



unesco

Institute for Statistics
Global Alliance to
Monitor Learning

PROTOCOLE D'ETABLISSEMENT DES RAPPORTS SUR L'INDICATEUR 4.1.1

Mars 2022

Acronymes et abréviations

ACER	<i>Australian Council for Educational Research</i> (Conseil australien pour la recherche en éducation)
BIE	Bureau international de l'éducation
CITE	Classification internationale type de l'éducation
CMC	Cadre mondial de contenus
DSC	Descripteurs des seuils de compétence
EGMA	Évaluation des compétences fondamentales en mathématiques
EGRA	Évaluation des compétences fondamentales en lectures
ENA	Évaluation nationale de l'apprentissage
ERCE	Étude régionale comparative et explicative
ETN	Évaluations transnationales
GAML	Alliance mondiale pour le suivi de l'apprentissage
IEA	Association internationale pour l'évaluation du rendement scolaire
ISU	Institut de statistique de l'UNESCO
LLECE	Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation
MICS	Enquête d'indicateurs multiples par conglomérats
MILO	Suivi des impacts sur les résultats d'apprentissage
NAEP	<i>National Assessment of Educational Progress</i> (Évaluation nationale des acquis de l'apprentissage des États-Unis)
PAL Network	<i>Network People's Action for Learning Network</i> (Réseau d'action du peuple pour l'apprentissage)
PASEC	Programme d'analyse des systèmes éducatifs de la CONFEMEN
OAC	Outil d'alignement de contenus
OAP	Outil d'alignement de procédures
OCDE	Organisation de coopération et de développement économiques
ODD	Objectifs de développement durable
PILNA	Évaluation des compétences en lecture, écriture et calcul dans les îles du Pacifique
PIRLS	<i>Programme international de recherche en lecture scolaire</i>
PISA	<i>Programme international pour le suivi des acquis des élèves</i>
PISA4D	PISA pour le développement
SACMEQ	Consortium d'Afrique australe et orientale pour le pilotage de la qualité de l'éducation
SEA-PLM	Métrie de l'apprentissage à l'école primaire en Asie du Sud-Est
SEAMEO	Organisation des ministres de l'Éducation d'Asie du Sud-Est
SMC	Seuil minimal de compétence
TIMSS	Tendances de l'étude internationale sur les mathématiques et les sciences
TRI	Théorie des réponses aux items

Table des matières

Acronymes et abréviations	2
1. OBJECTIFS ET STRUCTURE	6
2. LES DÉFIS	6
3. COHÉRENCE DES RAPPORTS	7
4. CADRE MONDIAL DE CONTENU	8
4.1 Pourquoi ?	8
4.2 Produits	8
4.3 Résultats escomptés :	9
5. ALIGNEMENT DES PROCÉDURES	9
5.1 Pourquoi ?	9
5.2 Comment ?	9
5.3 Produits	9
5.4 Résultats escomptés	9
6. CADRE DE COMPÉTENCE ET SEUIL MINIMAL, STRATÉGIES DE RACCORDEMENT ET ÉTABLISSEMENT DE RAPPORTS INTERMÉDIAIRES	10
6.1 Pourquoi ?	10
6.2 Comment ?	10
6.3 Produits	10
6.4 Produits et résultats escomptés	11
7. Établissement de rapports intermédiaires	11
8. Protocole d'établissement du rapport de l'indicateur ODD 4.1.1 pour la publication des données de l'ISU	12
8.1 Alignement sur les seuils minimaux de compétences :	13
8.2 Source de publication lorsque plusieurs sources sont disponibles	13
8.3 Mappage de l'année d'études et du point de mesure	14
1. 2e ou 3e année : plus une année lorsque l'enseignement primaire dure plus de 4 ans selon les niveaux de la CITE dans le pays, à l'exception de la 4e année TIMSS/PIRLS, qui correspond à la fin de l'enseignement primaire lorsque celui-ci dure six ans ou moins.	14
2. Fin de l'enseignement primaire : Plus ou moins un an à partir de la dernière année d'enseignement primaire, à l'exception de l'année 4 de l'enquête TIMSS/PIRLS, qui correspond à la fin de l'enseignement primaire lorsque celui-ci dure six ans ou moins, selon la cartographie des niveaux de la CITE dans le pays .	14
3. Fin de l'enseignement secondaire inférieur : Plus deux ou moins un de la dernière année du premier cycle de l'enseignement secondaire, selon la cartographie des niveaux de la CITE dans le pays.	14
8.4 Mappage des domaines à la lecture ou les mathématiques	14
8.5 Protocole d'établissement des rapports au moyen d'une ENA	15
8.6 Indicateurs de l'indice de parité (Indicateur ODD 4.5.1)	15
8.7 Résumé du protocole d'établissement des rapports	16
Annexe : Méthodologies alternatives de raccordement	18
Étalonnage pédagogique ou « modération sociale »	18
L'approche statistique	19

Liste des tableaux

Tableau 1. Récapitulatif du processus et axe du GAML.....	8
Table 2. L'établissement de rapports intermédiaires en bref.....	12
Table 3. Établissement du rapport 2019 de l'indicateur 4.1.1	17
Table 4. Bilan de l'état d'avancement et de l'exécution	22

Liste des figures

Figure 1. Indicateur 4.1.1.....	13
Figure 2. Illustration simplifiée de la méthodologie d'étalonnage pédagogique	18
Figure 3. Illustration simplifiée de la méthodologie de raccordement par tests	20
Figure 4. Illustration simplifiée de la méthodologie de raccordement par items.....	21

Résumé analytique

4.1.1 Proportion des enfants et des jeunes : (a) en 2^{ème} ou 3^{ème} année, (b) à la fin du cycle primaire, et (c) à la fin du premier cycle du secondaire qui possèdent les aptitudes minimales en (i) lecture et (ii) en mathématiques, par sexe.

