Annexe A

CADRE TECHNIQUE DE L'ENQUÊTE PISA 2015

L'ensemble des tableaux de l'annexe A sont disponibles en ligne

Annexe A1 : Indices extraits des questionnaires « Élève » et « Établissement »

Annexe A2 : Population cible, échantillons et définition des établissements de l'enquête PISA
http://dx.doi.org/10.1787/888933433129

Annexe A3 : Note technique sur les analyses du présent volume

Annexe A4 : Assurance qualité

Annexe A5 : Changements de mode d'administration et de procédure de mise à l'échelle lors de l'évaluation PISA 2015 et implications pour les analyses de tendances
http://dx.doi.org/10.1787/888933433162

Annexe A6 : Étude de l'effet du mode d'administration de l'enquête PISA 2015 dans le cadre de l'essai de terrain

Note concernant P-S-J-G (Chine)
P-S-J-G (Chine) fait référence aux quatre provinces chinoises participant à l'enquête PISA : Pékin, Shanghai, Jiangsu et Guangdong.

Note concernant CABA (Argentine)
CABA (Argentine) fait référence à la région Ciudad Autónoma de Buenos Aires, dont les données ont été adjugées.

Note concernant ERYM
ERYM fait référence à l'ex-République yougoslave de Macédoine.

Note concernant Chypre

Note de la Turquie : Les informations figurant dans ce document qui font référence à « Chypre » concernent la partie méridionale de l'Ile. Il n'y a pas d'autorité unique représentant à la fois les Chypriotes turcs et grecs sur l'Ile. La Turquie reconnaît la République Turque de Chypre Nord (RTCN). Jusqu'à ce qu'une solution durable et équitable soit trouvée dans le cadre des Nations Unies, la Turquie maintiendra sa position sur la « question chypriote ».

Note de tous les États de l'Union européenne membres de l'OCDE et de l'Union européenne : La République de Chypre est reconnue par tous les membres des Nations Unies sauf la Turquie. Les informations figurant dans ce document concernent la zone sous le contrôle effectif du gouvernement de la République de Chypre.

Note concernant les données d'Israël

Les données statistiques concernant Israël sont fournies par et sous la responsabilité des autorités israéliennes compétentes. L'utilisation de ces données par l'OCDE est sans préjudice du statut des hauteurs du Golan, de Jérusalem-Est et des colonies de peuplement israéliennes en Cisjordanie aux termes du droit international.

ANNEXE A1
INDICES EXTRAITS DES QUESTIONNAIRES « ÉLÈVE » ET « ÉTABLISSEMENT »

Description des indices

Cette section décrit les indices utilisés dans le présent volume et dérivés des questionnaires « Élève » et « Établissement » de l'évaluation PISA 2015.

Plusieurs indicateurs PISA sont basés sur des indices qui résument les réponses des élèves, de leurs parents et des enseignants ou des représentants de leur établissement (le chef d'établissement, le plus souvent) à une série de questions. Ces questions ont été sélectionnées dans des constructs plus importants sur la base de considérations théoriques et de recherches antérieures. Le *Cadre d'évaluation et d'analyse de l'enquête PISA 2015* (OCDE, 2016) fournit une description approfondie de ce cadre conceptuel. Des équations structurelles ont été modélisées pour confirmer les dimensions théoriques prévues et valider leur comparabilité entre pays. À cet effet, un modèle individuel a été préparé pour chaque pays et un modèle collectif a été réalisé à l'échelle de l'OCDE. Pour une description détaillée d'autres indices PISA et des méthodes utilisées, voir le rapport technique sur l'enquête PISA 2015 (*PISA 2015 Technical Report* [OCDE, à paraître en anglais uniquement]).

Il existe trois types d'indices : les indices simples, les nouveaux indices mis à l'échelle et les indices de tendance mis à l'échelle.

Les **indices simples** sont les variables calculées après traitement arithmétique ou recodage d'un ou de plusieurs items. Les réponses aux items sont utilisées pour calculer des variables pertinentes, par exemple la transposition des codes de la CITP-08 en « indice socio-économique le plus élevé des deux parents (HISEI) » ou le calcul du taux d'encadrement sur la base des informations recueillies dans le questionnaire « Établissement ».

Les **nouveaux indices mis à l'échelle et les indices de tendance mis à l'échelle** sont les variables calculées après mise à l'échelle de plusieurs items. Sauf mention contraire, lorsqu'un indice se fonde sur les réponses à plusieurs items, il est mis à l'échelle selon un modèle de réponse à l'item à deux paramètres (un modèle de crédit partiel généralisé est utilisé si les items sont assortis de plus de deux catégories de réponse), et ses valeurs correspondent aux estimations pondérées des réponses les plus vraisemblables (*weighted likelihood estimate*, WLE) (Warm, 1985). Pour des informations plus détaillées sur la construction de chaque indice mis à l'échelle, consulter le rapport technique sur l'enquête PISA 2015 (*PISA 2015 Technical Report* [OCDE, à paraître en anglais uniquement]). Ce processus se déroule généralement en trois étapes :

1. Les paramètres d'items sont estimés à partir d'échantillons d'élèves de taille constante dans chaque pays et économie ; seuls les cas comportant un minimum de trois réponses valides aux items composant l'indice ont été inclus. Dans le cas des **indices de tendance**, une méthode d'étalonnage commune a été utilisée pour mettre en correspondance les procédures : les pays/économies ayant pris part aux enquêtes PISA 2006 et PISA 2015 ont participé aux échantillons pour l'étalonnage des paramètres d'items ; chaque évaluation, et, lors de chaque évaluation, chaque pays/économies ont contribué à l'estimation à parts égales.
2. Les estimations sont calculées compte tenu de tous les élèves et de tous les établissements sur la base de l'ancrage des paramètres d'item obtenus lors de l'étape précédente.
3. Pour les **nouveaux indices mis à l'échelle**, les estimations pondérées des réponses les plus vraisemblables sont normalisées de sorte que la valeur moyenne pour l'effectif d'élèves de l'OCDE est égale à 0 et l'écart-type, à 1 (pondération équivalente des pays lors du processus de normalisation). Une équivalence est créée au niveau des **indices de tendance** de sorte que, dans les pays de l'OCDE, les valeurs de la moyenne et de l'écart-type des estimations de l'évaluation PISA 2006, après nouvelle mise à l'échelle, correspondent aux estimations initiales incluses dans la base de données PISA 2006. Les indices de tendance sont donc rapportés sur la même échelle que celle utilisée initialement pour l'évaluation PISA 2006, afin que les valeurs puissent être directement comparables à celles comprises dans la base de données PISA 2006.

Des codes séquentiels sont attribués à chaque catégorie de réponse, dans l'ordre où ces catégories figurent dans les questionnaires « Élève », « Établissement » ou « Parents ». Dans certains indices ou échelles, les codes ont été inversés, ainsi que le précise la présente section. Les valeurs négatives d'un indice n'impliquent pas forcément que les élèves ont répondu par la négative aux questions qui y sont associées. Ces valeurs signifient uniquement qu'ils ont répondu moins positivement que ne l'ont fait, en moyenne, tous les élèves de l'OCDE. De même, les valeurs positives d'un indice indiquent qu'ils ont répondu plus favorablement, ou plus positivement, que ne l'ont fait en moyenne tous les élèves des pays de l'OCDE. Dans les descriptions suivantes, les termes figurant entre crochets < > ont été remplacés par un équivalent approprié dans les versions nationales des questionnaires « Élève », « Parents » et « Établissement ». Ainsi, l'expression <diplôme de niveau 5A de la CITE> a été traduite aux États-Unis par <Bachelor's degree, post-graduate certificate program, Master's degree program or first professional degree program>. De même, au Luxembourg, l'expression <cours en langue de l'évaluation> a été remplacée par <cours d'allemand> et <cours de français> respectivement dans les versions allemande et française des instruments d'évaluation.

Outre les indices simples et les indices mis à l'échelle décrits ici, un certain nombre de variables dérivées des questionnaires ont été employées dans le présent volume et correspondent à des items uniques ne participant pas à la construction des indices. Ces variables qui n'ont pas été recodées sont précédées du préfixe <ST> si elles sont dérivées du questionnaire « Élève » et du préfixe <SC> si elles sont dérivées du questionnaire « Établissement ». Les questionnaires contextuels et la base de données internationale de l'enquête PISA, où sont enregistrées toutes les variables, peuvent être consultés sur le site de l'enquête PISA (www.oecd.org/pisa).

Indices simples de niveau élève

Âge des élèves

L'âge d'un élève (ÂGE) correspond à la différence entre le mois et l'année de l'évaluation, et le mois et l'année de naissance des élèves. Les données relatives à l'âge des élèves ont été obtenues à partir du questionnaire (ST003) et des formulaires de suivi des élèves. Si, pour un élève donné, on ignore le mois de l'évaluation, le calcul est effectué sur la base du mois médian pour le pays en question.

Niveau de formation des parents

Les réponses des élèves aux questions ST005, ST006, ST007 et ST008 concernant la formation des parents ont été codifiées conformément à la CITE de 1997 (OCDE, 1999). Les indices sur la formation des parents ont été élaborés selon le recodage des niveaux de formation classés dans l'une des catégories suivantes : (0) pas de formation ; (1) <Niveau CITE 1> (enseignement primaire) ; (2) <Niveau CITE 2> (premier cycle de l'enseignement secondaire) ; (3) <Niveau CITE 3B ou 3C> (filière préprofessionnelle ou professionnelle du deuxième cycle de l'enseignement secondaire) ; (4) <Niveau CITE 3A> (deuxième cycle de l'enseignement secondaire) et/ou <Niveau CITE 4> (enseignement post-secondaire non tertiaire) ; (5) <Niveau CITE 5B> (enseignement tertiaire à vocation professionnelle) ; et (6) <Niveau CITE 5A> et/ou <Niveau CITE 6> (enseignement tertiaire théorique et troisième cycle). Les indices avec ces catégories ont été fournis pour la mère (MISCED) et le père (FISCED) d'un élève. En outre, l'indice du niveau de formation le plus élevé des parents (HISCED) correspond au niveau le plus élevé de la CITE atteint par l'un ou l'autre parent. Cet indice est également recodé en une estimation du nombre d'années d'études (PARED). La correspondance entre les niveaux de formation et les années d'études est disponible dans le rapport technique sur l'évaluation PISA 2015, *PISA 2015 Technical Report* (OCDE, à paraître en anglais uniquement).

Statut professionnel le plus élevé des parents

Les données sur le statut professionnel du père et de la mère d'un élève proviennent de réponses à des questions ouvertes. Les réponses ont été codées à l'aide des codes de la CITP à quatre chiffres (OIT, 2007), puis cartographiées selon l'indice socio-économique international du statut professionnel (ISEI) (Ganzeboom et Treiman, 2003). Tant dans l'enquête PISA 2015 que dans l'édition 2012, les versions 2008 de la nouvelle Classification CITP et du nouvel indice ISEI ont été préférées aux versions de 1988 utilisées lors des quatre évaluations précédentes (Ganzeboom, 2010). Trois indices sont calculés en fonction de ces informations : le statut professionnel du père (BFMJ2), le statut professionnel de la mère (BMMJ1), et le statut professionnel le plus élevé des parents (HISEI), qui correspond à l'indice ISEI le plus élevé des deux parents ou à l'indice ISEI du seul parent en cas de famille monoparentale. Pour les trois indices, une valeur élevée de l'indice ISEI dénote un statut professionnel élevé.

Statut au regard de l'immigration

La base de données PISA contient trois variables nationales relatives au pays natal des élèves, à celui de leur mère et de leur père (COBN_S, COBN_M, et COBN_F). Les items ST019Q01TA, ST019Q01TB et ST019Q01TC ont été recodés et classés dans l'une des catégories suivantes : (1) le pays natal est le même que le pays de l'évaluation ; et (2) autre. L'indice du statut au regard de l'immigration (IMMIG) a été calculé en fonction de ces variables classées dans l'une des catégories suivantes : (1) les élèves autochtones (élèves dont au moins un des parents est né dans le pays) ; (2) les élèves immigrés de la deuxième génération (élèves nés dans le pays de l'évaluation, mais dont l'un des parents, ou les deux, sont nés à l'étranger) ; et (3) les élèves immigrés de la première génération (élèves nés à l'étranger de parents nés à l'étranger). Les données sont déclarées manquantes si les élèves n'ont pas répondu à la question les concernant ou aux deux questions concernant leurs parents.

Langue parlée en famille

Les élèves ont indiqué la langue qu'ils parlaient le plus souvent à la maison (ST022), et la base de données inclut une variable calculée (LANGN) contenant un code pays pour chaque langue. En outre, une variable comparable au niveau international (ST022Q01TA) a été dérivée de cette information et classée dans l'une des catégories suivantes : (1) la langue parlée en famille est identique à la langue de l'évaluation ; et (2) la langue parlée en famille est une autre langue.

Redoublement

La variable du redoublement (REPEAT) a été calculée sur la base du recodage des variables ST127Q01TA, ST127Q02TA, et ST127Q03TA. Une valeur de la variable REPEAT égale à « 1 » signifie que l'élève a redoublé au moins un niveau de la CITE, et une valeur égale à « 0 » signifie que l'option « Non, jamais » a été sélectionnée au moins une fois, puisque aucune des catégories d'années d'études redoublées n'a été choisie. L'indice est déclaré manquant si l'élève n'a coché aucune des trois catégories à aucun niveau.

Programme de cours

L'enquête PISA permet la collecte de données concernant les programmes de cours proposés aux élèves de 15 ans dans chaque pays. Ces informations proviennent des formulaires de suivi des élèves et du questionnaire « Élève ». Dans la base de données définitive, tous les programmes nationaux sont intégrés dans une variable distincte (PROGN) dont les six premiers chiffres se rapportent au code du centre national et les deux derniers chiffres, au code national du programme de cours. Tous les programmes de cours ont été classés selon la Classification internationale type de l'éducation (CITE) (OCDE, 1999). Les indices suivants sont dérivés des données relatives aux programmes de cours :

- Le niveau du programme (ISCEDL) indique si les élèves fréquentent un établissement du premier ou deuxième cycle de l'enseignement secondaire (niveaux 2 et 3 de la CITE).
- La variable ISCEDD désigne la typologie du programme de cours (A = programmes à vocation générale qui permettent d'accéder au niveau d'enseignement suivant ; B = programmes qui permettent d'accéder à des études à vocation professionnelle au niveau d'enseignement suivant ; C = programmes qui permettent d'accéder directement au marché du travail ; et M = programmes mixtes qui combinent tout ou partie de ces caractéristiques).
- L'orientation du programme (ISCEDO) indique si son contenu pédagogique est général, préprofessionnel ou professionnel.

Perspectives de carrière dans une profession scientifique

Dans l'enquête PISA 2015, on a demandé aux élèves de répondre à une question (ST114) concernant le « type de métier [qu'ils espèrent] exercer quand [ils auront] environ 30 ans ». Les réponses à cette question ouverte ont été codées à l'aide des codes de la CITP à quatre chiffres (OIT, 2007), dans la variable OCOD3. Cette variable a permis d'établir l'indice des perspectives de carrière dans une profession scientifique.

Les perspectives d'exercer une profession scientifique sont définies comme celles impliquant la poursuite des études en sciences au-delà de la scolarité obligatoire, vraisemblablement dans des établissements d'enseignement tertiaire. Le classement des professions dans la catégorie scientifique ou non scientifique repose sur la classification des professions CITP-08 à quatre chiffres.

On a estimé que seules les professions intellectuelles (grand groupe 2 de la CITP) et les professions intermédiaires (grand groupe 3 de la CITP) répondaient à la définition des perspectives de carrière dans une profession scientifique. Au sens large, plusieurs professions de directeurs, cadres de direction et gérants (grand groupe 1 de la CITP) présentent clairement un rapport avec la science, dont les suivantes : directeurs et cadres de direction, recherche-développement ; directeurs et cadres de direction, hôpital ; directeurs et cadres de direction, bâtiment ; et autres professions classées dans le sous-grand groupe 13 intitulé « Directeurs et cadres de direction, production et services spécialisés ». Néanmoins, lorsqu'une expérience et une formation en rapport avec la science est un critère important pour exercer un poste de direction, on considère qu'il ne s'agit pas là d'emplois de premier échelon et que les élèves de 15 ans qui souhaitent se destiner à une carrière scientifique ne s'attendraient pas à occuper un tel poste à l'âge de 30 ans.

De nombreux agriculteurs et ouvriers qualifiés de l'agriculture, de la sylviculture et de la pêche (grand groupe 6 de la CITP) pourraient également être considérés comme exerçant une profession scientifique. D'ailleurs, aux États-Unis, le classement des professions en rapport avec les sciences, la technologie, l'ingénierie et les mathématiques (STIM), intitulé « O*NET Online » (2016), tient compte desdites professions. Cependant, elles ne nécessitent généralement pas de formation ou d'études scientifiques une fois la scolarité obligatoire terminée. Pour ces motifs, seules les principales catégories de professions qui exigent les niveaux de compétence 3 et 4 de la CITP ont été prises en compte dans le cadre des perspectives de carrière dans une profession scientifique.

Au sein des professions intellectuelles et professions techniques intermédiaires, la frontière séparant les professions scientifiques des professions non scientifiques est parfois floue, et les diverses classifications utilisent différentes méthodes de séparation.

La classification utilisée dans ce rapport comprend les quatre catégories d'emplois suivantes[1] :

1. ***Spécialistes des sciences techniques :*** toutes les professions intellectuelles et scientifiques (sous-grand groupe 21), à l'exception des concepteurs modélistes de produits et de vêtements (2163), ainsi que des concepteurs graphiques et des multimédia-graphistes (2166).
2. ***Spécialistes de la santé :*** toutes les professions médicales du sous-grand groupe 22 (médecins, infirmiers, vétérinaires par exemple), à l'exception des professions liées à la médecine traditionnelle et médecine complémentaire (sous-groupe 223).
3. ***Spécialistes des technologies de l'information et des communications :*** toutes les professions des technologies de l'information et des communications (sous-grand groupe 25).
4. ***Professions intermédiaires à caractère scientifique,*** y compris :
   - les techniciens des sciences physiques et techniques (sous-groupe 311)
   - les techniciens et travailleurs assimilés des sciences de la vie (sous-groupe 314)
   - les techniciens de la sécurité aérienne (3155)

- les techniciens de la médecine et de la pharmacie (sous-groupe 321), à l'exception des techniciens de prothèses médicales et dentaires (3214)
- les techniciens de télécommunications (3522).

Comparaison entre la présente classification et les autres classifications

Quelques différences sont constatées lorsque l'on compare les trois classifications des perspectives de carrière dans une profession scientifique pour les élèves de 15 ans, toutes fondées sur la version de 1988 de la Classification internationale type des professions (CITP-88), avec la présente classification basée sur la CITP-08. Certaines s'expliquent par la mise à jour du codage des professions (tel qu'examiné dans la prochaine section) ; les autres différences sont résumées dans le tableau A1.1.

Établir une classification comparable pour la CITP-88

La même question ouverte a été intégrée dans le questionnaire de l'enquête PISA 2006 (identifiant en 2006 : ST30), mais les réponses des élèves ayant été codées dans la base de données PISA 2006 en fonction de la CITP-88, il n'est pas possible de garantir une classification strictement comparable. Pour rendre compte des évolutions dans le temps, la correspondance décrite dans le tableau A1.2 a servi à établir une classification analogue reposant sur les données de l'enquête PISA 2006 :

Tableau A1.1 Différences de définition quant aux perspectives de carrière dans une profession scientifique

	La présente classification	OCDE (2007)	Sikora et Pokropek (2012)	Kjærnsli et Lie (2011)
Postes de direction en rapport avec la science	absent	présent	présent	absent
Psychologues	absent	présent	présent	absent
Sociologues et spécialistes du travail social	absent	présent	absent	absent
Photographes et techniciens d'appareils enregistreurs d'images et de son, techniciens de matériels d'émissions de radio, de télévision et de télécommunications	absent	présent	présent	absent
Professions intermédiaires de la statistique, des mathématiques et assimilées	absent	absent	présent	absent
Contrôleurs du moyen de transport aérien (par exemple, contrôleurs de la circulation aérienne)	absent	présent	présent	absent
Contrôleurs du moyen de transport maritime (officiers de pont, etc.)	absent	absent	présent	absent
Assistants médicaux, assistants en médecine dentaire, assistants vétérinaires, personnel infirmier et sages-femmes (niveau intermédiaire)	absent	présent	présent	absent
Assistants informaticiens, techniciens de matériels informatiques et techniciens de robots industriels	absent	absent	absent	présent
Techniciens de la sécurité aérienne	présent	présent	présent	absent
Techniciens et assistants pharmaciens et préparateurs en pharmacie	présent	présent	présent	absent
Diététiciens et spécialistes de la nutrition	présent	présent	présent	absent

Tableau A1.2 Tableau de correspondance entre la CITP-08 et la CITP-88 pour les perspectives de carrière dans une profession scientifique

Groupe	CITP-08	CITP-88
Spécialistes des sciences techniques	21xx (sauf 2163 et 2166)	21xx (sauf 213x), 221x
Spécialistes de la santé	22xx (sauf 223x)	22xx (sauf 221x), 3223, 3226
Spécialistes des technologies de l'information et des communications	25xx	213x
Professions intermédiaires à caractère scientifique	311x, 314x, 3155, 321x (sauf 3214), 3522	311x, 3133, 3145, 3151, 321x, 3228

Dans le but de calculer l'indice des perspectives de carrière dans une profession scientifique, les principales différences à prendre en compte entre la CITP-88 et la CITP-08 sont les suivantes :

- Les inspecteurs de sécurité incendie et d'immeubles (CITP-88 : 3151) correspondent principalement aux techniciens du génie civil dans la CITP-08.
- Les diététiciens et spécialistes de la nutrition (CITP-88 : 3223) sont classés parmi les spécialistes dans la CITP-08. Par souci d'uniformité, cette profession de la CITP-88 a été classée parmi les professions médicales.
- Les kinésithérapeutes et assimilés (CITP-88 : 3226) forment deux catégories distinctes dans la CITP-08, les kinésithérapeutes intégrant la catégorie des spécialistes. Étant donné que les élèves espérant exercer la profession de kinésithérapeute sont bien plus nombreux que ceux souhaitant devenir travailleurs assimilés, cette profession de la CITP-88 a été classée parmi les professions médicales.
- Plusieurs professions médicales classées comme « Professions intermédiaires de la médecine moderne » dans la CITP-88, intègrent la catégorie des professions médicales dans la CITP-08 (par exemple, orthophonistes, optométristes). Tandis que les professions médicales sont, en général, intégrées aux carrières scientifiques, ce n'est pas le cas des professions médicales intermédiaires. En appliquant la classification à la CITP-88, l'ensemble du code a été exclu des carrières scientifiques.
- Les techniciens de télécommunications (CITP-08 : 3522) ne constituent pas une profession distincte dans la CITP-88, où on les retrouve parmi les techniciens en électronique et en télécommunications (CITP-88 : 3114).
- Les spécialistes des technologies de l'information et des communications forment un sous-grand groupe à part (25) dans la CITP-08, mais sont classés parmi les spécialistes des sciences physiques, mathématiques et techniques dans la CITP-88.

Indices mis à l'échelle de niveau Élève

Nouveaux indices mis à l'échelle

Intérêt pour les sciences

L'indice de l'intérêt pour les grands sujets scientifiques (INTBRSCI) a été élaboré à partir des réponses données par les élèves à une question créée pour l'enquête PISA 2015 (ST095). Les élèves ont indiqué sur une échelle de Likert composée de cinq échelons (« Pas intéressé(e) », « Peu intéressé(e) », « Intéressé(e) », « Très intéressé(e) » et « Je ne sais pas ce que c'est ») dans quelle mesure ils s'intéressaient aux sujets scientifiques suivants : la biosphère (par exemple, les écosystèmes, le développement durable) ; le mouvement et les forces (par exemple, la vitesse, la friction, les forces magnétiques et gravitationnelles) ; l'énergie et sa transformation (par exemple, la conservation, les réactions chimiques) ; l'Univers et son histoire ; et la façon dont la science peut aider à éviter certaines maladies. La dernière catégorie de réponse (« Je ne sais pas ce que c'est ») a été recodée comme manquante aux fins de la création de l'indice INTBRSCI. Les valeurs plus élevées de l'indice dénotent des degrés d'assentiment plus élevés avec ces affirmations.

Convictions épistémiques à propos de la science

L'indice des convictions épistémiques à propos de la science (EPIST) a été élaboré à partir des réponses données par les élèves à une question créée pour l'enquête PISA 2015 et portant sur la position des élèves concernant les méthodes scientifiques (ST131). Sur une échelle de Linkert composée de quatre échelons (« Pas du tout d'accord », « Pas d'accord », « D'accord » et « Tout à fait d'accord »), les élèves ont indiqué leur degré d'assentiment avec les affirmations suivantes : « Une bonne façon de savoir si quelque chose est vrai, c'est de faire une expérience » ; « Il arrive que les concepts en <sciences> changent » ; « Les bonnes réponses sont basées sur des éléments de preuve issus de nombreuses expériences différentes » ; « Il est préférable de répéter des expériences plusieurs fois pour être sûr des résultats » ; « Parfois les scientifiques changent d'avis sur ce qui est considéré comme vrai en <sciences> » ; et « Il arrive que des concepts de <sciences> figurant dans des livres de sciences changent ». Les valeurs plus élevées de l'indice dénotent des degrés d'assentiment plus élevés avec ces affirmations.

Indices de tendance mis à l'échelle

Plaisir de la science

L'indice du plaisir de la science (JOYSCIE) a été élaboré sur la base d'une question visant à analyser l'évolution des tendances (ST094) et extraite de l'enquête PISA 2006 (identifiant en 2006 : ST16), qui demandait aux élèves d'indiquer, sur une échelle de Linkert composée de quatre échelons (« Tout à fait d'accord », « D'accord », « Pas d'accord » et « Pas du tout d'accord ») leur degré d'assentiment avec les affirmations suivantes : « Je trouve généralement agréable d'apprendre des notions de <sciences> » ; « J'aime lire des textes qui traitent de <sciences> » ; « Cela me plaît d'avoir à résoudre des problèmes en <sciences> » ; « Je prends plaisir à acquérir de nouvelles connaissances en <sciences> » ; et « Cela m'intéresse d'apprendre des choses sur les <sciences> ». La variable calculée JOYSCIE a été assimilée à l'échelle correspondante dans la base de données PISA 2006, permettant ainsi une comparaison des tendances entre les éditions 2006 et 2015 de l'enquête PISA. Les valeurs plus élevées de l'indice dénotent des degrés d'assentiment plus élevés avec ces affirmations.

Efficacité perçue en sciences

L'indice de l'efficacité perçue en sciences (SCIEEFF) a été élaboré sur la base d'une question visant à analyser l'évolution des tendances (ST129) extraite de l'enquête PISA 2006 (identifiant en 2006 : ST17). Sur une échelle composée de quatre échelons (« J'y arriverais facilement », « J'y arriverais avec un peu d'effort », « Cela ne me serait pas facile d'y arriver seul(e) » et « Je n'y arriverais pas »), on a demandé aux élèves comment ils répondraient aux items de sciences suivants : « Identifier la question scientifique qui est à la base d'un article de journal portant sur un problème de santé » ; « Expliquer pourquoi les tremblements de terre sont plus fréquents dans certaines régions que dans d'autres » ; « Décrire le rôle des antibiotiques dans le traitement des maladies » ; « Déterminer quelle est la question scientifique liée au traitement des déchets » ; « Prévoir en quoi des changements apportés à l'environnement affecteront la survie de certaines espèces » ; « Interpréter des informations scientifiques fournies sur l'étiquette des produits alimentaires » ; « Discuter de la façon dont des données nouvelles pourraient modifier votre point de vue sur la probabilité qu'il existe de la vie sur Mars » ; et « Déterminer quelle est la meilleure de deux explications sur la formation de pluies acides ». Le codage des réponses a été inversé de sorte que les valeurs les plus élevées de l'indice correspondent aux plus hauts niveaux d'efficacité perçue en sciences. La variable calculée SCIEEFF a été assimilée à l'échelle correspondante dans la base de données PISA 2006, permettant ainsi une comparaison des tendances entre les éditions 2006 et 2015 de l'enquête PISA.

Activités scientifiques

L'indice des activités scientifiques (SCIEACT) a été élaboré sur la base d'une question visant à analyser l'évolution des tendances (ST146) et provenant de l'enquête PISA 2006 (identifiant en 2006 : ST19). Sur une échelle composée de quatre échelons (« Très souvent », « Régulièrement », « Parfois » et « Jamais ou presque jamais »), on a demandé aux élèves à quelle fréquence ils effectuaient les activités scientifiques suivantes : « Regarder des programmes télévisés sur des thèmes de <science au sens large> » ; « Acheter ou emprunter des livres sur des thèmes de <science au sens large> » ; « Surfer sur des sites Web traitant de thèmes de <science au sens large> » ; « Lire des revues de <science au sens large> ou des articles scientifiques dans les journaux » ; « Fréquenter un <club de sciences> » ; « Simuler des phénomènes naturels à l'aide de programmes informatiques ou dans des laboratoires virtuels » ; « Simuler des procédés techniques à l'aide de programmes informatiques ou dans des laboratoires virtuels » ; « Visiter des sites Web d'organisations écologiques » ; et « Suivre les actualités des organisations scientifiques, environnementales, écologiques sur des blogs et des micro-blogs ». Le codage des réponses a été inversé de sorte que les valeurs les plus élevées de l'indice correspondent aux plus hauts niveaux d'activités scientifiques des élèves. La variable calculée SCIEACT a été assimilée à l'échelle correspondante dans la base de données PISA 2006, permettant ainsi une comparaison des tendances entre les éditions 2006 et 2015 de l'enquête PISA.

