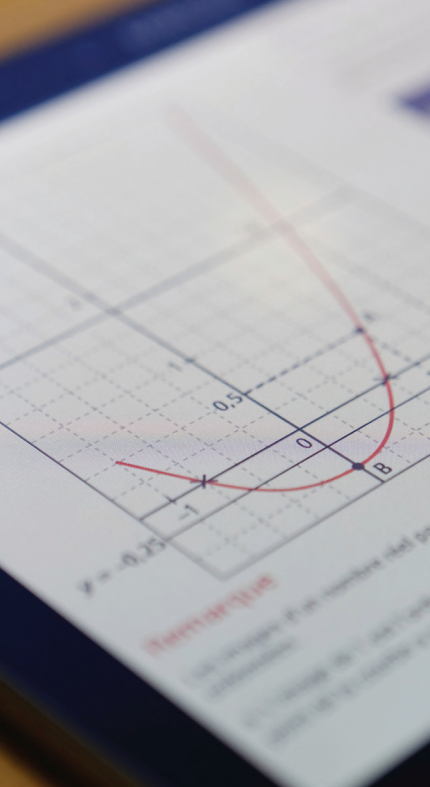




RAPPORT SCIENTIFIQUE

Nov. 2016

COMPARAISON DES ÉVALUATIONS **PISA ET TIMSS**

ACQUIS DES ÉLÈVES : COMPRENDRE LES
ÉVALUATIONS INTERNATIONALES

VOLUME 1



 **cnesco**
 **conseil national**
 **d'évaluation**
 **du système scolaire**

Acquis des élèves : Comprendre les évaluations internationales PISA et TIMSS

Analyse comparative des contenus en mathématiques et en sciences

Volume 1 : Analyse comparative des cadres de référence des deux enquêtes



Novembre 2016

Ce rapport est édité par le Conseil national d'évaluation du système scolaire.

Directrice de la publication

Nathalie Mons, présidente du Cnesco

Rédacteur en chef

Jean-François Chesné, directeur scientifique du Cnesco

Auteurs

Antoine Bodin, IREM d'Aix-Marseille (Institut de recherche sur l'enseignement des mathématiques)

Cécile de Hosson, professeure à l'université Paris Diderot – Paris 7

Nicolas Décamp, maître de conférences à l'université Paris Diderot – Paris 7

Nadine Grapin, maître de conférences à l'université Paris Est-Créteil

Pierre Vrignaud, professeur émérite de psychologie à l'université Paris-Ouest Nanterre – La Défense

Remerciements des auteurs

Ce travail a bénéficié de la collaboration de Franck Salles, coordinateur national de recherche pour TIMSS *Advanced* 2015, et co-manager du projet PISA 2012 pour la France, et de Youssef Souidi, chargé de mission au Cnesco.

Remerciements du Cnesco

Le Cnesco remercie l'OCDE et l'IEA pour l'avoir autorisé à accéder respectivement aux exercices des enquêtes de PISA et de TIMSS.

Le Cnesco remercie la Depp pour avoir mis à sa disposition les tests 2015 des enquêtes PISA sciences, TIMSS CM1 et TIMSS *Advanced*.

Table des matières

Table des matières	3
Liste des tableaux	5
Liste des figures	7
Préambule	9
Introduction.....	11
Chapitre 1 : Présentation générale des enquêtes PISA et TIMSS.....	13
A. L'IEA et les enquêtes sur l'enseignement des mathématiques et des sciences	14
B. L'OCDE et les enquêtes PISA	14
C. Objectifs généraux et évolution des objectifs dans le temps	15
1. Les objectifs de TIMSS.....	15
2. Les objectifs de PISA	16
D. Préparation et organisation des enquêtes.....	16
1. Méthodes psychométriques communes à PISA et à TIMSS	16
2. Spécificités des enquêtes PISA.....	21
3. Spécificités des enquêtes TIMSS.....	23
4. Rôle et place du numérique dans les enquêtes	23
Chapitre 2 : Le cadre de référence des enquêtes PISA.....	27
A. Organisation générale du cadre de référence	28
B. Le cadre de référence de PISA pour le volet mathématique	29
1. Processus et compétences.....	30
2. Les groupes de compétences (PISA 2000 à PISA 2009)	33
3. Mathématisation et modélisation	35
4. Les domaines de contenus.....	39
5. Les contextes.....	41
6. Relation des autres volets de PISA avec les mathématiques.....	44
C. Le cadre de référence de PISA en littératie scientifique dans PISA 2015	46
1. Contextualisation des questions de science	47
2. Une nouveauté : les niveaux de charge cognitive	47
Chapitre 3 : Le cadre de référence des enquêtes TIMSS	51
A. Organisation générale du cadre de référence de TIMSS	52
B. Le cadre de référence de TIMSS pour les mathématiques	54

1.	Les domaines de contenus.....	54
2.	Les domaines cognitifs.....	55
3.	Tableaux synthétiques des plans d'évaluation en mathématiques de TIMSS 2015.....	56
4.	Distribution des questions de TIMSS 2015 selon les domaines et les formats	57
5.	Le cadre de référence pour la physique (TIMSS <i>Advanced</i> 2015)	60
Chapitre 4 : Structure et format des exercices d'évaluation de PISA et de TIMSS en mathématiques.		61
A.	Aspects communs à PISA et à TIMSS	62
B.	Les types de questions de PISA et de TIMSS.....	64
Chapitre 5 : La présentation des résultats des enquêtes PISA et TIMSS		67
A.	Lecture des résultats et des échelles utilisées	68
B.	Comparaisons temporelles	71
C.	Niveaux des élèves et valeurs plausibles	74
D.	Comparaison des résultats aux enquêtes	74
Synthèse : Comparaison des programmes PISA et TIMSS.....		79
Références et bibliographie.....		83
Liste des sigles et acronymes		87

Liste des tableaux

Tableau 1 : Place des différents domaines cognitifs de PISA au cours des enquêtes	22
Tableau 2 : Processus, compétences et aptitudes fondamentales dans PISA (2000-2015).....	32
Tableau 3 : Tableau synthétique des catégories de PISA 2015 en littératie mathématique	41
Tableau 4 : Répartition des questions en fonction des processus dans PISA 2012	42
Tableau 5 : Répartition des questions en fonction des contextes dans PISA 2012.....	43
Tableau 6 : Répartition des formats de questions en fonction des processus dans PISA 2012	43
Tableau 7 : Plan d'évaluation de TIMSS 2015 - Mathématiques quatrième et huitième année scolaire	56
Tableau 8 : Plan d'évaluation de TIMSSADV 2015 - Mathématiques avancées.....	57
Tableau 9 : Contenus et formats des questions de TIMSS4 2015 (CM1) selon les domaines.....	58
Tableau 10 : Contenus et formats des questions de TIMSSADV 2015 selon les domaines	58
Tableau 11 : Domaines cognitifs des questions de TIMSS4 2015 (CM1) selon les processus.....	59
Tableau 12 : Domaines cognitifs des questions de TIMSSADV 2015 selon les processus.....	59
Tableau 13 : Pourcentages des domaines thématiques et cognitifs de l'enquête TIMSS <i>Advanced</i> 2015 en physique	60
Tableau 14 : « Scores » en mathématiques aux enquêtes PISA en France, en Finlande et en Allemagne	68
Tableau 15 : Exemple de présentation des résultats de PISA 2012	69
Tableau 16 : Évolution du taux de réponses correctes à la question « Conduites » de l'enquête PISA au niveau de la France et de l'OCDE.....	72
Tableau 17 : Évolution du taux de réponses correctes à la question « Récipients » de l'enquête PISA au niveau de la France et de l'OCDE.....	73
Tableau 18 : Corrélations entre les résultats des élèves dans les différents domaines de PISA 2012	74
Tableau 19 : Corrélations entre les scores des pays dans les différents domaines de PISA 2012	75
Tableau 20 : Différences entre les programmes PISA et TIMSS.....	80

Liste des figures

Figure 1 : Les relations entre niveaux de compétence des élèves et difficulté des questions (<i>items</i>)	19
Figure 2 : La Fleur des compétences – Projet KOM (1999)	31
Figure 3 : Illustration du poids relatif de chacune des compétences dans les groupes de compétences	34
Figure 4 : Le cycle de mathématisation avant 2012	36
Figure 5 : Le cycle de modélisation à partir de 2012	37
Figure 6 : Un modèle de la littératie mathématique en pratique	38
Figure 7 : Grille permettant de classer les questions en fonction des connaissances, des compétences et de la charge cognitive mobilisée	48
Figure 8 : Différentes entrées pour l'étude d'un curriculum selon l'IEA.....	52
Figure 9 : Illustration du format habituel des exercices de PISA (contenu fictif)	62
Figure 10 : Résultats en mathématiques en 4 ^e année de TIMSS 2011. Moyenne et distribution pour chaque pays.....	70
Figure 11 : Résultats de PISA 2012 en culture scientifique avec la représentation de l'erreur de mesure	71
Figure 12 : Corrélation entre les scores de littératie scientifique et de littératie mathématique à PISA 2012	75
Figure 13 : Corrélation entre les scores en compréhension de l'écrit et de littératie mathématique à PISA 2012	75
Figure 14 : Résultats en mathématiques des pays ayant participé à PISA et à TIMSS8 en 2003	76
Figure 15 : Résultats en sciences des pays ayant participé à PISA et à TIMSS8 en 2003	77
Figure 16 : Résultats en mathématiques des pays ayant participé à PISA 2012 et à TIMSS8 2011	78
Figure 17 : Résultats en sciences des pays ayant participé à PISA 2012 et à TIMSS8 2011	78

Préambule

Ce rapport du Cnesco est consacré aux programmes internationaux PISA (*Programme for International Student Assessment*) et TIMSS (*Trends in International Mathematics and Science Study*) d'évaluation des acquis des élèves. Il paraît à quelques semaines de la publication des résultats des enquêtes 2015. Alors que PISA est sans doute le plus important programme international mis en place pour comparer les systèmes éducatifs et le plus commenté dans le débat public, TIMSS est moins connu. La France n'avait pas participé à ce programme, consacré spécifiquement aux mathématiques et aux sciences, depuis 20 ans. Les résultats de l'édition de TIMSS 2015 devraient donc être largement commentés, d'autant qu'ils concernent à la fois les élèves de CM1 et de terminale scientifique (TIMSS *Advanced*).

Sa réalisation par le Cnesco s'est faite conformément à l'une des trois missions que lui assigne la loi d'orientation et de programmation pour la refondation de l'École de la République du 8 juillet 2013, à savoir « Évaluer les méthodologies mises en œuvre par les évaluateurs internes au ministère de l'Éducation nationale et celles mises en œuvre par les organismes internationaux ». Son objectif est d'analyser les cadres conceptuels et leurs opérationnalisations utilisés par les enquêtes internationales PISA et TIMSS. Étant donné la spécificité de TIMSS et l'importance donnée aux sciences par PISA en 2015, le rapport est centré sur les mathématiques et les sciences. Cette analyse porte sur deux grands axes : le premier est d'ordre technique, il concerne la préparation des enquêtes, les processus d'élaboration de leurs « instruments », c'est-à-dire tous les supports qui servent à leur mise en œuvre, le mode de passation des tests par les élèves, et la manière de rendre compte des résultats (la construction des « échelles ») ; c'est l'objet du premier volume du rapport, dont certaines parties sont d'ailleurs valables pour les autres domaines de PISA (comme par exemple la compréhension de l'écrit). Le second axe du rapport est d'ordre didactique : il s'intéresse aux contenus des tests qui sont proposés aux élèves, et analyse ce qui est demandé aux élèves en termes de connaissances et de compétences à mettre en œuvre ; le second volume du rapport lui est consacré.

Ce rapport n'a donc pas pour objet de présenter les résultats 2015 des élèves français : ils étaient en effet inconnus des auteurs lors de sa rédaction. Il donne toutefois des clés de lecture de ces résultats. Ce rapport doit permettre à tous – décideurs, chercheurs, formateurs, enseignants – de mieux appréhender les enquêtes PISA, TIMSS et TIMSS *Advanced* en soulignant leurs apports respectifs, et leurs limites éventuelles.

Pour rédiger ce rapport, le Cnesco a fait appel à plusieurs experts. Antoine Bodin, directeur du CEDEC (Centre pour le Développement et l'Évaluation des Compétences - École de la deuxième Chance de Marseille), en a rédigé la plus grande partie. Il connaît bien ces enquêtes puisqu'il a été à la fois membre du groupe d'experts mathématiques de PISA (*Mathematics Expert Group*) et membre du groupe d'experts de TIMSS pour les mathématiques (*Subject Matter Advisory Committee*). Ces enquêtes étant réalisées à l'aide de méthodes complexes issues de la psychométrie, le Cnesco a également sollicité Pierre Vrignaud, professeur émérite de psychologie à l'université Paris-Ouest Nanterre-La Défense, spécialiste de l'évaluation scolaire.

Étant donné l'importance accordée à la « culture scientifique » dans PISA 2015 et à la physique dans TIMSS *Advanced*, le Cnesco a fait appel à Nicolas Décamp et Cécile de Hosson, tous deux enseignants-chercheurs à l'université Paris Diderot et chercheurs en didactique des sciences au laboratoire de didactique André

Revuz. Ils ont eu la possibilité inédite d'étudier l'intégralité des questions¹ de PISA et de TIMSS qui ont été proposées aux élèves en 2015. Le rapport ne présente cependant que des questions dites « libérées » (rendues publiques ou soumises à usage restreint avec autorisation).

Afin d'approfondir l'analyse des exercices de mathématiques proposés aux élèves de CM1 dans le cadre de l'enquête TIMSS *grade 4* (TIMSS4), Nadine Grapin, maître de conférences à l'université Paris-Est Créteil a apporté son expertise en didactique des mathématiques.

Enfin, ce rapport a bénéficié des remarques et conseils de Franck Salles, coordinateur national de recherche pour TIMSS *Advanced* 2015, et co-manager du projet PISA 2012 pour la France.

¹ Le Cnesco remercie l'Organisation de Coopération et de Développement Économiques (OCDE), l'Association internationale sur l'évaluation et la réussite scolaire (IEA) et la Direction de l'évaluation, de la prospective et de la performance du ministère de l'Éducation nationale (Depp) pour avoir rendu possible ce travail exhaustif.

Introduction

Dans un monde de plus en plus interconnecté, dans le secteur éducatif comme dans les autres, chacun peut désormais connaître ce qui se passe en dehors des frontières de son propre pays, et chercher à comprendre les éventuelles différences. C'est le cas des chercheurs qui s'intéressent aux invariants ou aux facteurs de différenciation entre les systèmes éducatifs. C'est également le cas des décideurs politiques et des responsables des systèmes éducatifs qui cherchent dans les comparaisons internationales des éléments qui seraient de nature à les aider dans leurs actions au service de l'amélioration de leurs propres systèmes (*benchmarking*²). Ils cherchent aussi à s'assurer que le niveau des élèves de leurs pays ne s'éloigne pas trop des standards internationaux (Bottani & Vrignaud, 2005).

Ces éléments d'évaluation viennent compléter et mettre en perspective ceux qui sont obtenus par des sources nationales ; en France, ce sont, en particulier, ceux des évaluations menées auprès des élèves par la direction de l'évaluation de la programmation et de la prospective (Depp).³

Les programmes PISA (*Programme for International Student Assessment* ; en français, Programme international pour le suivi des acquis des élèves) et TIMSS (*Trends in International Mathematics and Science Study* ; en français, Tendances internationales dans l'enseignement des mathématiques et des sciences)⁴ sont sans doute les plus importants des programmes mis en place pour comparer les systèmes éducatifs. À eux deux ils couvrent la quasi-totalité des pays et des systèmes éducatifs de la planète ; en tout cas, plus de 90 % de son économie.

Tous les 3 ans depuis l'année 2000, les résultats de PISA sont présentés et commentés par l'OCDE (Organisation de coopération et de développement économiques) pour chacun des pays concernés, puis, dans ces pays, par les instances gouvernementales. Les médias tant professionnels que généralistes présentent et commentent à leur tour les résultats obtenus, et produisent ou renforcent un effet palmarès plus ou moins accentué ; les décideurs s'appuient sur ces résultats pour alerter sur tel ou tel point, pour annoncer de nouvelles mesures et, dans de nombreux pays, pour influencer sur les programmes d'enseignement et sur la formation des enseignants.

TIMSS est moins connu en France car après avoir fait partie de la première enquête en 1995, le pays ne participait plus à ce programme. Cependant, après 20 ans d'absence, la France a rejoint l'enquête en 2015, et l'on ne manquera certainement pas d'en entendre parler lorsque les résultats seront publiés (le 29 novembre 2016).

De nombreuses questions se posent à propos de ces programmes et des enquêtes qu'ils conduisent ; questions qui intéressent à des titres divers tous les acteurs de notre société.

Pourquoi ces programmes ? Qui en est responsable ? En quoi sont-ils comparables ? Leurs objectifs sont-ils similaires ou sont-ils complémentaires ? Quel est le niveau de fiabilité des méthodes utilisées et des

² La présente étude évite au maximum l'emploi de termes de la langue anglaise. Lorsqu'il semblera nécessaire d'utiliser de tels termes, ils seront placés en italique et entre guillemets. « *Benchmarking* » fait partie de ces mots qui n'ont pas d'équivalent en français.

³ Signalons qu'au cours de ces dix dernières années, les méthodes statistiques utilisées par la Depp se sont considérablement rapprochées de celles utilisées des enquêtes internationales : utilisation des modèles de réponses à l'item, construction d'échelles, etc.

⁴ Voir en annexe la signification des sigles et acronymes.

résultats publiés? Quelle est la part du contexte culturel dans les évaluations ? En quoi ces enquêtes peuvent-elles effectivement contribuer à l'amélioration de notre système éducatif ?

Ces questions et d'autres de nature plus philosophique ou politique ont suscité de nombreux débats, en France et dans beaucoup d'autres pays. La bibliographie donne quelques pistes, mais la production dans ce domaine est vaste et l'information facile à trouver sur le Web. L'objet du présent rapport se limite à éclairer les cadres de référence (ou cadres conceptuels) des programmes PISA et TIMSS, et leurs opérationnalisations.

Au fil des années, les programmes PISA et TIMSS sont devenus des organisations munies de moyens humains, méthodologiques et technologiques importants.

Au cours de l'année 2015, la France a participé à ces enquêtes pour divers niveaux et filières de la scolarité (CM1, terminale S) et âge (15 ans). En complément de nombreux documents et articles publiés antérieurement (*cf.* bibliographie), il a paru important de proposer à un large public, sous forme synthétique, des clés de lecture et d'interprétation des résultats qui seront publiés à partir de la fin du mois de novembre 2016.

Après une présentation générale des programmes PISA et TIMSS (chapitre 1), ce premier volume du rapport se poursuit avec une présentation détaillée des cadres de référence des deux programmes (chapitres 2 et 3). Ces cadres sont en quelque sorte des cahiers des charges définis pour piloter l'élaboration des « instruments » utilisés (tests et questionnaires). Une connaissance minimale de ces cadres est indispensable pour quiconque ne souhaite pas se limiter à la lecture des classements et des échelles de type palmarès qui sont régulièrement publiés.

Il est également important de s'intéresser à la manière dont sont présentés les exercices dans les deux enquêtes ainsi qu'aux types de questions posées. Cela permet de mieux appréhender les difficultés des élèves tout comme les différentes philosophies de PISA et de TIMSS (chapitre 4). Enfin, le dernier chapitre sera consacré à la présentation et à la lecture des résultats (chapitre 5).

Le second volume du rapport présente et analyse des exercices utilisés pour l'évaluation des connaissances et des compétences des domaines scientifique et mathématique.

Chapitre 1 : Présentation générale des enquêtes PISA et TIMSS

Les enquêtes internationales sont de plus en plus nombreuses et touchent tous les domaines. Dans celui de l'éducation, elles concernent tous les niveaux, de l'élémentaire à l'université et la plupart des matières. Elles concernent même les adultes qui ne sont plus scolarisés. Les revues et ouvrages spécialisés se font régulièrement l'écho de ces enquêtes en insistant souvent davantage sur les résultats que sur les objectifs et les méthodes utilisées. Parmi ces enquêtes, la plus connue est assurément l'enquête PISA de l'OCDE, dont les résultats font désormais la une des médias tous les trois ans. Pour les adultes, citons le Programme pour l'évaluation internationale des compétences des adultes (PIAAC) ainsi que celui sur la littératie et les savoir-faire des adultes (ALL). Ces deux programmes sont aussi menés par l'OCDE et partagent avec PISA des objectifs, des analyses, et des méthodes. Au niveau du CM1, la France participe au Programme international de recherche en lecture scolaire (PIRLS), mené par l'Association internationale sur l'évaluation et la réussite scolaire (IEA). D'autres enquêtes internationales concernent le préscolaire, la citoyenneté, les compétences numériques (digitales), les enseignants (Enquêtes TALIS), etc. Le lecteur trouvera une information exhaustive sur l'ensemble de ces programmes sur les sites de l'IEA et de l'OCDE (voir adresses utiles en fin de bibliographie).

En ce qui concerne PISA et TIMSS on trouvera dans les nombreux documents en accès libre sur Internet, et partiellement en français, des descriptions plus ou moins détaillées de leurs objectifs et de leurs méthodologies (cf. bibliographie). Certains des documents publiés par l'OCDE à propos de PISA le sont en français. Toutefois, il s'agit dans la plupart des cas de traductions des documents rédigés en anglais. Ces traductions masquent souvent des différences de culture existant entre les conceptions anglo-saxonnes et les conceptions latines en matière d'éducation (cf. chapitre 3). Pour atténuer l'effet de ces traductions, ou la reprise d'analyses fondées sur ces traductions, qui peuvent conduire à des interprétations erronées, la présente étude se base essentiellement sur les documents originaux en langue anglaise.

A. L'IEA et les enquêtes sur l'enseignement des mathématiques et des sciences

L'IEA (*International Association for the Evaluation of Educational Achievement*) est une association internationale à but non lucratif, indépendante des États et dont les membres sont des organismes de recherche universitaires ou gouvernementaux. L'IEA conduit des enquêtes internationales à grande échelle dans le domaine de l'éducation.

TIMSS s'inscrit dans l'histoire des enquêtes de l'IEA, laquelle débute dans les années 60 avec, en particulier, FIMS en 1964 (Première Étude Internationale sur l'Enseignement des Mathématiques), puis SIMS en 1982 (Seconde Étude). TIMSS a d'abord été, en 1995 et en 1999, l'acronyme anglais de « Troisième Étude internationale sur l'Enseignement des Mathématiques et des Sciences » de l'IEA, pour devenir en 2003 celui du programme « Tendances Internationales dans l'Enseignement des Mathématiques et des Sciences ». Depuis 1995, les enquêtes TIMSS ont lieu tous les quatre ans pour les élèves qui sont dans leur quatrième ou huitième année de scolarité (en France, élèves de CM1 et de quatrième). Ces enquêtes ont aussi lieu de façon plus irrégulière pour les élèves en fin d'études secondaires (limitées aux seuls élèves des classes à orientation scientifiques après 1995). La dernière étude TIMSS a eu lieu en 2015, simultanément pour ces trois populations. Toutefois, la France n'a participé que pour le niveau CM1 et pour le niveau des classes terminales scientifiques (terminales S).

Le nombre de pays ou de systèmes éducatifs participant aux enquêtes TIMSS est irrégulier ; de plus, les pays peuvent ne participer que pour un ou deux des niveaux concernés. En 2015, 57 pays ont participé à TIMSS : 48 pays pour la quatrième année de scolarité (CM1), 40 pays pour la huitième année, et 9 pays (États-Unis, France, Italie, Liban, Norvège, Portugal, Russie, Slovaquie et Suède) pour le niveau terminal scientifique. La France a participé aux enquêtes FIMS et SIMS pour les élèves de quatrième, puis à TIMSS en 1995 pour les élèves de quatrième et pour ceux en fin d'études secondaires (toutes sections, y compris les classes de baccalauréat professionnel). Elle a ensuite interrompu sa participation à TIMSS et ne l'a reprise qu'en 2015, mais seulement pour les élèves de CM1 (*grade 4*)⁵ et pour ceux de terminale scientifique (*TIMSS Advanced*)⁶.

B. L'OCDE et les enquêtes PISA

L'OCDE regroupe 35 pays parmi les plus riches de la planète. Sa mission « *est de promouvoir les politiques qui amélioreront le bien-être économique et social partout dans le monde* ». L'OCDE étant convaincue de l'importance de l'éducation tant pour le développement économique des pays que pour le bien-être des individus, elle ne peut manquer de s'intéresser de près à ce domaine.

En 1997, dans le cadre de sa réflexion sur les compétences clés, l'OCDE a lancé le programme PISA. La première étude a eu lieu en 2000 ; les autres ont suivi à intervalles de 3 ans. En 2015 a donc eu lieu la sixième étude du cycle PISA ; la septième, en préparation, aura lieu en 2018. Les enquêtes PISA portent sur la littératie, dans un sens qui sera précisé plus loin ; elles concernent l'ensemble des élèves dont l'âge est compris entre 15 ans 3 mois et 16 ans 2 mois, quelle que soit la place qu'ils occupent dans le système éducatif (en 2012, en France, la moyenne d'âge des élèves qui ont passé les épreuves PISA était de 15 ans 8

⁵ Noté TIMSS4 par la suite.

⁶ Noté TIMSSADV dans la suite de ce rapport.

mois). Conçues en premier lieu pour les pays de l'OCDE, lesquels sont tenus d'y participer, les enquêtes PISA se sont rapidement ouvertes à tous les pays qui le souhaitaient et qui en avaient les moyens. Ainsi 71 pays ou systèmes économiques ont participé à PISA 2015 (dont les 34 pays de l'OCDE⁷).

C. Objectifs généraux et évolution des objectifs dans le temps

1. Les objectifs de TIMSS

TIMSS est d'abord conçu pour contribuer aux recherches sur l'enseignement des mathématiques et des sciences. Il s'agit d'étudier, dans une perspective comparatiste, les curriculums⁸ mathématiques et scientifiques des pays participant aux enquêtes. TIMSS distingue plusieurs niveaux de curriculum, et s'intéresse d'abord aux contenus d'enseignement, à leur présence dans les programmes et aux acquis des élèves par rapport à ces contenus. TIMSS cherche d'abord à mieux connaître les systèmes éducatifs en ce qui concerne l'enseignement des mathématiques et des sciences et non, directement, à améliorer ces systèmes. En témoigne la publication d'une encyclopédie régulièrement mise à jour résumant les curriculums mathématiques et scientifiques, ainsi que les politiques éducatives de près de 80 pays ou systèmes éducatifs (Mullis & *al.* (ed.), 2012). De ce point de vue TIMSS mène bien des enquêtes comparatives et non des évaluations.

