

Pierre de Rosette Note de Synthèse

Établir une concordance entre les évaluations régionales (ERCE et PASEC) et internationales (TIMSS/PIRLS)



UNESCO

L'Acte constitutif de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO) a été adopté par 20 pays lors de la Conférence de Londres en novembre 1945 et est entré en vigueur le 4 novembre 1946.

L'objectif premier de l'UNESCO est de contribuer au maintien de la paix et de la sécurité dans le monde en resserrant, par l'éducation, la science et la culture, la collaboration entre nations afin d'assurer le respect universel de la justice, de la loi, des droits de l'homme et des libertés fondamentales pour tous, sans distinction de race, de sexe, de langue ou de religion, que la Charte des Nations Unies reconnaît à tous les peuples.

L'UNESCO a cinq fonctions principales inscrites dans son mandat : 1) des études prospectives sur l'éducation, la science, la culture et la communication pour le monde de demain ; 2) le progrès, le transfert et le partage des connaissances par des activités de recherche, de formation et d'enseignement ; 3) des actions normatives en vue de la présentation et l'adoption d'instruments internes et de recommandations réglementaires ; 4) l'expertise par le biais de la coopération technique aux États membres en faveur de leurs projets et politiques de développement ; 5) l'échange d'informations spécialisées

Institut de statistique de l'UNESCO

L'Institut de statistique de l'UNESCO (ISU) est le bureau de la statistique de l'UNESCO et le dépositaire des Nations Unies pour les statistiques mondiales en matière d'éducation, de science et de technologie, de culture et de communication. L'ISU a été créé en 1999 avec pour mission d'améliorer le programme statistique de l'UNESCO et d'élaborer et mettre à disposition des statistiques actualisées, précises et pertinentes pour les politiques, comme l'exige le contexte social, politique et économique actuel de plus en plus complexe et changeant.

Publié en 2022 par :
Institut de statistique de l'UNESCO
C.P 250 Succursale H
Montréal, Québec H3G 2K8
Canada

Courriel : uis.publications@unesco.org
<http://www.uis.unesco.org>
Réf : UIS/2022/LO/TD/16
ISBN : 978-92-9189-299-0
© UNESCO-UIS 2022



Cette publication est disponible en accès libre sous la licence Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). En utilisant le contenu de cette publication, les utilisateurs acceptent de se conformer aux conditions d'utilisation du service d'archive des publications en accès libre de l'UNESCO (<http://fr.unesco.org/open-access/terms-useccby-sa-fr>). Les appellations employées dans cette publication et la présentation des données qui y figurent n'impliquent de la part de l'UNESCO aucune prise de position quant au statut juridique de tout pays, territoire, ville ou zone ou de leurs autorités, ni quant au tracé de leurs frontières ou limites. Les idées et opinions exprimées dans cette publication sont celles des auteurs ; ils ne sont pas nécessairement ceux de l'UNESCO et n'engagent pas l'Organisation.

Conception et mise en page par: FreireCreative
Conception de la couverture par: büro Svenja

Remerciements

Le rapport d'analyse de la Pierre de Rosette fut un projet collaboratif de l'Institut de statistique de l'UNESCO (ISU). L'Association internationale pour l'évaluation des résultats scolaires (AIE) a été le partenaire technique de ce projet avec le TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College. Le soutien technique et de mise en œuvre a été fourni par CONFEMEN et LLECE.

Andres Sandoval Hernandez est l'auteur de ce rapport. Luis Crouch (RTI), Silvia Montoya (ISU), Dirk Hastedt (AIE) et Oliver Neuschmidt (AIE) ont servi d'examineurs.

Ce rapport est basé sur des recherches financées par la Fondation Bill & Melinda Gates. Les résultats et les conclusions qu'il contient sont ceux des auteurs et ne reflètent pas nécessairement les positions ou les politiques de la Fondation Bill & Melinda Gates.

Ce rapport contient des extraits des rapports techniques de Rosetta Stone fournis par le TIMSS & PIRLS International Study Center sur le développement des concordances TIMSS & PIRLS pour ERCE et PASEC. Les rapports techniques sont disponibles sur <https://timssandpirls.bc.edu/Rosetta-Stone-Reports/index.html> et toutes autres questions à ce sujet devraient être adressées au TIMSS & PIRLS International Study Center: <https://timssandpirls.bc.edu/>

INDICE

Tableaux	4
Introduction	5
Méthodes	7
A. Conception complexe des ILSA	7
B. Méthodes du projet Pierre de Rosette	8
Les tables de concordance	11
Proportion d'élèves atteignant l'ODD 4.1.1	14
Messages clés	19
Lectures complémentaires	21
Liste des acronymes	22

TABLEAUX

TABLEAU 1. Table de concordance pour les mathématiques ERCE et PASEC	12
TABLEAU 2. Table de concordance pour la lecture ERCE et PASEC	13
TABLEAU 3. Définition des référentiels ERCE, PASEC et TIMSS/PIRLS servant de MPL pour l'année d'études 6 (ERCE et PASEC) et l'année d'études 4 (TIMSS/PIRLS)	15
TABLEAU 4. Proportion d'élèves dans les pays ERCE et PASEC atteignant l'ODD 4.1.1 pour les mathématiques	17
TABLEAU 5. Proportion d'élèves dans les pays ERCE et PASEC atteignant l'ODD 4.1.1 pour la lecture.	18

INTRODUCTION¹

Alors qu'il reste moins de dix ans pour atteindre les 17 objectifs de développement durable (ODD) établis dans l'Agenda 2030, nos dirigeants mondiaux ont déclaré une Décennie d'Action et de réalisation pour le développement durable. La Décennie d'Action appelle à accélérer les solutions durables à tous les plus grands défis du monde - allant de la pauvreté et du genre au changement climatique, à l'inégalité et à l'amélioration de la qualité de l'éducation pour tous.

Ce document se concentre sur l'ODD 4, en particulier l'indicateur 4.1.1. L'ODD 4 établit que d'ici 2030, nous devons « assurer une éducation de qualité inclusive et équitable et promouvoir des opportunités d'apprentissage tout au long de la vie pour tous », et l'indicateur 4.1.1 opérationnalise cet objectif comme la demande de « garantir que toutes les filles et tous les garçons suivent un enseignement primaire et secondaire gratuit, équitable et de qualité, l'enseignement débouchant sur des résultats d'apprentissage pertinents et efficaces »². La communauté internationale a convenu de mesurer les progrès de cet objectif à partir du pourcentage d'enfants et de jeunes atteignant au moins un niveau minimal de compétences en lecture et en mathématiques à trois niveaux de la structure scolaire, et selon le sexe: (a) en 2e/3e année; (b) à la fin du primaire; et (c) à la fin du premier cycle du secondaire.

Il existe une grande variété d'évaluations aux niveaux national, régional et international qui produisent des informations sur le niveau de compétence en lecture et en mathématiques des enfants et des jeunes. Ces évaluations internationales à grande échelle (ILSA pour son sigle en anglais) dans le domaine de l'éducation ont été considérées par beaucoup comme la source d'informations la plus appropriée pour mesurer et suivre les progrès de plusieurs indicateurs de l'ODD 4. Les ILSA offrent des informations systématiques sur plus de 100 systèmes éducatifs, fournissent des informations sur le niveau de littéracie et de numératie des enfants et des jeunes et disposent de mécanismes d'assurance qualité des données sans équivalents. De plus, les ILSA produisent des données strictement comparables entre les pays et dans le temps. Cependant, chacune de ces évaluations a un cadre d'évaluation différent, est mesurée sur une échelle différente et est conçue pour éclairer la prise de décision dans différents contextes éducatifs.

Par conséquent, si des informations sont déjà disponibles, elles doivent être intégrées dans une approche d'évaluation globale. Une telle approche permettrait des comparaisons entre les systèmes éducatifs non seulement au sein d'une région géographique, mais globalement et dans le temps. Cela permettrait de tirer des enseignements des initiatives politiques de pays ayant des caractéristiques similaires (par exemple le niveau de développement économique, la structure du système éducatif) mais situés dans différentes parties du monde. Si les

1 Ce résumé analytique contient des extraits des rapports techniques Pierre de Rosette fournis par le Centre d'études international TIMSS & PIRLS sur l'élaboration des concordances TIMSS & PIRLS pour l'ERCE et le PASEC. Les rapports sont disponibles sur <https://timssandpirls.bc.edu/Rosetta-Stone-Reports/index.html>, et d'autres questions doivent être adressées à <mailto:timss@bc.edu> » timss@bc.edu.