Le format d'établissement des rapports de l'indicateur visé à communiquer deux informations :

- I. Le pourcentage d'élèves qui satisfont les normes minimales de compétence dans les domaines pertinents (mathématiques et lecture) à chaque point de mesure (2^{ème} ou 3^{ème} années d'études du cycle primaire, à la fin du cycle primaire et à la fin du premier cycle du secondaire); et
- II. Quand peut-on considérer que différents programmes sont comparables, et dans quelles conditions peut-on considérer qu'un pourcentage est comparable au pourcentage communiqué par un autre pays ?

L'élaboration de l'indicateur nécessite les intrants suivants :

- Quel contenu faut-il mesurer, et quel est le pourcentage de couverture qu'une évaluation donnée doit couvrir pour la rendre comparable à d'autres ?
- Quelles sont les procédures adéquates pour garantir la qualité des données collectées ? et
- Une échelle de compétence qui permettrait de renseigner toutes les évaluations (et sa fonction de conversion du seuil minimal ou la procédure d'alignement) et une définition du seuil minimal de chaque domaine qui permettrait d'estimer le pourcentage d'élèves qui atteignent le seuil minimal de compétence.

Le programme idéal d'établissement des rapports sur l'ODD 4.1.1 devra suivre trois étapes : cadre conceptuel, cadre méthodologique et cadre d'établissement des rapports. Chaque étape comporte plusieurs sous-étapes complexes. Avant d'accepter la responsabilité de devenir le dépositaire des rapports sur l'ODD 4.1.1, l'ISU a achevé la plupart des travaux relatifs à différents niveaux et types d'évaluation.

Reconnaissant que beaucoup de travaux ont déjà été réalisés, l'ISU privilégie ce qui n'a pas été fait et incite les autres à faire de même. Le tableau ci-dessous, et ce document en général, résume les travaux réalisés à ce jour. Ils sont indiqués dans la deuxième colonne du tableau. Cette note de synthèse examine l'axe des travaux de l'Institut de statistique de l'UNESCO (ISU) (deuxième colonne), et les colonnes de droite, et elle présente le protocole qui guide les modalités actuelles de mise en œuvre de l'établissement des rapports de l'ODD 4.1.1.

1. Objectifs et structure

Le document expliquera le flux des travaux, les activités et les produits dans le contexte du programme de travail plus large sur l'indicateur 4.1.1 de l'Alliance mondiale pour le suivi de l'apprentissage (GAML). Nous les présentons dans un ordre logique plutôt que chronologique.

Chaque activité et résultat contribue à la conception d'outils visant à garantir un niveau minimal de cohérence aux rapports établis par les systèmes éducatifs sur l'ODD 4.1.1, en préservant la flexibilité nécessaire aux systèmes éducatifs pour poursuivre les programmes d'évaluation adaptés à leur contexte et à leurs besoins.

Le format d'établissement des rapports visé à communiquer deux informations :

- I. Le pourcentage d'élèves qui satisfont les niveaux de compétences minimaux dans les domaines pertinents (mathématiques et lecture) à chaque point de mesure (2^{ème} ou 3^{ème} années d'études du cycle primaire, à la fin du cycle primaire et à la fin du premier cycle du secondaire) ; et
- II. Quand peut-on considérer que différents programmes sont comparables, et dans quelles conditions peut-on considérer qu'un pourcentage est comparable au pourcentage communiqué par un autre pays ?

L'élaboration de l'indicateur nécessite les intrants suivants :

- Quel contenu faut-il mesurer, et quel est le pourcentage de couverture qu'une évaluation donnée doit couvrir pour la rendre comparable à d'autres ?
- Quelles sont les procédures adéquates pour garantir la qualité des données collectées ? et
- Une échelle de compétence qui permettrait de renseigner toutes les évaluations (et sa fonction de conversion du seuil minimal ou la procédure de raccordement) et une définition du seuil minimal de chaque domaine qui permettrait d'estimer le pourcentage d'élèves qui atteignent le seuil minimal de compétence.

La section suivante expose les défis. La section 3 approfondit le contexte et définit la logique du flux des travaux. Les sections 4, 5 et 6 approfondissent encore chaque étape du processus en appliquant la même logique et le même format. Les sections 7 et 8 décrivent le protocole d'établissement des rapports de l'indicateur à l'intention de l'ISU.

2. Les défis

Les défis à relever pour assurer la cohérence des rapports établis à l'échelle mondiale dépassent de loin la définition des indicateurs. Dans bien des cas, il n'existe pas de « guichet unique » ou de source d'information unique pour assurer la cohérence d'un indicateur spécifique dans tous les contextes internationaux. Même lorsqu'il existe un accord sur la métrique à utiliser pour l'établissement des rapports, un processus d'harmonisation peut néanmoins être nécessaire pour assurer la cohérence de la couverture des données.

Deux extrêmes existent : au moins en théorie, la plus grande confiance se traduirait par l'utilisation d'un programme d'évaluation en parfaite adéquation pour établir les rapports alors que, toujours

en théorie, la plus grande flexibilité se traduirait par un alignement minimal pour établir les rapports. Les deux extrêmes sont insatisfaisants pour des raisons trop complexes à expliquer dans ce document. L'ISU a adapté une approche intermédiaire : offrir une flexibilité pour établir les rapports, mais en renforçant l'alignement et la comparabilité au fil du temps, sans jamais atteindre les deux extrêmes que sont une évaluation en parfaite adéquation ou plusieurs évaluations. Ainsi, un programme d'évaluation en avance, qui suivra certains guides de comparabilité ainsi que certains guides d'assurance qualité et de procédure, pourra établir des rapports dans les domaines pertinents. La flexibilité de cette approche implique l'élaboration d'outils pour guider les travaux des pays, outils qui, s'ils sont complétés par des activités de développement des capacités, garantiront que l'établissement des rapports de l'indicateur 4.1.1 stimulera le partage des connaissances et le renforcement des capacités mondiales en matière d'utilisation des programmes d'évaluation comme leviers d'amélioration du système.

3. Cohérence des rapports

L'objectif est de définir les critères et de concevoir les outils qui pourraient servir de points de référence, d'outils de transparence et de références normatives.

