



Organisation  
des Nations Unies  
pour l'éducation,  
la science et la culture

A photograph of a classroom with several students sitting at desks, focused on their work. The students are wearing dark blazers over white collared shirts. The background is a warm, wood-paneled wall.

# La promesse de l'évaluation à grande échelle des apprentissages

Reconnaître les limites pour libérer les potentialités

---

## Secteur de l'éducation de l'UNESCO

L'éducation est la priorité première de l'UNESCO car c'est un droit humain fondamental et la base pour construire la paix et faire progresser le développement durable. L'UNESCO est l'institution des Nations Unies spécialisée pour l'éducation et son Secteur de l'éducation assure un rôle moteur aux niveaux mondial et régional dans ce domaine, renforce les systèmes nationaux d'éducation et répond aux défis mondiaux actuels par le biais de l'éducation, en mettant tout particulièrement l'accent sur l'égalité des genres et l'Afrique.



Organisation  
des Nations Unies  
pour l'éducation,  
la science et la culture

Secteur de  
l'éducation

## L'agenda mondial Éducation 2030

En tant qu'institution des Nations Unies spécialisée pour l'éducation, l'UNESCO est chargée de diriger et de coordonner l'agenda Éducation 2030, qui fait partie d'un mouvement mondial visant à éradiquer la pauvreté, d'ici à 2030, à travers 17 Objectifs de développement durable. Essentielle pour atteindre chacun de ces objectifs, l'éducation est au coeur de l'Objectif 4 qui vise à « **assurer l'accès de tous à une éducation de qualité, sur un pied d'égalité, et promouvoir les possibilités d'apprentissage tout au long de la vie** ». Le Cadre d'action Éducation 2030 définit des orientations pour la mise en œuvre de cet objectif et de ces engagements ambitieux.



---

Publié en 2019 par l'Organisation des Nations pour l'éducation, la science et la culture,  
7, place de Fontenoy, 75352 Paris 07 SP, France

© UNESCO 2019

ISBN 978-92-3-200187-0

Œuvre publiée en libre accès sous la licence Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). Les utilisateurs du contenu de la présente publication acceptent les termes d'utilisation de l'Archive ouverte de libre accès UNESCO (<http://www.unesco.org/open-access/terms-use-ccbysa-fr>).



La présente licence s'applique exclusivement aux contenus textes de la publication. L'utilisation de contenus n'étant pas clairement identifiés comme appartenant à l'UNESCO devra faire l'objet d'une demande préalable d'autorisation auprès de l'UNESCO ([publication.copyright@unesco.org](mailto:publication.copyright@unesco.org)).

Titre original : *The promise of large-scale learning assessments: Acknowledging limits to unlock opportunities*

Publié en 2019 par l'Organisation des Nations pour l'éducation, la science et la culture

Les désignations 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 des pays, territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

Les idées et les opinions exprimées dans cette publication sont celles des auteurs ; elles ne reflètent pas nécessairement les points de vue de l'UNESCO et n'engagent en aucune façon l'Organisation.

Crédit de couverture : Getty / monkeybusinessimages

Mise en page et impression dans les ateliers de l'UNESCO

Imprimé en France

**La promesse de**

**l'évaluation à grande échelle**

**des apprentissages**

Reconnaître les limites pour libérer les potentialités

# Avant-propos

Les « tests standardisés » sont généralement associés aux examens de fin d'année que les élèves doivent réussir pour passer au niveau supérieur du cursus scolaire. Aujourd'hui, ces tests sont de plus en plus utilisés pour mesurer la performance des systèmes éducatifs, comme en témoigne l'importance croissante des évaluations à grande échelle des apprentissages (en anglais : Large-scale learning assessments, LSLA). Mises en œuvre au niveau national ou transnational, ces initiatives ont appelé l'attention sur les questions relatives aux niveaux d'apprentissage, mais aussi sur les facteurs qui déterminent cet apprentissage, notamment la qualification des enseignants, la qualité de l'environnement scolaire, le soutien et les conseils des parents, ainsi qu'un climat socioaffectif sain à l'intérieur et en dehors des écoles.

Les LSLA n'ont jamais connu un tel succès. Elles sont désormais considérées comme essentielles pour améliorer la qualité de l'éducation et contribuent grandement au renforcement des politiques et stratégies visant à développer un apprentissage efficace et pertinent. De plus en plus utilisées pour garantir l'excellence en matière d'éducation, elles sont au centre des discussions sur la qualité et l'efficacité des systèmes éducatifs aux niveaux national, régional et mondial. Néanmoins, la recherche de l'*excellence* et de la *qualité* ne doit pas se faire au détriment de l'*équité*. Le Programme de développement durable à l'horizon 2030 définit sans ambiguïté l'équité et l'inclusion comme l'axe central des politiques de développement de l'éducation. Il s'agit de faire en sorte que tous les enfants, jeunes et adultes aient droit à un apprentissage utile et pertinent. Si les LSLA sont un outil essentiel et indispensable en vue de la réalisation de ces objectifs, elles peuvent toutefois, dans certaines circonstances, compromettre involontairement les engagements pris au titre de l'agenda Éducation 2030 afin d'assurer à tous un apprentissage pertinent, équitable et de qualité.

Cette étude intitulée *La promesse de l'évaluation à grande échelle des apprentissages : reconnaître les limites pour libérer les potentialités* se penche sur les aspects plus problématiques de telles évaluations standardisées. S'appuyant sur la vaste expérience de l'UNESCO dans ce domaine – qu'il s'agisse de participer directement à la mise en œuvre des évaluations, de relayer les connaissances ou de tisser des réseaux – elle présente le point de vue critique de l'Organisation à l'égard de ces initiatives. Elle vise à rééquilibrer le débat sur les LSLA en examinant leurs effets bénéfiques tout en mettant en garde contre les risques et les écueils qui peuvent en résulter. Les analyses contenues dans la présente publication portent sur les LSLA conduites dans l'enseignement scolaire de type classique.

Il y a lieu d'espérer que les réflexions exposées dans cette étude seront utiles à divers groupes d'acteurs. En particulier, les gouvernements qui organisent des LSLA ou prévoient d'en lancer une, les organisations internationales et régionales qui facilitent par leur assistance technique la planification et la mise en œuvre de ces initiatives, les chercheurs qui en font une analyse

critique, les donateurs gouvernementaux qui les commandent de plus en plus souvent, et les entreprises privées qui pourraient envisager de se lancer dans ce marché de niche. Les analyses contenues dans cette publication sont destinées à mieux faire connaître et comprendre les LSLA en étudiant un certain nombre de questions clés. Pourquoi les LSLA se sont-elles imposées au cours des dix dernières années ? De quelles façons peuvent-elles aider à tenir les engagements pris, à l'horizon 2030, pour assurer à tous un apprentissage inclusif, équitable et de qualité ? Ou comment peuvent-elles au contraire compromettre ces engagements ? Justifient-elles le temps et les ressources qui leur sont consacrés ? Ciblent-elles réellement l'ensemble des enfants et contribuent-elles vraiment à promouvoir l'équité ? Mesurent-elles des connaissances et des savoir-faire pertinents, ou les compétences indispensables dans les sociétés modernes ?

La tendance actuelle à une participation croissante aux LSLA va certainement perdurer – ces évaluations devenant non seulement plus nombreuses mais aussi plus sophistiquées du fait de leur numérisation et de leur adaptabilité croissantes. Si cette perspective semble prometteuse pour l'avenir, nous devons néanmoins rester vigilants face aux risques qu'elle peut comporter. À quoi ressembleront les évaluations de demain et comment seront exploitées les données ainsi obtenues ? Comment éviter que les évaluations à grande échelle continuent, de manière involontaire, de faire obstacle aux efforts en faveur d'un apprentissage de qualité pour tous ou, pis encore, d'aggraver les inégalités sociales en matière d'apprentissage ? Pour la suite, nous devons nous assurer de comprendre pleinement les limites des LSLA – qu'il s'agisse de leur conception ou de leurs possibles usages – et en tirer les leçons pour l'élaboration des évaluations à venir.



**Stefania Giannini,**  
Sous-Directrice générale pour l'éducation, UNESCO

# Remerciements

La présente publication a été conçue par l'Équipe de recherche et de prospective en éducation, avec le concours du personnel des différents services du Secteur de l'éducation de l'UNESCO, y compris les bureaux régionaux et les instituts. Elle a été coordonnée par Maya Prince sous la direction de Sobhi Tawil.

L'UNESCO remercie particulièrement les experts qui ont contribué aux recherches et analyses : Anne-Berit Kavli (Norwegian Directorate of Education and Training) et Clara Fontdevila (Universitat Autònoma de Barcelona).

L'équipe souhaite également remercier Camilla Addey (Teachers College, Université Columbia) et Seamus Hegarty (Université de Warwick) pour la relecture du manuscrit.

Merci également à Zi Hu, Yutong Wang et Youngsun Cho, stagiaires à l'UNESCO, qui ont aidé à mener à bien les recherches préliminaires.

# Table des matières

Avant-propos .....	3
Remerciements .....	5
Liste des encadrés, figures et tableaux .....	7
Sigles et acronymes .....	8
Résumé exécutif .....	10
Introduction .....	14
<b>1. Le développement des évaluations .....</b>	<b>16</b>
Mise en contexte des évaluations à grande échelle des apprentissages .....	16
Un succès croissant .....	20
Quels sont les facteurs à l'origine de l'essor des évaluations à grande échelle des apprentissages ? .....	23
<b>2. L'évaluation de l'apprentissage dans le Programme 2030 .....</b>	<b>28</b>
Un mandat renouvelé en ce qui concerne les données relatives aux apprentissages .....	28
Promouvoir le nouvel agenda mondial de l'éducation .....	32
<b>3. L'apprentissage n'est pas seulement une affaire de mesure : les limites de la conception des évaluations à grande échelle des apprentissages .....</b>	<b>36</b>
Une conception limitée de l'éducation et de l'apprentissage .....	36
Une compréhension restreinte de la qualité de l'éducation .....	39
Une vision étrécie de ce qui présente de la valeur valeur dans les apprentissages .....	41
Limitation des catégories d'apprenants qui retiennent notre attention .....	49
<b>4. Au-delà des données considérées comme une fin en soi : les limites de l'utilisation des évaluations à grande échelle des apprentissages .....</b>	<b>53</b>
Les utilisations des évaluations à grande échelle des apprentissages : termes et concepts .....	53
Sous-exploitation des données d'évaluation .....	56
Surexploitation des données d'évaluation .....	59
Couplage des évaluations avec des cadres de réédition des comptes .....	65
Réflexions en guise de conclusion .....	70
Pour des évaluations qui tiennent leurs promesses .....	70
L'avenir des évaluations .....	71
Références .....	73
Annexe 1 : Évaluations transnationales à grande échelle des apprentissages .....	82
Annexe 2 : Évaluations nationales utilisant un cadre méthodologique commun .....	86

# Liste des encadrés, figures et tableaux

<b>Encadré 1</b>	Avantages attribués à la participation à des initiatives transnationales . . . .	24
<b>Encadré 2</b>	Initiatives régionales et mondiales récentes visant à promouvoir les évaluations à grande échelle des apprentissages . . . . .	33
<b>Encadré 3</b>	L'Alliance mondiale pour le suivi de l'apprentissage . . . . .	34
<b>Encadré 4</b>	Respect et coexistence – Test de l'étude internationale sur l'instruction civique et l'éducation à la citoyenneté 2009 . . . . .	45
<b>Encadré 5</b>	L'aménagement des tests – une arme à double tranchant ? . . . . .	51
<b>Encadré 6</b>	Protéger les élèves, les enseignants et les écoles des sanctions liées aux performances . . . . .	68
<b>Figure 1</b>	Les quatre piliers de l'éducation . . . . .	37
<b>Figure 2</b>	Pourcentage d'élèves selon le niveau de compétences atteint en langue et mathématiques – Début de scolarité . . . . .	64
<b>Tableau 1</b>	Récapitulatif des modalités d'évaluation . . . . .	17
<b>Tableau 2</b>	Importance nouvelle accordée à l'apprentissage – choix de publications . . . . .	26
<b>Tableau 3</b>	Cibles et indicateurs de l'ODD 4 relatifs à l'apprentissage . . . . .	29
<b>Tableau 4</b>	Données contextuelles collectées pendant les tests sur les acquis de l'apprentissage . . . . .	31
<b>Tableau 5</b>	Exemples d'initiatives politiques que les évaluations à grande échelle des apprentissages peuvent guider, par niveau d'intervention . . . . .	55



## Sigles et acronymes

<b>ACER</b> .....	Conseil australien de recherche pédagogique
<b>ADEA</b> .....	Association pour le développement de l'éducation en Afrique
<b>ANLAS</b> .....	<i>Analysis of National Learning Assessment Systems</i>
<b>ASER</b> .....	<i>Annual Status of Education Report</i>
<b>BIE</b> .....	Bureau international d'éducation de l'UNESCO
<b>CONFEMEN</b> .....	Conférence des ministres de l'Éducation des États et gouvernements de la Francophonie
<b>DFAT</b> .....	Ministère canadien des affaires étrangères et du commerce
<b>DFID</b> .....	Ministère britannique du développement international
<b>EGMA</b> .....	<i>Early Grade Mathematics Assessment</i> [Évaluation en mathématiques dans les petites classes]
<b>EGRA</b> .....	<i>Early Grade Reading Assessment</i> [Évaluation de la lecture dans les petites classes]
<b>EQAP</b> .....	<i>Educational Quality and Evaluation Programme</i>
<b>EDD</b> .....	Éducation en vue du développement durable
<b>ERCE</b> .....	<i>Estudio Regional Comparativo y Explicativo</i> [Étude régionale comparative et explicative]
<b>GAML</b> .....	Alliance mondiale pour le suivi de l'apprentissage
<b>GEM</b> .....	Rapport mondial de suivi sur l'éducation
<b>GPE</b> .....	Partenariat mondial pour l'éducation
<b>ICCS</b> .....	<i>International Civic and Citizenship Education Study</i>
<b>ICILS</b> .....	<i>International Computer and Information Literacy Study</i>
<b>IEA</b> .....	Association internationale pour l'évaluation du rendement scolaire
<b>IIEP</b> .....	Institut international de l'UNESCO pour la planification de l'éducation
<b>ISU</b> .....	Institut de statistique de l'UNESCO
<b>KIX</b> .....	<i>Knowledge and Innovation Exchange</i> [Partage de connaissances et d'innovations]
<b>LAMP</b> .....	Programme d'évaluation et de suivi de l'alphabetisation
<b>LaNA</b> .....	<i>Literacy and Numeracy Assessment</i>
<b>LLECE</b> .....	<i>Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación</i> [Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation]
<b>NALA</b> .....	<i>Network for African Learning Assessment</i>
<b>NAEP</b> .....	<i>National Assessment of Educational Progress</i>
<b>NEQMAP</b> .....	Réseau de suivi de la qualité de l'éducation dans la région Asie-Pacifique
<b>OAA</b> .....	<i>Optimizing Assessment for All</i>
<b>OCDE</b> .....	Organisation de coopération et de développement économiques
<b>ODD</b> .....	Objectif(s) de développement durable
<b>ONG</b> .....	Organisation non gouvernementale
<b>ONU</b> .....	Organisations des Nations Unies
<b>PAL</b> .....	<i>People's Action for Learning</i>
<b>PASEC</b> .....	Programme d'analyses des systèmes éducatifs de la CONFEMEN

<b>PERCE</b> .....	<i>Primer Estudio Regional Comparativo y Explicativo</i> [Première Étude régionale comparative et explicative]
<b>PIB</b> .....	Produit intérieur brut
<b>PILNA</b> .....	<i>Pacific Islands Literacy and Numeracy Assessment</i>
<b>PIRLS</b> .....	Programme international de recherche en lecture scolaire
<b>PISA</b> .....	Programme international pour le suivi des acquis des élèves
<b>RTI</b> .....	<i>Research Triangle Institute</i>
<b>SACMEQ</b> .....	Consortium de l'Afrique australe et orientale pour le pilotage de la qualité de l'éducation
<b>SEAMEO</b> .....	Organisation des ministres de l'éducation des pays du Sud-Est asiatique
<b>SERCE</b> .....	<i>Segundo Estudio Regional Comparativo y Explicativo</i> [Deuxième Étude régionale comparative et explicative]
<b>SEA-PLM</b> .....	<i>Southeast Asia Primary Learning Metrics</i>
<b>SIMCE</b> .....	<i>Sistema de Medición de la Calidad de la Educación</i> [Système de mesure de la qualité de l'éducation]
<b>TALENT</b> .....	Réseau africain des éducateurs pour la transformation de l'enseignement-apprentissage
<b>TERCE</b> .....	<i>Tercer Estudio Regional Comparativo y Explicativo</i> [Troisième Étude régionale comparative et explicative]
<b>TIC</b> .....	Technologies de l'information et de la communication
<b>TIMSS</b> .....	Étude internationale sur les tendances de l'enseignement des sciences et des mathématiques
<b>UNESCO</b> .....	Organisation des Nations Unies pour l'éducation, la science et la culture
<b>UNICEF</b> .....	Fonds des Nations Unies pour l'enfance
<b>USAID</b> .....	Agence des États-Unis pour le développement international
<b>VIH/sida</b> .....	Virus de l'immunodéficience humaine/Syndrome d'immunodéficience acquise

## Résumé exécutif

Les évaluations à grande échelle des apprentissages (*Large-scale learning assessments*, LSLA) se sont largement imposées au cours des dernières décennies. Ces épreuves standardisées, organisées au niveau national ou transnational offrent un instantané des acquis d'un groupe d'apprenants lors d'une année donnée et pour un nombre limité de domaines d'apprentissage.