Le but premier de TIMSS a toujours été et reste aujourd'hui la constitution d'une base de données mise à la disposition des chercheurs concernés par l'éducation mathématique et scientifique. De ce fait, dans de nombreux pays, les centres chargés des enquêtes TIMSS sont des institutions universitaires. Le rapport international publié par l'IEA est par conséquent peu étoffé et laisse la place à de nombreuses analyses secondaires. L'OCDE, quant à elle, tend à être exhaustive sur le traitement statistique des données produites par les enquêtes qu'elle conduit.

TIMSS est largement indépendant des États, lesquels sont libres d'y participer ou non. Toutefois, TIMSS entretient des relations étroites avec les gouvernements et avec les responsables des systèmes éducatifs qui participent à ses enquêtes. Cette participation n'est pas gratuite, ce qui donne de fait un droit de regard des gouvernements sur le déroulement des enquêtes. L'IEA ne fournit pas à ces gouvernements des évaluations toutes faites et ne cherche pas à s'immiscer dans les politiques éducatives. Il se limite à fournir des informations leur permettant de construire ou de compléter leurs évaluations nationales tout en les replaçant dans un contexte international.

⁷ La Lettonie, qui participait à PISA 2015 comme pays partenaire, est devenue membre de l'OCDE en 2016.

⁸ Bien que le mot commence à être utilisé dans la littérature de langue française, précisons que le curriculum concerne, certes, les programmes d'enseignement, mais aussi tout ce qui contribue à l'expérience scolaire des élèves.

2. Les objectifs de PISA

Les enquêtes PISA, organisées par l'OCDE, obéissent d'abord à des préoccupations de natures économique et sociétale. En particulier, l'OCDE part de l'idée que le développement des économies des pays dépend largement de la qualité de leur système éducatif. En 1997, l'OCDE a assigné à PISA la mission de « *déterminer dans quelle mesure les élèves qui approchent du terme de leur scolarité obligatoire possèdent les savoirs et les savoir-faire indispensables pour participer à la vie de la société* » (OCDE, 2005).

Pour l'OCDE, et donc pour PISA, les connaissances n'ont d'intérêt pour l'ensemble des citoyens que dans la mesure où ces derniers seront capables de les utiliser pour résoudre les problèmes qu'ils sont susceptibles de rencontrer dans la « vie réelle ». Les enquêtes PISA sont essentiellement destinées à informer les décideurs nationaux et à les aider à orienter leurs politiques éducatives. De plus, l'OCDE, à partir de ses propres objectifs et des résultats des enquêtes PISA, émet régulièrement des recommandations, incitant les pays à faire évoluer leur système éducatif. PISA, comme TIMSS, produit une base de données très importante qui, elle aussi, est mise à la disposition des chercheurs.

D. Préparation et organisation des enquêtes

1. Méthodes psychométriques communes à PISA et à TIMSS

La préparation et l'organisation des enquêtes PISA et TIMSS sont des opérations complexes qui mobilisent de très nombreux acteurs sur toute la planète. Ainsi, pour PISA 2015, entre 4 500 et 10 000 élèves ont passé les tests dans chacun des 71 pays ou systèmes économiques participants (plus de 500 000 élèves au total). Dans chaque pays, un échantillon représentatif d'élèves est soumis aux épreuves ; en France, cet échantillon est habituellement de l'ordre de 5 000 élèves, répartis dans quelque 200 établissements publics et privés⁹. Pour des informations plus précises sur la question de l'échantillonnage dans le cas français, on peut se référer aux notes d'information de la Depp (ou DEP) citées en référence. En ce qui concerne PISA, dans chaque établissement de l'échantillon, une ou plusieurs classes (selon la taille de l'établissement) est sélectionnée au hasard, et dans chacune de ces classes un nombre précis d'élèves est choisi au hasard. Pour TIMSS, ce sont des classes entières qui sont sélectionnées et qui répondent aux questionnaires.

Il existe tout un champ lexical propre aux évaluations des élèves, qui peut constituer une source d'incompréhension pour les non-spécialistes. L'encadré ci-après fixe le vocabulaire adopté dans ce rapport, qui n'est pas forcément celui des rapports nationaux ou internationaux :

⁹ Enquêtes 2015 en France : échantillon PISA : 6 000 élèves dans 250 établissements. TIMSS4 : 5 000 élèves dans 150 établissements. TIMSSADV : 8 000 élèves dans 150 établissements (source Depp).

Vocabulaire

- **Exercice ou exercice d'évaluation** : une unité de test d'évaluation pouvant comporter plusieurs questions (dans la terminologie PISA, un exercice est un « item »).
- **Question** : tout élément d'un exercice qui demande une réponse et qui sera l'objet d'un codage relevant de consignes de corrections standardisées.
- **Cahiers d'exercices ou cahiers de tests** : les cahiers regroupant des exercices d'évaluation, tels qu'ils ont été présentés aux élèves. L'ensemble de ces cahiers forme les questionnaires cognitifs.
- **Questionnaires de contexte** (ou contextuels) : questionnaires non cognitifs destinés à fournir des informations à la fois sur les élèves (niveau socio-économique et culturel, motivation, confiance en soi, etc.) ou sur les établissements (équipement, niveau d'autonomie, etc.)
- **Items** : les énoncés des questions des questionnaires contextuels qui appellent une réponse ou un choix.

Pour PISA, les cahiers de tests sont composites et comportent des exercices de plusieurs domaines (mathématiques, compréhension de l'écrit et sciences). Les élèves d'une même classe n'ont pas tous les mêmes exercices à traiter, du fait de la méthode dite « des cahiers tournants » (voir encadré ci-dessous), mais ils répondent aux mêmes items des questionnaires contextuels. Ainsi, pour PISA 2012, les exercices des différents domaines étaient répartis dans 13 cahiers différents. 2 cahiers seulement étaient proposés à chaque élève, sur une durée de 2 heures. De plus, 30 minutes supplémentaires étaient occupées par la passation des questionnaires contextuels. Pour TIMSS et TIMSSADV, les cahiers de tests sont organisés par matière : les cahiers de mathématiques et les cahiers de sciences ou de physique sont séparés. Des exercices communs à plusieurs cahiers de tests permettent des raccords, ce qui exige des méthodes probabilistes assez complexes utilisées pour le traitement des résultats (voir encadré ci-dessous).

Point méthodologique : pourquoi utiliser la méthode des cahiers tournants ?

Il apparaît difficile de concilier deux exigences : celle de présenter un grand nombre d'exercices pour mieux assurer la représentativité des compétences évaluées et celle de ne pas accroître de manière excessive le temps et la charge de travail des élèves. Pour concilier ces deux exigences, la solution consiste à ne pas administrer tous les exercices à tous les élèves. Il faut néanmoins disposer d'un lien (ancrage, *linking*) entre toutes les questions des exercices si l'on veut pouvoir les placer sur une même échelle. Les exercices seront donc répartis dans des blocs de durée équivalente, chaque élève passant un nombre réduit de blocs. La méthode psychométrique dite des plans incomplets (*Balanced Incomplete Block design* ou BIB) va permettre de réaliser des combinaisons optimales des ensembles de blocs d'exercices.

Du fait que les élèves ne passeront qu'une partie des exercices génère une situation où les exercices non administrés aux élèves sont considérés comme des données manquantes. Ces données manquantes sont considérées comme structurelles (engendrées par la structure du plan de collecte des données) et non engendrées par le fait qu'un élève n'a pas répondu volontairement ou involontairement à une partie des questions.

Comme il faut de plus neutraliser l'ordre de passation (le bloc passé en premier sera sans doute mieux réussi que le bloc passé en dernier par des élèves « fatigués »), chacun des blocs sera placé dans chacune des positions possibles dans les livrets. Cette combinaison des différents blocs et de leur ordre dans les livrets est appelée « cahiers tournants ». L'utilisation des cahiers tournants a été mise au point pour les enquêtes américaines et elle est systématiquement mise en œuvre dans les enquêtes PISA et TIMSS.

Des modèles probabilistes adaptés, appelés « modèles de réponse à l'item », ont été développés pour permettre une estimation fiable des paramètres des items et des paramètres de compétence des élèves sans que ces estimations soient perturbées par les données manquantes et la structure des cahiers tournants.

Pour PISA, l'utilisation de deux formats (papier et informatique) permet de créer davantage de blocs, ce qui autorise un *design* plus complexe engendrant davantage de combinaisons d'exercices. Bien que le principe de construction des épreuves soit le même que celui dit des « cahiers tournants », il serait inapproprié d'employer ce terme dans la mesure où il s'agit d'un matériel au format informatique.

La préparation de chaque enquête commence trois ans avant la passation des tests dont les résultats seront publiés. Cette durée se justifie notamment par la nécessité de vérifier à l'aide de méthodes psychométriques que le choix des items posés est pertinent, assurant ainsi leur validité. Il est également nécessaire d'identifier les biais éventuels induits par l'utilisation de versions linguistiques différentes des questionnaires.

Lors d'une phase de pré-test, un premier travail est réalisé sur les questions d'évaluation. Il consiste à vérifier leur fiabilité de manière générale en réalisant des analyses tant au niveau national qu'international : ces questions sont-elles homogènes, c'est-à-dire mesurent-elles la même compétence que les autres questions rattachées à la même échelle ? Les indicateurs obtenus (issus de la théorie psychométrique dite classique et de méthodes dites de réponses à l'item)¹⁰ sont fournis pour chaque question aux responsables internationaux et nationaux. L'ensemble de ces indicateurs est alors publié pour chaque question dans les documents des deux enquêtes. Ces informations conduisent à éliminer certaines questions après la phase de pré-test ou à gérer leur insertion dans la construction des échelles de manière spécifique (attribution d'une valeur spécifique du paramètre de difficulté pour la question dans un pays donné).

Puis, de manière plus spécifique, on va s'assurer de la possibilité d'émettre des comparaisons internationales en vérifiant l'absence de ce qu'il est convenu d'appeler des « biais culturels » : ce serait le cas si une question se comportait de manière différente selon les pays. Il s'agit alors d'identifier ce que la théorie nomme le « Fonctionnement Différentiel des Items ». On va ici vérifier qu'il n'apparaît pas d'interaction entre le niveau de difficulté des questions et chacun des pays, c'est-à-dire qu'un pays donné n'a pas un effet sur le niveau de difficulté des questions différent de celui de l'ensemble des pays enquêtés. Ce pourrait être le cas, par exemple, si la traduction d'une question à partir de l'anglais était incorrecte, ou si le contexte dans lequel s'inscrit la question était peu familier pour les élèves de certains pays. Dans les dernières versions des enquêtes TIMSS et PISA, les procédures et les tests statistiques utilisés pour cette identification sont les mêmes. Une fois cette phase de pré-test terminée, on procède à l'estimation des paramètres de difficulté des questions et de compétence des élèves.

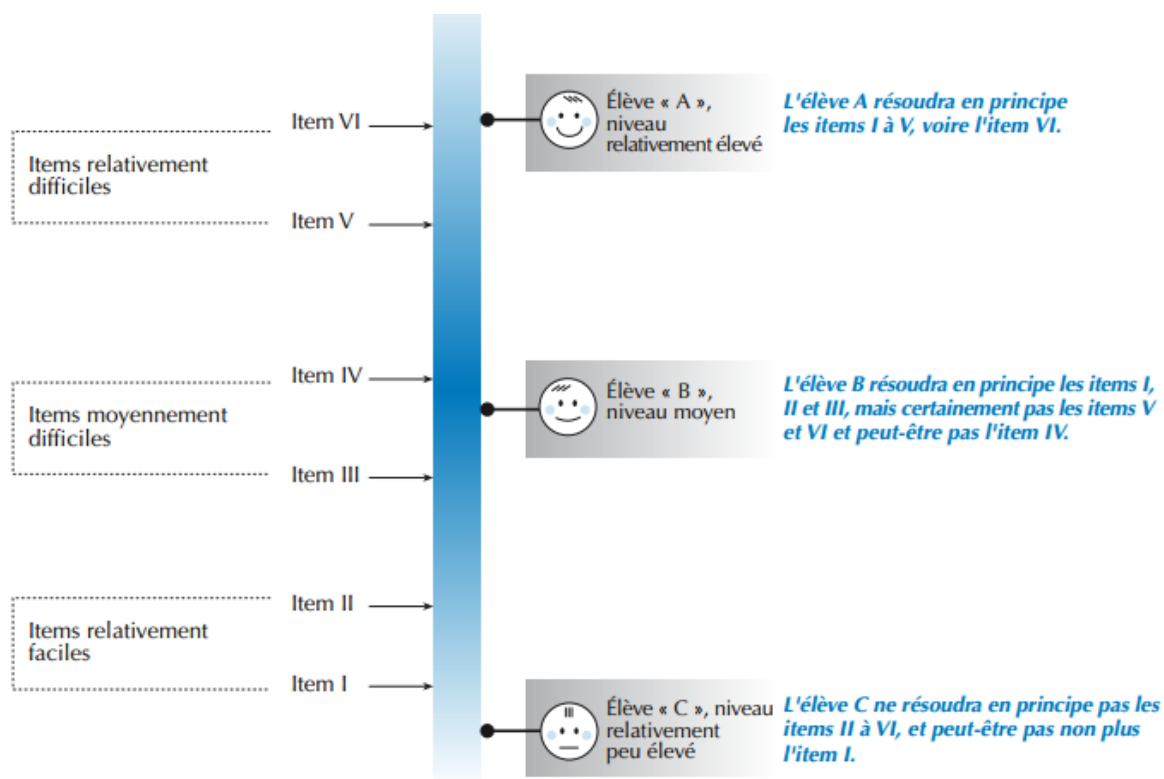
¹⁰ L'alpha de Cronbach est un indicateur particulièrement utilisé pour mesurer la cohérence globale d'un test.

La construction d'une « double échelle » pour présenter les résultats

Pour les deux enquêtes, les scores des pays sont donnés sur une échelle qui permet de représenter à la fois les estimations des niveaux de compétences des élèves et les difficultés des questions. La construction d'une telle échelle s'appuie sur les propriétés des « modèles de réponse à l'item » (Vrignaud, 2008).

Ces modèles sont probabilistes : on postule de façon assez raisonnable que la probabilité qu'un élève donne une réponse correcte à une question est fonction à la fois de sa compétence et de la difficulté de la question. Un élève très compétent a ainsi une probabilité élevée de répondre correctement à une question difficile, un sujet peu compétent aura une probabilité faible de répondre à une question difficile, mais aura une probabilité élevée de répondre à une question facile. Il s'agit d'estimer les niveaux de difficulté des questions et les niveaux de compétence des élèves sur une même échelle, permettant *in fine* de faire correspondre chaque question à un niveau de compétence (score obtenu) et de décrire les compétences associées à chaque type d'élève (ou intervalle de score). La figure 1 présente cette relation entre questions et compétences.

Figure 1 : Les relations entre niveaux de compétence des élèves et difficulté des questions (*items*)



Source : OCDE (2004)

Plus précisément, la compétence d'un élève est découpée en plusieurs niveaux qui seront définis à partir des questions classées dans chacun de ces niveaux : cinq pour TIMSS et six (ou sept) pour PISA. Ce découpage indique la probabilité de réussite d'un élève à une question en fonction de son niveau de compétence. Un élève d'un niveau de compétence donné a plus de chances de répondre correctement à des questions placées à des niveaux inférieurs de l'échelle (et en a d'autant plus que le niveau de difficulté des questions diminue) et au contraire, il en a moins de répondre correctement à des questions situées à des niveaux supérieurs de l'échelle.

Point méthodologique : les algorithmes d'estimation

De manière schématique, les algorithmes d'estimation permettent de calibrer dans un premier temps les questions (c'est-à-dire qu'on estime leur difficulté), puis dans un second temps, les paramètres de compétence des élèves. Pour PISA, l'échantillon de calibration est constitué d'un tirage aléatoire de 500 élèves de chacun des pays participants qui sont soumis à une pré-enquête.

Le modèle retenu est un développement du modèle de Rasch, le modèle de crédit partiel de Masters (1982). Il permet d'estimer les différents niveaux de difficulté d'une question dans le cas de questions polytomiques (comportant plusieurs niveaux de réussite) selon la performance du sujet. Pour les questions dichotomiques (réussite/échec) le modèle logistique simple est utilisé. Pour TIMSS, les modèles de réponse à l'item diffèrent selon que les items se présentent sous forme de QCM, de successions de vrai/faux ou de possibilité de donner pour les élèves une réponse partiellement juste : modèle à trois paramètres pour les QCM, modèle à deux paramètres pour les questions à réponse construite codées de manière dichotomique, modèle de crédit partiel pour les questions polytomiques à réponse construite. Les données de toute la population sont utilisées pour la calibration avec des pondérations pour équilibrer le poids des échantillons de taille différente selon les pays.

Les deux enquêtes suivent des modèles différents pour l'estimation des paramètres. La différence principale porte sur le nombre de paramètres du modèle : modèle dit de Rasch à un paramètre pour PISA et modèle à deux, voire trois paramètres pour TIMSS. Comme le rapport technique de PISA 2015 n'est pas encore disponible, il n'est pas possible de donner toutes les informations sur le modèle utilisé, mais il semble que l'ensemble des données depuis 2000 ait été recalibré avec un modèle à deux paramètres pour permettre l'étude des tendances avec le modèle adopté en 2015. Ces points pourront être précisés et comparés avec les méthodes utilisées dans la version de TIMSS 2015 dès que le rapport technique aura été publié.

Ainsi, étant donné le caractère sophistiqué de ces études, on comprend qu'elles demandent un temps de préparation important. La richesse des données obtenues exige d'y consacrer tout autant de temps. En effet, le travail d'exploitation des données se poursuit en général trois années après la passation des tests. Les programmes PISA et TIMSS s'inscrivent dans le temps long, avec un fort souci de continuité et de cohérence entre les différents cycles des enquêtes. Par exemple, dans les deux programmes, une partie des questions a été passée de la même façon depuis 15 ou 20 ans : il est donc possible, et intéressant, d'étudier l'évolution des résultats aussi bien question par question qu'en s'intéressant aux échelles globales.

Enfin, pour PISA comme pour TIMSS, les cahiers de tests (questionnaires cognitifs) sont accompagnés de questionnaires de contexte de divers types (voir chapitre 3). Ces questionnaires comportent eux aussi des items qui sont suivis au fil des enquêtes ; ils permettent des comparaisons dans le temps, et en particulier d'associer des variations dans les résultats à des variations curriculaires ou sociétales. Les données issues de l'ensemble de ces questionnaires (cognitifs et contextuels) sont mises en relation et nourrissent nombre d'enquêtes et rapports.

En France, l'organisation et l'exploitation de ces deux enquêtes sont confiées à la direction de l'évaluation, de la prospective et de la performance du ministère de l'Éducation nationale (Depp)¹¹.

¹¹ Voir à ce propos dans la bibliographie « Documents officiels du ministère de l'Éducation nationale », en particulier les notes d'information de la Depp, et Salles (2012).

2. Spécificités des enquêtes PISA

L'OCDE, sur appel d'offres, sélectionne des organismes publics ou privés qui forment le « consortium¹² » en charge de mener l'enquête¹³ et met en place un comité directeur (*Governing Board*) de PISA. Ce comité de 71 membres (en 2015) comporte un représentant officiel de chacun des pays participants (membre OCDE ou non). De plus, chaque pays nomme un directeur national de l'étude. Le comité directeur, en accord avec l'OCDE, nomme ensuite les membres de divers groupes d'experts (experts choisis pour leurs compétences, qui ne représentent pas tous les pays participants et qui ne dépendent pas de leurs gouvernements) :

- Groupe d'experts « mathématiques » (10 membres ; aucun francophone et même aucun latin de 2000 à 2009) ;
- Groupe d'experts « sciences » (12 membres) ;
- Groupe d'experts « compréhension de l'écrit » (7 membres) ;
- Groupe d'experts « résolution de problèmes » (8 membres) ;
- Groupe d'experts « questionnaires » (8 membres) ;
- Groupe d'experts techniques (11 membres).

En tout, environ 60 experts (leur nombre peut changer légèrement d'une enquête à l'autre) dont 8 de langue française dont 5 Français, auxquels il faut ajouter une centaine d'experts membres du consortium ou indépendants (dont 5 francophones). Les exercices d'évaluation de PISA sont proposés par les membres du consortium et par les pays. Les groupes d'experts les classent et les évaluent. Une partie de ces exercices est soumise à une pré-enquête (*field trial*) un an avant l'enquête principale. C'est cette pré-enquête qui permet un premier paramétrage des questions et le choix de celles qui seront utilisées lors de l'étude principale.

Le concept central des enquêtes PISA est celui de littératie. Dans la littérature de langue anglaise, le terme « *literacy* » apparaît dès les années 1950. Il ne s'est d'abord appliqué qu'à la question de la maîtrise de la langue, avant de se propager à d'autres domaines. Lorsqu'il est utilisé dans son sens premier, ce terme est associé à celui de « *numeracy* » qui, est encore largement utilisé lorsqu'il est question de savoir-faire de base relevant des mathématiques. En français, le terme littératie est maintenant couramment utilisé, mais il peut avoir des significations assez différentes allant de la simple alphabétisation à la culture au sens large. PISA évalue la littératie en même temps qu'elle la définit dans un sens particulier. En 2000, l'OCDE donnait la définition suivante de la littératie :

La littératie est l'« aptitude à comprendre et à utiliser l'information écrite dans la vie courante, à la maison, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités. » (OCDE, 2000)

Pour l'étude PIAAC, l'OCDE reprend la définition de l'UNESCO :

La littératie est la capacité à identifier, comprendre, interpréter, créer, communiquer et calculer, en utilisant des matériaux imprimés et écrits associés à des contextes variables. (UNESCO, 2005)

¹² Pour PISA 2012, le consortium était placé sous la direction de Australian Council for Educational Research (ACER). Depuis 2015, ACER a été remplacé par Educational Testing Service (ETS) –États-Unis :

¹³ Ici il faut mentionner le changement opéré en 2015, année à partir de laquelle Pearson entre dans le consortium, notamment pour la conception des cadres d'évaluation. Le conflit d'intérêt entre le concepteur du cadre PISA et le marchand international de manuels scolaires qu'est Pearson peut inquiéter.

Cette définition, qui ne limite pas le champ de la littératie à la compréhension de l'écrit, est sans doute la plus proche de celle implicitement utilisée pour les enquêtes PISA.

Depuis l'année 2000, les enquêtes PISA se focalisent sur les trois premiers types de littératie que l'OCDE a définis : compréhension de l'écrit (*reading literacy*), littératie mathématique, et littératie scientifique, auxquelles se sont ajoutées les littératies financière (*financial*) et numérique (*digital*). Lors de chaque cycle triennal, l'une des trois premières littératies occupe à tour de rôle une place majeure (voir tableau 1). Environ les deux tiers des deux heures allouées à la passation des questionnaires cognitifs (un peu moins en 2015) y sont consacrés.

Tableau 1 : Place des différents domaines cognitifs de PISA au cours des enquêtes

LES DOMAINES COGNITIFS DE PISA ¹⁴	Cycles des enquêtes PISA					
	2000	2003	2006	2009	2012	2015
Compréhension de l'écrit <i>Reading literacy</i>	M	m	m	M+	m+	m++
Littératie mathématique <i>Mathematical literacy</i>	m	M	m	m	M+	m++
Littératie scientifique <i>Science literacy</i>	m	m	M	m	m	M++
Résolution de problèmes¹⁵ <i>Problem solving</i>		m			m	
Résolution collaborative de problèmes en ligne <i>Collaborative problem solving</i>						m++
Littératie financière <i>financial literacy</i>					m	m++

Légende du tableau

- M** : domaine majeur
- M+** : domaine majeur avec une partie sur papier et une partie informatisée (optionnelle)
- M++** : domaine majeur, uniquement informatisée
- m** : domaine mineur
- m+** : domaine mineur avec une partie sur papier et une partie informatisée.
- m++** : domaine mineur, uniquement informatisée

Ainsi, en 2009, comme en 2000, PISA a porté principalement sur la compréhension de l'écrit. En 2012, comme en 2003, elle a porté principalement sur la littératie mathématique. En 2015, comme en 2006, elle a porté principalement sur la littératie scientifique. En 2018 ce sera à nouveau la compréhension de l'écrit, etc. La dernière enquête PISA pour laquelle la littératie mathématique a été le domaine majeur date de 2012. Elle peut être lue en continuité avec l'enquête de 2003. Pour la littératie scientifique qui a été la majeure de PISA en 2015, les résultats ne seront connus qu'en décembre 2016.

¹⁴ Le cadre de référence parle de domaines cognitifs ou de domaines de l'évaluation.

¹⁵ Voir chapitre 3.

Plus précisément, en 2015, l'évaluation PISA, qui s'est focalisée sur le domaine scientifique, a cependant repris environ 80 % des questions posées en 2012 dans le domaine mathématique, cela pour permettre de suivre l'évolution des tendances. Le cadre de référence de 2015 est le même que celui de 2012, à la différence près du passage d'une passation papier-crayon à une passation entièrement informatisée.

3. Spécificités des enquêtes TIMSS

L'organisation des enquêtes TIMSS s'appuie aussi sur un consortium, mais elle est nettement plus simple que celle de PISA. En effet, l'IEA confie les enquêtes à un centre international situé dans une université (*Boston College*, USA) qui, en accord avec l'IEA, délègue certaines parties de l'étude à d'autres organismes ou centres de recherche (*i.e.* traitements statistiques à Hambourg et Ottawa, etc.). Comme pour PISA, chaque pays participant désigne un coordonnateur national de l'enquête. L'équipe en charge de l'enquête, toujours en accord avec l'IEA, nomme les groupes d'experts internationaux :

- Groupe d'experts pour les mathématiques (14 membres) ;
- Groupe d'experts pour les sciences (14 membres) ;
- Groupe d'experts pour la compréhension des textes (11 membres).

Soit 39 experts de diverses nationalités, mais aucun francophone depuis 1995.

Les consortiums qui organisent TIMSS et PISA (mais aussi PIRLS et d'autres enquêtes internationales) ont un organisme en commun : ETS¹⁶ (*Educational Testing Service*).

Une spécificité de TIMSS est que, conformément aux habitudes du monde de la recherche, tous les documents de TIMSS sont signés de leurs auteurs, ce qui n'est pas le cas pour PISA.

4. Rôle et place du numérique dans les enquêtes

Depuis la création des enquêtes PISA et TIMSS, la place du numérique dans nos sociétés s'est accrue. Ainsi, les questionnaires de contexte de ces deux enquêtes contiennent désormais des questions relatives au numérique. Au-delà de ces informations, cette place grandissante du numérique a également des conséquences sur le mode de passation des tests ainsi que sur l'étendue et la nature des compétences évaluées.