2 Voir <https://en.unesco.org/education2030-sdg4/targets>

similitudes entre les systèmes éducatifs sont importantes pour permettre des comparaisons justes et valables, les différences contextuelles, historiques et culturelles permettent d'envisager le même problème sous différents angles et, par conséquent, de développer de nouvelles approches et solutions. Dans le même temps, des évaluations régionales et nationales, en se concentrant sur un test créé pour un contexte spécifique, facilitent le développement de comparaisons plus méticuleuses qui, à leur tour, peuvent éclairer les politiques de manière plus détaillée. De cette manière, la combinaison d'évaluations nationales, régionales et mondiales a le potentiel d'accroître la profondeur et l'étendue de nos efforts pour mesurer et suivre l'ODD 4.1.1.

Pour ces raisons, l'Institut de statistique de l'UNESCO (ISU) a dirigé un programme méthodologique qui vise à construire une stratégie de mesure robuste offrant des informations comparables pour autant de pays que possible afin de mesurer et de suivre les progrès vers l'ODD 4.1.1. Ce programme pionnier associe deux évaluations régionales et deux évaluations mondiales, reliées d'une manière qui ne nécessite en aucune manière de reconcevoir les évaluations régionales, permettant de potentiels ajustements tout en évitant de suivre une démarche descendante. C'est l'idée clé que l'UNESCO et d'autres agences ont essayé de maintenir tout au long du processus de l'ODD 4 en ne forçant pas le monde à adopter une seule évaluation.

Ce programme, mené par l'International Association for the Evaluation of Educational Achievement (IEA) et le TIMSS & PIRLS International Study Center, Lynch School of Education du Boston College, tire son nom de la célèbre découverte archéologique qui a permis la traduction entre différentes langues écrites : la Pierre de Rosette. L'étude Pierre de Rosette est conçue pour mesurer les progrès mondiaux de l'ODD 4.1.1 en reliant différents programmes d'évaluation nationaux et régionaux aux enquêtes TIMSS (*Trends in International Mathematics and Science Study*) et PIRLS (*Progress in International Reading Literacy Study*)³. L'objectif est de fournir aux pays qui ont participé aux évaluations régionales ou nationales mais pas à TIMSS et PIRLS des informations sur les proportions d'élèves du primaire qui ont atteint un niveau minimal de compétence en lecture et en mathématiques (ODD 4.1.1) qui permettent d'établir des comparaisons internationales.

Ce document présente une synthèse des premiers résultats de l'étude Pierre de Rosette, qui consiste à établir une table de concordance qui projette les distributions des scores estimés à partir de deux évaluations régionales vers les distributions sur TIMSS et PIRLS. Ces évaluations sont l'Étude comparative et explicative régionale de l'UNESCO (ERCE ; Estudio Regional Comparativo y Explicativo) dans les pays d'Amérique latine et des Caraïbes et le Programme d'Analyse des Systèmes Éducatifs de la Confemén (PASEC) dans les pays francophones de l'Afrique subsaharienne.

La suite du présent document est divisée en trois sections principales. La première section décrit les méthodes et les analyses conduites pour établir la concordance entre les évaluations régionales et TIMSS et PIRLS. Les tableaux de concordance qui en résultent et les directives pour les utiliser et les interpréter sont présentés dans la deuxième section. La troisième section montre la proportion d'élèves qui atteignent l'ODD 4.1.1 dans chacun des pays de l'ERCE et du PASEC selon les tables de concordance produites par l'étude Pierre de Rosette. La dernière section propose un ensemble de messages clés pour les décideurs politiques, les praticiens et les chercheurs.

3 Voir IEA, 2017, à la section Lectures complémentaires.

MÉTHODES

Bien qu'ils aient des cadres et des approches différents, les ILSA tels que l'ERCE, le PASEC, le TIMSS et le PIRLS partagent un ensemble de procédures et de méthodes largement similaires. Par exemple, les ILSA utilisent un échantillonnage complexe (plan d'échantillonnage stratifié, à plusieurs degrés et en grappes) et une évaluation complexe (cahiers tournants). Par conséquent, le processus de production d'une concordance entre eux est également complexe. Cette section commence par une brève description des procédures adoptées par les ILSA, suivie d'un aperçu des méthodes utilisées pour produire les tables de concordance. Il convient de noter que ce document offre une version simplifiée des deux types d'ILSA ; il est fortement recommandé de consulter les rapports techniques des ILSA correspondants et les rapports d'analyse de l'étude Pierre de Rosette pour les détails techniques complets (voir la liste des lectures à la fin de ce document).

A. Conception complexe des ILSA

L'objectif de la plupart des ILSA est de produire des informations représentatives au niveau national sur les résultats scolaires d'une population cible d'élèves dans une matière donnée. Pour ce faire, les ILSA sont confrontées à deux principaux défis pratiques. Premièrement, un test ne peut pas être administré à tous les élèves d'un système éducatif et deuxièmement, un seul élève ne peut pas répondre au test complet. Tester tous les élèves serait très coûteux et créerait de nombreuses complications logistiques, et les élèves testés ne peuvent pas répondre à tous les items car il leur faudrait entre huit et dix heures pour compléter un test couvrant tous les sujets inclus dans le cadre conceptuel de l'évaluation. Ainsi, seul un échantillon d'élèves répond au test, et chacun des élèves échantillonnés ne répond qu'à une partie du test. Ces défis sont surmontés en utilisant un modèle complexe d'échantillonnage et d'évaluation.

Le plan d'échantillonnage complexe des ILSA est généralement caractérisé par un échantillonnage en grappes stratifié à plusieurs degrés. La stratification consiste à organiser les listes d'écoles en groupes (c'est-à-dire en strates) selon certaines caractéristiques, comme leur localisation géographique ou le type d'écoles (par exemple : écoles publiques ou privées). La première étape d'échantillonnage consiste alors à sélectionner des écoles au sein de ces strates. La liste des écoles est également ordonnée selon la taille des écoles, de sorte que les plus grandes écoles (c'est-à-dire celles qui ont plus d'élèves) ont une plus grande probabilité d'être sélectionnées (probabilité proportionnelle à la taille). La deuxième étape consiste à échantillonner soit des élèves, soit des classes (c'est-à-dire des grappes) au sein des écoles échantillonnées. L'objectif étant de fournir des estimations pour l'ensemble de la population cible, ces probabilités inégales doivent être ajustées à l'aide de poids d'échantillonnage et de réplifications (d'autres méthodes, telles que les modèles multiniveaux, peuvent également être utilisées en combinaison avec les poids d'échantillonnage).

Le recours à une conception complexe du plan d'évaluation est indispensable pour les ILSA, car un grand nombre d'items est nécessaire pour évaluer avec une grande fiabilité les compétences des élèves, qui recouvrent

un vaste domaine de contenus. Pour permettre la passation de nombreux items dans un temps de test limité, la méthode dite des « cahiers tournants » est utilisée. Plus précisément, les items sont regroupés en blocs selon le contenu et les domaines cognitifs qu'ils mesurent. Les blocs sont ensuite répartis sur un certain nombre de cahiers de sorte que chaque cahier contienne deux blocs ou plus, mais qu'aucun cahier ne contienne tous les items. Le test est conçu de manière à ce que les différents cahiers soient reliés par des blocs communs d'items. En tant que tel, aucun élève ne prend tous les items et de nombreux élèves passent différents items. Cependant, un score comparable peut toujours être estimé grâce à une méthode de mise à l'échelle connue sous le nom de théorie de la réponse aux items (IRT en anglais). Grâce à l'IRT, la compétence de chaque élève et la difficulté de chaque item sont estimées. Cela permet d'estimer un score qui est comparable entre les élèves même lorsqu'ils ont répondu à différents items du test. Ensuite, pour estimer les scores, un modèle dit « de population » est utilisé. Ce modèle prend les estimations du modèle IRT et les autres informations disponibles concernant les élèves pour estimer le score que chaque élève aurait obtenu s'il avait passé tous les items du test. Étant donné que ce score est exprimé sous la forme d'une distribution de probabilité, plusieurs tirages au sort (généralement cinq) sont réalisés. Chaque tirage au sort, également connu sous le nom de valeur plausible (PV pour son sigle en anglais), est considéré comme une valeur représentative de la distribution des scores potentiels pour tous les élèves de l'échantillon qui partagent les mêmes caractéristiques et les mêmes modèles de réponses aux items (par exemple, les élèves qui ont tendance à répondre correctement aux items de même difficulté).