Motivation instrumentale à l'idée d'apprendre en sciences

L'indice de la motivation instrumentale à l'idée d'apprendre en sciences (INSTSCIE) a été élaboré sur la base d'une question visant à analyser l'évolution des tendances (ST113) extraite de l'enquête PISA 2006 (identifiant en 2006 : ST35). Sur une échelle de Linkert composée de quatre échelons (« Tout à fait d'accord », « D'accord », « Pas d'accord » et « Pas du tout d'accord »), les élèves ont indiqué leur degré d'assentiment avec les affirmations suivantes : « Cela vaut la peine de faire des efforts dans le(s) <cours de sciences>, car cela m'aidera dans le métier que je veux faire plus tard » ; « Ce que j'apprends dans le(s) <cours de sciences> est important pour moi, car j'en ai besoin pour ce que je veux faire plus tard » ; « Cela vaut la peine d'étudier pour le(s) <cours de sciences>, car ce que j'apprends améliorera mes perspectives de carrière professionnelle » ; et « Beaucoup de ce que j'apprends dans le(s) <cours de sciences> m'aidera à trouver un emploi ». Le codage des réponses a été inversé de sorte que les valeurs les plus élevées de l'indice correspondent aux plus hauts niveaux de motivation instrumentale. La variable calculée INSTSCIE a été assimilée à l'échelle correspondante dans la base de données PISA 2006, permettant ainsi une comparaison des tendances entre les éditions 2006 et 2015 de l'enquête PISA.

Mise à l'échelle des indices liés à l'indice PISA de statut économique, social et culturel

Comme lors des précédentes évaluations, l'indice PISA de statut économique, social et culturel (SESC) a été dérivé de trois indices relatifs à la situation familiale : le niveau de formation le plus élevé des parents (PARED), le statut professionnel le plus élevé des parents (HISEI), et le patrimoine familial (HOMEPOS), y compris la bibliothèque familiale. Les indices PARED et HISEI sont des indices simples, décrits ci-dessus. L'indice HOMEPOS constitue une mesure approximative de la richesse familiale.

Patrimoine familial

Dans l'enquête PISA 2015, les élèves ont indiqué s'ils disposaient chez eux de 16 articles ménagers (ST011), y compris trois items nationaux permettant de mesurer la richesse familiale en fonction de chaque contexte national. En outre, les élèves ont fait part de la quantité d'équipements et de livres dont ils disposaient chez eux (ST012, ST013).

L'indice HOMEPOS est un indice synthétique de l'ensemble des articles ménagers et équipements à disposition (ST011, ST012 et ST013). À l'occasion de l'enquête PISA 2015, l'échelle du patrimoine familial a été calculée différemment que lors des précédentes enquêtes, en vue de mettre en adéquation le modèle basé sur la théorie de réponse à l'item et celui utilisé pour les échelles cognitives et non cognitives. Les catégories associées au nombre de livres dans la bibliothèque familiale restent inchangées dans l'enquête PISA 2015. Le codage des items de la question ST011 (1 = « oui » et 2 = « non ») est inversé, de sorte qu'un niveau élevé révèle la présence de l'indicateur.

Calcul de l'indice SESC

En vue de calculer l'Indice PISA de statut économique, social et culturel (SESC), des valeurs prévues ont été imputées en lieu et place des indices PARED, HISEI ou HOMEPOS manquants, avec l'inclusion d'une composante aléatoire, sur la base d'une régression des deux autres variables. En cas d'absence de données pour plus d'une des trois variables, le calcul de l'indice SESC n'est pas effectué et l'indice est déclaré manquant.

L'indice PISA de statut économique, social et culturel est dérivé d'une analyse en composantes principales de variables normalisées (la moyenne de l'OCDE de toutes ces variables est égale à 0, et leur écart-type à 1), mesurant l'indice PISA de statut économique, social et culturel d'après les valeurs des facteurs de la première composante principale. Les pays et économies de l'OCDE, ainsi que les pays et économies partenaires, ont contribué, de manière égale, à l'analyse en composantes principales, tandis que dans les précédentes évaluations, seuls les pays de l'OCDE apportaient leur contribution. Toutefois, aux fins de compte-rendu, les valeurs de l'indice SESC ont été modifiées, de sorte que le score d'un l'élève moyen de l'OCDE est égal à 0 et l'écart-type à 1 dans les pays de l'OCDE dont le coefficient de pondération est équivalent.

L'analyse en composantes principales a également été menée séparément dans chaque pays et économie participant, afin de déterminer si les composantes de l'indice se comportent de la même façon dans les différents pays et économies.

Calcul d'un indice de tendance pour l'indice SESC

Tandis qu'un indice PISA de statut économique, social et culturel (SESC) est intégré dans toutes les précédentes bases de données PISA, les composantes de l'indice SESC et le modèle de mise à l'échelle ont évolué au fil des évaluations, ce qui signifie que les scores de cet indice ne sont pas directement comparables entre les diverses évaluations. Afin de pouvoir analyser l'évolution des tendances, l'indice SESC a été calculé lors de l'évaluation PISA 2015, puis recalculé pour les évaluations précédentes selon une méthodologie similaire[2].

Avant de pouvoir réaliser une estimation des scores de l'évolution des tendances, les trois composantes doivent être légèrement modifiées :

- Comme lors de l'enquête PISA 2012, le système de codage des professions utilisé dans le processus de création de l'indice HISEI a évolué, passant de la CITP-88 à la CITP-08, et les codes des professions pour les évaluations précédentes ont été cartographiés non plus selon l'ancien système, mais sur la base du système actuel (voir également le chapitre 3 du rapport technique sur l'évaluation PISA 2012).
- Afin de permettre une comparaison de l'indice PARED à travers les différentes évaluations, le même système de mappage de la CITE appliqué à l'indice PARED a été utilisé dans toutes les évaluations.
- Pour permettre une meilleure comparaison de l'indice HOMEPOS à travers les diverses évaluations, la variable *Bibliothèque familiale* (ST013Q01TA) a été recodée en une variable à quatre catégories (25 livres ou moins, 26-100 livres, 101-500 livres, plus de 500 livres). La mise à l'échelle de l'indice HOMEPOS pour le calcul de l'évolution des tendances a été élaborée en trois étapes. Premièrement, les paramètres d'items internationaux pour tous les items (à l'exception des items nationaux, c'est-à-dire des items ST011Q17NA, ST011Q18NA et ST011Q19NA) administrés lors de l'enquête PISA 2015 ont été obtenus sur la base d'un étalonnage parallèle des données de l'évaluation de 2015. Exception faite du recodage de la variable ST013Q01TA, cette étape est identique à la mise à l'échelle classique de l'indice HOMEPOS dans l'évaluation PISA 2015 (voir ci-dessus). Deuxièmement, les items uniques extraits des évaluations précédentes (c'est-à-dire de 2000 à 2012) ont été mis à l'échelle, en appliquant les paramètres de 2015 à la majorité des items administrés en 2015, tout en permettant une libre estimation d'un nombre restreint de paramètres d'items mais en devant appliquer un coefficient de pondération équivalent entre les pays lors des diverses évaluations. Les items nationaux (c'est-à-dire les items ST011Q17NA, ST011Q18NA et ST011Q19NA) se voient associer des paramètres uniques spécifiques à chaque évaluation et à chaque pays. Dans la troisième et dernière étape, les valeurs d'indice (estimations pondérées des réponses les plus vraisemblables, WLE) ont été générées pour tous les élèves ayant participé aux évaluations précédentes (de 2000 à 2012). Étant donné que, d'évaluation en évaluation, 17 des 27 items pris en compte dans le calcul de l'évolution des tendances de l'indice HOMEPOS disposent des mêmes paramètres d'items, on peut estimer que les scores obtenus se trouvent sur une échelle conjointe ce qui permet de comparer les pays lors des différentes évaluations et d'utiliser ces scores dans le calcul de l'évolution de l'indice SESC.

L'analyse en composantes principales visant à obtenir les scores de l'évolution de l'indice SESC a été calculée comme susmentionné, sauf que le calcul a été effectué sur toutes les évaluations en utilisant ces trois composantes pouvant être comparées (évolution de l'indice HISEI, évolution de l'indice PARED et évolution de l'indice HOMEPOS).

Indices mis à l'échelle de niveau Établissement

Ressources des établissements

L'enquête PISA 2015 comportait une question composée de huit items relatifs aux ressources des établissements, et mesurant la façon dont les chefs d'établissement percevaient des facteurs pouvant potentiellement affecter l'enseignement dispensé dans leur établissement (« L'enseignement que votre établissement est à même de dispenser est-il affecté par les problèmes suivants ? »). Les quatre catégories de réponses étaient les suivantes : « Pas du tout », « Très peu », « Dans une certaine mesure » et « Beaucoup ». Une question similaire a été utilisée lors des précédentes évaluations mais, dans l'édition 2015, le nombre d'items a été réduit et

les items reformulés pour que la question soit axée sur deux variables dérivées. L'indice de pénurie de personnel (STAFFSHORT) a été calculé sur la base des quatre items suivants : un manque de personnel enseignant ; un personnel enseignant inadéquat ou peu qualifié ; un manque de personnel auxiliaire ; et un personnel auxiliaire inadéquat ou peu qualifié. L'indice de pénurie de matériel pédagogique (EDUSHORT) a été mis à l'échelle sur la base des quatre items suivants : un manque de matériel pédagogique (par exemple : manuels scolaires, équipement informatique, matériel de bibliothèque ou de laboratoire) ; du matériel pédagogique inadéquat ou de mauvaise qualité (par exemple : manuels scolaires, équipement informatique, matériel de bibliothèque ou de laboratoire ; un manque d'infrastructures (par exemple : bâtiments, terrain, chauffage/climatisation, systèmes d'éclairage et acoustiques) ; et des infrastructures inadéquates ou de mauvaise qualité (par exemple : bâtiments, terrain, chauffage/climatisation, systèmes d'éclairage et acoustiques). Si les valeurs de ces indices sont positives, la mesure dans laquelle les chefs d'établissement estiment que la quantité ou la qualité insuffisante des ressources affecte l'enseignement dispensé dans leur établissement est supérieure à la moyenne de l'OCDE ; si les valeurs des indices sont négatives, la mesure dans laquelle ils estiment que leur établissement souffre d'un manque ou d'une inadéquation des ressources est inférieure à la moyenne de l'OCDE.

Pourcentage d'observations manquantes pour les variables utilisées dans le présent volume

Sauf mention contraire, aucun ajustement n'est réalisé en cas de non-réponses aux questionnaires dans les analyses comprises dans le présent volume. Les pourcentages et les estimations basés sur les indices font référence au pourcentage de l'échantillon comportant des réponses valides aux items des questionnaires correspondants. Le tableau A1.3, disponible en ligne, indique le pourcentage de l'échantillon participant à des analyses qui repose sur des variables des questionnaires « Élève » ou « Établissement ». Il convient de comparer les résultats avec prudence lorsque ce pourcentage varie fortement dans le temps ou entre les pays et économies.

Tableau disponible en ligne

Tableau A.3. Pourcentage pondéré d'élèves participants inclus dans les analyses basées sur les questionnaires (http://dx.doi.org/10.1787/888933433112)

Notes

1. Au Royaume-Uni (exception faite de l'Écosse), les perspectives de carrière ont été codées uniquement sur la base du niveau à trois chiffres. En conséquence, les concepteurs modélistes de produits et de vêtements (CITP-08 : 2163), les concepteurs graphiques et les multimédia-graphistes (2166) intègrent la catégorie des spécialistes des sciences techniques ; les techniciens de prothèses médicales et dentaires (3214) entrent dans celle des professions intermédiaires à caractère scientifique ; et les techniciens de télécommunications (3522) sont exclus. Pour les élèves qui envisagent d'exercer une profession scientifique, ces carrières ne représentent qu'un faible pourcentage, de sorte que les résultats ne sont pas sévèrement affectés.

2. En raison de cette procédure, il existe deux indices pour 2015 (l'indice SESC et l'indice SESC de tendance). Le coefficient de corrélation de Pearson entre ces deux indices est r = .989 parmi l'ensemble des pays et économies participant à l'enquête PISA 2015. Sont inclus 22 pays/économies où le coefficient de corrélation est r >.990, 50 pays/économies où ce coefficient est r = [.960, .990], et un dernier pays (la Géorgie) où ce coefficient est r = .946. Dans les chapitres 6 et 7, par souci de cohérence entre les tableaux, les données de 2015 relatives aux tendances de l'indice SESC se basent sur l'indice SESC de 2015, et non sur l'indice SESC de tendance de 2015.

Références

Ganzeboom, H.B.G. (2010), « A new international socio-economic index [ISEI] of occupational status for the International Standard Classification of Occupation 2008 [ISCO-08] constructed with data from the ISSP 2002-2007; with an analysis of quality of occupational measurement in ISSP », document présenté lors de l'Annual Conference of International Social Survey Programme, Lisbonne, 1er mai 2010.

Ganzeboom, H. B.G. et **D.J. Treiman** (2003), « Three Internationally Standardised Measures for Comparative Research on Occupational Status », pp. 159-193 in J.H.P. Hoffmeyer-Zlotnik et C. Wolf (éd.), *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, Kluwer Academic Press, New York.

Kjærnsli, M. et **S. Lie** (2011), « Students' Preference for Science Careers: International Comparisons Based on PISA 2006 », *International Journal of Science Education*, vol. 33/1, pp. 121-44, http://dx.doi.org/10.1080/09500693.2010.518642.

OCDE (à paraître), *PISA 2015 Technical Report*, PISA, Éditions OCDE, Paris.

OCDE (2016), *Cadre d'évaluation et d'analyse de l'enquête PISA 2015 : Compétences en sciences, en compréhension de l'écrit, en mathématiques et en matières financières*, PISA, Éditions OCDE, Paris, http://dx.doi.org/10.1787/9789264259478-fr.

OCDE (2007), *PISA 2006 : Les compétences en sciences, un atout pour réussir : Volume 1 Analyse des résultats*, Éditions OCDE, Paris, http://dx.doi.org/10.1787/9789264040137-fr.

OCDE (1999), *Nomenclature des systèmes d'éducation : Guide d'utilisation de la CITE-97 dans les pays de l'OCDE – Édition 1999*, Éditions OCDE, Paris.

Sikora, J. et **A. Pokropek** (2012), « Gender Segregation of Adolescent Science Career Plans in 50 Countries », *Science Education*, vol. 96/2, pp. 234-64, http://dx.doi.org/10.1002/sce.20479.

Warm, T.A. (1985), « Weighted Maximum Likelihood Estimation of Ability in Item Response Theory with Tests of Finite Length », *Technical Report CGI-TR-85-08*, U.S. Coast Guard Institute, Oklahoma City.

O*NET OnLine (n.d), « All STEM disciplines », page web, www.onetonline.org/find/stem?t=0 (consultée le 4 octobre 2016).

ANNEXE A2
POPULATION CIBLE, ÉCHANTILLONS ET DÉFINITION DES ÉTABLISSEMENTS DE L'ENQUÊTE PISA

Définition de la population cible de l'enquête PISA

L'enquête PISA 2015 évalue le rendement cumulé de l'éducation et de l'apprentissage à un moment où la plupart des jeunes adultes suivent toujours une formation initiale.

Pour une enquête internationale, garantir la comparabilité des populations nationales cibles représente un défi majeur.

Les différences en ce qui concerne la nature et l'importance de l'enseignement et des structures d'accueil préprimaires, l'âge minimum de la scolarité obligatoire et la structure institutionnelle des systèmes d'éducation des différents pays rendent impossible toute définition de niveaux de scolarité comparables au niveau international. C'est pourquoi les comparaisons internationales des performances éducatives définissent généralement les populations en se basant sur un groupe d'âge cible. Certaines enquêtes internationales antérieures ont défini leur population cible sur la base de l'année d'études la plus représentative d'une cohorte d'âge particulière. Cette méthode présente comme inconvénient que de légères variations de la répartition par âge des élèves dans les différents niveaux aboutissent souvent à la sélection d'années d'études différentes selon les pays ou les divers systèmes d'éducation au sein des pays, ce qui soulève des interrogations majeures concernant la comparabilité des résultats entre les pays, et dans certains cas, au sein des pays. En outre, les élèves de l'âge souhaité n'étant pas tous représentés dans les échantillons basés sur l'année d'études, les résultats risquent davantage d'être biaisés si les élèves non représentés dans l'échantillon sont inscrits dans l'année d'études supérieure dans certains pays, et inférieure dans d'autres. Cette situation est susceptible d'entraîner l'exclusion d'élèves qui disposent d'un niveau de performance potentiellement supérieur dans les premiers pays et d'élèves qui disposent d'un niveau de performance potentiellement inférieur dans les seconds pays.

L'enquête PISA a contourné ce problème en définissant sa population cible à un âge précis, c'est-à-dire indépendamment des structures institutionnelles des systèmes nationaux d'éducation. PISA évalue les élèves âgés de 15 ans et 3 mois (révolus) à 16 ans et 2 mois (révolus) au début de la période d'évaluation, avec une variation d'un mois, inscrits dans un établissement d'enseignement en 7e année ou dans une année d'études supérieure, quels que soient leur année d'études ou le type de leur établissement, ou leur mode de scolarisation à temps plein ou à temps partiel. Le présent rapport désigne généralement les établissements d'enseignement sous le terme générique d'établissements, bien que certains d'entre eux, en particulier ceux qui proposent des formations professionnelles, aient une autre dénomination dans certains pays. Conformément à cette définition, les élèves avaient en moyenne 15 ans et 9 mois au moment de l'évaluation dans les pays de l'OCDE. Cette moyenne varie de 2 mois et 18 jours (0.20 an), la moyenne minimale étant de 15 ans et 8 mois, et la moyenne maximale, de 15 ans et 10 mois.

La population cible étant définie en fonction de l'âge, l'enquête PISA permet de recueillir des résultats sur les connaissances et les compétences d'un groupe d'individus nés dans une période de référence comparable, mais susceptibles d'avoir vécu des expériences d'apprentissage différentes tant dans le cadre scolaire qu'extrascolaire. L'enquête PISA définit ces connaissances et compétences comme le rendement de l'éducation à un âge commun à tous les pays. En fonction de la politique nationale en matière d'âge obligatoire, de sélection et de promotion scolaire, l'éventail d'années d'études de ces élèves varie dans les divers systèmes, filières ou voies d'éducation. Il est crucial de tenir compte de ces différences lors de la comparaison des résultats de l'évaluation PISA entre les pays, car des différences constatées entre des élèves de 15 ans peuvent disparaître lors de l'éventuelle convergence ultérieure des expériences pédagogiques des élèves.

Lorsqu'un pays obtient un score significativement supérieur à un autre pays en culture scientifique, en compréhension de l'écrit ou en culture mathématique, il ne faut pas en conclure automatiquement que son système d'éducation ou ses établissements sont plus efficaces. En revanche, on peut tout à fait en déduire que l'impact cumulé des expériences d'apprentissage vécues de la prime enfance à l'âge de 15 ans, tant à la maison qu'à l'école ou ailleurs, a produit de meilleurs résultats dans les domaines de littératie évalués par l'enquête PISA.

Les ressortissants nationaux scolarisés à l'étranger sont exclus de la population cible de l'enquête PISA, contrairement aux ressortissants étrangers scolarisés dans les pays participants.

Lors de l'évaluation PISA 2015, les pays souhaitant disposer de résultats par année d'études à des fins d'analyse nationale se sont vu proposer une option permettant d'ajouter un échantillon basé sur l'année d'études à l'échantillon basé sur l'âge.

Représentativité des échantillons

Tous les pays et économies se sont efforcés d'obtenir une représentativité optimale des effectifs d'élèves de 15 ans dans leurs échantillons nationaux et y ont inclus les élèves fréquentant des établissements d'enseignement spécialisé. Par conséquent, l'enquête PISA 2015 a enregistré des taux de représentativité sans précédent dans une enquête internationale de cette nature.

Les normes d'échantillonnage utilisées dans l'enquête PISA permettent aux pays d'exclure jusqu'à 5 % d'élèves et d'établissements de leur population cible. Tous les pays sauf douze, en l'occurrence le Royaume-Uni (8.22 %), le Luxembourg (8.16 %), le Canada (7.49 %), la Norvège (6.75 %), la Nouvelle-Zélande (6.54 %), la Suède (5.71 %), l'Estonie (5.52 %), l'Australie (5.31 %), le Monténégro (5.17 %), la Lituanie (5.12 %), la Lettonie (5.07 %) et le Danemark (5.04 %), ont respecté ces normes. Le taux global d'exclusion est même inférieur à 2 % dans 29 pays et économies. Une fois les exclusions d'élèves pour raisons linguistiques comptabilisées (c'est-à-dire soustraites du taux total d'exclusion), le taux d'exclusion du Danemark, de la Lettonie, de la Nouvelle-Zélande et de la Suède passe sous la barre des 5 %. Pour plus de détails, consultez la page www.pisa.oecd.org.

Les exclusions contenues dans les limites précisées ci-dessus comprennent :

- Au niveau des établissements : les établissements géographiquement inaccessibles ou dans lesquelles l'enquête PISA a été jugée impossible à réaliser ; et les établissements accueillant exclusivement des élèves relevant des catégories définies sous la rubrique des exclusions « intra-établissement », tels que les écoles pour non-voyants. Le pourcentage d'élèves de 15 ans inscrits dans ces établissements doit être inférieur à 2.5 % de la population nationale cible théorique (0.5 % maximum dans la première catégorie d'établissements et 2 % maximum dans la deuxième). La magnitude, la nature et la justification des exclusions réalisées au niveau des établissements sont documentées dans le rapport technique sur l'enquête PISA 2015 (*PISA 2015 Technical Report* [OCDE, à paraître en anglais uniquement]).
- Au niveau des élèves : les élèves atteints d'un handicap mental ; les élèves atteints d'un handicap fonctionnel ; les élèves ne maîtrisant pas la langue de l'évaluation ; autre – catégorie définie par les centres nationaux et approuvée par le centre international ; et les élèves suivant un enseignement donné principalement dans une langue pour laquelle aucune ressource n'est disponible. Les élèves ne peuvent être exclus au seul motif d'un faible niveau de compétences ou de problèmes de discipline courants. Le pourcentage d'individus âgés de 15 ans exclus au sein des établissements doit être inférieur à 2.5 % de la population nationale cible théorique.

Le tableau A2.1 présente la population cible des pays qui ont participé à l'enquête PISA 2015. Des informations plus détaillées sur la population cible et la mise en œuvre des normes d'échantillonnage se trouvent dans le rapport technique sur l'enquête PISA 2015 (*PISA 2015 Technical Report* [OCDE, à paraître en anglais uniquement]).

- La **colonne 1** indique la population totale d'individus âgés de 15 ans, calculée sur la base des informations les plus récentes, soit celles de 2014 (l'année précédant celle de l'évaluation) dans la plupart des pays.
- La **colonne 2** indique le nombre total d'individus âgés de 15 ans inscrits dans un établissement d'enseignement en 7e année ou dans une année d'études supérieure (selon la définition ci-dessus), qui représente la « population admissible ».
- La **colonne 3** indique la population nationale cible théorique. Les pays ont été autorisés à exclure jusqu'à 0.5 % des élèves de leur population admissible a priori, essentiellement pour des raisons pratiques. Les exclusions a priori suivantes dépassent ces limites, mais ont été approuvées par le Consortium PISA : la Belgique a exclu 0.21 % de ses élèves cumulant études et activité professionnelle ; le Canada a exclu 1.22 % d'élèves vivant dans des territoires et réserves d'autochtones ; le Chili a exclu 0.04 % d'élèves, parce qu'ils vivent sur l'île de Pâques, dans l'archipel Juan Fernandez ou en Antarctique ; et les Émirats arabes unis ont exclu 0.04 % d'élèves dont les informations n'étaient pas disponibles. Aux États-Unis, l'État de Massachusetts, dont les données ont été adjugées, a exclu 13.11 % d'élèves et la Caroline du Nord a exclu 5.64 % d'élèves, parce qu'ils étaient scolarisés dans des établissements privés. Dans ces deux régions, la population cible théorique incluait uniquement les jeunes de 15 ans scolarisés en 7e année au moins dans des établissements publics.
- La **colonne 4** indique le nombre d'élèves scolarisés qui ont été exclus de la population nationale cible théorique soit au moment de la constitution de l'échantillon, soit ultérieurement, lors de la collecte des données sur le terrain.
- La **colonne 5** indique la population nationale cible théorique, déduction faite des élèves scolarisés dans des établissements exclus. Ce chiffre s'obtient en soustrayant le nombre de la colonne 4 du nombre de la colonne 3.
- La **colonne 6** indique le pourcentage d'élèves scolarisés dans des établissements exclus, un chiffre obtenu en divisant le nombre de la colonne 4 par le nombre de la colonne 3, puis en multipliant le total obtenu par 100.
- La **colonne 7** indique le nombre d'élèves qui ont participé à l'enquête PISA 2015. Dans certains cas, ce nombre exclut les jeunes de 15 ans soumis à une évaluation dans le cadre d'options nationales supplémentaires.
- La **colonne 8** indique le nombre pondéré d'élèves participants, c'est-à-dire le nombre d'élèves de la population nationale cible représentés par l'échantillon PISA.
- Tous les pays se sont efforcés d'obtenir une représentativité optimale de la population cible dans les établissements échantillonnés. Dans les établissements échantillonnés, tous les élèves admissibles, c'est-à-dire les individus de 15 ans, quelle que soit leur année d'études, ont d'abord été répertoriés. Les élèves échantillonnés à exclure ont été inclus dans la documentation d'échantillonnage et une liste indiquant les motifs de leur exclusion a été établie. La **colonne 9** indique le nombre total d'élèves exclus, dont les différentes catégories sont définies et classées dans le tableau A2.2.

[Partie 1/1]