Pourquoi introduire la passation numérique des questionnaires ?

On peut souhaiter passer d'un mode de passation « papier/crayon » à un mode de passation numérique pour au moins deux types de raisons. D'abord, pour des raisons logistiques : cela réduit les coûts, notamment en rendant plus simple le processus de correction des QCM et en économisant l'impression de

¹⁶ Fondée en 1947, ETS est, selon Wikipédia, « la plus grande organisation à but non lucratif privée de mesure et d'évaluation éducative au monde » ; c'est elle qui organise les évaluations nationales des États-Unis ; c'est elle qui a fourni la première base d'items de TIMSS et c'est elle qui a développé la plupart des outils d'analyse des résultats des enquêtes. Malgré des évolutions récentes, ETS reste essentiellement pilotée par des préoccupations d'ordre psychométrique, domaine dans lequel sa suprématie est reconnue dans le monde entier.

milliers de cahiers. L'utilisation du numérique consistera alors en une simple transcription des questionnaires papier-crayons sous format numérique. Toutefois, il est difficile de postuler que cette transcription ne fasse pas intervenir des compétences supplémentaires, notamment celles qui consistent à maîtriser la navigation dans un environnement numérique ou encore à mettre de nouveaux outils au service de la résolution d'un problème. C'est alors la question de l'équivalence du mode de passation qui est posée. Il s'agit d'une question particulièrement importante dans une perspective de comparaison temporelle des résultats : lorsque le mode de passation diffère entre deux vagues, dans quelle mesure une telle comparaison est-elle fiable ? Ensuite, la place du numérique s'étant accrue, on peut souhaiter évaluer la capacité des élèves à utiliser des outils spécifiques au numérique pour résoudre des tâches. La passation numérique permet alors d'élargir le champ des compétences évaluées, comme par exemple des problèmes d'optimisation via un simulateur en sélectionnant des variables pertinentes (par manipulation d'un curseur et/ou en cochant une case). L'informatisation de la passation a également permis de diversifier les formats de réponse. Par exemple, les fonctions « glisser-déplacer » ou « sélectionner une image » permettent aux élèves de répondre à certaines questions de façon non verbale. Il est donc bien important de distinguer ces deux types de passation numérique et de ne pas perdre de vue qu'elles engagent des compétences parfois différentes : résoudre une équation « à la main » n'est pas équivalent à une résolution à l'aide d'un logiciel de calcul formel.

La passation numérique dans les enquêtes TIMSS et PISA

Mis à part les traitements de données, TIMSS utilise le numérique pour ses questionnaires contextuels à destination des enseignants et des chefs d'établissement, lesquels peuvent être renseignés en ligne, mais ne l'a pas intégré dans le matériel destiné aux élèves (évaluations cognitives ou questionnaires contextuels).

Il n'en va pas de même pour PISA qui attache une attention de plus en plus grande à la question du numérique. Les questionnaires contextuels contiennent nombre d'items sur la place de l'informatique dans l'enseignement, dont les réponses ont alimenté de nombreux rapports. Dès PISA 2000, le cadre de référence insistait sur l'importance de l'utilisation de l'ordinateur tant dans le domaine de la compréhension de l'écrit (*reading literacy*) que dans celui de la littératie mathématique. Dans ce domaine, la description de chaque catégorie de contenus incluait déjà un paragraphe sur l'utilisation des outils numériques (tableurs, logiciels de géométrie dynamique, etc.) dans l'enseignement des mathématiques. La passation des questionnaires sur papier, mais aussi les différences dans les équipements et dans les pratiques des pays participants a longtemps empêché la prise en compte de ces outils dans les exercices d'évaluation.

PISA 2009, dont le domaine majeur était la compréhension de l'écrit, a introduit « une nouvelle composante concernant la capacité des élèves à lire et à comprendre des textes sur support électronique, afin de tenir compte de l'importance que l'informatique a pris dans les sociétés modernes » (OCDE, 2009). Cela s'est traduit par la passation sur ordinateur d'exercices en compréhension de l'écrit spécialement conçus à cet effet. 19 pays, dont la France, ont participé à ce volet optionnel de l'enquête.

En ce qui concerne le domaine mathématique, PISA 2012 a proposé une version optionnelle sur écran des questionnaires cognitifs, version complémentaire de la version sur papier : 32 pays, dont la France, ont choisi l'option sur écran. Les questions utilisées, conçues pour un passage sur écran étaient différentes de

celles de la version sur papier. Un classement alternatif des pays avait alors été publié¹⁷. Les travaux menés par l'OCDE pour comparer les deux modes de passation ont conclu à une certaine équivalence et à la possibilité de passer au tout numérique sans nuire à la qualité des échelles de scores ni à la poursuite de leur comparabilité dans le temps en effectuant éventuellement les ajustements statistiques nécessaires. Signalons encore que l'épreuve de résolution de problèmes (*problem solving*) de 2012 a été passée uniquement sur ordinateur avec une utilisation importante de l'interactivité permise par ce média.

Un pas supplémentaire a été franchi dans l'utilisation du numérique avec PISA 2015 puisque l'ensemble de l'évaluation a été passée dans un environnement informatique dans tous les pays de l'OCDE. Toutefois des versions sur papier ont été produites pour d'autres pays qui n'auraient pas pu ou voulu passer au tout numérique. Concernant le domaine mathématique, dans la mesure où l'ensemble des exercices posés étaient repris de PISA 2012, il n'a pu s'agir que d'une adaptation des exercices à l'écran, sans autre utilisation des outils numériques. Pour la littérature scientifique pour laquelle l'ensemble de la passation a été informatisée en 2015, PISA a repris et numérisé 6 blocs de questions de PISA 2006 et y a ajouté 8 blocs totalement nouveaux, dont une part adaptée au caractère interactif rendu possible par une passation sur écran. En outre, un nouveau domaine est venu remplacer le domaine résolution de problèmes : la résolution collaborative de problèmes. Cette fois, ce sont les ressources (ou les simulations) de la communication à distance qui sont mises à profit pour résoudre des problèmes dans le cadre d'une collaboration (simulée) avec autrui. Finalement, c'est depuis ce dernier cycle que l'ensemble des domaines de l'évaluation sont passés, officiellement, en mode numérique.

L'OCDE semble avoir pris au sérieux le problème éventuel d'équivalence du mode de passation. Des études sur la question ont été réalisées en 2012 (OCDE, 2015), puis en 2014, dans la version expérimentale de PISA 2015 où des élèves avaient passé les mêmes épreuves dans les deux formats. Les informations recueillies permettront d'apporter des réponses sur le niveau d'équivalence existant entre l'épreuve papier/crayon et l'épreuve au format informatique. Des écarts plus ou moins importants selon les pays entre les résultats des deux modes de passation ont été constatés, ce qui conduit nécessairement à introduire des corrections pour rendre comparables les échelles obtenues dans chacun des cas et, surtout, pour assurer la continuité avec les enquêtes précédentes. Pour ce qui est du domaine des sciences, hors le mode de passation, le cadre conceptuel de l'OCDE pour 2015 embrasse peu la particularité du mode de passation informatique dans l'évaluation de la « culture scientifique ».

Que nous dit la recherche sur cette question ?

La Depp a mené deux expérimentations sur la question. La première concernait les compétences de base des élèves en fin de primaire. Les matières évaluées étaient alors le français et les mathématiques. Il en est ressorti que le mode de passation avait peu d'influence sur l'évaluation de l'orthographe. En revanche, la compréhension de textes longs était mieux réussie sur papier que sur numérique, des manipulations supplémentaires de navigation informatique étant nécessaires dans le second cas¹⁸.

La seconde évaluation concernait, quant à elle, les acquis des élèves en fin de primaire et de collège dans le cadre de l'évaluation CEDRE mathématiques. En fin de primaire, on observe que les questions impliquant une lecture directe, sans difficulté majeure, sont mieux réussies sur support numérique. En revanche,

¹⁷ http://www.keepeek.com/Digital-Asset-Management/oecd/education/students-computers-and-learning_9789264239555-en#page24

¹⁸ http://cache.media.education.gouv.fr/file/revue_86-87/57/8/depp-2015-EF-86-87-evaluation-sous-forme-numerique-est-elle-comparable-evaluation-type-papier-crayon_424578.pdf

lorsque des étapes intermédiaires sont nécessaires (utiliser un brouillon, un instrument de mesure...), c'est le support papier qui est le plus efficace. En fin de collège, on retrouve le résultat selon lequel une question qui nécessite des étapes intermédiaires (schématisation, instruments de mesure, calculs intermédiaires...) est moins bien réussie sur support numérique¹⁹.

Ainsi, postuler une équivalence parfaite entre passation numérique et papier pour une même question est une hypothèse assez forte. Les résultats des expérimentations menées par l'OCDE en 2012 et en 2014 devraient permettre de préciser cette conclusion.

¹⁹ Idem.

Chapitre 2 : Le cadre de référence des enquêtes PISA

Le but de ce chapitre est d'éclairer le cadre de référence de PISA qui « présente les principes directeurs du cycle d'évaluation [...] et définit les contenus que les élèves doivent acquérir, les processus qu'ils doivent appliquer et les contextes dans lesquels leurs savoirs, savoir-faire et compétences seront évalués. Il fournit par ailleurs des exemples de tâches permettant d'illustrer les divers domaines d'évaluation » (OCDE, 2003). Certains exercices sont présentés dans ce chapitre à titre d'exemples, le second volume du rapport regroupant l'ensemble des exercices analysés.

A. Organisation générale du cadre de référence

Le contenu du cadre évolue au fil des enquêtes pour intégrer de nouveaux domaines et de nouveaux modes d'évaluation, pour tenir compte de l'évolution de certaines conceptions et aussi pour tenir compte des remarques et critiques émises par les représentants des pays participants, par les chercheurs et par les autres groupes concernés. Pour un domaine donné (mathématiques, compréhension de l'écrit, ...) il n'est toutefois significativement modifié que lorsque ce domaine est la majeure du cycle concerné (2003 et 2012 pour les mathématiques). Ainsi, avant chacun des cycles de PISA, un document cadre est publié qui précise ou rappelle :

- les objectifs de l'étude à venir ;
- les notions clés ;
- les objets de l'évaluation selon les types de littératie concernés (écrit, mathématique, scientifique, etc.) ;
- les dimensions de l'évaluation (contenus, processus, contextes, etc.), préparant ainsi les échelles qui seront utilisées dans les rapports ;
- l'équilibre de l'évaluation (nombre et répartition des questions à préparer dans les diverses catégories, c'est à dire les tables de spécification).

Le cadre présente aussi les questionnaires de contexte. Ces questionnaires, qui ne sont pas, en général, liés à un domaine particulier, constituent une partie très importante de l'étude. C'est aussi à partir de ces questionnaires que sont construits toute une série d'indicateurs qui seront utilisés dans les rapports PISA mais aussi dans de nombreuses autres études de l'OCDE. Parmi ces indicateurs, citons l'indice de niveau socio-économique et culturel des élèves, les indices concernant la motivation des élèves, la confiance en soi, l'équipement des établissements, leur niveau d'autonomie. Ce sont ainsi des dizaines d'indices qui sont construits à partir des réponses des élèves, et de celles des chefs d'établissement. Les rapports PISA sont fortement centrés sur l'analyse des réponses à ces questionnaires et sur les mises en relation des résultats des tests cognitifs avec les réponses à ces questionnaires. Ainsi, pour le cycle PISA 2015, 6 questionnaires ont été proposés :

- le questionnaire « Établissement », administré aux chefs d'établissement (12 pages) ;
- le questionnaire « Élève », administré à tous les élèves participants (17 pages) ;
- deux questionnaires facultatifs à l'intention des élèves :
 - le questionnaire sur le parcours scolaire (second questionnaire élèves) (9 pages),
 - le questionnaire sur la maîtrise des technologies de l'information et de la communication (questionnaire TIC - 5 pages) ;
- deux autres questionnaires facultatifs :
 - le questionnaire « Parents » (9 pages) ;
 - le questionnaire « Enseignants » (16 pages).

Tous ces questionnaires représentent, dans leur version en français, 68 pages pour plus de 200 items. En 2015, comme en 2012, la France n'a administré que les deux premiers questionnaires. C'est à partir de ces questionnaires²⁰ que PISA et l'OCDE réalisent une série de corrélations entre les résultats des élèves et leurs caractéristiques individuelles ou des caractéristiques des contextes d'apprentissage, qui peuvent

²⁰ Ils contiennent cependant des parties destinées à être mises en relation avec certains domaines de l'étude (par exemple les items relatifs au niveau d'anxiété par rapport aux mathématiques).

parfois être interprétés comme des liens de causalité, malgré les précautions mentionnées dans les rapports. Dans certains cas, le cadre de référence présente aussi des exemples d'exercices choisis parmi ceux des enquêtes précédentes et ceux utilisés dans les pré-enquêtes (« *field trials* »).

B. Le cadre de référence de PISA pour le volet mathématique

Le cadre de référence pour le volet mathématique a été précisé pour PISA 2003 ; il vaut pour l'ensemble des cycles PISA depuis la première enquête (2000). Le groupe d'experts mathématiques de PISA, responsable du développement du cadre de référence, était alors présidé par le professeur Jan de Lange de l'Institut Freudenthal d'Utrecht (Pays-Bas). Avec Mogens Niss de l'université de Roskilde (Danemark), Jan de Lange apporta la majeure partie des idées développées ensuite, en particulier celles liées à l'éducation mathématique en lien avec le réel (« *Realistic Mathematical Education* ») et celles concernant les compétences mathématiques. Contrairement donc à ce qui a souvent été écrit, les influences principales, du moins en ce qui concerne le volet mathématique de PISA, ont été largement nord-européennes (issues du courant constructiviste), et non nord-américaines.

Pour le domaine mathématique, le cadre de référence de PISA 2003 a repris sans modification notable ce qui avait été développé à partir de 1997 pour PISA 2000, jusqu'en 2012. Ce cadre a été révisé pour PISA 2012. Le cadre de 2015 a repris celui de 2012 avec pour seules modifications celles concernant le mode de passation : passage de l'administration de tests papier-crayon à l'administration sur ordinateur. De ce fait, il est pertinent de se limiter à la description des cadres 2003 et 2012, et aux modifications apportées lors du passage de l'un à l'autre. Pour le domaine de la littératie scientifique, le cadre de 2006 (première majeure) a été révisé pour PISA 2015 (seconde majeure).

Avant d'aller plus loin, rappelons ici que l'enjeu des cadres de référence est d'abord de définir et d'éclairer la notion de littératie dans les différents domaines de l'enquête telle que l'OCDE la conçoit afin de permettre une évaluation aussi valide que possible de cette littératie.

Pour rappel, la définition initiale de la littératie mathématique pour PISA était la suivante :

PISA 2003 : La littératie mathématique est la capacité d'un individu à identifier et à comprendre le rôle que les mathématiques jouent dans le monde, à produire des jugements bien fondés, à utiliser et à impliquer les mathématiques, en fonction des besoins de sa vie en tant que citoyen constructif, impliqué et réfléchi. (Traduction d'Antoine Bodin de la définition de 2003)

Cette définition a souvent été souvent critiquée par des commentateurs qui estimaient qu'elle accordait une place insuffisante aux contenus mathématiques. En réponse à ces critiques, la définition de 2012, maintenue en 2015, met davantage l'accent sur la place et l'importance dans la littératie mathématique des contenus, des concepts et des démarches strictement mathématiques.

PISA 2012 : la littératie mathématique est la capacité d'un individu à **formuler, employer et interpréter** des mathématiques dans une variété de contextes. Cela inclut la capacité à raisonner mathématiquement et à utiliser des concepts, des procédures, des faits et des outils mathématiques pour décrire, expliquer et prévoir des phénomènes. Elle aide les individus à reconnaître le rôle que les mathématiques jouent dans le monde, à produire des jugements bien fondés et à prendre les décisions nécessaires en citoyens constructifs, engagés et réfléchis. (OECD 2012 - Traduction d'A. Bodin)

Cette définition prépare à la fois le classement des questions selon trois catégories de processus (formuler, employer, interpréter, voir chapitre 2 B.3), et la présentation des résultats selon ces catégories. La littératie mathématique est désormais analysée selon trois aspects interdépendants :

1. **Les processus mathématiques**, qui décrivent ce que font des individus pour établir un lien entre le contexte du problème et les mathématiques et, donc, pour résoudre le problème, ainsi que les capacités qui sous-tendent ces processus ;
2. **Les contenus mathématiques** qui pourraient être utilisés pour le traitement des questions ;
3. **Les contextes** dans lesquels les questions s'inscrivent.

Ces trois aspects et les catégories qu'ils sous-tendent sont utilisés pour établir la table de spécification des questions, qui précise les pourcentages de questions à prévoir dans chacune des catégories (voir chapitre 2 B.3), et pour rapporter les résultats (constitution d'échelles).

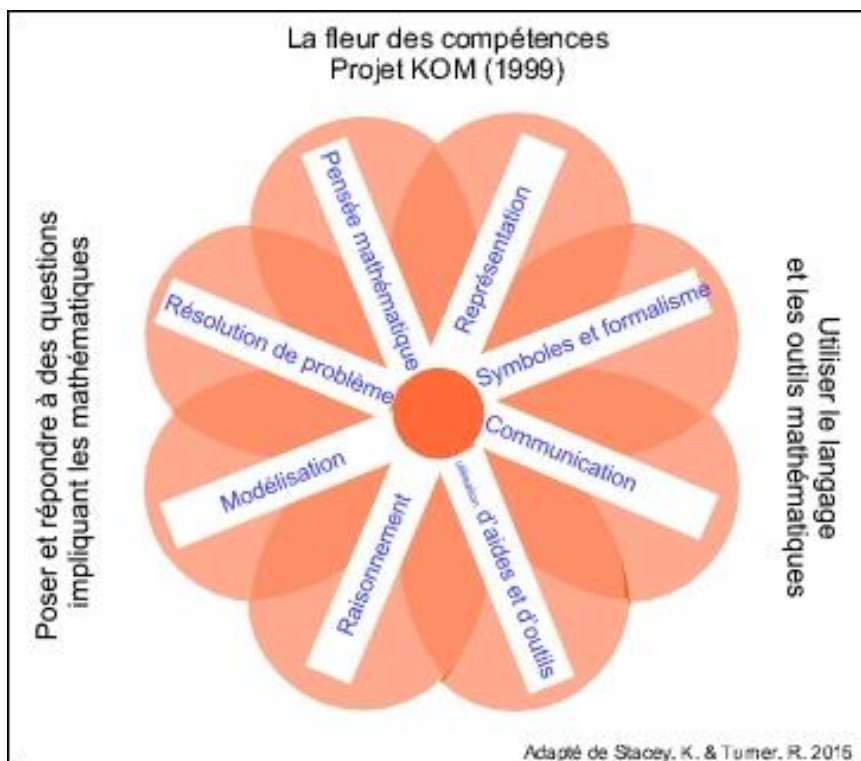
1. Processus et compétences

Ainsi que nous l'avons dit, les enquêtes PISA s'intéressent autant aux connaissances qu'à la façon de les mobiliser. De ce fait, la question des processus est l'élément central du cadre de référence. Cette question, par ailleurs, est étroitement liée à celle des compétences.

Le sujet est complexe et a été l'objet de nombreuses polémiques. Soumise aux critiques externes, mais aussi à la réflexion interne et ayant été l'objet de nombreuses réunions de travail, la façon de décrire, d'organiser et d'évaluer les processus et les compétences est celle qui a le plus évolué entre 2000 et 2015. Il est difficile de bien comprendre les rapports et les analyses de PISA sans connaître au moins les grandes lignes de cette évolution.

Au départ, le modèle, issu de recherches et d'expériences menées aux Pays-Bas et au Danemark identifiait huit « compétences » qui se recouvraient partiellement. La représentation en forme de fleur proposée ci-dessous exprimait bien le fait que ces compétences n'étaient pas supposées indépendantes.

Figure 2 : La Fleur des compétences – Projet KOM (1999)



Ce modèle était bien adapté à l'enseignement et, certainement à l'évaluation en classe, mais il l'était moins pour une évaluation standardisée dans laquelle, pour satisfaire à des critères psychométriques, il convenait de faire émerger des variables aussi indépendantes que possible.

Pour PISA 2000, et ensuite jusqu'à PISA 2015, sauf en ce qui concerne la **modélisation** sur laquelle nous reviendrons, cette liste de « compétences » est restée stable à quelques détails de dénomination près, ainsi que l'on peut le voir dans le tableau suivant.

Tableau 2 : Processus, compétences et aptitudes fondamentales dans PISA (2000-2015)

La fleur des compétences 1999	PISA 2000 à PISA 2009	PISA 2012 et PISA 2015
	Processus en 2000 Compétences en 2003, 2006 et 2009 (<i>Processes puis competencies</i>)	Les aptitudes fondamentales (<i>Fundamental capabilities</i>)
• Pensée mathématique	• Pensée et raisonnement mathématique	• Raisonnement et argumentation
• Raisonnement	• Argumentation	
• Communication	• Communication	• Communication
• Modélisation	• Modélisation	• Mathématisation
• Résolution de problèmes	• Création et résolution de problèmes.	• Conception de stratégies de résolution de problèmes
• Représentation	• Représentation	• Représentation
• Symboles et formalisme	• Utilisation d'un langage et d'opérations de nature symbolique, formelle et technique	• Utilisation d'opérations et d'un langage symbolique, formel et technique
• Utilisation d'aides et d'outils	• Utilisation d'instruments et d'outils	• Utilisation d'outils mathématiques

Source : Cadres de référence de PISA 2000 à 2015.

Cette liste concerne les opérations mentales et les activités qui sont en jeu dans l'activité mathématique et cherche à les classer. De ce fait, parler de **processus** était sans doute plus adéquat que de parler de **compétences**. Les critiques portées sur l'utilisation du terme compétence dans ce contexte a conduit, en 2012, à lui substituer les « *fundamental capabilities* », expression qu'on traduira par la suite par **aptitudes fondamentales**²¹.

²¹ Note d'Antoine Bodin : le terme utilisé en anglais est « *capabilities* ». Dans sa présentation en français, l'OCDE le traduit par « facultés », ce qui donne une connotation à caractère psychologique peu sensible aux apprentissages. Une traduction par « aptitudes » paraît plus adaptée, mais cela reste discutable.

2. Les groupes de compétences (PISA 2000 à PISA 2009)

De PISA 2000 à PISA 2009, pour alléger les critères concernant les questions et pour limiter le nombre de variables destinées à constituer les échelles de performance, il avait paru nécessaire de regrouper les compétences (processus en 2000) en trois groupes hiérarchisés :

- **Groupe 1 : Reproduction** (reproduction, définitions et calculs) ;
- **Groupe 2 : Connexions** (connexions et intégration pour la résolution de problèmes) ;
- **Groupe 3 : Réflexion** (mathématisation, pensée mathématique, généralisation et *insight*²²).

Ces groupes étaient définis de la façon suivante (OCDE, 2003b) :

Le groupe « Reproduction »

« Les compétences classées dans ce groupe impliquent essentiellement la reproduction de connaissances déjà bien exercées – en particulier, celles qui sont les plus communément sollicitées dans les tests d'évaluation normalisés et les évaluations périodiques en classe : connaissance de faits, représentations de problèmes courants, identification d'équivalences, mémorisation de propriétés et d'objets mathématiques familiers, exécution de procédures routinières, application d'algorithmes et de savoir-faire techniques usuels, utilisation d'énoncés contenant des symboles et des formules standard, et réalisation de calculs. »

Les items²³ utilisés pour évaluer les compétences du groupe reproduction peuvent être décrits au moyen de deux expressions clés : la reproduction d'acquis et l'exécution d'opérations de routine. »

Le groupe « Connexions »

« Les compétences du groupe connexions sont dans le prolongement de celles du groupe reproduction, dans la mesure où elles servent à résoudre des problèmes qui ne sont plus de simples routines, mais qui continuent à impliquer un cadre familier ou quasi-familier. »

Les items²⁴ relevant de ce groupe de compétences exigent habituellement que l'élève fasse preuve de sa capacité à intégrer et relier des éléments appartenant à diverses idées majeures ou à divers domaines mathématiques, ou qu'il puisse mettre en relation diverses représentations d'un problème. »

Les items²⁵ utilisés pour évaluer les compétences du groupe connexions peuvent être décrits au moyen des expressions clés suivantes : l'intégration, la mise en relation et un (modeste) degré de transfert au-delà de l'acquis. »

²² La traduction de « *insight* » dans les documents en français de l'OCDE est « compréhension en profondeur ». En réalité, la notion d'*insight* a davantage à voir avec la pensée latérale, l'intuition et la créativité qu'avec la compréhension au sens habituel. Le terme est en général considéré comme intraduisible. Il figure cependant aujourd'hui dans le Larousse (en ligne) et dans le dictionnaire de l'office canadien de la langue française avec la définition « Saisie soudaine de la solution d'un problème après une période plus ou moins longue de tâtonnement ».

²³ Lire « questions »

²⁴ Idem.

²⁵ Idem.

Le groupe « Réflexion »

« Les activités cognitives associées à ce groupe demandent aux élèves de faire preuve d'une démarche mentale réfléchie lors du choix et de l'utilisation de processus pour résoudre un problème. Elles sont en rapport avec les capacités auxquelles les élèves font appel pour planifier des stratégies de solution et les appliquer dans des situations-problème qui contiennent plus d'éléments que celles du groupe connexions, et qui sont plus « originales » (ou peu familières).

Les items utilisés pour évaluer les compétences du groupe réflexion peuvent être décrits au moyen des expressions clés suivantes : le raisonnement approfondi, l'argumentation, l'abstraction, la généralisation et la modélisation appliqués à de nouveaux contextes. »

Le terme de « groupe » ou de « classe » employé ici était trompeur, dans la mesure où chacune des huit compétences étaient supposées contribuer plus ou moins à chacun des groupes. On pouvait alors proposer la représentation ci-dessous.

Figure 3 : Illustration du poids relatif de chacune des compétences dans les groupes de compétences

Groupes de compétences (*)			Compétences
1 Reproduction	2 Connexions	3 Réflexion	
			Pensée et raisonnement mathématique
			Argumentation
			Communication
			Modélisation
			Création et résolution de problèmes
			Représentation
			Utilisation d'un langage et d'opérations de nature symbolique, formelle et technique
			Utilisation d'instruments et d'outils

Source : OCDE (2013c)

Lecture : (*) Il s'agit des compétences jusqu'à 2012, mais l'adaptation aux aptitudes fondamentales de 2012 est immédiate.

La coloration plus ou moins foncée des cases indique l'implication plus ou moins grande de chaque compétence dans le groupe concerné.