Elles connaissent un succès grandissant et sont aujourd'hui utilisées dans plus de la moitié des pays du globe. Leur champ s'est élargi. D'abord utilisées pour mesurer les compétences en lecture et en mathématiques, elles ciblent désormais un plus grand nombre de domaines, y compris les compétences numériques, les notions élémentaires en matière d'informatique et d'information, les aptitudes socioaffectives, ou la compréhension des questions et concepts relatifs au civisme et à la citoyenneté.

Les LSLA occupent aujourd'hui une place centrale dans les débats sur l'éducation, aussi bien à l'échelon local qu'au niveau mondial. L'intérêt qu'elles suscitent a été encore avivé par l'importance nouvelle que le Programme de développement à l'horizon 2030 accorde à la nécessité d'offrir à tous l'accès à un apprentissage équitable, efficace et pertinent, d'où un effort massif d'exploitation des données pour améliorer les politiques et les stratégies. De plus en plus, les LSLA sont incluses dans les programmes de travail des organisations internationales et régionales, et bénéficient de l'aide financière et de l'assistance technique des organismes donateurs.

Les bienfaits et avantages des LSLA ont été abondamment analysés dans la littérature spécialisée, et la capacité des données d'évaluation d'éclairer de diverses façons la formulation des politiques est de plus en plus reconnue. De ce fait, les LSLA ont créé au cours des dernières décennies des attentes de plus en plus fortes en tant que moteur de la réforme des politiques. On compte en effet les mettre au service de multiples fins – suivi, reddition des comptes, définition des priorités, analyses, etc. Le potentiel de ces outils va bien au-delà de l'établissement de rapports. Ils peuvent signaler les domaines dans lesquels des améliorations sont nécessaires et aider à identifier les interventions les plus appropriées, les plus prometteuses et les plus efficaces.

Toutefois, les évaluations de l'apprentissage tant nationales que transnationales suscitent aussi des préoccupations. Des données en nombre croissant appellent l'attention sur toutes sortes d'effets non voulus, voire négatifs. S'appuyant sur des sources diverses, notamment la littérature et l'expérience de tout un éventail d'organisations internationales, de partenaires de développement et de spécialistes de l'évaluation, la présente publication se penche sur les possibles conséquences involontaires des LSLA. Certaines préoccupations sont directement liées aux caractéristiques de leur conception même, tandis que d'autres ont trait à l'utilisation des données pour éclairer (ou influencer) l'établissement des priorités et la formulation des politiques.

## **L'apprentissage n'est pas seulement une affaire de mesure : les limites de la conception des évaluations à grande échelle des apprentissages**

Un premier ensemble de conséquences indésirables des LSLA est directement imputable à leur conception. Quatre grandes dimensions de notre philosophie peuvent être négativement affectées : notre manière de concevoir l'éducation et l'apprentissage, notre vision de ce qu'est une éducation de qualité, l'éventail des compétences, connaissances et attitudes auxquelles nous accordons de la valeur, et notre attachement à l'inclusion et au respect de la diversité.

### **Une conception limitée de l'éducation et de l'apprentissage**

L'éducation est censée servir de multiples fins en matière de développement personnel et social, pour les individus comme pour la collectivité. Cette conception des finalités multiples et intégrées de l'éducation a trouvé sa juste traduction dans les quatre piliers de l'éducation mis en avant par la Commission Delors en 1996. Il est permis d'affirmer qu'elle est plus pertinente encore aujourd'hui pour progresser vers un développement humain et social durable dans un monde en pleine mutation. Pourtant, les LSLA portent en général sur un ensemble limité de dimensions de l'apprentissage et ne retiennent qu'une part de ces finalités multiples de l'éducation. Cela tient en grande partie à leur conception, qui se concentre sur un petit nombre de savoir-faire et de compétences mesurables pour permettre les comparaisons entre systèmes. Même si cela n'est pas en soi problématique, cela risque au bout du compte de limiter notre compréhension de ce qui a de la valeur dans l'éducation – en particulier si les LSLA acquièrent une importance disproportionnée par rapport à d'autres formes d'évaluation.

### **Une interprétation plus étroite de la qualité de l'éducation**

Les acquis des élèves, tels que mesurés par les LSLA, constituent une des principales dimensions et un des principaux indicateurs de la qualité de l'éducation. On ne saurait toutefois les interpréter isolément ou les considérer comme le seul critère valide de la qualité de l'éducation. Il convient de les apprécier au regard d'autres dimensions pertinentes, notamment les retombées positives du point de vue de la cohésion sociale, les variables relatives au processus d'enseignement et d'apprentissage, et les caractéristiques de la salle de classe et de l'environnement scolaire. Or, les progrès accomplis dans ces domaines ne se mesurent pas nécessairement en termes d'acquis. De fait, il est difficile de dire si une telle évolution serait possible ou même souhaitable. Si nous voulons appréhender les différentes dimensions d'une éducation de qualité, il nous faut combiner les résultats des tests standardisés avec toute une série d'autres indicateurs qualitatifs et quantitatifs.

## Une vision étriquée de ce qui présente de la valeur

La limitation du nombre de domaines ciblés et mesurés apparaît comme une caractéristique intrinsèque des LSLA. D'où certaines inquiétudes quant à la capacité de ces enquêtes d'évaluer correctement le large éventail des compétences cultivées par l'éducation. Même dans les domaines pris en considération, les LSLA ne sont pas suffisamment attentives aux nombreuses connaissances, compétences, attitudes et valeurs que l'éducation est censée faire acquérir. Certaines organisations explorent de nouvelles stratégies pour mesurer les compétences non retenues, mais la viabilité et la désirabilité de ces mesures restent sujettes à caution. La meilleure façon de remédier à la non-prise en compte de certains domaines du savoir serait sans doute de combiner les LSLA avec d'autres formes d'évaluation de façon que l'attention ne soit pas exclusivement centrée sur les résultats de l'apprentissage.

## Une attention insuffisante aux apprenants les plus vulnérables

Les LSLA sont très prometteuses en tant que moyen de promouvoir l'équité. Toutefois, leur conception et leur gestion ne respectent pas toujours l'impératif d'inclusion. Les enfants, jeunes et adultes vulnérables et marginalisés peuvent être exclus au moment de la constitution des échantillons, mais aussi au niveau de la classe ou de l'établissement scolaire – du fait même de la conception ou de la formulation des items. Des efforts croissants sont faits pour prendre en compte la totalité des élèves et mesurer correctement leurs progrès, mais l'entreprise demeure difficile, car les aménagements spéciaux risquent de créer à la fois des incitations perverses et des problèmes de comparabilité.

## Au-delà des données considérées comme une fin en soi : les limites de l'utilisation des évaluations à grande échelle des apprentissages

Les limites inhérentes aux LSLA tiennent aussi à l'exploitation – ou à la non-exploitation – de leurs résultats. Les données produites par ces évaluations servent à des fins diverses, notamment le suivi de l'évolution des tendances, l'évaluation et l'orientation des politiques, la sensibilisation à certains enjeux pour influencer sur la définition des grandes priorités, et le contrôle des performances de divers acteurs.

Quelles que soient les bonnes intentions voulant utiliser les données pour améliorer le processus éducatif, trois grands sujets de préoccupation se font jour : la sous-exploitation des données, leur surexploitation, et l'association (ou la subordination) de ces deux tendances à des cadres redditionnels.

## Sous-exploitation des données d'évaluation

Les LSLA ne peuvent devenir un moteur du changement et du progrès que si les parties concernées s'appuient sur elles pour agir. Or, divers facteurs font obstacle à un usage efficace

des données d'évaluation pour éclairer les débats sur les grandes orientations et guider l'action politique, notamment l'institutionnalisation limitée des pratiques d'évaluation et l'appropriation insuffisante des financements et de la conception, de la gestion et de la diffusion des données. La mise en œuvre des évaluations nécessitant des ressources considérables, il faut impérativement que les données obtenues soient utilisées, faute de quoi l'exercice est futile ou perçu comme un gaspillage des deniers publics.

## **Surexploitation des données d'évaluation**

Accorder une attention ou un crédit excessif aux résultats des évaluations peut détourner des priorités et fausser le comportement de toute une série d'acteurs. Cela se produit dans certaines circonstances et à la faveur d'un certain nombre de processus. Premièrement, l'impact des données d'évaluation sur la formulation des politiques est perdu si les mesures politiques visent uniquement ou principalement à améliorer les chiffres et non à engager des réformes de fond. Deuxièmement, les classements au sein d'un pays et entre pays peuvent créer une dynamique de compétition qui va inciter à adopter erronément des pratiques visant à copier tel ou tel pays plutôt qu'à tirer les leçons de son expérience. Troisièmement, les données d'évaluation peuvent pousser à des analyses de causalité qui risquent d'aboutir à des conclusions mal fondées sur les effets de l'apprentissage. Enfin, l'utilisation des données d'évaluation dans les systèmes de financement axés sur les résultats qui ont de plus en plus la faveur des donateurs présente un risque important – celui d'inspirer des stratégies ayant pour objet exclusif de modifier les résultats des évaluations.

## **Combiner évaluation et cadres de réédition des comptes**

Bon nombre des risques liés aux LSLA résultent en fait de ce que ces évaluations sont utilisées de pair avec des cadres de réédition des comptes très exigeants. La littérature spécialisée a abondamment analysé les effets non voulus de tels dispositifs. Parmi les effets négatifs les plus fréquemment examinés figurent la tendance à enseigner en fonction des tests et l'appauvrissement des programmes d'enseignement. Mais certes, les conséquences des évaluations dépourvues d'enjeux trop élevés sont encore mal comprises et reconnues et demandent à être étudiées plus avant.

Globalement, les LSLA sont riches de possibilités, mais comportent certains risques. Si leurs limites ne les rendent pas inutiles ou intrinsèquement problématiques, la vigilance est néanmoins de rigueur lors de la conception des outils et de l'exploitation des données obtenues. Puisque les LSLA deviennent un exercice obligé dans la plupart des systèmes éducatifs, il importe de déterminer les domaines dans lesquels des enquêtes, réflexions et recherches plus poussées seraient nécessaires. Par exemple, les questions relatives à la transparence et à l'appropriation des données, à l'utilité des autres modalités d'évaluation à des fins de diagnostic, et au potentiel des nouvelles formes de LSLA.

# Introduction

En matière d'éducation, le Programme de développement à l'horizon 2030 attache une importance nouvelle à la nécessité d'offrir à tous un apprentissage équitable, efficace et pertinent. Cet objectif politique central suppose que des efforts accrus soient faits pour surveiller l'évolution des résultats des processus éducatifs, comme incite également à le faire la reconnaissance croissante du potentiel des données d'évaluation pour diverses autres fins en rapport avec les politiques.

La littérature spécialisée confirme dans des travaux toujours plus nombreux les avantages offerts par les évaluations à grande échelle des apprentissages (LSLA). Ces systèmes d'évaluation sont un instrument de suivi essentiel aux niveaux mondial, national ou même local. Leur diffusion dans le monde entier au cours des deux dernières décennies témoigne de ce rôle crucial. De fait, ils ont contribué à l'amélioration des politiques et stratégies nationales visant à assurer à tous un apprentissage efficace et pertinent. S'ils continuent de remplir des fonctions toujours plus variées, leurs possibles effets imprévus et le risque que les données d'évaluation soient mal utilisées pour influencer sur les politiques et les pratiques en matière d'éducation suscitent toutefois des préoccupations. La présente publication apporte sa pierre au débat mondial en examinant ce que l'on sait des avantages, des limites et des éventuels effets pervers des LSLA. Elle appelle à plus de prudence dans la conception et la mise en œuvre des LSLA, ainsi que dans l'utilisation des données obtenues. Ces précautions sont nécessaires si nous voulons tenir notre engagement collectif de garantir à tous un apprentissage efficace et pertinent.

## Les questions essentielles

La publication s'efforce de répondre aux questions suivantes : En quoi consistent les LSLA ? Quels sont les facteurs à l'origine du succès croissant de ces initiatives ? Quelles sont les relations entre les LSLA et les engagements mondiaux à l'horizon 2030 ? Et, par dessus tout, quelles peuvent être les conséquences indésirables pour les politiques et les pratiques de la conception des LSLA et de l'utilisation des données ainsi obtenues ? L'objectif global de cette étude est d'aider à comprendre comment mieux concevoir et utiliser ces évaluations à grande échelle pour faire en sorte que l'éducation soit équitable et de qualité. Bref, de comprendre comment libérer les potentialités des LSLA.

## Structure de l'étude

### Chapitre 1

Le premier chapitre s'attache à clarifier le concept de l'évaluation à grande échelle des apprentissages (LSLA) en vue d'en préciser la place dans le champ plus général des instruments de suivi de l'éducation. Il retrace brièvement l'évolution des LSLA au cours des dernières décennies et décrit comment leur champ s'est élargi, et comment elles ont crû en nombre, en importance et en visibilité. Il examine également les raisons possibles de cette expansion, ainsi que l'évolution de la culture de l'évaluation et le changement d'optique, de l'accès à l'apprentissage, dans la théorie mondiale des politiques de l'éducation.

### Chapitre 2

Ce chapitre situe les LSLA par rapport aux principes et engagements qui sous-tendent le Cadre d'action Éducation 2030, et par rapport au discours relatif à l'apprentissage des acteurs mondiaux des politiques de l'éducation. Il examine comment l'adoption de l'objectif de développement durable concernant l'éducation (ODD 4) a ravivé l'attention portée aux acquis scolaires et comment les LSLA s'articulent avec ce nouveau mandat et le soutiennent. Il analyse enfin la manière dont les partenaires de développement ont contribué à faire de l'évaluation un moyen privilégié d'accélérer les progrès vers la réalisation de l'ODD 4.

### Chapitre 3

Ce chapitre présente une première série de conséquences involontaires liées à la conception et au développement des LSLA. Il examine quatre grands impacts négatifs pouvant résulter de la conception même des LSLA : la manière dont nous concevons l'éducation et l'apprentissage, notre compréhension de ce qu'est une éducation de qualité, l'éventail des compétences, connaissances et attitudes auxquelles nous attachons de la valeur, et notre engagement en faveur de l'inclusion et du respect de la diversité.

### Chapitre 4

Ce chapitre examine une deuxième série de conséquences non voulues de l'exploitation des données issues des LSLA par toutes sortes de parties prenantes. Trois grands sujets de préoccupation sont passés en revue. Premièrement, la sous-exploitation des données d'évaluation, et en particulier les limites qui résultent d'une appropriation insuffisante des données. Deuxièmement, la surexploitation de ces données, c'est-à-dire l'attention ou le crédit excessifs accordés aux résultats des évaluations, avec des effets de distorsion. Enfin, l'association (ou la subordination) de ces deux tendances à des cadres de réédition des comptes présentant des enjeux élevés.

De manière générale, cette publication s'emploie à mettre en lumière les tensions entre les possibilités offertes par les LSLA en tant qu'outils au service d'une éducation de qualité, équitable et responsable, et les conséquences négatives qu'elles peuvent avoir dans la pratique du fait de leur conception ou de leur utilisation à des fins impropres. En conclusion, ces tensions appellent à utiliser les LSLA de manière prudente et réfléchie afin de parer aux possibles effets néfastes.



# 1. Le développement des évaluations à grande échelle des apprentissages

Ce chapitre décrit à grands traits le développement progressif des évaluations à grande échelle des apprentissages (LSLA) au cours des dernières décennies, en ce qui concerne tout particulièrement leur portée, leur nombre, leur influence et leur visibilité. Après quelques éclaircissements d'ordre conceptuel, visant à replacer les LSLA dans leur contexte, il examine la diffusion et l'évolution des LSLA nationales et transnationales. Enfin, il se penche sur les facteurs qui stimulent la demande et l'intérêt pour ces outils, au niveau mondial comme à l'échelon national.

