédiction, une explication ou des informations sont requises	Formuler ou choisir une prédiction, une explication ou des informations supplémentaires basées sur une compréhension d'un concept scientifique ou des renseignements non fournis dans la question ou le matériel de stimulus

allant de la capacité d'appliquer des concepts scientifiques aisés à assimiler, et de mettre en œuvre dans des situations familières, des démarches telles que :

– identifier les questions auxquelles une recherche scientifique permet ou ne permet pas de répondre ;

– identifier l'information dont on a besoin pour vérifier une affirmation ou explorer un problème, dans des situations où il y a une variable à modifier et une à contrôler ;

– expliquer pourquoi certaines conclusions ou affirmations sont erronées, dans des situations où il n'y a pas de contrôle sur une variable qui aurait dû être contrôlée ;

– présenter à autrui les grandes lignes de l'expérience en reliant les faits aux conclusions d'une manière compréhensible ;

– faire des prédictions et proposer des explications à partir de concepts généralement simple à assimiler.

… à la capacité d'appliquer des concepts plus exigeants sur le plan cognitif, et de mettre en œuvre, dans des situations plus complexes, des comportements tels que :

– être conscient du caractère provisoire des connaissances scientifiques et du fait qu'en mettant des théories à l'épreuve, on peut être conduit à les modifier et à améliorer notre compréhension des phénomènes ;

– identifier les informations à recueillir et les conditions dans lesquelles elles doivent être recueillies, pour vérifier une explication ou explorer un problème dans des situations complexes ;

– critiquer la pertinence des informations fournies pour étayer une affirmation ou un argument. Argumenter pour et contre une affirmation ou une conclusion en se référant aux données disponibles, dans des cas où il n'existe pas de relation causale claire ;

– argumenter pour ou contre une conclusion en utilisant des connaissances ou des données scientifiques, ou l'information disponible ;

– faire des prédictions et proposer des explications fondées sur une compréhension de concepts scientifiques plus complexe et abstraits.

Les analyses effectuées sur les résultats des essais de terrain fourniront des indications plus détaillées sur l'échelle de culture scientifique. Ces résultats permettront d'identifier les items que l'on peut regrouper et ceux qui se répartissent entre différents points de l'échelle. Ces données empiriques constitueront un test permettant de vérifier le bien-fondé de la dimension développement proposée ici sur la base de notre jugement et de ce que l'on sait déjà du développement cognitif des élèves.

En 2006, lorsque la durée des tests permettra une couverture plus complète des concepts et des champs d'application scientifiques, il sera probablement possible d'établir, en outre, une sous-échelle portant sur la compréhension des concepts scientifiques (savoir-faire 5), tels qu'ils auront été évalués en les appliquant aux situations proposées. Ce type d'échelle décrira le développement des acquis des élèves allant de la preuve d'une compréhension correcte mais incomplète, portant le plus souvent sur des concepts faciles à assimiler *à* la preuve d'une compréhension plus complète, portant le plus souvent sur des concepts complexes.

Il se peut également qu'on dispose, en 2006, de suffisamment d'informations sur les démarches scientifiques décrites par la figure 14 pour envisager d'établir des sous-échelles relatives à ces diverses démarches. Ici encore, cela dépendra des considérations statistiques, conceptuelles et politiques. S'il s'avère possible de présenter des sous-échelles, les pays auront l'avantage de pouvoir comparer en détail les résultats obtenus par leur enseignement des sciences avec les objectifs atteints.

La présentation du contenu des réponses correctes et des réponses incorrectes aux différents items est un élément d'information important pour accompagner les statistiques sur les items. Ces catégories de réponses correctes seront élaborées sur la base des essais sur le terrain et non pas imposées a priori, sans référence aux types de réponses réellement fournies par les élèves.

OCDE 1999

La présentation de types de réponses aux items spécifiques sera aussi nécessaire pour illustrer l'échelle(s) et les titres appropriés qui lui (leur) sont associés. Cela impliquera de publier un échantillon d'items provenant de ceux utilisés au cours des enquêtes.