Le terme de processus revient avec PISA 2012, mais cette fois pour désigner des catégories « sous-tendant les aptitudes fondamentales » (c'est-à-dire, à peu près, les compétences de 2003). La différence essentielle avec les anciens groupes de compétences est que ces processus ne sont pas hiérarchisés, ce qui permettra de les rapporter à des échelles *a priori* indépendantes.

Les processus retenus sont :

- **Formuler** : Formuler des situations de façon mathématique ;
- **Employer** : Employer des concepts, faits, procédures et raisonnements mathématiques ;
- **Interpréter** : Interpréter, appliquer et évaluer des résultats mathématiques.

« Ces trois verbes, « formuler », « employer » et « interpréter », constituent à eux seuls une structure signifiante qui permet de définir les processus mathématiques qui décrivent ce que les individus font pour établir un lien entre le contexte d'un problème et les mathématiques et, donc, pour résoudre le problème. Les épreuves de mathématiques du cycle PISA 2012 permettront pour la première fois de rendre compte des résultats des élèves en fonction de ces processus mathématiques, une structure qui fournira des catégories utiles et pertinentes pour l'action publique. » (OCDE, 2013c)

Le paragraphe suivant précise le sens à donner à ces termes.

3. Mathématisation et modélisation

La littératie mathématique selon PISA se résume donc à la capacité à traiter des situations qui peuvent se prêter à un traitement mathématique ; de plus ces situations doivent être susceptibles d'être rencontrées dans le monde « réel »²⁶. Ces situations se présentent donc sous la forme de problèmes dont la résolution ne se limite pas à la simple application d'une connaissance mathématique. Si l'on considère, ce qui est le cas pour la plupart des mathématiciens, que « faire des mathématiques » c'est résoudre des problèmes et que, « *la résolution de problèmes est la source et le critère du savoir* » (Vergnaud, 1981), il n'y a là rien de particulièrement choquant, même si les élèves en France sont sans doute plus souvent confrontés à des exercices d'application qu'à des problèmes pouvant relever du cycle complet de modélisation (ou de mathématisation) défini ci-dessous²⁷.

Une expression revient souvent dans les enquêtes PISA : celle de « vie réelle » (*real life*). L'évaluation du domaine mathématique porte alors sur l'aptitude à utiliser ses connaissances et ses savoir-faire dans des situations dans lesquelles l'utilisation des connaissances mathématiques, aussi minimales soient-elles, suppose un traitement préalable passant d'abord par la compréhension de la situation proposée, très rarement située dans le domaine mathématique. Suit la traduction en langage mathématique, le traitement mathématique, et finalement l'interprétation des résultats par un retour au « monde réel ».

C'est ce que PISA a appelé, jusqu'à PISA 2012, le cycle de mathématisation, défini de la façon suivante (OCDE, 2003b) :

1. *Commencer par un problème relevant de la réalité ;*
2. *Organiser le problème en fonction de concepts mathématiques ;*
3. *Effacer progressivement la réalité au travers de divers processus, tels que la formulation d'hypothèses concernant l'identification des principales caractéristiques du problème, la*

²⁶ L'expression « *real life problems* » a été officiellement traduit par « problèmes de la vie courante ». Cependant, la vie « réelle » ne se limite pas à la vie courante telle qu'elle est souvent comprise.

²⁷ Il est toutefois à noter une évolution des énoncés des épreuves finales de mathématiques du diplôme national du brevet depuis 2013.

généralisation et la formalisation (dont l'objectif est de faire ressortir les caractéristiques mathématiques de la situation et de transformer le problème réel en un problème mathématique qui soit le reflet fidèle de la situation) ;

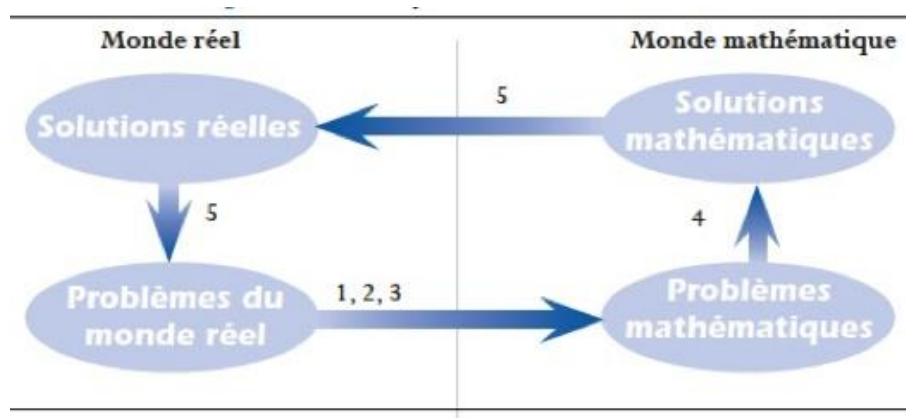
4. *Résoudre le problème mathématique ;*

5. *Comprendre la solution mathématique et l'appliquer à la situation réelle (ce qui implique aussi d'identifier les limites de la solution).*

On voit que PISA excluait *a priori* du périmètre de ses évaluations la compétence à résoudre des problèmes intra-mathématiques, ce qui l'opposait résolument à TIMSS. Cette présentation du cycle de mathématisation, qui date de 2000, n'a pas vraiment été modifiée par les changements qui ont suivi. Les points 1 à 5 se retrouvent en effet inchangés dans la présentation de 2012-2015. Toutefois, PISA admet que, pour des raisons psychométriques, il est rarement possible de produire des questions indépendantes qui mettraient en jeu l'ensemble du cycle. Certaines questions peuvent donc ne mettre en jeu que les étapes 1 et 2 ou 2,4 et 5, etc. Dans un ou deux cas, on trouve une question ne mettant en jeu que l'étape 4, mais ce n'est pas du tout dans l'esprit de PISA.

Dans la figure 4 ci-dessous, les numéros 1 à 5 renvoient à la description précédente.

Figure 4 : Le cycle de mathématisation avant 2012



Source : OCDE (2003)

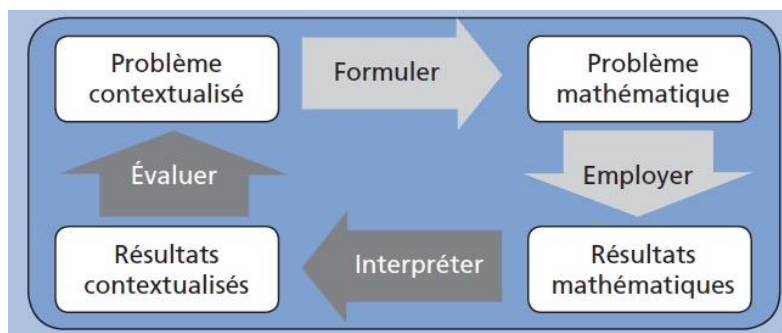
Le point 5 apparaît deux fois : « une première fois lors du passage de la solution mathématique à une solution réelle, et une seconde fois, lorsque cette solution est mise à son tour en relation avec le problème original appartenant au monde réel » (OCDE, 2003). À l'évidence, le terme de mathématisation utilisé ci-dessus pour désigner l'ensemble du processus de résolution d'un problème n'est pas bien choisi. En effet, ce terme renvoie habituellement aux seuls points 2 et 3 de la figure 4.

Avec PISA 2012, on assiste à un changement de vocabulaire conduisant pratiquement à permuter les termes de mathématisation et de modélisation. La mathématisation devient l'une des sept aptitudes fondamentales tandis que la modélisation est promue au rang de principe organisationnel des nouveaux processus :

1. Le triplet (1 ; 2 ; 3) devient le processus **FORMULER**.
2. Le point 4 devient le processus **EMPLOYER**.
3. Le couple (5 ; 6) devient le processus **INTERPRÉTER**

La séquence Formuler, Employer, Interpréter, devient le cycle de modélisation.

Figure 5 : Le cycle de modélisation à partir de 2012



Source : OCDE (2013)

Ainsi que le montre la présentation du cycle de mathématisation de PISA 2003, ces catégories de processus ne contredisent pas les catégories utilisées lors des cycles précédents ; elles les affinent et, surtout, elles définissent des variables « mesurables » qui pourront générer des échelles de compétences lors des traitements des résultats (On notera le changement de sens des flèches du diagramme, mais cela est sans signification particulière).

Ces trois processus (formuler, employer et interpréter) sont des composantes majeures du cycle de modélisation mathématique et aussi des composantes majeures de la définition de la littératie mathématique. Chacun d'eux fait appel aux aptitudes mathématiques fondamentales, lesquelles, à leur tour, font appel aux connaissances mathématiques détaillées de l'élève à propos de sujets spécifiques (OECD, 2013b – traduction d'A. Bodin²⁸).

FORMULER : Cette catégorie concerne la formulation des situations de façon mathématique. Cela « renvoie à la capacité des individus d'identifier et de reconnaître des possibilités d'utiliser les mathématiques dans le contexte d'un problème, puis de structurer sous forme mathématique un problème présenté jusqu'à un certain point sous une forme contextualisée » (*ibidem*). C'est ce que l'on appelle habituellement la mathématisation des situations.

EMPLOYER : Il s'agit d'employer des concepts, des faits, des procédures et des raisonnements mathématiques. C'est là qu'interviennent les connaissances proprement mathématiques. PISA reconnaît ainsi que, non seulement ces connaissances sont nécessaires, mais de plus qu'il faut savoir les utiliser à bon escient. Lors des cycles précédents, il avait souvent été reproché à PISA de ne pas suffisamment prendre en compte les contenus d'enseignement. Avec cette catégorie, PISA tente de prendre en compte cette critique et d'afficher plus clairement la place des connaissances dans le cycle de modélisation.

²⁸ Voici la version officielle en langue anglaise : « *These processes of formulating, employing, and interpreting mathematics are key components of the mathematical modelling cycle and also key components of the definition of mathematical literacy. These three processes each draw on fundamental mathematical capabilities, which in turn draw on the problem solver's detailed mathematical knowledge about individual topics.* »

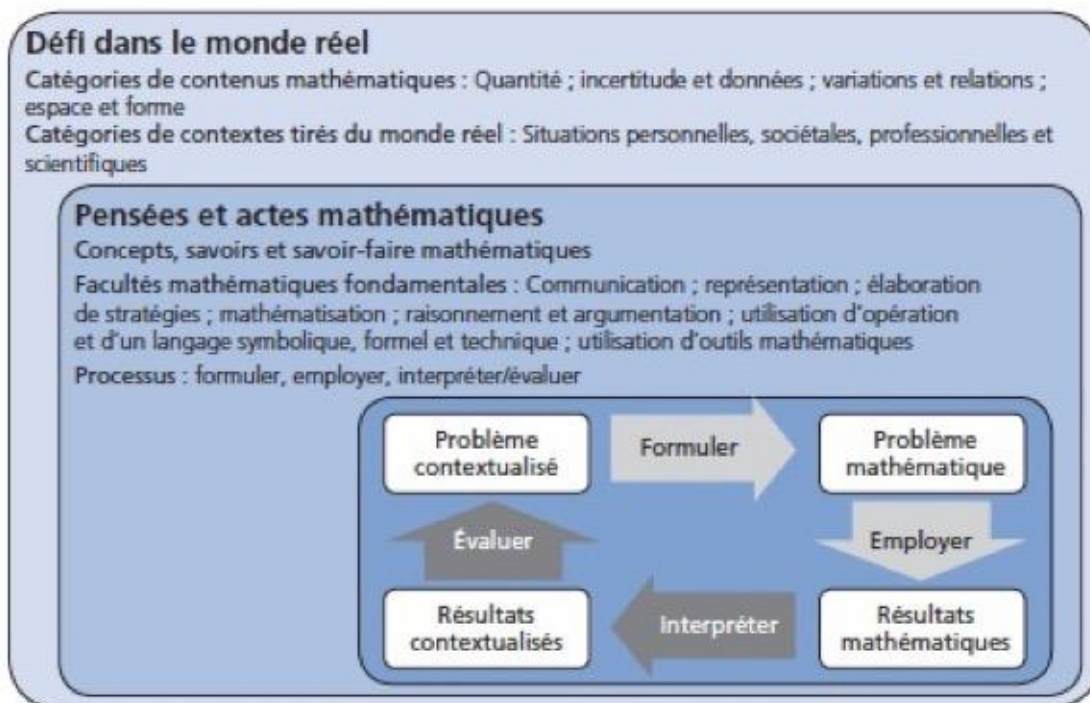
La traduction officielle en langue française est : « *Ces processus qui consistent à formuler, à employer et à interpréter de façon mathématique sont des composantes majeures du cycle de modélisation mathématique et des composantes majeures de la définition de culture mathématique. Ces trois processus reposent sur des facultés mathématiques fondamentales, qui reposent à leur tour sur les connaissances mathématiques détaillées de l'individu à propos de thématiques spécifiques.* »

INTERPRÉTER : Il s'agit d'interpréter, d'appliquer et d'évaluer des résultats mathématiques. Cela « renvoie à la capacité des individus de réfléchir à des solutions, des résultats ou des conclusions mathématiques, et de les interpréter dans le cadre de problèmes tirés du monde réel » (*ibidem*). Cette catégorie concerne donc le sens et la portée que l'élève est capable de donner à ses calculs, à ses raisonnements et à ses résultats.

Finalement, PISA résume son cadre de référence pour l'évaluation de la littératie mathématique par le diagramme ci-dessous. On y voit apparaître l'ensemble des éléments-clés suivants :

- L'ancrage dans le « monde réel » ;
- Les contenus mathématiques ;
- Les contextes (voir plus loin) ;
- Les aptitudes fondamentales ;
- Les processus ;
- Le cycle de modélisation.

Figure 6 : Un modèle de la littératie mathématique en pratique



Source : OCDE (2013)

4. Les domaines de contenus

Plutôt que de s'attacher à un découpage traditionnel en France (et scolaire) des contenus, le cadre de référence de PISA met l'accent sur des grandes idées mathématiques.²⁹

« Les contenus se rapportent à quatre idées majeures (les variations et les relations ; l'espace et les formes ; la quantité ; l'incertitude et les données) qui sont liées aux disciplines mathématiques (telles que l'arithmétique, l'algèbre et la géométrie), et qui se chevauchent de façon complexe. » (OCDE, 2013c)

Il s'agit là d'une approche de type phénoménologique qui s'oppose à l'approche disciplinaire. Dans cette approche, les phénomènes sont premiers et les concepts et les théories mathématiques sont élaborés pour les aborder et pour les traiter.

On le sait, « les curriculums de mathématiques sont habituellement organisés en chapitres », ou domaines enseignés. PISA justifie son approche non curriculaire en remarquant que « ces divisions ont pour effet de compartimenter les mathématiques, et d'accorder une importance exagérée aux techniques de calcul et aux formules ».

« Au début du XX^e siècle, on pouvait raisonnablement envisager les mathématiques comme un ensemble formé d'une douzaine de matières distinctes (arithmétique, géométrie, algèbre, calcul, etc.). De nos jours, le nombre de matières à évoquer serait plutôt de soixante à soixante-dix. Certains domaines, comme l'algèbre ou la topologie, ont été scindés en divers sous-domaines. D'autres, comme la théorie de la complexité ou la théorie des systèmes dynamiques, sont des sujets d'étude entièrement neufs. Pour être pertinentes, les enquêtes sur les mathématiques doivent donc être à l'image des structures complexes du monde qui nous entoure. » (PISA, 2000)

Les catégories de contenus retenues par PISA sont donc les suivantes :

- **Variations et relations ;**
- **Espace et formes ;**
- **Quantité ;**
- **Incertitude et données.**

Ces catégories sont restées inchangées depuis PISA 2000. Toutefois, sans que cela en modifie l'esprit, leur présentation a évolué pour mieux mettre en évidence l'importance des contenus enseignés et l'adéquation existant entre ces catégories et les objectifs définis par les programmes des pays participants³⁰.

- La catégorie Variations et relations englobe tous les types de changements que l'on peut rencontrer. Cela concerne aussi bien les changements continus, que ceux, de nature discrète, qui procèdent par sauts (tel l'évolution au cours d'une journée du nombre de voyageurs d'une ligne d'autobus : chaque arrêt correspond à un saut). La modélisation de ces changements peut impliquer des fonctions, des équations, ainsi que des représentations graphiques et symboliques.

²⁹ Le découpage de PISA est dérivé des recommandations de la Société Mathématique Américaine - AMS (cf. « *On the shoulders of the Giants : new approaches to numeracy* » qui décrit les idées majeures reprises par PISA). Il s'agit d'une organisation du type de celle connue en France sous le titre d'organisation « par problématiques » (cf. Gras, R. et APMEP).

³⁰ La présentation des contenus de PISA 2012 s'attache aussi à mettre en évidence la place aujourd'hui prise par les concepts et outils numériques.

- La catégorie Espace et formes fait référence aux objets et aux phénomènes que l'on rencontre dans notre environnement. La géométrie est évidemment la discipline mathématique de référence de cette catégorie, mais ne s'y réduit pas. Alors que la géométrie est souvent vue comme une théorie formelle et abstraite, dans laquelle la notion de démonstration tient une place privilégiée, la catégorie Espace et formes procède essentiellement d'une approche sensualiste et phénoménologique.
- La catégorie Quantité concerne l'attribution de nombres et de mesures aux phénomènes observés. Sont donc concernés ici le dénombrement, le calcul et, en particulier le calcul mental, le mesurage, la question des ordres de grandeur etc. Le sens des nombres ainsi que les différentes représentations des nombres sont des éléments clés de cette catégorie.
- La catégorie Incertitude et données intègre les phénomènes dont on ne peut pas prédire l'issue de façon certaine. Très nombreux, ils ont pris une place très importante dans l'activité scientifique comme dans la vie des sociétés. Le cas qui vient immédiatement à l'esprit est celui des jeux de hasard, mais de nombreux phénomènes physiques ou sociétaux obéissent à des processus probabilistes dont l'analyse s'appuie sur des données statistiques. Bien sûr, la statistique et les probabilités sont les théories mathématiques de référence de cette catégorie, mais la « grande idée » retenue ici est celle d'incertitude.

Ainsi, PISA a cherché à échapper à la logique des contenus traditionnels. Cela n'a pas été sans susciter de nombreux débats qui l'ont en particulier amené à introduire des références plus explicites à ces contenus dans le cadre de référence de 2012. Toutefois, dès l'enquête 2000, les contenus traditionnels n'avaient pas été totalement ignorés. On lit en effet dans le cadre de référence de PISA 2000 :

« L'aspect des 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. » (OCDE, 1999)

Ces références ont ensuite disparu des cadres de référence jusqu'à celui de 2012 dans lequel ils réapparaissent en force.

5. Les contextes

Voulant inscrire ses questions d'évaluation de la littératie non seulement dans le « monde réel », mais aussi dans la vie réelle des élèves, PISA distingue et hiérarchise les types de situations dans lesquelles les problèmes sont rencontrés. Jusqu'à PISA 2012, les cadres de référence précisent que les situations peuvent être situées dans une grande variété de contextes et distinguent des types de situations dans lesquels les contextes peuvent varier (adaptation libre de OECD, 2003) :

- Situation la plus proche de l'élève : sa vie personnelle ou familiale ;
- Situation plus éloignée : sa vie scolaire et le monde du travail et des loisirs ;
- Situation encore plus éloignée : la communauté locale et la société nationale ou mondiale ;
- Et enfin, beaucoup plus loin : les situations de nature scientifique.

Avec le cadre de référence de 2012, les types de situations deviennent les **catégories de contextes** :

- Contextes personnels ;
- Contextes professionnels ;
- Contextes sociétaux ;
- Contextes scientifiques.

Le tableau 3 récapitule les trois processus que les situations proposées aux élèves visent à évaluer, les quatre domaines mathématiques sur lesquels ces situations portent et les quatre types de contextes dans lesquels elles se placent.

Tableau 3 : Tableau synthétique des catégories de PISA 2015 en littératie mathématique

Domaines de l'évaluation : trois aspects	Catégories de chacun des domaines de l'évaluation
PROCESSUS	FORMULER
	EMPLOYER
	INTERPRÉTER
CONTENUS	VARIATIONS ET RELATIONS
	ESPACE ET FORMES
	QUANTITÉ
	INCERTITUDE ET DONNÉES
CONTEXTES	PERSONNELS
	PROFESSIONNELS
	SOCIÉTAUX
	SCIENTIFIQUES

Source : adapté de OCDE (2013c)

Le cadre de référence de PISA a en particulier pour objectif de prévoir la répartition des questions d'évaluation selon les diverses catégories de processus, de contenus, et de contextes. Ce sont ces catégories qui feront l'objet du rapport final et qui constitueront autant d'échelles dans les rapports de synthèse, devenant ainsi des « dimensions de l'étude » pour les statisticiens. Il a aussi pour objectif de produire des plans de répartitions des questions selon les divers formats (QCM, questions ouvertes,...). L'ensemble de ces deux plans constitue la table de spécification. Cette table est définie *a priori* et devient alors le cahier des charges qui commande l'élaboration et la sélection des questions. Par construction, il n'y a que des différences minimales entre les tables de spécification et les tableaux que l'on peut construire après l'évaluation.

Par exemple³¹, voici ces tableaux pour PISA 2012 :

Tableau 4 : Répartition des questions en fonction des processus dans PISA 2012

	Variations et relations	Espace et formes	Quantité	Incertitude	Total	Poids
Formuler	8	13	6	5	32	30 %
Employer	16	12	16	6	50	46 %
Interpréter	5	2	6	14	27	24 %
Total	29	27	28	25	109	
Poids	27 %	25 %	25 %	23 %		

Source : adapté de OECD (2014b)

³¹ Les questions de 2015 constituent une grande partie de celles de 2012 (environ 80 %).

Tableau 5 : Répartition des questions en fonction des contextes dans PISA 2012

	Personnel	Professionnel	Sociétal	Scientifique	Total	Poids
Formuler	6	9	9	8	32	30 %
Employer	10	10	14	16	50	46 %
Interpréter	5	5	13	4	27	24 %
Total	21	24	36	28	109	
Poids	19 %	22 %	33 %	26 %		

Source : adapté de OECD (2014b)

Tableau 6 : Répartition des formats de questions en fonction des processus dans PISA 2012

	Questions à choix multiples	Questions à ouvertes	Total	Poids
Formuler	10	22	32	30 %
Employer	18	32	50	46 %
Interpréter	17	10	27	24 %
Total	45	64	109	
Poids	41 %	59 %		

Source : adapté de OECD (2014b)

6. Relation des autres volets de PISA avec les mathématiques

Les compétences mathématiques, ne trouvent pas à s'exercer que dans des tâches désignées comme relevant strictement du domaine mathématique. Cela est vrai dans la « vie réelle » qui est la référence principale de PISA ; cela est aussi vrai dans les autres domaines des enquêtes.

Ainsi, certains exercices contiennent aussi des éléments qui pourraient les faire passer pour des exercices classés par PISA dans le domaine mathématique : c'est par exemple le cas pour les trois premières questions de l'exercice « Lac Tchad » présenté ci-dessous.

Le lac Tchad

La figure 1 présente les changements de niveau du lac Tchad, situé au Sahara, en Afrique du Nord. Le lac Tchad a complètement disparu vers 20 000 av. J.-C., pendant la dernière ère glaciaire. Il a réapparu vers 11 000 av. J.-C.. À présent, son niveau est à peu près le même que celui qu'il avait en 1 000 apr. J.-C.

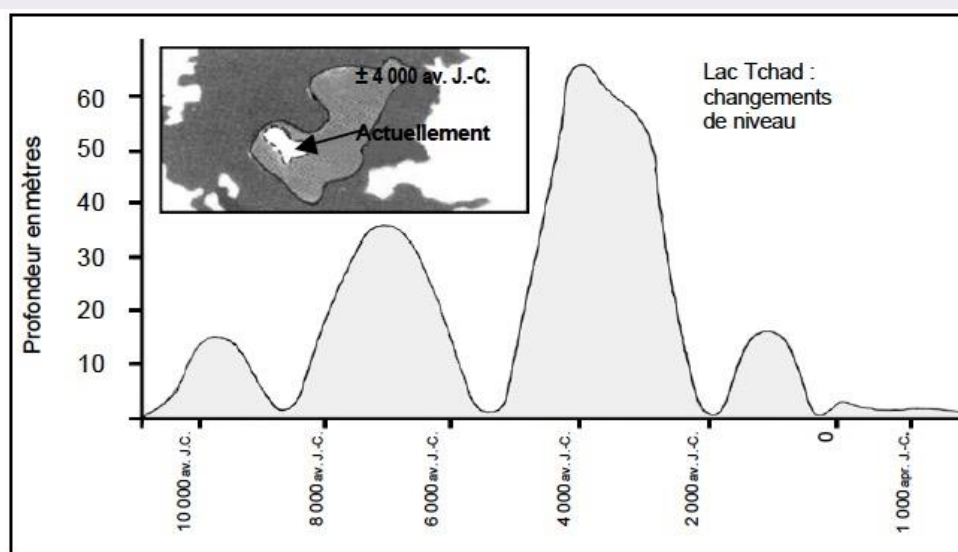


Figure 1

Le lac Tchad - Question 1

Quelle est la profondeur du lac Tchad à présent ?

- A Environ deux mètres.
- B Environ quinze mètres.
- C Environ cinquante mètres.
- D Il a complètement disparu.
- E L'information n'est pas donnée.

Le lac Tchad - Question 2

À peu près en quelle année commence le graphique présenté par la figure 1 ?

Le lac Tchad - Question 3

Pourquoi l'auteur a-t-il choisi de faire commencer le graphique à ce moment ?

Cet exercice est classé dans le domaine de la compréhension de l'écrit. D'autres exercices de ce domaine supposent des lectures de tableaux ou d'organigrammes complexes. Cela n'est aucunement contradictoire avec le projet de PISA, mais cela peut expliquer des proximités à première vue étonnantes dans les résultats des enquêtes selon les domaines (chapitre 5 D).

Pour un lecteur francophone, l'expression « résolution de problèmes³² » peut faire référence aux mathématiques. En réalité, ce domaine, introduit avec PISA 2003, s'est en quelque sorte construit en opposition avec elles. Considérant que les problèmes de la vie n'étaient pas toujours du ressort des mathématiques ou des sciences, l'OCDE a décidé de créer un domaine distinct des autres en excluant, autant que faire se peut, tout ce qui pourrait se rapporter aux mathématiques ou aux sciences. Les questions de ce domaine font en général appel à la logique courante, aux graphes et aux organigrammes. Cependant, la logique courante (connecteurs logiques « et », « ou », « non », « négation d'une assertion », ...) faisait explicitement partie des programmes français de mathématiques du lycée dans les années 1970 ; après en avoir été bannie dans les années 1980 ; elle y est revenue discrètement dans les années 2000 et fut réhabilitée, de façon allégée et moins formalisée, dans les programmes de 2009. Elle est présente au collège, mais seulement implicitement, dans les classes de mathématiques comme dans celles d'autres disciplines. Quant aux graphes et organigrammes, ils sont aussi souvent présents dans les classes de mathématiques. La théorie des graphes est un champ important des mathématiques discrètes et des éléments de cette théorie sont d'ailleurs enseignés dans certaines classes de lycée.