La complexité de l'échantillonnage ainsi que la complexité de la conception de l'évaluation impliquent une certaine incertitude (ou erreur de mesure) qui doit être prise en compte dans tous les calculs. Il est important de noter qu'aucune conception d'évaluation n'élimine complètement l'erreur de mesure. En pratique, il s'agit d'optimiser l'incertitude engendrée par l'erreur de mesure par rapport à des considérations de coût et pratiques. Cette incertitude est prise en compte dans toutes les analyses par l'utilisation de poids d'échantillonnage et de réplifications (pour le plan d'échantillonnage complexe) et par l'utilisation de modèles IRT et de méthodes PV (pour le plan d'évaluation complexe).

B. Méthodes de l'étude Pierre de Rosette

Cette section décrit les instruments et la conception de l'étude Pierre de Rosette et la procédure d'analyse suivie pour la construction des tables de concordance. L'étude Pierre de Rosette, comme la plupart des ILSA, utilise une conception d'évaluation complexe dans laquelle chaque élève ne reçoit qu'un sous-ensemble d'items de la liste. L'étude Pierre de Rosette est composée de deux parties d'évaluation. La première partie est l'évaluation de la source (c'est-à-dire ERCE ou PASEC), y compris les items cognitifs et le questionnaire de contexte. La deuxième partie est la pièce maîtresse de l'étude, l'évaluation Pierre de Rosette (cible), consistant en des cahiers de test avec des blocs d'items de TIMSS et PIRLS. Au total, huit blocs tournants d'items de mathématiques (TIMSS) et quatre passages tournants en lecture (PIRLS) ont été utilisés. Les deux parties de l'évaluation ont été administrées sous forme d'évaluations de type papier-crayon aux mêmes élèves. Chaque élève a reçu les cahiers de l'évaluation source (PASEC ou ERCE) le premier jour et un cahier Pierre de Rosette le deuxième jour.

Pour établir des tables de concordance, l'analyse des données a été réalisée en quatre étapes. Premièrement, la qualité des données a été évaluée en fonction de statistiques descriptives et de la variabilité de la non-réponse.

Deuxièmement, des modèles IRT ont été utilisés pour construire des échelles comparables ERCE, PASEC et Pierre de Rosette (c.-à-d. TIMSS et PIRLS) parmi les populations d'élèves. Troisièmement, la méthodologie des PV a été utilisée séparément pour ERCE, PASEC et Pierre de Rosette pour produire des scores en mathématiques et en lecture. Quatrièmement, des tables de concordance ont été établies sur la base des scores moyens en mathématiques et en lecture des évaluations source (ERCE et PASEC) et cible (Pierre de Rosette – TIMSS et PIRLS). L'analyse a été réalisée sur les données de trois pays du PASEC (Burundi, Guinée et Sénégal) et de deux pays de l'ERCE (Colombie et Guatemala) en utilisant les poids d'échantillonnage fournis par les équipes de l'ERCE et du PASEC au centre d'études international de TIMSS et PIRLS.

i. Évaluation de la qualité des données

La qualité des données a été évaluée à l'aide de statistiques d'items classiques. La difficulté des items a été évaluée en estimant le pourcentage d'élèves ayant répondu correctement à chaque item, et la discrimination des items en estimant la corrélation point-bisériale entre la réussite à l'item et le score total. Les résultats de cette analyse ont confirmé que les données recueillies dans le cadre de l'étude Pierre de Rosette avaient des propriétés adéquates pour passer aux étapes suivantes de l'analyse. La difficulté des items se situe généralement dans une fourchette acceptable (25 % à 95 % de réponses correctes) et est similaire pour les items ERCE, PASEC, TIMSS et PIRLS. Les valeurs de discrimination sont également similaires dans toutes les études et se situent dans des plages acceptables (coefficients de corrélation non inférieurs à 0,20), avec des valeurs typiques comprises entre 0,24 et 0,41 pour le PASEC et entre 0,36 et 0,50 pour l'ERCE.

ii. Construction des échelles comparables pour ERCE, PASEC et Pierre de Rosette (c'est-à-dire TIMSS et PIRLS)

Des modèles IRT unidimensionnels ont été utilisés pour mettre à l'échelle les items de liaison Pierre de Rosette (c'est-à-dire les items TIMSS et PIRLS) ainsi que les items PASEC et ERCE (séparément pour les mathématiques et la lecture). Les modèles IRT unidimensionnels ont montré des niveaux élevés de comparabilité (c'est-à-dire des proportions élevées d'items avec un bon ajustement) entre les pays ERCE et les échelles Pierre de Rosette (avec des valeurs allant de 91,2 % à 95,1 %), ainsi qu'entre les pays PASEC et les échelles Pierre de Rosette (avec des valeurs allant de 81,2 % à 86,3 %). Les modèles ont également montré des niveaux élevés de comparabilité entre les pays pour les échelles ERCE et PASEC, fournissant une base solide pour établir une concordance (avec des valeurs allant de 91,5 % à 98,7 %).

Après que les items PASEC, ERCE et Pierre de Rosette ont été mis à l'échelle avec des modèles IRT unidimensionnels séparés, des modèles IRT multidimensionnels ont été utilisés pour examiner à quel point les construits mesurés dans les différentes évaluations sont similaires ou différents. Plus précisément, ces modèles multidimensionnels ont été utilisés pour étudier la relation entre les échelles de mathématiques ERCE et PASEC et les échelles de mathématiques TIMSS, et entre les échelles de lecture ERCE et PASEC et les échelles de lecture PIRLS. Les valeurs des paramètres des items dans ces modèles multidimensionnels ont été fixés sur les valeurs des paramètres des items obtenues à partir de l'ensemble des modèles unidimensionnels décrits précédemment.

Les résultats des modèles IRT multidimensionnels ont indiqué que les échelles correspondantes Pierre de Rosette et PASEC d'une part, et Pierre de Rosette et ERCE d'autre part mesurent des construits qui ne sont pas identiques (comme prévu puisqu'ils ont été développés sur la base de cadres différents) mais qui sont assez similaires (avec des corrélations latentes entre les dimensions supérieures à 0,80 et inférieures à 0,90) pour permettre d'établir une concordance pertinente à travers la projection des distributions de scores. La corrélation latente varie de 0,82 à 0,90 pour l'ERCE et de 0,81 à 0,83 pour le PASEC.

iii. Calcul des scores de mathématiques et de lecture

Une fois les scores IRT estimés, un modèle de population a été utilisé pour estimer la distribution des compétences ou des PV. Ce modèle de population utilise les estimations du modèle IRT et les autres informations disponibles sur les élèves pour estimer une distribution des scores possibles que chaque élève aurait eus s'il avait passé l'ensemble du test. L'objectif final de la modélisation de la population est de déterminer des distributions de compétences à partir desquelles des PV peuvent être tirées au sort. Suivant les procédures utilisées dans TIMSS et PIRLS, cinq PV ont été tirées au sort dans la distribution des compétences pour chaque domaine et chaque élève.

iv. Construction des tables de concordance

La dernière étape consiste à construire les tables de concordance Pierre de Rosette, qui établissent une relation entre les scores des évaluations source et cible. Dans Pierre de Rosette, une gamme de scores TIMSS et PIRLS plausibles est projetée à partir des scores ERCE et PASEC en mathématiques et en lecture, respectivement. Autrement dit, les mathématiques ERCE, la lecture ERCE, les mathématiques PASEC et la lecture PASEC représentent les évaluations sources et TIMSS et PIRLS représentent les évaluations cibles.