D'autres types de présentation des résultats sont souhaitables, et seront possibles après l'enquête majeure en sciences, en 2006. L'un de ces modes de présentation consisterait à présenter les performances des élèves en regroupant les items proposés dans des épreuves différentes pour mesurer les divers champs d'application de la science. Ce type d'information serait précieux pour déterminer si telle ou telle problématique importante dans le monde d'aujourd'hui a reçu une attention suffisante et effective.

Autres questions

L'empiètement des domaines

Lorsque l'information destinée à une épreuve sur la culture scientifique prend la forme d'un texte relativement long, il est possible de procéder à l'évaluation de certains aspects de la lecture. De même, lorsque l'information se présente sous la forme de tableaux, de diagrammes, de graphiques, etc., il sera possible d'évaluer la capacité de lire ce type d'information ; là où une manipulation de chiffres est requise, on pourra évaluer certains aspects relevant des mathématiques. Les épreuves de ce type feront partie de la section « intégrée » du test. D'autres épreuves seront destinées à la seule compréhension de concepts scientifiques et à la combinaison de démarches et de connaissances scientifiques.

Pour les mêmes raisons que celles exposées dans le chapitre sur la « culture mathématique », il ne sera pas possible de faire de lien psychométrique entre OCDE/PISA et IEA/TIMSS.

Enquêtes scientifiques majeures et mineures

Les enquêtes de sciences des années 2000 et 2003, au cours desquelles la science sera un domaine « mineur », sont destinées à servir de base à l'établissement de séries temporelles. Le nombre d'épreuves sera limité pour ces deux années (même si la conception de l'enquête autorise la rotation de différents ensembles d'items auprès de sous-échantillons d'élèves), ce qui implique qu'il y aura moins d'épreuves concernant chacun des champs d'application des sciences qu'en l'an 2006. Dès lors, les enquêtes mineures en sciences porteront sur tous les savoir-faire décrits à la figure 14, mais sur une partie seulement de ceux et des champs d'application inventoriés par les figures 15 et 16. En 2006, année de l'enquête majeure en sciences, une plus grande couverture des savoir-faire et des champs d'application sera possible dans l'évaluation.

BIBLIOGRAPHIE

ALMOND, R.G. et MISLEVY, R.J. (1998),
Graphical Models and Computerized Adaptive Testing, TOEFL Technical Report n° 14, Educational Testing Service, Princeton, New Jersey, mars.

BAKER, L. (1991),
« Metacognition, reading and science education », in C. M. Santa et D. E. Alvermann (dir. pub.), *Science Learning : Processes and Applications*, International Reading Association, Newark, DE, pp. 2-13.

BENNETT, R.E. (1993),
« On the meanings of constructed response », in R.E. Bennett (dir. pub.), *Construction vs. Choice in Cognitive Measurement : Issues in Constructed Response, Performance Testing, and Portfolio Assessment,* Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 1-27.

BINKLEY, M. et LINNAKYLÄ, P. (1997),
« Teaching reading in the United States and Finland », in M. Binkley, K. Rust et T. Williams (dir. pub.), *Reading Literacy in an International Perspective,* U.S. Department of Education, Washington, DC.

BRUNER, J. (1990),
Acts of Meaning, Harvard University Press, Cambridge, MA.

BYBEE, R. W. (1997),
« Towards an understanding of scientific literacy », in W. Graber et C. Bolte (dir. pub.), Scientific Literacy – An *International Symposium,* IPN, Kiel.

Conseil de l'Europe (1996),
Modern Languages: Learning, Teaching, Assessment. A Common European Framework of Reference, CC LANG (95) 5 rev. IV, Strasbourg.