Les outils à concevoir ont le potentiel de devenir la norme au regard de laquelle les pays, les régions, les institutions, les organismes internationaux et les professionnels évalueront leurs programmes et certificats et établiront des comparaisons internationales, s'ils le souhaitent. Ce processus se déroule déjà de manière informelle de multiples façons et/ou il est intégré *de facto* dans les différentes évaluations internationales (et nationales).

Le flux de travaux est conçu suivant la structure de mise en œuvre de toute évaluation d'apprentissage. Le **Tableau 1** récapitule les domaines pertinents des travaux et situe le contexte des travaux (réalisés et en cours) relatifs aux trois étapes principales d'élaboration de la méthode d'établissement des rapports de l'ODD 4. Ce tableau approfondit et détaille le contexte de la documentation liminaire présentée à ce jour, et la dernière colonne indique l'axe des travaux actuels de l'ISU et de ses partenaires. La colonne 2 est identique à celle du premier tableau ci-dessus.

Tableau 1. Récapitulatif du processus et axe du GAML

Phase / Outils	Points abordés	Principaux éléments	Axe des travaux de l'ISU
Cadre conceptuel	Quoi/qui évaluer : concept Qui évaluer ? Population : scolarisée et non scolarisée ? Quelles informations contextuelles collecter ?	<ul style="list-style-type: none"> • Domaine et sous-domaine : couverture minimale • Population cible • Questionnaire général 	Cadre Mondial de Contenu (CMC) Outil d'Alignement de Contenus (OAC)
Cadre méthodologique	Quelles sont les procédures d'intégrité des données ?	<ul style="list-style-type: none"> • Conception de l'enquête • Cadre d'échantillonnage • Conception opérationnelle • Production des données • Analyse des données 	Conseils de bonnes pratiques Outil d'alignement de procédures
Cadre d'établissement des rapports	Quel format pour établir les rapports ? Quel est le seuil minimal ? Comment raccorder ou « harmoniser » ?	Modèle d'établissement des rapports <ul style="list-style-type: none"> • Échelle ou cadre de compétences • Raccordement • Définition d'une stratégie d'établissement de rapports intermédiaires 	Cadre de compétences et seuil minimal Stratégies de raccordement Stratégie d'établissement de rapports intermédiaires

Source : Institut de statistique de l'UNESCO (ISU)

4. Cadre mondial de contenu

Cette section décrit plus en détail les travaux à réaliser ou en cours pour la ligne 1 de la colonne 3 du tableau 1 ci-dessus.

1.1 Pourquoi ?

Les cadres conceptuels des programmes d'évaluation sont différents. Par exemple, selon le curriculum des pays, la couverture du contenu d'une année d'études donnée est généralement différente dans les domaines pertinents. Les domaines peuvent être définis différemment. Parfois, les programmes évaluent des compétences différentes, utilisent des contenus différents pour évaluer le même domaine, et ce de manière différente, même pour la même année d'études.

Pour évaluer le degré d'alignement de différentes évaluations et commencer à jeter les bases d'une comparaison mondiale, l'ISU et le Bureau International de l'Éducation (BIE-UNESCO) ont élaboré ensemble un Cadre Mondial de Contenu (CMC) relatif aux domaines de la lecture et des mathématiques.

1.2 Produits

Il y a trois produits finis :

1. Le Cadre Mondial de Contenu (CMC) des [Mathématiques](#) et de la [Lecture](#) pour servir de référence ;

2. L'[Outil d'alignement de contenus](#) (OAC) y compris les critères d'alignement ;
3. Une [plateforme](#) pour aider les pays à s'auto-évaluer.

1.3 **Résultats escomptés :**

Garantir l'intégrité des données en matière de comparabilité.

5. Alignement des procédures

Cette section décrit plus en détail les travaux à réaliser ou en cours pour la ligne 2 de la colonne 3 du tableau 1 ci-dessus.

1.4 **Pourquoi ?**

La fiabilité et la cohérence des activités et des procédures sont des composantes essentielles de toute évaluation à grande échelle, afin de maximiser la qualité des données et de minimiser l'impact des variations de procédure. Les évaluations internationales à grande échelle, ainsi que de nombreuses évaluations régionales à grande échelle, fournissent des exemples de normes de procédure, qui ont pour but de veiller à la cohérence de la procédure dans les contextes internationaux. De nombreuses évaluations nationales établissent également clairement des directives de procédure pour appuyer la cohérence en matière d'opérationnalisation.

La mise en œuvre d'une évaluation nécessite la prise de nombreuses décisions d'ordre méthodologique, notamment les décisions relatives aux formats du test et à l'échantillonnage. Il n'est pas nécessaire d'appliquer des procédures et des formats identiques à toutes les évaluations. Cependant, un ensemble minimal de procédures est requis pour protéger l'intégrité des données et assurer la fiabilité et la comparabilité raisonnable des résultats dans un pays donné dans le temps, mais aussi entre les pays à un point donné dans le temps.

1.5 **Comment ?**

En définissant des procédures minimales qui garantissent que l'intégrité des données est suffisante pour établir les rapports et comparer les résultats de différents programmes d'évaluation.

1.6 **Produits**

1. [Principes des bonnes pratiques en évaluation des apprentissages](#)
2. [Guide rapide : établir le bien-fondé d'une évaluation de l'apprentissage](#)
3. [Guide rapide : mettre en œuvre une évaluation nationale de l'apprentissage](#)
4. [Outil d'alignement de procédures](#)
5. [Plateforme en ligne de l'outil d'alignement de procédures](#)

1.7 **Résultats escomptés**

Des données comparables du point de vue des procédures.

6. Cadre de compétence et seuil minimal, stratégies de raccordement et établissement de rapports intermédiaires

1.8 Pourquoi ?

Cette section décrit plus en détail les travaux à réaliser ou en cours pour la **ligne 3** de la **colonne 3 du tableau 1 ci-dessus**.

Les programmes utilisent généralement différentes échelles pour établir les rapports. L'analyse des résultats reste limitée à un test particulier, à une méthodologie et à une échelle. Bien que les méthodologies des évaluations internationales et régionales aient tendance à converger, il reste difficile de situer les évaluations des résultats d'apprentissage de chaque niveau et chaque domaine sur un cadre commun de références.