La première étape consiste à évaluer la relation entre les données des évaluations de la source et de la cible. Les analyses de corrélation entre les scores moyens des mathématiques ERCE et TIMSS ($r = 0,80$ à $0,82$), mathématiques PASEC et TIMSS ($r = 0,70$ à $0,80$), lecture ERCE et PIRLS ($r = 0,78$ à $0,82$) et lecture PASEC et PIRLS ($r = 0,73$ à $0,81$) indiquent que les échelles ERCE et Pierre de Rosette et les échelles PASEC et Pierre de Rosette mesurent des constructions différentes mais similaires ; c'est-à-dire que les corrélations sont raisonnablement élevées pour construire une concordance.

Les scores et les niveaux de concordance ont ensuite été identifiés sur la base des scores moyens estimés du PASEC et de l'ERCE en utilisant les données combinées du Burundi, de la Guinée et du Sénégal pour le PASEC et de la Colombie et du Guatemala pour l'ERCE. Les fourchettes de scores des scores moyens des échelles mathématiques et lecture du PASEC et de l'ERCE ont été arrondies soit vers le haut soit vers le bas pour couvrir la quasi-totalité des données des pays participants, et pour être le plus symétrique possible autour de la moyenne globale (500 pour le PASEC, 700 pour l'ERCE). Pour les deux échelles PASEC, mathématiques et lecture, les scores varient d'environ 200 à 800. Pour les deux échelles ERCE, mathématiques et lecture, les scores varient d'environ 440 à 940. Dans les deux cas, ces fourchettes couvrent près de 99,5 % des données.

Pour les échelles ERCE et PASEC, 20 points ont été spécifiés comme intervalle de score pour inclure suffisamment de niveaux de score ou de compétence et pour conserver autant d'informations que possible. Par conséquent, il y a 26 niveaux de score dans la fourchette de score de 440 à 940 et 200 à 800, respectivement.

Pour chaque niveau de score de concordance identifié, un modèle prédictif d'appariement a été utilisé pour sélectionner cinq donneurs (c'est-à-dire des élèves ayant passé le test) pour chaque test et dans chaque pays, afin que chaque pays contribue de manière égale à chacune des tables de concordance. Chacun des donneurs a fait don de cinq PV sur les tests cibles. La moyenne et l'écart-type des PV des donneurs à partir des données de liaison de Pierre de Rosette ont été calculés sur la base du total des PV données à chaque niveau de concordance. Il faut noter que ces étapes ont été mises en œuvre séparément pour l'ERCE mathématiques et lecture et pour le PASEC mathématiques et lecture.

Des tables de concordance préliminaires pour les mathématiques ERCE, la lecture ERCE, les mathématiques PASEC et la lecture PASEC ont été créées en attribuant la moyenne et l'écart-type estimés de chaque ensemble de PV en fonction des données de liaison Pierre de Rosette (TIMSS et PIRLS), respectivement, à chaque niveau de score de concordance dans l'intervalle spécifié de mathématiques ERCE, de lecture ERCE, de mathématiques PASEC et de lecture PASEC. Enfin, pour chaque point de score de concordance, la moyenne des PV données a été lissée en appliquant une moyenne mobile simple, en utilisant une fenêtre de sept points de score. L'écart-type des PV de chaque point de score a été lissé de la même manière que les moyennes des PV, en utilisant une moyenne géométrique mobile des variances de chaque ensemble des cinq moyennes de PV données regroupées au niveau de score correspondant dans la table. La racine carrée de cette variance lissée devient l'écart-type conditionnel lissé.

LES TABLES DE CONCORDANCE

Les tableaux 1 et 2 montrent les tables de concordance finales pour les mathématiques ERCE/PASEC et la lecture ERCE/PASEC, respectivement. Les première et cinquième colonnes de chaque tableau indiquent les niveaux de score de concordance de l'évaluation source (ERCE ou PASEC), soit en mathématiques, soit en lecture. Les deuxième et sixième colonnes montrent la distribution prédictive, les moyennes projetées et les écarts-types conditionnels du score d'évaluation de la source sur l'échelle TIMSS ou PIRLS. Les troisième, quatrième, septième et huitième colonnes montrent les valeurs d'échelle pour des centiles de cette distribution conditionnelle compte tenu de l'évaluation de la source (ERCE ou PASEC) sur les échelles TIMSS ou PIRLS. Les valeurs inférieures et supérieures aux 68ème et 95ème centiles sont fournies.

TABLEAU 1. Table de concordance pour les mathématiques ERCE et PASEC

Score de mathématiques ERCE	Score projeté sur l'échelle TIMSS		Borne inférieure		Borne supérieure	
	Moyenne	ET	95%	68%	68%	95%
400	290	64	162	226	354	417
420	304	63	178	241	367	430
440	319	62	194	256	381	443
460	318	63	192	255	381	444
480	322	62	198	260	384	446
500	326	62	201	264	389	451
520	334	62	211	273	396	458
540	342	64	214	278	406	470
560	357	63	231	294	419	482
580	371	62	247	309	433	495
600	389	61	266	327	450	511
620	403	61	282	342	463	524
640	420	58	303	361	478	537
660	432	57	317	375	489	546
680	449	53	344	397	502	555
700	465	52	362	414	517	569
720	481	51	379	430	532	583
740	497	49	399	448	547	596
760	515	50	415	465	565	616
780	531	50	431	481	581	631
800	548	48	453	500	596	643
820	563	46	471	517	609	655
840	576	46	484	530	622	668
860	590	45	500	545	635	680
880	599	46	508	554	645	691
900	608	46	516	562	654	699
920	617	48	520	568	665	713
940	624	51	522	573	675	726
960	638	50	538	588	688	739
980	653	49	554	603	702	751

Score de mathématiques ERCE	Score projeté sur l'échelle TIMSS		Borne inférieure		Borne supérieure	
	Moyenne	ET	95%	68%	68%	95%
220	197	76	44	121	274	350
240	210	76	58	134	286	362
260	223	75	72	148	298	374
280	224	75	74	149	300	375
300	229	77	75	152	306	382
320	231	76	78	155	307	383
340	235	75	85	160	310	385
360	241	71	99	170	312	383
380	254	69	115	184	323	392
400	265	68	130	197	333	401
420	273	68	137	205	341	410
440	284	65	154	219	349	414
460	297	62	172	234	359	421
480	315	61	194	254	376	437
500	336	63	209	273	399	462
520	344	65	215	280	409	473
540	355	64	227	291	419	483
560	371	63	245	308	434	497
580	382	66	251	317	448	514
600	395	70	256	326	465	535
620	403	71	260	332	475	546
640	417	71	274	345	488	559
660	437	69	299	368	506	575
680	453	69	316	385	522	591
700	469	67	335	402	536	602
720	484	64	357	421	548	612
740	500	57	386	443	556	613
760	513	52	408	461	566	618
780	526	52	422	474	578	630
800	539	51	436	487	590	641

ET = écart-type

TABLEAU 2. Table de concordance pour la lecture ERCE et PASEC

Score de lecture ERCE	Score projeté sur l'échelle PIRLS		Borne inférieure		Borne supérieure	
	Moyenne	ET	95%	68%	68%	95%
400	284	84	116	200	369	453
420	296	83	130	213	380	463
440	309	82	144	226	391	473
460	321	81	158	240	402	483
480	329	80	169	249	409	489
500	342	77	187	265	419	497
520	347	76	195	271	424	500
540	373	74	224	298	447	521
560	378	74	229	303	452	526
580	395	68	258	326	463	532
600	410	65	279	344	475	540
620	421	62	297	359	483	545
640	436	62	311	373	498	560
660	455	63	328	392	518	581
680	467	64	338	402	531	595
700	482	60	362	422	542	602
720	488	63	361	425	551	615
740	501	65	372	437	566	631
760	522	64	393	457	586	650
780	526	63	399	463	590	653
800	536	61	414	475	597	658
820	546	60	425	486	607	667
840	555	60	436	495	614	674
860	565	59	447	506	623	682
880	568	56	456	512	624	680
900	576	56	463	519	632	688
920	588	58	473	530	646	703
940	596	59	477	537	655	715
960	608	58	492	550	666	725
980	620	57	506	563	678	735