Tableau A2.1 Populations cibles et échantillons PISA

		Informations sur la population et l'échantillon												Indices de couverture		
		Population totale de jeunes âgés de 15 ans	Population totale de jeunes de 15 ans scolarisés en 7e année d'études ou à un niveau supérieur	Total dans la population nationale cible théorique	Total des exclusions au niveau des établissements	Total de la population nationale cible théorique après toutes les exclusions des établissements et avant les exclusions d'élèves au sein des établissements	Taux d'exclusion au niveau des établissements (%)	Nombre d'élèves participants	Nombre pondéré d'élèves participants	Nombre d'élèves exclus	Nombre pondéré d'élèves exclus	Taux d'exclusion au sein des établissements (%)	Taux global d'exclusion (%)	Indice de couverture 1 : couverture de la population nationale théorique	Indice de couverture 2 : couverture de la population nationale d'individus scolarisés	Indice de couverture 3 : couverture de la population de jeunes âgés de 15 ans
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
OCDE	Australie	282 888	282 547	282 547	6 940	275 607	2.46	14 530	256 329	681	7 736	2.93	5.31	0.947	0.947	0.906
	Autriche	88 013	82 683	82 683	790	81 893	0.96	7 007	73 379	84	866	1.17	2.11	0.979	0.979	0.834
	Belgique	123 630	121 954	121 694	1 597	120 097	1.31	9 651	114 902	39	410	0.36	1.66	0.983	0.981	0.929
	Canada	396 966	381 660	376 994	1 590	375 404	0.42	20 058	331 546	1 830	25 340	7.10	7.49	0.925	0.914	0.835
	Chili	255 440	245 947	245 852	2 641	243 211	1.07	7 053	203 782	37	1 393	0.68	1.75	0.983	0.982	0.798
	République tchèque	90 391	90 076	90 076	1 814	88 262	2.01	6 894	84 519	25	368	0.43	2.44	0.976	0.976	0.935
	Danemark	68 174	67 466	67 466	605	66 861	0.90	7 161	60 655	514	2 644	4.18	5.04	0.950	0.950	0.890
	Estonie	11 676	11 491	11 491	416	11 075	3.62	5 587	10 834	116	218	1.97	5.52	0.945	0.945	0.928
	Finlande	58 526	58 955	58 955	472	58 483	0.80	5 882	56 934	124	1 157	1.99	2.78	0.972	0.972	0.973
	France	807 867	778 679	778 679	28 742	749 937	3.69	6 108	734 944	35	3 620	0.49	4.16	0.958	0.958	0.910
	Allemagne	774 149	774 149	774 149	11 150	762 999	1.44	6 522	743 969	54	5 342	0.71	2.14	0.979	0.979	0.961
	Grèce	105 530	105 253	105 253	953	104 300	0.91	5 532	96 157	58	965	0.99	1.89	0.981	0.981	0.911
	Hongrie	94 515	90 065	90 065	1 945	88 120	2.16	5 658	84 644	55	1 009	1.18	3.31	0.967	0.967	0.896
	Islande	4 250	4 195	4 195	17	4 178	0.41	3 374	3 966	131	132	3.23	3.62	0.964	0.964	0.933
	Irlande	61 234	59 811	59 811	72	59 739	0.12	5 741	59 082	197	1 825	3.00	3.11	0.969	0.969	0.965
	Israël	124 852	118 997	118 997	2 310	116 687	1.94	6 598	117 031	115	1 803	1.52	3.43	0.966	0.966	0.937
	Italie	616 761	567 268	567 268	11 190	556 078	1.97	11 583	495 093	246	9 395	1.86	3.80	0.962	0.962	0.803
	Japon	1 201 615	1 175 907	1 175 907	27 323	1 148 584	2.32	6 647	1 138 349	2	318	0.03	2.35	0.976	0.976	0.947
	Corée	620 687	619 950	619 950	3 555	616 395	0.57	5 581	569 106	20	1 806	0.32	0.89	0.991	0.991	0.917
	Lettonie	17 255	16 955	16 955	677	16 278	3.99	4 869	15 320	70	174	1.12	5.07	0.949	0.949	0.888
	Luxembourg	6 327	6 053	6 053	162	5 891	2.68	5 299	5 540	331	331	5.64	8.16	0.918	0.918	0.876
	Mexique	2 257 399	1 401 247	1 401 247	5 905	1 395 342	0.42	7 568	1 392 995	30	6 810	0.49	0.91	0.991	0.991	0.617
	Pays-Bas	201 670	200 976	200 976	6 866	194 110	3.42	5 385	191 817	14	502	0.26	3.67	0.963	0.963	0.951
	Nouvelle-Zélande	60 162	57 448	57 448	681	56 767	1.19	4 520	54 274	333	3 112	5.42	6.54	0.935	0.935	0.902
	Norvège	63 642	63 491	63 491	854	62 637	1.35	5 456	58 083	345	3 366	5.48	6.75	0.933	0.933	0.913
	Pologne	380 366	361 600	361 600	6 122	355 478	1.69	4 478	345 709	34	2 418	0.69	2.38	0.976	0.976	0.909
	Portugal	110 939	101 107	101 107	424	100 683	0.42	7 325	97 214	105	860	0.88	1.29	0.987	0.987	0.876
	République slovaque	55 674	55 203	55 203	1 376	53 827	2.49	6 350	49 654	114	912	1.80	4.25	0.957	0.957	0.892
	Slovénie	18 078	17 689	17 689	290	17 399	1.64	6 406	16 773	114	247	1.45	3.07	0.969	0.969	0.928
	Espagne	440 084	414 276	414 276	2 175	412 101	0.53	6 736	399 935	200	10 893	2.65	3.16	0.968	0.968	0.909
	Suède	97 749	97 210	97 210	1 214	95 996	1.25	5 458	91 491	275	4 324	4.51	5.71	0.943	0.943	0.936
	Suisse	85 495	83 655	83 655	2 320	81 335	2.77	5 860	82 223	107	1 357	1.62	4.35	0.956	0.956	0.962
	Turquie	1 324 089	1 100 074	1 100 074	5 746	1 094 328	0.52	5 895	925 366	31	5 359	0.58	1.10	0.989	0.989	0.699
	Royaume-Uni	747 593	746 328	746 328	23 412	722 916	3.14	14 157	627 703	870	34 747	5.25	8.22	0.918	0.918	0.840
	États-Unis	4 220 325	3 992 053	3 992 053	12 001	3 980 052	0.30	5 712	3 524 497	193	109 580	3.02	3.31	0.967	0.967	0.835
Partenaires	Albanie	48 610	45 163	45 163	10	45 153	0.02	5 215	40 896	0	0	0.00	0.02	1.000	1.000	0.841
	Algérie	389 315	354 936	354 936	0	354 936	0.00	5 519	306 647	0	0	0.00	0.00	1.000	1.000	0.788
	Argentine	718 635	578 308	578 308	2 617	575 691	0.45	6 349	394 917	21	1 367	0.34	0.80	0.992	0.992	0.550
	Brésil	3 803 681	2 853 388	2 853 388	64 392	2 788 996	2.26	23 141	2 425 961	119	13 543	0.56	2.80	0.972	0.972	0.638
	P-S-J-G (Chine)	2 084 958	1 507 518	1 507 518	58 639	1 448 879	3.89	9 841	1 331 794	33	3 609	0.27	4.15	0.959	0.959	0.639
	Bulgarie	66 601	59 397	59 397	1 124	58 273	1.89	5 928	53 685	49	433	0.80	2.68	0.973	0.973	0.806
	Colombie	760 919	674 079	674 079	37	674 042	0.01	11 795	567 848	9	507	0.09	0.09	0.999	0.999	0.746
	Costa Rica	81 773	66 524	66 524	0	66 524	0.00	6 866	51 897	13	98	0.19	0.19	0.998	0.998	0.635
	Croatie	45 031	35 920	35 920	805	35 115	2.24	5 809	40 899	86	589	1.42	3.63	0.964	0.964	0.908
	Chypre*	9 255	9 255	9 253	109	9 144	1.18	5 571	8 785	228	292	3.22	4.36	0.956	0.956	0.949
	République dominicaine	193 153	139 555	139 555	2 382	137 173	1.71	4 740	132 300	4	106	0.08	1.79	0.982	0.982	0.685
	ERYM	16 719	16 717	16 717	259	16 458	1.55	5 324	15 847	8	19	0.12	1.67	0.983	0.983	0.948
	Géorgie	48 695	43 197	43 197	1 675	41 522	3.88	5 316	38 334	35	230	0.60	4.45	0.955	0.955	0.787
	Hong-Kong (Chine)	65 100	61 630	61 630	708	60 922	1.15	5 359	57 662	36	374	0.65	1.79	0.982	0.982	0.886
	Indonésie	4 534 216	3 182 816	3 182 816	4 046	3 178 770	0.13	6 513	3 092 773	0	0	0.00	0.13	0.999	0.999	0.682
	Jordanie	126 399	121 729	121 729	71	121 658	0.06	7 267	108 669	70	1 006	0.92	0.97	0.990	0.990	0.860
	Kazakhstan	211 407	209 555	209 555	7 475	202 080	3.57	7 841	192 909	0	0	0.00	3.57	0.964	0.964	0.912
	Kosovo	31 546	28 229	28 229	1 156	27 073	4.10	4 826	22 333	50	174	0.77	4.84	0.952	0.952	0.708
	Liban	64 044	62 281	62 281	1 300	60 981	2.09	4 546	42 331	0	0	0.00	2.09	0.979	0.979	0.661
	Lituanie	33 163	32 097	32 097	573	31 524	1.79	6 525	29 915	227	1 050	3.39	5.12	0.949	0.949	0.902
	Macao (Chine)	5 100	4 417	4 417	3	4 414	0.07	4 476	4 507	0	0	0.00	0.07	0.999	0.999	0.884
	Malaisie	540 000	448 838	448 838	2 418	446 420	0.54	8 861	412 524	41	2 344	0.56	1.10	0.989	0.989	0.764
	Malte	4 397	4 406	4 406	63	4 343	1.43	3 634	4 296	41	41	0.95	2.36	0.976	0.976	0.977
	Moldavie	31 576	30 601	30 601	182	30 419	0.59	5 325	29 341	21	118	0.40	0.99	0.990	0.990	0.929
	Monténégro	7 524	7 506	7 506	40	7 466	0.53	5 665	6 777	300	332	4.66	5.17	0.948	0.948	0.901
	Pérou	580 371	478 229	478 229	6 355	471 874	1.33	6 971	431 738	13	745	0.17	1.50	0.985	0.985	0.744
	Qatar	13 871	13 850	13 850	380	13 470	2.74	12 083	12 951	193	193	1.47	4.17	0.958	0.958	0.934
	Roumanie	176 334	176 334	176 334	1 823	174 511	1.03	4 876	164 216	3	120	0.07	1.11	0.989	0.989	0.931
	Russie	1 176 473	1 172 943	1 172 943	24 217	1 148 726	2.06	6 036	1 120 932	13	2 469	0.22	2.28	0.977	0.977	0.953
	Singapour	48 218	47 050	47 050	445	46 605	0.95	6 115	46 224	25	179	0.39	1.33	0.987	0.987	0.959
	Taipei chinois	295 056	287 783	287 783	1 179	286 604	0.41	7 708	251 424	22	647	0.26	0.67	0.993	0.993	0.852
	Thaïlande	895 513	756 917	756 917	9 646	747 271	1.27	8 249	634 795	22	2 107	0.33	1.60	0.984	0.984	0.709
	Trinité-et-Tobago	17 371	17 371	17 371	0	17 371	0.00	4 692	13 197	0	0	0.00	0.00	1.000	1.000	0.760
	Tunisie	122 186	122 186	122 186	679	121 507	0.56	5 375	113 599	3	61	0.05	0.61	0.994	0.994	0.930
	Émirats arabes unis	51 687	51 518	51 499	994	50 505	1.93	14 167	46 950	63	152	0.32	2.25	0.978	0.977	0.908
	Uruguay	53 533	43 865	43 865	4	43 861	0.01	6 062	38 287	6	32	0.08	0.09	0.999	0.999	0.715
	Viet Nam	1 803 552	1 032 599	1 032 599	6 557	1 026 042	0.63	5 826	874 859	0	0	0.00	0.63	0.994	0.994	0.485

Remarques : Pour une explication détaillée des valeurs présentées dans ce tableau, consulter le rapport technique de l'enquête PISA 2015 (*PISA 2015 Technical Report* [OCDE, à paraître en anglais uniquement]).
La valeur correspondant à la population nationale totale de jeunes de 15 ans scolarisés présentée dans la colonne 2 peut parfois être supérieure à la valeur correspondant au nombre total de jeunes de 15 ans présentée dans la colonne 1 en raison de l'utilisation de sources de données différentes.
Pour le Mexique, en 2015, la population totale de jeunes de 15 ans scolarisés en 7e année d'études ou à un niveau supérieur est une estimation de la taille de la population cible du cadre d'échantillonnage à partir duquel les élèves de 15 ans ont été sélectionnés pour l'évaluation PISA. Lorsque le Mexique a fourni ces informations à PISA, la statistique officielle de cette population s'établissait à 1 573 952.
* Voir les notes au début de cette annexe.
StatLink http://dx.doi.org/10.1787/888933433129

[Partie 1/2]

Tableau A2.2 Exclusions

		Nombre d'élèves exclus (non pondéré)					
		Nombre d'élèves exclus pour cause de handicap fonctionnel (Code 1)	Nombre d'élèves exclus pour cause de handicap mental (Code 2)	Nombres d'élèves exclus pour des raisons linguistiques (Code 3)	Nombre d'élèves exclus pour d'autres motifs (Code 4)	Nombre d'élèves exclus en raison de l'absence de matériel disponible dans la langue d'instruction (Code 5)	Nombre total d'élèves exclus
		(1)	(2)	(3)	(4)	(5)	(6)
OCDE	Australie	85	528	68	0	0	681
	Autriche	8	15	61	0	0	84
	Belgique	4	18	17	0	0	39
	Canada	156	1 308	366	0	0	1 830
	Chili	6	30	1	0	0	37
	République tchèque	2	9	14	0	0	25
	Danemark	18	269	156	70	1	514
	Estonie	17	93	6	0	0	116
	Finlande	2	90	17	8	7	124
	France	5	21	9	0	0	35
	Allemagne	4	25	25	0	0	54
	Grèce	3	44	11	0	0	58
	Hongrie	3	13	9	30	0	55
	Islande	9	66	47	9	0	131
	Irlande	25	57	55	60	0	197
	Israël	22	68	25	0	0	115
	Italie	78	147	21	0	0	246
	Japon	0	2	0	0	0	2
	Corée	3	17	0	0	0	20
	Lettonie	7	47	16	0	0	70
	Luxembourg	4	254	73	0	0	331
	Mexique	4	23	3	0	0	30
	Pays-Bas	1	13	0	0	0	14
	Nouvelle-Zélande	23	140	167	0	3	333
	Norvège	11	253	81	0	0	345
	Pologne	11	20	0	3	0	34
	Portugal	4	99	2	0	0	105
	République slovaque	7	71	2	34	0	114
	Slovénie	33	36	45	0	0	114
	Espagne	9	144	47	0	0	200
	Suède	154	0	121	0	0	275
	Suisse	8	42	57	0	0	107
	Turquie	1	23	7	0	0	31
	Royaume-Uni	77	690	102	0	1	870
	États-Unis	16	120	44	13	0	193
Partenaires	Albanie	0	0	0	0	0	0
	Algérie	0	0	0	0	0	0
	Argentine	10	10	1	0	0	21
	Brésil	20	99	0	0	0	119
	P-S-J-G (Chine)	6	25	2	0	0	33
	Bulgarie	39	6	4	0	0	49
	Colombie	3	4	2	0	0	9
	Costa Rica	3	1	0	9	0	13
	Croatie	2	75	9	0	0	86
	Chypre*	12	164	52	0	0	228
	République dominicaine	1	3	0	0	0	4
	ERYM	7	1	0	0	0	8
	Géorgie	3	25	7	0	0	35
	Hong-Kong (Chine)	0	35	1	0	0	36
	Indonésie	0	0	0	0	0	0
	Jordanie	43	17	10	0	0	70
	Kazakhstan	0	0	0	0	0	0
	Kosovo	9	13	27	0	0	50
	Liban	0	0	0	0	0	0
	Lituanie	12	213	2	0	0	227
	Macao (Chine)	0	0	0	0	0	0
	Malaisie	10	22	9	0	0	41
	Malte	8	27	6	0	0	41
	Moldavie	12	8	1	0	0	21
	Monténégro	14	23	5	0	258	300
	Pérou	4	9	0	0	0	13
	Qatar	76	110	7	0	0	193
	Roumanie	1	1	1	0	0	3
	Russie	3	10	0	0	0	13
	Singapour	3	15	7	0	0	25
	Taipei chinois	3	19	0	0	0	22
	Thaïlande	1	19	2	0	0	22
	Trinité-et-Tobago	0	0	0	0	0	0
	Tunisie	0	0	3	0	0	3
	Émirats arabes unis	16	24	23	0	0	63
	Uruguay	2	4	0	0	0	6
	Viet Nam	0	0	0	0	0	0

Codes d'exclusion :
Code 1 : Handicap fonctionnel : l'élève souffre d'un handicap physique permanent de modéré à grave.
Code 2 : Handicap mental : l'élève souffre d'un handicap mental ou de troubles émotionnels, et d'un retard cognitif identifié lors de tests ou diagnostiqué par des professionnels.
Code 3 : Maîtrise insuffisante de la langue de l'évaluation : l'élève a une langue maternelle différente des langues de l'évaluation dans le pays, où il réside depuis moins d'un an.
Code 4 : Autres motifs : définis par les Centres nationaux et approuvés par le Centre international.
Code 5 : Absence de matériel disponible dans la langue de l'évaluation.
Remarque : Pour une explication détaillée des valeurs présentées dans ce tableau, consulter le rapport technique de l'enquête PISA 2015 (*PISA 2015 Technical Report* [OCDE, à paraître en anglais uniquement]).

* Voir les notes au début de cette annexe.

StatLink http://dx.doi.org/10.1787/888933433129

[Partie 2/2]

Tableau A2.2 Exclusions

		Nombre d'élèves exclus (pondéré)					
		Nombre d'élèves exclus pour cause de handicap fonctionnel (Code 1)	Nombre d'élèves exclus pour cause de handicap mental (Code 2)	Nombres d'élèves exclus pour des raisons linguistiques (Code 3)	Nombre d'élèves exclus pour d'autres motifs (Code 4)	Nombre d'élèves exclus en raison de l'absence de matériel disponible dans la langue d'instruction (Code 5)	Nombre total pondéré d'élèves exclus
		(7)	(8)	(9)	(10)	(11)	(12)
OCDE	Australie	932	6 011	793	0	0	7 736
OCDE	Autriche	74	117	675	0	0	866
OCDE	Belgique	33	192	185	0	0	410
OCDE	Canada	1 901	18 018	5 421	0	0	25 340
OCDE	Chili	194	1 190	9	0	0	1 393
OCDE	République tchèque	40	140	188	0	0	368
OCDE	Danemark	122	1 539	551	421	11	2 644
OCDE	Estonie	29	176	13	0	0	218
OCDE	Finlande	18	858	156	67	58	1 157
OCDE	France	562	2 144	914	0	0	3 620
OCDE	Allemagne	423	2 562	2 357	0	0	5 342
OCDE	Grèce	43	729	193	0	0	965
OCDE	Hongrie	57	284	114	554	0	1 009
OCDE	Islande	9	67	47	9	0	132
OCDE	Irlande	213	526	516	570	0	1 825
OCDE	Israël	349	1 070	384	0	0	1 803
OCDE	Italie	3 316	5 199	880	0	0	9 395
OCDE	Japon	0	318	0	0	0	318
OCDE	Corée	291	1 515	0	0	0	1 806
OCDE	Lettonie	21	115	38	0	0	174
OCDE	Luxembourg	4	254	73	0	0	331
OCDE	Mexique	842	4 802	1 165	0	0	6 810
OCDE	Pays-Bas	33	469	0	0	0	502
OCDE	Nouvelle-Zélande	233	1 287	1 568	0	24	3 112
OCDE	Norvège	105	2 471	790	0	0	3 366
OCDE	Pologne	876	1 339	0	203	0	2 418
OCDE	Portugal	29	818	13	0	0	860
OCDE	République slovaque	44	567	12	288	0	912
OCDE	Slovénie	84	71	92	0	0	247
OCDE	Espagne	511	7 662	2 720	0	0	10 893
OCDE	Suède	2 380	0	1 944	0	0	4 324
OCDE	Suisse	91	540	726	0	0	1 357
OCDE	Turquie	43	4 094	1 222	0	0	5 359
OCDE	Royaume-Uni	2 724	27 808	4 001	0	214	34 747
OCDE	États-Unis	7 873	67 816	26 525	7 366	0	109 580
Partenaires	Albanie	0	0	0	0	0	0
Partenaires	Algérie	0	0	0	0	0	0
Partenaires	Argentine	579	770	18	0	0	1 367
Partenaires	Brésil	1 743	11 800	0	0	0	13 543
Partenaires	P-S-J-G (Chine)	438	2 970	201	0	0	3 609
Partenaires	Bulgarie	347	51	35	0	0	433
Partenaires	Colombie	181	309	17	0	0	507
Partenaires	Costa Rica	22	5	0	71	0	98
Partenaires	Croatie	13	501	75	0	0	589
Partenaires	Chypre*	16	212	65	0	0	292
Partenaires	République dominicaine	24	82	0	0	0	106
Partenaires	ERYM	15	4	0	0	0	19
Partenaires	Géorgie	19	170	41	0	0	230
Partenaires	Hong-Kong (Chine)	0	363	11	0	0	374
Partenaires	Indonésie	0	0	0	0	0	0
Partenaires	Jordanie	656	227	122	0	0	1 006
Partenaires	Kazakhstan	0	0	0	0	0	0
Partenaires	Kosovo	28	37	104	0	0	174
Partenaires	Liban	0	0	0	0	0	0
Partenaires	Lituanie	40	1 000	10	0	0	1 050
Partenaires	Macao (Chine)	0	0	0	0	0	0
Partenaires	Malaisie	663	1 100	580	0	0	2 344
Partenaires	Malte	8	27	6	0	0	41
Partenaires	Moldavie	66	51	1	0	0	118
Partenaires	Monténégro	27	38	6	0	261	332
Partenaires	Pérou	224	520	0	0	0	745
Partenaires	Qatar	76	110	7	0	0	193
Partenaires	Roumanie	31	63	26	0	0	120
Partenaires	Russie	425	2 044	0	0	0	2 469
Partenaires	Singapour	22	115	43	0	0	179
Partenaires	Taipei chinois	78	568	0	0	0	647
Partenaires	Thaïlande	114	1 830	163	0	0	2 107
Partenaires	Trinité-et-Tobago	0	0	0	0	0	0
Partenaires	Tunisie	0	0	61	0	0	61
Partenaires	Émirats arabes unis	30	75	47	0	0	152
Partenaires	Uruguay	10	22	0	0	0	32
Partenaires	Viet Nam	0	0	0	0	0	0

Codes d'exclusion :
Code 1 : Handicap fonctionnel : l'élève souffre d'un handicap physique permanent de modéré à grave.
Code 2 : Handicap mental : l'élève souffre d'un handicap mental ou de troubles émotionnels, et d'un retard cognitif identifié lors de tests ou diagnostiqué par des professionnels.
Code 3 : Maîtrise insuffisante de la langue de l'évaluation : l'élève a une langue maternelle différente des langues de l'évaluation dans le pays, où il réside depuis moins d'un an.
Code 4 : Autres motifs : définis par les Centres nationaux et approuvés par le Centre international.
Code 5 : Absence de matériel disponible dans la langue de l'évaluation.
Remarque : Pour une explication détaillée des valeurs présentées dans ce tableau, consulter le rapport technique de l'enquête PISA 2015 (*PISA 2015 Technical Report* [OCDE, à paraître en anglais uniquement]).

* Voir les notes au début de cette annexe.

StatLink http://dx.doi.org/10.1787/888933433129

- La **colonne 10** indique le nombre pondéré d'élèves exclus, c'est-à-dire le nombre total d'élèves exclus de la population nationale cible, représenté par le nombre d'élèves exclus de l'échantillon, qui est également défini et classé par catégories d'exclusion dans le tableau A2.2. Les cinq catégories d'élèves exclus sont : les élèves atteints d'un handicap mental (élèves souffrant de troubles mentaux ou émotionnels, ainsi que d'un retard cognitif qui les empêche de se soumettre aux conditions de test de l'enquête PISA ; les élèves atteints d'un handicap fonctionnel permanent modéré à grave les empêchant de se soumettre aux conditions de test de l'enquête PISA ; les élèves dont la maîtrise de la langue de l'évaluation est insuffisante (élèves ne sachant parler ou lire aucune des langues d'évaluation du pays et ne pouvant surmonter cet obstacle linguistique dans les conditions de l'évaluation [généralement, les élèves qui ont suivi moins d'une année de cours dans la langue d'évaluation peuvent être exclus]) ; autre motif, dont la catégorie a été définie par les centres nationaux et approuvée par le centre international ; et les élèves suivant un enseignement donné principalement dans une langue pour laquelle aucune ressource n'est disponible.
- La **colonne 11** indique le pourcentage d'élèves exclus au sein des établissements. Ce taux est calculé en divisant le nombre pondéré d'élèves exclus (colonne 10) par le nombre pondéré d'élèves exclus et participants (somme des colonnes 8 et 10), et en multipliant le tout par 100.
- La **colonne 12** indique le taux global d'exclusion qui représente le pourcentage pondéré de la population nationale cible théorique exclue de PISA, soit au niveau des établissements soit au niveau des élèves au sein des établissements. Ce taux est calculé comme suit : la somme du taux d'établissements exclus (colonne 6 divisée par 100) et du taux d'exclusion intra-établissement (la colonne 11 divisée par 100), multipliée par un, moins le pourcentage d'élèves exclus dans les établissements (le nombre de la colonne 6 divisé par 100). Ce résultat est ensuite multiplié par 100.
- La **colonne 13** indique la mesure dans laquelle l'échantillon PISA est représentatif de la population nationale cible théorique. L'Australie, le Canada, le Danemark, l'Estonie, la Lettonie, la Lituanie, le Luxembourg, le Monténégro, la Norvège, la Nouvelle-Zélande, le Royaume-Uni et la Suède sont les seuls pays où la représentativité de l'échantillon est inférieure à 95 %.
- La **colonne 14** indique la mesure dans laquelle l'échantillon PISA est représentatif des effectifs d'élèves de 15 ans. Cet indice mesure la proportion totale de la population nationale scolarisée représentée par la proportion d'élèves non exclus des échantillons d'élèves. Il tient compte à la fois des exclusions d'élèves et d'établissements. Les valeurs proches de 100 indiquent que l'échantillon PISA est représentatif de l'ensemble du système d'éducation défini dans le cadre de l'évaluation PISA 2015. Cet indice correspond au nombre pondéré d'élèves participants (colonne 8) divisé par le nombre pondéré d'élèves participants et exclus (somme des colonnes 8 et 10), multiplié par la population nationale cible théorique (colonne 5), divisé par la population admissible (colonne 2), multiplié par 100.
- La **colonne 15** présente l'indice de la représentativité de la population d'individus de 15 ans. Cet indice correspond au nombre total d'élèves participants (colonne 8) divisé par la population totale d'élèves de 15 ans (colonne 1).

Cette forte représentativité contribue à la comparabilité des résultats de l'évaluation. En effet, même en partant du principe que les élèves exclus auraient systématiquement obtenu des scores plus faibles que les élèves ayant participé et que cette corrélation est moyennement forte, un taux d'exclusion de l'ordre de 5 % aurait vraisemblablement abouti à une surestimation des scores moyens des pays inférieure à 5 points de score (sur une échelle de compétence dont la moyenne internationale s'établit à 500 points et l'écart-type, à 100 points de score). Cette estimation se base sur les calculs suivants : si la corrélation entre la propension à l'exclusion et la performance des élèves est de 0.3, les scores moyens risquent d'être surestimés de 1 point de score si le taux d'exclusion est de 1 %, de 3 points de score si le taux d'exclusion est de 5 %, et de 6 points de score si le taux d'exclusion est de 10 %. Si la corrélation entre la propension à l'exclusion et la performance des élèves est de 0.5, les scores moyens risquent d'être surestimés de 1 point de score si le taux d'exclusion est de 1 %, de 5 points de score si le taux d'exclusion est de 5 %, et de 10 points de score si le taux d'exclusion est de 10 %. Ce calcul repose sur un modèle partant de l'hypothèse d'une répartition bivariée normale pour la performance et la propension à la participation. Pour plus d'informations, consultez le rapport technique sur l'enquête PISA 2015 (*PISA 2015 Technical Report* [OCDE, à paraître en anglais uniquement]).

Procédures d'échantillonnage et taux de participation

Quelle que soit l'enquête, la précision de ses résultats dépend de la qualité des données utilisées par les échantillons nationaux ainsi que des procédures d'échantillonnage. Des normes, des procédures, des instruments et des mécanismes de vérification de la qualité ont été élaborés dans le cadre de l'enquête PISA afin de garantir la comparabilité des informations recueillies dans les échantillons nationaux et des comparaisons de résultats fiables.

La plupart des échantillons PISA ont été conçus comme des échantillons stratifiés à deux degrés (les échantillons conçus différemment par les pays sont présentés dans le rapport technique sur l'enquête PISA 2015 (*PISA 2015 Technical Report* [OCDE, à paraître en anglais uniquement]). Au premier degré, on a échantillonné des établissements scolarisant des élèves de 15 ans. Les établissements ont été échantillonnés systématiquement selon des probabilités proportionnelles à leur taille, laquelle était fonction du nombre estimé d'élèves admissibles (âgés de 15 ans). Dans chaque pays, 150 établissements minimum (dans la mesure où le pays en comptait autant) ont été sélectionnés, même si, dans de nombreux cas, un échantillon plus important a

été constitué à des fins d'analyse nationale. Dans le même temps, des établissements de remplacement ont été identifiés pour le cas où un établissement échantillonné choisirait de ne pas participer à l'évaluation PISA 2015.

En Islande, au Luxembourg, à Macao (Chine), à Malte et au Qatar, tous les établissements et les élèves admissibles ont été échantillonnés.

Les experts du Consortium PISA ont effectué la sélection des échantillons dans la plupart des pays participants et l'ont suivie de près dans les pays qui ont sélectionné leurs propres échantillons. Le second degré du processus d'échantillonnage a consisté à sélectionner les élèves dans les établissements échantillonnés. Une fois les établissements sélectionnés, la liste des élèves de 15 ans qui y étaient scolarisés a été dressée. Dans cette liste, 42 élèves ont été sélectionnés de manière aléatoire (tous les élèves de 15 ans ont été sélectionnés si la liste comptait moins de 42 élèves). Le nombre d'élèves à échantillonner pour chaque établissement pouvait être compris entre 20 et 42 élèves.

Les normes de qualité concernant les données de l'enquête PISA imposent un taux de participation minimum, tant pour les établissements que pour les élèves, afin de minimiser les biais potentiels liés à la participation. Dans les pays respectant ces normes, il est probable que le biais résultant de la non-participation serait négligeable, c'est-à-dire généralement inférieur à l'erreur d'échantillonnage.

Un taux de participation minimum de 85 % a été fixé pour les établissements sélectionnés initialement. Toutefois, lorsque le taux initial de participation des établissements se situait entre 65 % et 85 %, le recours à des établissements de remplacement a permis d'obtenir un taux de participation acceptable. Cette procédure risquant d'augmenter les biais liés à la participation, les pays ont été encouragés à convaincre le plus grand nombre possible d'établissements de l'échantillon initial de participer. Les établissements dont le taux de participation des élèves est compris entre 25 % et 50 % ne sont pas considérés comme des établissements participants, mais leurs résultats ont été inclus dans la base de données et ont été pris en compte dans les différentes estimations. En revanche, les résultats des établissements dont le taux de participation des élèves est inférieur à 25 % n'ont pas été inclus dans la base de données.

L'évaluation PISA 2015 exigeait également un taux de participation de 80 % minimum pour les élèves sélectionnés dans les établissements participants. Ce taux de participation minimum devait être respecté à l'échelle nationale, pas nécessairement dans chaque établissement participant. Des séances de rattrapage ont été imposées dans les établissements où un nombre insuffisant d'élèves avait participé aux séances d'évaluation initiale. Le taux de participation des élèves a été calculé dans tous les établissements de l'échantillon initial et dans tous les établissements participants, qu'ils figurent dans l'échantillon initial ou qu'ils soient des établissements de remplacement, au niveau des élèves ayant participé à la première séance d'évaluation ou à l'éventuelle séance de rattrapage. Les élèves ayant participé à la première séance d'épreuves cognitives ou à une séance de rattrapage sont considérés comme des élèves participants. Ceux qui ont seulement répondu au questionnaire ont été inclus dans la base de données internationale et pris en compte dans les statistiques présentées dans ce rapport s'ils ont indiqué au moins la profession de leur père ou de leur mère.

Le tableau A2.3 indique le taux de participation des élèves et des établissements, avant et après le recours aux établissements de remplacement.

- La **colonne 1** indique le taux pondéré de participation des établissements avant le recours aux établissements de remplacement. Il se calcule en divisant la colonne 2 par la colonne 3.
- La **colonne 2** indique le nombre pondéré d'établissements participants avant le recours aux établissements de remplacement (pondéré en fonction des effectifs d'élèves).
- La **colonne 3** indique le nombre pondéré d'établissements échantillonnés avant le recours aux établissements de remplacement (soit des établissements participants ou non participants, pondéré en fonction des effectifs d'élèves).
- La **colonne 4** indique le nombre non pondéré d'établissements participants avant le recours aux établissements de remplacement.
- La **colonne 5** indique le nombre non pondéré d'établissements participants et non participants avant le recours aux établissements de remplacement.
- La **colonne 6** indique le taux pondéré de participation des établissements après le recours aux établissements de remplacement. Il se calcule en divisant la colonne 7 par la colonne 8.
- La **colonne 7** indique le nombre pondéré d'établissements participants après le recours aux établissements de remplacement (pondéré en fonction des effectifs d'élèves).
- La **colonne 8** indique le nombre pondéré d'établissements échantillonnés après le recours aux établissements de remplacement (soit des établissements participants et non participants, pondéré en fonction des effectifs d'élèves).
- La **colonne 9** indique le nombre non pondéré d'établissements participants après le recours aux établissements de remplacement.