## Mise en contexte des évaluations à grande échelle des apprentissages

Les LSLA n'occupent une place importante dans les débats mondiaux, nationaux, voire locaux, sur l'éducation que depuis quelques décennies. Il convient toutefois de les concevoir comme une variante bien particulière faisant partie intégrante de systèmes d'évaluation plus vastes et d'une grande complexité. Cette première section clarifie certains concepts en vue de définir ce que sont les LSLA et de les situer dans leur contexte plus général.

Pour plus de simplicité et de clarté, cette étude est centrée sur les LSLA auxquelles sont soumis les écoliers – même si elles ne sont pas nécessairement mises en œuvre en milieu scolaire. Les questions qui y sont débattues concernent plus particulièrement l'enseignement primaire et secondaire. Il y a lieu toutefois de noter que les LSLA destinées à d'autres groupes d'âge se heurtent à un ensemble distinct de difficultés et de limites qui, quel que soit leur intérêt, n'entrent pas dans le champ de cette réflexion.<sup>1</sup>

### Évaluation de l'apprentissage ou évaluation *pour* l'apprentissage ?

L'évaluation est un élément essentiel de tout processus d'apprentissage. Les pédagogues distinguent fréquemment évaluation *pour* l'apprentissage et évaluation *de* l'apprentissage. Entrent dans la première catégorie les évaluations en salle de classe ou formatives sur lesquelles les enseignants se fondent pour adapter leurs méthodes pédagogiques ou pour noter les élèves et les classer au terme d'une certaine période d'instruction. Ces évaluations individuelles sont conçues pour mesurer et contrôler les connaissances, les compétences, les attitudes et les valeurs acquises dans tout un ensemble de matières. Elles comprennent des évaluations diagnostiques servant à repérer les élèves en difficulté et à adapter les activités d'enseignement et d'apprentissage en fonction de leurs besoins (Clarke, 2012). Les évaluations *de* l'apprentissage

---

1 C'est le cas par exemple des débats sur la manière de mesurer l'alphabétisation des adultes (Addey, 2014) ou sur les incidences en termes d'évaluation du concept d'apprentissage tout au long de la vie.

comprennent quant à elles les examens ou évaluations de nature sommative utilisés pour sélectionner les élèves d'une classe ou d'une tranche d'âge donnée et certifier qu'ils sont à même de poursuivre leur scolarité ou les orienter vers une formation ou un emploi. Ce sont des formes d'évaluation standardisées qui sont compatibles et en accord avec les programmes d'enseignement – mais généralement centrées sur les matières principales. **Le tableau 1** récapitule ces différentes modalités d'évaluation. Il est à noter que la distinction entre évaluation *pour* l'apprentissage et évaluation *de* l'apprentissage (et la distinction correspondante entre évaluation formative et évaluation sommative) est établie à des fins d'analyse et n'est pas strictement pertinente dans la pratique – un nombre croissant d'évaluations combinant l'une et l'autre fonctions (Clarke, 2012).

**Tableau 1. Récapitulatif des modalités d'évaluation**

<b>Évaluation interne ou en milieu scolaire</b>	Évaluation formative en salle de classe	
	Évaluation sommative en salle de classe	
<b>Évaluation externe ou standardisée</b>	Examens centralisés (appelés aussi examens publics ou examens de fin de cycle ou de fin d'études) – enjeu élevé pour les élèves	
	Évaluations à grande échelle ou à l'échelle du système éducatif (enjeu généralement faible pour les élèves)	Évaluations transnationales
		Évaluations nationales ou à l'échelon sous-national

Cases grisées : Évaluation pour l'apprentissage  
Cases blanches : Évaluation de l'apprentissage

Source : Basé sur Clarke, 2012 ; Fiske, 2000 ; OCDE, 2013.

## Les systèmes de mesure de l'apprentissage

Les évaluations *de* l'apprentissage et les évaluations *pour* l'apprentissage sont toutes deux des dimensions essentielles des systèmes de mesure des apprentissages. Ces systèmes sont constitués d'un ensemble de politiques, de structures, de pratiques et d'outils visant à générer des données sur les résultats de l'apprentissage (Clarke, 2012 ; UNESCO, 2017a). Ils regroupent différents types d'évaluation qui livrent des informations concernant les individus, les établissements ou le système éducatif qui sont utiles aux enseignants, aux élèves, aux planificateurs, aux responsables de l'élaboration des politiques et à d'autres parties prenantes. Ces systèmes peuvent aider les élèves à mieux apprendre, les enseignants à améliorer leurs cours, les planificateurs à décider de l'allocation des ressources et les responsables de l'élaboration des politiques et les gouvernements à corriger les déficiences du système éducatif ou à évaluer les programmes d'enseignement.

Appliquées à l'échelle d'un système éducatif, les LSLAs fournissent un instantané des acquis d'un groupe d'apprenants donné au cours d'une année donnée et dans un nombre limité de domaines. On distingue souvent les évaluations nationales et celles qui ont un caractère transnational.

## Qu'est-ce qu'une évaluation de l'apprentissage à grande échelle ?

Les LSLA sont une sous-catégorie particulière d'évaluation des résultats de l'apprentissage. Appliquées à l'échelle d'un système éducatif, elles fournissent un instantané des acquis d'un groupe d'apprenants donné (défini par la tranche d'âge ou le niveau d'études) au cours d'une année donnée et dans un nombre limité de domaines. On distingue souvent les évaluations nationales et celles qui ont un caractère transnational (régional ou international)<sup>2</sup>. Le contenu, les procédures de mise en œuvre, le calendrier et le système de notation des LSLA sont uniformisés et standardisés – de fait, on parle souvent de *tests standardisés* – en particulier dans la littérature et les pays anglo-saxons. Ces évaluations reposent en général sur des échantillons, bien que, ces dernières décennies, un nombre croissant de pays aient opté pour un système fondé sur un recensement (Verger *et al.*, 2018). Elles peuvent être centrées sur les écoles ou sur les ménages et être conçues ou non en fonction des programmes d'enseignement ; en général, les résultats ont des conséquences pour les enseignants et les écoles, mais n'en ont que très peu ou pas du tout pour les élèves qui passent les épreuves.

Les autorités responsables de l'éducation utilisent généralement les LSLA pour déterminer le niveau global des acquis des élèves. Ces évaluations aident les gouvernements à suivre l'évolution des résultats de l'apprentissage au fil du temps et à repérer les inégalités en matière d'acquis scolaires entre différents groupes au sein de la population. En mettant en évidence les corrélations entre tels ou tels facteurs et les résultats de l'apprentissage et en donnant une vision plus précise des interactions entre diverses variables, les données fournies par les LSLA permettent aussi de mieux comprendre la dynamique qui détermine la performance de systèmes éducatifs. De la sorte, elles éclairent la formulation des politiques et stratégies visant à améliorer les connaissances et les compétences des élèves, ainsi que l'équité en matière d'apprentissage.

## Rigueur dans la conception et la mise en œuvre

Une caractéristique distinctive (mais non exclusive) des LSLA est le fait qu'elles sont conçues et mises en œuvre avec le souci de répondre à des normes d'un certain niveau<sup>3</sup>. Si aucune norme internationale n'a été adoptée qui définirait les critères auxquels devrait satisfaire une LSLA pour être considérée comme solide (c'est-à-dire livrant des données fiables), les concepteurs de tests, les statisticiens et les psychométriciens s'accordent pour l'essentiel sur les impératifs techniques. L'élaboration et la conduite des LSLA doivent respecter au moins trois principes.

Premièrement, les LSLA doivent être techniquement solides. En d'autres termes, la méthodologie, l'analyse et l'interprétation des données issues de l'évaluation doivent reposer sur des principes scientifiques. L'évaluation nécessite par exemple des instruments contextuels bien construits, des outils d'évaluation valides et fiables, un contrôle de la qualité linguistique des

---

<sup>2</sup> On trouvera à l'annexe 1 une liste des évaluations transnationales et des informations plus détaillées à leur sujet.

<sup>3</sup> Si leur conception et leur gestion normalisées sont l'une des caractéristiques les plus distinctives des LSLA, il y a lieu de noter qu'elles ne sont pas une particularité *exclusive* de cette modalité d'évaluation. Ainsi, la normalisation est également un trait essentiel de la plupart des examens nationaux.

tests qui sont traduits, des procédures d'échantillonnage rigoureuses et le choix de méthodes d'analyse appropriées (ACER, 2017)<sup>4</sup>.

Deuxièmement, les évaluations doivent être conduites sur le terrain selon des modalités normalisées. Les données doivent par exemple être collectées dans des conditions techniques similaires et de manière efficace et sûre – indépendamment du contexte. Tous les agents participant aux opérations doivent avoir été convenablement formés et les tests être organisés conformément aux procédures établies.

Troisièmement, les LSLA doivent être conçues de manière à garantir un traitement éthique, impartial et inclusif de la population cible. Les principes d'équité et d'éthique doivent être respectés à chacune des étapes de l'évaluation, notamment :

- l'élaboration – s'efforcer par exemple de réduire autant que possible les biais de mesure<sup>5</sup> ;
- la mise en œuvre – il convient par exemple d'avoir à l'esprit le bien-être des participants lorsque l'on décide de la durée des épreuves ;
- l'analyse et la diffusion – il importe par exemple de toujours respecter le droit à la vie privée des participants en rendant les données anonymes préalablement à leur analyse et à leur publication.

Les évaluations doivent en outre être inclusives. Elles doivent être pertinentes pour un nombre aussi grand que possible de membres de la population cible et conçues pour s'adapter efficacement à tous les apprenants (UNESCO, 2017b). L'échantillon doit donc être constitué d'un ensemble représentatif d'apprenants, dans lequel il n'est pas tenu compte des attributs individuels, notamment le milieu d'origine (socioéconomique ou autre), la situation géographique ou les performances scolaires. De même, ces principes d'équité et d'inclusion doivent s'appliquer, par exemple, à la traduction et à l'adaptation des matériels de test en fonction de différents groupes de population ou à l'intention d'enfants ayant des besoins spéciaux.

4 À l'heure actuelle, il n'existe pas de norme internationale énonçant les caractéristiques d'une évaluation de l'apprentissage « techniquement solide ». Toutefois, l'étude intitulée *The Principles of Good Practice in Learning Assessment* réalisée par l'ACER en collaboration avec l'Institut de statistique de l'UNESCO et l'Alliance mondiale pour le suivi de l'apprentissage (ACER-ISU, 2017) propose une analyse indépendante des bonnes pratiques dans laquelle trouvent leur place les activités en matière de LSLA menées partout dans le monde. Pour plus d'informations, voir <http://ISU.unesco.org/sites/default/files/documents/principles-good-practice-learning-assessments-2017-en.pdf>.

5 Il se produit un « biais de mesure » lorsqu'un test ou un questionnaire de contexte traite systématiquement un groupe particulier de manière discriminatoire pour des raisons sans rapport avec le domaine d'apprentissage évalué ou les données relatives au profil des participants.

## Les promesses des LSLA

Les multiples avantages des LSLA ont été abondamment analysés dans la littérature spécialisée. Le potentiel des données d'évaluation s'agissant d'appuyer diverses interventions des pouvoirs publics est de fait de mieux en mieux reconnu. On a ainsi noté que, même si les évaluations s'inscrivent dans la phase de suivi et d'évaluation du cycle des politiques, leur utilité va bien au-delà. Comme le font valoir Best *et al.* (2013), les LSLA fournissent aussi des éléments essentiels qui éclairent l'établissement des grands objectifs (c'est-à-dire l'identification des questions prioritaires), orientent la formulation des politiques et guident leur application. De même, Tobin *et al.* (2015) décrivent l'impact des données issues des LSLA sur toute une série de politiques relatives à l'éducation – tout particulièrement les politiques appliquées à l'échelle du système, comme les réformes des programmes et l'établissement de normes de performance, les décisions relatives à l'allocation des ressources, et les stratégies en matière d'enseignement et d'apprentissage qui influent sur les pratiques dans les écoles et les salles de classe.<sup>6</sup> Comme le notent ces auteurs, les LSLA peuvent servir toutes sortes d'actions, y compris promouvoir la qualité, l'équité et la responsabilité. Les possibilités offertes par les LSLA sont également de plus en plus mises en avant par les organismes spécialisés dans l'éducation, qui ont plaidé pour un investissement soutenu dans ce type de mécanismes (voir, par exemple, ISU, 2018a ou Banque mondiale, 2018).

Le potentiel des données d'évaluation s'agissant d'appuyer diverses interventions des pouvoirs publics est de fait de mieux en mieux reconnu. Les LSLAs fournissent aussi des éléments essentiels qui éclairent l'établissement des grands objectifs, orientent la formulation des politiques et guident leurs applications.

Dans l'ensemble, ces travaux montrent clairement que les LSLA ont de multiples effets bénéfiques et devraient être un élément essentiel de tout système éducatif. L'absence de données sur l'apprentissage crée des risques variés pouvant compromettre les progrès vers plus d'équité et de qualité. Néanmoins, les efforts visant à déterminer sous la forme d'une synthèse les limites, les problèmes et les risques liés aux LSLA sont plus rares et restent embryonnaires. La présente publication a précisément pour objet d'étudier ces limites de manière plus systématique et de déterminer quelles sont les conditions optimales pour que les LSLA tiennent leurs promesses et contribuent à améliorer la qualité et l'équité dans l'éducation.

## Un succès croissant

Au cours des deux dernières décennies, les LSLA, nationales ou transnationales, ont connu un succès croissant dans le monde entier, et non plus seulement dans les pays riches. De fait, un nombre de plus en plus grand de pays à revenu intermédiaire – ou à faible revenu – investissent dans la mise en place de systèmes d'évaluation nationaux et participent à des évaluations transnationales (UNESCO, 2015b). Entre 2013 et 2017, des évaluations représentatives au niveau

---

<sup>6</sup> De fait, un certain nombre d'évaluations internationales à grande échelle livrent des analyses individuelles ou transnationales sur certains sujets. Voir à l'annexe 1 au présent rapport les liens vers des sites Web contenant des synthèses et analyses de recherches réalisées dans le cadre des évaluations transnationales mises en place.

national (transnationales ou nationales) ont été menées à bien dans au moins 177 pays pour ce qui concerne l'enseignement primaire et au moins 150 pays pour ce qui est de l'enseignement secondaire (ISU, 2019a). La présente section décrit succinctement la diffusion et l'évolution des deux catégories d'évaluation.

## Évaluations nationales

À la différence des examens qui sanctionnent les acquis des élèves et livrent accès au niveau d'études supérieur, les évaluations nationales ont pour principal objet de renseigner les responsables nationaux et sous-nationaux de l'élaboration des politiques sur les résultats d'apprentissage d'une population cible donnée (Greeney, 2008 ; Clarke, 2012). Les évaluations nationales portent généralement sur un petit nombre de matières (Greeney, 2008) – principalement la lecture, les mathématiques et, parfois, les sciences, les sciences sociales et les langues étrangères (Benavot et Köseleci, 2015)<sup>7</sup>. Le nombre de pays pratiquant de telles évaluations est passé de 70 en 1990-1999 à 135 pays en 2013 (Benavot et Köseleci, 2015), du fait principalement de leur diffusion dans les pays en développement. Ces chiffres comprennent les initiatives citoyennes et les EGRA (Évaluations de la lecture dans les petites classes – 2<sup>e</sup> et 3<sup>e</sup> années de scolarité)<sup>8</sup>, qui ont été organisées dans plus de 70 pays en développement et plus d'une centaine de langues. Dans une douzaine de pays, il s'agissait au départ d'enquêtes portant sur un échantillon restreint devenues des évaluations représentatives à l'échelon national (ISU, 2019a)<sup>9</sup>. Même s'il est impossible à l'heure où nous écrivons de chiffrer avec précision les évaluations nationales déjà en place, de vastes efforts sont en cours pour recenser les initiatives pays par pays – la base de données de l'ISU sur les évaluations de l'apprentissage<sup>10</sup> représentant la tentative la plus complète à ce jour.

## Évaluations transnationales

De même, le nombre de pays<sup>11</sup> participant à des évaluations transnationales a presque doublé depuis les premières éditions (1999-2000), et des estimations récentes font état de 137 pays engagés dans une ou plusieurs évaluations de ce type (ISU, 2018b)<sup>12</sup>. C'est ainsi que le nombre de pays prenant part à l'Étude internationale sur les tendances de l'enseignement des sciences et des mathématiques (TIMSS) est passé de 42 en 1995 à 64 en 2019 ; dans le cas du Programme international pour le suivi des acquis des élèves (PISA), la progression a été de 42 pays en 2000/2001 à 79 en 2018 – à quoi il faut ajouter sept pays supplémentaires engagés dans

7 Parmi les évaluations nationales des apprentissages les plus développées et les plus solidement établies, on peut citer le National Assessment of Educational Progress (NAEP) aux États-Unis et le SIMCE (*Sistema de Medición de la Calidad de la Educación* ou Système de mesure de la qualité de l'éducation) au Chili.