Score de lecture ERCE	Score projeté sur l'échelle PIRLS		Borne inférieure		Borne supérieure	
	Moyenne	ET	95%	68%	68%	95%
220	146	72	2	74	218	290
240	161	72	17	89	233	304
260	175	72	31	103	247	319
280	178	72	34	106	249	321
300	181	72	38	110	253	325
320	190	71	47	118	261	332
340	196	72	52	124	267	339
360	205	71	63	134	276	347
380	216	72	73	145	288	359
400	228	72	84	156	300	372
420	238	76	87	163	314	390
440	253	74	104	179	327	401
460	265	73	120	193	338	411
480	280	71	139	209	351	422
500	297	71	155	226	369	440
520	317	73	172	244	390	462
540	330	72	186	258	402	474
560	351	66	219	285	417	482
580	364	66	232	298	430	496
600	377	68	241	309	446	514
620	392	69	255	323	461	529
640	405	67	271	338	471	538
660	420	63	295	357	483	545
680	444	66	312	378	511	577
700	456	69	319	388	525	593
720	473	71	332	402	544	615
740	486	70	346	416	555	625
760	492	72	347	420	565	637
780	507	72	362	434	579	651
800	521	72	377	449	593	665

ET = écart-type

Les tables de concordance doivent être interprétées avec prudence car elles ne sont pas des prédictions parfaites de la performance d'un élève à un test cible (par exemple, TIMSS ou PIRLS). Elles ne fournissent pas de lien direct entre les évaluations et dépendent des caractéristiques de l'échantillon. Par conséquent, l'incertitude de la prédiction doit être prise en considération lors de l'utilisation et de l'interprétation des tables de concordance. Par exemple, un score PASEC en mathématiques de 500 n'entraîne pas un score TIMSS de 336 mais une distribution des scores TIMSS avec une moyenne de 336. En d'autres termes, en supposant des distributions de scores conditionnelles approximativement normales, 68 % des PV générées sur l'échelle TIMSS se situeraient probablement dans la plage de scores de 273 à 399 (si un élève ayant des compétences similaires a passé l'évaluation TIMSS) et 95 % des PV générées sur l'échelle TIMSS se situeraient probablement entre 209 et 462, comme le montre le tableau 1.

PROPORTION D'ÉLÈVES ATTEIGNANT L'ODD 4.1.1

Au-delà des déductions qu'il est possible de faire concernant la plage de scores probables obtenue sur les échelles TIMSS ou PIRLS à partir d'un score ERCE ou PASEC, il est également possible de générer les PV probables pour les élèves eux-mêmes (qui ont participé à ERCE ou PASEC mais pas à TIMSS ou PIRLS) sur les échelles TIMSS et PIRLS en utilisant les moyennes projetées et les écarts-type des tables de concordance. Les rapports d'analyse Pierre de Rosette fournissent des instructions détaillées sur la manière de générer des PV aléatoires à l'aide des tables de concordance. Ces PV projetées peuvent ensuite être utilisées pour estimer le pourcentage d'élèves atteignant un seuil donné, par exemple le niveau de compétence minimum global (MPL, minimum proficiency level) pour la fin du primaire, qui marque la réalisation de l'ODD 4.1.1.

Il convient de noter que les critères de référence ERCE et PASEC servant de MPL sont censés correspondre au MPL mondial et, par conséquent, indiquer la réalisation de l'ODD 4.1.1. Cependant, ils ont été fixés indépendamment par les équipes techniques de chaque évaluation. Par conséquent, ils sont qualitativement différents les uns des autres. Les descriptions de ce que les élèves savent et peuvent faire lorsqu'ils atteignent les MPL fixés pour l'ERCE et le PASEC, ainsi que pour le MPL mondial (fixé par un consensus d'experts internationaux et aligné sur les références TIMSS/PIRLS⁴) se trouve dans le tableau 3.

4 Voir ACER, 2019 et UIS, 2022 à la section Lectures complémentaires.

TABLEAU 3. Définition des référentiels ERCE, PASEC et TIMSS/PIRLS servant de MPL pour l'année d'études 6 (ERCE et PASEC) et l'année d'études 4 (TIMSS/PIRLS)

Étude	Mathématiques	Lecture
ERCE	<p>Niveau 3 (789 points)</p> <p>Les élèves de ce niveau ont montré qu'ils étaient capables de:</p> <ul style="list-style-type: none"> • Résoudre des problèmes plus complexes nécessitant l'interprétation d'informations et impliquant deux opérations ou plus, dont la multiplication ou la division. • Interpréter le sens des variations proportionnelles dans des situations contextualisées. • Identifier des fractions équivalentes (avec un dénominateur autre que 10) et calculer des additions et des soustractions de fractions avec le même dénominateur. • Relier les nombres décimaux à des fractions simples ou à des nombres fractionnaires simples (par exemple avec le dénominateur 2) et calculer ou estimer des additions et des soustractions de nombres décimaux. • Déterminer les termes intermédiaires manquants d'une séquence présentée dans une situation contextualisée, en interprétant son schéma de formation. • Identifier les relations de perpendicularité et de parallélisme dans le plan. • Résoudre des problèmes complexes impliquant le calcul ou l'estimation d'aires et de périmètres de figures géométriques. • Résoudre des problèmes impliquant des mesures (mesures de volume de masse et de temps) et convertir des unités de mesure. • Résoudre des problèmes nécessitant la lecture et l'interprétation d'informations à partir de tableaux et de graphiques ou l'identification de graphiques représentant des informations fournies dans différents formats. 	<p>Niveau 3 (754 points)</p> <p>Les élèves de ce niveau ont montré qu'ils étaient capables de:</p> <ul style="list-style-type: none"> • Faire des inférences à partir de liens entre des idées spécifiques ou secondaires et situées dans différentes parties d'un ou plusieurs textes. • Dédire le thème central d'un paragraphe ou d'une partie du texte, en établissant une relation avec le texte dans son ensemble. • Faire des inférences (ex. inférer le thème central, les caractéristiques ou les sentiments des personnages, le conflit et le dénouement) qui nécessitent de comprendre le texte globalement et d'intégrer les idées implicites qui y sont présentes. • Interpréter des expressions en langage figuré à partir d'indices implicites dans le texte ou qui mettent à l'épreuve les connaissances de l'élève sur le monde qui l'entoure. • Relier des informations visuelles et verbales dans un texte.
PASEC	<p>Niveau 3 (609 points)</p> <p>Les élèves maîtrisent la chaîne verbale (compter jusqu'à 60 en deux minutes) et sont capables de lire des chiffres, de comparer des nombres, de compléter des suites de nombres et de réaliser des opérations (additions et soustractions) sur des nombres supérieurs à cinquante. Ils peuvent raisonner sur des problèmes basiques avec des nombres inférieurs à 20.</p>	<p>Niveau 4 (595 points)</p> <p>Lecteur intermédiaire : vers une lecture autonome pour comprendre des phrases et des textes. Les élèves ont atteint un niveau de déchiffrage de l'écrit et de compréhension orale qui leur permet de comprendre des informations explicites dans des phrases et des textes courts. Ils sont capables de croiser leurs compétences de décodage et leur maîtrise du langage oral pour restituer le sens littéral d'un texte court.</p>
TIMSS/PIRLS	<p>TIMSS intermédiaire (475 points)</p> <p>Les élèves peuvent appliquer des connaissances mathématiques de base dans des situations simples. Ils peuvent calculer avec des nombres entiers à trois et quatre chiffres dans une variété de situations. Ils ont une certaine compréhension des nombres décimaux et des fractions. Les élèves peuvent identifier et dessiner des formes avec des propriétés simples. Ils peuvent lire, étiqueter et interpréter des informations dans des graphiques et des tableaux.</p>	<p>PIRLS Bas (400 points)</p> <p>Lorsqu'ils lisent principalement des textes littéraires plus simples, les élèves peuvent :</p> <ul style="list-style-type: none"> • Localiser et récupérer des informations, des actions ou des idées explicitement énoncées. • Faire des inférences directes sur les événements et les raisons des actions. • Commencer à interpréter les événements de l'histoire et les idées centrales. <p>Lorsqu'ils lisent principalement des textes informatifs plus simples, les élèves peuvent :</p> <ul style="list-style-type: none"> • Localiser et reproduire des informations explicites à partir de textes et d'autres formats (par exemple, tableaux, diagrammes). • Commencer à faire des inférences directes sur les explications, les actions et les descriptions.