Conseil des ministres de l'Éducation, Canada (1994),
Student Achievement Indicators Program: Reading and Writing, Toronto.

de LANGE, J. et VERHAGE, H. (1992),
Data Visualization, Sunburst, Pleasantville, NY.

de LANGE, J. (1987),
Mathematics, Insight and Meaning, OW et OC, Utrecht.

DEVLIN, K. (1994, 1997),
Mathematics: The Science of Patterns, Scientific American Library, New York.

DOLE, J., DUFFY, G., ROEHLER, L. et PEARSON, P. (1991),
« Moving from the old to the new: Research on reading comprehension instruction », *Review of Educational Research,* 16 (2), pp. 239-264.

EHRLICH, M.F. (1996),
« Metacognitive monitoring in the processing of anaphoric devices in skilled and less-skilled comprehenders », in C. Cornoldi et J. Oakhill (dir. pub.), *Reading Comprehension Difficulties: Processes and Interventions,* Lawrence Erlbaum Associates, Mahwah, NJ, pp. 221-249.

EHRLICH, M. F., KURTZ-COSTES, B. et LORIDANT, C. (1993),
« Cognitive and motivational determinants of reading comprehension in good and poor readers », *Journal of Reading Behavior, 25,* pp. 365-381.

EINSTEIN, A. (1933),
« Preface to M. Plank », *Where is Science Going ?,* Allen and Unwin, Londres.

OCDE 1999

ELLEY, W.B. (1992) ;
How in the World do Students Read ?, International Association for the Evaluation of Educational Achievement, La Haye.

FREDERICKSON, N. (1984),
« The real test bias », *American Psychologist, 39,* pp. 193-202.

FREUDENTHAL, H. (1973),
Mathematics as an Educational Task, Reidel, Dordrecht.

FREUDENTHAL, H. (1983),
Didactical Phenomenology of Mathematical Structures, Reidel, Dordrecht.

GRAEBER, W. et BOLTE, C. (dir. pub.) (1997),
Scientific Literacy – An International Symposium, IPN, Kiel.

GRONLUND, N.E. (1968),
Constructing Achievement Tests, Prentice Hall, Englewood Cliffs.

GRÜNBAUM, B. (1985),
« Geometry strikes again », *Mathematics Magazine,* 58 (1), pp 12-18.

HAWKING, S.W. (1988),
A Brief History of Time, Bantam Press, Londres.

HUBBARD, R. (1989),
« Notes from the underground : Unofficial literacy in one sixth grade », *Anthropology and Education Quarterly, 20,* pp. 291-307.

JONES, S. (1995),
« Application des capacités de lecture et d'écriture », in *Littératie, Économie et Société, – Résultats de la première enquête internationale sur l'alphabétisation des adultes,* OCDE et Statistique Canada, Paris et Ottawa, pp. 101-129.

KIRSCH, I. (1995),
« Rendement lié à la compréhension de documents d'information variés selon trois échelles : définitions et résultats », dans *Littératie, Économie et Société – Résultats de la première enquête internationale sur l'alphabétisation des adultes,* OCDE et Statistique Canada, Paris et Ottawa, pp. 33-64.

KIRSCH, I.S. et MOSENTHAL, P.B. (1989-1991),
« Understanding documents. A monthly column », *Journal of Reading,* International Reading Association, Newark, DE.

KIRSCH, I. S. et MOSENTHAL, P. B. (1994),
« Interpreting the IEA reading literacy scales », in M. Binkley, K. Rust et M. Winglee (dir. pub.), *Methodological Issues in Comparative Educational Studies: The Case of the IEA Reading Literacy Study,* US Department of Education, National Center for Education Statistics, Washington, DC, pp. 135-192.

KIRSCH, I., JUNGEBLUT, A. et MOSENTHAL, P. B. (1998),
« The measurement of adult literacy », in T. S. Murray, I. S. Kirsch, et L. Jenkins (dir. pub.), *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey,* US Department of Education, National Center for Education Statistics, Washington, DC.

LANGER, J. (1995),
Envisioning Literature, International Reading Association, Newark, DE.