8 Même si les Évaluations des compétences en lecture dans les petites classes ont une portée nationale, ou même souvent sous-nationale (s'agissant d'instruments propres à un pays et non conçus à des fins de comparaisons entre pays), le fait qu'elles reposent sur un cadre méthodologique commun a conduit certains analystes à les classer dans la catégorie des évaluations internationales ou hybrides (voir Wagner, 2011 ; Raudonyte, 2019).

9 Pour plus d'informations sur les évaluations d'initiative citoyenne ou réalisées dans les petites classes, voir l'annexe 2 au présent rapport et le site Web du Global Reading Network : <https://www.globalreadingnetwork.net/>

10 <http://uis.unesco.org/fr/uis-learning-outcomes>.

11 Il y a lieu de noter que, si les évaluations transnationales comparent en général différents pays, les régions à l'intérieur d'un même pays peuvent constituer aussi des unités d'analyse différenciées, dans le cas en particulier de systèmes fédéraux, ou de pays dans lesquels les autorités locales jouissent d'un très grand degré d'autonomie administrative.

12 Voir à l'annexe 1 une liste des évaluations transnationales, assortie d'informations plus détaillées.

le programme PISA pour le développement (Raudonyte, 2019). Il convient aussi de noter que si la plus grosse part (60 %) des pays participants se situe encore en Europe et en Amérique du Nord (ISU, 2019a), la participation s'est étendue au fil des ans à l'ensemble des régions du monde. Parmi les pays du Sud, les prochaines éditions d'évaluations transnationales rallient de nouvelles vagues de participants et de nouvelles évaluations régionales ont vu le jour. Dans l'Afrique subsaharienne, par exemple, la République démocratique du Congo, le Gabon, la Guinée, Madagascar et le Mali rejoindront le Programme d'analyses des systèmes éducatifs (PASEC) de la CONFEMEN en 2019, tandis qu'en Amérique latine, la Bolivie, Cuba et El Salvador prendront part à la quatrième édition de l'Étude régionale comparative et explicative (ERCE)<sup>13</sup>. Dans la région Asie et Pacifique, l'évaluation de mesure des apprentissages primaires en Asie du Sud-Est (SEA-PLM) sera réalisée pour la première fois en 2019 dans 11 pays de la région.

Conçues initialement pour mesurer l'acquisition des compétences fondamentales par les élèves de l'enseignement primaire et du premier cycle de l'enseignement secondaire, les évaluations transnationales de l'apprentissage ont été étendues aussi aux enfants en bas âge et aux adultes et s'intéressent à un large éventail de domaines d'apprentissage, notamment l'initiation à l'informatique et à l'information, la compréhension des questions et concepts relatifs à la citoyenneté, et les comportements et l'engagement civiques.

### **Des défections, mais de courte durée**

Bien que la participation aux évaluations transnationales des apprentissage progresse, il est intéressant de noter que certains pays y renoncent, pour des raisons d'ordre politique, financier ou technique (Addey et Sellar, 2019)<sup>14</sup>. El Salvador, par exemple, a cessé de participer à toutes les évaluations internationales au milieu des années 2000, y compris à l'Étude internationale sur les tendances de l'enseignement des sciences et des mathématiques (TIMSS), au programme du Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation (LLECE) et au Programme d'évaluation et de suivi de l'alphabétisation (LAMP), après que le gouvernement nouvellement élu ait mis fin à toutes les activités d'évaluation à l'échelle du système éducatif. L'Argentine s'est retirée du Programme international de recherche en lecture scolaire (PIRLS) après la première évaluation en 2001. Ces changements s'expliquent peut-être par le fait que certains pays sont plus intéressés par les comparaisons avec leurs homologues régionaux – comme l'ont montré Ferrer et Fiszbein (2015) dans le cas de pays latino-américains. On peut aussi supposer que le programme PIRLS n'apportait aucune valeur ajoutée du fait de l'existence d'une évaluation régionale (LLECE) mesurant elle aussi les compétences en lecture des élèves des petites classes de l'enseignement primaire<sup>15</sup>. Ces tendances donnent à penser que la

---

13 À la date de publication, il était encore trop tôt pour connaître les nouveaux pays participant au Programme international de recherche en lecture scolaire (PIRLS) 2021 et à l'enquête PISA 2021.

14 À une échelle différente, on a pu observer aussi des mouvements de retrait vis-à-vis d'initiatives communautaires et citoyennes, principalement aux États-Unis. Ces mouvements se caractérisent par des campagnes incitant à refuser de se soumettre à des tests standardisés. Ils ont été conduits généralement par des parents d'enfants d'âge scolaire (voir Mitra *et al.*, 2016, ou Pizmony-Levy et Saraisky, 2016). Cette question n'entre toutefois pas dans le cadre de la présente publication.

15 Il importe toutefois de noter que les pays d'Amérique latine sont intéressés par la comparaison du niveau de leurs élèves du secondaire à celui des écoliers de pays fortement industrialisés au moyen d'évaluations qui, comme le programme PISA, ciblent les enfants âgés de 15 ans (Ferrer et Fiszbein, 2015).

participation aux évaluations transnationales est motivée ou induite par des raisons à la fois instrumentales et symboliques.

Des pays ont abandonné aussi leurs évaluations nationales. Tel est le cas de l'Afrique du Sud, qui avait organisé une Évaluation annuelle des acquis de tous les élèves des niveaux 1 à 9. L'évaluation était liée au contrôle des enseignants et des établissements scolaires, de sorte que la décision d'y mettre fin a été controversée. L'Ouganda, en revanche, a renoncé à ses évaluations annuelles en raison de difficultés financières qui l'ont poussé à ne plus les organiser que tous les trois ans. Toutefois, ces cas demeurent exceptionnels. La tendance générale va dans le sens d'un accroissement du nombre de LSLA et d'un élargissement de leur champ.

De fait, il apparaît que des pays qui se sont retirés de certaines initiatives transnationales pourraient fort bien participer aux éditions à venir. L'Autriche, la Bosnie-Herzégovine et les Philippines prendront part à l'évaluation TIMSS 2019 après plusieurs années de retrait. En Amérique latine, la Bolivie, Cuba et El Salvador participeront à l'étude ERCE 2019, après en avoir boudé une ou plusieurs éditions. L'Algérie était restée à l'écart des évaluations transnationales depuis sa participation à l'étude TIMSS 2011, mais a pris part au programme PISA en 2015.

## Quels sont les facteurs à l'origine de l'essor des évaluations à grande échelle des apprentissages ?

Les raisons qui poussent à mettre en place une évaluation nationale des acquis ou à participer à des initiatives transnationales sont multiples et dépendent de toutes sortes de facteurs souvent liés entre eux. Le postulat selon lequel la participation à des LSLA serait simplement fonction de la richesse ou du niveau de développement, ou servirait une fin particulière dans la sphère des relations internationales, est donc à écarter. La présente section analyse quatre grands facteurs à l'origine de la multiplication et de la diffusion des LSLA : le fait qu'on leur attribue un plus grand nombre d'avantages, l'évolution de la culture de l'évaluation en général, une réorientation de la politique de la communauté internationale en matière d'éducation, et les priorités et demandes des bailleurs de fonds du développement.

### Quels sont les avantages perçus ?

Certaines LSLA sont commandées par les donateurs bilatéraux et multilatéraux (Benavot et Köseleci, 2015) et/ou répondent à un souci de résultat et de contrôle (Kamens et McNeely, 2010), mais la participation est motivée aussi par des considérations liées aux avantages que l'on attribue au processus d'évaluation (voir **Encadré 1**). Cela est particulièrement vrai des initiatives transnationales. Au nombre de ces avantages figurent le développement de relations internationales positives, l'amélioration des performances économiques et la collecte de données pouvant éclairer les programmes d'enseignement et les stratégies pédagogiques. D'autres avantages, moins expressément reconnus, ont un caractère plus

Les raisons qui poussent à mettre en place une évaluation nationale des acquis ou à participer à des initiatives transnationales sont multiples et dépendent de toutes sortes de facteurs souvent liés entre eux.



symbolique, comme la participation à un « rite de reconnaissance internationale » qui incite à prendre exemple sur des sociétés dites « de référence », ou l'avancement des objectifs politiques personnels ou de la carrière de responsables nationaux de l'élaboration des politiques (Addey, 2014; Addey *et al.*, 2017)<sup>16</sup>. Si les raisons varient, la tendance générale est semble-t-il à une diminution du nombre de pays qui croient pouvoir accéder au statut de « société optimale » sans avoir progressé sur le plan des niveaux d'éducation atteints, *ni* fait les efforts de suivi correspondants (Kamens et McNeely, 2010).

---

### Encadré 1. Avantages attribués à la participation à des initiatives transnationales

- Assurer l'accès à des réseaux régionaux et internationaux, pour les chercheurs comme pour les décideurs ;
- Susciter le dialogue entre différents groupes et faire avancer le débat relatif à l'éducation ;
- Produire des données relatives à l'évolution de l'éducation et favoriser le recours à des indicateurs de l'éducation techniques et complexes ;
- Inciter à réformer les dispositifs réglementaires et les comportements, notamment en matière d'enseignement, d'apprentissage et d'évaluation ;
- Attirer l'attention des médias et améliorer la transparence en ce qui concerne les performances du système éducatif et le développement du capital humain aux niveaux national et transnational ;
- Développer les capacités des professionnels qui participent aux évaluations, sur le plan technique et opérationnel.

Sources : Basé sur Addey, 2014 ; Addey *et al.*, 2017 ; Diaz et Flores, 2008 ; Fenwick et Edwards, 2014 ; Lockheed *et al.*, 2015 ; Wagner *et al.*, 2012 ; OCDE, 2016b.

---

## Une culture de l'évaluation qui évolue

Le développement des LSLA est également lié à une évolution de la culture de l'évaluation. Au fil du temps, les évaluations conduites par l'OCDE et l'Association internationale pour l'évaluation du rendement scolaire (IEA), qui étaient au départ des recherches visant à analyser en profondeur tel ou tel système éducatif, sont devenues des initiatives transnationales axées sur la gestion des politiques et mettant l'accent sur la transparence et la concurrence (Pizmony-Levy, 2013 ; Addey, 2018). Cette évolution s'observe aussi ailleurs. Le programme PASEC est un exemple de ce passage d'évaluations nationales à des études comparatives. Entre 1991 et 2012, le PASEC a mené à bien des évaluations individuelles dans 24 pays et territoires de l'Afrique francophone, du Moyen-Orient et de l'Asie de l'Est. Jusque-là, les données n'étaient pas comparables entre pays. Ce n'est qu'à partir de 2014 que le PASEC a changé de mode opératoire et réalisé sa première évaluation transnationale portant sur 10 pays. Cette évolution des approches s'explique

---

16 En ce sens, l'essor des LSLA ne peut être dissocié de changements plus vastes et, tout particulièrement, de l'émergence de nouveaux paradigmes politiques hégémoniques tels que la « nouvelle gestion publique » (Addey *et al.*, 2017). Les modèles politiques préconisés par cette conception particulière des services publics mettent en avant la fonction évaluative de l'État (dans le domaine de l'éducation ou d'autres domaines). Alors que les pays sont de plus en plus vivement pressés de signaler leur engagement en faveur de ces programmes, l'adhésion aux LSLA est perçue comme un moyen relativement peu coûteux et (politiquement) avantageux de le faire (Verger *et al.*, 2018).

par l'intensification de la mondialisation et témoigne de l'intérêt croissant porté aux actions internationales (Greeney, 2008) et de la place croissante des LSLA dans le discours mondial sur l'éducation.

## **De l'accès à l'apprentissage : évolution du discours mondial sur les politiques de l'éducation**

L'objet principal du discours mondial sur l'éducation – en particulier dans les années qui ont conduit à l'adoption du Programme 2030 – s'est déplacé, de l'*accès aux études scolaires* à l'*apprentissage*. Il s'agit de fait d'une réorientation fondamentale, la préoccupation n'étant plus les *intrants* de l'éducation, tels que le taux d'inscription, les équipements scolaires, les matériels utilisés dans les salles de classe et la formation des enseignants, mais les *produits* de l'éducation – c'est-à-dire les connaissances, les compétences, les attitudes et les valeurs dont l'acquisition par les élèves est un *résultat* de l'éducation. C'est ce qui ressort des récents documents de stratégie des grands partenaires de développement internationaux (voir **tableau 2**) qui, avec d'autres, plaident pour un changement de paradigme, l'accent n'étant plus mis sur un *accès* équitable, mais sur un *apprentissage pour tous pertinent et efficace* (UNESCO-BIE, 2013 ; Barrett, 2016). L'équité et la *qualité* sont désormais au centre des débats mondiaux sur l'éducation et les LSLA apparaissent comme des outils susceptibles d'aider les pays dans leurs efforts pour assurer la qualité et l'équité de l'éducation.<sup>17</sup>

L'équité et la qualité sont désormais au centre des débats mondiaux sur l'éducation et les LSLA apparaissent comme des outils susceptibles d'aider les pays dans leurs efforts pour assurer la qualité et l'équité de l'éducation.

Les négociations auxquelles a donné lieu l'agenda Éducation 2030 ont contribué de manière décisive à affirmer et consolider le lien entre éducation de qualité, résultats de l'apprentissage et LSLA, en particulier dans le cadre du choix des indicateurs relatifs aux différentes cibles. Certes, les divers rapports et déclarations qui ont conduit à l'adoption de ce nouveau programme d'action ont mis en avant des conceptions différentes de la qualité. Certains en ont retenu une définition étroite, en l'assimilant aux compétences en lecture, écriture et calcul, mais la conception qui s'est finalement imposée se distingue par sa portée très étendue. Néanmoins, la traduction de la liste finale des cibles en indicateurs a entraîné une relative restriction de l'ambition initiale en faisant de la mesure de la maîtrise de la lecture et mathématiques un critère voisin (King, 2017 ; Unterhalter, 2019).

17 Si les LSLA sont au centre des débats internationaux, il est à noter que ces outils retiennent l'attention et font l'objet d'un travail de conceptualisation depuis au moins les premiers jours de l'initiative Éducation pour tous. Le rapport publié en 2000 sous le titre *L'évaluation des acquis scolaires* dans la série *Situation et tendances* (Fiske, 2000) s'interrogeait déjà sur le rôle essentiel de l'évaluation des acquis des élèves comme moyen de promouvoir une éducation de qualité pour tous. En ce sens, il convient de relever que le présent rapport se nourrit de la plupart des questions autour desquelles tournent les débats d'aujourd'hui.

**Tableau 2. Importance nouvelle accordée à l'apprentissage – choix de publications**

Publication/Initiative	Organisation	Objectif
Apprentissage pour tous : investir dans l'acquisition de connaissances et de compétences pour promouvoir le développement. Banque mondiale 2020 Éducation Stratégie (2011)	Groupe de la Banque mondiale	Faire en sorte que tous les enfants et les jeunes aillent à l'école, mais aussi acquièrent les connaissances et les compétences qui leur sont nécessaires pour mener une vie saine et productive et obtenir un emploi satisfaisant.
APPRENDRE pour réaliser la promesse de l'éducation. Rapport sur le développement dans le monde 2018 (2018)	Groupe de la Banque mondiale	Résultats de l'apprentissage
Initiative mondiale pour l'éducation avant tout (2012-2016)	Initiative mondiale du Secrétaire général de l'ONU pour l'éducation	Scolarisation de tous les enfants ; qualité de l'apprentissage ; citoyenneté mondiale.
Comité de réflexion sur la métrique des apprentissages. Access Plus Learning (2012-2016)	Institut de statistique de l'UNESCO et le Centre pour l'éducation universelle à la Brookings Institution	Renforcer les systèmes d'évaluation et l'utilisation des données d'évaluation.
La génération d'apprenants. Investir dans l'éducation pour un monde en pleine évolution (2016)	Commission internationale pour le financement de possibilités d'éducation dans le monde)	Mobiliser des ressources et les déployer pour faire en sorte que tous les enfants et les jeunes aient accès à l'éducation et aux compétences qui leur seront nécessaires pour devenir des adultes productifs et pour réussir.

*Note :* Cette sélection n'est pas exhaustive. Elle ne vise qu'à donner des exemples de l'accent mis sur l'apprentissage par des initiatives, mécanismes de financement et stratégies organisationnelles variés.