Les proportions d'élèves des pays ERCE et PASEC qui atteignent l'ODD 4.1.1 pour les mathématiques (seuil intermédiaire dans TIMSS) selon ERCE (niveau de performance 3) et PASEC (niveau de performance 3), ainsi que selon les tables de concordance produites par l'étude Pierre de Rosette, sont présentées dans le tableau 4. Les proportions d'élèves dans les pays ERCE et PASEC qui atteignent l'ODD 4.1.1 pour la lecture (seuil bas dans PIRLS) selon ERCE (niveau de performance 3) et PASEC (niveau de performance 4), ainsi que selon les tables de concordance produits par l'étude Pierre de Rosette, sont présentées dans le tableau 5. Les erreurs standard de ces pourcentages sont calculées à l'aide de poids appropriés et de formules PV pour inclure les erreurs de mesure et d'échantillonnage (voir le Rapport d'analyse Pierre de Rosette pour plus de détails).

Il est important de noter que les pourcentages estimés sur la base de Pierre de Rosette sont dans de nombreux cas considérablement différents de ceux rapportés sur la base des scores PASEC et ERCE. Dans la plupart des cas, les pourcentages sont plus élevés lorsque les estimations sont basées sur Pierre de Rosette pour l'ERCE et plus faibles pour le PASEC. Ces écarts pourraient être dus à des différences dans les cadres d'évaluation et les construits mesurés dans chacun. Comme mentionné ci-dessus, les MPL fixés par chaque évaluation sont sensiblement différents. Par exemple, alors que l'ERCE considère que le MPL est atteint lorsque les élèves peuvent «interpréter des expressions en langage figuré à partir d'indices implicites dans le texte», le PASEC considère que le MPL est atteint lorsque les élèves peuvent « [...] combiner leurs compétences en décodage et leur maîtrise de la langue orale pour saisir le sens littéral d'un court passage » (voir tableau 3).

En outre, le PASEC a été développé dans et pour le contexte des pays francophones d'Afrique subsaharienne, tandis que l'ERCE a été développé dans et pour le contexte des pays d'Amérique latine. D'autre part, TIMSS et PIRLS ont été conçus comme des initiatives mondiales. Bien que ces évaluations mesurent à la fois les mathématiques et la lecture, la définition même des construits varie entre eux ; ils ont des points de départ différents et fixent des aspirations différentes pour leurs systèmes éducatifs.

Enfin, les populations cibles des évaluations sont différentes : 6e année pour le PASEC et l'ERCE et 4e année pour le TIMSS et le PIRLS. D'autres facteurs pourraient également jouer un rôle important dans l'explication de ces différences, comme le moment où la transition des langues locales vers la langue d'évaluation (par exemple le français ou l'espagnol) est introduite dans le système scolaire. Dans de nombreux pays, cette transition se produit en 1ère année, tandis que dans d'autres, elle n'a lieu qu'en 4e année. À ce stade, il s'agit là d'hypothèses non vérifiées, et davantage de recherches (et des échantillons plus importants de pays et d'élèves participant à Pierre de Rosette) sont nécessaires pour les tester et fournir des réponses claires.

TABLEAU 4. Proportion d'élèves dans les pays ERCE et PASEC atteignant l'ODD 4.1.1 pour les mathématiques

Pays	Élèves au-dessus du MPL Total		Élèves au-dessus du MPL Garçons		Élèves au-dessus du MPL Filles	
	%	ES	%	ES	%	ES
ERCE (6^{ème} année) MPL						
Argentine	13.1	(0.9)	14.2	(1.1)	12.0	(0.9)
Brésil	28.7	(1.5)	30.5	(1.7)	26.7	(1.7)
Colombie	16.5	(1.2)	17.6	(1.4)	15.3	(1.7)
Costa Rica	20.8	(1.7)	22.6	(1.9)	19.0	(1.9)
Cuba	20.6	(1.5)	20.5	(1.6)	20.8	(1.7)
République Dominicaine	2.1	(0.5)	2.1	(0.5)	2.1	(0.6)
Equateur	22.8	(1.4)	23.0	(1.5)	22.4	(1.6)
El Salvador	7.4	(0.6)	8.0	(0.7)	6.8	(0.8)
Guatemala	6.6	(0.9)	7.6	(1.2)	5.6	(0.8)
Honduras	11.1	(1.9)	11.9	(2.0)	10.4	(1.9)
Mexique	37.9	(1.3)	36.9	(1.6)	38.9	(1.7)
Nicaragua	3.0	(0.5)	3.8	(0.7)	2.3	(0.5)
Panama	3.3	(0.4)	3.4	(0.5)	3.2	(0.5)
Paraguay	5.6	(0.7)	5.8	(0.8)	5.3	(0.9)
Pérou	38.7	(1.5)	38.3	(1.8)	39.2	(1.8)
Uruguay	37.8	(1.5)	38.1	(1.8)	37.5	(1.7)
PASEC (6^{ème} année) MPL						
Bénin	19.3	(2.6)	18.9	(2.4)	19.8	(3.0)
Burkina Faso	25.2	(1.2)	26.5	(1.5)	24.1	(1.4)
Burundi	18.2	(1.5)	23.7	(1.8)	13.8	(1.5)
Cameroun	11.2	(1.0)	10.5	(1.0)	12.1	(1.4)
Congo	7.8	(0.9)	6.9	(1.0)	8.6	(1.3)
Côte d'Ivoire	2.6	(0.6)	2.9	(0.7)	2.3	(0.7)
Gabon	23.1	(1.9)	26.5	(2.4)	19.9	(2.0)
Guinée	6.9	(1.2)	7.5	(1.4)	6.2	(1.4)
Madagascar	6.3	(1.7)	5.7	(1.7)	7.0	(1.9)
Niger	8.0	(1.3)	8.1	(1.5)	7.8	(1.4)
RDC	3.2	(1.1)	3.4	(1.3)	3.0	(1.0)
Sénégal	27.6	(1.9)	27.8	(2.2)	27.4	(2.2)
Tchad	1.9	(0.6)	1.8	(0.6)	2.0	(0.7)
Togo	16.0	(1.2)	16.1	(1.4)	15.8	(1.3)
Pierre de Rosette (TIMSS 4^{ème} année) MPL						
	42.7	(1.4)	44.2	(1.6)	41.2	(1.6)
	56.9	(1.5)	58.2	(1.9)	55.6	(1.8)
	48.2	(1.8)	49.4	(2.1)	46.9	(2.1)
	56.4	(1.6)	57.9	(2.2)	54.9	(1.9)
	42.2	(1.9)	42.0	(2.0)	42.7	(2.1)
	22.2	(1.4)	20.9	(1.4)	23.5	(2.0)
	53.1	(1.5)	53.1	(1.7)	53.0	(1.7)
	35.7	(1.0)	36.5	(1.7)	34.9	(1.6)
	30.1	(1.2)	32.4	(1.5)	27.8	(1.5)
	38.5	(2.2)	40.4	(2.3)	36.7	(2.6)
	66.1	(1.5)	64.9	(1.7)	67.3	(1.6)
	30.2	(1.2)	33.0	(1.8)	27.5	(1.5)
	25.2	(1.0)	25.4	(1.3)	25.0	(1.2)
	27.1	(1.3)	27.8	(1.6)	26.4	(1.6)
	66.3	(1.3)	66.1	(1.6)	66.6	(1.7)
	65.8	(1.4)	66.6	(2.0)	65.0	(1.9)
Pierre de Rosette (TIMSS 4^{ème} année) MPL						
	8.9	(1.4)	8.8	(1.5)	9.0	(1.7)
	10.2	(0.8)	10.9	(0.9)	9.5	(1.1)
	7.8	(0.7)	9.6	(1.0)	6.4	(0.8)
	5.1	(0.8)	5.0	(0.9)	5.1	(1.0)
	3.7	(0.5)	3.5	(0.7)	3.9	(0.7)
	1.5	(0.4)	1.6	(0.4)	1.4	(0.5)
	8.8	(1.1)	9.8	(1.8)	7.9	(1.1)
	3.1	(0.7)	3.6	(0.9)	2.5	(0.6)
	3.0	(0.7)	2.9	(0.8)	3.0	(0.8)
	3.5	(0.7)	3.6	(0.8)	3.4	(0.8)
	1.9	(0.7)	2.1	(0.9)	1.6	(0.6)
	12.0	(1.3)	12.6	(1.5)	11.5	(1.3)
	1.2	(0.3)	1.3	(0.4)	1.0	(0.5)
	7.0	(0.6)	7.0	(0.7)	6.9	(0.7)