La question la plus importante pour définir une échelle est celle des références ou niveaux de compétence inscrits dans l'échelle numérique et leurs points de découpage sur cette échelle. Ces références sont généralement associées à des descripteurs des seuils de compétence, qui décrivent de façon assez détaillée les compétences caractéristiques des élèves à tous les points de découpage de l'échelle. Normalement, la déclaration de politique générale ou la définition de la politique donnent un sens à la succession de notes de passage et de seuils de compétence, mais plus important, elles définissent ce qui constitue le seuil *minimal* de compétence (demandé par les indicateurs de l'ODD 4.1.1) qui a trait au contenu.¹

1.9 Comment ?

Une échelle sur laquelle tous les programmes d'évaluation de l'apprentissage pourraient être situés et la définition d'une stratégie de raccordement à cette échelle. La définition d'une échelle implique :

- Une métrique qui est arbitraire.
- La définition d'un ensemble de seuils de compétence ou d'une référence incluant le seuil minimal.
- Les déclarations politiques associées aux ensembles de références.

1.10 Produits

Les produits finis sont :

- [Une échelle pour chaque domaine et point de mesure](#) (références et définition du seuil minimal de compétence ou de chaque domaine et point de mesure).
- [Un portefeuille de stratégies de raccordement](#) et les outils qui permettent de situer les *seuils* de compétence des évaluations *sur une échelle*.
- [L'établissement de rapports intermédiaires](#) sur la stratégie et le protocole.

¹ Extrait de la [National Assessment of Educational Progress \(NAEP\)](#) sur la déclaration : « *Les définitions politiques sont des déclarations générales qui donnent un sens aux niveaux.* »

1.11 Produits et résultats escomptés

Une échelle de compétence qui comprend la définition des niveaux de performance requis des élèves pour être compétents, la définition du nombre de niveaux de performance, de déterminer les étiquettes, et de rédiger les descriptions des niveaux de la métrique de compétence.²

7. Établissement de rapports intermédiaires

Définition de l'activité : proposer une stratégie d'établissement de rapports jusqu'à l'achèvement de l'alignement des contenus et des procédures.

Champ d'application : l'ISU a défini une stratégie d'établissement de rapports intermédiaires qui repose sur la vision de la stratégie à long terme d'établissement des rapports de l'ISU.

Actuellement, l'ISU accepte toutes les données des évaluations internationales et nationales, en ajoutant des notes de bas de page et des qualificatifs qui précisent la provenance des données et aident les utilisateurs à comprendre les limitations de ces données.

Table 2. Évaluations actuellement utilisées pour les rapport

Évaluation	2 ^{ème} ou 3 ^{ème} années d'études du cycle primaire	Fin du cycle primary	Fin du premier cycle du secondary
LLECE/UNESCO	X	X	
PASEC	X	X	
PILNA	X	X	
PIRLS	X	X	
PISA			x
PISA-D			x
SACMEQ IV		X	
SEA-PLM		X	
TIMSS 4eme années- Math	X	X	
TIMSS 8eme années- Math			x
Évaluations nationales sur mesures.	Soumis à approbation statistique. Ou un lien de politique solide. Sur une base provisoire si un niveau de compétence a été identifié dont le descripteur est aligné sur la MPL mondiale		
Modules qui mesurent un seul du niveau de compétence			
MICS6**	x*		
MILO		X	
PAL Network**	X	X	
EGRA.EGMA**	Uniquement si le niveau aligné est administré sur le terrain et soumis à représentativité nationale		

* Provisoire et sous réserve d'alignement avec le Minimum Niveaux du Competences (MNC) et d'évaluation de toutes les caractéristiques d'évaluation

** Basé sur la Population

² L'élaboration initiale de l'échelle de compétence s'appuiera sur l'avis d'experts et leur analyse des données existantes et des descripteurs au niveau politique.

Table 2.a Considérations de mappage

Seuil minimal de compétence	Tel que défini par le global minimum proficiency level
Évaluations liées à la 2 ^e ou 3 ^e année d'étude	Évaluations administrées en deuxième ou troisième année selon le niveau CITE du pays plus un an lorsque le primaire dure plus de 4 ans, à l'exception de TIMSS/PIRLS grade 4, qui sont cartographiées à la fin du primaire lorsque celui-ci dure six ans ou moins.
Évaluations liées à la fin du primaire	Dernière année d'études selon la CITE tel que défini. Plus une ou moins une année d'études selon le niveau CITE de chaque pays.
Évaluations liées à la fin du second cycle de l'éducation secondaire	Dernière année/année d'études du premier cycle du secondaire selon le niveau CITE du pays Plus deux / moins une année d'études pour la dernière année du premier cycle du secondaire.

Note : TIMSS/PIRLS 4^{ème} année d'études : ces résultats sont affectés à la fin du primaire quand dans un pays donné, selon les niveaux de la Classification Internationale Type de l'Éducation (CITE), il existe 6 années d'études ou moins au primaire. Quand le primaire a plus de 6 années d'études, on leur affecte les 3^{ème}/4^{ème} années d'études. * seulement s'ils peuvent établir des rapports au niveau demandé.