## Importance accordée à la production et à l'utilisation des données

Aujourd'hui, les partenaires de développement et organismes donateurs multilatéraux subordonnent de plus en plus l'octroi de leurs prêts et aides à une forme quelconque d'évaluation de l'apprentissage (Kamens et McNeely, 2010). Les bailleurs de fonds s'appuient bien souvent sur les résultats des LSLA pour apprécier l'efficacité des politiques et pratiques en matière d'éducation. Les données d'évaluation peuvent aider à suivre l'évolution des déficits d'apprentissage au sein de différents groupes et entre eux, au regard de facteurs tels que le sexe, le niveau de revenu, le lieu de résidence et la langue parlée dans le foyer. Le plan stratégique pour 2020 du Partenariat mondial pour l'éducation (GPE) a par exemple adopté le principe de financements fondés sur les résultats pour aider les gouvernements à mettre en place de solides systèmes d'évaluation, et il les encourage à mener des évaluations de l'apprentissage de toutes formes pour vérifier les acquis des apprenants (GPE, 2016). Pour avoir accès aux aides du GPE, tous les candidats doivent faire la preuve qu'ils disposent de systèmes de gestion des données de bonne qualité ou ont adopté une stratégie pour s'en doter (GPE, 2018). De la même façon, la stratégie « L'apprentissage pour tous » de la Banque mondiale soutient les gouvernements dans leurs efforts pour accroître la quantité de données disponibles sur l'apprentissage et les

compétences, et pour instaurer une culture du suivi et de l'évaluation des résultats (Banque mondiale, 2011). Le Cadre d'apprentissage de l'OCDE pour 2030 (OCDE, 2018a) et la Stratégie de l'UNESCO pour l'éducation 2014-2021 attachent également une importance toute particulière à l'évaluation de l'apprentissage, considérée comme un moyen de vérifier les acquis fondamentaux et transmissibles, et à l'utilisation des données pour éclairer les décisions visant à améliorer l'enseignement et l'apprentissage.

Des priorités similaires s'observent dans les stratégies de développement de certains grands donateurs bilatéraux finançant l'éducation. C'est ainsi que, dans sa nouvelle Stratégie pour l'éducation, le Ministère britannique du développement international souligne sa volonté de coopérer avec les décideurs pour améliorer la qualité et l'utilisation des données relatives à l'éducation, en donnant la priorité à l'adoption de bons outils nationaux et transnationaux de mesure des acquis (DFID UK, 2018). L'amélioration des systèmes d'évaluation de l'apprentissage et l'utilisation des données à l'appui de politiques et de pratiques solides figurent également en bonne place parmi les objectifs de la stratégie du Ministère australien des affaires étrangères et du commerce pour 2015-2020 (DFAT, 2015). De même, les gouvernements du Canada, des États-Unis, de la Finlande, de la Norvège et de la Suède mentionnent expressément les évaluations des apprentissages au nombre des outils de suivi essentiels (NORAD, 2018 ; ASDI, 2018 ; DFAT Canada, 2014 ; USAID, 2011 ; Ministère de l'éducation, Finlande, 2006).

## 2. L'évaluation de l'apprentissage dans le Programme 2030

Si les évaluations à grande échelle des apprentissages (LSLA) et le suivi des acquis ont récemment retenu l'attention de la communauté des acteurs du développement, l'adoption de l'ODD 4 a conféré à ces aspects de l'éducation une importance sans précédent. Le présent chapitre examine comment les LSLA s'articulent avec l'agenda Éducation 2030, et comment divers partenaires de développement encouragent les acteurs internationaux à s'intéresser davantage aux données relatives aux apprentissages en tant que moyen d'accélérer les progrès vers la réalisation de l'ODD 4.

### Un mandat renouvelé en ce qui concerne les données relatives aux apprentissages

L'agenda Éducation 2030 a contribué à renforcer l'intérêt pour les évaluations et les données relatives aux acquis, et à présenter ainsi les LSLA comme indispensables au suivi des apprentissages et des progrès dans la réalisation des nouvelles cibles. La présente section analyse la place centrale des LSLA dans la promotion de cette vision nouvelle de l'éducation – comme des outils de suivi et d'établissement de rapports, mais aussi comme des moyens d'orienter et d'encourager les politiques et d'assurer à terme la réalisation de l'ODD 4.

#### La vision de l'agenda Éducation 2030

Quel est le rôle des LSLA dans l'agenda Éducation 2030 ?

Comme indiqué dans la Déclaration d'Incheon adoptée en 2015, cet agenda est inspiré par une vision humaniste de l'éducation et du développement fondée sur le respect de droits de l'homme et la dignité humaine, la justice sociale, l'inclusion, la protection, la diversité culturelle, linguistique et ethnique, ainsi que sur une responsabilité et une redevabilité partagées. La Déclaration réaffirme le principe selon lequel l'éducation est un bien public, un droit humain fondamental et un préalable à l'exercice d'autres droits. Elle décrit l'éducation comme une condition essentielle de l'épanouissement de chacun et du développement durable dans un contexte où la croissance économique est guidée par un souci fondamental de justice sociale et de gestion responsable de l'environnement. Les éléments du Programme de développement à l'horizon 2030 relatifs à l'éducation ont recentré l'attention sur l'équité et l'inclusion, ainsi que sur la qualité et la pertinence de l'éducation, dans la perspective de l'apprentissage tout au long de la vie. Telle est, en essence, la teneur de l'Objectif de développement durable 4 (ODD 4), qui appelle à « assurer une éducation inclusive et équitable de qualité et promouvoir des possibilités d'apprentissage tout au long de la vie pour tous » (UNESCO, 2016c, page 9).

L'agenda Éducation 2030 a contribué à renforcer l'intérêt pour les évaluations et les données relatives aux acquis, et à présenter ainsi les LSLA comme indispensables au suivi des apprentissages et des progrès dans la réalisation des nouvelles cibles.

## Les résultats et produits de l'éducation au cœur de l'ODD 4

L'ODD 4 met l'accent sur les résultats et produits du processus éducatif. Il privilégie l'acquisition des connaissances, aptitudes fondamentales et transférables et compétences pertinentes qui constituent le socle d'un apprentissage tout au long de la vie utile dans le monde du travail et dans la vie civique, sociale et culturelle (UNESCO, 2016a). C'est ce dont témoignent clairement les cibles et indicateurs mondiaux, axés sur les résultats, qui sont proposés pour suivre les progrès dans la réalisation de l'ODD 4 (voir **tableau 3**).

Compte tenu de la place centrale de l'apprentissage dans le nouvel agenda, les données s'y rapportant apparaissent de plus en plus comme un élément essentiel des mécanismes de suivi visant à mesurer la mise en œuvre de l'ODD 4 – aux niveaux mondial, régional et national.

**Tableau 3. Cibles et indicateurs de l'ODD 4 relatifs à l'apprentissage**

<b>Cible 4.1. D'ici à 2030, faire en sorte que toutes les filles et tous les garçons suivent, sur un pied d'égalité, un cycle complet d'enseignement primaire et secondaire gratuit et de qualité les dotant d'acquis véritablement utiles.</b>	
Indicateur 4.1.1	Pourcentage d'enfants et de jeunes (a) de 2e et de 3e années ; (b) au terme de l'enseignement primaire ; et (c) au terme du premier cycle du secondaire ayant atteint au moins un niveau minimum de compétence (i) en lecture et (ii) en mathématiques, par sexe.
<b>Cible 4.2. D'ici à 2030, faire en sorte que toutes les filles et tous les garçons aient accès à des services de développement et de prise en charge de la petite enfance et à une éducation préprimaire de qualité qui les préparent à suivre un enseignement primaire.</b>	
Indicateur 4.2.1	Proportion d'enfants de moins de 5 ans dont le développement est en bonne voie en matière de santé, d'apprentissage et de bien-être psychosocial, par sexe.
<b>Cible 4.4. D'ici à 2030, augmenter nettement le nombre de jeunes et d'adultes disposant des compétences, notamment techniques et professionnelles nécessaires à l'emploi, à l'obtention d'un travail décent et à l'entrepreneuriat.</b>	
Indicateurs 4.4.1	Proportion de jeunes et d'adultes ayant des compétences dans le domaine des technologies de l'information et des communications, par type de compétence.
4.4.2	Pourcentage de jeunes et d'adultes qui ont acquis au moins un niveau minimum de compétences en matière d'alphabétisation numérique.
<b>Cible 4.6. D'ici à 2030, faire en sorte que tous les jeunes et une proportion considérable d'adultes, hommes et femmes, sachent lire, écrire et compter.</b>	
Indicateur 4.6.1	Proportion de la population d'un groupe d'âge donné ayant les compétences voulues à au moins un niveau d'aptitude fixé (a) en alphabétisme et (b) en arithmétique fonctionnels, par sexe.
<b>Cible 4.7. D'ici à 2030, faire en sorte que tous les apprenants acquièrent les connaissances et les compétences nécessaires pour promouvoir le développement durable, y compris notamment par l'éducation en vue du développement durable et de modes de vie durables, les droits de l'homme, l'égalité des genres, la promotion d'une culture de la paix et de la non-violence, la citoyenneté mondiale et l'appréciation de la diversité culturelle et de la contribution des cultures au développement durable.</b>	
Indicateurs 4.7.4	Pourcentage d'élèves par tranche d'âge (ou d'un niveau d'études) montrant une compréhension adéquate des questions relatives à la citoyenneté mondiale et à la durabilité.
4.7.5	Pourcentage d'élèves de 15 ans montrant une maîtrise des connaissances de la science environnementale et de la géoscience.

Source : UNESCO 2016a.

Note : La liste complète des 10 cibles de l'ODD 4 et des 43 indicateurs correspondants est disponible à l'adresse suivante : Comprendre l'objectif de développement durable 4<sup>18</sup>. La base de données de l'ISU contient les données relatives aux 43 indicateurs thématiques : <http://data.ISU.unesco.org/>.

18 Voir Comprendre l'ODD 4 : [https://unesdoc.unesco.org/ark:/48223/pf0000246300\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000246300_fre).

## Un outil indispensable de suivi des indicateurs de l'ODD 4

Un certain nombre des cibles et indicateurs mondiaux de l'ODD 4 étant axés sur les résultats, les LSLA sont apparus comme un moyen de produire des données récentes et fiables permettant de vérifier si les engagements relatifs à l'apprentissage étaient tenus (ISU, 2018c). De fait, les différentes cibles liées à des résultats qui sont présentées dans le tableau 3 sont conçues pour faire l'objet d'un suivi à l'aide d'une forme quelconque d'évaluation des résultats de l'apprentissage au niveau des systèmes éducatifs. C'est pourquoi les LSLA sont considérées comme des outils indispensables – du moins jusqu'à ce que soient mises au point de nouvelles méthodologies intégrant d'autres sources de données (telles que données administratives, examens nationaux, recensements ou enquêtes auprès des ménages) et d'autres types d'estimations et/ou de projections.

## Au-delà des cibles de l'agenda mondial pour l'éducation

Les méthodes de mesure de l'apprentissage et les LSLA jouent un rôle central dans l'agenda Éducation 2030 car elle permettent de rendre compte de toute une série de cibles axées sur les résultats, mais elles offrent bien d'autres possibilités encore<sup>19</sup>. La métrique de l'apprentissage livre aussi des indications essentielles sur les motivations des élèves qui réussissent, ainsi que sur l'efficacité des politiques visant à soutenir et améliorer leurs progrès.

Les LSLA permettent de mesurer les progrès au fil du temps. Les tests sont accompagnés d'une série de questionnaires qui sont distribués aux élèves, aux enseignants, aux chefs d'établissement et/ou aux parents pour obtenir des informations concernant le contexte socioéconomique et démographique, ainsi que sur les pratiques ou l'expérience qui influent sur les résultats de l'apprentissage (voir **tableau 4**). De plus, certaines évaluations livrent d'importants renseignements sur le système éducatif, notamment des informations qui ne sont en général pas disponibles dans les bases de données internationales sur l'éducation, comme les modalités de l'éducation de la petite enfance, les politiques nationales en matière de scolarisation, de passage aux degrés supérieurs et d'examens, et le nombre d'heures de cours consacrées à telle ou telle matière.

Toutes ces informations permettent aux pays d'évaluer la bonne santé de leur système éducatif, de déterminer les facteurs corrélés avec les résultats (par exemple qualifications des enseignants, ressources pédagogiques, accès) et de comprendre comment la combinaison de diverses variables donne une image plus claire de la dynamique des systèmes éducatifs. Ensemble, les données apportent des indications sur les aspects en rapport avec la réussite de l'élève et les liens entre celle-ci et la santé, l'équité, la qualité des intrants scolaires, le degré de soutien parental au foyer, et la typologie et la situation géographique des écoles. En ce sens, les LSLA jouent un rôle essentiel dans la mesure de l'équité – l'un des principes essentiels qui guident l'agenda Éducation 2030.

---

19 Les activités des différentes organisations chargées de produire des données d'évaluation témoignent de ce potentiel. L'UNESCO et l'IEA ont ainsi réalisé une brochure sur les résultats de l'enquête PIRLS 2016 qui confirme comment les données fournies par les LSLA sur les résultats de l'apprentissage et autres aspects peuvent aider à suivre les progrès accomplis dans la réalisation de l'ODD 4.

Si les évaluations transnationales des apprentissages sont généralement assorties de divers questionnaires sur les facteurs contextuels concernant le foyer, la communauté, l'école et les élèves, cela est rarement le cas des évaluations nationales. La plupart d'entre elles s'accompagnent d'un questionnaire sur l'élève (8 sur 10), et la moitié d'un questionnaire sur l'école ou les enseignants, mais seulement 20 % de ces évaluations collectent des informations sur le milieu familial et les pratiques ou le soutien des parents – autant d'éléments déterminants du succès scolaire (ISU, 2019a). Il existe encore d'autres facteurs auxquels aucun des deux types d'évaluation ne s'intéresse. La collecte de données sur la santé et les handicaps, l'origine ethnique ou le statut de minorité est souvent strictement réglementée et n'est pas facile à réaliser dans le cadre des évaluations nationales ou transnationales de l'apprentissage.

**Tableau 4. Données contextuelles collectées pendant les tests sur les acquis de l'apprentissage**

Questionnaire pour l'élève	Questionnaire pour l'enseignant	Questionnaire pour les parents	Questionnaire pour l'école (rempli par le chef d'établissement)
<ul style="list-style-type: none"> <li>• Sexe, âge, langue parlée au foyer</li> <li>• Equipement du foyer [par exemple bureau de travail, Internet, chambre à coucher, nombre et types de livres]</li> <li>• Pratiques et soutien parentaux [par exemple utilisation d'un ordinateur, lecture récréative, activités avec les parents]</li> <li>• Santé et bien-être [absences de l'élève, sentiments à l'égard de l'école, nutrition correcte au foyer]</li> <li>• L'établissement [harcèlement, plaisir pris aux différentes leçons, impressions concernant les enseignants, utilisation de la bibliothèque]</li> <li>• Renseignements donnés par l'élève lui-même concernant son niveau en lecture et en mathématiques.</li> </ul>	<ul style="list-style-type: none"> <li>• Sexe, âge, niveau d'instruction, principaux domaines d'études et de spécialisation</li> <li>• Expérience de l'enseignement et formation professionnelle</li> <li>• Attitudes et impressions concernant l'école, ses ressources, son environnement et les facteurs de réussite scolaire</li> <li>• Observations concernant les obstacles (comportements, pratique, ressources) à l'apprentissage dans les classes</li> <li>• Pratiques pédagogiques et évaluation de l'enseignement</li> <li>• Participation des familles aux activités de l'école.</li> </ul>	<ul style="list-style-type: none"> <li>• Type et fréquence des activités menées aux côtés des enfants [lecture, chant, calcul, jeux, etc.]</li> <li>• Situation en matière de migration et langue(s) parlé(es) au foyer</li> <li>• Inscription des enfants à des programmes d'éducation de la petite enfance et durée de cette éducation</li> <li>• Âge d'entrée des enfants à l'école primaire</li> <li>• Impressions concernant le niveau des enfants en lecture, mathématiques ou science</li> <li>• Activités d'apprentissage extrascolaires</li> <li>• Impressions sur l'école fréquentée par l'enfant</li> <li>• Pratique de la lecture, plus haut niveau d'études et type d'emploi des parents.</li> </ul>	<ul style="list-style-type: none"> <li>• Expérience et niveau d'études du chef d'établissement</li> <li>• Nombre d'élèves inscrits et distribution par milieu socio-économique</li> <li>• Situation géographique de l'école</li> <li>• État de l'infrastructure scolaire</li> <li>• Services de restauration</li> <li>• Nombre d'heures de cours</li> <li>• Ressources et moyens techniques [par exemple bibliothèque, salle d'informatique, fournitures, équipements destinés aux élèves handicapés]</li> <li>• Discipline et sécurité dans l'établissement</li> <li>• Contribution des parents et de la communauté à la vie de l'établissement.</li> </ul>

Source : Les auteurs. Pour plus d'exemples de questionnaires de contexte, voir Cresswell et al., 2015 ; OCDE, 2017 ; ISU, 2016b ; IEA, 2015 ; IEA, 2016 ; ISU, 2016b.

Note : Aucun questionnaire ne contient l'ensemble des items présentés dans ce tableau, qui ne vise qu'à illustrer la profusion d'informations qu'il est possible de collecter et d'analyser, outre les données relatives aux acquis, au moyen de tels questionnaires.