Enfin, il est légitime de s'interroger sur la présence de connaissances mathématiques dans les exercices de littérature scientifique. La section qui suit ainsi que le chapitre 1 du second volume répondront à cette question.

³² Bien que l'objectif du présent rapport ne soit pas de commenter les résultats des enquêtes, on peut signaler que contrairement aux résultats en littérature mathématique, les résultats des élèves français en résolution de problèmes se situent nettement au-dessus de la moyenne de l'OCDE.

C. Le cadre de référence de PISA en littératie scientifique dans PISA 2015

Sans reprendre l'intégralité de la partie consacrée au cadre d'évaluation de la littératie scientifique du document publié par l'OCDE³³ en 2016, il est utile d'indiquer les lignes de force des contextes, des savoirs, des savoir-faire et des attitudes à l'égard de la science retenus dans les épreuves de résolution d'exercices scientifiques proposés aux élèves et de donner au lecteur du présent rapport des éléments pour saisir la manière dont la performance des élèves en sciences est évaluée et présentée par PISA 2015.

Le cadre défini pour les sciences dans PISA 2015 trouve sa source dans celui construit pour PISA 2006, même s'il a été affiné voire élargi³⁴. Il fait de la science une pierre angulaire de l'éducation à la citoyenneté en soutenant l'idée que les programmes (scolaires) de formation doivent faire en sorte « que tous les jeunes deviennent des utilisateurs informés et critiques de la connaissance scientifique ». Selon l'OCDE³⁵, pour comprendre les questions en rapport avec la science et la technologie³⁶, « il faut posséder trois compétences spécifiques au domaine scientifique » (OCDE, 2016, pp. 21-22).

- **Expliquer des phénomènes de manière scientifique** (reconnaître, proposer, évaluer des thèses expliquant une série de phénomènes naturels et technologiques) ;
- **Évaluer et concevoir des recherches scientifiques** (décrire et évaluer des études scientifiques, et proposer des moyens de répondre à des questions de manière scientifique) ;
- **Interpréter des données de manière scientifique** (analyser et évaluer des données, des thèses et des arguments présentés sous diverses formes, et en tirer des conclusions scientifiques appropriées).

Il est précisé que « ces compétences requièrent des connaissances » qui ne doivent pas être exclusivement des **connaissances scientifiques** (même si celles-ci demeurent nécessaires). En effet, pour exploiter les 2^e et 3^e compétences ci-dessus, il paraît nécessaire de posséder certaines « idées relatives à la science », c'est-à-dire « connaître les procédures normalisées à la base des diverses méthodes et pratiques utilisées pour créer le savoir scientifique » (*i.e.* **connaissances procédurales**). Ces compétences requièrent également des connaissances permettant « d'évaluer la crédibilité des thèses avancées et de connaître la signification des termes essentiels tels que « hypothèse », « théorie » et « données » (*i.e.* **connaissances épistémologiques**³⁷). Ces deux connaissances (procédurales et épistémologiques) doivent permettre au

³³ Les lignes qui suivent se réfèrent exclusivement au document complet de l'OCDE : <http://www.oecd.org/fr/publications/cadre-d-evaluation-et-d-analyse-de-l-enquete-pisa-2015-9789264259478-fr.htm> (plus particulièrement entre les pages 19 et 48).

³⁴ À titre d'exemple, en 2006 certaines questions de PISA se donnaient pour but d'évaluer la « valeur accordée à la démarche scientifique » ; tandis qu'en 2015, des questions de PISA visent à évaluer la « valeur accordée aux méthodes scientifiques » (OCDE, p. 25). À noter : la « culture scientifique » constituait le domaine « majeur » de PISA 2006 et le domaine « mineur » des éditions 2009 et 2012. En 2015, la « culture scientifique » est à nouveau domaine majeur d'évaluation.

³⁵ Nous tenons à faire remarquer ici que les auteurs cités dans le document de l'OCDE présentant le cadre conceptuel de l'évaluation PISA 2015 du domaine « culture scientifique » sont, pour nombre d'entre eux, chercheurs en *science education* [éducation scientifique - nous dirions en France, didactique des sciences] et qu'ils publient régulièrement dans les journaux relevant de ce champ de recherche tels que *International Journal of Science Education* ou *Journal of Research in Science Teaching* (pour ne citer que ces deux-là).

³⁶ Le concept de « culture scientifique » renvoie à la fois à la science, définie comme une activité « cherchant à répondre à une question spécifique à propos du monde naturel » et à la technologie qui « a pour but de trouver la solution optimale à un problème humain » (OCDE, p. 20).

³⁷ PISA utilise le terme « *epistemic* » (traduit par épistémique en français) pour ce type de connaissances.

citoyen d'identifier « les questions auxquelles la science peut apporter une réponse, déterminer si des procédures appropriées ont été utilisées pour justifier les thèses avancées et distinguer des questions scientifiques de questions de valeurs ou de considérations économiques » (OCDE, 2016).

Dans ce contexte, « l'enquête PISA 2015 vise à évaluer dans quelle mesure les jeunes de 15 ans sont capables d'utiliser à bon escient les compétences décrites ci-dessus dans un éventail de contextes personnels, locaux/nationaux (classés dans la même catégorie) et mondiaux. Dans les épreuves PISA 2015 ces compétences sont évaluées uniquement sur la base de connaissances que les élèves de 15 ans ont, **selon toute vraisemblance**, déjà acquises » (OCDE, 2016, p. 22, **souligné par les auteurs de cette section**). Ces considérations accordent à l'application des connaissances scientifiques dans des situations « tirées de la vie courante » une place de choix.

1. Contextualisation des questions de science

Les épreuves PISA 2015 de sciences ont pour but d'évaluer des compétences et des connaissances dans des contextes spécifiques (à trois niveaux distincts personnel, local/national et mondial), relevant des domaines suivants : « santé et maladie », « ressources naturelles », « qualité de l'environnement », « risques », « frontières de la science et de la technologie » (OCDE, 2016, p. 25).

Finalement, comme pour les mathématiques, le cadre d'évaluation retenu pour les épreuves de culture scientifique dans l'enquête PISA 2015 repose sur quatre « aspects » interdépendants : les **contextes** au sein desquels les questions se déploient (voir *supra*) et à partir desquels l'individu interrogé doit être capable de déployer les trois **compétences** scientifiques attendues (expliquer, évaluer, interpréter), sachant que « sa façon de faire est influencée par » des **attitudes** (intérêt pour la science, valeur accordée aux méthodes scientifiques, sensibilisation à l'environnement) et des **connaissances** (scientifiques, procédurales et épistémologiques)³⁸.

2. Une nouveauté : les niveaux de charge cognitive

Le document cadre de 2016 de l'OCDE précise que « les épreuves PISA doivent permettre de faire la distinction entre les questions plus faciles et plus difficiles en termes de performances [*i.e* : degré de difficulté], mais également de recueillir des informations sur la mesure dans laquelle les élèves, quel que soit leur niveau de compétence, peuvent résoudre des problèmes dont la charge cognitive varie » (OCDE, p. 41). La notion de « charge cognitive » (appelée également « profondeur des connaissances ») renvoie aux « types de processus mentaux requis » (pour réaliser une tâche donnée). À titre d'exemple, une question peut être difficile parce que les élèves n'ont pas beaucoup de connaissances à son sujet, alors que le processus mental associé relève de la simple remémoration. À l'inverse, une question peut demander un gros effort intellectuel aux élèves si ceux-ci doivent relier de nombreux fragments de connaissances dont il est pourtant facile de se souvenir. PISA 2015 se donne pour but de permettre l'opérationnalisation d'une

³⁸ Le document « cadre » de l'OCDE présente des grilles qui explicitent les différentes compétences en jeu dans les questions PISA 2015 relevant de la culture scientifique et la répartition souhaitée des questions de sciences entre les compétences. Il présente également les grilles relatives aux connaissances.

hiérarchie cognitive. C'est-à-dire que les questions doivent pouvoir être identifiées selon leur degré de charge cognitive :

- **Charge cognitive peu élevée** : effectuer une procédure en une étape, par exemple, se remémorer un fait, un terme, un principe ou un concept, ou localiser un seul fragment d'information dans un graphique ou un tableau (8 % des questions³⁹).
- **Charge cognitive moyenne** : utiliser et appliquer des connaissances concernant des concepts pour décrire ou expliquer des phénomènes, choisir des procédures appropriées de plus d'une étape, structurer/représenter des données, interpréter ou utiliser des groupes de données ou des graphiques simples (30 % des questions).
- **Charge cognitive élevée** : analyser des informations ou des données complexes ; résumer ou évaluer des faits ; justifier et raisonner sur la base de plusieurs sources ; élaborer un plan ou une série d'étapes pour aborder un problème⁴⁰ (61 % des questions).

Figure 7 : Grille permettant de classer les questions en fonction des connaissances, des compétences et de la charge cognitive mobilisée

		Compétences			Profondeur des connaissances		
		Expliquer des phénomènes de manière scientifique	Évaluer et concevoir des recherches scientifiques	Interpréter des données et des faits de manière scientifique	Faible	Moyenne	Élevée
Connaissances	Connaissances scientifiques						
	Connaissances procédurales						
	Connaissances épistémiques						

Source : OCDE (2016)

Lecture : pour être complète, cette grille devrait également inclure une 4^e dimension rendant compte des « attitudes » évaluées (degré d'intérêt des élèves pour la science - voir ci-dessus).

Au final, on parvient à la formalisation d'un projet de description⁴¹ de six « niveaux de culture scientifique » (voir pp. 46-47 du cadre conceptuel) qui représentent des « niveaux de performances » repérés par les critères suivants :

³⁹ Les pourcentages sont arrondis, ce qui explique que leur somme soit légèrement différente de 100 %.

⁴⁰ Il est fait référence ici aux niveaux de Webb modifiés (*Depth of knowledge*) de 1997.

⁴¹ Ce projet est, comme son nom l'indique, construit en amont de l'enquête afin d'élaborer un cahier des charges clair aux concepteurs des exercices. En réalité, il s'appuie fortement sur les résultats de l'enquête 2006, et présentera certainement peu de différences avec l'échelle de compétences construite *a posteriori*.

- Le nombre et la complexité des connaissances visées ;
- La mesure dans laquelle les élèves possèdent les connaissances scientifiques, procédurales et épistémologiques requises, et dans laquelle ces connaissances leur sont familières ;
- Les opérations cognitives à effectuer pour répondre aux questions (se remémorer, analyser, évaluer, etc.) ;
- La mesure dans laquelle la production de la réponse dépend de modèles ou de concepts scientifiques abstraits.

Il est donc indéniable qu'un effort considérable a été porté par les auteurs du cadre conceptuel de l'évaluation PISA 2015 de la culture scientifique pour lier les tâches attendues des élèves et des éléments de culture scientifique dont la définition fait l'objet d'un travail de délimitations (épistémologique, procédurale, contextuelle, comportementale et cognitive) d'une remarquable finesse. La notion de « culture scientifique » prend un sens précis et critérié et donc, opérationnalisable dans des tâches qu'il s'agit maintenant d'analyser au regard de ces critères et des attendus explicites.

Enfin, considérant que les attitudes des élèves ont un rôle important dans l'intérêt qu'ils accordent au domaine scientifique, PISA évalue également les attitudes vis-à-vis de la science à travers trois domaines dans un questionnaire « Élève » :

- l'intérêt pour la science et la technologie ;
- la sensibilisation à l'environnement ;
- la valeur accordée aux méthodes scientifiques.

Chapitre 3 : Le cadre de référence des enquêtes TIMSS

Le cadre de référence de TIMSS a les mêmes fonctions que celui de PISA (auquel il a servi de modèle). Pour rappel, les documents de TIMSS ne sont pas à ce jour traduits en français ; toutes les citations utilisées sont donc des traductions d'Antoine Bodin (comme pour les exercices – sauf ceux de 2015, que les élèves français ont passés).

A. Organisation générale du cadre de référence de TIMSS

Les cadres de référence successifs de TIMSS ont peu changé depuis la première enquête de 1995 : ils ont juste été affinés et précisés. La France n'ayant participé à TIMSS, partiellement, qu'en 1995 et en 2015, le rapport se limite essentiellement à ces deux enquêtes. D'une façon générale, les cadres de référence de TIMSS sont nettement moins détaillés que ceux de PISA, et sont essentiellement centrés sur le modèle de curriculum sur lequel toutes les enquêtes se fondent, tant en mathématiques qu'en sciences. Outre ce modèle de curriculum, les cadres de référence définissent les contenus sur lesquels portent les enquêtes et les capacités cognitives supposées pouvant être mises en œuvre par les élèves. Comme cela est le cas pour PISA, ils présentent aussi les questionnaires contextuels.

Le programme TIMSS est résolument orienté vers la recherche. Il s'est constitué, après FIMS et SIMS, comme un observatoire permanent des curriculums mathématiques et scientifiques d'une grande partie des pays du monde.

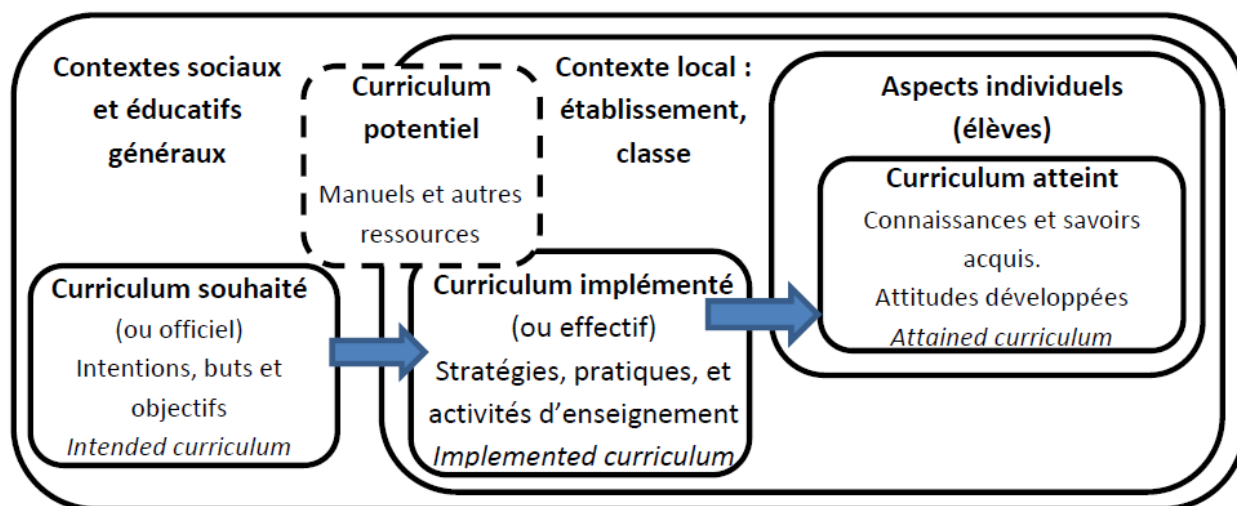
Le modèle développé par l'IEA pour les enquêtes internationales distingue trois aspects du curriculum :

- 1 le curriculum souhaité (« *intended curriculum* ») ;
- 2 le curriculum implémenté (« *implemented curriculum* ») ;
- 3 le curriculum atteint (« *attained curriculum* »).

Les enquêtes TIMSS cherchent à mieux connaître et faire connaître chacun de ces aspects des curriculums et à éclaircir les liens qu'ils entretiennent entre eux.

La figure ci-dessous illustre les relations entre ces trois aspects d'un curriculum.

Figure 8 : Différentes entrées pour l'étude d'un curriculum selon l'IEA



Source : adapté de Robitaille, D.F. ; Schmidt, W.H. ; Raizen, S. ; McKnight, C. ; Britton, E. & Nicol, C. (1993) et de Schmidt, W.H. (1996).

Le curriculum souhaité : pour chaque pays ou système éducatif, il s'agit du curriculum officiel ; celui que l'on peut déduire des programmes, des instructions, et des plans de formation des enseignants. À cela on peut ajouter le curriculum potentiel, constitué des manuels scolaires et des ressources de plus en plus

nombreuses mises à la disposition des enseignants (y compris, bien sûr, les ressources numériques et les équipements divers).

TIMSS attache une grande importance au curriculum souhaité, à la fois pour adapter ses enquêtes aux pratiques du moment et pour alimenter son encyclopédie. Pour l'étude de 1995, un groupe de travail a eu pour mission d'étudier cet aspect du curriculum à partir de documents de tous types envoyés au centre de recherche par les pays participants. La tâche était ambitieuse et n'a pu être que partiellement menée à bien (voir cependant Schmidt, 1996). Pour les enquêtes ultérieures, le curriculum souhaité est déduit de questionnaires renseignés par chacun des pays participants.

Le curriculum implémenté : évidemment, selon les pays, la distance est plus ou moins grande entre les intentions affichées et la réalité de l'enseignement donné. TIMSS cherche à s'approcher de cette réalité en s'appuyant sur des questionnaires relatifs à l'organisation des établissements scolaires et aux pratiques d'enseignement. Il s'agit de quatre questionnaires destinés respectivement :

- aux élèves ;
- aux élèves et à leurs parents (sauf pour TIMSSADV) ;
- aux professeurs ;
- aux chefs d'établissements et aux directeurs d'écoles.

En France, en 2015, au niveau CM1, les 4 questionnaires ont été administrés en format papier. Pour TIMSSADV, les questionnaires élèves (questionnaires distincts pour les volets mathématique et physique) ont été administrés sur papier ; les questionnaires enseignants (professeurs des écoles, professeurs de mathématiques et de physique) et le questionnaire établissement ont été administrés en ligne.

De plus, le chapitre de l'encyclopédie, qui résume les curriculums mathématique et scientifique ainsi que les politiques éducatives de près de 80 pays, (voir chapitre 1 D.1), est écrit en concertation étroite avec les représentants de chaque pays.

Le curriculum atteint : le curriculum souhaité et le curriculum implémenté précisent les objectifs que chaque pays cherche à obtenir et la façon dont il s'organise pour l'obtenir. L'objet des questionnaires cognitifs est de permettre la description du curriculum atteint et d'en permettre une analyse comparative avec les curriculums souhaité et implémenté.

On voit à quel point la philosophie de TIMSS diffère de celle de PISA. Le programme TIMSS part de l'enseignement tel qu'il se conçoit dans chacun des pays et tel qu'il s'y déroule. Il part des disciplines telles qu'elles se sont constituées et telles qu'elles sont enseignées. Cherchant à mettre à plat les éléments des curriculums puis à mettre ces éléments en relation, TIMSS dégage des tendances (le « T » de TIMSS signifie maintenant « tendances » (*trends*)) et met à la disposition des décideurs, des chercheurs et du public des éléments de comparaison de nature à les aider à orienter leur action et leurs réflexions. PISA, quant à lui, part du « monde réel » et même, du monde tel qu'il sera demain selon l'OCDE. Il cherche en quelque sorte à anticiper les besoins de la société et de l'humain par rapport à la société. Les disciplines, si elles interviennent, ne sont que des moyens qu'il s'agit de mettre au service de ces besoins. Une de ces deux approches est-elle préférable à l'autre ? Ce rapport n'a pas pour objectif de répondre à cette question, mais d'explicitier ces différences afin de permettre une lecture éclairée des rapports relatifs à ces enquêtes ou des interprétations qui en sont faites. Enfin, il ne faudrait pas déduire de ce qui précède que, contrairement à PISA, TIMSS ne se soucierait pas de l'amélioration des systèmes éducatifs ; simplement, les chemins empruntés ne sont pas du même ordre.

L'approche de TIMSS avec la description des trois curriculums rejoint celle adoptée par Chevallard (1991) dans le cadre de la transposition didactique. Ainsi les organisations mathématiques (OM) peuvent être définies dans ce cadre en fonction des étapes de la transposition : OM à enseigner (en lien avec le curriculum intentionnel), OM enseignée (curriculum mis en œuvre) et OM apprise (pour le curriculum atteint). Cette approche didactique permet, comme l'a opérationnalisé Grapin (2015), d'analyser le contenu d'évaluations externes et d'interpréter les résultats des élèves au regard des savoirs en jeu et de l'enseignement dispensé.

B. Le cadre de référence de TIMSS pour les mathématiques

1. Les domaines de contenus

En 1995, le syllabus était simplement une liste couvrant l'ensemble des mathématiques enseignées ou enseignables de la maternelle à l'université. Bien qu'accompagnée de catégories que l'on nommerait aujourd'hui compétences cognitives et perspectives (enjeux de la formation mathématique), cette présentation témoignait d'un ancrage très fort sur la discipline, ancrage qui subsiste aujourd'hui. Par la suite, et en particulier pour TIMSS 2015, le syllabus a laissé la place à des présentations de contenus spécifiés selon les niveaux (quatrième année scolaire, huitième année et année de terminale scientifique). Pour chacun des niveaux étudiés, la liste des contenus définit et précise plusieurs domaines⁴² dont l'organisation générale a peu varié au fil des enquêtes.

Voici la répartition de ces domaines pour TIMSS 2015 :

- **Pour la quatrième année scolaire :**
 - Nombres
 - Formes géométriques et mesures
 - Représentation de données
- **Pour la huitième année scolaire :**
 - Nombres
 - Algèbre
 - Géométrie
 - Données et probabilités
- **Pour l'année de fin d'études secondaires à orientation scientifique :**
 - Algèbre
 - Analyse
 - Géométrie

L'évolution depuis 1995 a essentiellement consisté, conformément à une tendance observée mondialement, à passer du simple syllabus (liste de chapitres et de thèmes de cours) à une présentation en termes de savoir-faire introduits par des verbes d'action (Faire, Calculer, Identifier, etc.). Ces contenus et ces savoir-faire étant le produit d'une étude comparée et détaillée des contenus des programmes des pays participant à l'étude, ils sont, de fait, en adéquation avec ces programmes dans leurs grandes lignes. Le

⁴² Le terme domaine n'a pas le même sens dans TIMSS que dans PISA.

cadre de référence de TIMSS 2015 justifie l'organisation des contenus par des considérations de nature pédagogique et remplace le syllabus par une liste de savoir-faire adaptés au niveau considéré. La France n'ayant pas participé à TIMSS 1995 pour le niveau CM1 et n'ayant pas participé à TIMSS 2015 pour le niveau quatrième, le présent rapport ne cherchera pas à comparer les contenus sollicités par ces deux enquêtes. En ce qui concerne TIMSSADV, en 1995, le domaine « Algèbre » était nommé « Nombres et équations », mais c'est surtout le détail des contenus des différents domaines qui a évolué depuis cette date.

Pour TIMSSADV 2015, chacun des domaines de contenu est constitué de thèmes (Tableau 7), et chaque thème à son tour comprend plusieurs sous-thèmes. Relativement au temps total alloué à l'évaluation, le poids en temps, relatif à chaque thème est identique. La comparaison des contenus pris en compte en 1995 et en 2015 paraît difficile à partir de celle des cadres de référence, en raison des différences de présentation. Une telle comparaison ne pourrait être faite qu'à partir de l'analyse comparée de l'ensemble des exercices utilisés dans ces deux enquêtes. On note cependant, conformément aux évolutions curriculaires dans le monde, un plus grand développement de tout ce qui touche aux fonctions, et en particulier à l'étude qualitative des fonctions en relation avec leurs représentations, et l'introduction de la fonction exponentielle. On note aussi la place plus importante donnée à l'analyse, avec toutefois la disparition des équations différentielles (notion qui n'avait d'ailleurs fait l'objet d'aucune question en 1995).

On pourra s'étonner du fait que les statistiques et les probabilités n'apparaissent pas dans les contenus de ce niveau. TIMSS l'explique par le fait que de grandes différences entre les pays existent dans ces domaines, ce qui rendait très problématique leur inclusion dans l'étude. Ce fait est certainement à noter en ce qui concerne la participation de la France à TIMSSADV 2015, compte tenu de la place maintenant prise par les probabilités et les statistiques dans les programmes scolaires français au collège et au lycée.

2. Les domaines cognitifs

Les domaines cognitifs définis par TIMSS et utilisés pour présenter les résultats sont :

- Connaître
- Appliquer
- Raisonner

Il s'agit ici d'aspects de l'activité mathématique plus ou moins entremêlés : il est en effet difficile d'appliquer sans connaître et sans raisonner.

Voici comment TIMSS 2015 décrit ces domaines dans ses cadres de référence :

Quatrième et huitième année scolaire (en France : CM1 et quatrième)

« Le premier domaine, « connaître », couvre les faits, les concepts et les procédures que les élèves ont besoin de connaître, tandis que le deuxième, « appliquer », met l'accent sur la capacité des élèves à appliquer les connaissances et la compréhension des concepts pour résoudre des problèmes ou pour répondre à des questions. Le troisième domaine, « raisonner », va au-delà de la solution des problèmes de routine pour englober des situations inhabituelles, des contextes complexes et des problèmes à plusieurs étapes. » (Traduit de Mullis et al., 2013)

Fin d'études secondaires à orientation scientifique (TIMSSADV 2015)

« La dimension cognitive des mathématiques se compose de trois domaines basés sur les processus de pensée que les élèves sont censés utiliser pour traiter les questions de mathématiques développées pour l'évaluation TIMSSADV 2015. Le premier domaine, « connaître », concerne la capacité des élèves de se rappeler et reconnaître les faits, les procédures et les concepts nécessaires à une solide base mathématique. Le deuxième domaine, « appliquer », met l'accent sur l'utilisation de ces connaissances pour modéliser et mettre en œuvre des stratégies pour résoudre des problèmes. Le troisième domaine, « raisonner », comprend les capacités à analyser, synthétiser, généraliser, et justifier par des arguments ou des preuves mathématiques. Les situations nécessitant un raisonnement sont souvent peu familières ou complexes.

Bien qu'il y ait une certaine hiérarchie entre les trois domaines cognitifs (de « connaître » à « appliquer » et « raisonner »), chaque domaine contient des éléments représentant une gamme complète de difficultés. » (Traduit de Mullis et al., 2014).

3. Tableaux synthétiques des plans d'évaluation en mathématiques de TIMSS 2015

Les tableaux suivants présentent une vue d'ensemble des plans d'évaluation de TIMSS 2015. Pour chaque niveau de l'enquête, chacun des domaines constitue une dimension de l'évaluation. Ces dimensions, conformément aux « méthodes de réponse à l'item » (voir Chapitre 1 D.1), produiront autant d'échelles qui seront présentées dans les rapports internationaux.