EGMA : Évaluation des compétences fondamentales en mathématiques

EGRA : Évaluation des compétences fondamentales en lectures

LLECE : Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation

MILO : Suivi des impacts sur les résultats d'apprentissage

MICS 6 : Enquête de Multiples Indicateurs par Conglomerats

PAL Network: *People's Action for Learning* (Réseau d'action du peuple pour l'apprentissage)

PASEC : Programme d'analyse des systèmes éducatifs de la CONFEMEN

PILNA : Évaluation des compétences en lecture, écriture et calcul dans les îles du Pacifique

PIRLS : Programme international de recherche en lecture scolaire

PISA : Programme international pour le suivi des acquis des élèves

PISA D : PISA pour le développement

SACMEQ : Consortium d'Afrique australe et orientale pour le pilotage de la qualité de l'éducation

SEA-PLM : Les mesures d'apprentissage primaire en Asie du Sud-Est

TIMSS : Tendances de l'étude internationale sur les mathématiques et les sciences

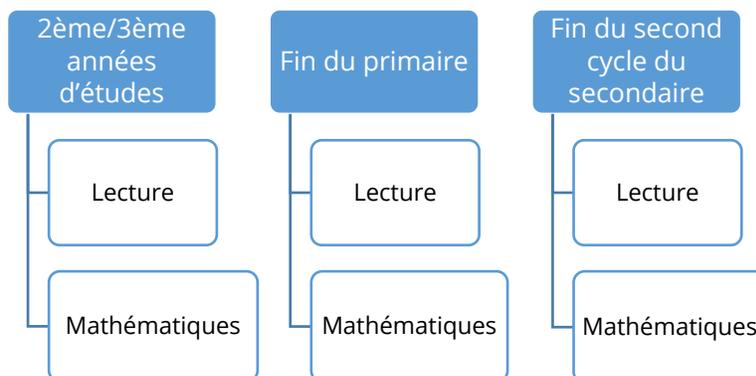
Source : Institut de statistique de l'UNESCO (ISU)

8. Protocole d'établissement du rapport de l'indicateur ODD 4.1.1 pour la publication des données de l'ISU

L'indicateur 4.1.1 a trois points de mesure et deux domaines donnant lieu à 6 sous-indicateurs comme indiqué dans la figure ci-dessous.

L'indicateur 4.1.1 comprend le 6 indicateurs suivants :

Figure 1. Indicateur 4.1.1



Source : Institut de statistique de l'UNESCO (ISU)

8.1 **Alignement sur les seuil minimal de compétences :**

Pour chaque évaluation, l'alignement du programme doit être effectué en utilisant :

- CNA : selon l'alignement convenu de la réunion de consensus et sa révision
- Évaluation nationale des apprentissages (NLA) : à condition que le pays n'ait pas de CNA. Utilisez le lien statistique ou le lien politique.

8.2 **Source de publication lorsque plusieurs sources sont disponibles**

Pour établir les rapports mondiaux de tous les indicateurs énumérés dans la liste ci-dessous, il convient de donner la priorité aux sources de données sélectionnées en respectant l'ordre suivant des évaluations, sous réserve qu'un mappage ait guidé la première sélection de sources.

- Évaluations internationales
- Évaluations régionales
- Évaluations nationales

L'évaluation qui établit la meilleure correspondance entre les années d'études sera toujours l'option privilégiée. Tant que la mise en équivalence n'est pas terminée, une évaluation transnationale des rapport serait la seconde option privilégiée, car elle permet la comparaison avec d'autres pays. L'évaluation nationale sera choisie que si aucun autre programme de rapports n'est disponible dans le délai requis.

Voici un exemple pour illustrer le processus de sélection : les résultats d'apprentissage de l'indicateur ODD 4.1.1.a-Lecture du Honduras ont été produits par des évaluations régionales et internationales de l'apprentissage.

Année	Sources de données
2011	PIRLS 2011 (4 ^{ème} année d'études)
2013	ERCE 2013 (3 ^{ème} année d'études)
2016	Évaluations nationales de l'apprentissage (ENA) (3 ^{ème} année d'études)

Les données qui doivent être utilisées pour établir les rapports sur l'ODD 4.1.1.a - Lecture du Honduras doivent être prises dans ERCE 2013. Plus précisément, PIRLS cible la 4^{ème} année d'études, soit une année d'études de plus que l'année d'études visée par l'indicateur (2^{ème}/3^{ème} année d'études). L'évaluation ERCE et l'ENA évaluent toutes deux les élèves de 3^{ème} année d'études, mais ERCE l'emporte sur l'ENA conformément à l'ordre de priorité des sources d'information indiquées ci-dessus, car cela facilite la comparaison entre les différents pays qui utilisent le même outil (ERCE). PIRLS devrait être utilisé pour informer sur la fin du primaire. Cependant à la fin du primaire, le Honduras dispose d'informations plus récentes issues d'une évaluation transnationale (ERCE). Par conséquent, ERCE devrait être utilisé pour le rapport de 6^e année.

Jusqu'à l'achèvement du processus visant à mettre en adéquation les évaluations internationales, régionales et nationales, il est important de n'utiliser qu'une seule source d'information de façon à pouvoir suivre les progrès sur une base comparable. La restriction est un niveau et un domaine. Un pays peut utiliser différentes sources pour établir les rapports d'une année donnée et de niveaux différents (par ex. : ERCE pour les 2^{ème}/3^{ème} année d'études et PISA/TIMSS pour le premier cycle du secondaire).

8.3 *Mappage de l'année d'études et du point de mesure*

1. **2e ou 3e année : plus une année lorsque l'enseignement primaire dure plus de 4 ans selon les niveaux de la CITE dans le pays, à l'exception de la 4e année TIMSS/PIRLS, qui correspond à la fin de l'enseignement primaire lorsque celui-ci dure six ans ou moins.**
2. **Fin de l'enseignement primaire : Plus ou moins un an à partir de la dernière année d'enseignement primaire, à l'exception de l'année 4 de l'enquête TIMSS/PIRLS, qui correspond à la fin de l'enseignement primaire lorsque celui-ci dure six ans ou moins, selon la cartographie des niveaux de la CITE dans le pays .**
3. **Fin de l'enseignement secondaire inférieur : Plus deux ou moins un de la dernière année du premier cycle de l'enseignement secondaire, selon la cartographie des niveaux de la CITE dans le pays.**

8.4 *Mappage des domaines à la lecture ou les mathématiques*

1. Lecture :
 - a. Le pays n'a aucune évaluation de la lecture. Les domaines alternatifs, comme la langue ou l'écriture, doivent être utilisés si la lecture n'est pas disponible pour établir les rapports.

- b. Différentes langues peuvent être utilisées pour les rapports ? Quand les résultats sont disponibles dans des langues différentes, la langue officielle ou la plus pertinente du pays doit être utilisée.
2. Mathématiques : les domaines alternatifs peuvent aussi être envisagés et utilisés pour établir les rapports.