[Partie 1/1]

Tableau A2.3 Taux de réponse

		Échantillon initial, avant recours à des établissements de remplacement					Échantillon final, après recours à des établissements de remplacement					Échantillon final : nombre d'élèves au sein des établissements après recours à des établissements de remplacement				
		Taux pondéré de participation des établissements avant recours à des établissements de remplacement (%)	Nombre pondéré d'établissements participants (également pondéré en fonction des effectifs d'élèves)	Nombre pondéré d'établissements échantillonnés (participants et non participants) (également pondéré en fonction des effectifs d'élèves)	Nombre d'établissements participants (non pondéré)	Nombre d'établissements participants et non participants (non pondéré)	Taux pondéré de participation des établissements après recours à des établissements de remplacement (%)	Nombre pondéré d'établissements participants (également pondéré en fonction des effectifs d'élèves)	Nombre pondéré d'établissements échantillonnés (participants et non participants) (également pondéré en fonction des effectifs d'élèves)	Nombre d'établissements participants (non pondéré)	Nombre d'établissements participants et non participants (non pondéré)	Taux pondéré de participation des élèves après recours à des établissements de remplacement (%)	Nombre d'élèves évalués (pondéré)	Nombre d'élèves échantillonnés (évalués et absents) (pondéré)	Nombre d'élèves évalués (non pondéré)	Nombre d'élèves échantillonnés (évalués et absents) (non pondéré)
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
OCDE	Australie	94	260 657	276 072	720	788	95	262 130	276 072	723	788	84	204 763	243 789	14 089	17 477
	Autriche	100	81 690	81 730	269	273	100	81 690	81 730	269	273	87	63 660	73 521	7 007	9 868
	Belgique	83	98 786	118 915	244	301	95	113 435	118 936	286	301	91	99 760	110 075	9 635	10 602
	Canada	74	283 853	381 133	703	1 008	79	299 512	381 189	726	1 008	81	210 476	260 487	19 604	24 129
	Chili	92	215 139	232 756	207	232	99	230 749	232 757	226	232	93	189 206	202 774	7 039	7 515
	République tchèque	98	86 354	87 999	339	344	98	86 354	87 999	339	344	89	73 386	82 672	6 835	7 693
	Danemark	90	57 803	63 897	327	371	92	58 837	63 931	331	371	89	49 732	55 830	7 149	8 184
	Estonie	100	11 142	11 154	206	207	100	11 142	11 154	206	207	93	10 088	10 822	5 587	5 994
	Finlande	100	58 653	58 782	167	168	100	58 800	58 800	168	168	93	53 198	56 934	5 882	6 294
	France	91	679 984	749 284	232	255	94	706 838	749 284	241	255	88	611 563	693 336	5 980	6 783
	Allemagne	96	764 423	794 206	245	256	99	785 813	794 206	253	256	93	685 972	735 487	6 476	6 944
	Grèce	92	95 030	103 031	190	212	98	101 653	103 218	209	212	94	89 588	94 986	5 511	5 838
	Hongrie	93	83 897	89 808	231	251	99	88 751	89 825	244	251	92	77 212	83 657	5 643	6 101
	Islande	99	4 114	4 163	122	129	99	4 114	4 163	122	129	86	3 365	3 908	3 365	3 908
	Irlande	99	61 023	61 461	167	169	99	61 023	61 461	167	169	89	51 947	58 630	5 741	6 478
	Israël	91	105 192	115 717	169	190	93	107 570	115 717	173	190	90	98 572	108 940	6 598	7 294
	Italie	74	383 933	516 113	414	532	88	451 098	515 515	464	532	88	377 011	430 041	11 477	12 841
	Japon	94	1 087 414	1 151 305	189	200	99	1 139 734	1 151 305	198	200	97	1 096 193	1 127 265	6 647	6 838
	Corée	100	612 937	615 107	168	169	100	612 937	615 107	168	169	99	559 121	567 284	5 581	5 664
	Lettonie	86	14 122	16 334	231	269	93	15 103	16 324	248	269	90	12 799	14 155	4 845	5 368
	Luxembourg	100	5 891	5 891	44	44	100	5 891	5 891	44	44	96	5 299	5 540	5 299	5 540
	Mexique	95	1 311 608	1 373 919	269	284	98	1 339 901	1 373 919	275	284	95	1 290 435	1 352 237	7 568	7 938
	Pays-Bas	63	121 527	191 966	125	201	93	178 929	191 966	184	201	85	152 346	178 985	5 345	6 269
	Nouvelle-Zélande	71	40 623	56 875	145	210	85	48 094	56 913	176	210	80	36 860	45 897	4 453	5 547
	Norvège	95	58 824	61 809	229	241	95	58 824	61 809	229	241	91	50 163	55 277	5 456	6 016
	Pologne	88	314 288	355 158	151	170	99	352 754	355 158	168	170	88	300 617	343 405	4 466	5 108
	Portugal	86	87 756	102 193	213	254	95	97 516	102 537	238	254	82	75 391	91 916	7 180	8 732
	République slovaque	93	50 513	54 499	272	295	99	53 908	54 562	288	295	92	45 357	49 103	6 342	6 900
	Slovénie	98	16 886	17 286	332	349	98	16 896	17 286	333	349	92	15 072	16 424	6 406	7 009
	Espagne	99	404 640	409 246	199	201	100	409 246	409 246	201	201	89	356 509	399 935	6 736	7 540
	Suède	100	93 819	94 097	202	205	100	93 819	94 097	202	205	91	82 582	91 081	5 458	6 013
	Suisse	93	75 482	81 026	212	232	98	79 481	81 375	225	232	92	74 465	80 544	5 838	6 305
	Turquie	97	1 057 318	1 091 317	175	195	99	1 081 935	1 091 528	187	195	95	874 609	918 816	5 895	6 211
	Royaume-Uni	84	591 757	707 415	506	598	93	654 992	707 415	547	598	89	517 426	581 252	14 120	16 123
	États-Unis	67	2 601 386	3 902 089	142	213	83	3 244 399	3 893 828	177	213	90	2 629 707	2 929 771	5 712	6 376
Partenaires	Albanie	100	43 809	43 919	229	230	100	43 809	43 919	229	230	94	38 174	40 814	5 213	5 555
	Algérie	96	341 463	355 216	159	166	96	341 463	355 216	159	166	92	274 121	296 434	5 494	5 934
	Argentine	89	508 448	572 941	212	238	97	556 478	572 941	231	238	90	345 508	382 352	6 311	7 016
	Brésil	93	2 509 198	2 692 686	806	889	94	2 533 711	2 693 137	815	889	87	1 996 574	2 286 505	22 791	26 586
	P-S-J-G (Chine)	88	1 259 845	1 437 201	248	268	100	1 437 652	1 437 652	268	268	97	1 287 710	1 331 794	9 841	10 097
	Bulgarie	100	56 265	56 483	179	180	100	56 600	56 600	180	180	95	50 931	53 685	5 928	6 240
	Colombie	99	664 664	673 817	364	375	100	672 526	673 835	371	375	95	535 682	566 734	11 777	12 611
	Costa Rica	99	66 485	67 073	204	206	99	66 485	67 073	204	206	92	47 494	51 369	6 846	7 411
	Croatie	100	34 575	34 652	160	162	100	34 575	34 652	160	162	91	37 275	40 803	5 809	6 354
	Chypre*	97	8 830	9 126	122	132	97	8 830	9 126	122	132	94	8 016	8 526	5 561	5 957
	République dominicaine	99	136 669	138 187	193	195	99	136 669	138 187	193	195	94	122 620	130 700	4 731	5 026
	ERYM	100	16 426	16 472	106	107	100	16 426	16 472	106	107	95	14 999	15 802	5 324	5 617
	Géorgie	97	40 552	41 595	256	267	99	41 081	41 566	262	267	94	35 567	37 873	5 316	5 689
	Hong-Kong (Chine)	75	45 603	60 716	115	153	90	54 795	60 715	138	153	93	48 222	51 806	5 359	5 747
	Indonésie	98	3 126 468	3 176 076	232	236	100	3 176 076	3 176 076	236	236	98	3 015 844	3 092 773	6 513	6 694
	Jordanie	100	119 024	119 024	250	250	100	119 024	119 024	250	250	97	105 868	108 669	7 267	7 462
	Kazakhstan	100	202 701	202 701	232	232	100	202 701	202 701	232	232	97	187 683	192 921	7 841	8 059
	Kosovo	100	26 924	26 924	224	224	100	26 924	26 924	224	224	99	22 016	22 333	4 826	4 896
	Liban	67	40 542	60 882	208	308	87	53 091	60 797	270	308	95	36 052	38 143	4 546	4 788
	Lituanie	99	31 386	31 588	309	311	100	31 543	31 588	310	311	91	27 070	29 889	6 523	7 202
	Macao (Chine)	100	4 414	4 414	45	45	100	4 414	4 414	45	45	99	4 476	4 507	4 476	4 507
	Malaisie	51	229 340	446 237	147	230	98	437 424	446 100	224	230	97	393 785	407 396	8 843	9 097
	Malte	100	4 341	4 343	59	61	100	4 341	4 343	59	61	85	3 634	4 294	3 634	4 294
	Moldavie	100	30 145	30 145	229	229	100	30 145	30 145	229	229	98	28 754	29 341	5 325	5 436
	Monténégro	100	7 301	7 312	64	65	100	7 301	7 312	64	65	94	6 346	6 766	5 665	6 043
	Pérou	100	468 406	470 651	280	282	100	469 662	470 651	281	282	99	426 205	430 959	6 971	7 054
	Qatar	99	13 333	13 470	166	168	99	13 333	13 470	166	168	94	12 061	12 819	12 061	12 819
	Roumanie	99	171 553	172 652	181	182	100	172 495	172 495	182	182	99	162 918	164 216	4 876	4 910
	Russie	99	1 181 937	1 189 441	209	210	99	1 181 937	1 189 441	209	210	97	1 072 914	1 108 068	6 021	6 215
	Singapour	97	45 299	46 620	175	179	98	45 553	46 620	176	179	93	42 241	45 259	6 105	6 555
	Taipei chinois	100	286 778	286 778	214	214	100	286 778	286 778	214	214	98	246 408	251 424	7 708	7 871
	Thaïlande	99	739 772	751 010	269	273	100	751 010	751 010	273	273	97	614 996	634 795	8 249	8 491
	Trinité-et-Tobago	92	15 904	17 371	141	163	92	15 904	17 371	141	163	79	9 674	12 188	4 587	5 745
	Tunisie	99	121 751	122 767	162	165	99	121 838	122 792	163	165	86	97 337	112 665	5 340	6 175
	Émirats arabes unis	99	49 310	50 060	473	477	99	49 310	50 060	473	477	95	43 774	46 263	14 167	15 014
	Uruguay	98	42 986	43 737	217	221	99	43 442	43 737	219	221	86	32 762	38 023	6 059	7 026
	Viet Nam	100	996 757	996 757	188	188	100	996 757	996 757	188	188	100	871 353	874 859	5 826	5 849

* Voir les notes au début de cette annexe.

StatLink http://dx.doi.org/10.1787/888933433129

[Partie 1/1]

Tableau A2.4a Pourcentage d'élèves par année d'études

		Tous les élèves											
		7e année		8e année		9e année		10e année		11e année		12e année et niveaux supérieurs	
		%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.
OCDE	Australie	0.0	(0.0)	0.1	(0.0)	11.2	(0.3)	74.6	(0.4)	14.0	(0.4)	0.1	(0.0)
	Autriche	0.0	(0.0)	2.0	(0.6)	20.8	(0.9)	71.2	(1.0)	5.9	(0.3)	0.0	(0.0)
	Belgique	0.6	(0.1)	6.4	(0.5)	30.7	(0.7)	61.0	(0.9)	1.3	(0.1)	0.0	(0.0)
	Canada	0.1	(0.0)	0.7	(0.1)	10.8	(0.5)	87.6	(0.6)	0.8	(0.1)	0.0	(0.0)
	Chili	1.7	(0.3)	4.1	(0.6)	24.0	(0.7)	68.1	(1.0)	2.1	(0.2)	0.0	(0.0)
	République tchèque	0.5	(0.1)	3.9	(0.3)	49.4	(1.2)	46.2	(1.2)	0.0	(0.0)	0.0	c
	Danemark	0.2	(0.1)	16.4	(0.6)	81.9	(0.7)	1.4	(0.5)	0.0	c	0.0	c
	Estonie	0.8	(0.2)	21.3	(0.6)	76.6	(0.6)	1.3	(0.3)	0.0	c	0.0	(0.0)
	Finlande	0.5	(0.1)	13.6	(0.4)	85.7	(0.4)	0.0	(0.0)	0.2	(0.1)	0.0	c
	France	0.0	(0.0)	1.0	(0.2)	23.1	(0.6)	72.5	(0.7)	3.2	(0.2)	0.1	(0.1)
	Allemagne	0.5	(0.1)	7.7	(0.4)	47.3	(0.8)	43.1	(0.8)	1.5	(0.5)	0.0	(0.0)
	Grèce	0.2	(0.1)	0.7	(0.2)	3.8	(0.8)	95.3	(0.9)	0.0	c	0.0	c
	Hongrie	1.7	(0.3)	8.5	(0.5)	75.8	(0.7)	14.0	(0.5)	0.0	c	0.0	c
	Islande	0.0	c	0.0	c	0.0	c	100.0	c	0.0	c	0.0	c
	Irlande	0.0	(0.0)	1.8	(0.2)	60.6	(0.7)	26.5	(1.1)	11.1	(0.9)	0.0	c
	Israël	0.0	c	0.1	(0.0)	16.4	(0.9)	82.7	(0.9)	0.9	(0.3)	0.0	c
	Italie	0.1	(0.0)	1.0	(0.2)	15.2	(0.6)	77.2	(0.7)	6.6	(0.3)	0.0	c
	Japon	0.0	c	0.0	c	0.0	c	100.0	(0.0)	0.0	c	0.0	c
	Corée	0.0	c	0.0	c	9.1	(0.8)	90.4	(0.8)	0.5	(0.1)	0.0	c
	Lettonie	0.9	(0.2)	11.7	(0.5)	84.4	(0.6)	2.9	(0.3)	0.0	(0.0)	0.0	c
	Luxembourg	0.3	(0.1)	7.9	(0.1)	50.9	(0.1)	40.3	(0.1)	0.6	(0.0)	0.0	c
	Mexique	2.3	(0.3)	4.8	(0.4)	31.9	(1.4)	60.3	(1.6)	0.5	(0.1)	0.2	(0.0)
	Pays-Bas	0.1	(0.0)	2.8	(0.3)	41.6	(0.6)	54.8	(0.6)	0.8	(0.2)	0.0	(0.0)
	Nouvelle-Zélande	0.0	c	0.0	c	0.0	(0.0)	6.2	(0.3)	88.8	(0.5)	5.0	(0.5)
	Norvège	0.0	c	0.0	c	0.6	(0.1)	99.3	(0.2)	0.1	(0.1)	0.0	c
	Pologne	0.6	(0.1)	4.9	(0.3)	93.8	(0.4)	0.6	(0.2)	0.0	c	0.0	c
	Portugal	3.2	(0.3)	8.4	(0.5)	22.9	(0.9)	65.1	(1.2)	0.4	(0.1)	0.0	c
	République slovaque	2.2	(0.4)	4.6	(0.4)	42.6	(1.3)	50.6	(1.2)	0.1	(0.0)	0.0	c
	Slovénie	0.0	c	0.3	(0.1)	4.8	(0.3)	94.6	(0.4)	0.3	(0.1)	0.0	c
	Espagne	0.1	(0.0)	8.6	(0.5)	23.4	(0.6)	67.9	(0.9)	0.1	(0.1)	0.0	c
	Suède	0.1	(0.1)	3.1	(0.4)	94.9	(0.8)	1.8	(0.7)	0.1	(0.1)	0.0	c
	Suisse	0.5	(0.1)	11.8	(0.7)	61.3	(1.2)	25.9	(1.3)	0.5	(0.1)	0.0	(0.0)
	Turquie	0.6	(0.1)	2.6	(0.4)	20.7	(1.0)	72.9	(1.2)	3.0	(0.3)	0.1	(0.0)
	Royaume-Uni	0.0	c	0.0	c	0.0	c	1.6	(0.3)	97.4	(0.4)	1.0	(0.3)
	États-Unis	0.0	(0.0)	0.5	(0.3)	9.6	(0.7)	72.4	(0.9)	17.3	(0.6)	0.1	(0.0)
Partenaires	Albanie	0.2	(0.1)	1.0	(0.2)	35.8	(2.3)	61.7	(2.3)	1.2	(0.7)	0.0	(0.0)
	Algérie	18.8	(1.0)	23.5	(1.1)	35.1	(1.5)	19.4	(2.1)	3.2	(0.7)	0.0	c
	Brésil	3.5	(0.2)	6.4	(0.4)	12.5	(0.5)	35.9	(0.9)	39.2	(0.8)	2.5	(0.2)
	P-S-J-G (Chine)	1.1	(0.2)	9.2	(0.7)	52.7	(1.7)	34.6	(2.0)	2.2	(0.5)	0.1	(0.0)
	Bulgarie	0.5	(0.2)	3.0	(0.6)	92.2	(0.8)	4.3	(0.4)	0.0	c	0.0	c
	Colombie	5.3	(0.4)	12.3	(0.6)	22.7	(0.6)	40.2	(0.7)	19.5	(0.6)	0.0	c
	Costa Rica	6.2	(0.7)	14.0	(0.7)	33.0	(1.2)	46.5	(1.6)	0.2	(0.1)	0.1	(0.1)
	Croatie	0.0	c	0.2	(0.2)	79.2	(0.5)	20.6	(0.4)	0.0	c	0.0	c
	Chypre*	0.0	c	0.3	(0.0)	5.8	(0.1)	93.1	(0.1)	0.7	(0.1)	0.0	c
	République dominicaine	7.1	(0.8)	13.8	(1.2)	20.6	(0.8)	41.9	(1.1)	14.2	(0.7)	2.4	(0.3)
	ERYM	0.1	(0.1)	0.1	(0.1)	70.2	(0.2)	29.7	(0.2)	0.0	c	0.0	c
	Géorgie	0.1	(0.0)	0.8	(0.2)	22.0	(0.8)	76.0	(0.9)	1.1	(0.3)	0.0	c
	Hong-Kong (Chine)	1.1	(0.1)	5.6	(0.4)	26.0	(0.7)	66.7	(0.7)	0.6	(0.5)	0.0	c
	Indonésie	2.1	(0.3)	8.1	(0.7)	42.1	(1.5)	45.5	(1.6)	2.3	(0.4)	0.0	(0.0)
	Jordanie	0.2	(0.1)	0.6	(0.1)	6.6	(0.4)	92.6	(0.4)	0.0	c	0.0	c
	Kosovo	0.0	(0.1)	0.6	(0.1)	24.9	(0.8)	72.4	(0.9)	2.1	(0.2)	0.0	c
	Liban	3.7	(0.5)	8.3	(0.8)	16.6	(1.1)	62.3	(1.4)	9.0	(0.8)	0.1	(0.1)
	Lituanie	0.1	(0.0)	2.6	(0.2)	86.3	(0.4)	11.0	(0.4)	0.0	(0.0)	0.0	c
	Macao (Chine)	2.9	(0.1)	12.2	(0.2)	29.7	(0.2)	54.5	(0.1)	0.6	(0.1)	0.0	c
	Malte	0.0	c	0.0	c	0.3	(0.1)	6.1	(0.2)	93.6	(0.1)	0.1	(0.0)
	Moldavie	0.2	(0.1)	7.6	(0.5)	84.5	(0.8)	7.5	(0.8)	0.0	(0.0)	0.0	c
	Monténégro	0.0	c	0.0	c	83.7	(0.1)	16.3	(0.1)	0.0	c	0.0	c
	Pérou	2.5	(0.3)	6.6	(0.4)	15.9	(0.5)	50.2	(0.8)	24.8	(0.8)	0.0	c
	Qatar	0.9	(0.1)	3.5	(0.1)	16.3	(0.1)	60.7	(0.1)	18.0	(0.1)	0.6	(0.0)
	Roumanie	1.4	(0.3)	8.9	(0.5)	74.8	(0.9)	14.9	(0.7)	0.0	c	0.0	c
	Russie	0.2	(0.1)	6.6	(0.3)	79.7	(1.5)	13.4	(1.5)	0.1	(0.0)	0.0	c
	Singapour	0.0	(0.0)	1.9	(0.3)	7.9	(0.8)	90.0	(1.0)	0.1	(0.0)	0.1	(0.0)
	Taipei chinois	0.0	c	0.0	c	35.4	(0.7)	64.6	(0.7)	0.0	c	0.0	c
	Thaïlande	0.2	(0.1)	0.6	(0.2)	23.8	(1.0)	72.9	(1.0)	2.4	(0.4)	0.0	c
	Trinité-et-Tobago	3.3	(0.2)	10.8	(0.3)	27.3	(0.3)	56.5	(0.3)	2.2	(0.2)	0.0	c
	Tunisie	4.3	(0.3)	10.6	(0.8)	19.6	(1.3)	60.9	(1.7)	4.6	(0.4)	0.0	c
	Émirats arabes unis	0.6	(0.1)	2.5	(0.3)	10.6	(0.7)	53.4	(0.8)	31.4	(0.8)	1.5	(0.1)
	Uruguay	7.5	(0.6)	9.7	(0.5)	20.7	(0.7)	61.3	(1.2)	0.8	(0.1)	0.0	c
	Viet Nam	0.3	(0.1)	1.7	(0.4)	7.7	(1.8)	90.4	(2.2)	0.0	(0.0)	0.0	c
	Argentine**	1.6	(0.4)	9.7	(0.8)	27.4	(1.2)	58.5	(1.6)	2.8	(0.3)	0.0	c
	Kazakhstan**	0.1	(0.1)	2.7	(0.3)	60.4	(1.7)	36.2	(1.8)	0.6	(0.1)	0.0	c
	Malaisie**	0.0	c	0.0	c	3.2	(0.6)	96.4	(0.7)	0.4	(0.3)	0.0	c

* Voir les notes au début de cette annexe.
** Couverture trop limitée pour garantir la comparabilité (voir l'annexe A4).
StatLink http://dx.doi.org/10.1787/888933433129

[Partie 1/1]