Tableau 7 : Plan d'évaluation de TIMSS 2015 - Mathématiques quatrième et huitième année scolaire

	DOMAINES	THÈMES
CONTENUS	NOMBRES	Nombres entiers
		Fractions et décimaux
		Expressions, équations simples et relations
	FORMES GÉOMÉTRIQUES ET MESURES	Points, lignes, et angles
		Formes en deux ou trois dimensions
	REPRÉSENTATIONS DE DONNÉES	Lecture, interprétation et représentation
COMPÉTENCES COGNITIVES	CONNAÎTRE	Rappeler
		Reconnaître
		Classifier, ranger
		Calculer
		Extraire
		Mesurer
	APPLIQUER	Déterminer
		Représenter/Modéliser
		Mettre en œuvre
	RAISONNER	Analyser
		Intégrer/Synthétiser
		Évaluer
		Tirer des conclusions
		Généraliser
		Justifier

Tableau 8 : Plan d'évaluation de TIMSSADV 2015 - Mathématiques avancées

	DOMAINES	THÈMES
CONTENUS	ALGÈBRE	Expressions et opérations
		Équations et inéquations
		Fonctions
	ANALYSE	Limites
		Dérivées
		Intégrales
	GÉOMÉTRIE	Géométrie classique et géométrie analytique
		Trigonométrie
COMPÉTENCES COGNITIVES	CONNAÎTRE	Rappeler
		Reconnaître
		Calculer
		Extraire
	APPLIQUER	Choisir les méthodes
		Représenter/Modéliser
		Mettre en œuvre
	RAISONNER	Analyser
		Intégrer/Synthétiser
		Évaluer
		Tirer des conclusions
		Généraliser
		Justifier

4. Distribution des questions de TIMSS 2015 selon les domaines et les formats

Comme cela a été expliqué pour PISA, les domaines de contenus et les domaines cognitifs de TIMSS (processus dans PISA) sont utilisés pour établir des tables de spécification *a priori* (répartition souhaitée des questions dans chacun des domaines et des formats de questions). Ces tables spécifient le poids relatif de chacun des domaines dans l'évaluation.

Dans la pratique, la répartition finale est toujours très proche de la répartition souhaitée.

Voici cette répartition finale pour TIMSS 2015 (mathématiques 4^e année et « *Advanced* »).

Tableau 9 : Contenus et formats des questions de TIMSS4 2015 (CM1) selon les domaines

	Questions à choix multiples	Questions ouvertes	Nombre total de questions	Score total ⁴³	Pourcentage du score total
Nombres	46	43	89	95	50 %
Formes géométriques et mesures	35	21	56	59	35 %
Représentation de données	8	16	24	28	15 %
Total	89	80	169	182	
Pourcentage du score total	50 %	50 %			

Source : Mullis & al. (2016a).

Tableau 10 : Contenus et formats des questions de TIMSSADV 2015 selon les domaines

	Questions à choix multiples	Questions ouvertes	Nombre total de questions	Score total ⁴⁴	Pourcentage du score total
Algèbre	19	18	37	43	35 %
Analyse	21	13	34	44	35 %
Géométrie	19	12	30	36	30 %
Total	59	43	102	123	
Pourcentage du score total	50 %	50 %			

Source : Mullis & al. (2016b).

⁴³ Certaines questions sont codées sur 1 ou 2 points.

⁴⁴ idem.

Tableau 11 : Domaines cognitifs des questions de TIMSS4 2015 (CM1) selon les processus

	Questions à choix multiples	Questions ouvertes	Nombre total de questions	Score total ⁴⁵	Pourcentage du score total
Connaître	37	27	64	65	40 %
Appliquer	36	36	72	80	40 %
Raisonner	16	17	33	37	20 %
Total	89	80	169	182	
Pourcentage du score total	50 %	50 %			

Source : Mullis & *al.* (2016a).

Tableau 12 : Domaines cognitifs des questions de TIMSSADV 2015 selon les processus

	Questions à choix multiples	Questions ouvertes	Nombre total de Questions	Score total ⁴⁶	Pourcentage du score total
Connaître	27	6	33	36	35 %
Appliquer	22	18	40	50	35 %
Raisonner	10	19	39	37	30 %
Total	59	43	102	123	
Pourcentage du score total	50 %	50 %			

Source : Mullis & *al.* (2016b).

⁴⁵ Certaines questions sont codées sur 1 ou 2 points.

⁴⁶ Certaines questions sont codées sur 1 ou 2 points.

5. Le cadre de référence pour la physique (TIMSS Advanced 2015)

Les grandes lignes retenues pour l'évaluation de la physique dans TIMSS Advanced 2015 sont rappelées ici, à partir du document plus complet⁴⁷ disponible sur le site internet consacré aux études TIMSS & PIRLS.

Dans la mesure où la population d'élèves concernée par l'enquête TIMSS Advanced 2015 constitue le futur vivier de scientifiques et d'ingénieurs, l'IEA considère qu'il est important de connaître le degré de maîtrise des concepts scientifiques de cette catégorie d'élèves (on ne discutera que du cas de la physique). Les enquêtes équivalentes précédentes (TIMSS Advanced) menés par l'IEA remontent à 2008 et 1995. En 2015 les études TIMSS et TIMSS Advanced ont été conduites pour la première fois simultanément, ce qui permet, si on y ajoute l'enquête PISA d'avoir une vue assez complète de l'enseignement des mathématiques et des sciences, de l'école primaire au lycée.

Contrairement au cadre d'évaluation de PISA 2015 dans lequel est défini pour chaque exercice un contexte (« santé et maladie », « ressources naturelles », « qualité de l'environnement », etc.) sans que la discipline scolaire ne soit nécessairement précisée, dans l'enquête TIMSS la discipline (« physique ») est explicite et le cadre d'évaluation indique également les différents domaines thématiques explorés : « mécanique et thermodynamique », « électricité et magnétisme », « phénomènes ondulatoires et physique nucléaire/atomique ». Par ailleurs, chaque question relève d'un des trois domaines cognitifs suivants : « savoir », « appliquer », « raisonner » domaines que l'on pourrait respectivement mettre en regard des trois degrés de charge cognitive de PISA : « peu élevé », « moyenne », « élevée ». Les pourcentages de questions relevant de ces domaines sont donnés dans le tableau ci-après.

Tableau 13 : Pourcentages des domaines thématiques et cognitifs de l'enquête TIMSS Advanced 2015 en physique

Domaines thématiques	Pourcentage
Mécanique et thermodynamique	40 %
Electricité et magnétisme	25 %
Phénomènes ondulatoires et physique atomique/nucléaire	35 %

Domaines cognitifs	Pourcentage
Connaître	30 %
Appliquer	40 %
Raisonner	30 %

Source : repris de TIMSS Advanced 2015 Framework, p. 18

⁴⁷ TIMSS Advanced 2015 Assessment Frameworks. Mullis, I.V.S. & Martin, M.O. (Eds.). (2014). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Document disponible à l'adresse suivante: <http://timssandpirls.bc.edu/timss2015-advanced/frameworks.html>

Chapitre 4 : Structure et format des exercices d'évaluation de PISA et de TIMSS en mathématiques.

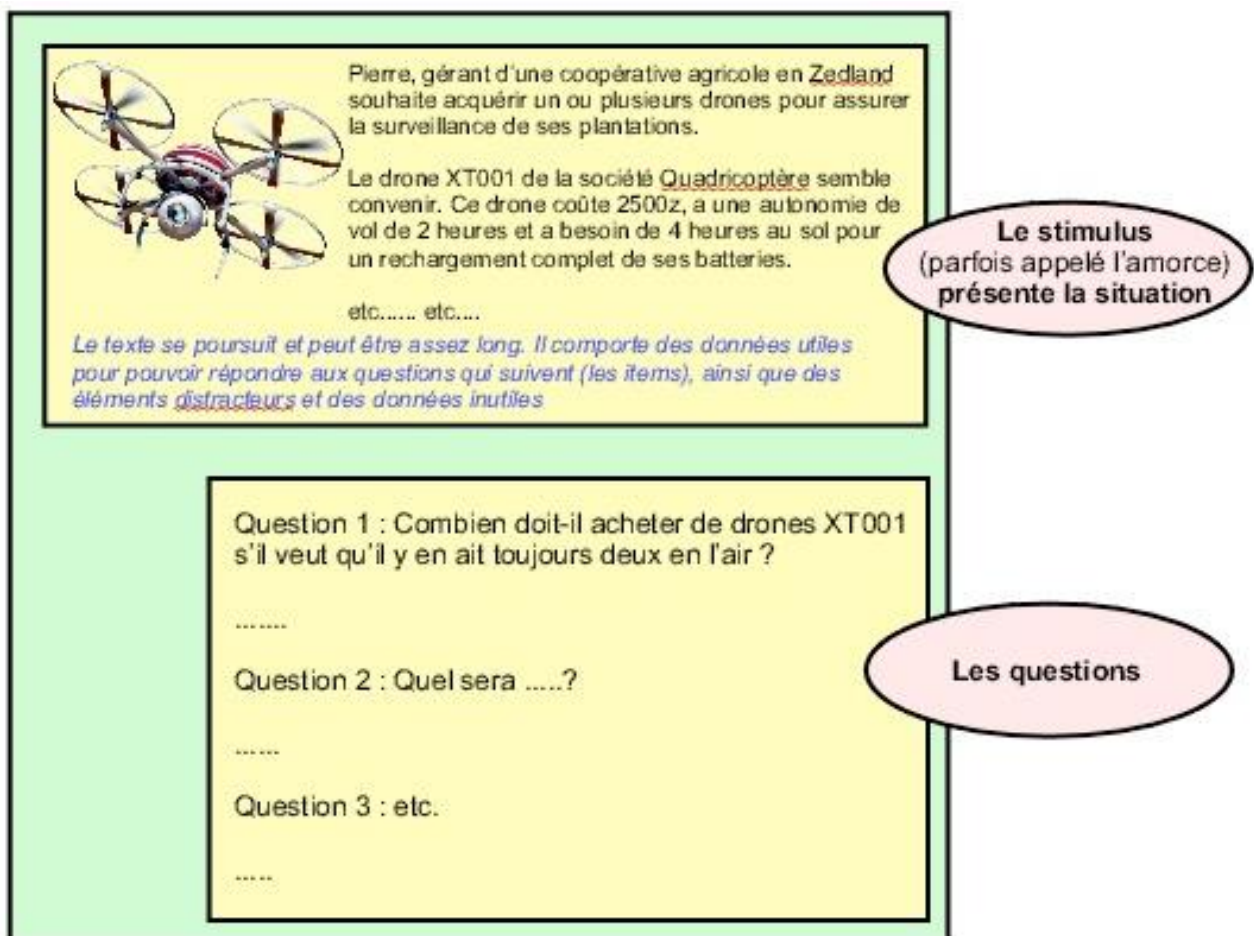
Pour comprendre les différences qui existent entre les enquêtes PISA et TIMSS, et ce qu'elles évaluent, il est également nécessaire de se pencher sur la manière dont sont présentés les énoncés des exercices : font-ils référence à des situations concrètes ? Ou sont-ils purement mathématiques ? Il faut aussi s'intéresser aux types de questions posées : demande-t-on aux élèves d'argumenter leur réponse ? Ou simplement d'en choisir une parmi d'autres dans le cadre d'un QCM et sans exiger d'explication ? Ce chapitre se propose de répondre à ces questions.

A. Aspects communs à PISA et à TIMSS

Les exercices de PISA comportent systématiquement une première partie destinée à présenter la situation sur laquelle porteront les questions auxquelles l'élève aura à répondre. Il s'agit de mettre les questions en contexte et de les ancrer sur le « monde réel ». Cette partie constitue le stimulus de l'exercice et peut comporter un texte de présentation qui peut être assez long.

Comme cela est en général le cas dans les situations rencontrées hors de la classe, le stimulus ne cherche pas à être pur, c'est-à-dire qu'il ne se limite pas à présenter les informations dont la prise en compte sera nécessaire pour répondre aux questions qui lui sont associées. Conçu comme la mise en scène d'une situation, le stimulus comprend des informations superflues (textes, images, graphiques, etc.). Pour répondre aux questions qui sont associées au stimulus, l'élève devra d'abord comprendre la situation, puis la mathématiser, ce qui suppose qu'il fasse des choix pertinents et suffisants parmi les informations apportées par le stimulus ; c'est le processus « Formuler » de PISA.

Figure 9 : Illustration du format habituel des exercices de PISA (contenu fictif)



La figure 9 illustre le format habituel des exercices mathématiques de PISA⁴⁸, mais signalons que le même format est utilisé pour les autres domaines des enquêtes.

Les questions associées à un exercice donné sont supposées indépendantes. Elles le sont en effet dans la mesure où l'élève peut répondre à l'une quelconque d'entre elles sans même avoir pris connaissance des autres. Toutefois, l'élève est amené à faire un aller et retour entre la question à laquelle il doit répondre et le stimulus, et il n'est pas exclu qu'il puisse être influencé par le contexte et par les difficultés qu'il a pu avoir à répondre aux questions précédentes.

Une partie des exercices de TIMSS sont construits sur le même modèle que ceux de PISA⁴⁹. Toutefois le contenu du stimulus est réduit au minimum et même disparaît assez souvent dans le cas de questions liées à un même stimulus, ces questions sont toutefois considérées comme indépendantes. Les exercices de TIMSS sont le plus souvent placés d'emblée dans le monde mathématique et, lorsqu'ils ne le sont pas, leur stimulus est beaucoup plus réduit que dans le cas des exercices de PISA. De ce fait, ils sont beaucoup plus en accord avec les habitudes scolaires de nombreux pays. Cette remarque permet d'ailleurs de distinguer certains pays qui, tels la Russie, réussissent mieux à TIMSS qu'à PISA, de ceux qui, tels la Norvège ou la Nouvelle-Zélande, réussissent mieux à PISA qu'à TIMSS.

Plus généralement, une étude comparant les résultats de TIMSS grade 8 et de PISA (Wu, 2009) a amené son auteure à sélectionner parmi les questions de TIMSS 2003 grade 8 ceux qui auraient pu être utilisés par PISA, non en ce qui concerne le niveau, mais seulement en ce qui concerne la forme. Elle avait conclu que plus de 50 % des questions de TIMSS n'auraient pas pu être utilisées par PISA car trop formelles et trop intra-mathématiques. Le même travail mené pour ce rapport en étudiant les questions de TIMSS 2015 pour le grade 8 a conduit à peu près au même pourcentage. Pour TIMSSADV 2015, environ 15 % des questions conviendraient à l'esprit de PISA⁵⁰.

En réponse aux critiques faites par les représentants des pays et par de nombreux observateurs, PISA tend à mettre davantage l'accent sur les contenus mathématiques, ce qui se traduit par la proportion importante de questions classées dans la catégorie « Employer » (près de 50 %). Dans le même temps et pour répondre à des critiques opposées, TIMSS tend à augmenter légèrement la proportion de questions ancrées sur le « monde réel ».

PISA et TIMSS utilisent la même méthode pour organiser la passation de leurs questions d'évaluation : la méthode dite des « cahiers tournants » (chapitre 1. D.1). Cette méthode, qui permet de baser l'enquête sur un nombre de questions trois fois plus important que le nombre de questions que passera chaque élève, a aussi le grand avantage de limiter considérablement les problèmes de copiage entre élèves. Pour PISA, le temps moyen laissé pour répondre à une question est d'environ 2 minutes. Les élèves disposent d'un temps limité pour effectuer l'ensemble de leur cahier de test, mais le temps passé sur chacun des exercices n'est pas minuté. Pour TIMSS, la situation est à peu près la même : la vitesse de traitement y est aussi favorisée. Le temps moyen alloué à chaque question est de 2 minutes pour le niveau CM1 et un peu plus long (2 min 40 s) pour TIMSSADV. De ce fait, dans un cas comme dans l'autre, tous les élèves ne vont pas jusqu'au bout des cahiers de tests. Comme les unités d'évaluation sont réparties différemment dans les

⁴⁸ Rappelons que les définitions que nous utilisons pour les termes Exercice, Question, Item ont été précisées au chapitre 1 D.1.

⁴⁹ Des exercices issus de TIMSS sont présentés et analysés dans le volume 2 de ce rapport.

⁵⁰ Il est également intéressant de constater que le format de la plupart des questions de TIMSSADV 2015 est également assez éloigné de celui des épreuves du baccalauréat S en France, qui sont désormais nettement plus en relation avec le « monde réel » que celles de TIMSS.

cahiers, le même exercice peut se retrouver en début de cahier, au milieu, ou à la fin, ce qui a pour effet de réduire considérablement l'effet d'ordre observé dans des évaluations plus classiques où ce sont les mêmes derniers exercices qui ne sont pas atteints par une partie plus ou moins importante des élèves.

PISA libère environ 25 % de ses questions d'évaluation de la littératie mathématique et encore moins, à ce jour, de littératie scientifique. TIMSS, de son côté, s'appuyant sur une banque d'exercices très importante, en libère un peu plus de 40 %, aussi bien en mathématiques qu'en sciences⁵¹.

B. Les types de questions de PISA et de TIMSS

La différence principale entre les questions de PISA et celles de TIMSS réside dans leur ancrage : ancrage presque toujours dans le « monde réel » pour PISA ; ancrage le plus souvent dans le monde mathématique pour TIMSS. Pour PISA, cela résulte en une charge de lecture de textes nettement plus importante pour PISA que pour TIMSS. Dans les deux enquêtes, les questions sont soit des questions à choix multiples (QCM) soit des questions à réponses ouvertes. TIMSS et TIMSS *Advanced* répartissent de façon égale les points attribués aux QCM et aux questions à réponses ouvertes, alors que PISA en attribue un peu moins aux QCM (environ un tiers-deux tiers pour les mathématiques).

On distingue les QCM à choix simple, c'est-à-dire les questions pour lesquelles une seule des réponses proposées est exacte, les autres étant les distracteurs, et les QCM à choix complexe, c'est-à-dire les questions où plusieurs des réponses proposées peuvent être exactes. Pour TIMSS, tous les QCM sont à choix simple parmi quatre réponses proposées. Pour PISA, environ les deux tiers des QCM sont à choix simple parmi un nombre de réponses proposées allant de 4 à 6. Les autres QCM sont à choix complexe.

Dans les deux enquêtes, dans le domaine mathématique, les questions ouvertes sont semi-ouvertes à réponses courtes, voire très courtes. Dans les deux tiers des cas pour PISA, la réponse attendue se limite en effet à un mot ou à un nombre et peut être codée par une personne sans compétence particulière ou automatiquement dans le cas d'une passation sur ordinateur. Dans les autres cas, la réponse attendue peut comporter une ou plusieurs phrases, mais est toujours limitée à 3 ou 4 lignes en réponse à la consigne « Montrez votre travail », qui ne demande pas une réponse structurée, encore moins une démonstration construite.

En France, certains commentaires des résultats de PISA ont pu laisser croire que la présence de QCM désavantagerait les élèves français. En fait, ce sont plutôt les questions à réponses ouvertes et en particulier celles qui demandent une justification, en général sous la forme « Montrez votre travail », qui les défavorisent. Dans ce cas, les élèves français se croient en effet obligés de faire des phrases et de produire une rédaction correcte de leur réponse. Or les codages se satisfont de quelques mots plus ou moins déconnectés et d'opérations non argumentées. Cela fait souvent dire que les élèves français ne savent pas justifier leur démarche, alors que ce pourrait être au contraire leur souci d'une explication conforme aux habitudes scolaires qui les désavantage.

Ainsi, lorsque la réponse attendue est dite simplement ouverte, un seul mot ou un seul nombre peut suffire ; lorsque cette réponse est dite complexe, une ligne ou deux, non rédigées, suffisent en général. Il y a donc là une différence importante avec les standards habituels de l'enseignement des mathématiques en

⁵¹ Pour TIMSS 2015, l'IEA donne accès à environ 40 % des questions, mais leur utilisation est soumise à autorisation.

France. Certains pays avaient un enseignement plus adapté aux évaluations internationales ; d'autres ont fait le choix de s'y adapter, aussi bien d'une façon générale qu'en ce qui concerne l'entraînement aux types de questions de ces enquêtes. Mais une étude récente de l'OCDE, s'appuyant justement sur les résultats de PISA 2012, met en évidence le fait que les élèves qui réussissent le mieux au volet mathématique de PISA sont ceux qui ont reçu un enseignement de type formel – donc ceux qui y sont *a priori* le moins préparés - et non ceux qui ont reçu un enseignement essentiellement centré sur les situations de la vie réelle (OECD, 2016b)⁵².

Cela semble peu connu en France, et encore moins utilisé, mais il faut noter que l'OCDE a publié en 2009 un ouvrage titré « *Take the test* » et dont la traduction du sous-titre est : « Échantillon de questions des évaluations PISA de l'OCDE ». L'ouvrage (en anglais) s'adresse directement aux élèves, aux parents et aux enseignants et contient 300 pages de questions des différents domaines de PISA ; ce sont des questions qui ont été utilisées dans les enquêtes, simplement expérimentées lors de pré-enquêtes ou présentées comme exemples (OECD, 2009). De même, on trouve sur le Web des banques d'exercices pour s'entraîner à TIMSS⁵³.

Un point commun aux questions des deux enquêtes est que les descripteurs utilisés sont limités à quelques mots clés, à savoir ceux désignant les contenus, les processus et les formats. Avant l'enquête, les questions sont aussi accompagnées des paramètres statistiques déduits de leur expérimentation (indices de difficulté, de discrimination, etc.) ; après l'enquête, on a en plus accès aux taux de réussite observés dans les différents pays participants.

Les consignes de codage des réponses se limitent au contrôle de l'exactitude de la réponse et donc ne fournissent aucune indication sur les démarches suivies par les élèves, ou sur les difficultés qu'ils ont pu rencontrer et les types d'erreurs qu'ils ont pu faire. De plus, ainsi que le remarquent Roditi et Salles (2015) à propos de PISA :

« Les classifications utilisées par l'OCDE ne permettent ni de recenser précisément les connaissances acquises des élèves ni d'estimer le niveau d'acquisition de ces connaissances ».

⁵² En France, au moment de la passation des épreuves PISA 2012, 66 % des élèves étaient au lycée, où la plupart d'entre eux recevaient un enseignement plutôt « formel ». On ne peut cependant pas dire quel type d'enseignement ces élèves avaient reçu au collège.

⁵³ Par exemple : http://www.edinformatics.com/timss/timss_intro.htm

On pourrait faire un constat analogue pour TIMSS. Cela est compréhensible pour des enquêtes de cette ampleur, destinées au moins pour PISA à fournir des indicateurs généraux pour le pilotage des systèmes éducatifs. Des études complémentaires peuvent être cependant menées dans chacun des pays pour compléter l'information. Ces études ont en effet lieu, comme celle menée en France à la Depp par Roditi et Salles (2015), mais il manque la dimension internationale qui permettrait de pointer des types d'erreurs différentes selon les pays.

Mener une enquête du type PISA ou TIMSS avec des ambitions d'usage didactique à la hauteur des ambitions psychométriques serait certainement trop lourd et potentiellement source de nombreuses confusions, mais la focalisation sur les qualités psychométriques des questions et des échelles entraîne à choisir un certain type de questions, à ne pas en choisir d'autres, et à adopter un mode de codage limité.

Chapitre 5 : La présentation des résultats des enquêtes PISA et TIMSS

Les résultats aux enquêtes internationales sont souvent présentés sous forme de scores à partir desquels se fait un classement des différents pays. Mais de quelle manière sont construits ces scores ? Comment les interpréter ? Le taux de réussite aux différentes questions n'est-il pas aussi instructif ? Pourquoi la position précise d'un pays dans le classement est-elle à prendre avec précaution ? Enfin, doit-on s'attendre à des résultats sensiblement différents entre les trois domaines principaux évalués par PISA pour un même pays ? Ce chapitre se donne pour objectif à la fois de comprendre comment sont construits ces scores mais aussi de les interpréter correctement.

A. Lecture des résultats et des échelles utilisées⁵⁴

Il est difficile de comprendre les résultats des enquêtes PISA si l'on a pas quelques notions sur la façon dont les données sont traitées et dont les scores sont calculés. En effet, que signifie par exemple la phrase suivant ? « En mathématiques, en 2003, le score de la France était égal à 511, tandis que celui de la Finlande était de 544, soit un écart de 33 points ». Ce type de formulation apparaît souvent pourtant, aussi bien dans les rapports et les commentaires officiels que dans la presse. Le tableau suivant rassemble ces « scores » pour les enquêtes de PISA 2000, 2003, 2006 et 2012.

Tableau 14 : « Scores » en mathématiques aux enquêtes PISA en France, en Finlande et en Allemagne⁵⁵

	Maths 2000	Maths 2003	Maths 2006	Maths 2012
FRANCE	517	511	496	495
FINLANDE	536	544	548	519
ALLEMAGNE	490	503	504	514

Sources : Rapports PISA de 2000 à 2012 (OCDE)

Les résultats centraux et les plus diffusés, en particulier dans les médias, des enquêtes internationales sont les positions moyennes des pays et des populations participantes sur une échelle (dont la moyenne est fixée à 500 et l'écart-type à 100). Outre ce résultat dit de tendance centrale, sont fournies des informations sur la dispersion des résultats. Les présentations de ces résultats globaux diffèrent pour les deux enquêtes : elles sont essentiellement fournies sous forme de tableaux pour PISA (Tableau 15).

⁵⁴ Cette partie est reprise de Bodin (2008b) et légèrement adaptée.

⁵⁵ La comparaison avec la Finlande s'imposait dans la mesure où ce pays est souvent cité comme exemple à suivre. L'Allemagne est choisie également comme pays ayant réagi dès 2003 à ses « mauvais » résultats à PISA.