Des questionnaires et autres relevés de données connexes sur le système éducatif aident également à recueillir des informations utiles pour vérifier la concrétisation des engagements politiques au fil du temps. Il est notamment possible de collecter des données détaillées sur l'impact de certaines mesures concernant la langue d'instruction, ou de connaître l'évolution des politiques touchant les enseignants et le poids relatif de telle ou telle matière dans les programmes d'enseignement. Bref, ces questionnaires de contexte peuvent fournir de précieuses informations non seulement sur les résultats d'apprentissage de groupes particuliers, mais aussi sur des facteurs contribuant à cet apprentissage, tels que les enseignants, les salles de classe, le soutien parental et les ressources scolaires. Cette profusion de données joue ensuite un rôle essentiel en guidant et orientant les politiques à de multiples niveaux (mondial, national, local), favorisant ainsi à terme la réalisation de l'ODD 4.

## Promouvoir le nouvel agenda mondial de l'éducation

De nombreuses initiatives conçues pour faciliter la mesure des résultats de l'apprentissage ont vu le jour au cours des dernières décennies. La présente section passe en revue l'éventail des projets et programmes récemment mis en œuvre à différents niveaux pour appuyer le développement et le renforcement des LSLA.

### Soutien aux LSLA : les initiatives mondiales et régionales

Nous l'avons vu, les LSLA deviennent rapidement des outils indispensables pour suivre les résultats obtenus en matière d'apprentissage dans le cadre du Programme 2030. Leur utilisation a été soutenue par des donateurs gouvernementaux, des organisations internationales et régionales, des partenaires de développement et d'autres acteurs privés qui reconnaissent l'intérêt considérable de ces instruments pour l'amélioration des systèmes éducatifs.

Les organisations internationales et régionales plaident en faveur de l'utilisation des LSLA pour faciliter l'amélioration du système éducatif. Des efforts considérables sont faits pour aider les pays à produire des données et à les exploiter dans des domaines essentiels pour l'amélioration des processus et des résultats de l'apprentissage.

Ces organisations reconnaissent aussi l'importance des données d'évaluation pour l'élaboration de stratégies nationales de développement de l'éducation, et la nécessité d'intensifier les efforts nationaux, régionaux et mondiaux visant à remédier aux vastes lacunes dans les données relatives aux résultats de l'apprentissage observées dans le monde (ISU, 2018a).

Les organisations internationales et régionales plaident en faveur de l'utilisation des LSLA pour faciliter l'amélioration du système éducatif. L'une de ces initiatives solidement en place est, par exemple, le programme d'évaluation régional géré par le *Laboratorio Latino-americano de Evaluación de la Calidad de la Educación* (Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation, LLECE) au Bureau

régional de l'UNESCO pour l'éducation en Amérique latine et dans les Caraïbes. Depuis 1997, ce ne sont pas moins de trois évaluations régionales qui ont été réalisées dans la région, sur les mathématiques, les sciences et la lecture. Plus récemment, des efforts considérables sont faits

pour aider les pays à produire des données et à les exploiter dans des domaines essentiels pour l'amélioration des processus et des résultats de l'apprentissage (voir **Encadré 2**). Ces initiatives et approches nouvelles représentent de grandes avancées dans la réflexion sur l'utilisation des données relatives aux apprentissages en vue d'améliorer les systèmes éducatifs.

## **Encadré 2. Initiatives régionales et mondiales récentes visant à promouvoir les évaluations à grande échelle des apprentissages**

### **SOUTIEN TECHNIQUE**

- L'Institut international de l'UNESCO pour la planification de l'éducation (IIPÉ) intensifie son soutien aux États membres (principalement sous la forme de recherches et de formation) en vue de les encourager à utiliser les données d'évaluation de l'apprentissage aux fins de la planification et de la gestion des systèmes éducatifs (UNESCO-IIPÉ, 2017).
- Le Centre pour l'éducation universelle à la Brookings Institution a lancé l'initiative OAA (Optimiser l'évaluation pour tous) afin d'aider les pays à améliorer l'évaluation, l'enseignement et l'apprentissage des compétences indispensables au XXI<sup>e</sup> siècle, en collaboration avec l'initiative NEQMAP du Bureau de l'UNESCO de Bangkok et l'initiative TALENT du Bureau de l'UNESCO de Dakar (sur lesquelles on trouvera ci-après de plus amples informations)<sup>20</sup>.
- L'Association pour le développement de l'éducation en Afrique (ADEA) a créé en 2016 le Réseau pour l'évaluation des apprentissages en Afrique (NALA) dans le cadre d'un programme de recherche visant à promouvoir les évaluations se rapportant à l'ODD 4, et plus particulièrement les évaluations des compétences non cognitives.
- L'Analyse des systèmes nationaux d'évaluation des apprentissages (ANLAS) réalisée par le Conseil australien de recherche pédagogique (ACER) avec le concours du GPE a pour objet d'aider les pays à mener à bien une analyse approfondie de leur système d'évaluation.

### **RÉSEAUX ET PARTAGE DES CONNAISSANCES**

- Le Réseau de suivi de la qualité de l'éducation dans la région Asie-Pacifique (NEQMAP) vise à améliorer la capacité des États membres de s'appuyer sur l'évaluation des acquis des élèves pour renforcer leur système éducatif.
- En Afrique subsaharienne, le Réseau des éducateurs africains pour la transformation de l'enseignement-apprentissage (TALENT) créé par l'UNESCO coopère lui aussi directement avec les États membres en vue de renforcer des éléments essentiels pour l'amélioration de l'apprentissage – les programmes d'enseignement, les politiques et formations concernant les enseignants et l'évaluation des résultats de l'apprentissage.
- En 2016, l'IIPÉ a lancé le Portail de l'apprentissage, conçu pour fournir des données actualisées, pertinentes et impartiales sur tous les aspects de l'apprentissage, de l'enseignement primaire jusqu'à la fin de l'enseignement secondaire.

### **FINANCEMENT**

- L'initiative pour l'évaluation des apprentissages mise en place par le Partenariat mondial pour l'éducation (GPE) vise à faire en sorte, au niveau national, que les investissements dans les systèmes d'évaluation de l'apprentissage se fondent sur une analyse des besoins, articulée avec le processus de formulation des politiques et systématiquement prise en compte dans les plans du secteur de l'éducation. Le GPE a récemment lancé un mécanisme de partage de connaissances et d'innovations (*Knowledge and Innovation Exchange – KIX*), lui aussi spécifiquement axé sur l'appui aux systèmes d'évaluation des acquis. Ce mécanisme s'appuie sur l'initiative Évaluation pour l'apprentissage (A4L), lancée en 2017, qui avait pour objet de renforcer les capacités et d'aider les pays à développer leurs systèmes d'évaluation.

Source : Les auteurs.

Note : Pour plus d'information sur les activités de l'UNESCO dans le domaine de l'évaluation des acquis scolaires, voir la brochure « Évaluation de l'apprentissage à l'UNESCO : assurer un apprentissage efficace et pertinent pour tous » (UNESCO, 2017c).

<sup>20</sup> Pour plus d'informations sur l'initiative OAA, voir [https://www.brookings.edu/wp-content/uploads/2017/12/overview\\_optimizing-assessment-for-all.pdf](https://www.brookings.edu/wp-content/uploads/2017/12/overview_optimizing-assessment-for-all.pdf).

Depuis quelques années, les organisations internationales se montrent de plus en plus soucieuses d'appuyer la création de solides systèmes de collecte de données. Avant même l'adoption du Programme 2030, quantité d'initiatives visant à soutenir la mesure des résultats de l'apprentissage avaient fleuri, proposant diverses options pour une métrique mondiale de la maîtrise de la lecture, de l'écriture et du calcul. On citera notamment le Comité de réflexion sur la métrique des apprentissages (2012-2016)<sup>21</sup>, le Learning Metrics Partnership (Partenariat pour la métrique des apprentissages) (2014)<sup>22</sup> et le groupe d'études sur la mesure des résultats de l'apprentissage du Centre pour le développement mondial (2013)<sup>23</sup>. En 2016, l'Alliance mondiale pour le suivi de l'apprentissage (GAML) a été lancée sous la houlette de l'Institut de statistique de l'UNESCO (ISU) afin de consolider différents processus. La GAML offre un espace de dialogue sur les politiques possibles et les compétences techniques requises en ce qui concerne la mise en place de stratégies, méthodologies et outils de suivi des résultats de l'apprentissage au niveau mondial (voir Encadré 3). L'établissement de l'Alliance représente une avancée majeure vers une meilleure synergie du fait de la collaboration qu'elle instaure entre toutes sortes de parties prenantes, dont beaucoup se disputent le « marché » de la métrique de l'apprentissage.

---

### Encadré 3. L'Alliance mondiale pour le suivi de l'apprentissage

L'Alliance mondiale pour le suivi de l'apprentissage (GAML) est une initiative multipartite conduite par l'ISU qui vise à soutenir les stratégies nationales de mesure de l'apprentissage et permettre l'établissement de rapports internationaux sur le cadre d'indicateurs relatifs à l'ODD 4. Elle regroupe les États membres des Nations Unies, des experts techniques internationaux et les partenaires d'exécution (États membres, donateurs, société civile, organismes des Nations Unies et secteur privé) en vue d'améliorer l'évaluation des acquis au niveau mondial. La tâche est confiée à des « équipes spéciales » qui réfléchissent aux questions de comparabilité mondiale des données relatives aux résultats de l'apprentissage.

La GAML s'emploie à bâtir un consensus mondial sur les mesures et rapports internationaux concernant l'apprentissage sur la base des approches méthodologiques et des pratiques les mieux adaptées aux fins recherchées. Il élabore : (1) tous les indicateurs de l'ODD 4 relatifs à l'apprentissage et aux compétences, ainsi que les outils et normes méthodologiques nécessaires pour assurer la comparabilité au niveau mondial ; (2) les normes, directives et outils conçus pour aider les pays à renforcer la mise en œuvre de leurs évaluations et à en évaluer la qualité; et (3) les outils de développement des capacités qui aident les pays à collecter, analyser et utiliser les données issues des évaluations des acquis. Le travail méthodologique de la GAML contribue à d'autres processus mondiaux.

Source : Adapté du site web ISU-GAML <http://gaml.ISU.unesco.org/>.

---

21 Pour plus d'informations concernant le Comité de réflexion sur la métrique des apprentissages : <https://www.brookings.edu/product/learning-metrics-task-force/>

22 Pour plus d'informations concernant le Partenariat pour la métrique des apprentissages : [http://uis.unesco.org/sites/default/files/documents/learning-metrics-partnership-a-capacity-support-and-policy-strengthening-initiative-to-develop-and-use-common-learning-metrics-mathematics-reading-2014-en\\_1.pdf](http://uis.unesco.org/sites/default/files/documents/learning-metrics-partnership-a-capacity-support-and-policy-strengthening-initiative-to-develop-and-use-common-learning-metrics-mathematics-reading-2014-en_1.pdf)

23 Pour plus d'informations concernant le groupe d'études et ses travaux : <https://www.cgdev.org/publication/ft/schooling-not-education-using-assessment-change-politics-non-learning>.

## Une tendance mondiale croissante qui suscite aussi des préoccupations

Les LSLA sont de l'avis général des instruments propres à aider les pays à assurer l'équité et la *qualité* dans l'éducation. Il faut donc s'attendre à ce que la multiplication et la diffusion de telles évaluations (nationales et transnationales) que l'on observe partout dans le monde aillent croissant. Et cela d'autant plus que, ces dix dernières années, les acteurs mondiaux et régionaux de l'éducation ont aussi « reconsidéré leurs identités et leurs programmes de travail » pour tenir compte des LSLA (Addey, 2018) tandis que les partenaires de développement encourageaient toujours plus les pays à s'engager dans de telles initiatives.

Du fait de ces tendances, le débat mondial actuel sur les politiques de l'éducation tourne en grande partie autour des mesures et comparaisons des résultats de l'apprentissage. Un certain nombre d'analystes ont toutefois mis en garde contre le risque que cette focalisation accrue n'influence les politiques et les pratiques éducatives au détriment des priorités plus vastes de l'ODD 4. Certaines de leurs préoccupations ont trait à la conception même des LSLA et à leurs caractéristiques intrinsèques, d'autres aux analyses, interprétations et (mauvais) usages auxquelles pourraient donner lieu les données collectées. Les chapitres qui suivent examinent ces deux écueils, en prêtant une attention particulière à leurs incidences sur les efforts de concrétisation de l'engagement mondial en faveur d'un apprentissage équitable, efficace et pertinent pour tous.

Un certain nombre d'analystes ont toutefois mis en garde contre le risque que cette focalisation accrue n'influence les politiques et les pratiques éducatives au détriment des priorités plus vastes de l'ODD 4. Certaines de leurs préoccupations ont trait à la conception même des LSLA et à leurs caractéristiques intrinsèques, d'autres aux analyses, interprétations et (mauvais) usages auxquelles pourraient donner lieu les données collectées.

# 3. L'apprentissage n'est pas seulement une affaire de mesure : les limites de la conception des évaluations à grande échelle des apprentissages

Le présent chapitre analyse les conséquences involontaires des évaluations à grande échelle des apprentissages (LSLA) qui tiennent à leur conception même. Il passe en revue quatre grandes dimensions de notre philosophie qui risquent d'être négativement affectées du fait même de la nature et de la conception des LSLA : notre manière de concevoir l'éducation et l'apprentissage, notre compréhension générale de ce qu'est une éducation de qualité, l'éventail des connaissances, compétences, attitudes et valeurs prises en considération, et notre attachement à l'inclusion et au respect de la diversité.

## Une conception limitée de l'éducation et de l'apprentissage

Nous allons voir pour commencer comment une place excessive faite aux LSLA peut avoir pour effet d'appauvrir les concepts d'éducation et d'apprentissage. Cette section passe d'abord en revue les nombreuses visées de l'éducation et les multiples dimensions de l'apprentissage. Elle examine ensuite comment l'accent mis sur les mesures quantitatives et la comparabilité au niveau des systèmes éducatifs conduit souvent à négliger d'autres modes d'évaluation mieux à même de rendre compte de certains domaines d'apprentissage, mais aussi, ce qui est plus grave, risque de détourner l'esprit d'une vision intégrée de l'éducation et de ses fins plus vastes.

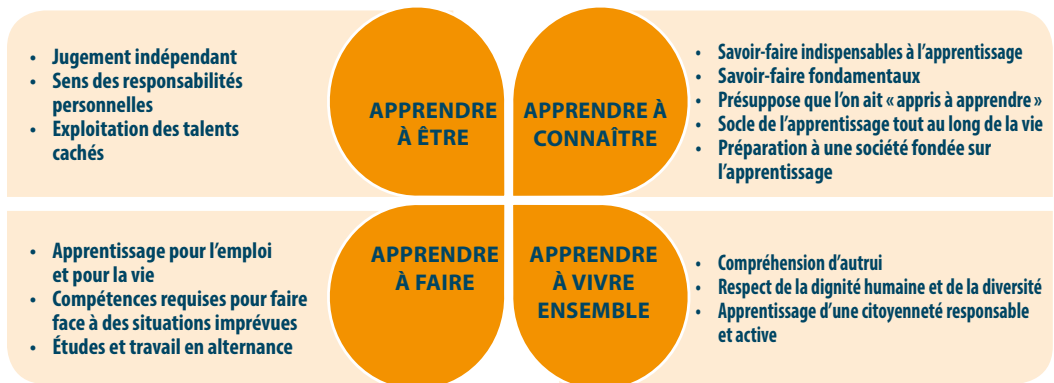
### Rappel des nombreuses visées de l'éducation

Il importe de rappeler que l'éducation sert de multiples objectifs pour les individus comme pour les communautés. Au-delà de l'acquisition des savoir-faire fondamentaux et professionnels, ainsi que des connaissances et compétences générales, l'éducation a aussi pour fonction d'inculquer les valeurs de respect de la vie, de la dignité humaine et de la diversité culturelle qu'exige l'harmonie sociale dans un monde caractérisé par la diversité (UNESCO, 2015a). Elle fait prendre conscience des choix qui doivent être faits quotidiennement pour vivre une existence saine et épanouie, et contribuer aux fondements sociaux, économiques et environnementaux du développement durable. L'éducation est indissociable des diverses dimensions du développement humain individuel et collectif, à l'égard desquelles elle joue un rôle pertinent. Il faut donc intégrer les multiples finalités sociales, économiques, culturelles, éthiques, environnementales, spirituelles et politiques de l'éducation et trouver le juste équilibre entre elles.