Tableau A2.4b Pourcentage d'élèves par année d'études et par sexe

		Garçons												Filles									
		7e année		8e année		9e année		10e année		11e année		12e année et niveaux supérieurs		7e année		8e année		9e année		10e année		11e année		12e année et niveaux supérieurs	
		%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.	%	Er.-T.
OCDE	Australie	0.0	(0.0)	0.2	(0.1)	13.2	(0.4)	73.5	(0.5)	13.1	(0.5)	0.0	(0.0)	0.0	(0.0)	0.1	(0.0)	9.2	(0.3)	75.7	(0.5)	14.9	(0.6)	0.1	(0.1)
	Autriche	0.1	(0.1)	2.0	(0.4)	21.6	(1.2)	71.1	(1.2)	5.2	(0.4)	0.0	(0.0)	0.0	c	2.0	(0.9)	20.0	(1.0)	71.4	(1.3)	6.6	(0.4)	0.0	(0.0)
	Belgique	0.7	(0.1)	6.7	(0.5)	33.6	(1.0)	57.9	(1.1)	1.2	(0.2)	0.0	c	0.6	(0.1)	6.2	(0.5)	27.7	(0.8)	64.2	(1.1)	1.3	(0.1)	0.0	(0.0)
	Canada	0.1	(0.1)	1.0	(0.2)	11.7	(0.6)	86.5	(0.6)	0.7	(0.1)	0.0	(0.0)	0.1	(0.0)	0.4	(0.1)	9.9	(0.6)	88.8	(0.6)	0.8	(0.1)	0.0	(0.0)
	Chili	2.2	(0.5)	4.8	(0.8)	26.4	(0.9)	64.8	(1.3)	1.8	(0.2)	0.1	(0.1)	1.2	(0.4)	3.5	(0.7)	21.5	(0.8)	71.4	(1.1)	2.4	(0.3)	0.0	c
	République tchèque	0.6	(0.2)	5.5	(0.5)	52.3	(1.5)	41.5	(1.6)	0.0	(0.0)	0.0	c	0.4	(0.2)	2.2	(0.3)	46.2	(1.5)	51.2	(1.6)	0.0	c	0.0	c
	Danemark	0.3	(0.1)	21.9	(0.9)	76.6	(1.0)	1.2	(0.5)	0.0	c	0.0	c	0.1	(0.1)	10.8	(0.5)	87.3	(0.7)	1.7	(0.6)	0.0	c	0.0	c
	Estonie	1.3	(0.3)	23.7	(0.9)	74.2	(0.8)	0.8	(0.3)	0.0	c	0.0	(0.0)	0.2	(0.1)	18.8	(0.8)	79.1	(0.8)	1.9	(0.4)	0.0	c	0.0	c
	Finlande	0.4	(0.1)	15.5	(0.6)	83.9	(0.6)	0.0	(0.0)	0.2	(0.1)	0.0	c	0.5	(0.1)	11.5	(0.5)	87.7	(0.5)	0.0	c	0.3	(0.2)	0.0	c
	France	0.0	c	1.0	(0.2)	26.1	(0.9)	69.6	(1.0)	3.1	(0.3)	0.2	(0.1)	0.1	(0.1)	1.0	(0.2)	20.1	(0.6)	75.4	(0.8)	3.3	(0.3)	0.1	(0.0)
	Allemagne	0.7	(0.2)	9.0	(0.5)	50.1	(1.0)	38.8	(1.0)	1.4	(0.4)	0.0	(0.0)	0.3	(0.1)	6.3	(0.6)	44.3	(0.9)	47.5	(1.0)	1.6	(0.6)	0.0	c
	Grèce	0.4	(0.2)	1.1	(0.3)	4.7	(1.0)	93.8	(1.2)	0.0	c	0.0	c	0.1	(0.1)	0.2	(0.1)	2.8	(0.8)	96.9	(0.8)	0.0	c	0.0	c
	Hongrie	1.8	(0.4)	10.1	(0.6)	75.6	(0.9)	12.5	(0.6)	0.0	c	0.0	c	1.6	(0.4)	6.9	(0.8)	76.0	(0.9)	15.5	(0.7)	0.0	c	0.0	c
	Islande	0.0	c	0.0	c	0.0	c	100.0	c	0.0	c	0.0	c	0.0	c	0.0	c	0.0	c	100.0	c	0.0	c	0.0	c
	Irlande	0.0	c	2.2	(0.3)	62.8	(0.9)	24.1	(1.2)	10.9	(1.0)	0.0	c	0.0	(0.0)	1.4	(0.2)	58.2	(0.9)	29.0	(1.4)	11.3	(1.1)	0.0	c
	Israël	0.0	c	0.1	(0.1)	18.0	(1.2)	80.9	(1.3)	1.1	(0.6)	0.0	c	0.0	c	0.1	(0.0)	14.9	(0.8)	84.4	(0.8)	0.7	(0.1)	0.0	c
	Italie	0.2	(0.1)	1.3	(0.3)	18.1	(0.8)	75.0	(0.9)	5.4	(0.4)	0.0	c	0.1	(0.0)	0.7	(0.2)	12.2	(0.8)	79.3	(1.0)	7.7	(0.5)	0.0	c
	Japon	0.0	c	0.0	c	0.0	c	100.0	c	0.0	c	0.0	c	0.0	c	0.0	c	0.0	c	100.0	c	0.0	c	0.0	c
	Corée	0.0	c	0.0	c	10.1	(1.4)	89.4	(1.4)	0.5	(0.1)	0.0	c	0.0	c	0.0	c	8.0	(0.8)	91.5	(0.8)	0.5	(0.1)	0.0	c
	Lettonie	1.5	(0.4)	14.7	(0.8)	81.8	(0.9)	1.9	(0.3)	0.0	(0.0)	0.0	c	0.4	(0.2)	8.7	(0.7)	87.0	(0.7)	3.9	(0.4)	0.0	c	0.0	c
	Luxembourg	0.2	(0.1)	9.4	(0.2)	52.4	(0.3)	37.3	(0.2)	0.7	(0.1)	0.0	c	0.3	(0.1)	6.4	(0.2)	49.4	(0.2)	43.3	(0.2)	0.6	(0.1)	0.0	c
	Mexique	3.1	(0.5)	5.9	(0.6)	32.2	(1.5)	58.0	(1.6)	0.6	(0.2)	0.2	(0.0)	1.5	(0.3)	3.7	(0.4)	31.6	(1.7)	62.5	(1.7)	0.4	(0.1)	0.2	(0.1)
	Pays-Bas	0.0	(0.0)	3.8	(0.4)	45.3	(0.8)	50.2	(0.8)	0.8	(0.3)	0.0	c	0.1	(0.0)	1.9	(0.3)	38.0	(0.7)	59.3	(0.7)	0.7	(0.2)	0.0	(0.0)
	Nouvelle-Zélande	0.0	c	0.0	c	0.0	c	6.9	(0.5)	88.6	(0.8)	4.5	(0.5)	0.0	c	0.0	c	0.0	(0.0)	5.4	(0.4)	89.1	(0.6)	5.5	(0.6)
	Norvège	0.0	c	0.0	c	0.8	(0.2)	99.1	(0.2)	0.1	(0.1)	0.0	c	0.0	c	0.0	c	0.3	(0.1)	99.6	(0.1)	0.1	(0.1)	0.0	c
	Pologne	0.9	(0.2)	6.8	(0.5)	92.1	(0.6)	0.2	(0.2)	0.0	c	0.0	c	0.4	(0.1)	3.0	(0.3)	95.6	(0.5)	1.1	(0.3)	0.0	c	0.0	c
	Portugal	4.2	(0.4)	10.5	(0.7)	25.4	(1.0)	59.6	(1.4)	0.3	(0.1)	0.0	c	2.1	(0.4)	6.4	(0.5)	20.5	(0.9)	70.5	(1.2)	0.5	(0.1)	0.0	c
	République slovaque	2.4	(0.4)	4.8	(0.5)	43.5	(1.6)	49.4	(1.8)	0.0	c	0.0	c	1.9	(0.5)	4.3	(0.6)	41.7	(1.8)	51.9	(1.8)	0.1	(0.1)	0.0	c
	Slovénie	0.0	c	0.5	(0.2)	5.4	(0.7)	93.9	(0.7)	0.2	(0.1)	0.0	c	0.0	c	0.2	(0.1)	4.1	(0.6)	95.3	(0.6)	0.4	(0.2)	0.0	c
	Espagne	0.1	(0.1)	10.7	(0.7)	25.4	(0.8)	63.7	(1.1)	0.1	(0.1)	0.0	c	0.0	c	6.5	(0.5)	21.3	(0.8)	72.1	(1.0)	0.1	(0.1)	0.0	c
	Suède	0.1	(0.1)	3.5	(0.5)	95.0	(0.9)	1.4	(0.7)	0.1	(0.1)	0.0	c	0.2	(0.1)	2.6	(0.4)	94.9	(1.0)	2.3	(0.9)	0.1	(0.1)	0.0	c
	Suisse	0.7	(0.2)	13.4	(0.8)	60.7	(1.1)	24.7	(1.2)	0.5	(0.1)	0.0	c	0.3	(0.1)	10.1	(0.8)	62.0	(1.7)	27.2	(1.9)	0.5	(0.2)	0.0	(0.0)
	Turquie	0.8	(0.3)	3.1	(0.6)	25.4	(1.2)	68.4	(1.6)	2.2	(0.4)	0.1	(0.1)	0.4	(0.2)	2.1	(0.4)	16.1	(1.1)	77.5	(1.3)	3.8	(0.4)	0.1	(0.0)
	Royaume-Uni	0.0	c	0.0	c	0.0	c	1.9	(0.5)	97.3	(0.6)	0.9	(0.3)	0.0	c	0.0	c	0.0	c	1.4	(0.2)	97.5	(0.3)	1.1	(0.3)
	États-Unis	0.0	c	0.5	(0.4)	11.6	(0.8)	72.4	(1.0)	15.3	(0.7)	0.2	(0.1)	0.1	(0.1)	0.5	(0.2)	7.6	(0.6)	72.4	(0.9)	19.4	(0.7)	0.1	(0.0)
Partenaires	Albanie	0.2	(0.2)	0.9	(0.2)	41.2	(2.7)	56.3	(2.6)	1.3	(0.9)	0.0	(0.0)	0.1	(0.1)	1.1	(0.3)	30.4	(2.1)	67.1	(2.2)	1.2	(0.5)	0.1	(0.0)
	Algérie	24.4	(1.3)	25.7	(1.2)	32.6	(1.5)	14.7	(1.9)	2.6	(0.7)	0.0	c	12.6	(1.1)	21.0	(1.2)	37.9	(2.0)	24.6	(2.5)	3.9	(0.8)	0.0	c
	Brésil	4.6	(0.3)	7.8	(0.6)	13.9	(0.6)	36.5	(1.0)	35.3	(0.9)	1.8	(0.2)	2.4	(0.2)	5.0	(0.4)	11.1	(0.6)	35.3	(0.9)	43.0	(0.9)	3.1	(0.2)
	P-S-J-G (Chine)	1.2	(0.2)	9.9	(0.7)	55.4	(1.7)	31.6	(1.9)	1.9	(0.5)	0.1	(0.0)	1.1	(0.2)	8.4	(0.8)	49.6	(1.8)	38.1	(2.2)	2.6	(0.5)	0.1	(0.1)
	Bulgarie	0.6	(0.2)	4.1	(0.8)	91.8	(1.0)	3.5	(0.4)	0.0	c	0.0	c	0.4	(0.2)	1.8	(0.4)	92.7	(0.7)	5.2	(0.4)	0.0	c	0.0	c
	Colombie	7.2	(0.6)	14.3	(0.8)	25.2	(0.8)	37.1	(0.9)	16.2	(0.8)	0.0	c	3.6	(0.4)	10.5	(0.7)	20.5	(0.9)	42.9	(1.0)	22.5	(0.8)	0.0	c
	Costa Rica	7.8	(0.8)	16.7	(0.8)	34.3	(1.2)	41.2	(1.5)	0.1	(0.0)	0.0	c	4.7	(0.7)	11.4	(0.7)	31.8	(1.4)	51.6	(1.8)	0.3	(0.1)	0.2	(0.1)
	Croatie	0.0	c	0.2	(0.1)	80.5	(0.5)	19.4	(0.5)	0.0	c	0.0	c	0.0	c	0.3	(0.2)	78.0	(0.7)	21.7	(0.7)	0.0	c	0.0	c
	Chypre*	0.0	c	0.3	(0.1)	6.6	(0.2)	92.4	(0.2)	0.6	(0.1)	0.0	c	0.0	c	0.3	(0.1)	5.1	(0.2)	93.8	(0.2)	0.8	(0.1)	0.0	c
	République dominicaine	10.3	(1.1)	16.4	(1.5)	23.3	(1.2)	37.2	(1.4)	11.1	(0.8)	1.7	(0.3)	4.0	(0.6)	11.2	(1.1)	18.1	(0.8)	46.5	(1.1)	17.2	(0.8)	3.0	(0.3)
	ERYM	0.2	(0.2)	0.2	(0.2)	70.9	(0.3)	28.8	(0.2)	0.0	c	0.0	c	0.0	c	0.0	c	69.4	(0.3)	30.6	(0.3)	0.0	c	0.0	c
	Géorgie	0.1	(0.0)	0.9	(0.2)	23.0	(1.0)	75.2	(1.0)	0.8	(0.2)	0.0	c	0.1	(0.1)	0.7	(0.2)	20.9	(0.9)	76.8	(1.0)	1.5	(0.4)	0.0	c
	Hong-Kong (Chine)	1.3	(0.2)	6.4	(0.5)	28.5	(0.8)	63.3	(0.9)	0.5	(0.4)	0.0	c	1.0	(0.2)	4.7	(0.4)	23.5	(0.8)	70.2	(0.9)	0.6	(0.6)	0.0	c
	Indonésie	2.5	(0.4)	8.9	(0.9)	44.3	(1.9)	42.1	(2.0)	2.1	(0.4)	0.0	(0.0)	1.7	(0.3)	7.2	(1.0)	39.8	(1.9)	48.9	(2.1)	2.4	(0.4)	0.0	c
	Jordanie	0.1	(0.1)	0.5	(0.1)	6.6	(0.7)	92.9	(0.7)	0.0	c	0.0	c	0.2	(0.1)	0.7	(0.1)	6.6	(0.6)	92.4	(0.6)	0.0	c	0.0	c
	Kosovo	0.1	(0.1)	0.5	(0.1)	26.4	(0.9)	71.5	(1.0)	1.6	(0.3)	0.0	c	0.0	c	0.7	(0.2)	23.5	(1.0)	73.3	(1.0)	2.5	(0.3)	0.0	c
	Liban	4.0	(0.6)	8.2	(0.9)	17.2	(1.4)	63.5	(1.7)	6.9	(0.7)	0.2	(0.1)	3.4	(0.6)	8.3	(1.0)	16.1	(1.2)	61.2	(1.8)	10.8	(1.2)	0.1	(0.1)
	Lituanie	0.2	(0.1)	3.5	(0.3)	87.4	(0.6)	8.8	(0.5)	0.0	(0.0)	0.0	c	0.0	(0.0)	1.7	(0.2)	85.1	(0.7)	13.1	(0.6)	0.0	(0.0)	0.0	c
	Macao (Chine)	4.3	(0.2)	16.4	(0.3)	30.8	(0.2)	48.2	(0.2)	0.4	(0.1)	0.0	c	1.6	(0.2)	8.0	(0.2)	28.7	(0.3)	60.8	(0.3)	0.9	(0.2)	0.0	c
	Malte	0.0	c	0.0	c	0.5	(0.1)	6.8	(0.3)	92.7	(0.2)	0.0	c	0.0	c	0.0	c	0.1	(0.0)	5.4	(0.2)	94.4	(0.2)	0.1	(0.1)
	Moldavie	0.3	(0.1)	8.2	(0.7)	86.3	(0.9)	5.0	(0.9)	0.1	(0.1)	0.0	c	0.2	(0.1)	7.0	(0.6)	82.8	(1.2)	10.1	(1.2)	0.0	c	0.0	c
	Monténégro	0.0	c	0.0	c	85.2	(0.2)	14.8	(0.2)	0.0	c	0.0	c	0.0	c	0.0	c	82.2	(0.2)	17.8	(0.2)	0.0	c	0.0	c
	Pérou	3.0	(0.5)	7.5	(0.5)	17.9	(0.7)	48.7	(0.9)	22.9	(1.0)	0.0	c	1.9	(0.3)	5.6	(0.5)	14.0	(0.6)	51.7	(1.0)	26.8	(0.9)	0.0	c
	Qatar	0.8	(0.1)	3.6	(0.1)	18.0	(0.2)	59.3	(0.2)	17.6	(0.2)	0.6	(0.1)	1.0	(0.1)	3.4	(0.1)	14.5	(0.1)	62.1	(0.2)	18.4	(0.2)	0.6	(0.1)
	Roumanie	1.7	(0.4)	10.7	(0.8)	74.3	(1.0)	13.3	(0.7)	0.0	c	0.0	c	1.1	(0.4)	7.2	(0.8)	75.3	(1.1)	16.4	(0.8)	0.0	c	0.0	c
	Russie	0.2	(0.1)	7.2	(0.5)	80.1	(1.7)	12.4	(1.7)	0.0	(0.0)	0.0	c	0.1	(0.1)	6.0	(0.4)	79.3	(1.5)	14.4	(1.6)	0.1	(0.1)	0.0	c
	Singapour	0.1	(0.0)	1.8	(0.3)	8.9	(0.9)	89.1	(1.1)	0.1	(0.1)	0.0	(0.0)	0.0	(0.0)	2.0	(0.4)	6.9	(0.8)	90.8	(1.1)	0.2	(0.1)	0.1	(0.0)
	Taipei chinois	0.0	c	0.0	c	36.5	(1.3)	63.5	(1.3)	0.0	c	0.0	c	0.0	c	0.0	c	34.3	(1.3)	65.7	(1.3)	0.0	c	0.0	c
	Thaïlande	0.2	(0.1)	0.8	(0.3)	25.4	(1.2)	71.4	(1.2)	2.3	(0.4)	0.0	c	0.3	(0.1)	0.5	(0.2)	22.5	(1.3)	74.1	(1.3)	2.6	(0.4)	0.0	c
	Trinité-et-Tobago	3.7	(0.3)	14.2	(0.5)	30.8	(0.5)	48.9	(0.5)	2.4	(0.2)	0.0	c	2.8	(0.2)	7.5	(0.4)	23.8	(0.4)	63.9	(0.5)	2.0	(0.3)	0.0	c
	Tunisie	5.9	(0.5)	13.8	(1.0)	22.0	(1.4)	54.0	(1.9)	4.3	(0.5)	0.0	c	3.0	(0.3)	7.8	(0.7)	17.5	(1.4)	67.0	(1.8)	4.8	(0.5)	0.0	c
	Émirats arabes unis	0.7	(0.1)	2.9	(0.4)	11.4	(1.1)	54.0	(1.3)	29.6	(1.0)	1.4	(0.2)	0.4	(0.1)	2.2	(0.5)	9.9	(0.9)	52.8	(0.9)	33.1	(1.1)	1.6	(0.2)
	Uruguay	9.2	(0.8)	11.2	(0.7)	22.5	(0.9)	56.5	(1.5)	0.5	(0.1)	0.0	c	6.0	(0.7)	8.3	(0.6)	19.0	(0.8)	65.6	(1.1)	1.1	(0.2)	0.0	c
	Viet Nam	0.5	(0.2)	2.3	(0.6)	11.1	(2.6)	86.1	(3.2)	0.0	c	0.0	c	0.1	(0.0)	1.1	(0.4)	4.6	(1.2)	94.2	(1.4)	0.0	(0.0)	0.0	c
	Argentine**	2.3	(0.6)	11.5	(0.9)	27.8	(1.3)	56.0	(1.8)	2.4	(0.3)	0.0	c	1.0	(0.3)	8.1	(0.9)	26.9	(1.4)	60.8	(1.7)	3.2	(0.3)	0.0	c
	Kazakhstan**	0.1	(0.1)	3.1	(0.4)	62.8	(2.3)	33.5	(2.4)	0.5	(0.1)	0.0	c	0.1	(0.1)	2.3	(0.3)	57.8	(1.7)	39.0	(1.8)	0.7	(0.1)	0.0	c
	Malaisie**	0.0	c	0.0	c	4.2	(0.8)	95.4	(0.9)	0.4	(0.3)	0.0	c	0.0	c	0.0	c	2.3	(0.5)	97.2	(0.6)	0.4	(0.4)	0.0	c

* Voir les notes au début de cette annexe.
** Couverture trop limitée pour garantir la comparabilité (voir l'annexe A4).
StatLink http://dx.doi.org/10.1787/888933433129

- La **colonne 10** indique le nombre non pondéré d'établissements participants et non participants après le recours aux établissements de remplacement.
- La **colonne 11** indique le taux pondéré de participation des établissements après le recours aux établissements de remplacement. Il se calcule en divisant la colonne 12 par la colonne 13.
- La **colonne 12** indique le nombre pondéré d'élèves évalués.
- La **colonne 13** indique le nombre pondéré d'élèves échantillonnés (incluant les élèves évalués et les élèves absents le jour de l'évaluation).
- La **colonne 14** indique le nombre non pondéré d'élèves évalués. Il faut rappeler que les élèves scolarisés dans des établissements où le taux de participation des élèves était inférieur à 50 % n'ont pas été pris en compte dans le calcul de ces taux (pondérés et non pondérés).
- La **colonne 15** indique le nombre non pondéré d'élèves échantillonnés (incluant les élèves évalués et les élèves absents le jour de l'évaluation). Rappelons que les élèves scolarisés dans des établissements où moins de la moitié des élèves admissibles ont répondu aux épreuves n'ont pas été inclus dans le calcul de ces taux (pondérés et non pondérés).

Définition des établissements

Certains pays ont échantillonné des sous-groupes d'établissements et non des établissements, ce qui est susceptible d'altérer l'estimation des composantes de la variance entre les établissements. En Allemagne, en Autriche, en Hongrie, au Japon, en République tchèque, en Roumanie et en Slovénie, les établissements proposant plus d'un programme d'études ont été divisés en unités distinctes pour chaque programme. Aux Pays-Bas, les établissements proposant à la fois les programmes d'études du premier et du deuxième cycle du secondaire ont été divisés en unités pour chaque niveau de programmes. En Communauté flamande de Belgique, les établissements comptant plusieurs sites (campus) scolaires ont été échantillonnés par site, alors qu'en Communauté française de Belgique, ils ont été échantillonnés par unité administrative. En Australie, les établissements comptant plusieurs sites ont été échantillonnés par site. En Argentine et en Croatie, les établissements comptant plusieurs sites ont été échantillonnés par localisation. En Espagne, les établissements du Pays basque comptant plusieurs modèles linguistiques ont été divisés par modèle linguistique lors de l'échantillonnage. Au Luxembourg, un établissement se situant à la frontière avec l'Allemagne a été divisé en fonction du pays de résidence des élèves. De plus, les écoles internationales y ont été divisées en deux sous-groupes : un premier dans lequel la langue d'instruction des élèves figurait parmi les trois langues officielles du pays, et un deuxième reprenant les établissements exclus parce qu'aucun matériel n'était disponible dans une des langues d'instruction. Aux Émirats arabes unis, certains établissements ont été divisés en unités en fonction des programmes ou du sexe des élèves. En Suède, à la suite de la réorganisation de l'enseignement, certains établissements ont été divisés en unités et un chef d'établissement a été désigné pour chaque unité. Au Portugal, les établissements ont été réorganisés pour créer des groupements scolaires, les unités au sein d'un même groupement se partageant les enseignants et le chef d'établissement.

Années d'études

Les élèves évalués lors de l'enquête PISA 2015 sont scolarisés dans des années d'études différentes. Le pourcentage d'élèves par année d'études est représenté par pays dans le tableau A2.4a et selon le sexe dans chaque pays dans le tableau A2.4b.

Référence

OCDE (à paraître en anglais uniquement), *PISA 2015 Technical Report*, PISA, Éditions OCDE.

ANNEXE A3
NOTE TECHNIQUE SUR LES ANALYSES DU PRÉSENT VOLUME

Méthodes et définitions

Risque relatif

Le risque relatif est une mesure de l'association entre un antécédent et un résultat. Le risque relatif est simplement le ratio de deux risques, à savoir le risque d'observer le résultat en présence de l'antécédent et le risque d'observer ce même résultat en l'absence de l'antécédent. Les conventions utilisées ci-après sont présentées dans le graphique A3.1.

Graphique A3.1 ■ **Légendes utilisées dans un tableau à deux entrées**

p_{11}	p_{12}	$p_{1.}$
p_{21}	p_{22}	$p_{2.}$
$p_{.1}$	$p_{.2}$	$p_{..}$

La valeur p_{ij} représente les probabilités de chaque cellule et est égale au nombre d'observations dans une cellule donnée divisé par le nombre total d'observations. Les valeurs $p_{i.}$ et $p_{j.}$ représentent respectivement les probabilités marginales pour chaque ligne et chaque colonne. Les probabilités marginales sont égales aux fréquences marginales divisées par le nombre total d'élèves.

En partant du principe que les lignes représentent l'antécédent, la première correspondant à la présence de l'antécédent et la deuxième, à l'absence de l'antécédent, et que les colonnes représentent le résultat, la première correspondant à la présence du résultat et la deuxième, à l'absence du résultat, le risque relatif (RR) se calcule comme suit :

$$RR = \frac{(p_{11}/p_{1.})}{(p_{21}/p_{2.})}$$

Rapport de cotes

Une notation identique peut être utilisée pour définir le rapport de cotes, qui mesure également la probabilité relative d'observer un résultat particulier dans les deux groupes. Le rapport de cote (OR) relatif à l'observation du résultat en présence d'un antécédent est définit comme suit :

$$OR = \frac{(p_{11}/p_{12})}{(p_{21}/p_{22})}$$

où la valeur p_{11}/p_{12} représente la « cote », c'est-à-dire la probabilité d'observer le résultat en présence de l'antécédent, et la valeur p_{21}/p_{22} la probabilité d'observer ce même résultat en l'absence de l'antécédent.

La régression logistique peut être utilisée pour réaliser une estimation du rapport de cotes : le coefficient logit exponentiel pour une variable binaire est équivalent au rapport de cotes. Un rapport de cotes « généralisé », qui prend en compte d'autres différences entre les groupes, peut être estimé en introduisant des variables de contrôle dans la régression logistique.

Statistiques basées sur des modèles multiniveau

Les statistiques basées sur des modèles multiniveau comprennent des composantes de la variance (variance intra-établissement et inter-établissements), l'indice d'inclusion dérivé de ces composantes et des cœfficients de régression, qui sont mentionnés le cas échéant. Les modèles multiniveau sont généralement des modèles de régression à deux niveaux (niveau Élève et niveau Établissement), avec des résidus répartis normalement, et les estimations sont calculées à l'aide de la méthode de la probabilité la plus vraisemblable. Si la variable dépendante est la performance en sciences, en compréhension de l'écrit ou en mathématiques, dix valeurs plausibles du score de chaque élève sur l'échelle de culture mathématique sont imputées. L'estimation des modèles a été réalisée à l'aide du module « mixte » Stata® (version 14.1).

Dans les modèles multiniveau, une pondération a été appliquée tant au niveau Élève qu'au niveau Établissement. Elle vise à ajuster les différences de probabilité inhérentes à la sélection d'élèves dans un échantillon. Ces différences résultent de facteurs tant au niveau Établissement qu'au niveau Élève, car l'échantillonnage s'effectue à deux niveaux dans l'enquête PISA. Dans les modèles multiniveau, les pondérations finales des élèves (W_FSTUWT) sont utilisées. Les pondérations intra-établissement correspondent aux pondérations finales des élèves, après mise à l'échelle pour obtenir l'équivalence avec la taille de l'échantillon d'établissements. Les pondérations inter-établissements correspondent à la somme des pondérations finales des élèves (W_FSTUWT) dans chaque établissement. La définition des pondérations inter-établissements est la même que dans les rapports initiaux de l'enquête PISA 2012.

L'indice d'inclusion est défini et estimé comme suit :

$$100 * \frac{\sigma_w^2}{\sigma_w^2 + \sigma_b^2}$$

ou σ_w^2 et σ_b^2 représentent respectivement les estimations de la variance intra-établissement et inter-établissements.

Dans les modèles multiniveau, les résultats en général, et l'estimation de la variance inter-établissements en particulier, dépendent de la façon dont les établissements sont définis et organisés dans les pays, ainsi que des unités choisies pour prélever les échantillons. Dans certains pays, les établissements ont été échantillonnés comme unités administratives (même s'ils comptent plusieurs implantations différentes, comme en Italie), alors que dans d'autres, les établissements ont été échantillonnés comme des composantes de groupes scolaires plus larges qui accueillent des jeunes âgés de 15 ans, comme des bâtiments scolaires ou encore comme des entités administratives (dirigées par un chef d'établissement). Le rapport technique sur l'enquête PISA 2015 (*PISA Technical Report* [OCDE, à paraître en anglais uniquement]) et l'annexe A2 expliquent la façon dont les établissements sont définis. En Slovénie, par exemple, l'unité primaire d'échantillonnage est définie comme un groupe d'élèves qui suivent le même programme de cours dans un établissement (une filière dans un établissement). Dans ce cas particulier, la variance inter-établissements correspond à la variance intra-établissement entre filières. L'utilisation de variables de stratification dans la sélection des établissements peut aussi influer sur l'estimation de la variance inter-établissements, en particulier si les variables de stratification sont associées à des différences inter-établissements.

Par ailleurs, la variance intra-établissement comprend la variation de la performance entre classes et entre élèves en raison du mode d'échantillonnage des élèves.

Ampleur de l'effet

Il est parfois utile de comparer pour un indice donné les différences existantes entre des sous-groupes d'élèves (les garçons et les filles, par exemple) parmi les pays. Dans ce type de cas, la variation de la répartition de l'indice entre les groupes et les pays peut poser problème. L'une des solutions est de calculer l'ampleur de l'effet généré par les différences de répartition. L'ampleur de l'effet mesure la différence entre, par exemple, l'efficacité perçue des filles et des garçons en sciences dans un pays donné, par rapport à la variation moyenne de l'efficacité perçue de tous les élèves en sciences dans ledit pays.

Conformément aux pratiques établies, le tableau I.3.6 définit comme faible une ampleur de l'effet inférieure à 0.20, comme moyenne une ampleur de l'effet de l'ordre de 0.50, et comme élevée une ampleur de l'effet supérieure à 0.80.

L'ampleur de l'effet entre deux sous-groupes est calculée comme suit :

$$\frac{m_1 - m_2}{\sqrt{\sigma^2}}$$

ou m_1 et m_2 représentent respectivement les valeurs moyennes des sous-groupes 1 et 2, et la valeur σ^2 représente la variance globale (entre les groupes et au sein même de ceux-ci).

Indices de concentration

Indice de concentration réelle

L'indice de concentration réelle d'élèves issus de l'immigration scolarisés dans chaque pays/économie (ou indice de concentration réelle) correspond à la proportion minimale d'élèves, qu'ils soient ou non issus de l'immigration, qui devraient changer d'établissement si tous les établissements devaient accueillir une proportion identique d'élèves issus de l'immigration et, par conséquent, une proportion identique d'élèves autochtones. Cet indice se définit comme suit :

$$CC = \frac{\sum_{i=1}^{I} N_i \, | p_i - p |}{N}$$

où N_i équivaut au nombre d'élèves dans l'établissement *i*, *N* équivaut au nombre d'élèves dans la population et *I* équivaut au nombre d'établissements. La valeur $p_i = A_i/N_i$ représente la proportion d'élèves issus de l'immigration qui sont scolarisés dans l'établissement *i*, et la valeur $p = A/N$ la proportion d'élèves issus de l'immigration au sein de la population.

L'indice de concentration réelle *S* est corrélé à l'indice de ségrégation élaboré par Gorard et Taylor (2002), qui correspond au pourcentage d'élèves issus de l'immigration qui devraient changer d'établissement si tous les établissements accueillaient la même proportion d'élèves issus de l'immigration, compte tenu de leur taille initiale. L'indice de ségrégation de Gorard et Taylor est défini comme suit :

$$S = 0.5 \times \sum_{i=1}^{I} \left| \frac{A_i}{A} - \frac{N_i}{N} \right|$$

L'indice de concentration réelle peut se déduire directement de l'indice de ségrégation : $CC = 2pS$. L'indice de ségrégation de Gorard et Taylor dépend fortement du pourcentage d'élèves issus de l'immigration au sein de la population. Si les élèves issus de l'immigration sont très peu nombreux au sein d'un pays et s'ils sont principalement scolarisés dans des écoles internationales, le pourcentage d'élèves qui devraient changer d'établissement serait alors proche de 100 %. L'indice de concentration réelle est moins sensible à ce cas extrême, mais reste sensible au pourcentage global d'élèves issus de l'immigration au sein de la population.

Lorsque l'indice de concentration réelle est calculé pour un échantillon représentatif, il est important de prendre en compte les pondérations et l'erreur d'échantillonnage. L'indice de concentration réelle peut également être écrit sous la forme d'une moyenne d'élèves :

$$\frac{\sum_{i=1}^{I} N_i \,|\, p_i - p|}{N} = \frac{1}{N}\sum_{i=1}^{I} N_i \,|\, p_i - p| = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{N_i} |p_i - p| \cdot$$

Il peut donc être facilement généralisé à des échantillons pondérés, en remplaçant simplement la dernière expression par une moyenne pondérée :

$$\frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i} w_{ij}|p_i - p|}{\sum_{i=1}^{I}\sum_{j=1}^{n_i} w_{ij}} .$$

L'indice de concentration réelle peut alors être mesuré au niveau des élèves comme la différence absolue entre le pourcentage d'élèves issus de l'immigration dans un établissement donné et le pourcentage d'élèves issus de l'immigration dans le pays pondéré en fonction des pondérations finales des élèves. Pour cet indice, les erreurs-type sont obtenues en remplaçant les pondérations finales par le poids des 80 échantillons répétés dans le calcul.

Indice de concentration maximale

L'indice de concentration maximale représente la valeur maximale théorique de la concentration d'élèves issus de l'immigration dans un établissement donné, en fonction de la taille de cet établissement et du nombre d'élèves issus de l'immigration dans le pays. Cet indice correspond à la proportion minimale d'élèves, qu'ils soient ou non issus de l'immigration, qui devraient changer d'établissement si tous les établissements devaient accueillir une proportion identique d'élèves issus de l'immigration, dans la situation contrefactuelle où les élèves issus de l'immigration seraient scolarisés dès le départ dans les établissements les plus grands. Dans ce scénario hypothétique, la concentration d'élèves issus de l'immigration est maximale en ce sens que ces élèves sont uniquement scolarisés dans un nombre restreint d'établissements (compte tenu de la taille des établissements et de la population issue de l'immigration).

Le calcul de l'indice nécessite tout d'abord de classer par ordre décroissant les établissements au sein de chaque pays en fonction de leurs pondérations d'établissements respectives (calculées en effectuant la somme des pondérations finales des élèves dans un établissement donné). Dans un deuxième temps, tous les élèves issus de l'immigration sont répartis dans les établissements en fonction de ce classement, jusqu'à obtenir la somme pondérée d'élèves issus de l'immigration dans le pays donné. L'indice de concentration défini ci-dessus est ensuite calculé. Pour cet indice, les erreurs-type sont obtenues en remplaçant dans le calcul les pondérations finales par le poids des 80 échantillons répétés.

Définition des établissements à faible ou forte concentration

La classification des établissements présentant une concentration faible ou forte d'élèves issus de l'immigration est établie sur la base d'un seuil spécifique à chaque pays/économie. Ainsi, le nombre d'établissements à faible ou forte concentration ne dépend pas de la proportion d'élèves issus de l'immigration au sein de chaque système d'éducation. Le seuil est défini comme la répartition médiane (pondérée) du pourcentage d'élèves issus de l'immigration dans les établissements. Dans chaque pays, environ 50 % des élèves sont scolarisés dans des établissements à forte concentration, et 50 % des élèves sont scolarisés dans des établissements à faible concentration.

Erreurs-types et tests de signification

Les données statistiques présentées dans ce rapport correspondent à des estimations de la performance nationale réalisées sur la base d'échantillons d'élèves, et non à des valeurs qui auraient pu être calculées si tous les élèves de chaque pays avaient répondu à toutes les questions. Par conséquent, il importe de connaître le degré d'incertitude inhérent à ces estimations. Dans l'enquête PISA, chaque estimation est associée à un degré d'incertitude exprimé sous la forme d'une erreur-type. Le recours aux intervalles de confiance permet d'établir des inférences à propos des moyennes et des pourcentages d'une population d'une manière qui reflète l'incertitude associée aux estimations calculées sur la base d'échantillons. À partir d'une donnée statistique obtenue au moyen d'un échantillon et dans l'hypothèse d'une répartition normale, il est possible d'affirmer que le résultat

correspondant à l'échelle de la population se situe dans l'intervalle de confiance dans 95 cas sur 100 de la même mesure dans différents échantillons prélevés dans la même population.

Très souvent, le lecteur s'intéresse principalement aux écarts entre différentes valeurs au sein d'un même pays (le score des filles et des garçons, par exemple) ou entre plusieurs pays. Dans les tableaux et graphiques présentés dans ce rapport, les écarts sont déclarés statistiquement significatifs si des écarts de cette taille, plus petits ou plus grands, en valeur absolue s'observent dans moins de 5 % des cas en l'absence d'écarts réels dans les valeurs correspondantes.

Des tests de signification ont été réalisés pour évaluer la signification statistique des comparaisons présentées dans ce rapport.

Différences entre les sexes et différences entre des moyennes de sous-groupes

La signification statistique des différences de performance ou d'indice observées entre les sexes a été vérifiée. Les différences sont favorables aux garçons lorsqu'elles sont positives et favorables aux filles lorsqu'elles sont négatives. En règle générale, les différences indiquées en gras dans les tableaux du présent volume sont statistiquement significatives à un niveau de confiance de 95 %.

De même, la signification statistique des différences entre d'autres groupes d'élèves (par exemple, entre les élèves autochtones et les élèves issus de l'immigration) ou entre d'autres catégories d'établissements (par exemple, entre les établissements favorisés et les établissements défavorisés) a été vérifiée. Les sous-groupes sont généralement définis dans les tableaux ainsi que dans les textes accompagnant les analyses. Les établissements sont dits défavorisés sur le plan socio-économique s'ils se situent dans le quartile inférieur de la répartition de l'indice moyen PISA de statut économique, social et culturel (SESC) dans les établissements au sein de chaque pays et économie ; ils sont dits favorisés s'ils se situent dans le quartile supérieur de la répartition de cet indice. Toutes les différences indiquées en gras dans les tableaux de l'annexe B du présent rapport sont statistiquement significatives à un niveau de confiance de 95 %.

Différences de moyenne entre sous-groupes, après contrôle d'autres variables

Dans de nombreux tableaux, des comparaisons de valeurs sont faites entre sous-groupes « avant contrôle » et « après contrôle » d'autres variables telles que l'indice PISA de statut économique, social et culturel. Les différences ajustées sont estimées à l'aide d'une régression linéaire et leur signification est vérifiée à un niveau de confiance de 95 %. Les différences significatives sont indiquées en gras.

Différences de valeur entre les quartiles inférieur et supérieur des indices et échelles PISA

La signification statistique des différences de valeur d'indice ou d'échelle PISA entre le quartile supérieur et le quartile inférieur a été vérifiée. Les différences de valeur entre le quartile supérieur et le quartile inférieur des échelles de compétence ou d'indices à l'étude sont indiquées en gras si elles sont statistiquement significatives à un niveau de confiance de 95 %.