Tableau 15 : Exemple de présentation des résultats de PISA 2012

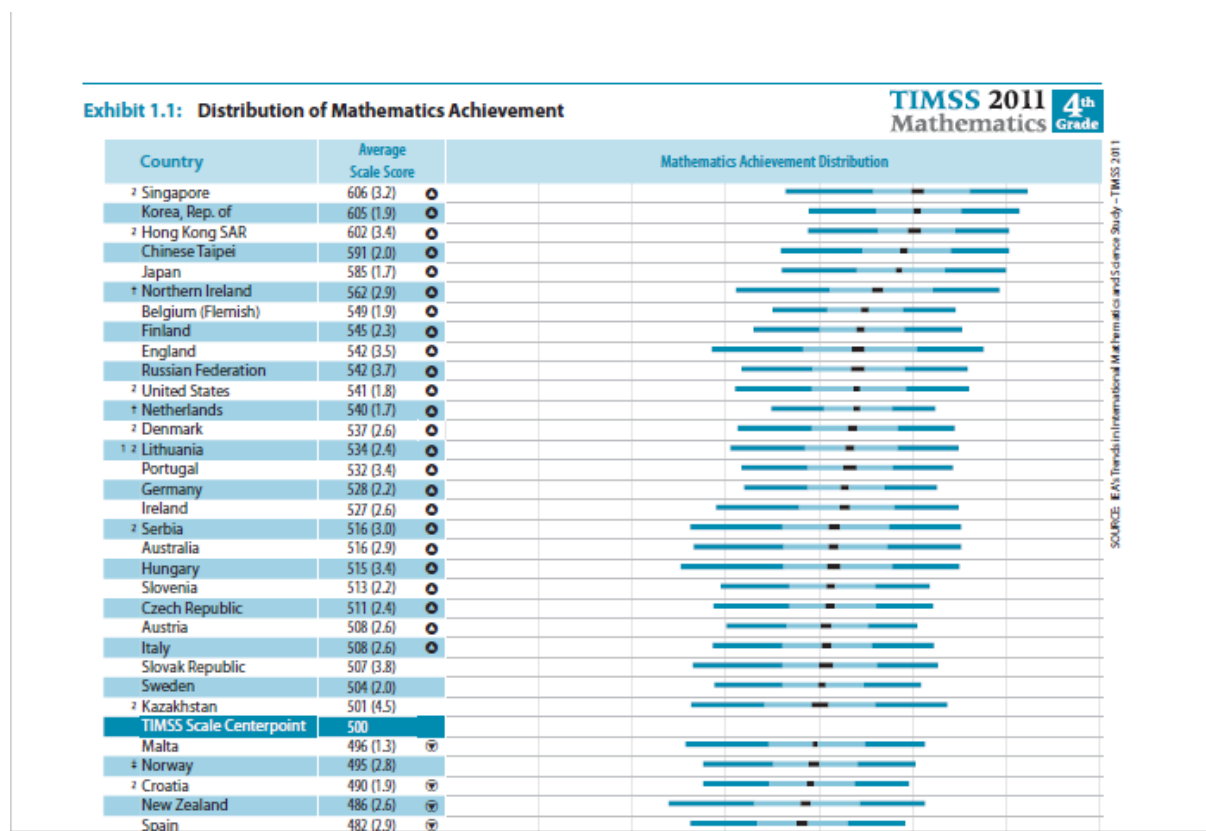
**Synthèse de la performance des pays et économies en mathématiques,
en compréhension de l'écrit et en sciences**

	Pays/économies dont le score moyen/le pourcentage d'élèves très performants sont supérieurs à la moyenne de l'OCDE				Pays/économies dont le pourcentage d'élèves peu performants est inférieur à la moyenne de l'OCDE			
	Pays/économies dont le score moyen et les pourcentages d'élèves très performants/peu performants ne s'écartent pas de la moyenne de l'OCDE dans une mesure statistiquement significative				Pays/économies dont le score moyen/le pourcentage d'élèves très performants sont inférieurs à la moyenne de l'OCDE			
	Pays/économies dont le pourcentage d'élèves peu performants est supérieur à la moyenne de l'OCDE							
	Mathématiques				Compréhension de l'écrit		Sciences	
	Score moyen lors de l'évaluation PISA 2012	Pourcentage d'élèves peu performants en mathématiques (sous le niveau 2)	Pourcentage d'élèves très performants en mathématiques (niveau 5 ou 6)	Variation annualisée (en points de score)	Score moyen lors de l'évaluation PISA 2012	Variation annualisée (en points de score)	Score moyen lors de l'évaluation PISA 2012	Variation annualisée (en points de score)
Moyenne OCDE	494	23.0	12.6	-0.3	496	0.3	501	0.5
Shanghai (Chine)	613	3.8	55.4	4.2	570	4.6	580	1.8
Singapour	573	8.3	40.0	3.8	542	5.4	551	3.3
Hong-Kong (Chine)	561	8.5	33.7	1.3	545	2.3	555	2.1
Taipei chinois	560	12.8	37.2	1.7	523	4.5	523	-1.5
Corée	554	9.1	30.9	1.1	536	0.9	538	2.6
Macao (Chine)	538	10.8	24.3	1.0	509	0.8	521	1.6
Japon	536	11.1	23.7	0.4	538	1.5	547	2.6
Liechtenstein	535	14.1	24.8	0.3	516	1.3	525	0.4
Suisse	531	12.4	21.4	0.6	509	1.0	515	0.6
Pays-Bas	523	14.8	19.3	-1.6	511	-0.1	522	-0.5
Estonie	521	10.5	14.6	0.9	516	2.4	541	1.5
Finlande	519	12.3	15.3	-2.8	524	-1.7	545	-3.0
Canada	518	13.8	16.4	-1.4	523	-0.9	525	-1.5
Pologne	518	14.4	16.7	2.6	518	2.8	526	4.6
Belgique	515	19.0	19.5	-1.6	509	0.1	505	-0.9
Allemagne	514	17.7	17.5	1.4	508	1.8	524	1.4
Vietnam	511	14.2	13.3	m	508	m	528	m
Autriche	506	18.7	14.3	0.0	490	-0.2	506	-0.8
Australie	504	19.7	14.8	-2.2	512	-1.4	521	-0.9
Irlande	501	16.9	10.7	-0.6	523	-0.9	522	2.3
Slovénie	501	20.1	13.7	-0.6	481	-2.2	514	-0.8
Danemark	500	16.8	10.0	-1.8	496	0.1	498	0.4
Nouvelle-Zélande	500	22.6	15.0	-2.5	512	-1.1	516	-2.5
République tchèque	499	21.0	12.9	-2.5	493	-0.5	508	-1.0
France	495	22.4	12.9	-1.5	505	0.0	499	0.6
Royaume-Uni	494	21.8	11.8	-0.3	499	0.7	514	-0.1

Source : Principaux résultats de l'enquête PISA 2012 : ce que les élèves de 15 ans savent et ce qu'ils peuvent faire avec ce qu'ils savent (OCDE, 2014).

Les informations sont plus graphiques pour TIMSS (Figure 10), mais le lecteur trouvera dans les différents documents publiés par l'IEA les informations attendues sur les tendances centrales et les distributions.

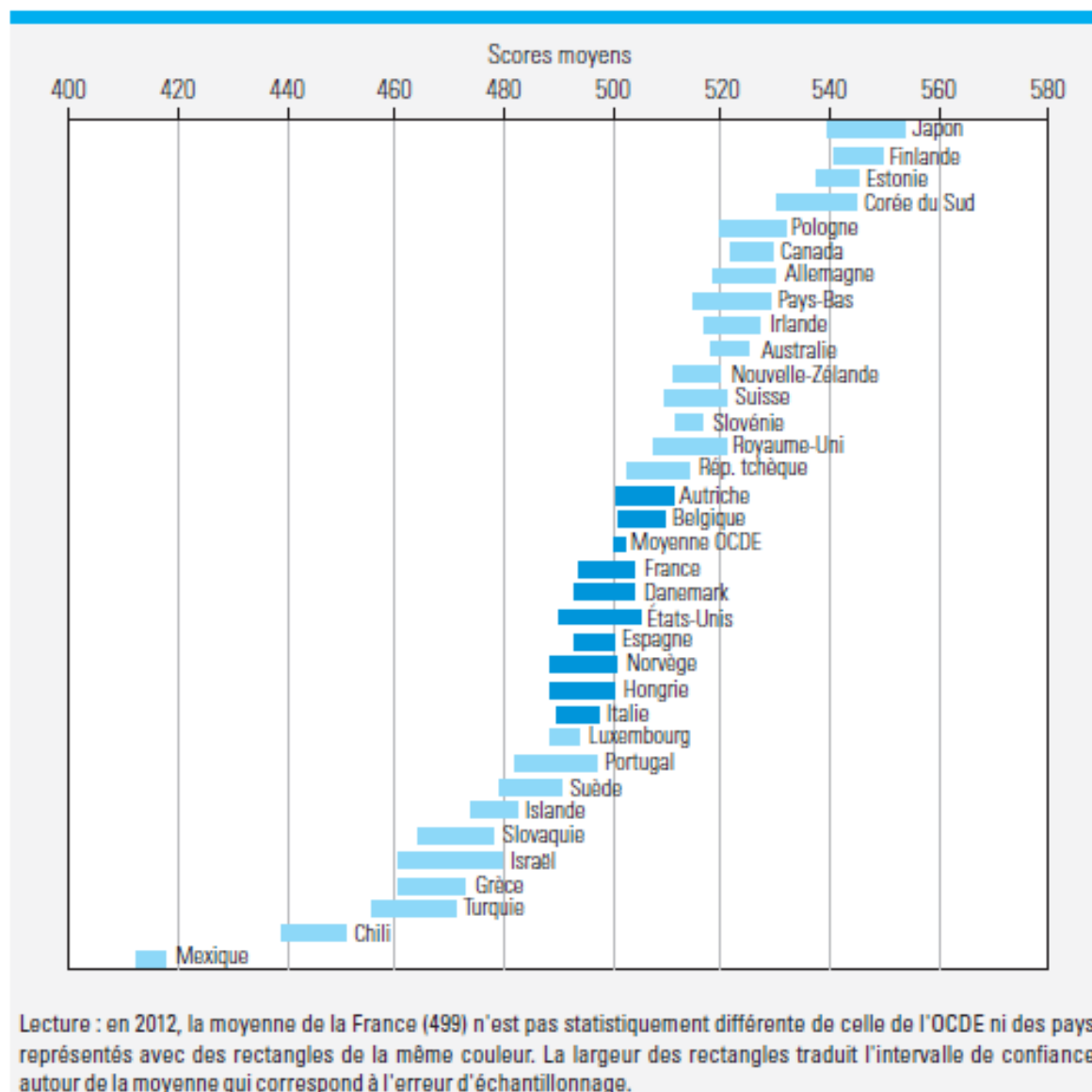
Figure 10 : Résultats en mathématiques en 4^e année de TIMSS 2011. Moyenne et distribution pour chaque pays



Source : TIMSS 2011 International Results in Mathematics », In V.S. Mullis, Michael O. Martin, Pierre Foy, & Alka Arora (2012)

Sont également données les erreurs-types de mesure qui permettent d'identifier dans les comparaisons l'existence de différences statistiquement significatives entre les pays. En effet, le score étant une estimation statistique, il est important de tenir compte de ces erreurs-types pour déterminer si un pays a un score significativement différent d'un autre pays (Figure 10). À titre d'exemple, en 2012, pour PISA, la France était classée 18^e parmi les pays de l'OCDE en littératie mathématique, mais son score ne différait pas significativement des pays classés entre la 14^e et la 23^e place (OCDE, 2014).

Figure 11 : Résultats de PISA 2012 en culture scientifique avec la représentation de l'erreur de mesure



Source : MENESR-Depp. L'évolution des acquis des élèves de 15 ans en compréhension de l'écrit et en culture scientifique - Premiers résultats de l'évaluation internationale PISA 2012. *Note d'Information n°13.30*, décembre 2013.

B. Comparaisons temporelles

L'utilisation d'items d'ancrage

Le fait que les enquêtes soient reproduites avec une certaine périodicité permet d'étudier l'évolution dans le temps des compétences des élèves d'un âge donné (PISA) ou étant scolarisés à un niveau scolaire donné (TIMSS). On peut alors comparer les performances de générations différentes dans un plan d'études dit transversal (*cross-section*). En effet, il ne s'agit pas d'une étude longitudinale où une même population d'élèves est suivie au fil du temps, comme par exemple dans les panels d'élèves mis en place par la Depp.

La comparaison de résultats d'élèves dans un plan transversal nécessite de pouvoir placer sur une même échelle les compétences des échantillons différents qui n'ont pas passé en général un test identique. Pour

placer les résultats de ces différents échantillons sur une même échelle, on utilise une procédure dite de chaînage (*linking*). Pour cela, on utilise des questions dites d’ancrage qui sont les mêmes d’une passation à l’autre, et dont on a vérifié qu’elles n’étaient pas sensibles aux éventuels changements de contextes. Les estimations des paramètres de ces questions d’ancrage ou de tendance (*trend items*) permettent de placer les résultats des cohortes sur une même échelle, ce qui permet d’apprécier l’existence et l’ampleur des évolutions d’une enquête à l’autre.

L’étude locale du taux de réponses correctes aux questions d’ancrage

Outre la façon dont les programmes PISA et TIMSS utilisent des questions d’ancrage pour relier de façon fiable leurs enquêtes successives, il est possible également de s’intéresser aux pourcentages de réponses correctes des élèves. Ces taux de réussite sont des scores « vrais », c’est-à-dire qu’ils correspondent aux réponses des élèves ayant effectivement répondu aux questions, alors que le niveau de difficulté d’une question, utilisé pour déterminer le niveau de compétence des élèves et validé d’un point de vue psychométrique, est lui, plus complexe.

Par exemple, la question M273Q01T dénommée « Conduites » est utilisée depuis la première enquête de PISA. Cette question du domaine géométrie n’étant pas libérée, il n’est pas possible de la présenter ici. Les résultats sont cependant intéressants à observer.

Tableau 16 : Évolution du taux de réponses correctes à la question « Conduites » de l’enquête PISA au niveau de la France et de l’OCDE

	2000	2003	2006	2009	2012	2015
OCDE	54 %	52 %	54 %	53 %	51 %	?
FRANCE	52 %	57 %	51 %	51 %	49 %	?
France GARÇONS	59%	63%	56%	54%	55%	?
France FILLES	45 %	51 %	47 %	47 %	44 %	?

Sources : Bases de données des enquêtes PISA de 2000 à 2012.

On remarque immédiatement que, pour cette question, la moyenne de réussites dans l’OCDE est restée stable au fil des enquêtes et que, mis à part le résultat de 2003, pour lequel on peut soupçonner un effet d’échantillonnage,⁵⁶ il en a été de même dans le cas français. Compte tenu des marges d’erreurs de l’enquête⁵⁷, les différences observées entre 2000 et 2012 ne sont pas significatives. D’une façon générale, on sait que pour la France, dans le domaine de la littératie mathématique, PISA a enregistré une baisse de

⁵⁶ Tous les résultats sont accompagnés d’une marge d’erreur reconnue et assumée (due aux fluctuations d’échantillonnage).

⁵⁷ PISA publie tous ses résultats avec des intervalles de confiance au seuil de 95 %. Malheureusement, ces erreurs de mesure ne sont pas toujours prises en compte par les commentateurs. 2003 et 2012 sont ici considérées comme références principales pour la littératie mathématique (domaine majeur).

niveau entre 2003 et 2012 ; toutefois, pour les questions posées, il n'y a pas de baisse systématique. Le taux de réussite à certaines questions a baissé de façon importante tandis que pour d'autres il est resté stable, et que pour d'autres enfin il s'est accru.

Le tableau 17 présente les résultats d'une question du domaine quantité : question PM192Q01T dénommée « Récipients » (non libérée).

Tableau 17 : Évolution du taux de réponses correctes à la question « Récipients » de l'enquête PISA au niveau de la France et de l'OCDE

	2000	2003	2006	2009	2012	2015
OCDE	37,8 %	40,4 %	40,3 %	41,1 %	40,4 %	?
FRANCE	36,5 %	36 %	34,4 %	40,3 %	39,1 %	?
France GARÇONS	43,4 %	42,8 %	37,2 %	48,3 %	45,2 %	?
France FILLES	29,3 %	23,7 %	31,8 %	33,1 %	33,0 %	?

Sources : Bases de données des enquêtes PISA de 2000 à 2012.

Comme pour la question précédente, on observe une certaine stabilité dans les résultats, avec, même, une tendance à la hausse.

Au-delà des résultats globaux, on observe que ces deux questions discriminent de façon importante les filles par rapport aux garçons ; cela de façon stable dans le temps (différence de réussite : entre 10 et 14 points). Pour ces deux questions, il est vrai assez souvent discriminantes par rapport au genre, certains pays ont sur ce point une différence beaucoup plus faible, voire nulle ou négative.

En ce qui concerne le cas français, on trouve aussi des questions auxquelles les filles réussissent mieux que les garçons. Une étude approfondie serait nécessaire, mais il est déjà établi que le contexte de la question influence les résultats des élèves selon les groupes auxquels ils appartiennent (garçons, filles, milieux défavorisés, favorisés, âge, ...). On pourra se reporter sur ce point aux études EVAPM⁵⁸.

Ces remarques faites à partir de deux exercices de PISA pourraient être faites de la même façon pour des questions de TIMSS, du moins dans le cas où les enquêtes et les questions auront été suivies dans le temps. En France, ce ne sera possible, pour l'instant, que pour les questions de TIMSSADV reprises de l'enquête de 1995.

⁵⁸ Évaluation des Apprentissages Mathématiques. EVAPM est un dispositif continu d'observation et de recherche, développé depuis 1986 dans le cadre de l'APMEP (Association des professeurs de mathématiques de l'enseignement public).

C. Niveaux des élèves et valeurs plausibles

Le résultat à un test n'est qu'une mesure imparfaite d'une compétence (qui est quant à elle inobservable). De fait, la réponse à une question peut être influencée par des facteurs tels que la condition physique et mental le jour J, ou des conditions de passation du questionnaire peu propices à la concentration (salle particulièrement bruyante, surchauffée, ...). Les algorithmes développés au cours des vingt dernières années ont précisément cherché à rendre aussi fiable que possible l'estimation du niveau de compétence de chacun des pays. Le développement de ces algorithmes aboutit à l'estimation de valeurs dites plausibles qui seront estimées conditionnellement aux réponses des élèves et à leurs caractéristiques (recueillies à l'aide du questionnaire de contexte).

Bien que les résultats les plus diffusés montrent une seule valeur moyenne pour chaque échantillon, au niveau des élèves, on n'obtient pas en réalité un score unique, mais plusieurs, (actuellement cinq), qu'on appelle des valeurs plausibles. Il s'agit d'imputer à chaque élève sa compétence à partir d'un ensemble de valeurs incomplètes. L'idée centrale est que l'on ne va pas, comme il est usuel, calculer un score en faisant le total des points obtenus à chacune des questions, mais obtenir une estimation probabiliste de la compétence d'un élève ayant telles caractéristiques socio-démographiques et ayant donné telles réponses aux questions auxquelles il a répondu. Pour chaque élève, cinq valeurs plausibles sont imputées à partir de la distribution construite pour cet élève. Cette méthode permet une estimation plus robuste, en particulier des valeurs des paramètres au niveau des populations. Elle nécessite toutefois des procédures particulières lorsqu'on réalise des analyses sur ces données dans la mesure où il faudra gérer cinq valeurs de compétences par élève, ce qui nécessite d'employer des méthodes prenant en compte cette variabilité.

D. Comparaison des résultats aux enquêtes

Comparaison des résultats des trois domaines des enquêtes PISA

On observe des niveaux de corrélation très élevés entre les résultats des différents domaines (mathématiques, résolution de problèmes, compréhension de l'écrit, ...) aux enquêtes PISA, comme le montre le tableau 18 pour 2012.

Tableau 18 : Corrélations entre les résultats des élèves dans les différents domaines de PISA 2012
(coefficient de corrélation linéaire de Bravais-Pearson)

	Littératie scientifique	Compréhension de l'écrit	Résolution de problèmes
Littératie mathématique	0,89	0,85	0,85
Littératie scientifique		0,85	0,80
Compréhension de l'écrit			0,82

Source : PISA 2012 (indices extraits du rapport technique)

Ces niveaux de corrélation sont beaucoup plus élevés que ceux qui ont été observés pendant plus de 20 ans dans les enquêtes EVAPM entre les différents domaines mathématiques (nombres, algèbre, analyse, géométrie, données) à tous les niveaux de la sixième à la terminale. Peu d'études s'intéressent aux corrélations des résultats individuels des élèves. La plupart comparent les résultats globaux des pays : dans ce cas, les corrélations sont toujours très importantes entre les résultats de PISA, le PIB du pays, l'indice de niveau socio-économique et culturel, et entre les résultats de PISA et de TIMSS (voir paragraphe suivant), même si des différences apparaissent, détaillées par PISA et par l'OCDE dans leurs rapports.

Tableau 19 : Corrélations entre les scores des pays dans les différents domaines de PISA 2012

	Littératie scientifique	Compréhension de l'écrit	Résolution de problèmes
Littératie mathématique	0,93	0,90	0,83
Littératie scientifique		0,93	0,78
Compréhension de l'écrit			0,79

Source : PISA 2012 (indices extraits du rapport technique)

D'ailleurs, si l'on observe les corrélations entre les pays, elles sont encore plus fortes que les corrélations entre les individus, au point que certains se demandent si l'évaluation d'un seul de ces domaines, à savoir la compréhension de l'écrit, ne serait pas suffisante pour fournir finalement une information que les responsables des systèmes éducatifs attendent. Les figures 12 et 13 dans lesquelles chaque point représente un pays de l'OCDE illustrent parfaitement cette question.

Figure 12 : Corrélation entre les scores de littératie scientifique et de littératie mathématique à PISA 2012

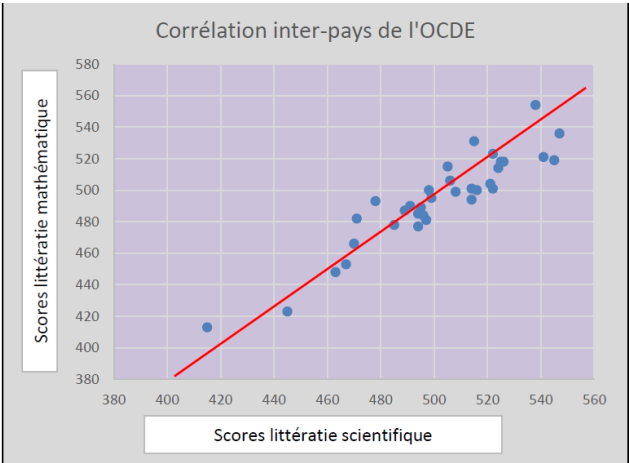
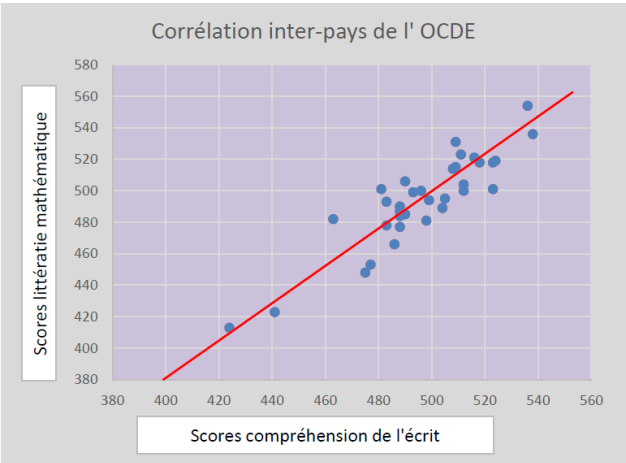


Figure 13 : Corrélation entre les scores en compréhension de l'écrit et de littératie mathématique à PISA 2012



Ces observations confirment que PISA évalue un domaine particulier, la littératie. Ce domaine, qui s'appuie sur des domaines disciplinaires, semble en réalité davantage dépendant du développement de processus cognitifs transdisciplinaires au développement desquels l'école participe, que de la maîtrise et de la spécificité des connaissances disciplinaires elles-mêmes.

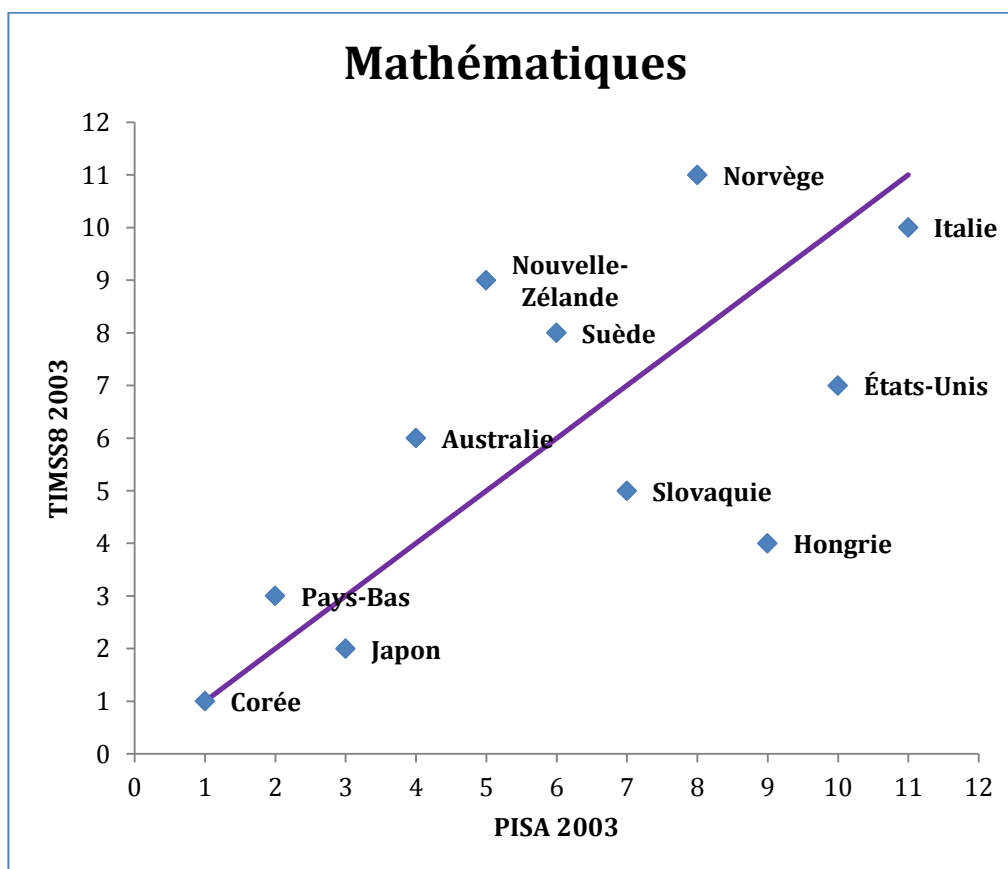
Comparaison des résultats des enquêtes PISA et des enquêtes TIMSS

L'analyse comparative des enquêtes PISA et TIMSS menée dans ce rapport a bien souligné des différences d'objectifs, de champs d'évaluation et de types d'exercices proposés aux élèves. On peut se demander si les pays qui participent aux deux enquêtes obtiennent des résultats comparables ou au contraire si des différences marquantes de résultats apparaissent selon qu'on s'intéresse à l'une ou l'autre de ces deux enquêtes.

Pour cela, nous avons comparé les résultats PISA avec ceux de TIMSS8, c'est-à-dire l'enquête de TIMSS qui s'adresse à des élèves de huitième année de scolarité, et que la France n'a pas passée, les âges des élèves ayant passé ces deux enquêtes étant proches.