LINNAKYLÄ, P. (1992),
« Recent trends in reading literacy research in Finland », in P. Belanger, C. Winter et A. Sutton (dir. pub.), *Literacy and Basic Education in Europe on the Eve of the 21st Century,* Council of Europe, Strasbourg, pp. 129-135.

LUNDBERG, I. (1991),
« Reading as an individual and social skill », in I. Lundberg et T. Hoien (dir. pub.), *Literacy in a World of Change,* Center for Reading Research/UNESCO, Stavanger.

MACCARTHEY, S.J. et RAPHAEL, T.E. (1989),
Alternative Perspectives of Reading/Writing Connections, College for Education, Institute for Research on Teaching. Occasional paper #130, Michigan State University.

MILLAR, R. et OSBORNE, J. (1998),
Beyond 2000: Science Education for the Future, King's College London School of Education, Londres.

MYERS, M. et PARIS, S.G. (1978),
« Children's metacognitive knowledge about reading », *Journal of Educational Psychology, 70,* pp. 680-690.

PARIS, S., WASIK, B. et TURNER, J. (1991),
« The development of strategic readers », in R. Barr, M. Kamil et P. Mosenthal (dir. pub.), *Handbook of Reading Research,* vol. II, Longman, New York.

SENECHAL, M. (1990),
« Shape », in L.A. Steen (dir. pub.), *On the Shoulders of the Giant – New Approaches to Numeracy,* National Academy Press, Washington DC, pp 139-182.

SHAFER, M.C. et ROMBERG, T. A (à paraître),
« Assessment in Classroom that Promote Understanding », in E. Fennema et T.A. Romberg (dir. pub.), *Mathematics Classrooms that Promote Understanding,* Erlbaum, Mahwah, New Jersey.

SHAMOS, M.H. (1995),
The Myth of Scientific Literacy, Rutgers University Press, New Brunswick.

SMITH, M.C. (1996),
« Differences in adults'reading practices and literacy proficiencies », *Reading Research Quarterly, 31,* pp.196-219.

STICHT, T.G. (dir. pub.) (1975),
Reading for Working: A Functional Literacy Anthology, Human Resources Research Organization, Alexandria, VA.

STIGGINS, R.J. (1982),
« An analysis of the dimensions of job-related reading », *Reading World, 82,* pp. 237-247.

STREEFLAND. L. (1990),
Fractions in Realistic Mathematics Education, A Paradigm of Developmental Research, Reidel Dordrecht.

STUART, I. (1990),
« Change », in L.A. Steen (dir. pub.), *On the Shoulders of the Giant – New Approaches to Numeracy,* National Academy Press, Washington D.C., pp. 183-218.

TAUBE, K. et MEJDING, J. (1997),
« A nine-country study: What were the differences between the low and high performing students in the IEA Reading Literacy Study ? », in M. Binkley, K. Rust et T. Williams (dir. pub.), *Reading Literacy in the International Perspectives,* U.S. Department of Education, National Center for Education Statistics, Washington, DC, pp. 63-100.

TRAUB, R.E. (1993),
« On the equivalence of the traits assessed by multiple-choice and constructed-response tests », in R.E. Bennett (dir. pub.), *Construction vs. Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment,* Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 29-44.

TRAVERS, K.J. et WESTBURY, I. (1989),
The IEA Study of Mathematics, Vol. 1, Analysis of mathematics curricula, Pergamon Press, Oxford.

TREFFERS, A. (1986),
Three Dimensions, Reidel, Dordrecht.

TREFFERS, A. et GOFFREE, F. (1985),
« Rational analysis of realistic mathematics education », in L. Streefland (dir. pub.), *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education (PME),* OW et OC, Utrecht, pp. 79-122.

UNESCO (1993*),*
International Forum on Scientific and Technological Literacy for All, Final Report, UNESCO, Paris.