MPL = niveau de compétence minimum ; ES = erreur standard

TABLEAU 5. Proportion d'élèves dans les pays ERCE et PASEC atteignant l'ODD 4.1.1 pour la lecture

Pays	Élèves au-dessus du MPL Total		Élèves au-dessus du MPL Garçons		Élèves au-dessus du MPL Filles	
	%	ES	%	ES	%	ES
ERCE (6^{ème} année) MPL						
Argentine	31.7	(1.2)	28.8	(1.3)	34.5	(1.4)
Brésil	43.3	(1.5)	40.2	(1.7)	46.7	(1.8)
Colombie	37.3	(1.7)	35.2	(2.2)	39.5	(2.0)
Costa Rica	53.8	(1.7)	51.1	(2.0)	56.4	(1.9)
Cuba	44.4	(1.3)	38.6	(1.6)	50.4	(1.6)
République Dominicaine	16.3	(1.1)	12.1	(1.0)	20.6	(1.5)
Equateur	26.0	(1.0)	24.1	(1.2)	27.9	(1.3)
El Salvador	29.2	(1.2)	26.3	(1.3)	32.3	(1.6)
Guatemala	15.9	(1.0)	15.5	(1.4)	16.3	(1.1)
Honduras	16.1	(1.5)	14.0	(1.5)	18.1	(1.9)
Mexique	41.4	(1.4)	37.8	(1.8)	45.0	(1.6)
Nicaragua	12.8	(0.8)	11.7	(0.8)	14.0	(1.1)
Panama	17.4	(1.1)	15.3	(1.1)	19.4	(1.2)
Paraguay	18.7	(1.1)	15.8	(1.2)	21.5	(1.4)
Pérou	48.8	(1.2)	44.6	(1.5)	53.4	(1.5)
Uruguay	43.5	(1.3)	40.3	(1.5)	46.7	(1.5)
PASEC (6^{ème} année) MPL						
Bénin	45.5	(2.7)	44.0	(2.6)	47.3	(3.1)
Burkina Faso	33.0	(1.4)	32.6	(1.5)	33.3	(1.6)
Burundi	4.5	(1.0)	4.7	(1.0)	4.4	(1.1)
Cameroun	30.3	(1.7)	28.2	(1.7)	32.7	(2.2)
Congo	33.6	(1.8)	30.2	(1.9)	37.1	(2.2)
Côte d'Ivoire	22.1	(1.9)	21.0	(2.2)	23.3	(1.9)
Gabon	76.3	(1.8)	73.7	(2.1)	78.9	(1.9)
Guinée	22.2	(1.7)	21.9	(1.9)	22.6	(2.1)
Madagascar	6.3	(2.0)	5.4	(1.8)	7.1	(2.4)
Niger	14.4	(1.6)	13.1	(1.8)	15.9	(1.7)
RDC	9.2	(1.5)	9.9	(1.4)	8.5	(1.8)
Sénégal	41.1	(2.2)	38.2	(2.3)	43.5	(2.7)
Tchad	7.6	(1.2)	7.8	(1.2)	7.4	(1.4)
Togo	19.4	(1.1)	18.5	(1.3)	20.2	(1.3)
Pierre de Rosette (PIRLS 4^{ème} année) MPL						
	78.9	(1.1)	77.6	(1.3)	80.2	(1.3)
	86.2	(0.8)	84.4	(1.1)	88.1	(1.0)
	83.7	(1.1)	82.3	(1.5)	85.1	(1.5)
	90.6	(0.9)	89.8	(0.9)	91.5	(1.3)
	86.6	(1.1)	83.6	(1.6)	89.6	(1.1)
	65.7	(1.2)	61.1	(1.7)	70.4	(1.6)
	76.5	(1.1)	75.9	(1.5)	77.1	(1.2)
	80.2	(1.1)	78.3	(1.3)	82.1	(1.3)
	67.0	(1.4)	67.0	(2.0)	66.9	(2.3)
	72.1	(1.5)	71.7	(1.8)	72.5	(1.6)
	84.1	(0.9)	83.0	(1.1)	85.1	(1.1)
	71.4	(1.2)	70.4	(1.7)	72.5	(1.5)
	68.7	(1.4)	67.2	(1.7)	70.0	(1.5)
	69.7	(1.3)	66.9	(1.6)	72.4	(1.6)
	85.7	(1.0)	84.5	(1.4)	87.1	(1.4)
	85.3	(0.9)	83.7	(1.0)	87.0	(1.3)
Pierre de Rosette (PIRLS 4^{ème} année) MPL						
	37.4	(2.2)	36.2	(2.5)	38.7	(2.8)
	27.8	(1.0)	27.8	(1.5)	27.7	(1.5)
	9.9	(0.8)	10.5	(0.9)	9.4	(1.0)
	25.0	(1.3)	24.0	(1.5)	26.1	(1.8)
	28.0	(1.5)	25.6	(1.9)	30.5	(1.9)
	19.3	(1.4)	19.2	(1.7)	19.5	(1.6)
	56.7	(1.8)	55.2	(2.3)	58.3	(2.2)
	19.5	(1.6)	19.3	(1.9)	19.7	(1.9)
	8.1	(1.2)	7.1	(1.3)	9.1	(1.4)
	13.4	(1.2)	12.6	(1.5)	14.4	(1.5)
	11.0	(1.2)	11.9	(1.4)	10.0	(1.4)
	34.2	(1.9)	32.9	(2.3)	35.3	(2.3)
	8.8	(1.1)	9.2	(1.3)	8.2	(1.5)
	17.5	(0.9)	16.7	(1.2)	18.2	(1.3)