8.5 **Protocole d'établissement des rapports au moyen d'une ENA**

Pour établir les rapports de l'ODD 4.1.1 en utilisant une ENA, les pays doivent s'assurer de la conformité aux principes suivant :

- a. L'ENA utilise la Théorie des Réponses des Items (TRI).
- b. Les résultats sont disponibles en pourcentage d'élèves par seuils de compétence.
- c. Le descripteur d'un seuil de compétence est aligné sur le [seuil minimal mondial de compétence](#)
- d. Le contenu de l'ENA couvre suffisamment le Cadre mondial de contenu dans le domaine pertinent (lecture et/ou mathématiques). L'outil d'alignement de contenus permet ce mappage et les directives sont disponibles dans le rapport de l'[OAC](#).
- e. L'ENA se conforme au niveau minimum des bonnes pratiques. Comme indiqué précédemment, l'outil d'alignement de procédures permet aux pays d'évaluer le niveau de conformité avec un ensemble de bonnes pratiques et, plus important, si elles atteignent le niveau minimal. Un rapport est rédigé à la fin de la procédure. (Rapport [OAP](#))
- f. Des notes de bas de page sont ajoutées aux points de données (par ex. : nom de l'évaluation nationale, seuil minimal de compétence, et niveau d'études).

8.6 **Indicateurs de l'indice de parité (Indicateur ODD 4.5.1)**

Les indicateurs de l'indice de parité doivent être calculés en utilisant la [formule de l'indice ajusté de parité](#).

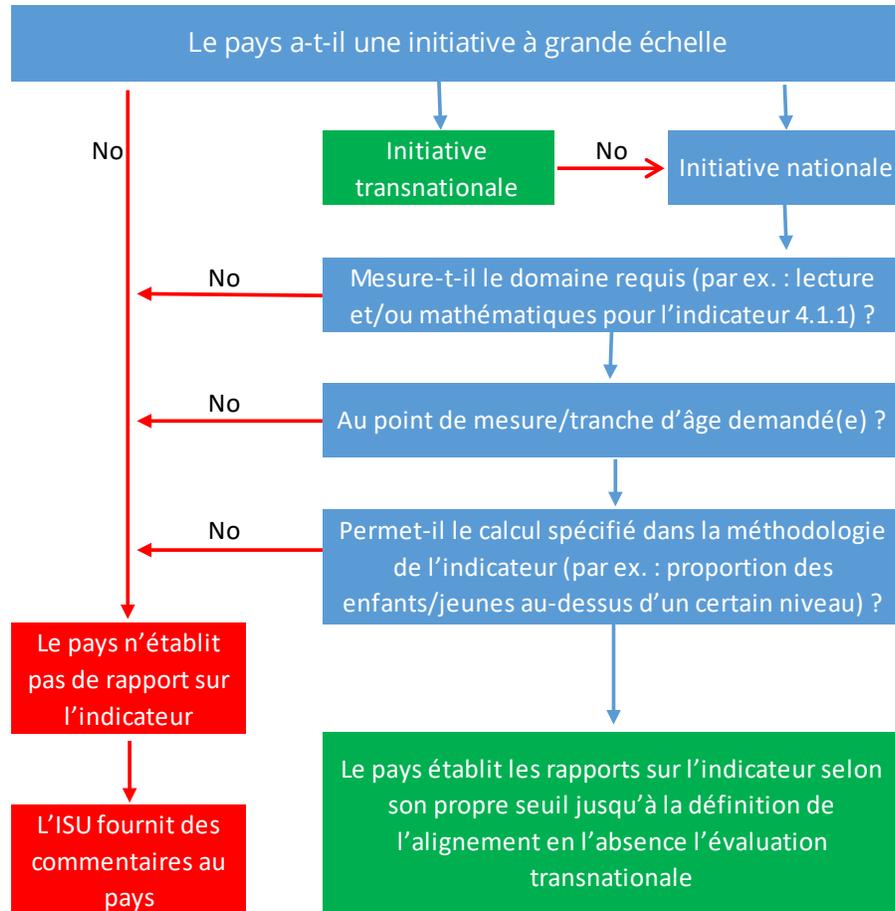
8.7 **Résumé du protocole d'établissement des rapports**

Table 3. Établissement du rapport 2019 de l'indicateur 4.1.1

Évaluation	Sources de données	Alignement au SMC mondial	Priorité à l'établissement des rapports par niveau d'enseignement	Notes de bas de page
Évaluations internationales : PISA, PISA D, TIMSS/PIRLS	Oui	Selon le consensus dégagé lors de la réunion	1	[Nom et année de l'évaluation]
Évaluations régionales : LLECE, PASEC, SACMEQ, PILNA, MILO, SEA-PLM	Oui	Selon le consensus dégagé lors de la réunion	2	[Nom et année de l'évaluation]
Évaluations nationales de l'apprentissage (ENA)	Oui	Proposer et valider l'alignement	3	« Évaluations nationales de l'apprentissage (ENA) : » [nom évaluation] et ; « Niveau d'études » [numéro niveau d'études] et ; Seuil minimal de compétence : « [nom ou numéro du SMC utilisé pour le rapport] »
EGRA/EGMA	Oui	Selon le consensus dégagé lors de la réunion	4	[Nom et année de l'évaluation]
MICS	Oui	Selon le consensus dégagé lors de la réunion	4	[Nom et année de l'évaluation]
PAL NETWORK (par ex. : ASER, UWEZO, etc.)	Oui	Selon le consensus dégagé lors de la réunion	4	[Nom et année de l'évaluation]

Source : Institut de statistique de l'UNESCO (ISU)