WARD, W.C., DUPREE, D. et CARLSON, S.B. (1987),
A Comparison of Free-response and Multiple-choice Questions in the Assessment of Reading Comprehension (RR-87-20), Educational Testing Service, Princeton, NJ.

WERLICH, E. (1976),
A Text Grammar of English, Quelle and Meyer, Heidelberg.

ZIMAN, J.M. (1980),
Teaching and Learning about Science and Society, Cambridge University Press.

OCDE 1999

Annexe 1

MEMBRES DU GROUPE FONCTIONNEL D'EXPERTS

Lecture

Irwin Kirsch (président)
Educational Testing Service
Princeton, New Jersey, États-Unis

Marilyn Binkley
National Center for Education Statistics
Washington, DC, États-Unis

Alan Davies
University of Edinburgh
Écosse, Royaume-Uni

Stan Jones
Statistique Canada
Nova Scotia, Canada

John de Jong
CITO, National Institute for Educational Measurement
Arnhem, Pays-Bas

Dominique Lafontaine
Université de Liège
Liège, Belgique

Pirjo Linnakylä
University of Jyvaskyla
Jyvaskyla, Finlande

Martine Rémond
Institut National de Recherche Pédagogique
Paris, France

Wolfgang Schneider
University of Würzburg
Würzburg, Allemagne

Ryo Watanabe
National Institute for Educational Research
Tokyo, Japon

Mathématiques

Jan de Lange (président)
Utrecht University
Utrecht, Pays-Bas

Raimondo Bolletta
Centro Europeo dell'Educazione
Frascati, Italie

OCDE 1999

Sean Close
St Patricks College
Dublin, Irlande

Maria Luisa Moreno
Instituto Nacional de Calidad y Evaluación (INCE)
Madrid, Espagne

Mogens Niss
Roskilde University
Roskilde, Danemark

Kyung Mee Park
Korea Institute of Curriculum and Evaluation
Séoul, Corée

Thomas Romberg
University of Wisconsin-Madison
Madison, Wisconsin, États-Unis

Peter Schüller
Federal Ministry of Education and Cultural Affairs
Vienne, Autriche

Sciences

Wynne Harlen (Président)
The Scottish Council for Research in Education
Edinburgh, Royaume-Uni

Peter Fensham
Monash University
Melbourne, Australie

Raul Gagliardi
Genève, Suisse

Donghee Shin
Korea Institute of Curriculum and Evaluation
Séoul, Corée

Svein Lie
University of Oslo
Oslo, Norvège

Manfred Prenzel
Institut für die Pädagogik der Naturwissenschaften
(Institut des sciences de l'éducation – IPN)
Kiel, Allemagne

Senta Raizen
National Center for Improving Science Education
Washington, DC, États-Unis

Elizabeth Stage
University of California
Oakland, Californie, États-Unis

Annexe 2

QUESTIONS POUR LES CYCLES A VENIR DU PROGRAMME OCDE/PISA

Exposition à l'apprentissage et mesure des attitudes des élèves pertinentes pour l'évaluation de la culture scientifique

Au-delà des acquis cognitifs des élèves produits par l'enseignement des sciences, il convient de recueillir des informations sur le développement affectif qu'il induit. Dans le cadre du programme OCDE/PISA, il sera pertinent de relever divers types d'informations sur la participation des élèves à des activités à contenu scientifique et sur leur opinion quant à l'utilité des activités scientifiques pour eux-mêmes et pour la société en général. Par ailleurs, comme il se peut que les tâches utilisées dans cette enquête soient nouvelles pour certains élèves, il sera important de s'informer sur leur familiarité avec le contenu et la forme des tâches et des items. Dans certains cas, ces informations peuvent fournir des variables explicatives ; dans d'autres cas, elles fourniront des données complémentaires importantes à propos des effets produits par l'enseignement des sciences.