## Réaffirmer la pertinence des quatre piliers

L'apprentissage est un processus complexe d'acquisition ou de développement de connaissances, de compétences et d'attitudes qui, toutes, reposent sur des valeurs (UNESCO, 2017c). Cette approche intégrée est bien résumée par les quatre piliers de l'apprentissage (UNESCO, 1996) : apprendre à connaître, à faire, à être, et à vivre ensemble (voir figure 1). Apprendre à faire signifie acquérir les savoir-faire fondamentaux et transférables qui sont le socle de l'apprentissage tout au long de la vie et préparent à prendre une part active à la vie de la société. Apprendre à *faire* signifie apprendre à appliquer les connaissances et compétences acquises dans des situations familières ou imprévisibles, au travail et dans sa vie. Apprendre à être signifie développer sa propre personnalité, et acquérir l'aptitude à juger par soi-même et le sens de ses responsabilités. Apprendre à vivre ensemble implique l'aptitude à comprendre les autres et le sens du respect de la dignité humaine, de la diversité et des autres cultures, indispensables à une citoyenneté responsable et active.

**Figure 1. Les quatre piliers de l'éducation**



Source : Adapté de UNESCO, 1996.

## L'évaluation au-delà de la mesure et l'apprentissage au-delà de l'évaluation

Le champ des programmes d'enseignement et des résultats attendus des pratiques d'apprentissage est vaste. De par leur conception, toutefois, les LSLA se limitent généralement à un petit nombre de domaines d'apprentissage aisés à mesurer et qui se prêtent à des comparaisons entre systèmes ou entre pays. Ces domaines mesurables représentent les « savoir-faire fondamentaux » et/ou connaissances générales de base, comme la lecture et les mathématiques (et parfois les sciences) qui sont considérés comme essentiels pour acquérir d'autres compétences, ainsi que pour progresser

Le champ des programmes d'enseignement et des résultats attendus des pratiques d'apprentissage est vaste. De par leur conception, toutefois, les LSLA se limitent généralement à un petit nombre de domaines d'apprentissage aisés à mesurer et qui se prêtent à des comparaisons entre systèmes ou entre pays.

dans ces mêmes disciplines et aller plus loin. De ce fait, les LSLA tendent à se désintéresser de diverses dimensions de l'apprentissage qui sont importantes au regard des finalités multiples de l'éducation<sup>24</sup>. Certes, cela n'est peut-être pas foncièrement problématique. Toutefois, dans un contexte dominé par les LSLA, le risque est que nous attachions de la valeur à ce que nous mesurons plutôt que de mesurer les apprentissages auxquels nous attachons de la valeur (Biesta, 2009).

Il en est ainsi en particulier parce que le discours mondial sur l'éducation tend de plus en plus à assimiler évaluation et mesures quantitatives visant à comparer les systèmes éducatifs, en passant sous silence d'autres formes d'évaluation. L'évaluation est essentielle pour l'apprentissage. Toutefois, il est permis de se demander s'il est possible ou même souhaitable de rendre compte d'aspects importants de l'apprentissage social, civique et personnel au moyen des méthodes quantitatives traditionnelles. Les compétences cruciales dans les domaines de la communication, de la collaboration, de la créativité, de l'éthique et des relations sociales, par exemple, dépendent largement de facteurs culturels et sont fortement liés à la personnalité. Évaluer convenablement ces compétences suppose donc que les apprenants fassent la preuve de leurs attitudes et de leurs perceptions – et non pas simplement de leurs connaissances et de leur compréhension. Des formes d'évaluation différentes, comme l'examen d'un dossier, l'observation ou même des variantes des méthodes d'évaluation formative non conçues à des fins de comparaison permettraient sans doute de mieux apprécier ces qualités. Mais il n'y a guère ou pas de place pour de telles méthodes différentes aux fins du suivi des apprentissages au niveau du système éducatif.

Enfin, nous devons être conscients que, bien que l'apprentissage touche tous les domaines, les discussions concernant son évaluation ont tendance à ne s'intéresser qu'à cette part de l'apprentissage qui est systématique et organisée ainsi qu'aux acquis que l'on peut mesurer. Or, si l'apprentissage peut être volontaire, intentionnel, systématique et organisé, il est aussi souvent moins institutionnalisé et peut s'opérer, intentionnellement ou non, à la faveur de diverses activités de la vie quotidienne, sur le lieu de travail ou dans le cadre de programmes d'éducation non formelle ou de la communauté locale et répondre à un projet personnel, familial ou social (UNESCO, 2015a). L'importance accrue que le discours sur l'éducation accorde aux LSLA fait courir le risque que le concept d'apprentissage ne désigne plus que les acquis mesurables.

### **Compréhension limitée de ce qu'est une « bonne » éducation**

Du fait de cette dynamique, il est possible que, par leur participation et leur attention accrues aux LSLA, les organismes internationaux et gouvernements nationaux contribuent à créer un monde où ces évaluations déterminent en grande partie notre perception de ce qui présente de la valeur dans l'éducation et dans ses buts (Hamilton, 2017). Cette dynamique finit par restreindre

---

24 Tel a été le cas des évaluations axées sur les aspects utilitaires de l'apprentissage, et notamment sa contribution au développement économique – tout particulièrement le programme PISA. D'autres programmes d'évaluation se sont caractérisés dans le passé par une approche plus inclusive ou plus exhaustive et ont été, de ce fait, attentifs à un plus grand nombre de dimensions de l'apprentissage. Pour des raisons qui n'entrent pas dans le cadre de la présente étude, toutefois, les programmes d'évaluation existants témoignent d'une convergence croissante dans les domaines mesurés qui se traduit par une attention accrue à la maîtrise de la lecture, de l'écriture et des mathématiques.

notre compréhension générale de ce qui constitue une éducation de qualité (Biesta, 2009), et risque de détourner la réflexion d'importantes dimensions de l'éducation qui ne doivent pas être oubliées. Le plus préoccupant est que ces évolutions n'affectent pas seulement la théorie. Comme nous le verrons dans le chapitre suivant, elles risquent en effet d'avoir des effets concrets lorsque des programmes politiques les institutionnalisent, et ont le pouvoir de transformer les pratiques et l'identité des enseignants, les enseignements formels et invisibles dispensés dans les écoles, et les priorités de différents acteurs de l'éducation.

L'évaluation est essentielle pour l'apprentissage. Toutefois, il est permis de se demander s'il est possible ou même souhaitable de rendre compte d'aspects importants de l'apprentissage social, civique et personnel au moyen des méthodes quantitatives traditionnelles.

## Une compréhension restreinte de la qualité de l'éducation

La présente section examine comment l'importance croissante que l'on accorde aux résultats de l'apprentissage, en tant que dimension essentielle d'une éducation de qualité, peut détourner l'attention d'autres paramètres essentiels (par exemple les résultats sur le plan social, l'environnement offert par l'école et la salle de classe, les procédures d'enseignement et d'apprentissage), et/ou le rôle de macro-facteurs influant sur la réussite de l'apprenant. Après une brève présentation de la nature pluridimensionnelle du concept de qualité de l'éducation, elle analyse l'approche équilibrée qui est nécessaire pour prendre en compte la totalité des différents éléments et indicateurs d'une éducation de qualité.

### La qualité de l'éducation : un concept pluridimensionnel

Les approches traditionnelles utilisées pour mesurer la qualité de l'éducation s'appuyaient sur des indicateurs indirects de l'offre et des intrants correspondant à des facteurs qui influent sur le progrès de l'enseignement et de l'apprentissage, comme le nombre d'élèves par enseignant, le pourcentage d'enseignants ayant reçu une formation ou les dépenses par élève en pourcentage du PIB<sup>25</sup>. Plus récemment, les résultats de l'apprentissage sont devenus un indicateur privilégié pour mesurer la qualité de l'éducation. Le *Rapport mondial de suivi sur l'éducation* (Rapport GEM), par exemple, a récemment proposé un cadre révisé de conceptualisation de la qualité de l'éducation (UNESCO, 2017d) qui met notamment l'accent sur les résultats de l'apprentissage pour les individus et la société, tels que déterminés par les scores obtenus aux tests des LSLA.

Comme le souligne toutefois le Rapport GEM, les résultats individuels ne peuvent être interprétés isolément ou comme seule mesure valide de la qualité de l'éducation. D'autres dimensions de l'apprentissage doivent être prises en considération aussi. Il s'agit notamment des *retombées pour la société* (par exemple le fait que soient encouragés les comportements

25 Ces approches sont passées en revue dans Tawil *et al.* (2012). Certaines se réfèrent, par exemple, au modèle de Santiago établi par l'UNESCO (UNESCO Santiago, 2007), qui propose cinq dimensions de la qualité de l'éducation tenant compte du point de vue de divers acteurs de l'éducation ; citons encore le cadre d'analyse défini par Tikly et Barret (2007), qui préconise d'analyser le contexte historique, culturel, socioéconomique et politique pour comprendre comment opèrent les systèmes éducatifs ; le modèle établi par Nikel et Lowe (2010), qui proposent sept dimensions conceptuelles les amenant à comparer la qualité de l'éducation à une sorte de « tissu » qui est d'autant plus solide qu'il est « étiré » ou maintenu en tension.



propices à un développement social et environnemental durable, la citoyenneté mondiale, le vivre-ensemble et le respect de la diversité) ainsi que *l'environnement offert par les écoles et les salles de classe*. Ce dernier paramètre comprend le processus d'enseignement-et d'apprentissage, la présence d'enseignants convenablement formés et motivés, ainsi que la rigueur du processus pédagogique, la capacité de l'école d'exercer un vigoureux leadership, la participation de la communauté et le caractère démocratique de la gouvernance, ainsi que des structures et intrants matériels adéquats<sup>26</sup>.

Plus généralement, comprendre les niveaux de résultat de l'apprentissage ou leur évolution au fil du temps exige la prise en considération de *macro-facteurs*, tels que le contexte économique, politique et social ou la structure et l'organisation du système éducatif (politiques, financement, programmes d'enseignement, normes et reddition des comptes, recrutement et incitations, perfectionnement professionnel). Faute de quoi, analyses et perspectives sont inévitablement faussées. L'amélioration de la qualité de l'éducation ne conduit pas nécessairement à de meilleurs résultats de l'apprentissage en raison de variation dans les acquis des élèves imputables à des facteurs liés au contexte scolaire, mais aussi aux caractéristiques du milieu familial et communautaire (Ungerleider, 2006). Si bon nombre de LSLA sont précisément conçues pour permettre de mieux comprendre l'impact de tels facteurs, le rôle central de ces facteurs peut parfois échapper à l'attention du public, des responsables politiques et des praticiens de l'éducation – en particulier si les résultats publiés privilégient les aspects comparatifs et, dans le cas d'évaluations transnationales, insistent sur le classement de tel pays par rapport à tel et tel autres pays. Les LSLA sont nécessaires pour suivre la performance des systèmes éducatifs et améliorer l'apprentissage et l'équité dans les écoles, mais elles pourraient être plus utiles encore si elles étaient considérées en même temps que toute une série d'autres facteurs, scolaires et extrascolaires

### **L'évaluation : un outil nécessaire mais insuffisant pour rendre compte de la qualité de l'éducation**

Pour évaluer la qualité de l'éducation, il est essentiel de bien comprendre diverses autres dimensions qui concernent l'apprenant, l'enseignant, l'école, le système éducatif et le contexte économique, social et politique (Colclough, 2012 ; UNESCO, 2016b ; Zadjia, 2014). Les données relatives à ces dimensions ne sont pas toujours fournies par les LSLA, dont les questionnaires connexes sont parfois insuffisants, mais par d'autres sources, telles que les enquêtes administratives ou auprès des ménages. Considérer et analyser les données issues des LSLA à la lumière de diverses autres données générées à l'échelon national permettrait donc d'obtenir un tableau plus juste des différentes dimensions d'une éducation de qualité.

Certes, les chercheurs et spécialistes travaillant pour des organismes gouvernementaux ou des cabinets d'évaluation procèdent fréquemment de la sorte. Mais les rapports moins spécialisés, tels que ceux qui sont mis en avant par les médias, donnent souvent moins d'importance à ces

---

26 Une position similaire est défendue par Unterhalter et Brighouse (2007), qui, adeptes de l'approche axée sur les capacités, notent que les résultats de l'apprentissage sont une mesure inappropriée ou insuffisante des principales dimensions de la qualité de l'éducation, en dehors de ses visées instrumentales, notamment les valeurs intrinsèques et de principe de l'éducation.

rapprochements (ou n'en tiennent aucun compte). De plus, les analyses équilibrées et pluridimensionnelles semblent se raréfier à mesure que les LSLA continuent de gagner du terrain. Les résultats des LSLA sont en effet de plus en plus acceptés comme des diagnostics complets de la qualité de l'éducation (Hanushek et Woessmann, 2015 ; Komatsu et Rappleye, 2017), et ces évaluations ont été décrites comme un « élément majeur » (ACER-ISU, 2017) ou « une référence essentielle » (Friedman *et al.*, 2016) en la matière. Si cela témoigne en général d'une juste préoccupation pour les acquis, considérés comme un résultat important des études scolaires, le risque est qu'une conception étroite de la qualité de l'éducation l'assimile aux seuls progrès mis en évidence par les LSLA (Sayed *et al.*, 2018 ; Smith, 2016).

En définitive, les niveaux d'acquisition des connaissances et compétences mesurés par les LSLA sont un paramètre nécessaire, mais non le seul à l'être pour mesurer la qualité de l'éducation.

Faire des résultats des évaluations des apprentissages l'unique moyen d'évaluer la qualité de l'éducation a de multiples conséquences. Cela risque de détourner l'attention d'importants aspects de la qualité de l'éducation, comme la nécessité d'instaurer une plus grande équité, d'offrir des environnements d'apprentissage sûrs, adéquats et efficaces, de déployer des enseignants qualifiés et de développer leurs capacités, de veiller à la pertinence des programmes d'enseignement, et de faire preuve d'un solide leadership au niveau des établissements. Cela peut aussi conduire à négliger d'autres dimensions sociales, économiques et contextuelles de la qualité de l'éducation. En définitive, les niveaux d'acquisition des connaissances et compétences mesurés par les LSLA sont un paramètre nécessaire, mais non *le seul à l'être pour mesurer la* qualité de l'éducation. Il est même permis de se demander si l'on peut correctement apprécier la qualité sans prendre en compte ces autres dimensions sociales, économiques et contextuelles.

## Une vision étriquée de ce qui présente de la valeur dans les apprentissages

Comme il a été dit plus haut, les LSLA peuvent avoir pour effet de réduire notre perception des multiples finalités de l'éducation et des dimensions de l'apprentissage. Cette section s'intéresse à l'un des mécanismes par lesquels s'opère cette réduction, à savoir la limitation du nombre de domaines d'apprentissage ciblés par la plupart des LSLA. Elle examine tout d'abord les différentes modalités de cette réduction des programmes d'enseignement, avant de donner des exemples des efforts qui sont faits pour prendre en compte, dans les LSLA, un éventail plus large de connaissances, de compétences, d'attitudes et de valeurs, et des obstacles auxquels se heurtent ces efforts. Enfin, elle s'interroge sur le bien-fondé de ces derniers et sur les risques qu'ils comportent.

### Limitation du nombre de domaines et de sous-domaines

Nous l'avons vu, le nombre limité de domaines et d'aptitudes évalués apparaît comme une conséquence intrinsèque de la conception des LSLA. Ces évaluations se concentrent par nécessité sur un petit nombre de matières et d'aptitudes – c'est-à-dire, plus précisément, sur les domaines d'apprentissage mesurables et qui se prêtent à des comparaisons à l'intérieur

d'un système éducatif ou entre pays, ou qui sont considérés comme correspondant à des compétences fondamentales ouvrant la voie à de plus amples acquis dans d'autres domaines.

Certains font donc valoir que le champ d'investigation des LSLA est doublement trop étroit. Premièrement, les LSLA n'évaluent pas adéquatement l'ensemble des connaissances dans un large éventail de domaines, tels que les arts ou l'initiation aux médias et à l'information dans nos cybersociétés, ni les compétences variées que les écoles sont censées faire acquérir, comme l'exercice de la citoyenneté ou des responsabilités à l'égard de l'environnement. Deuxièmement, au sein même des domaines évalués, les LSLA ne sont pas suffisamment attentives au spectre complet des connaissances disponibles. Dans le domaine des compétences linguistiques, par exemple, elles laissent souvent de côté des sous-domaines tels que la connaissance de la littérature, maîtrise de l'écrit et la compréhension orale, qui sont pourtant d'importants objectifs de cette discipline (Ungerleider, 2006).