Annexe : Méthodologies alternatives de raccordement

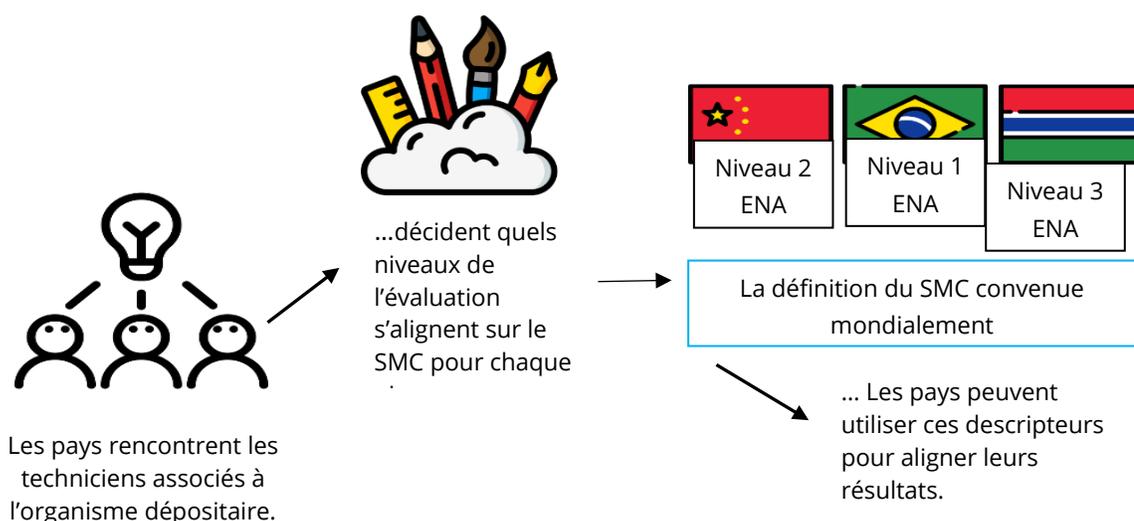
Étalonnage pédagogique ou « modération sociale »

Quoi ? Cette approche implique d'utiliser les définitions des Descripteurs de Seuil de compétence (DSC) inscrits dans une évaluation donnée, basés sur leurs propres cadres de compétence et items, pour identifier les descripteurs de compétence alignés sur la définition du seuil minimal mondial de compétence et les items de l'échantillon, provenant de différentes évaluations considérées par les experts comme se situant d'un côté ou de l'autre du seuil minimal mondial de compétence. L'alignement des DSC (celui de l'évaluation alternative à raccorder au seuil minimal mondial de compétence) est guidé par l'avis des experts, mais suivant un processus rigoureux de convergence entre les experts.

Pour plus de précision, nous proposons que la majorité des items et la richesse des descripteurs se situent autour du seuil minimal mondial. Ainsi, dans un sens, la plupart des « soutiens empiriques se situeraient seulement à un point de rupture (deux DSC), mais davantage de DSC pourraient être utilisés, afin de donner le contexte et d'accroître l'utilité pour les pays (spécialement lorsque certains pays atteignent le point le plus bas ou le point le plus haut si un seul point de rupture est utilisé).

Comment ? Sur le plan opérationnel, pour chaque évaluation (nationale ou internationale), un groupe de 8 à 10 experts d'une matière se réuniront à l'occasion d'un atelier socialement modéré. Durant l'atelier, les experts exprimeront des avis personnels et indépendants sur chaque item du test pour définir l'alignement et fixer les notes de passage initiales en se basant sur leur compréhension des descripteurs de seuil de compétence, l'expérience des populations d'élèves et leur méthode d'alignement et d'itération jusqu'à l'atteinte du niveau souhaité de convergence.

Figure 2. Illustration simplifiée de la méthodologie d'étalonnage pédagogique



Note : Les drapeaux des pays sont utilisés uniquement à des fins d'illustration et n'impliquent pas nécessairement la participation de ces pays au projet.

Source : Institut de statistique de l'UNESCO (ISU)

Quels niveaux d'enseignement ? L'étalonnage pédagogique pourrait s'appliquer à tous les niveaux d'enseignement, mais c'est la méthode qui est préférée pour aligner deux points de données dans le primaire, dans la mesure où la méthode psychométrique pure de raccordement est peu faisable dans les petites classes et que de nombreuses régions n'ont pas d'évaluations internationales et régionales à la fin du primaire.

L'approche statistique

A. Réétalonnage en menant un test parallèle sur un échantillon représentatif d'élèves

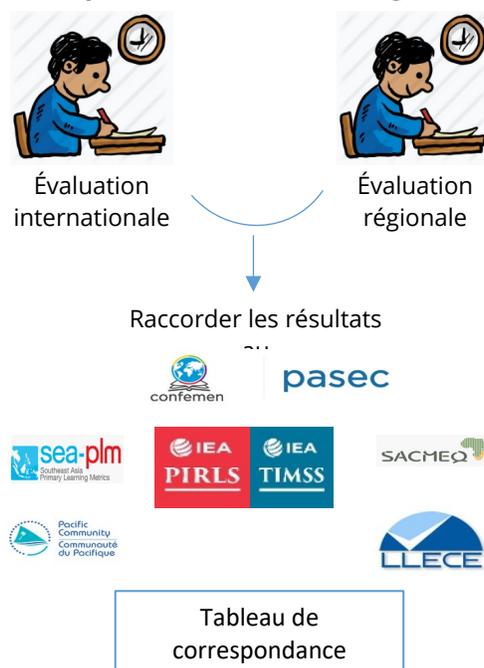
Quoi ? L'approche de raccordement par tests s'appuie sur la participation des pays aux évaluations transnationales (au niveau mondial et régional). La méthodologie permet d'exprimer deux évaluations, une internationale et l'autre régionale, sur la même échelle (et d'aligner cette échelle sur une échelle mondiale, ou de devenir de facto, la clé de l'échelle mondiale). L'alignement sera possible à travers la création de tableaux de correspondance des résultats qui permettront de traduire les résultats des évaluations régionales menées par les pays sur les échelles de rendement scolaire de TIMSS et PIRLS.

Comment ? Les élèves passeront deux évaluations, une internationale et l'autre régionale, et leurs résultats aux deux tests seront alignés de manière à établir un lien entre les résultats des évaluations régionales menées dans le primaire et les critères internationaux de référence de TIMSS et PIRLS pour la lecture, l'écriture et le calcul.