Dans les cycles à venir, on prévoit, soit à travers le questionnaire sur les élèves ou au sein du matériel de test, de recueillir des informations sur :

- La participation des élèves à des activités scientifiques à l'école ou en dehors de l'école (par exemple en dehors du programme scolaire – lire des magazines scientifiques, regarder des émissions télévisées scientifiques, s'impliquer dans des mouvements communautaires à visée environnementale).
- Les opinions des élèves sur l'utilité des notions scientifiques qu'ils ont acquises à l'école et en dehors de l'école, pour des décisions à prendre au niveau personnel et de la collectivité.
- Le jugement que les élèves portent sur le rôle de la science en tant que facteur pouvant créer ou aider à résoudre des problèmes.
- L'exposition à l'apprentissage dont les élèves ont bénéficié à propos des contenus spécifiques abordés par les épreuves d'évaluation.
- La familiarité des élèves avec le format des tâches et des items.

Ces informations seront à recueillir au moyen du questionnaire sur les élèves, à l'exception de celles relevant de « l'exposition à l'apprentissage » dont les élèves ont bénéficié à propos des contenus spécifiques abordés par les épreuves d'évaluation. Cette variable fera l'objet d'une dernière question posée lors du test des sciences. La question proposera la liste des items du test et demandera aux élèves d'indiquer pour chacun des items s'ils ont l'occasion de discuter à propos du contenu abordé dans la question, à l'école d'une part, et en situation non scolaire d'autre part.

Des formats alternatifs de l'item en mathématiques

Quand les mathématiques constitueront un domaine majeur de PISA en 2003, il sera possible d'étendre le champ des épreuves d'évaluation qui peuvent être utilisées. Cela sera particulièrement important si les programmes futurs mettent l'accent sur l'évaluation des compétences de classe 3.

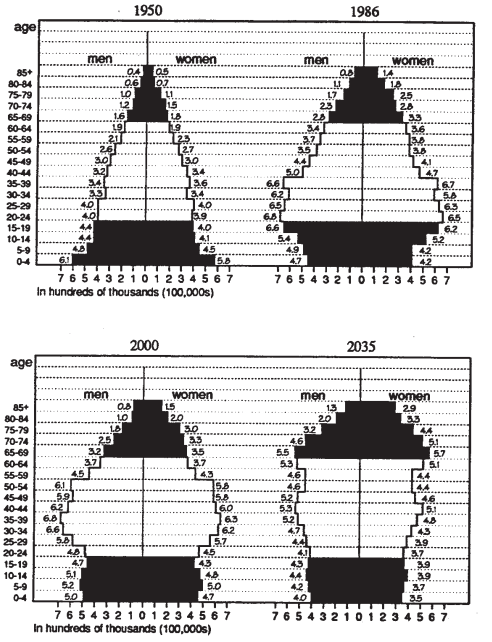
Étendre le champ des épreuves d'évaluation

Il est possible de faire appel à un test de dissertation pour étendre le champ des épreuves d'évaluation. Selon Gronlund (1968), les épreuves faisant appel à la dissertation sont inefficaces pour mesurer les connaissances acquises, mais elles offrent une liberté de réponse, nécessaire à l'évaluation des acquis complexes. Ces acquis complexes réfèrent à la capacité de créer, d'organiser, d'intégrer, d'exprimer, ainsi qu'à d'autres types de comportements semblables qui font appel à la production et la synthèse des idées.

OCDE 1999

Figure 20. **Une épreuve de « dissertation » mathématique**

Les quatre pyramides représentent la population des Pays-Bas.
Les deux premières représentent la population en 1950 et en 1986. Les deux dernières représentent la population prévue pour les années 2000 et 2035.



Les informations fournies à partir de ces pyramides de la population suggèrent-elles un vieillissement de la population des Pays-Bas ? Ayez recours à d'autres représentations graphiques pour appuyer votre raisonnement.

L'extension du champ des épreuves donne aux élèves la possibilité de s'impliquer plus à fond dans une problématique complexe. Contrairement à une série d'items liés à une source commune, ces épreuves permettent aux élèves de formuler leur réponse personnelle tout en étudiant un problème. C'est un cadre particulièrement approprié pour évaluer les compétences de classe 3.