Variation de la performance associée à la variation de 1 unité de l'indice

Dans de nombreux tableaux, la variation de la performance associée à la variation de 1 unité de l'indice à l'étude est indiquée. Les différences indiquées en gras s'écartent de 0 dans une mesure statistiquement significative à un niveau de confiance de 95 %.

Risque relatif et rapport de cotes

Les chiffres en gras dans les tableaux présentés à l'annexe B de ce rapport indiquent que le risque relatif/rapport de cotes s'écarte de 1 dans une mesure statistiquement significative, à un niveau de confiance de 95 %. Pour calculer la signification statistique autour de la valeur de 1 (l'hypothèse nulle), la statistique de risque relatif/rapport de cotes est supposée suivre une distribution logarithmique normale, plutôt qu'une distribution normale, en vertu de l'hypothèse nulle.

Pour de nombreux tableaux, des rapports de cotes « généralisés » (après contrôle d'autres variables) sont également présentés. Ces rapports de cotes ont été estimés à l'aide d'une régression logistique et leur signification a été vérifiée en fonction de l'hypothèse nulle d'un rapport de cotes égal à 1 (c'est-à-dire dans le cas où la somme des probabilités de tous les événements serait égale à 1 après contrôle d'autres variables).

Plage de classement

Pour calculer la plage de classement des pays, des simulations sont effectuées sur la base des moyennes et des écarts-types des pays afin de générer des valeurs possibles. Quelque 10 000 simulations sont effectuées, et 10 000 classements possibles sont produits pour chaque pays sur la base de ces valeurs. Les chiffres de chaque position dans le classement sont agrégés du plus grand au plus petit jusqu'à ce qu'ils soient égaux ou supérieurs à 9 500. La plage de classement de chaque pays, y compris les positions agrégées, est ensuite indiquée. Cela signifie que le niveau de confiance associé aux plages de classement s'établit à 95 % au moins et que l'on peut considérer sans risque que la distribution des positions est unimodale. Cette méthode utilisée lors l'enquête PISA 2015 l'a été lors de chaque évaluation depuis PISA 2003.

La différence principale entre les plages de classement des pays (voir, par exemple, le graphique I.2.14) et les comparaisons de scores moyens entre les pays (voir, par exemple, le graphique I.2.13) réside dans le fait que, contrairement aux comparaisons de scores, les plages tiennent compte des multiples comparaisons inhérentes au classement des pays/économies. C'est pourquoi il arrive que de légères différences s'observent entre la plage de classement et le nombre de pays se classant au-dessus d'un pays donné, sur la base des comparaisons des performances entre les pays, deux par deux. Ainsi, l'entité P-S-J-G (Chine) et la Corée ont obtenu le même score moyen et se situent dans le même groupe de pays dont le score moyen ne s'écarte pas du leur dans une mesure statistiquement significative (voir le graphique I.2.13). Toutefois, parmi les pays de l'OCDE, la Corée se classe entre la 9e et la 14e position, tandis que l'entité P-S-J-G (Chine) se situe dans une plage de classement plus vaste (entre la 8e et la 16e position), avec un niveau de confiance de 95 % (voir le graphique I.2.14). Comme on peut considérer sans risque que la distribution des estimations de position de chaque pays est unimodale, il y a lieu d'utiliser cette plage lors de l'examen du classement des pays.

Erreurs-types dans les estimations statistiques dérivées de modèles multiniveau

Concernant les estimations statistiques dérivées de modèles multiniveau (par exemple, les estimations relatives aux composantes de la variance et les coefficients calculés au moyen de modèles de régression à deux niveaux), les erreurs-types ne sont pas estimées à l'aide de la méthode habituelle de réplication, qui tient compte de la stratification et des taux d'échantillonnage de populations finies. Dans ce cas, les erreurs-types sont estimées sur la base d'un modèle : elles sont calculées dans l'hypothèse où les établissements et les élèves au sein de ceux-ci sont échantillonnés de manière aléatoire (les probabilités d'échantillonnage étant reflétées dans les pondérations des établissements et des élèves) dans une population théorique infinie d'établissements et d'élèves, conforme aux hypothèses paramétriques du modèle.

L'erreur-type de l'estimation de l'indice d'inclusion est calculée comme suit : une distribution approximative est dérivée des erreurs-types (déduites par modélisation) des composantes de la variance à l'aide de la « méthode delta ».

Erreurs-types dans les analyses sur l'évolution de la performance : erreur d'ancrage

Les erreurs-type dans les comparaisons de la performance au fil du temps tiennent compte de l'incertitude inhérente à la procédure de correspondance qui permet de comparer sur une même échelle les scores entre les différentes évaluations PISA. Cette source d'incertitude supplémentaire suscite des erreurs-types plus importantes (par rapport aux erreurs-types estimées avant l'inclusion de l'erreur d'ancrage) (voir la description technique de l'erreur d'ancrage à l'annexe A5).

Les chiffres en gras dans les tableaux sur l'évolution de la performance au fil du temps présentés à l'annexe B du présent rapport indiquent que la variation de la performance dans un groupe donné s'écarte de 0 dans une mesure statistiquement significative, à un niveau de confiance de 95 %. Les erreurs-types utilisées pour calculer la signification statistique de l'évolution indiquée incluent l'erreur d'ancrage.

Références

Gorard, S. et **C. Taylor** (2002), « What is segregation? A comparison of measures in terms of 'strong' and 'weak' compositional invariance », *Sociology*, vol.36/4, pp. 875-895, http://dx.doi.org/ 10.1177/003803850203600405.

OCDE (à paraître en anglais uniquement), *PISA 2015 Technical Report*, PISA, Éditions OCDE, Paris.

ANNEXE A4
ASSURANCE QUALITÉ

Des procédures d'assurance qualité ont été appliquées tout au long de l'évaluation PISA 2015, comme lors des évaluations précédentes. Les normes techniques de l'évaluation PISA 2015 (www.oecd.org/pisa/) spécifient la façon dont l'évaluation PISA doit être mise en œuvre dans chaque pays, économie ou entité infranationale. Les contractants internationaux suivent la mise en œuvre de l'évaluation dans chacun des pays, économies et entités infranationales, et vérifient que les normes y sont respectées.

Pour faciliter le processus à mener afin de garantir la qualité et l'équivalence linguistique des instruments de l'évaluation PISA 2015, on a vérifié que les items ne posaient aucun problème de traduction à partir de la version anglaise. Deux versions sources des instruments d'évaluation, en anglais et en français, ont été préparées (à l'exception des épreuves de culture financière et des manuels d'opérations, disponibles uniquement en anglais). Les pays ont traduit les versions sources anglaise et/ou française selon la méthode de la double traduction (la ou les versions sources sont traduites par deux personnes différentes, dont les deux traductions sont ensuite conciliées par une troisième personne). Des instructions détaillées de localisation des instruments (c'est-à-dire leur adaptation, leur traduction et leur validation) ont été fournies pour leur permettre de préparer les instruments de l'essai de terrain et les réviser en vue de la campagne définitive. Des directives de traduction et d'adaptation ont également été fournies. Une équipe indépendante de vérificateurs spécialisés, choisis et formés par le Consortium PISA, ont vérifié l'équivalence de chaque version nationale aux versions sources anglaise et/ou française. Ces vérificateurs sont des traducteurs qui ont pour langue maternelle la langue d'instruction du pays concerné et qui connaissent bien son système d'éducation. Pour plus d'informations sur les procédures de traduction, voir le rapport technique sur l'évaluation PISA 2015, *PISA 2015 Technical Report* (OCDE, à paraître en anglais uniquement).

L'évaluation a été mise en œuvre dans le respect de procédures normalisées. Le Consortium PISA a préparé des manuels détaillés expliquant le mode de mise en œuvre de l'évaluation, dans lesquels figuraient notamment des instructions précises concernant le travail des coordinateurs scolaires et des scripts que les administrateurs de test ont été priés de respecter durant les séances d'évaluation. Les propositions d'adaptation des procédures de mise en œuvre ou de modification du script des séances d'évaluation ont été soumises au Consortium PISA pour approbation. Le Consortium PISA a ensuite vérifié la traduction et l'adaptation de ces documents réalisées par chaque pays.

Dans le but d'établir la crédibilité de l'enquête PISA en tant qu'étude valide et non biaisée, et d'assurer autant que faire se peut l'application de procédures uniformes lors des séances d'évaluation, les administrateurs de test des pays et économies participants ont été recrutés en fonction des critères suivants : il a été exigé que l'administrateur de test ne soit pas le professeur de la langue d'évaluation, de mathématiques ou de sciences des élèves participant aux séances qu'il administrerait dans le cadre de l'enquête PISA ; il a été recommandé que l'administrateur de test ne soit pas membre du personnel de l'un des établissements dans lequel il administrerait des séances d'évaluation dans le cadre de l'enquête PISA ; et il a été jugé préférable que l'administrateur de test ne soit membre du personnel d'aucun des établissements constituant l'échantillon PISA. Les administrateurs de test ont assisté en personne aux séances de formation organisées à leur intention par les pays et économies participants.

Les pays et économies participants ont été priés de veiller à ce que les administrateurs de test préparent les séances d'évaluation en collaboration avec les coordinateurs scolaires. Les missions des administrateurs de test sont les suivantes : mettre à jour la liste de présence des élèves, c'est-à-dire un formulaire où sont consignés les élèves présents et les épreuves qui leur sont attribuées ; établir le rapport de séance, un formulaire où sont consignés l'horaire des séances, les événements qui les ont perturbées le cas échant, etc. ; vérifier que le nombre de carnets de test et de questionnaires remplis par les élèves correspond au nombre envoyé à l'établissement (dans les pays qui ont administré les instruments sur papier) ou que le nombre de clés USB correspond au nombre prévu (dans les pays qui ont administré les instruments sur ordinateur) ; faire parvenir les questionnaires « Établissement » et « Parents » aux intéressés (le cas échéant) ; et renvoyer tout le matériel de test (y compris les instruments vierges) au Centre national à l'issue des séances.

Le Consortium PISA, qui est responsable de la supervision des opérations, s'est chargé de toutes les phases du processus PISA de contrôle de la qualité : sélection des contrôleurs de la qualité dans chaque pays après entretien avec les candidats, organisation de la formation des contrôleurs de la qualité, sélection des établissements à inspecter,

et collecte des informations d'inspection. Les contrôleurs de la qualité sont des contractants indépendants basés dans les pays participants qui sont recrutés par le contractant international chargé des opérations. Ils se rendent dans un certain nombre d'établissements pour observer l'administration des épreuves et rendre compte du respect des procédures d'administration lors de la campagne définitive.

En règle générale, deux ou trois contrôleurs ont été recrutés dans chaque pays, où ils ont effectué une visite d'inspection dans une quinzaine d'établissements en moyenne. Des contrôleurs supplémentaires ont dû être recrutés dans les pays où des entités infranationales ont participé à l'évaluation PISA, puisqu'il faut un minimum de cinq établissements contrôlés pour que les données soient adjugées.

Toutes les données d'assurance de la qualité recueillies lors de l'évaluation PISA 2015 ont été saisies dans une base de données centrale d'adjudication portant sur la qualité des opérations sur le terrain, de la traduction et de l'impression des instruments, de l'échantillonnage des établissements et des élèves et du codage. Des rapports exhaustifs ont ensuite été rédigés à l'intention du Groupe d'adjudication PISA. Ce groupe a été constitué par le Groupe consultatif technique et le référent pour l'échantillonnage. Il a pour mission de passer en revue la base de données d'adjudication et de formuler des recommandations de traitement pour préserver la qualité des données PISA. Pour plus d'informations, voir le rapport technique sur l'évaluation PISA 2015, *PISA 2015 Technical Report* (OCDE, à paraître en anglais uniquement).

Les résultats de l'adjudication et des analyses plus approfondies menées par la suite confirment que les normes techniques PISA ont été respectées dans tous les pays et économies qui ont participé à l'évaluation PISA 2015, sauf dans ceux ci-dessous :

- En Albanie, l'évaluation PISA a été administrée dans le respect des normes et directives de l'OCDE. Toutefois, les résultats des élèves aux épreuves cognitives n'ont pu être mis en correspondance avec les informations qu'ils ont fournies dans le questionnaire contextuel à cause de la façon dont les données ont été saisies. C'est pourquoi l'Albanie n'a pu être incluse dans les analyses où les réponses des élèves au questionnaire sont mises en relation avec leurs résultats aux épreuves cognitives.
- En Argentine, l'évaluation PISA a été administrée dans le respect des normes et directives de l'OCDE. Toutefois, le pourcentage d'individus de 15 ans représentés dans l'échantillon a considérablement diminué. La restructuration des établissements d'enseignement secondaire qui a eu lieu en Argentine, mais pas dans la Ciudad Autónoma de Buenos Aires (dont les données ont été adjugées), a vraisemblablement influé sur le plan d'échantillonnage des établissements. En conséquence, les résultats de l'Argentine ne sont pas nécessairement comparables aux résultats des autres pays, ni à ses résultats lors d'évaluations antérieures.
- Au Kazakhstan, il est apparu que les correcteurs avaient été indulgents lors du codage. En conséquence, les données des items qu'ils ont corrigés ont été exclues de la base de données internationale pour cause de non-respect des normes PISA. Comme les items corrigés par les correcteurs constituent une part importante des constructs PISA, l'exclusion de ces items a significativement réduit la portée des épreuves PISA. Il s'ensuit que les résultats du Kazakhstan ne sont pas nécessairement comparables aux résultats des autres pays, ni à ses résultats lors d'évaluations antérieures.
- En Malaisie, l'évaluation PISA a été administrée dans le respect des normes et directives de l'OCDE. Toutefois, le taux de réponse pondéré des établissements initialement échantillonnés (51 %) est inférieur au taux de réponse spécifié dans les normes PISA (85 %). Il s'ensuit que les résultats de la Malaisie ne sont pas nécessairement comparables aux résultats des autres pays, ni à ses résultats lors d'évaluations antérieures.

Référence

OCDE (à paraître), *PISA 2015 Technical Report*, PISA, Éditions OCDE, Paris.

ANNEXE A5
CHANGEMENTS DE MODE D'ADMINISTRATION ET DE PROCÉDURE DE MISE À L'ÉCHELLE LORS DE L'ÉVALUATION PISA 2015 ET IMPLICATIONS POUR LES ANALYSES DE TENDANCES

Comparaison des performances en mathématiques, en compréhension de l'écrit et en sciences entre les évaluations PISA

La même échelle de culture scientifique a été utilisée lors des évaluations PISA de 2006, de 2009, de 2012 et de 2015, de sorte que les scores rapportés sur cette échelle sont directement comparables dans le temps. Il en va de même pour l'échelle de compréhension de l'écrit, qui est la même depuis l'évaluation PISA 2000 et, pour l'échelle de culture mathématique, qui est la même depuis l'évaluation PISA 2003. Il est possible de comparer les scores dans le temps parce que certains items sont administrés lors de chaque évaluation et que les échelles de compétence dérivées de paramètres d'items différents sont alignées grâce à une procédure d'étalonnage.

Toutes les estimations statistiques sont associées à une incertitude statistique, ce qui vaut aussi pour les paramètres utilisés pour étalonner les échelles PISA successives. L'erreur d'ancrage, qui reflète cette incertitude, est incluse dans l'erreur-type des estimations des tendances d'évolution de la performance PISA. (Pour plus de détails sur les erreurs d'ancrage, voir les sections ci-dessous.)

L'incertitude associée à l'étalonnage des échelles est le résultat des changements introduits au fil du temps dans la façon dont les épreuves sont administrées (les différences de conception des épreuves, par exemple) et mises à l'échelle (les différences liées aux échantillons d'étalonnage, par exemple). Elle reflète aussi la nature évolutive des cadres d'évaluation. L'enquête PISA revisite les cadres d'évaluation de la culture scientifique, de la compréhension de l'écrit et de la culture mathématique, tour à tour, tous les neuf ans, pour y répercuter les dernières avancées de la recherche sur les connaissances et compétences qu'il est important que les jeunes de 15 ans acquièrent pour participer pleinement à la vie des sociétés de demain.

Les changements introduits dans la conception et l'administration des épreuves peuvent influer dans une certaine mesure sur la façon dont les élèves répondent aux items. Les estimations de la difficulté des items sont différentes à cause des changements introduits dans les échantillons et les modèles de mise à l'échelle. Par voie de conséquence, le fait de rapporter les résultats d'une évaluation sur l'échelle de compétence d'une évaluation précédente est entouré d'incertitude. Avant 2015, toutes les évaluations PISA se sont différenciées les unes des autres à trois égards, à savoir :

- *La conception des épreuves*[1]. La façon dont les épreuves sont conçues peut influer à divers égards sur la façon dont les élèves y répondent. Les élèves peuvent par exemple estimer que le même item de compréhension de l'écrit n'est pas aussi difficile s'il est présenté au début des épreuves, comme lors de l'évaluation PISA 2000, que s'il est présenté à d'autres endroits des épreuves, comme lors d'évaluations ultérieures. De même, ils risquent de ne pas déployer autant d'efforts si l'épreuve de compréhension de l'écrit de 30 minutes est programmée au beau milieu d'épreuves de mathématiques et de sciences, ce qui est le plus souvent le cas lorsque la compréhension de l'écrit est un domaine mineur (comme en 2003, 2006 et 2012), et non un domaine majeur. Dans l'enquête PISA, ces effets ne sont pas systématiques et sont le plus souvent minimes, mais ils contribuent à l'incertitude entourant les estimations.
- *Les échantillons d'étalonnage*. Avant 2015, la difficulté des items PISA était estimée sur la seule base des réponses des élèves ayant passé les dernières épreuves en date. C'est un sous-groupe aléatoire de 500 élèves dans chaque pays et économie qui a été utilisé comme échantillon d'étalonnage lors des évaluations PISA 2009 et PISA 2012. Quant à l'échantillon d'étalonnage utilisé lors des évaluations PISA 2000, PISA 2003 et PISA 2006, il se limitait aux pays de l'OCDE (500 élèves par pays) (OCDE, 2009). Cela implique qu'à chaque fois que des items d'ancrage sont administrés, leur degré de difficulté est estimé ; il y a donc autant d'estimations (indépendantes) de leur degré de difficulté qu'il y a d'évaluations. Ces estimations du degré de difficulté sont différentes, et leur variabilité contribue à l'incertitude des comparaisons entre les évaluations PISA. Le fait de n'utiliser qu'un sous-échantillon des données PISA par pays accroît cette incertitude, mais c'était la seule manière de procéder étant donné la limitation de la puissance de calcul à l'époque des premières évaluations PISA.
- *Le nombre d'items d'ancrage communs aux évaluations*. Inclure davantage d'établissements et d'élèves dans l'échantillon permet de réduire l'incertitude entourant le score moyen des pays et les paramètres d'items, il en va de même pour l'incertitude entourant l'équivalence entre les échelles successives, qui diminue plus les items déjà administrés sont nombreux. Dans le domaine majeur (la culture scientifique, en 2015), les épreuves sont en partie constitués d'items communs à des évaluations précédentes, mais pas entièrement, car la batterie PISA d'items est progressivement renouvelée pour refléter les changements introduits dans les cadres d'évaluation, qui sont revus en fonction des dernières avancées de la recherche sur les connaissances et compétences qu'il est important que les jeunes de 15 ans acquièrent pour participer pleinement à la vie de la société.

Lors de l'évaluation PISA 2015, la conception des épreuves et la procédure de mise à l'échelle ont été améliorées à plusieurs égards pour réduire les trois sources d'incertitude évoquées ci-dessus. Dans la conception des épreuves PISA de 2015, on a en

particulier réduit, voire éliminé, la différence de couverture des *constructs* entre les domaines et la différence susceptible de découler du fait que des élèves considèrent que des domaines sont « majeurs » ou « mineurs ». Dans la version la plus souvent administrée des épreuves (la version informatisée dans les pays qui ont évalué la résolution collaborative de problèmes), par exemple, 86 % des élèves ont passé des épreuves dans deux domaines seulement, à raison d'une heure par domaine (33 % en sciences et en compréhension de l'écrit, 33 % en sciences et en mathématiques, et 22 % en sciences et en résolution collaborative de problèmes, avec une inversion des épreuves dans la moitié de chaque groupe) (voir OCDE [à paraître], pour plus de détails). Le nombre d'items déjà administrés lors d'évaluations précédentes a considérablement augmenté dans tous les domaines, mais plus encore dans les domaines mineurs. Par exemple, lorsque la compréhension de l'écrit était un domaine mineur (en 2003 et en 2006), un nombre d'items équivalant à 1 heure de test, soit 2 blocs de 30 minutes, a été utilisé pour étayer le lien avec l'évaluation PISA 2000 ; lorsque les mathématiques ont pour la deuxième fois été le domaine majeur, en 2012, un nombre d'items d'ancrage équivalant à 1.5 heure de test a été utilisé pour étayer le lien avec l'évaluation de 2003. En 2015, l'équivalent de trois heures de test a été utilisé en sciences (le domaine majeur), en compréhension de l'écrit et en mathématiques pour étayer le lien avec les échelles existantes.

Pour améliorer la procédure de mise à l'échelle, l'échantillon d'étalonnage a été basé sur toutes les réponses des élèves aux quatre dernières évaluations, dont une évaluation où chaque domaine a été le domaine majeur. Dans le domaine majeur, l'échantillon remonte à l'évaluation où il était déjà domaine majeur. Lors de la prochaine évaluation PISA (2018), le taux de couverture de l'échantillon d'étalonnage atteindra 75 % avec l'évaluation PISA 2015. Par voie de conséquence, l'incertitude associée à la ré-estimation des paramètres d'items lors de la mise à l'échelle sera nettement moindre, par comparaison aux évaluations d'avant 2015.

Ces améliorations devraient contribuer à réduire l'erreur d'ancrage entre l'évaluation de 2015 et les évaluations ultérieures, mais elles pourraient accroître l'erreur d'ancrage entre l'évaluation de 2015 et les évaluations précédentes, dont la conception des épreuves et la procédure de mise à l'échelle étaient différentes.

De plus, d'autres changements ont été introduits dans l'administration des épreuves et la procédure de mise à l'échelle à l'occasion de l'évaluation PISA 2015 :

- *Changement de mode d'administration*. Les épreuves ont en majorité été administrées sur ordinateur lors de l'évaluation PISA 2015. Tous les items d'ancrage utilisés lors de l'évaluation PISA 2015 ont été adaptés pour être administrés sur ordinateur. L'équivalence entre les items d'ancrage sur papier et sur ordinateur utilisés pour évaluer la performance des élèves en sciences, en compréhension de l'écrit et en mathématiques a été analysée dans un effectif diversifié d'élèves de tous les pays et économies participant à l'évaluation PISA 2015 dans le cadre d'un essai de terrain à grande échelle mené dans tous les pays et économies participants. Les résultats de cette étude sur l'effet du mode d'administration, concernant le degré d'équivalence des items (équivalence « scalaire » ou « métrique » ; voir par exemple Davidov, Schmidt et Billiet, 2011 ; Meredith, 1993), ont été pris en considération lors de la mise à l'échelle des réponses des élèves lors de la campagne définitive. Les paramètres des items dont l'invariance est scalaire ou métrique sont identiques dans l'ensemble de l'échantillon d'étalonnage, que les élèves aient passé les épreuves sur papier ou sur ordinateur (pour plus de détails, voir la section « Comparaison des résultats entre les épreuves PISA sur papier et sur ordinateur »).
- *Changement dans le modèle de mise à l'échelle*. Un modèle statistique plus souple a été utilisé lors de la mise à l'échelle des paramètres d'items. Ce modèle, dont la forme la plus large est le modèle de crédit partiel généralisé (c'est-à-dire un modèle à deux paramètres basé sur la théorie de réponse à l'item ; voir Birnbaum, 1968 ; Muraki, 1992), prévoit de retenir autant d'items d'ancrage avec des fonctions de probabilité à un paramètre que le permettent les données ; c'est la raison pour laquelle il est dit « hybride ». Les modèles à un paramètre utilisés lors de la mise à l'échelle des évaluations précédentes (Masters, 1982 ; Rasch, 1960) sont des variantes spéciales du modèle de 2015. La principale différence entre le modèle hybride et les modèles à un paramètre utilisés par le passé réside dans le fait que le score est calculé non selon une pondération égale de tous les items, mais selon une pondération optimale des items, sur la base de leur capacité à faire la distinction entre les élèves peu et très performants. Ce modèle est donc plus adapté vu la diversité des formats de réponse dans les épreuves PISA.
- *Changement dans le traitement du fonctionnement différentiel des items (Differential Item Functioning, DIF) entre les pays*. Dans des tests tels que les épreuves PISA, dont les items sont traduits dans de nombreuses langues, il arrive que des items ne fonctionnent pas de la même manière dans certains pays. Par exemple, il n'est pas toujours possible d'éviter des termes plus difficiles à traduire dans certaines langues que dans d'autres. Les interactions items-pays qui en résultent peuvent menacer la validité des résultats. Lors des évaluations précédentes, les paramètres d'items étaient identiques partout, à l'exception de quelques rares items considérés comme « douteux » qui avaient été déclarés « non administrés » dans certains pays (à cause par exemple d'erreurs de traduction ou de problèmes d'impression qui n'avaient été détectés qu'à un stade avancé du processus). En 2015, des écarts ont été autorisés par rapport aux paramètres internationaux dans un nombre (limité) d'items lors de l'étalonnage (Glas et Jehangir, 2014 ; Oliveri et von Davier, 2014 ; Oliveri et von Davier, 2011)[2]. Cette approche préserve la comparabilité des scores PISA entre les pays et au fil du temps, grâce à l'utilisation d'un nombre suffisant d'items invariants et à la réduction de la dépendance (limitée) du classement des pays aux items choisis pour constituer les épreuves, ce qui permet de mettre les pays sur un pied d'égalité. Le rapport technique sur l'évaluation PISA 2015 indique le nombre de paramètres spécifiques à chacun des pays et économies qui ont participé à l'évaluation PISA (OCDE, à paraître).

- *Changement dans le traitement des items non atteints.* Enfin, lors de l'évaluation PISA 2015, les items non atteints (c'est-à-dire ceux auxquels les élèves n'ont pas répondu à la fin des épreuves) ont été traités comme non administrés, alors que lors des évaluations précédentes, leur traitement était différent entre l'estimation des scores (le « codage »), où l'on considérait que les élèves n'y avaient pas répondu correctement, et l'estimation des paramètres d'items (la « mise à l'échelle »), où l'on considérait qu'ils n'avaient pas été administrés. Grâce à ce changement, le traitement des réponses d'élèves est le même lors de l'estimation des paramètres d'items et de l'estimation du niveau de compétence des élèves, et le risque d'avantager des pays où les élèves ont répondu au hasard à des items à choix multiple qu'ils n'ont pu terminer à temps par rapport aux pays où les élèves n'ont pas répondu à ces items a été supprimé[3]. Toutefois, ce nouveau traitement des items non atteints pourrait donner lieu à des scores plus élevés que ceux qui auraient été estimés par le passé dans les pays où le nombre d'items non atteints était important.

Présentation des résultats de l'évaluation PISA 2015 sur les échelles de compétence existantes

Cette section décrit la procédure adoptée pour transformer les résultats de l'évaluation PISA 2015 et les rapporter sur les échelles PISA définies lors de l'évaluation PISA 2000 en compréhension de l'écrit, lors de l'évaluation PISA 2003 en mathématiques et lors de l'évaluation PISA 2006 en sciences.

Dans l'estimation des paramètres d'items en 2015, basée sur les réponses des élèves aux épreuves de 2006, de 2009, de 2012 et de 2015, on a considéré que ces réponses provenaient de M populations distinctes, où M est le nombre total de pays et économies participant à l'enquête PISA multiplié par le nombre d'évaluations auxquelles ils ont participé (modèle multigroupe). Chaque population m_{ij} (où i identifie le pays et j identifie l'évaluation) se caractérise par un certain score moyen et une certaine variation du niveau de compétence[4]. Les scores moyens et les écarts-types font partie des paramètres estimés par le modèle de mise à l'échelle, comme les paramètres d'items. (Lors des évaluations précédentes, les estimations des scores étaient imputées dans une deuxième étape, distincte dans chacun des pays et économies. Cette étape de « codage » n'a été faite que lors de l'évaluation PISA 2015.) Le résultat de l'étape de mise à l'échelle est une échelle d'ancrage, sur la base de l'hypothèse de l'invariance du fonctionnement des items entre les évaluations de 2006, de 2009, de 2012 et de 2015, sur laquelle les scores moyens et les écarts-types des pays sont directement comparables dans le temps.

Pour aligner l'échelle construite lors de cette étape sur l'échelle numérique utilisée pour rapporter les résultats des évaluations PISA précédentes, les résultats ont été transformés de manière linéaire. Les paramètres de pente et d'intercept utilisés dans ce processus de transformation ont été définis sur la base de la comparaison des scores moyens et des écarts-types des pays et économies, estimés lors de l'étape de mise à l'échelle dans l'échelle logit, avec les scores moyens et écarts-types correspondants obtenus lors des évaluations PISA précédentes et publiés dans les rapports sur les évaluations PISA. Plus précisément, les résultats de sciences ont été transformés sur la base de la comparaison du score moyen de l'OCDE et de l'écart-type (national) par rapport au score moyen de l'OCDE avec le score moyen de l'OCDE et l'écart-type (national) de 2006. Cette transformation préserve l'attribut de l'échelle PISA de culture scientifique, dont la moyenne a été fixée à 500 points et l'écart-type, à 100 points, dans les pays de l'OCDE, lors de la première évaluation dont le domaine majeur était la culture scientifique. Une procédure similaire a été utilisée en mathématiques (alignement des scores moyens et des écarts-types des pays de l'OCDE sur ceux de la dernière évaluation dont la culture mathématique était le domaine majeur, à savoir en 2012) et en compréhension de l'écrit (alignement des résultats ré-estimés sur les résultats publiés de 2009).

Évaluation de l'impact des changements introduits dans la méthode de mise à l'échelle en 2015 sur l'estimation des tendances d'évolution

Il est possible d'estimer le score moyen que les pays et économies auraient obtenus lors d'une évaluation précédente si, à l'époque, les réponses des élèves avaient été mises à l'échelle selon la méthode utilisée en 2015. Cette section compare les scores moyens publiés dans les rapports sur les évaluations PISA précédentes (par exemple OCDE, 2014a) et les scores moyens recalculés selon la méthode de mise à l'échelle utilisée en 2015.