Quels niveaux d'enseignement ? On pourrait le faire à la fin du cycle primaire lorsque les évaluations régionales et mondiales sont administrées et que les capacités et l'intérêt existent à l'échelle internationale. Cela permettrait de tirer parti des seuils de compétence définis par l'Association internationale pour l'évaluation du rendement scolaire (IEA) pour ses évaluations : Tendances de l'étude internationale sur les mathématiques et les sciences (TIMSS) et Programme international de recherche en lecture scolaire (PIRLS). TIMSS mesure les tendances en mathématiques et en science tous les quatre ans depuis 1995. PIRLS mesure les tendances en lecture et en écriture tous les cinq ans depuis 2001. Avec 50 à 70 pays qui participent à chaque cycle d'évaluation, les échelles de rendement scolaire et les références internationales de TIMSS et de PIRLS sont bien établies et utilisées par les pays du monde entier.

Six évaluations régionales mènent des évaluations de la lecture et des mathématiques, et constituent le potentiel d'application de la méthodologie :

- SACMEQ – Consortium d'Afrique australe et orientale pour le pilotage de la qualité de l'éducation ;
- PASEC – Programme d'analyse des systèmes éducatifs ;
- LLECE – Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation ;
- SEA-PLM – Métrique de l'apprentissage à l'école primaire en Asie du Sud-Est ;
- PILNA – Évaluation des compétences en lecture, écriture et calcul dans les îles du Pacifique.
- MILO : Suivi des impacts sur les résultats d'apprentissage

Figure 3. Illustration simplifiée de la méthodologie de raccordement par tests

Source : Institut de statistique de l'UNESCO (ISU)

Cependant, l'ISU et l'IEA mettront en œuvre la première application de la méthodologie dans la région de l'Amérique latine, à travers une collaboration avec LLECE. Comme les tableaux de correspondance permettent de projeter une note TIMSS ou PIRLS à toutes les notes envisageables de l'évaluation régionale, il sera possible de déterminer quelles sont les notes de l'évaluation régionale qui correspondent à chaque référence internationale de TIMSS ou PIRLS.

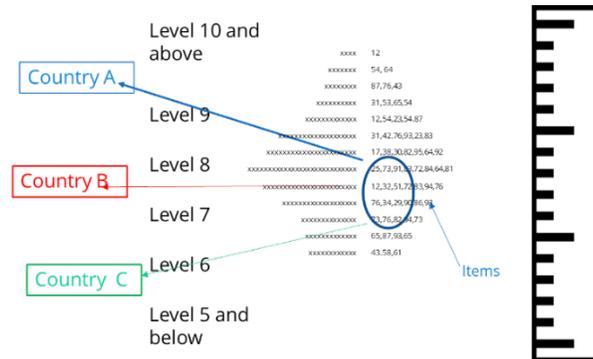
B. Réécalonnage fondé sur la psychométrie par items communs

Quoi ? Le raccordement par items est une méthode statistique qui consiste à donner à différents élèves des items que les évaluations partagent et à les utiliser pour étalonner le niveau de difficulté des items qu'elles ne partagent pas. Il repose sur l'étalonnage psychométrique d'items communs pour réaliser l'alignement de normes et de niveaux de difficulté.

- ✓ Il implique l'utilisation d'items communs de différents programmes d'évaluation.
- ✓ Une version a été proposée par le Conseil australien pour la recherche en éducation (ACER) dans le cadre d'une proposition générale de la progression de l'apprentissage, mais les options ne sont pas épuisées.³

³ Notez que l'échelle de référence est définie à partir des items de différentes évaluations.

Figure 4. Illustration simplifiée de la méthodologie de raccordement par items



Source : Institut de statistique de l'UNESCO (ISU).

Comment ? La méthodologie consiste à recueillir des centaines de questions de tests provenant d'un large éventail de programmes d'évaluation et de les ordonner ensuite au moyen de méthodes statistiques et d'avis d'experts en vue de comparer les données des différents tests et d'établir une échelle unique pour chaque domaine. Les questions des tests sont analysées pour identifier le type de compétences requises pour répondre correctement à chaque question. Chaque niveau des échelles d'établissement des rapports décrit les compétences de chaque segment sur la base des données empiriques des tests. Les échelles d'établissement des rapports découlent de ces descriptions et sont fondées sur des preuves empiriques de compétence démontrées à des points similaires dans le développement de la lecture et des mathématiques, associées aux avis des experts. Tous les items, ou un sous-ensemble utile d'items, seront ensuite étalonnés au moyen d'un ensemble d'items communs à chaque « niveau d'année d'études ». Cela dit, il est important de souligner que la méthodologie ne mesure pas le rendement scolaire. Il s'agit plutôt d'une approche visant à établir les rapports sur les résultats d'apprentissage de manière constructive et cohérente.

Table 4. Bilan de l'état d'avancement et de l'exécution

	Étalonnage pédagogique	par tests	par items
Seuils minimaux de compétence			
<i>Accompli à ce jour</i>	<ul style="list-style-type: none"> • Accord sur la définition • Alignement des évaluations internationales et basées sur la population • Analyse des tâches • Compilation des items 		
<i>Étapes restantes</i>			
• Quoi ?	Affinement des DSC et analyse des tâches		
• Quand ?	Début juillet.		
Raccorder les méthodologies			
<i>Accompli à ce jour</i>	Mise au banc d'essais de la boîte à outils en cours d'élaboration	<ul style="list-style-type: none"> ✓ Pays identifiés ✓ Organisation régionale ✓ Instruments détaillés ✓ Mise au banc d'essais terminée dans les pays d'Amérique latine 	Compilation des items
<i>Étapes restantes</i>			
• À exécuter par	<ul style="list-style-type: none"> ✓ Second trimestre 2019 ✓ 2 pays (USAID) ✓ 2 pays (UIS) ✓ 1 évaluation régionale (ISU) 	<ul style="list-style-type: none"> ✓ Principal domaine des évaluations régionales et mondiales ✓ Tableau de correspondance 	Banque d'items et plateforme d'échange
• Quand ?	2019-2020	2019-2021	2019-

Source : Institut de statistique de l'UNESCO (ISU)