MPL = niveau de compétence minimum; ES = erreur standard

MESSAGES CLÉS

- Le cadre des ODD demande explicitement aux systèmes nationaux de suivre la réalisation de leurs objectifs à l'aide d'indicateurs comparables, c'est-à-dire d'indicateurs qui peuvent être comparés entre les pays et dans le temps. Cependant, les organismes des Nations Unies ont été réticents à imposer un cadre de mesure spécifique de manière descendante. Outre les enjeux humanistes et politiques, une telle tâche exigerait du temps et des ressources excessives de la part de toutes les parties prenantes. Par ailleurs, de nombreuses théories du développement économique et des finances publiques suggèrent de minimiser les solutions trop standardisées. Un compromis optimal a consisté à développer des échelles pour les outils de mesure existants et de les rendre comparables a posteriori (ex-post), par des moyens analytiques et empiriques. C'est ce que Pierre de Rosette a fait, en gardant toujours à l'esprit que les contextes nationaux restent différents (tant les pays d'Afrique subsaharienne que ceux d'Amérique latine) et que les comparaisons mondiales sont parfois limitées. À notre connaissance, il s'agit d'un projet pionnier, sans doute le premier du genre dans le domaine de la mesure des apprentissages. Des efforts similaires, mais quelque peu différents, de l'ISU et d'autres partenaires sont également en cours.
- Des échelles comparables ont été établies entre les pays et entre les évaluations. Les résultats de Pierre de Rosette indiquent que, d'un point de vue statistique, les construits en lecture et en mathématiques mesurés par PASEC et ERCE sont suffisamment similaires aux construits mesurés par TIMSS et PIRLS pour permettre la concordance souhaitée. Ces tables de concordance constituent la première étape vers la compréhension des différences entre les ILSA, l'exercice devant s'étendre en termes de pays et d'échantillons d'écoles afin d'établir des inférences robustes. Les tables de concordance incluses dans ce document peuvent être utilisées pour faire des déductions sur la fourchette probable de scores obtenus sur les échelles TIMSS ou PIRLS à partir d'un score PASEC ou ERCE, permettant aux pays de comparer les résultats de leurs élèves à l'échelle mondiale et fournissant des données qui les aident à mesurer les progrès mondiaux vers l'ODD 4.1.1.
- Les résultats de l'étude Pierre de Rosette pour l'ERCE et le PASEC suggèrent que des tables de concordance peuvent être établies pour d'autres évaluations régionales (par exemple SAQMEC, SEA-PLM, PIL-NA). Les informations qui en résulteraient ainsi que davantage d'analyses seraient extrêmement utiles et précieuses car elles permettraient à toutes les évaluations régionales de se comparer non seulement au TIMSS et au PIRLS, mais aussi entre elles.
- Les informations fournies par Pierre de Rosette et les évaluations régionales comme le PASEC et l'ERCE doivent être considérées comme complémentaires. Les évaluations régionales et Pierre de Rosette jouent un rôle irremplaçable dans la stratégie mondiale de mesure et de suivi des progrès vers l'ODD 4.1.1. Ensemble, ils renforcent les possibilités d'approfondir et d'élargir l'analyse qui peut être effectuée et, par conséquent, améliorent la qualité et la pertinence des informations disponibles pour les décideurs et les autres parties prenantes.

- Pierre de Rosette ouvre des possibilités infinies d'analyses secondaires qui peuvent aider à améliorer les rapports mondiaux et faciliter les analyses comparatives des systèmes éducatifs du monde entier. Chaque évaluation reflète le programme d'études de la région/du pays concerné, tandis que la répartition des niveaux de compétence est adéquate pour les rapports mondiaux.
- Les tables de concordance présentées ici sont basées sur seulement deux pays pour l'ERCE et trois pays pour le PASEC. Pour les pays qui n'ont pas participé à l'étude Pierre de Rosette (c'est-à-dire qui n'ont pas administré les cahiers de liaison TIMSS/PIRLS), l'utilisation des tables de concordance fournies dans ce document est une extrapolation et s'accompagne d'une incertitude supplémentaire qui ne peut être entièrement prise en compte sans effectuer également une collecte de données Pierre de Rosette. Cependant – et on ne saurait trop insister sur ce point – cette approche est meilleure que toute alternative existante.
- L'augmentation de la taille des échantillons nationaux et l'ajout de plus de pays par évaluation régionale amélioreraient encore la concordance estimée et permettraient de mener des recherches pour expliquer les différences observées dans le pourcentage d'élèves atteignant l'ODD 4.1.1 lorsqu'il est estimé avec Pierre de Rosette par rapport à ERCE ou PASEC.
- Une réflexion plus approfondie sur l'établissement des niveaux de compétence minimaux pour chaque étude régionale, qui correspondent le mieux au niveau de compétence global convenu est nécessaire pour assurer des comparaisons plus justes des pourcentages d'élèves qui atteignent l'ODD 4.1.1 dans chaque système éducatif.
- Si les pays souhaitent établir une concordance entre leur évaluation nationale et TIMSS et PIRLS, l'approche Pierre de Rosette pourrait également être menée. Cependant, les coûts par pays seraient relativement élevés car les résultats ne seraient applicables qu'à un seul pays et les instruments de liaison devraient être adaptés à chaque évaluation nationale.
- Bien que l'ajout d'évaluations nationales à Pierre de Rosette puisse être une option pour les pays qui ne sont pas encore inclus dans une évaluation régionale, en raison de la complémentarité des évaluations régionales et de Pierre de Rosette discutée ci-dessus, il est clair que le fait d'avoir plus de pays participant aux deux évaluations régionales (par exemple PASEC et ERCE) et mondiales (par exemple, TIMSS et PIRLS) constituerait une stratégie beaucoup plus efficace.

LECTURES COMPLÉMENTAIRES

Conseil australien pour la recherche en éducation (ACER) (2019). Niveaux de compétence minimum décrits, déballés et illustrés. GAML/REF/2.

<https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/GAML6-REF-2-MLP-recommendations-ACER.pdf>

IEA (Association internationale pour l'évaluation du rendement scolaire) (2017). Mesurer l'ODD 4 : comment le PIRLS peut aider ; Comment les progrès de l'étude internationale sur la compréhension de l'écrit (PIRLS) aident à suivre les cibles de l'objectif de développement durable 4. Paris: Unesco. (ED/IPS/IGE/2017/05).

<https://unesdoc.unesco.org/ark:/48223/pf0000260607>

IEA (Association internationale pour l'évaluation du rendement scolaire) (2017). MO Martin, IVS Mullis et M. Hooper (eds), Methods and Procedures in PIRLS 2016. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College et AIE.

<https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>

IEA (Association internationale pour l'évaluation du rendement scolaire) (2020). MO Martin, M. von Davier et IVS Mullis (eds), Methods and Procedures: TIMSS 2019 Technical Report. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College et IEA.

<https://timssandpirls.bc.edu/timss2019/methods>

Khorramdel, L., L. Yin, P. Foy, J. Yoon Jung, U. Bezirhan et M. von Davier (2022). Établir une concordance entre ERCE et TIMSS/PIRLS. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College et International Association for the Evaluation of Educational Achievement (IEA).

<https://timssandpirls.bc.edu/Rosetta-Stone-Reports/index.html>

Khorramdel, L., L. Yin, P. Foy, J. Yoon Jung, U. Bezirhan et M. von Davier (2022). Établir une concordance entre PASEC et TIMSS/PIRLS. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College et International Association for the Evaluation of Educational Achievement (IEA).

<https://timssandpirls.bc.edu/Rosetta-Stone-Reports/index.html>

Programme d'Analyse des Systèmes Éducatifs (PASEC) (2017). Rapport technique de l'évaluation internationale PASEC 2014. Dakar : PASEC, Conférence des Ministres de l'Éducation des États et Gouvernements de la Francophonie (CONFEMEN).

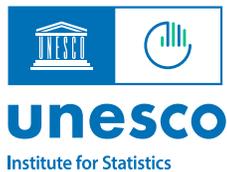
<https://www.pasec.confemen.org/wp-content/uploads/2017/12/RAPPORT-TECHNIQUE-PASEC2014.pdf>

UNESCO (2016). Tercer Estudio Regional Comparativo y Explicativo: Reporte Técnico. Santiago: UNESCO.
<https://unesdoc.unesco.org/ark:/48223/pf0000247123>

Institut de statistique de l'UNESCO (ISU) (2022), Metadata for Global Indicator 4.1.1
<https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2021/03/Minimum-Proficiency-Levels-MPLs.pdf>.

LISTE DES ACRONYMES

ERCE	Etude régionale comparative et explicative / <i>Estudio regional comparativo y explicativo</i>
IEA	International Association for the Evaluation of Educational Achievement
ILSA	Évaluations internationales à grande échelle (<i>International Large-scale Assessment</i>)
IRT	Théorie de la réponse aux items (<i>Item Response Theory</i>)
MPL	Niveau de compétence minimal
PASEC	Programme d'analyse des systèmes éducatifs de la Confemen
PIRLS	<i>Progress in International Reading Literacy Study</i>
PV	Valeur plausible (<i>Plausible Value</i>)
SAQMEC	Consortium d'Afrique australe et orientale pour le suivi de la qualité de l'éducation
ET	Écart-type : mesure de la dispersion des données par rapport à la moyenne
ES	Erreur standard : écart-type de la moyenne d'échantillons répétés d'une population
ODD	Objectifs de développement durable
SEA-PLM	Southeast Asia Primary Learning Metrics
TIMSS	Trends in International Mathematics and Science Study
ISU	Institut de statistique de l'UNESCO



TIMSS & PIRLS
International Study Center
Lynch School of Education
BOSTON COLLEGE

BILL & MELINDA
GATES *foundation*