Le tableau A5.1 montre les corrélations entre deux séries de scores moyens en 2006, 2009, 2012 et 2015 : ceux présentés dans les tableaux à l'annexe B et analysés dans ce rapport, et les scores moyens estimés sur la base des mêmes données, mais générés selon la méthode de mise à l'échelle de 2015, avec le modèle multigroupe décrit ci-dessus. Les différences de score moyen peuvent s'expliquer : par l'utilisation d'échantillons d'étalonnage plus importants que les séries de données de plusieurs évaluations ; par le nouveau traitement du fonctionnement différentiel des items entre les pays et des items non atteints ; ou par l'utilisation d'un modèle hybride de théorie de réponse à l'item au lieu des modèles à un paramètre retenus lors des évaluations précédentes. La colonne portant sur 2015 montre l'ampleur, négligeable, des différences dues à l'imputation des scores lors du codage.

Les coefficients de corrélation élevés indiqués dans ce tableau pour les années 2006, 2009 et 2012 (tous supérieurs à 0.993, sauf en compréhension de l'écrit en 2006, où le coefficient s'établit à 0.985) montrent que le rang des pays dans le classement PISA n'est guère affecté par les changements introduits en 2015 dans la méthode de mise à l'échelle. Les corrélations des estimations calculées selon diverses méthodes sont aussi d'une intensité plus forte que les corrélations des scores moyens des évaluations PISA successives, et nettement plus forte que les corrélations des scores moyens de deux évaluations dont le domaine majeur est le même (entre lesquelles l'intervalle est de neuf ans)[5]. Cela signifie que les changements méthodologiques peuvent au mieux expliquer une petite partie seulement de l'évolution de la performance PISA au fil du temps.

Tableau A5.1 Corrélation des moyennes nationales selon des méthodes alternatives de mise à l'échelle

Dans l'ensemble des pays/économies ayant participé à PISA 2015

	2006	2009	2012	2015
Sciences	0.9941	0.9961	0.9966	0.9997
Compréhension de l'écrit	0.9850	0.9949	0.9934	0.9992
Mathématiques	0.9953	0.9974	0.9973	0.9995

Remarque : Ce tableau présente le coefficient de corrélation entre les estimations moyennes indiquées à l'annexe B, sur la base de méthodes de mise à l'échelle spécifiques à chaque cycle, et les moyennes des répartitions ultérieures, obtenues selon la méthode de mise à l'échelle de 2015.

StatLink http://dx.doi.org/10.1787/888933433162

Comparaison des scores moyens selon la même méthode de mise à l'échelle

Une fois que les scores moyens des pays et économies générés lors de la mise à l'échelle des paramètres d'items sont transformés comme indiqué dans la section ci-dessus, ils peuvent être utilisés pour évaluer, dans chaque pays et économie, la sensibilité des tendances présentées dans le corps du rapport et dans les tableaux de l'annexe B aux changements apportés en 2015 à la méthode de mise à l'échelle et aux échantillons d'étalonnage[6]. Ces scores moyens transformés sont indiqués dans le tableau A5.1 en sciences, en compréhension de l'écrit et en mathématiques.

Dans une grande majorité des pays et économies, les différences entre les scores moyens indiqués à l'annexe B et les scores moyens indiqués dans les tableaux A5.3, A5.4 et A5.5 se situent largement dans l'intervalle de confiance associé à l'erreur d'ancrage (voir ci-dessous). Toutefois, certains pays et économies font figures d'exception (voir les graphiques A5.1, A5.2 et A5.3). L'analyse de l'évolution entre l'évaluation PISA 2015 et la dernière évaluation dont un domaine était le domaine majeur révèle quelques cas particuliers :

Sciences

- L'augmentation du score moyen en sciences constatée en Colombie est presque totalement imputable aux changements intervenus dans la méthode de mise à l'échelle. En effet, le score moyen n'aurait augmenté que de 3 points (soit une différence non significative) si les résultats de 2006 avaient été mis à l'échelle selon la méthode et l'échantillon d'étalonnage de 2015. Dans une moindre mesure, l'augmentation du score moyen constatée au Brésil, au Chili, en Indonésie et en Uruguay est également imputable aux changements intervenus dans l'échantillon d'étalonnage et la méthode de mise à l'échelle. La tendance n'aurait pas été aussi haussière (ni, très vraisemblablement, significative) si les scores moyens des évaluations précédentes avaient été générés selon la méthode de mise à l'échelle utilisée en 2015. Il n'est pas possible d'identifier avec certitude les différences entre la mise à l'échelle initiale des données de l'évaluation PISA 2006 et la nouvelle procédure de mise à l'échelle de l'évaluation PISA 2015 qui ont produit ces résultats. Il est toutefois probable que ces différences résident dans le nouveau traitement des items non atteints. Dans tous ces pays en effet, de nombreux élèves ne sont pas parvenus à répondre aux items en fin d'épreuves.
- Aux États-Unis, le score moyen en sciences a augmenté dans une mesure non significative (de 7 points) entre 2006 et 2015. Il aurait augmenté plus fortement, dans une mesure qui aurait très probablement été déclarée significative (+15 points), si les résultats de 2006 avaient été mis à l'échelle selon la méthode de mise à l'échelle et l'échantillon d'étalonnage de 2015. La variation est plus forte que celle constatée, certes, mais elle reste comprise dans l'intervalle de confiance même si la méthode de 2015 est utilisée.

Compréhension de l'écrit

- La diminution du score constatée en Corée entre les évaluations PISA 2009 et PISA 2015 (-22 points) est en grande partie due à la différence de méthode de mise à l'échelle. Si les résultats de compréhension de l'écrit de l'évaluation PISA 2009 avaient été mis à l'échelle selon la méthode utilisée lors de l'évaluation PISA 2015, la différence de score en Corée n'aurait été que de 9 points et aurait très vraisemblablement été déclarée non significative. Selon le modèle de mise à l'échelle de l'évaluation PISA 2015, les résultats antérieurs de la Corée sont quelque peu surestimés. Il n'est pas possible d'identifier avec certitude, sur la base de ces résultats, quel aspect de la méthode utilisée lors de l'évaluation PISA 2015 explique la différence. Toutefois, il est probable que le nouveau traitement du fonctionnement différentiel d'item soit en cause. En effet, la plupart des items accusant un niveau modéré de fonctionnement différentiel en Corée, c'est-à-dire ceux dont des paramètres spécifiques à la Corée ont été utilisés dans l'étalonnage de l'évaluation PISA 2015, sont des items auxquels les élèves coréens ont répondu correctement lors des évaluations précédentes dans une mesure supérieure à celle prévue par leurs paramètres internationaux. Cette tendance s'observe aussi, quoique dans une moindre mesure, en Thaïlande : la diminution constatée du score moyen (-12 points) aurait été déclarée non significative (-3 points) si la comparaison avait porté sur les résultats de 2009 après nouvelle mise à l'échelle.
- Au Danemark, le score moyen a augmenté dans une mesure non significative (de 5 points) entre les évaluations PISA 2009 et PISA 2015. Toutefois, sa progression aurait été déclarée significative (+15 points) si les résultats avaient été mis à l'échelle selon la méthode utilisée lors de l'évaluation PISA 2015.
- En Estonie, l'augmentation significative du score (18 points) n'aurait été que de 10 points si les résultats de l'évaluation PISA 2009 avaient été générés dans le modèle de mise à l'échelle utilisé lors de l'évaluation PISA 2015.
- Aux Pays-Bas, le score a diminué dans une mesure non significative (-5 points) entre les évaluations PISA 2009 et PISA 2015. Toutefois, il aurait augmenté de 4 points (un écart qui aurait très probablement été déclaré non significatif) si les résultats avaient été mis à l'échelle selon la méthode retenue lors de l'évaluation PISA 2015.

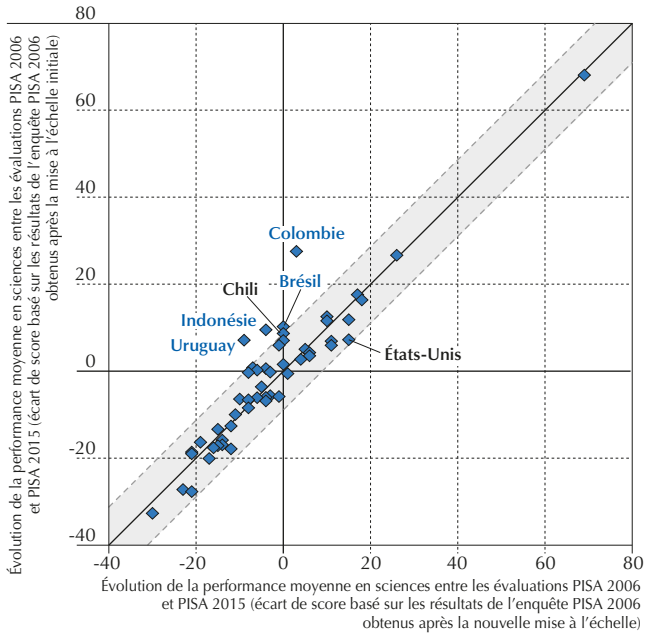
- L'augmentation du score moyen de compréhension de l'écrit constatée en Colombie, à Trinité-et-Tobago et en Uruguay est très vraisemblablement imputable aux changements intervenus dans la méthode de mise à l'échelle. Le score moyen aurait varié dans une mesure quasi nulle (et sa variation aurait été déclarée non significative) si les résultats de 2009 avaient été mis à l'échelle selon la méthode et l'échantillon d'étalonnage utilisés en 2015. De même, le score moyen n'aurait augmenté que de 21 points en Moldavie et de 15 points au Pérou (au lieu des 28 points annoncés) si la même méthode de mise à l'échelle avait été utilisée. Ces différences s'expliquent vraisemblablement par le nouveau traitement des items non atteints. Dans tous ces pays, de nombreux élèves ne sont pas arrivés aux items en fin d'épreuves.

Mathématiques

- La diminution du score constatée entre les évaluations PISA 2012 et PISA 2015 au Taipei chinois (-18 points) et au Viêtnam (-17 points) est en grande partie imputable à l'utilisation de méthodes différentes de mise à l'échelle. Si les résultats de mathématiques de l'évaluation PISA 2012 avaient été mis à l'échelle selon la méthode et l'échantillon d'étalonnage utilisés lors de l'évaluation PISA 2015, le score n'aurait diminué que de 3 points au Taipei chinois et de 4 points au Viêtnam, et ces écarts n'auraient très probablement pas été déclarés significatifs. Le nouveau traitement du fonctionnement différentiel d'item pourrait être la cause principale de ces différences.
- En Turquie, la diminution du score entre les évaluations PISA 2012 et PISA 2015 (-28 points) aurait été moindre (-18 points) si tous les résultats avaient été produits selon la méthode de mise à l'échelle utilisée en 2015. L'importance de la variation est amplifiée, mais sa direction et sa signification sont similaires dans les deux séries de données.
- En Albanie, l'augmentation du score en mathématiques entre les évaluations PISA 2012 et PISA 2015 (+19 points) aurait été moindre (+7 points) et aurait très vraisemblablement été déclarée non significative si la même méthode de mise à l'échelle avait été utilisée. Cette différence est probablement imputable au nouveau traitement des items non atteints. De même, l'augmentation non significative constatée en Uruguay (+9 points) aurait été quasi nulle (+1 point) si la même méthode de mise à l'échelle avait été utilisée.
- À Singapour, le score moyen a diminué de 9 points, un écart déclaré significatif vu l'erreur d'échantillonnage minime dans ce pays. Si les résultats de l'évaluation PISA 2012 avaient été générés à l'aide du modèle de mise à l'échelle utilisé lors de l'évaluation PISA 2015, le score aurait toutefois été inférieur de 7 points au score publié ; l'écart par rapport au score de l'évaluation PISA 2015 n'aurait en effet été que de -2 points si la même méthode de mise à l'échelle avait été utilisée lors des deux évaluations.

Tous les autres écarts entre les variations constatées et les variations estimées une fois que les résultats des évaluations PISA précédentes sont générés à l'aide de la méthode de mise à l'échelle utilisée lors de l'évaluation PISA 2015 sont moindres que les écarts prévus compte tenu des erreurs d'ancrage indiquées dans les sections suivantes de la présente annexe.

Graphique A5.1 ■ **Évolution de la performance en sciences entre 2006 et 2015, basée sur les résultats obtenus après la mise à l'échelle initiale et ceux obtenus après la nouvelle mise à l'échelle**

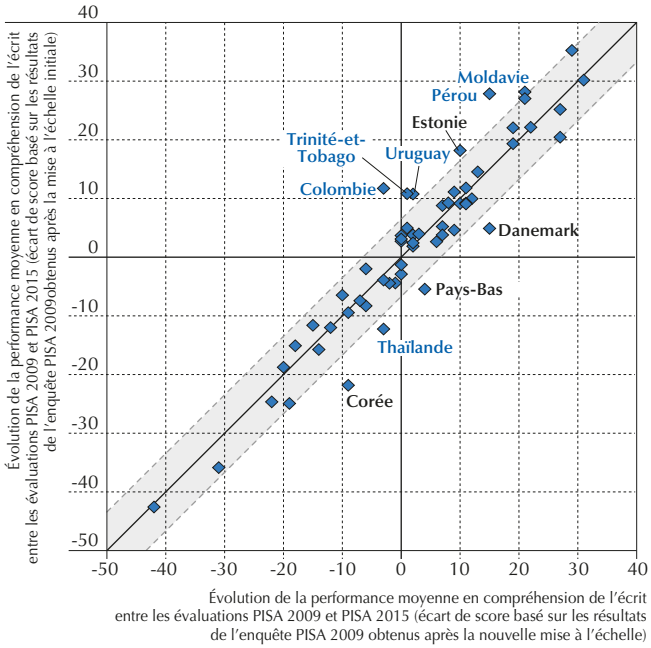


Remarque : La ligne en trait plein indique la diagonale, qui représente l'absence d'évolution entre les résultats initiaux et les résultats mis à l'échelle. La zone grisée indique l'intervalle de confiance de la diagonale, basé sur l'erreur d'ancrage des comparaisons des résultats de 2006 obtenus après la mise à l'échelle initiale et ceux de 2015 (voir le tableau A5.2).

Source : OCDE, Base de données PISA 2015, tableaux I.2.4a et A5.3.

StatLink http://dx.doi.org/10.1787/888933433132

Graphique A5.2 ■ Évolution de la performance en compréhension de l'écrit entre 2009 et 2015, basée sur les résultats obtenus après la mise à l'échelle initiale et ceux obtenus après la nouvelle mise à l'échelle

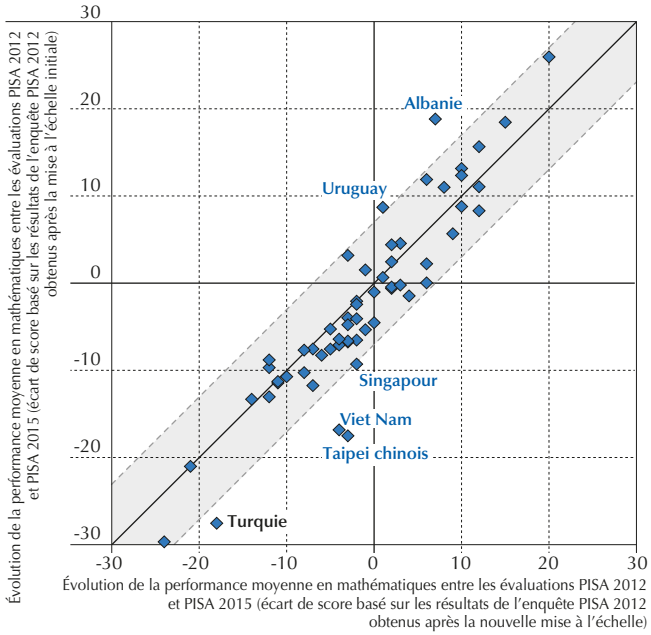


Remarque : La ligne en trait plein indique la diagonale, qui représente l'absence d'évolution entre les résultats initiaux et les résultats mis à l'échelle. La zone grisée indique l'intervalle de confiance de la diagonale, basé sur l'erreur d'ancrage des comparaisons des résultats de 2009 obtenus après la mise à l'échelle initiale et ceux de 2015 (voir le tableau A5.2).

Source : OCDE, Base de données PISA 2015, tableaux I.4.4a et A5.4.

StatLink http://dx.doi.org/10.1787/888933433149

Graphique A5.3 ■ Évolution de la performance en mathématiques entre 2012 et 2015, basée sur les résultats obtenus après la mise à l'échelle initiale et ceux obtenus après la nouvelle mise à l'échelle



Remarque : La ligne en trait plein indique la diagonale, qui représente l'absence d'évolution entre les résultats initiaux et les résultats mis à l'échelle. La zone grisée indique l'intervalle de confiance de la diagonale, basé sur l'erreur d'ancrage des comparaisons des résultats de 2012 obtenus après la mise à l'échelle initiale et ceux de 2015 (voir le tableau A5.2).

Source : OCDE, Base de données PISA 2015, tableaux I.5.4a et A5.5.

StatLink http://dx.doi.org/10.1787/888933433156

Comparaison des résultats entre les épreuves PISA sur papier et sur ordinateur

L'équivalence des items d'ancrage, qui a été estimée à l'échelon international, a été établie lors de l'étude à grande échelle sur l'effet du mode d'administration qui a été menée lors de l'essai de terrain de l'évaluation PISA 2015. Cette étude confirme, éléments probants à l'appui, que les résultats des élèves peuvent être rapportés sur la même échelle quel que soit le mode d'administration des épreuves. De plus, l'existence de paramètres nationaux spécifiques peut, dans une certaine mesure, expliquer les écarts nationaux par rapport à la norme internationale.

L'équivalence des items d'ancrage a tout d'abord été évaluée lors de l'essai de terrain (en 2014) dans des populations équivalentes prélevées de manière aléatoire dans les établissements. Dans les pays et économies envisageant d'administrer les épreuves PISA sur ordinateur en 2015, plus de 40 000 élèves se sont vu attribuer de manière aléatoire des épreuves sur papier ou sur ordinateur au sein des établissements échantillonnés, pour que le niveau de compétence des élèves soit comparable entre les deux modes d'administration. Cette approche a permis d'imputer toute différence dans les tendances de réponse des élèves entre les modes d'administration, en particulier les différences plus importantes que celles qui pouvaient être attendues à cause des variations aléatoires, à un impact du mode d'administration sur l'item plutôt qu'à la capacité des élèves d'utiliser le mode d'administration. L'essai de terrain a été conçu pour examiner les effets du mode d'administration à l'échelle internationale, mais pas dans chacun des échantillons ou sous-échantillons nationaux.

L'étude sur les effets du mode d'administration a été menée pour répondre à deux grandes questions :

- Les items conçus pour être administrés sur papier lors d'évaluations PISA précédentes évaluent-ils les mêmes compétences s'ils sont administrés sur ordinateur ? Par exemple, les items de sciences qui ont été adaptés pour être administrés sur ordinateur évaluent-ils uniquement des compétences en sciences ? Ou évaluent-ils à la fois des compétences en sciences et en informatique ?
- La difficulté des items est-elle la même selon qu'ils sont administrés sur papier ou sur ordinateur ?

Seuls les items dont il a été établi qu'ils évaluaient les mêmes compétences et qu'ils étaient du même degré de difficulté sur papier et sur ordinateur ont été déclarés totalement équivalents (invariance scalaire) et probants pour étayer des comparaisons de score entre les modes d'administration. Cette analyse d'équivalence des épreuves repose sur les données agrégées de tous les pays et économies. Des modèles explicatifs basés sur la théorie de réponse à l'item (*Item Response Theory*, IRT) ont été utilisés. Dans ces modèles, deux séries distinctes de paramètres sont utilisées pour déterminer dans quelle mesure les réponses des élèves sont révélatrices de leur niveau de compétence et estimer leur niveau de compétence. L'analyse d'équivalence a identifié trois groupes d'items :

- Relèvent du groupe 1 les items dont le degré de difficulté et les paramètres de discrimination sont identiques dans les deux modes d'administration ; il est établi que ces items sont parfaitement équivalents sur papier et sur ordinateur (*invariance scalaire*).
- Relèvent du groupe 2 les items dont les paramètres sont identiques, mais dont le degré de difficulté varie (*invariance métrique*). Le fait de répondre correctement à ces items est révélateur à certains égards du niveau de compétence dans le domaine d'évaluation ; mais la difficulté des items varie entre les modes d'administration, souvent à cause des interfaces, par exemple dans des formats où les élèves doivent dessiner à main levée ou construire des équations. Plusieurs de ces items se sont révélés soit plus difficiles, soit plus faciles sur ordinateur.
- Relèvent du groupe 3 les items dont les estimations de l'essai de terrain indiquent qu'ils évaluent des compétences différentes selon le mode d'administration (pas d'*invariance métrique*).

Les items classés dans le groupe 3 n'ont pas été inclus dans les épreuves administrées sur ordinateur lors de la campagne définitive (deux items de mathématiques ont été inclus dans les épreuves administrées sur papier uniquement). Des items classés dans les groupes 1 et 2 ont été administrés, et la stabilité de leurs paramètres a été analysée de manière plus approfondie entre les modes d'administration et les évaluations lors des opérations de mise à l'échelle de la campagne définitive. En fin de compte, les données confirment une équivalence parfaite (scalaire) entre les modes concernant 61 items de sciences, 65 items de compréhension de l'écrit et 51 items de mathématiques[7]. Ces items font office d'items d'ancrage lors de la mise à l'échelle et sont à la base des comparaisons entre les modes d'administration et entre les évaluations. L'équivalence métrique des autres items d'ancrage (24 items en sciences, 38 items en compréhension de l'écrit et 30 items en mathématiques) administrés lors de la campagne définitive de l'évaluation PISA 2015 a été confirmée, mais un paramètre de difficulté propre à chaque mode leur a été attribué. Il y a lieu de signaler au sujet de la comparaison des résultats entre les élèves qui ont passé les épreuves sur papier et sur ordinateur que ce sous-groupe d'items à invariance métrique fournit uniquement des informations sur le rang des élèves dans le classement selon le mode d'administration qui les concerne (et contribue à la précision de l'estimation de leur niveau de compétence), mais pas sur le rang des élèves et des pays dans le classement selon les deux modes d'administration. Les paramètres des items dont l'équivalence est scalaire sont identiques dans les épreuves sur papier et sur ordinateur dans les tableaux C2.1, C2.3 et C2.4 ; et les items dont l'équivalence est seulement métrique sont associés à des paramètres identiques en termes de pente, mais différents en termes de difficulté.

L'équivalence parfaite des items d'ancrages entre les modes d'administration, qui a été déterminée sur la base d'un échantillon représentatif de tous les élèves qui ont passé les épreuves PISA sur ordinateur, garantit que les résultats peuvent être comparés entre les épreuves sur papier et sur ordinateur, et que le lien entre les deux séries de résultats est probant. Cela implique entre autres que les élèves qui ont passé les épreuves PISA sur ordinateur en 2015 n'auraient pas obtenu un score significativement différent s'ils les avaient passées sur papier ; dans cette hypothèse, les pourcentages d'élèves aux divers niveaux de compétence n'auraient pas non plus été significativement différents.

L'annexe A6 fournit de plus amples informations sur l'analyse exploratoire des interactions entre les groupes d'items et les modes d'administration qui a été menée sur la base des données de l'essai de terrain. Les résultats de cette analyse, en particulier sur les interactions entre les modes d'administration et les sexes, sont encourageants, mais il y a lieu de tenir compte de la portée limitée des données de l'essai de terrain dans ce type d'exercice lors de l'interprétation des résultats.

Évaluation de la comparabilité des items d'ancrage et des nouveaux items dans les épreuves de sciences

De nouveaux items de sciences ont été conçus en vue de l'évaluation PISA 2015 pour refléter les changements introduits dans le cadre PISA de culture scientifique et le nouveau mode principal d'administration. Dans l'enquête PISA, la révision du cadre d'évaluation, qui coïncide avec la conception de nouveaux items, est fréquente : le cadre d'évaluation a été remanié en compréhension de l'écrit en 2009 et en mathématiques en 2012. La conception de nouveaux items de sciences a été guidée par la nécessité de couvrir de manière équilibrée tous les aspects du cadre de l'évaluation PISA 2015, en particulier les aspects qui ont été affinés ou mis en évidence par rapport au cadre de l'évaluation PISA 2006. Parmi ces aspects revisités, citons la distinction entre les connaissances procédurales et épistémiques, qui était implicite dans le cadre précédent, et la composante plus active de la culture scientifique. Cette dernière évolution se retrouve dans la nouvelle organisation de la culture scientifique autour des compétences « évaluer et concevoir des investigations scientifiques » et « interpréter des données et des faits de manière scientifique » (et « expliquer des phénomènes de manière scientifique »). Ces compétences sont liées aux compétences « identifier des questions d'ordre scientifique » et « utiliser des faits scientifiques » retenues auparavant, mais elles n'y sont de toute évidence pas totalement identiques.

Après la campagne définitive de 2015, la dimensionnalité a été évaluée pour déterminer s'il était possible de rapporter les résultats sur l'échelle existante de culture scientifique, celle élaborée en 2006. Lorsque les anciens et les nouveaux items de sciences ont été traités comme s'ils étaient liés à des dimensions latentes distinctes, une corrélation médiane de 0.92 a été établie entre ces dimensions (dans l'ensemble des groupes pays/langue) ; ce coefficient est relativement élevé (et similaire à celui qui s'observe dans les sous-échelles d'un même domaine). Les statistiques d'adéquation ont confirmé qu'un modèle unidimensionnel était plus adapté aux données qu'un modèle bidimensionnel, ce qui étaye la conclusion selon laquelle les anciens et les nouveaux items de sciences forment une échelle unidimensionnelle cohérente d'une bonne fiabilité. Le rapport technique sur l'évaluation PISA 2015, *PISA 2015 Technical Report* (OCDE, à paraître en anglais uniquement) fournit des informations plus détaillées sur les résultats de la mise à l'échelle.

Quantifier l'incertitude de la comparabilité de l'échelle dans l'erreur d'ancrage

Les erreurs-types des estimations de l'évolution de la performance entre les évaluations PISA tiennent compte de l'incertitude introduite par la mise en correspondance d'échelles produites selon un étalonnage différent. Ces erreurs-types ajustées, plus importantes que les erreurs-types estimées avant l'introduction de l'erreur d'ancrage, reflètent non seulement la précision des mesures et la variation de l'échantillonnage dans les résultats PISA habituels, mais aussi l'erreur d'ancrage présentée au tableau A5.2. Dans l'évaluation PISA 2015, l'erreur d'ancrage reflète non seulement l'incertitude liée à la sélection des items d'ancrage, mais également l'incertitude liée aux changements introduits dans la méthode de mise à l'échelle, par rapport aux évaluations précédentes.

Comme lors des évaluations précédentes, seule l'incertitude associée à la localisation des scores des évaluations précédentes sur l'échelle de compétence de 2015 est reflétée dans l'erreur d'ancrage. Comme cette incertitude relative à la localisation dans la répartition (un changement dans l'intercept) disparaît si l'on analyse les estimations invariantes en termes de localisation (par exemple les estimations de la variance, de la plage inter-quartiles, des écarts entre les sexes, des coefficients de régression, des coefficients de corrélation, etc.), les erreurs-types de ces estimations n'incluent pas l'erreur d'ancrage.

Erreur d'ancrage des scores entre deux évaluations PISA

Dans l'évaluation PISA 2015, les erreurs d'ancrage sont estimées sur la base de la comparaison des scores moyens des pays et économies dans chaque domaine après nouvelle mise à l'échelle (c'est-à-dire ceux indiqués dans les tableaux A5.3, A5.4 et A5.5) avec les scores moyens publiés qui ont été générés selon la méthode de mise à l'échelle retenue lors de l'évaluation dont ils relèvent. Cette nouvelle méthode d'estimation des erreurs d'ancrage a été utilisée pour la première fois lors de l'évaluation PISA 2015. Le nombre d'observations utilisé pour calculer chaque erreur d'ancrage est égal au nombre de pays dont les résultats des deux évaluations sont disponibles. Comme les données sous-tendant le calcul de l'erreur d'ancrage ne sont pas nombreuses, une estimation probante de l'écart-type a été utilisée, sur la base de la valeur statistique S_n (Rousseeuw et Croux, 1993).

Tableau A5.2 Erreurs d'ancrage dans les comparaisons entre PISA 2015 et les évaluations antérieures

Comparaison	Sciences	Compréhension de l'écrit	Mathématiques
Entre PISA 2000 et 2015		6.8044	
Entre PISA 2003 et 2015		5.3907	5.6080
Entre PISA 2006 et 2015	4.4821	6.6064	3.5111
Entre PISA 2009 et 2015	4.5016	3.4301	3.7853
Entre PISA 2012 et 2015	3.9228	5.2535	3.5462

Remarque : Les comparaisons entre les scores obtenus à l'évaluation PISA 2015 et ceux obtenus aux évaluations antérieures ne peuvent être réalisées qu'avec l'évaluation où le domaine concerné est devenu pour la première fois domaine majeur d'évaluation. Ainsi, il n'est par exemple pas possible de comparer la performance en sciences entre les évaluations PISA 2000 et PISA 2015.

StatLink http://dx.doi.org/10.1787/888933433162

Erreur d'ancrage dans les autres types de comparaison de la performance des élèves

Les erreurs d'ancrage de l'évolution de la performance, calculée à partir de régressions et de comparaisons basées sur des transformations non linéaires des scores, peuvent être estimées par simulation, compte tenu de l'erreur d'ancrage de la comparaison des scores entre deux évaluations PISA. Le tableau A5.6 indique les estimations de l'erreur d'ancrage dans les comparaisons du pourcentage d'élèves sous le niveau 2 et au niveau 5 ou 6, tandis que le tableau A5.7 indique l'ampleur de l'erreur d'ancrage associée à l'estimation de l'évolution moyenne de la performance par intervalle de trois ans.

L'estimation des erreurs d'ancrage concernant le pourcentage d'élèves sous le niveau 2 et au niveau 5 ou 6 repose sur l'hypothèse que l'importance de l'incertitude associée à la mise en correspondance d'échelles suit une distribution normale, avec une moyenne de 0 et un écart-type égal à l'erreur d'ancrage indiquée dans le tableau A5.2. Cinq cents erreurs sont tirées de cette distribution et ajoutées à la première valeur plausible des élèves en 2015 dans chacun des pays et économies, pour représenter les 500 scénarios possibles dans lesquels la seule source de différence par rapport à 2015 est l'incertitude de l'ancrage.