La figure 20 donne un exemple d'une telle épreuve.

Les épreuves orales

Dans certains pays, l'épreuve orale est ou a été une pratique courante, faisant même partie du système officiel national d'examens. Différentes formes d'épreuves orales existent, dont nous citons trois exemples ci-dessous :

– une discussion orale sur certains sujets mathématiques connus des élèves ;

– une discussion orale sur un sujet, comprenant une épreuve à faire à la maison, remise aux élèves avant la discussion ; et

– une discussion orale sur une épreuve à faire à la maison, après que celle-ci a été réalisée par les élèves.

Le format de l'épreuve orale est souvent utilisé dans la mise en oeuvre de buts plus ambitieux.

Les épreuves à deux étapes

Une combinaison de différents formats d'examen peut être considérée comme une épreuve se faisant en deux étapes. Une épreuve écrite suivi d'une épreuve orale sur le même sujet en est un exemple classique. Les épreuves à deux étapes combinent les avantages d'épreuves écrites traditionnelles en temps limité avec les possibilités qu'offrent des tâches plus ouvertes.

Les caractéristiques suivantes sont celles d'une épreuve à deux étapes. La première étape comprend une épreuve écrite ayant les caractéristiques suivantes :

– tous les élèves passent la même épreuve en même temps ;
– tous les élèves doivent finir l'épreuve dans la même limite de temps ;
– l'épreuve a pour but d'identifier ce que les élèves ne connaissent pas plutôt que ce qu'ils savent ;
– la question d'examen fait généralement appel à des tâches « simples » : calcul et compréhension ;
– les questions d'examen sont du type ouvert ;
– les notes sont aussi objectives que possible.

La seconde étape compense ce qui a manqué à la première :

– l'épreuve se fait sans aucune limite de temps ;
– l'épreuve peut se faire chez soi ;
– l'épreuve met l'accent sur ce que l'élève connaît plutôt que sur ce qu'il ignore ;
– les tâches plus « importantes » sont privilégiées : interprétation, raisonnement, communication, etc. ;
– la structure de l'épreuve est plus souple, consistant en des questions faisant appel à des réponses longues, dont des dissertations ;
– la notation peut être difficile et moins objective.

Les items de production

Les procédures d'évaluation devraient offrir aux élèves la possibilité de démontrer leurs compétences, et devraient être perçues en tant que partie intégrante du savoir-faire de l'apprentissage-enseignement. L'utilisation des *productions personnelles* n'est pas nouvelle, et beaucoup d'expériences de ce genre d'évaluation ont été faites. Treffers (1987) a introduit une différence entre les notions d'*interprétation* et de *production*. La production libre est celle qui offre le plus de possibilités en terme de liberté afin que les élèves puissent démontrer leurs raisonnement personnel et leurs compétences. Il peut s'agir de :

– résoudre des problèmes relativement étendus qui demandent de produire des réponses divergentes en raison de la grande variété de solutions qui peuvent s'y appliquer, souvent à des niveaux différents de mathématisation ;
– résoudre des problèmes incomplets, qui, avant qu'ils puissent être résolus, demandent aux élèves de chercher et d'obtenir des données et une bibliographie ;
– s'inventer soi-même des problèmes (faciles, moyens, difficiles) sous forme de copie d'examen ou de recueil de problèmes concernant un thème ou un sujet destinés à la prochaine promotion d'élèves (Streefland, 1990).

Ces possibilités de format, ainsi que d'autres, devraient être explorées plus avant afin d'être utilisées dans le cadre de PISA dans les années à venir.

OCDE 1999

LES ÉDITIONS DE L'OCDE, 2, rue André-Pascal, 75775 PARIS CEDEX 16
IMPRIMÉ EN FRANCE
(96 1999 05 2 P) ISBN 92-64-27053-1 – n° 50618 1999