Les figures suivantes ont été réalisées à partir des scores des pays qui ont participé à PISA 2003 et à TIMSS 2003 en mathématiques (Figure 14) et en sciences (Figure 15).

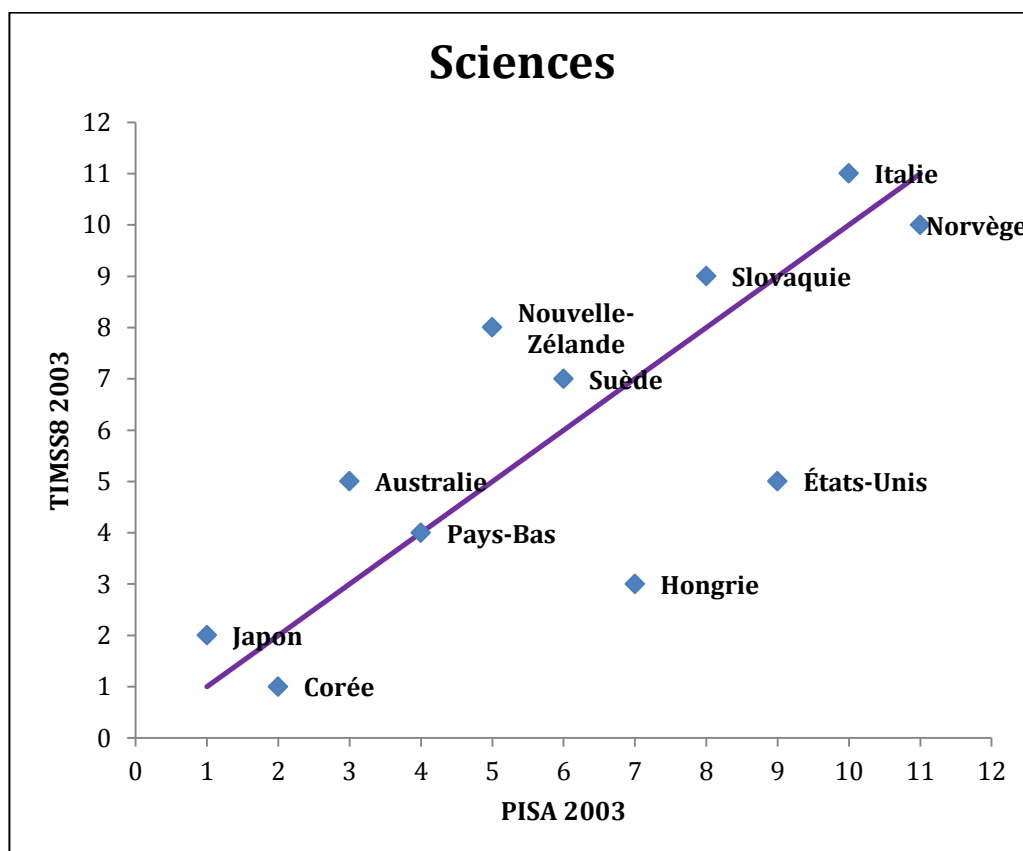
Figure 14 : Résultats en mathématiques des pays ayant participé à PISA et à TIMSS8 en 2003



Source : Rapports techniques de PISA 2003 et de TIMSS 2003

Lecture : Chacun des pays apparaissant sur le graphique a été reclassé parmi les pays ayant participé aux deux enquêtes selon son score. La Corée est ainsi 1^{re} à PISA 2003 et à TIMSS 2003, l'Australie 4^e à PISA et 6^e à TIMSS, les États-Unis 10^e à PISA et 7^e à TIMSS. La droite tracée en mauve correspond aux pays qui auraient le même classement aux deux enquêtes.

Figure 15 : Résultats en sciences des pays ayant participé à PISA et à TIMSS8 en 2003



Lecture : Chacun des pays apparaissant sur le graphique a été reclassé parmi les pays ayant participé aux deux enquêtes selon son score. La Corée est ainsi 2^e à PISA 2003 et à TIMSS 2003, l'Australie 3^e à PISA et 5^e à TIMSS, les États-Unis 9^e à PISA et 5^e à TIMSS. La droite tracée en mauve correspond aux pays qui auraient le même classement aux deux enquêtes.

On s'aperçoit que majoritairement, malgré les différences de conception, les résultats des pays aux deux enquêtes sont corrélés, que ce soit en mathématiques (le coefficient de corrélation est 0,66) ou en sciences (0,77), avec quelques pays plus « marqués » comme les États-Unis ou la Hongrie d'un côté, et l'Australie et la Nouvelle-Zélande de l'autre.

Nous avons réalisé le même travail de comparaison pour TIMSS 2011 et PISA 2012, et obtenu des résultats similaires, avec des corrélations élevées (0,66 en mathématiques et 0,80 en sciences), comme l'illustrent les figures 16 et 17.

Figure 16 : Résultats en mathématiques des pays ayant participé à PISA 2012 et à TIMSS8 2011

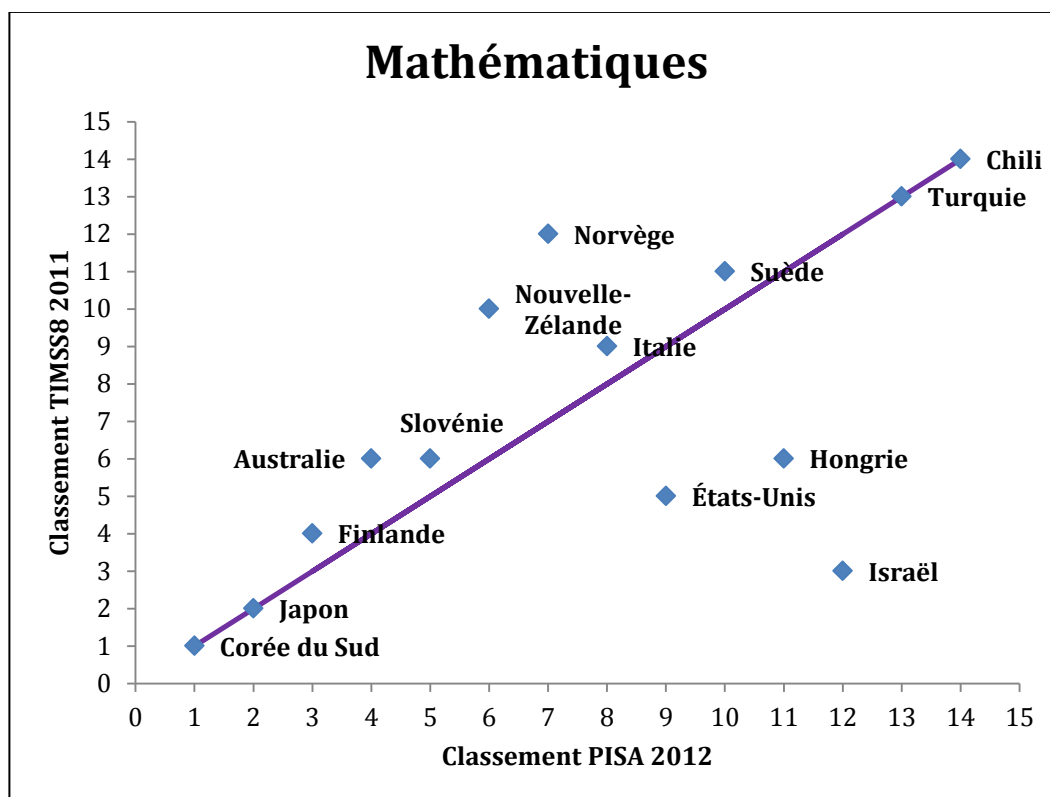
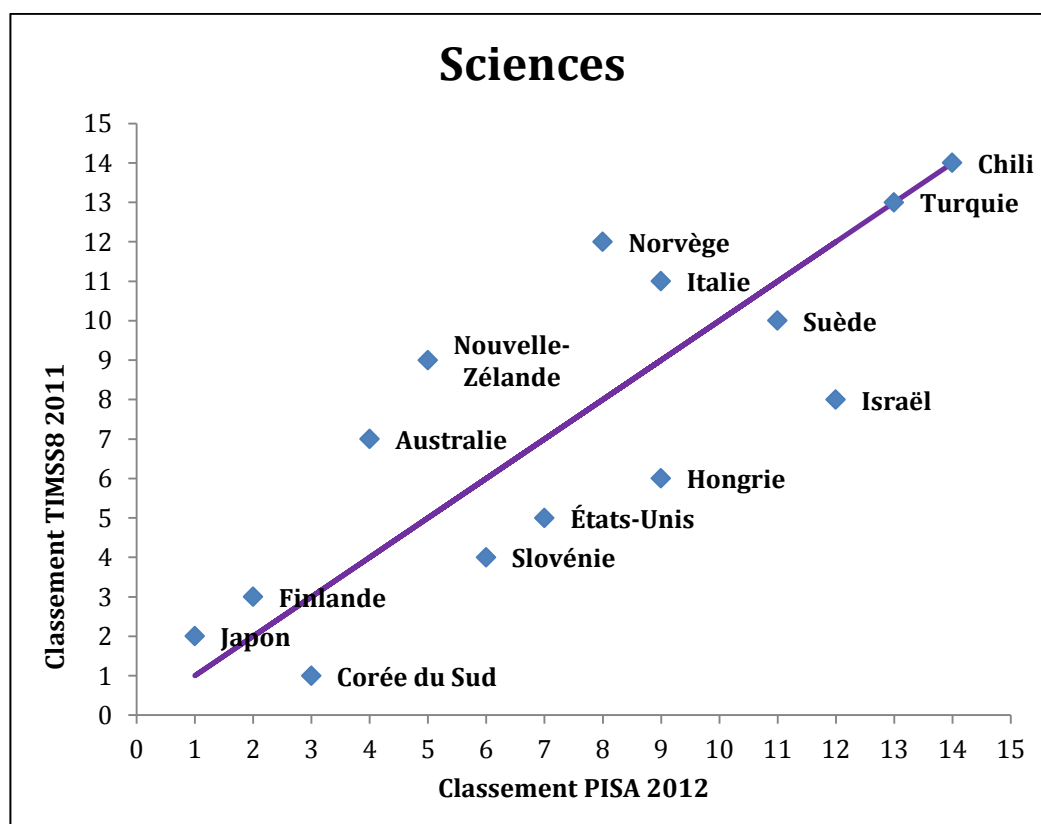


Figure 17 : Résultats en sciences des pays ayant participé à PISA 2012 et à TIMSS8 2011



Synthèse : Comparaison des programmes PISA et TIMSS

Les chapitres de ce premier volume du rapport ont mis en évidence des ressemblances et des différences entre les objectifs des programmes PISA et TIMSS, et entre les méthodes et les instruments qu'ils utilisent.

PISA ayant beaucoup emprunté à TIMSS et aux enquêtes de l'IEA, les deux programmes ont évidemment des points communs :

- PISA et TIMSS sont des programmes pluriannuels d'évaluation des élèves.
- PISA et TIMSS cherchent tous deux à obtenir la participation du plus grand nombre de pays possible à leurs enquêtes.
- Pour PISA comme pour TIMSS, les tests sont accompagnés de questionnaires de contexte (pour les élèves, les parents⁵⁹, les professeurs et les chefs d'établissement). Pour TIMSS, ces questionnaires alimentent surtout une base de connaissances sur les systèmes éducatifs tandis que PISA met plus systématiquement en relation les variables issues de ces questionnaires avec les résultats des tests et les transforme en indicateurs pour le pilotage des systèmes éducatifs.
- PISA et TIMSS pratiquent tous les deux une politique de données ouvertes. Sont ainsi en accès libre quasiment tout ce qui concerne l'organisation des études, les instruments utilisés et leurs modalités d'utilisation, les données recueillies, ainsi que l'ensemble des rapports, y compris les rapports techniques détaillant les méthodologies utilisées pour le traitement des données. Tous les questionnaires de contexte sont en accès libre. La seule exception concerne les questions d'évaluation : pour chacune des enquêtes, une partie de ces questions est gardée secrète pour pouvoir être utilisées à nouveau lors des enquêtes ultérieures afin d'autoriser des comparaisons temporelles.

⁵⁹ Sauf pour TIMSSADV

Les deux programmes présentent cependant des différences résumées dans le tableau 20.

Tableau 20 : Différences entre les programmes PISA et TIMSS

	PISA	TIMSS	Remarques
•	... est piloté par l'OCDE, organisation intergouvernementale qui, par définition, s'intéresse en premier lieu au développement de l'économie mondiale.	... est piloté par l'IEA qui est une association indépendante des États, centrée sur l'éducation et sur la recherche.	
•	... mène son programme dans l'ensemble des 34 pays de l'OCDE ainsi que dans un certain nombre de pays partenaires (37 pays en 2015).	La participation à TIMSS est volontaire et varie selon les niveaux évalués (57 pays en 2015 pour TIMSS 4 ^e ou 8 ^e année, et 9 pays pour TIMSS <i>Advanced</i>).	Dans les deux cas, la participation des pays est payante et l'évaluation se fait sur un échantillon d'élèves.
•	... fournit des résultats qui sont destinés, en premier lieu, aux responsables des systèmes éducatifs.	... fournit des résultats qui sont destinés, en premier lieu, aux chercheurs.	Chacun des deux programmes met en œuvre des enquêtes qui, à côté de leurs destinations premières, peuvent être utiles à de nombreux acteurs.
•	... est un programme d'évaluation pluriannuel qui existe depuis 2000 avec des cycles de 3 ans, et un domaine majeur à chaque enquête.	... est une étude comparative pluriannuelle qui existe depuis 1995 avec des cycles de 4 ans, et de façon plus irrégulière pour le niveau « terminale ».	
•	La France participe à PISA depuis sa première édition.	La France a participé à TIMSS lors de sa première édition, puis n'y a plus participé jusqu'en 2015.	
•	 évalue les élèves qui sont dans leur quinzième année au moment de la passation des tests.	... évalue les élèves de quatrième année scolaire (CM1 pour la France), de huitième année scolaire (4 ^e pour la France) et de fin d'études secondaires à orientation scientifique (terminale S pour la France) dans TIMSS <i>Advanced</i> .	En France, l'âge moyen des élèves qui ont passé les tests PISA en 2012 était de 15 ans et 8 mois. 55 % d'entre eux étaient en 2 ^{nde} GT, 11 % en 2 ^{nde} professionnelle et 30 % en 3 ^e ou 4 ^e générale (collège). La France n'a pas participé à TIMSS 2015 pour les élèves de 4 ^e .

	PISA	TIMSS	Remarques
•	... évalue les compétences en matière de littératie, dans ses différentes dimensions : compréhension de l'écrit, littératie mathématique, scientifique, financière, et résolution de problèmes	... évalue les connaissances et les savoir-faire scolaires relatifs aux mathématiques et aux sciences.	Dans les deux cas, les enquêtes portent aussi sur les processus que les élèves mettent en œuvre pour résoudre les exercices qui leur sont proposés.
•	... se focalise sur les compétences qui subsistent après l'école et qui sont de nature à permettre au citoyen de s'intégrer harmonieusement au monde actuel et à contribuer à son développement.	... est centrée sur les curriculums.	
•	... les questions sont pour la plupart ancrées sur des situations issues du monde réel.	...les questions sont pour la plupart directement ancrées dans le monde mathématique dans le cas des mathématiques ou dans un contexte scientifique pour ce qui est des sciences.	Dans les deux cas il s'agit de QCM ou de questions à réponse courte ou très courte.

Références et bibliographie

Documents officiels OCDE et PISA

OCDE (1999). Mesurer les connaissances et compétences des élèves Un nouveau cadre d'évaluation.

OECD (2000). Measuring student knowledge and skills. The PISA Assessment of Reading, Mathematical and Scientific literacy.

OCDE (2000b). Mesurer les connaissances et compétences des élèves. Lecture, Mathématiques et sciences : l'évaluation de PISA 2000.

OCDE (2002). Définitions and sélection des compétences (DESECO) : fondements théoriques et conceptuels. Document de stratégie.

OECD (2003). The PISA 2003 assessment framework. Mathematics, Reading, Science and Problem Solving Knowledge and Skills.

OCDE (2003b). Cadre d'évaluation de PISA 2003 – Connaissances et compétences en mathématiques, lecture, science et résolution de problèmes.

OCDE (2004) Apprendre aujourd'hui, réussir demain – Premiers résultats de PISA 2003.

OCDE (2005). Définition et sélection des compétences clés.

OCDE (2006). Compétences en sciences, lecture et mathématiques - Le cadre d'évaluation de PISA 2006.

OCDE (2009). Le cadre d'évaluation de PISA 2009. Les compétences clés en compréhension de l'écrit, en mathématiques et en sciences.

OECD (2009). Take the Test : sample Questions from OECD's PISA Assessments.

OECD (2013). Beyond PISA 2015 : a longer-term strategy of PISA. PISA Governing Board.

OECD (2013b). PISA 2012 Assessment and Analytical Framework Mathematics, Reading, Science, Problem Solving and Financial Literacy.

OCDE (2013c). Cadre d'évaluation et d'analyse du cycle PISA 2012

OCDE (2014). Principaux résultats de l'enquête PISA 2012 : ce que les élèves de 15 ans savent et ce qu'ils peuvent faire avec ce qu'ils savent.

OECD (2014). PISA 2012 Results: Creative Problem Solving Students' skills in tackling real-life problems (Volume V)

OECD (2014b). PISA 2012 Technical report.

OCDE (2016). Cadre d'évaluation et d'analyse de l'enquête PISA 2015.

OECD (2016b).PISA : Equations and Inequalities making mathematics accessible to all.

Documents officiels TIMSS

Schmidt, W.H. (1996). Many visions, many aims - A Cross-National Investigation of Curricular Intentions, Volume 1 - Mathematics. *Rapport de l'analyse des curricula de la Troisième Etude Internationale sur l'Enseignement des Mathématiques et des Sciences*, TIMSS de l'IEA, Kluwer Academics Publishers.

Robitaille, D.F., Schmidt, W. H., Raizen, S., McKnight, C., Britton, E. & Nico, C. (1993). TIMSS Third International Mathematics and Science Study, Monography n°1, Curriculum Frameworks for Mathematics and Science. Pacific Educational Press, U.B.C, Vancouver.

Mullis, I. & al. (1998). Mathematics and science achievement in the final year of secondary school : IEA's third international mathematics and science study (TIMSS).

Mullis, V. S., Martin, M. O., Foy, P. , & Arora, A. (2012) TIMSS 2011 International Results in Mathematics.

Mullis, I. & al (ed.) (2012). TIMSS 2011 Encyclopedia - Education Policy and Curriculum in Mathematics and Science (Vol 1 and 2). TIMSS & PIRLS International Study Center. Boston College.

Mullis, I. & Martin, M. (ed.) (2013). TIMSS 2015 Assessment Frameworks.

Mullis et al. (2016a). Methods and procedures in TIMSS 2015. Chapter 1 : developing the TIMSS 2015 achievement items.

Mullis et al. (2016b). Methods and procedures in TIMSS advanced 2015. Chapter 1 : developing the TIMSS 2015 achievement items.

Documents officiels du ministère de l'Éducation nationale (France)

DEP (1996). Les connaissances en mathématiques et en physique des élèves de terminale scientifique. *Note d'information*, n°96.50.

DEP (1996). Les connaissances des élèves en mathématiques et en sciences en terminale. *Note d'information*, n°96.49.

DEP (1997). Évaluation internationale en mathématiques et en sciences des élèves de cinquième et de quatrième. *Note d'information*, n°97.06.

DEP (2001). Les élèves de 15 ans - Premiers résultats d'une évaluation internationale des acquis des élèves (PISA). *Note d'information*, n°01.52.

Depp (2004). Les élèves de 15 ans. Premiers résultats de l'évaluation internationale PISA 2003. *Note d'information* n°04.12.

Depp (2007). L'évolution des acquis des élèves de 15 ans en culture mathématique et en compréhension de l'écrit. Premiers résultats de l'évaluation internationale PISA 2006. *Note d'information*, n°08.08.

Depp (2007). L'évaluation internationale PISA 2003 : compétences des élèves français en mathématiques, compréhension de l'écrit et sciences. *Éducation & formations*, n°180.

Depp (2008). Comparaisons Internationales. *Éducation & formations*, n°78.

Depp (2010). L'évolution des acquis des élèves de 15 ans en culture mathématique et en culture scientifique. Premiers résultats de l'évaluation internationale PISA 2009. *Note d'information*, n°10.23.

Depp (2013). L'évolution des acquis des élèves de 15 ans en compréhension de l'écrit et en culture scientifique - Premiers résultats de l'évaluation internationale PISA 2012. *Note d'information*, n°13.30.

Depp (2013) : Les élèves de 15 ans en France selon PISA 2012 en culture mathématique : baisse des performances et augmentation des inégalités par rapport à 2003. *Note d'information*, n°13.31.

Références

Artigue, M. & Winslow, C. (2010). International comparative studies on mathematics education : a viewpoint from the anthropological theory of didactics. *Recherches en Didactique des Mathématiques*, 30(1), 47-82.

Arzoumanian, P., Bessonneau, P. et Pastor, J-M. (2015). « Une évaluation sous forme numérique est-elle comparable à une évaluation de type « papier-crayon » ? » », *Education et formations*, n°86-87. MENESR-Depp.

Bodin, A. (1997). *Une présentation de la Troisième Étude Internationale sur l'enseignement des Mathématiques et des Sciences - Considérations sur la démarche, sur les résultats, sur l'intérêt de l'étude* - Dossier d'information sur TIMSS - IREM de Besançon.

Bodin, A. (2006a). Ce qui est vraiment évalué par PISA en mathématiques. Ce qui ne l'est pas. Un point de vue français. Communication faite à la conférence Franco-Finlandaise sur PISA 6-8 octobre 2005 - Paris *Bulletin de l'APMEP*, N°463.

Bodin, A. (2006b). Les mathématiques face aux évaluations nationales et internationales. De la première étude menée en 1960 aux études TIMSS et PISA en passant par les études de la DEP et d'EVAPM. *Communication séminaire de l'EHESS*. Repères IREM, N°65, octobre 2006.

Bodin, A. (2006c). Un point de vue sur PISA. *Gazette des mathématiciens* N°108 – Avril 2006 - Société Mathématique de France (SMF).

Bodin, A. (2007). What does Pisa really assess, in S. Hopman, G. Brinek, M. Retzl (éds): *PISA according to PISA*. Wien: Lit Verlag, 2007.

Bodin, A. (2008a). French Pisa Mathematics Results and Reactions - *Paper to the Second Iberian Mathematical Meeting* - Badajoz, Spain, October 3-5, 2008.

Bodin, A. (2008b). Lecture et utilisation de PISA pour les enseignants. *Petit x* ; n° 78, pp. 53-78, IREM de Grenoble.

Bottani, N., & Vrignaud, P. (2005). La France et les évaluations internationales. *Les rapports établis à la demande du Haut Conseil de l'Évaluation de l'École*, 16, 180 pages.

Chevallard, Y. (1991). *La transposition didactique - du savoir savant au savoir enseigné*. Grenoble : La pensée sauvage.

Douady, R (1986). Jeux de cadres et dialectique outil-objet - *Recherches en Didactique des Mathématiques*, Vol 7/2.

Grapin, N. (2015). *Étude de la validité de dispositifs d'évaluation et conception d'un modèle d'analyse multidimensionnelle des connaissances numériques des élèves de fin d'école*. Thèse, Université Paris-Diderot (Paris 7).

Roditi, E. & Salles, F. (2015). Nouvelles analyses de l'enquête Pisa 2012 en mathématiques. Un autre regard sur les résultats. *Éducation & Formations*, n°86-87.

Vergnaud, G. (1981). *Quelques orientations théoriques et méthodologiques des recherches françaises en didactique des mathématiques* » — R.D.M. Vol. 2.2, La Pensée Sauvage, Grenoble, pp. 215-232.

Vrignaud, P. (2008). Measuring literacy and the international assessment surveys: methodology is the answer, but what was the question? *Education & Formation, Special Issue*.

Wu, M. (2009). A Critical Comparison of the Contents of PISA and TIMSS Mathematics Assessments. University of Melbourne Assessment Research Centre.

Liste des sigles et acronymes

ACER	<i>Australian Council for Educational Research.</i>
ALL	<i>Adult Literacy and Life Skills Survey</i> Enquête sur la littératie et les compétences des adultes
CEDRE	Cycle des évaluations disciplinaires réalisées sur échantillon
DEP	Direction de l'évaluation et de la prospective
Depp	Direction de l'évaluation, de la prospective et de la performance
ETS	<i>Educational Testing Service</i>
EVAPM	Evaluation des apprentissages mathématiques
FIMS	<i>First International Mathematics Study</i> Première étude internationale sur l'enseignement des mathématiques
IEA	<i>The International Association for the Evaluation of Educational Achievement</i> Association internationale sur l'évaluation et la réussite scolaire
KOM	<i>Competencies and Mathematics Learning</i>
MENESR	Ministère de l'Éducation nationale, de l'enseignement supérieur et de la recherche
OECD	<i>Organisation for Economic Co-operation and Development</i>
OCDE	Organisation de coopération et de développement économiques
OM	Organisation mathématique
PIAAC	<i>Program for the International Assessment of Adult Competencies</i> Programme pour l'évaluation internationale des compétences des adultes
PIB	Produit Intérieur Brut
PIRLS	<i>Progress in Reading Literacy Study</i> Programme international de recherche en lecture scolaire
PISA	<i>Program for International Student Achievement.</i> Programme international pour l'évaluation des élèves
QCM	Questionnaire à choix multiples
SIMS	<i>Second International Mathematics Study</i> Seconde étude internationale sur l'enseignement des mathématiques et des sciences
TALIS	<i>Teaching and Learning International Survey</i> Enquête internationale sur l'enseignement et l'apprentissage
TIC	Technologies de l'information et de la communication
TIMSS	<i>Third International Mathematics and Science Study</i> Troisième étude internationale sur l'Enseignement des Mathématiques et des sciences <i>Trends in International Mathematics and Science Study</i> Tendances internationales dans l'enseignement des mathématiques et des sciences
TIMSS4	TIMSS <i>grade</i> 4 (niveau CM1 en France)
TIMSS8	TIMSS <i>grade</i> 8 (niveau quatrième en France)
TIMSSADV	TIMSS <i>Advanced</i> (terminale S en France)
UNESCO	Organisation des Nations unies pour l'éducation, la science et la culture



cnesco
**conseil national
d'évaluation
du système scolaire**

Carré Suffren
31-35 rue de la Fédération
75 015 Paris

Tél. 01 55 55 02 09

cnesco.communication@education.gouv.fr

 www.cnesco.fr

 @Cnesco

 Cnesco