L'estimation de l'intérêt (pourcentage d'élèves à un niveau de compétence spécifique) est calculée pour chacune des 500 répétitions, ce qui permet d'évaluer l'effet de l'erreur d'ancrage sur cette estimation. L'écart-type des 500 estimations répétées est utilisé comme erreur d'ancrage pour la variation du pourcentage d'élèves à un niveau de compétence spécifique. Comme l'effet de l'erreur d'ancrage sur cette estimation dépend de la forme et de la densité exacte de la répartition de la performance autour des points de limite, les erreurs d'ancrage varient entre les pays et entre les filles et les garçons au sein des pays dans les comparaisons des niveaux de compétence.

L'estimation des erreurs d'ancrage dans les tendances d'évolution basées sur des régressions repose aussi sur l'hypothèse que l'incertitude de l'ancrage suit une distribution normale avec une moyenne de 0 et un écart-type égal à l'erreur d'ancrage indiquée au tableau A5.2. Comme ce sont les tendances entre plus de deux évaluations qui nous intéressent, la covariance entre les erreurs d'ancrage doit être considérée en plus des erreurs d'ancrage indiquées au tableau A5.2. Pour simuler les données de plusieurs évaluations PISA, 2 000 observations ont été prélevées dans une distribution normale à variables multiples dont toutes les moyennes sont égales à 0 et dont la structure de variance/covariance est identifiée par l'erreur d'ancrage indiquée au tableau A5.2, ainsi que par celles estimées pour les échelles PISA antérieures, publiées dans le tableau 12.31 du rapport technique sur l'évaluation PISA 2012, *PISA 2012 Technical Report* (OCDE, 2014). Ces observations représentent 2 000 scénarios possibles où la tendance réelle est égale à 0, et où la tendance estimée reflète totalement l'incertitude de la comparabilité des scores entre les échelles. Les erreurs d'ancrage de l'évolution moyenne, par intervalle de trois ans, de la performance entre l'évaluation PISA 2015 et les évaluations antérieures dépendent du nombre d'évaluations retenu dans l'estimation, mais pas de la forme de la distribution de la performance dans chaque pays.

Comparaison de la performance : différence entre deux évaluations et évolution moyenne par intervalle de trois ans

Afin d'évaluer l'évolution de la performance, les analyses étudient la variation de la performance entre deux évaluations et son évolution moyenne par intervalle de trois ans. En compréhension de l'écrit, où jusqu'à six valeurs sont disponibles dans le temps, des tendances curvilignes sont également estimées.

Les comparaisons entre deux évaluations (par exemple, la variation de la performance d'un sous-groupe ou d'un pays ou d'une économie entre les évaluations PISA 2006 et PISA 2015) sont calculées de la manière suivante :

$$\Delta_{2015\text{-}t} = PISA_{2015} - PISA_{t}$$

Où $\Delta_{2015\text{-}t}$ représente l'écart de performance entre l'évaluation PISA 2015 et une évaluation PISA antérieure (on ne peut faire de comparaisons qu'à partir du moment où une matière a été le domaine majeur d'une évaluation ; il n'est donc pas possible de comparer la performance en mathématiques entre les évaluations PISA 2015 et PISA 2000, ni de comparer la performance en sciences entre les évaluations PISA 2015 et PISA 2000 ou PISA 2003) ; où $PISA_{2015}$ est le score en mathématiques, en

compréhension de l'écrit ou en sciences lors de l'évaluation PISA 2015 et $PISA_t$ est le score en mathématiques, en compréhension de l'écrit ou en sciences lors d'une évaluation antérieure. L'erreur-type de la variation de la performance $\sigma(\Delta_{2015\text{-}t})$ est égale à :

$$\sigma(\Delta_{2015-t}) = \sqrt{\sigma_{2015}^2 + \sigma_t^2 + erreur_{2015,t}^2}$$

Où σ_{2015} correspond à l'erreur-type du score $PISA_{2015}$, σ_t correspond à l'erreur-type du score $PISA_t$ et l'$erreur_{2015,t}$ correspond à l'erreur d'ancrage des comparaisons de la performance en mathématiques, en compréhension de l'écrit ou en sciences entre l'évaluation PISA 2015 et l'évaluation antérieure (t). La valeur de l'$erreur_{2015,t}$ est indiquée au tableau A5.2 dans la plupart des comparaisons et au tableau A5.6 dans les comparaisons des niveaux de compétence.

Un second ensemble d'analyses présenté dans le rapport sur l'évaluation PISA se rapporte à l'évolution moyenne de la performance par intervalle de trois ans. Il s'agit de la mesure moyenne dans laquelle la performance a évolué, par intervalle de trois ans (l'intervalle typique entre les deux évaluations PISA successives), dans les pays et économies depuis qu'ils participent à l'enquête PISA. Une évolution positive de x points en moyenne par intervalle de trois ans indique donc que le score a augmenté de x points tous les trois ans dans un pays ou une économie depuis la première évaluation PISA auquel il a participé et dont les données sont comparables. Dans les pays et économies qui ont uniquement participé aux évaluations PISA 2012 et PISA 2015, l'évolution moyenne de la performance par intervalle de trois est égale à la différence de score entre les deux évaluations[8].

L'évolution moyenne de la performance par intervalle de trois ans est calculée selon la régression suivante :

$$\text{PISA}_{i,t} = ß_0 + ß_1 temps_t + \varepsilon_{i,t}$$

Où $PISA_{i,t}$ est le rang du pays i sur l'échelle de culture scientifique, de compréhension de l'écrit ou de culture mathématique l'année t (le score moyen ou le centile du score dans la répartition), le $temps_t$ est une variable mesurant le temps par unité de trois ans et où $\varepsilon_{i,t}$ est le terme d'erreur indiquant l'incertitude d'échantillonnage et d'évaluation du rang $PISA_{i,t}$. Dans l'estimation, les erreurs d'ancrage et d'évaluation sont supposées être indépendantes dans le temps. Selon ces spécifications, l'estimation de $ß_1$ indique l'évolution moyenne de la performance par intervalle de trois ans. Tout comme une erreur d'ancrage est intégrée lors de la comparaison entre deux évaluations PISA, les erreurs-types de $ß_1$ comprennent également une erreur d'ancrage :

$$\sigma(ß_1) = \sqrt{\sigma_{s,i}^2(ß_1) + \sigma_t^2(ß_1)}$$

Où $\sigma_{s,i}(ß_1)$ est l'erreur d'échantillonnage et d'imputation associée à l'estimation de $ß_1$ et $\sigma_t^2(ß_1)$ est l'erreur d'ancrage associée à l'évolution moyenne de la performance par intervalle de trois ans (voir le tableau A5.2).

L'évolution moyenne de la performance par intervalle de trois ans est un indicateur plus probant de l'évolution des résultats scolaires dans un pays ou une économie, car elle est calculée sur la base des informations recueillies lors de toutes les évaluations. Elle est donc moins sensible aux mesures anormales qui peuvent affecter les comparaisons basées sur deux évaluations seulement. L'évolution moyenne de la performance par intervalle de trois est à la meilleure estimation de l'évolution des résultats des pays et économies au fil de leur participation aux évaluations PISA. Elle est calculée sur la base d'une régression des scores PISA l'année où les pays ou économies ont participé à l'évaluation PISA (par intervalle de trois ans). L'évolution moyenne de la performance par intervalle de trois ans tient également compte du fait que dans certains pays et économies, l'intervalle entre les différentes évaluations PISA est inférieur à trois ans. C'est le cas dans les pays et économies qui ont administré les épreuves de l'évaluation PISA 2000 ou PISA 2009 dans le cadre de PISA+ : les épreuves y ont été administrées en 2001 ou en 2002, et non en 2000, et en 2010, et non en 2009.

En compréhension de l'écrit, les tendances curvilignes sont estimées selon une méthode similaire, sur la base d'une fonction de régression quadratique dans les scores PISA du pays i entre les évaluations indexées par t :

$$PISA_{i,t} = ß_2 + ß_3 année_t + ß_4 année_t^2 + \varepsilon_{i,t}$$

Où l'$année_t$ est une variable mesurant le temps jusqu'à 2015 et l'$année_t^2$ est égale à l'$année_t$ au carré. Comme l'année est mise à l'échelle pour être égale à 0 en 2015, $ß_3$ indique le taux estimé de variation annuelle jusqu'en 2015 et $ß_2$, l'accélération ou le ralentissement de la tendance. Si $ß_4$ est positif, la tendance observée est en forme de « U » et les taux de variation de la performance sont plus élevés (plus positifs) dans les années plus proches de 2012 que dans les années précédentes. Si $ß_4$ est négatif, la tendance observée est en forme de « U » inversé et les taux de variation de la performance sont moins élevés (plus négatifs) dans les années plus proches de 2012 que dans les années précédentes. Tout comme une erreur d'ancrage est intégrée à l'estimation des erreurs-types de l'évolution moyenne de la performance par intervalle de trois ans, une erreur d'ancrage est intégrée dans les erreurs-types de $ß_3$ et de $ß_4$ (voir le tableau A5.2). Les tendances curvilignes sont uniquement estimées en

compréhension de l'écrit dans les pays et économies dont la performance peut être comparée entre cinq évaluations au moins, pour éviter un ajustement excessif des données.

Tendances ajustées

L'enquête PISA conserve les mêmes normes techniques d'une évaluation à l'autre. Cette constance permet de calculer des tendances dans des populations définies de manière cohérente, certes, mais le pourcentage de l'effectif d'individus de 15 ans que ces populations représentent et les caractéristiques démographiques de l'effectif d'élèves de 15 ans peuvent évoluer, par exemple à cause des flux migratoires.

Étant donné que les analyses tendancielles illustrent le rythme de la progression de cohortes d'élèves successives, il est important d'examiner la mesure dans laquelle elles sont influencées par l'évolution du taux de couverture des échantillons et des caractéristiques démographiques des élèves si l'on veut tirer des conclusions fiables de ces résultats. C'est pourquoi trois séries de tendances ont été développées : les tendances non ajustées, les tendances ajustées compte tenu de l'évolution des taux de scolarisation et les tendances ajustées compte tenu de l'évolution des caractéristiques démographique des échantillons. Les tendances ajustées montrent l'évolution de la performance telle qu'elle est estimée après neutralisation des effets de l'évolution concomitante des caractéristiques démographiques des échantillons.

Tendances ajustées compte tenu de l'évolution des taux de scolarisation

Pour neutraliser l'impact de l'évolution des taux de scolarisation (ou, plus précisément, de la variation de la mesure dans laquelle l'échantillon PISA est représentatif de la population totale d'individus de 15 ans ; voir l'indice de couverture 3 à l'annexe A2), l'hypothèse est que les jeunes de 15 ans qui ne sont pas représentés dans l'échantillon auraient tous obtenu des scores inférieurs au score médian des jeunes de 15 ans qui y sont représentés. Grâce à cette hypothèse, le score médian de tous les individus de 15 ans et le score des centiles supérieurs ont pu être calculés (dans les pays dont le taux de couverture de l'échantillon atteint 50 % au moins) sans qu'il faille préciser le score des individus de 15 ans non représentés dans l'échantillon.

Dans les faits, l'estimation des tendances ajustées compte tenu de l'évolution des taux de scolarisation requiert l'ajout, dans la base de données, d'une observation par pays et économie correspondant à l'effectif d'individus de 15 ans qui ne sont pas représentés par l'échantillon PISA. La pondération finale de niveau Élève de cette observation correspond à la différence entre l'effectif total d'individus de 15 ans (voir le tableau I.6.1 et l'annexe A2) et la somme des pondérations finales de niveau Élève des observations représentées dans l'échantillon (le nombre pondéré d'élèves participants). De même, chaque réplique de pondération de cette observation est égale à la différence entre la population totale d'individus de 15 ans et la somme des répliques de pondération correspondantes. Toutes les pondérations négatives résultant de cette procédure sont remplacées par une valeur nulle. Une valeur inférieure aux valeurs plausibles de l'échantillon PISA est imputée dans les variables de performance de cette observation.

Dans un deuxième temps, les scores dans le centile médian et les centiles supérieurs de la répartition sont calculés sur la base de cet échantillon majoré. Dans les quelques cas où le taux de couverture est inférieur à 50 %, l'estimation du score médian après ajustement est déclarée manquante.

Tendances ajustées compte tenu de l'évolution des caractéristiques démographiques de l'échantillon

Une procédure de repondération, analogue à une post-stratification, a été utilisée pour aligner les caractéristiques démographiques des échantillons antérieurs sur les caractéristiques de l'échantillon de l'évaluation PISA 2015.

Dans un premier temps, l'échantillon de chaque évaluation est divisé en cellules discrètes, selon le statut des élèves au regard de l'immigration (quatre catégories : élèves autochtones, élèves immigrés de la première génération, élèves immigrés de la deuxième génération, et statut manquant), leur sexe (deux catégories : garçons et filles) et leur âge (quatre catégories correspondant à quatre trimestres). Les quelques observations incluses dans les bases de données PISA antérieures dont le sexe ou l'âge est manquant ont été supprimées. Ce processus aboutit à la définition de 32 cellules discrètes maximum concernant l'effectif total. Toutefois, si le nombre d'observations inclus dans l'une de ces 32 cellules est inférieur à 10 dans un pays ou une économie ou lors d'une évaluation PISA, la cellule correspondante est combinée à une autre cellule similaire, selon un algorithme séquentiel, jusqu'à ce que toutes les cellules comportent au minimum 10 observations[9].

Dans un deuxième temps, les cellules sont repondérées de sorte que la somme des pondérations finales de niveau Élève de chaque cellule est constante entre les évaluations et égale à la somme des pondérations finales de niveau Élève dans l'échantillon de l'évaluation PISA 2015. On estime alors le score moyen et la répartition des élèves selon la performance sur la base de ces échantillons repondérés. Ces estimations correspondent à la performance qui aurait été observée si les caractéristiques des échantillons des évaluations précédentes avaient été les mêmes que celles de l'échantillon de l'évaluation PISA 2015 concernant les variables utilisées dans cette procédure de repondération.

Le tableau A5.2 indique, dans tous les pays et économies, le nombre de cellules utilisées durant la post-stratification et, à chaque évaluation, le nombre d'observations exclues des tendances pour tenir compte de l'évolution des caractéristiques démographiques de l'échantillon.

Le tableau A5.2 indique, dans tous les pays et économies, les moyennes des variables contextuelles utilisées dans l'ajustement.

Comparaison des échelles et des items non cognitifs entre les différentes évaluations PISA

Dans l'enquête PISA, les élèves et les chefs d'établissement remplissent un questionnaire contextuel, qui permet de recueillir des informations sur leurs caractéristiques. Comme certaines questions n'ont pas changé entre les évaluations PISA 2006 et PISA 2015, il est possible de comparer l'évolution des réponses à ces questions au fil du temps. Les réponses aux questions dont l'énoncé a légèrement ou fortement changé n'ont pas été comparées dans le temps (sauf mention contraire), car il est impossible de déterminer si les différences observées dans les réponses sont dues à des changements dans leur construct ou à des changements dans la façon dont leur construct est mesuré.

Par ailleurs, comme l'indique l'annexe A1, les items utilisés dans les questionnaires PISA servent à élaborer des indices. Dans l'enquête PISA, deux types d'indices sont utilisés : les indices simples et les indices mis à l'échelle.

Les indices simples recodent un ensemble de réponses aux items. Dans les analyses tendancielles, les valeurs observées lors de l'évaluation PISA 2006 sont comparées directement à celles de l'évaluation PISA 2015, comme les réponses aux items des questionnaires. C'est le cas des indices tels que le taux d'encadrement et le pourcentage d'élèves issus de l'immigration.

En revanche, les indices mis à l'échelle reposent sur des estimations pondérées des réponses les plus vraisemblables (*Warm Likelihood Estimate*, WLE ; Warm, 1989) dans la base de données, ainsi que sur un modèle de crédit partiel généralisé (*Generalised Partial Credit Model*, GPCM ; voir Muraki 1992). À chaque fois qu'une partie de question utilisée dans la construction d'un indice n'a pas changé entre les évaluations PISA 2006 et PISA 2015, la mise à l'échelle de l'indice correspondant se base sur un étalonnage concomitant des données des deux évaluations, suivi d'une transformation linéaire, le but étant de constituer une échelle basée sur l'échelle de l'indice lors de l'évaluation PISA 2006, qui a été dérivée à l'aide d'un modèle de crédit partiel (*Partial Credit Model*, PCM ; voir OCDE, 2009). Cette procédure, qui est analogue à celle utilisée pour les échelles cognitives, garantit la comparabilité des valeurs d'indice.

Afin d'évaluer la variation de ces items et de ces échelles, les analyses présentent la variation de l'estimation entre deux évaluations différentes, généralement les évaluations PISA 2006 et PISA 2015. La comparaison entre deux évaluations (la variation de l'indice du plaisir de l'apprentissage des sciences entre les évaluations PISA 2006 et PISA 2015, ou sa variation dans un sous-groupe) est calculée comme suit :

$$\Delta_{2015,2006} = PISA_{2015} - PISA_{2006}$$

Où $\Delta_{2015,t}$ correspond à l'écart entre l'indice de l'évaluation PISA 2015 et une évaluation antérieure, $PISA_{2015}$ correspond à la valeur de l'indice lors de l'évaluation PISA 2015 et $PISA_{2006}$ correspond à la valeur d'indice lors de l'évaluation PISA 2006. L'erreur-type de la variation de la valeur d'indice $\sigma(\Delta_{2015\text{-}2006})$ est égale à :

$$\sigma(\Delta_{2015-2006}) = \sqrt{\sigma^2_{2015} + \sigma^2_{2006}}$$

Où σ_{2015} est l'erreur-type de la valeur $PISA_{2015}$ et σ_{2006} est l'erreur-type de la valeur $PISA_{2006}$. Comme les erreurs-types de la variation des valeurs d'indice n'incluent ni l'incertitude de mesure, ni l'incertitude liée à la procédure d'étalonnage, elles sont quelque peu sous-estimées. Les erreurs-types de la variation des réponses à des items spécifiques ne sont pas entourées d'incertitude de mesure ou d'étalonnage.

Moyenne de l'OCDE

La moyenne de l'OCDE sert de valeur de référence tout au long de ce rapport. Elle correspond à la moyenne des résultats dans l'ensemble des pays de l'OCDE, chaque pays étant pondéré de façon égale. Certains pays de l'OCDE n'ont pas participé à des évaluations. Dans d'autres pays, les résultats de certaines évaluations ne sont pas comparables ou des questions ont été exclues des questionnaires ou ont été substantiellement modifiées entre des évaluations. Dans les graphiques et tableaux relatifs aux tendances, la moyenne de l'OCDE est calculée sur la base d'ensembles cohérents de pays de l'OCDE. Par exemple, la moyenne de l'OCDE-33 a été calculée sur la seule base des 33 pays de l'OCDE dont les observations ne sont pas manquantes lors des évaluations dont la moyenne elle-même n'est pas manquante. Cette restriction permet d'effectuer des comparaisons valables de la moyenne de l'OCDE au fil du temps.

Tableaux disponibles en ligne (StatLink http://dx.doi.org/10.1787/888933433162)

Tableau A5.3. Scores moyens en sciences depuis 2006 suite à l'application de la méthode de mise à l'échelle de 2015

Tableau A5.4. Scores moyens en compréhension de l'écrit depuis 2006 suite à l'application de la méthode de mise à l'échelle de 2015

Tableau A5.5. Scores moyens en mathématiques depuis 2006 suite à l'application de la méthode de mise à l'échelle de 2015

Tableau A5.6. Erreur d'ancrage pour les comparaisons des niveaux de compétence entre l'enquête PISA 2015 et les enquêtes précédentes

Tableau A5.7. Erreur d'ancrage pour les comparaisons de l'évolution moyenne par intervalle de 3 ans entre l'enquête PISA 2015 et les enquêtes précédentes

Tableau A5.8. Erreur d'ancrage pour la tendance curviligne entre l'enquête PISA 2015 et les enquêtes précédentes

Tableau A5.9. Cellules utilisées pour l'ajustement des scores en sciences, en compréhension de l'écrit et en mathématiques en fonction des échantillons de l'enquête PISA 2015

Tableau A5.10. Statistiques descriptives relatives aux variables d'ajustement des scores en sciences, en compréhension de l'écrit et en mathématiques en fonction des échantillons de l'enquête PISA 2015

Notes

1. Voir aussi les travaux de Carstensen (2013) sur l'influence de la conception des épreuves sur l'évaluation des tendances.

2. Le traitement limité du DIF lors des évaluations précédentes, combiné avec l'échantillon d'étalonnage spécifique à chacune d'entre elles, a été critiqué pour l'incohérence des estimations de tendances par rapport aux étalonnages nationaux sur la base d'échantillons concomitants (Urbach, 2013).

3. Dans l'évaluation PISA 2015, le nombre d'items non atteints est une information contextuelle qui est prise en considération lors de la production des valeurs plausibles : la corrélation entre les items non atteints et le niveau de compétence est modélisée et imputée dans les résultats.

4. Dans ce modèle, il est possible que certains pays et économies disposent des données de moins de quatre évaluations.

5. Le coefficient de corrélation des scores moyens des évaluations PISA 2009 et PISA 2012 des pays et économies qui ont participé à l'évaluation PISA 2015 s'établit à 0.985 en sciences (un domaine mineur lors des évaluations de 2009 et de 2012, dont les épreuves étaient strictement les mêmes), à 0.972 en compréhension de l'écrit (dont les épreuves de 2012 ne comportaient qu'une partie des épreuves PISA 2009) et à 0.981 en mathématiques (dont le cadre d'évaluation a été revu et les épreuves, étoffées, lors de l'évaluation PISA 2012). Les évaluations PISA 2009 et PISA 2012 sont les deux dont la conception des épreuves et la méthode de mise à l'échelle sont les plus similaires. La corrélation des scores moyens des évaluations PISA 2000 et PISA 2009 (dans les pays et économies qui ont participé à l'évaluation PISA 2015) s'établit à 0.955 en compréhension de l'écrit ; la corrélation des scores moyens des évaluations PISA 2003 et PISA 2012 s'établit à 0.953 en mathématiques ; et la corrélation des scores moyens des évaluation PISA 2006 et PISA 2015 s'établit à 0.947 en sciences (et à 0.944 selon les résultats indiqués dans le tableau A5.3, produits selon la même méthode de mise à l'échelle).

6. Les scores moyens des pays et économies générés lors de l'étape de mise à l'échelle sont ceux qui auraient été observés si seuls les élèves qui ont passé les épreuves dans les différents domaines avaient été pris en compte. Toutefois, comme les données de tous les élèves dans tous les domaines évalués dans les pays et économies sont imputés dans les modèles PISA, que les élèves aient reçu un carnet de test comportant des unités dans un domaine ou non, les scores générés durant la mise à l'échelle peuvent différer des scores moyens indiqués à l'annexe B. Précisons cependant que l'effet des scores imputés sur les scores moyens est négligeable, comme le révèle la comparaison des résultats de 2015, entre les estimations basées sur le modèle de mise à l'échelle, qui sont présentées dans les tableaux A5.3, A5.4 et A5.5, et les estimations basées sur le modèle où la totalité de la population est prise en compte, qui sont présentées dans les tableaux I.2.3, I.4.3 et I.5.3.

7. Lors de l'analyse des résultats nationaux, il convient de considérer que ces chiffres correspondent à une limite supérieure du nombre d'items à invariance scalaire, étant donné la possibilité d'écarts nationaux spécifiques aux évaluations par rapport à la norme internationale.

8. L'évolution moyenne de la performance par intervalle de trois ans est liée à la « variation annualisée » à laquelle les rapports sur des évaluations PISA précédentes font référence (OCDE, 2014b). L'évolution moyenne de la performance par intervalle de trois ans correspond à la variation annualisée multipliée par trois.

9. Les échantillons sont tout d'abord séparés selon le statut au regard de l'immigration (sauf si cela donne lieu à des groupes de moins de 10 observations), puis, dans les groupes définis selon ce statut, selon le sexe (sauf si cela donne lieu à des groupes de moins de 10 observations), et enfin, selon l'âge. S'il apparaît, à n'importe quel stade du processus, que des groupes comportent moins de 10 observations, les fusions suivantes sont effectuées ; à chaque stade, la séquence de fusions s'interrompt dès que tous les groupes comportent au moins 10 observations. Première étape (statut au regard de l'immigration, dans les groupes linguistiques définis auparavant) : fusion des catégories de statut manquant et d'autochtones ; fusion des catégories d'élèves immigrés de la première génération et de la deuxième génération ; fusion de toutes les catégories. Deuxième étape (sexe, dans les groupes d'élèves issus de l'immigration définis auparavant) : fusion des catégories de garçons et de filles. Troisième étape (âge, dans les groupes d'élèves issus de l'immigration et de garçons et de filles définis auparavant) : fusion du premier et du deuxième trimestre ; fusion du troisième et du quatrième trimestre, fusion de toutes les catégories.

Références

Birnbaum, A. (1968), *On the Estimation of Mental Ability*, Series Report 15, USAF School of Aviation Medicine, Randolph Air Force Base (TX).

Carstensen, C.H. (2013), « Linking PISA competencies over three cycles – Results from Germany », pp. 199-213 in *Research on PISA*, Springer, Pays-Bas, http://dx.doi.org/10.1007/978-94-007-4458-5_12.

Davidov, E., P. Schmidt et **J. Billiet** (éd.) (2011), *Cross-Cultural Analysis: Methods and Applications*, Routledge, New York.

Glas, C. et **K. Jehangir** (2014), « Modeling country specific differential item functioning », in *Handbook of International Large-Scale Assessment*, CRC Press, Boca Raton (FL).

Masters, G.N. (1982), « A Rasch model for partial credit scoring », *Psychometrika*, vol.47/2, pp. 149-74, http://dx.doi.org/10.1007/BF02296272.

Meredith, W. (1993), « Measurement invariance, factor analysis and factorial invariance », *Psychometrika*, vol. 58/4, pp. 525-43, http://dx.doi.org/10.1007/BF02294825.

Muraki, E. (1992), « A generalized partial credit model: Application of an EM algorithm », *Applied Psychological Measurement*, vol. 16/2, pp. 159-76, http://dx.doi.org/10.1177/014662169201600206.

OCDE (à paraître), *PISA 2015 Technical Report*, PISA, Éditions OCDE, Paris.

OCDE (2014a), *Résultats du PISA 2012 : Savoirs et savoir-faire des élèves (Volume I) : Performance des élèves en mathématiques, en compréhension de l'écrit et en sciences*, PISA, Éditions OCDE, Paris, http://dx.doi.org/10.1787/9789264208827-fr.

OCDE (2014b), *PISA 2012 Technical Report*, PISA, Éditions OCDE, Paris, http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.

OCDE (2009), *PISA 2006 Technical Report*, PISA, Éditions OCDE, Paris, http://dx.doi.org/10.1787/9789264048096-en.

Oliveri, M.E. et **M. von Davier** (2014), « Toward increasing fairness in score scale calibrations employed in international Large-Scale Assessments », *International Journal of Testing*, vol. 14/1, pp. 1-21, http://dx.doi.org/10.1080/15305058.2013.825265.

Oliveri, M.E. et **M. von Davier** (2011), « Investigation of model fit and score scale comparability in international assessments », *Pyschological Test and Assessment Modeling*, vol. 53/1, pp. 315-33.

Rasch, G (1960), P*robabilistic Models for Some Intelligence and Attainment Tests*, Nielsen & Lydiche, Copenhague.

Rousseeuw, P.J. et **C. Croux** (1993), « Alternatives to the median absolute deviation », *Journal of the American Statistical Association*, vol. 88/424, pp. 1273-83, http://dx.doi.org/10.1080/01621459.1993.10476408.

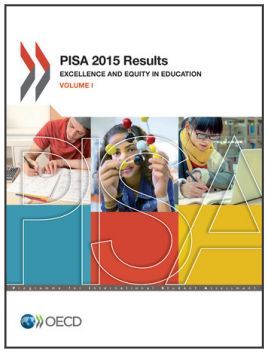
Urbach, D. (2013), « An investigation of Australian OECD PISA trend results », in *Research on PISA*, pp. 165-79, Springer, Pays-Bas, http://dx.doi.org/10.1007/978-94-007-4458-5_10.

Warm, T.A. (1989), « Weighted likelihood estimation of ability in item response theory », *Psychometrika*, vol. 54/3, pp. 427-450, http://dx.doi.org/10.1007/BF02294627.

ANNEXE A6
ÉTUDE DE L'EFFET DU MODE D'ADMINISTRATION DE L'ENQUÊTE PISA 2015 DANS LE CADRE DE L'ESSAI DE TERRAIN

Disponible en ligne uniquement.

À consulter sur : www.oecd.org/pisa



Extrait de :

PISA 2015 Results (Volume I)

Excellence and Equity in Education

Accéder à cette publication :
https://doi.org/10.1787/9789264266490-en

Merci de citer ce chapitre comme suit :

OCDE (2016), « Cadre technique de l'enquête PISA 2015 », dans *PISA 2015 Results (Volume I) : Excellence and Equity in Education*, Éditions OCDE, Paris.

DOI: https://doi.org/10.1787/9789264267534-13-fr

Cet ouvrage est publié sous la responsabilité du Secrétaire général de l'OCDE. Les opinions et les arguments exprimés ici ne reflètent pas nécessairement les vues officielles des pays membres de l'OCDE.

Ce document et toute carte qu'il peut comprendre sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région.

Vous êtes autorisés à copier, télécharger ou imprimer du contenu OCDE pour votre utilisation personnelle. Vous pouvez inclure des extraits des publications, des bases de données et produits multimédia de l'OCDE dans vos documents, présentations, blogs, sites Internet et matériel d'enseignement, sous réserve de faire mention de la source OCDE et du copyright. Les demandes pour usage public ou commercial ou de traduction devront être adressées à rights@oecd.org. Les demandes d'autorisation de photocopier une partie de ce contenu à des fins publiques ou commerciales peuvent être obtenues auprès du Copyright Clearance Center (CCC) info@copyright.com ou du Centre français d'exploitation du droit de copie (CFC) contact@cfcopies.com.

OCDE