### **Au-delà de la maîtrise de la lecture, de l'écriture et du calcul ?**

Telles qu'elles sont conçues, les LSLA tendent donc à sous-estimer les finalités multiples de l'éducation et les dimensions de l'apprentissage et, partant, les principes qui sont au cœur de l'agenda Éducation 2030. C'est là une conséquence de leur contenu, ainsi que de la manière dont elles sont réalisées et de la formulation des questionnaires utilisés. Les tests à choix multiples ou appelant une réponse courte sont conçus pour faciliter la notation par rapport à un ensemble défini de rubriques, ce qui limite les possibilités d'évaluer les perceptions, attitudes et jugements. De même, la plupart des évaluations ne cherchent pas nécessairement à recueillir des données sur des aspects indispensables à un mode de vie sain, ou sur des valeurs et des attitudes allant dans le sens de la démocratie, de la protection des droits de l'homme, du respect de la diversité culturelle, d'une citoyenneté active et responsable et de la viabilité de l'environnement<sup>27</sup>.

Certains types de connaissances, compétences, attitudes et valeurs, bien qu'étant des éléments essentiels de l'agenda Éducation 2030, ne sont pris en compte que partiellement (et non sans difficultés) par les LSLA. Un examen des items utilisés par ces instruments montre à quel point il est malaisé d'intégrer un large éventail de compétences (domaines d'apprentissage, connaissances, compétences, attitudes, valeurs) dans les méthodes de test habituelles.

### **Des modes de vie sains : des questions insuffisamment étudiées et qui perdent du terrain**

L'éducation est considérée comme l'un des moyens les plus efficaces d'améliorer la santé de l'individu et de faire en sorte que ces bénéfices soient transmis aux générations futures. Inversement, la santé est un facteur déterminant de l'apprentissage (UNESCO, 2016b ; 2016c). Pourtant, la plupart des LSLA ne cherchent pas expressément à promouvoir une éducation qui inciterait à un mode de vie sain. En ce qui concerne les évaluations nationales, la santé, la nutrition et le bien-être de enfants – ainsi que de leurs parents ou tuteurs – sont le plus souvent mesurés sous l'angle du développement de la petite enfance car ce sont censés être des conditions essentielles pour que l'enfant apprenne et soit prêt à aborder l'enseignement primaire

---

<sup>27</sup> La difficulté d'évaluer les acquis dans ces domaines ne peut être dissociée du fait que ces derniers constituent rarement des matières à part entière dans les programmes d'enseignement nationaux, mais sont le plus souvent abordés comme des thèmes transversaux.

(UNESCO *et al.*, 2017). Il n'y a guère d'indications que des LSLA ciblant les enfants en âge de faire des études primaires ou secondaires, ou même des adultes, fassent une place centrale à la notion de modes de vie sains. Cela est vrai en particulier des évaluations élaborées à l'échelon national. Certes, les questionnaires de contexte collectent fréquemment des données sur les pratiques et habitudes des élèves en matière de santé ou de nutrition. Ils ne contiennent généralement que des questions visant à déterminer si les apprenants se trouvent ou non *dans de bonnes conditions* pour apprendre, c'est-à-dire s'ils ont accès à des services de soins de santé adéquats, s'ils reçoivent une alimentation appropriée et suffisante à l'école, et ainsi de suite. Aucune information n'est en général collectée concernant leur attitude à l'égard de la prévention des maladies ou leurs opinions concernant des comportements tels que la pratique de l'hygiène et les habitudes alimentaires ou les comportements sexuels à risque. De même, si certains items ont trait à la santé, les tâches ont généralement pour but de mesurer des compétences spécifiques, comme la compréhension de l'écrit ou la résolution d'un problème comportant plusieurs étapes.

Par exemple, le test sur les compétences en science de l'enquête PISA 2015 intitulé « Courir par temps chaud » (OCDE, 2015)<sup>28</sup>, comprenait cinq questions sur l'interaction entre santé et environnement. Il s'agissait d'évaluer une série d'aptitudes : interpréter la figure, manipuler une variable et comparer les résultats de deux essais, inférer du texte et de la simulation la fonction de la transpiration, utiliser les connaissances procédurales pour expliquer comment les données collectées confortaient la réponse donnée, et développer une hypothèse, en extrapolant à partir des données qui pouvaient être directement obtenues au moyen de la simulation. L'apprenant n'était pas censé avoir une quelconque connaissance préalable sur le sujet, les réponses aux questions pouvant être déduites du stimulus. Il ne s'agissait pas non plus d'évaluer des décisions/actions/jugements concernant un exercice physique par temps chaud. Par conséquent, les questionnaires PISA s'appuient sur des matériaux relatifs à la santé pour évaluer les aptitudes en science et en résolution de problèmes, mais tous les indices sont fournis dans le stimulus, de sorte que les aspects touchant le jugement de l'élève sont exclus du champ de l'évaluation.

Certains efforts ont néanmoins été faits pour donner plus d'importance à la notion de modes de vie sains. C'est ainsi que, compte tenu du nombre élevé de cas de VIH/sida en Afrique subsaharienne, l'enquête SACMEQ 2007 comportait un module sur le VIH et le sida visant à déterminer les attitudes des enfants à l'égard des personnes vivant avec la maladie, leur perception des cours sur le VIH et le sida et leurs sources d'information préférées au sujet de la maladie<sup>29</sup>. On a également étudié le traitement réservé, dans la formation des enseignants, aux questions relatives au VIH et au sida, qui occupent une place centrale dans toute stratégie efficace de prévention par l'éducation (Dolata, 2011).

## **Éducation à la citoyenneté mondiale et éducation en vue du développement durable : des domaines de mieux en mieux définis et évalués**

L'éducation en vue du développement durable (EDD) et l'éducation à la citoyenneté mondiale sont deux dimensions essentielles de l'agenda Éducation 2030. Dès 1974, la Conférence générale

28 L'item de PISA 2015 est accessible en ligne au lien suivant : <http://www.oecd.org/pisa/PISA2015Questions/platform/index.html?user=&domain=SCI&unit=5623-RunningInHotWeather&lang=fra-ZZZ>

29 Notons toutefois que le module sur le VIH et le sida ne figurait plus dans la dernière enquête SACMEQ (2013-2014).

de l'UNESCO a adopté la *Recommandation sur l'éducation pour la compréhension, la coopération et la paix internationales et l'éducation relative aux droits de l'homme et aux libertés fondamentales*. La Recommandation de 1974 encourageait les États à prendre les mesures institutionnelles et pédagogiques nécessaires pour incorporer une série de principes (respect de la diversité culturelle, promotion de la paix et résolution non violente des conflits, équité, protection de la planète, etc.) dans leurs programmes d'enseignement (UNESCO, 1974).

Aussi utiles que soient ces efforts, concevoir des mesures permettant d'évaluer les connaissances, compétences, valeurs et attitudes correspondantes est une tâche complexe. Cela tient en particulier au fait que les outils existants se limitent pour beaucoup à des questionnaires d'auto-évaluation qui créent souvent un biais parce que les participants s'efforcent généralement de formuler leurs réponses d'une manière qui leur semble socialement acceptable. Qui plus est, il peut être difficile de prendre en compte dans les évaluations transnationales des normes culturellement pertinentes propres à des pays ou des régions particuliers tout en préservant la comparabilité transnationale des données collectées. Pour l'heure, l'expertise technique qui permettrait de remédier à ces difficultés semble encore embryonnaire. De plus, il n'est pas certain que l'EDD sera prise en compte par les évaluations nationales des apprentissages dans un proche avenir. En Europe, les connaissances relatives à la citoyenneté mondiale sont généralement incluses dans ces évaluations, mais les attitudes et compétences des élèves en la matière tendent à l'être moins souvent (De Coster et Sigalas, 2017).

La difficulté d'intégrer l'éducation à la citoyenneté dans les évaluations nationales vient en particulier de ce que l'enseignement et l'apprentissage en la matière font intervenir de multiples approches, les unes axées sur le sujet, les autres de nature transversale, et d'autres encore prenant la forme d'activités hors programmes ou entrant dans le cadre de la participation à la vie de l'établissement (De Coster et Sigalas, 2017). On s'efforce néanmoins de prendre en compte l'EDD. Certaines évaluations transnationales de l'apprentissage, comme le Programme international pour le suivi des acquis des élèves (PISA) de l'OCDE et l'Étude internationale sur l'instruction civique et l'éducation à la citoyenneté (ICCS) de l'Association internationale pour l'évaluation du rendement scolaire (IEA) tentent de mesurer certains aspects de l'EDD et de l'éducation à la citoyenneté. De même, l'étude ICCS, qui a été menée dans 38 pays, est conçue pour déterminer, sans référence aux programmes d'enseignement, comment les jeunes se préparent à leur rôle de citoyen (IEA, 2016b). L'étude rend compte des connaissances des élèves et de leur compréhension des concepts et thèmes liés au civisme et à la citoyenneté, ainsi que de leurs croyances, de leurs attitudes et de leurs comportements. L'**encadré 4** offre un exemple de l'approche adoptée par l'ICCS pour mesurer les connaissances relatives à la diversité des confessions religieuses et les attitudes et dispositions à l'égard de la tolérance. L'IEA a elle aussi encouragé l'élaboration de rapports régionaux analysant le contexte de l'instruction civique et de l'éducation à la citoyenneté, ainsi que la perception qu'ont les élèves des institutions publiques et du gouvernement, ainsi que leur vision de la coexistence pacifique et de la cohésion et de la diversité sociales<sup>30</sup>.

---

30 Pour mesurer ces dimensions, il est laissé aux élèves un espace libre où exprimer leurs opinions par écrit, ce qui accroît encore la difficulté de noter leurs réponses.

#### Encadré 4. Respect et coexistence – Test de l'étude internationale sur l'instruction civique et l'éducation à la citoyenneté 2009

Voici une vignette autocollante que les gens peuvent acheter sur Internet.



On y voit des symboles qui représentent différentes visions du monde. Ils sont assemblés comme pour former le mot anglais « coexist », qui signifie « vivre ensemble ».

CI2COM1

- 1** À quoi sert le plus probablement cette vignette ?
- À montrer qu'au fond, les différentes façons de penser sont semblables.
  - À montrer que les gens devraient bien réfléchir à leur croyance.
  - À montrer que les gens peuvent s'accepter les uns les autres, même s'ils ont des croyances différentes.
  - À montrer que les gens qui ont une vision différente du monde ne pourront jamais vivre heureux ensemble.

Source : Brese et al. 2011, page 7 (reproduit avec autorisation).

Dans les pays du Sud, le programme de mesure des apprentissages primaires en Asie du Sud-Est (SEA-PLM) est actuellement la seule évaluation transnationale qui comporte un module mesurant les attitudes et comportements en matière de citoyenneté mondiale. Ce module est centré sur les attitudes et les valeurs, plutôt que sur les acquis cognitifs, les comportements et les compétences, et il est conçu pour refléter des valeurs universelles, et pas seulement les valeurs des nations de l'Asie du Sud-Est. Ainsi, ce cadre d'évaluation récemment mis au point reconnaît les défis que soulève l'évaluation d'un domaine novateur, mais suppose expressément que l'évaluation soit conçue d'une manière culturellement appropriée et cohérente avec les particularités de la région de l'Asie du Sud-Est, tout en tenant compte de son hétérogénéité (Parker et Fraillon, 2016).

#### L'expression écrite aide à mesurer la créativité, la curiosité et la valeur accordée à la culture

Il y aurait lieu de noter que les LSLA n'évaluent pas la maîtrise de l'expression écrite, qui est une compétence fondamentale nécessaire à la communication, aux apprentissages ultérieurs et à une pleine participation aux activités économiques, politiques et sociales, ainsi qu'à de

nombreux aspects de la vie quotidienne<sup>31</sup>. À l'ère du numérique et des économies du savoir, la communication personnelle et sociale passe de plus en plus par l'écrit, y compris sur les téléphones mobiles et dans les médias sociaux. L'évaluation des compétences en matière d'écriture ou de leur utilisation en vue de mesurer les acquis dans des domaines tels que la créativité, la curiosité et l'appréciation des cultures, sort en général elle aussi du cadre des LSLA. Cela tient au fait qu'un texte écrit est, de par sa nature subjective, particulièrement difficile à mesurer d'une manière qui préserve la comparabilité entre systèmes.

Étant donné l'importance des compétences en matière d'écriture et le fait que ces compétences sont notées lors des examens, des efforts sont faits pour inclure cette dimension dans les LSLA. Bien qu'encore au stade expérimental, l'étude SEA-PLM représente la première initiative transnationale visant à mesurer la *maîtrise de l'écrit*, comprise comme la capacité de créer du sens, à travers la rédaction de textes écrits, pour s'exprimer et de communiquer avec autrui en réponse à des besoins personnels, sociétaux, économiques et civiques. Cette définition « reconnaît que la maîtrise de l'écrit est plus que la capacité d'orthographier correctement des mots et de les tracer de manière lisible. Elle laisse entendre que l'expression écrite se pratique dans un contexte donné, pour des destinataires particuliers et dans un but spécifique. Elle englobe toute une série de compétences cognitives telles que la capacité de gérer et d'organiser des idées, d'utiliser un vocabulaire et d'utiliser des connaissances relatives aux structures linguistique et textuelles » (SEA-PLM *et al.*, 2017, pages 30-31).

### **Mesurer la capacité de participer à une réflexion collective ainsi que les compétences socio-affectives**

En général, les LSLA excluent toute mesure de la réflexion collective des compétences correspondantes. Ces compétences sont cependant essentielles à la production et à la maturation d'idées nouvelles (Hannon *et al.*, 2011) qui, de plus en plus, nécessitent une coopération au sein d'un groupe ou entre plusieurs groupes d'individus réfléchissant et travaillant ensemble et s'appuyant sur les connaissances existantes pour en créer de nouvelles (OCDE, 2018a). L'aptitude au travail et à la réflexion en collaboration (c'est-à-dire la capacité de nouer des relations positives avec autrui, de coopérer, de travailler en équipe, de gérer et résoudre un conflit, etc.) est difficile à mesurer au moyen de tests individuels. Cette aptitude et cette capacité sont de plus en plus nécessaires aux jeunes pour s'adapter aux exigences en constante évolution du marché du travail.

S'il s'avère particulièrement malaisé de mesurer la prise de décisions, la coordination et la production au sein d'un groupe au moyen de tests individuels, des tentatives sont faites pour déterminer les capacités individuelles dans des situations de collaboration. Le programme PISA, par exemple, utilise des tests informatisés pour évaluer l'aptitude des élèves à la collaboration en les confrontant à des personnages virtuels représentant les membres d'une équipe disposant de compétences et capacités variées. Cette méthode offre le degré élevé de contrôle et de

---

31 Cela est indéniablement lié au fait que les campagnes d'instruction et d'alphabetisation de masse passées ont été centrées sur la lecture plus que sur l'écriture.

standardisation qu'exige ce type de mesure. Elle permet aussi d'observer des élèves dans un certain nombre de situations où ils sont appelés à collaborer et de procéder à des mesures dans le temps limité correspondant à un test (OCDE, 2017). De manière plus générale, ces efforts demeurent toutefois essentiellement embryonnaires du fait que les éventuelles difficultés sont encore mal comprises.

Étant donné leur valeur intrinsèque, les compétences socioaffectives sont un autre objet de mesure auquel on accorde une attention accrue en tant que facteur déterminant de progrès dans l'apprentissage. C'est ainsi que le Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation (LLECE) travaille actuellement à un module sur les compétences socio-affectives pour la prochaine évaluation régionale – ERCE 2019. Ce module, qui a fait l'objet d'un essai pilote en 2018, vise à répondre à la conception générale de la qualité de l'éducation défendue par l'agenda Éducation 2030 et à examiner les effets de ces compétences sur les résultats de l'apprentissage et les facteurs dont dépendent ces compétences.

### **Pertinence accrue des compétences numériques**

Le débat sur la mesure des compétences numériques n'a trouvé un écho auprès des acteurs mondiaux de l'éducation que tout récemment. Fait remarquable, cela est dû en grande partie à l'inclusion des compétences numériques et de l'initiation aux médias et à l'information dans la cible d'ODD 4.4, qui est centrée sur les compétences exigées des jeunes et des adultes dans le monde du travail. Cela explique pourquoi les travaux conceptuels et techniques menés au niveau mondial ont été quelque peu déconnectés des débats sur l'évaluation des enfants d'âge scolaire.

Malgré leur nombre croissant, les initiatives visant à évaluer les compétences numériques et la maîtrise des TIC, sont de nature extrêmement hétérogène, qu'il s'agisse des contenus ou des autorités responsables (ISU, 2018c). De plus, les évaluations qui sont expérimentées semblent l'être presque exclusivement dans des pays à revenu élevé, tandis que les initiatives transnationales ciblant les élèves des niveaux élémentaires sont rares. Fait exception l'Étude internationale sur l'initiation à l'informatique et à l'information de l'IEA. Mais les autres évaluations de ce type déjà en place portent en général sur la population adulte.

Quoi qu'il en soit, ces initiatives ont pour la plupart en commun de se concentrer sur un nombre assez faible de compétences, comme la capacité d'exploiter l'information numérique et certaines compétences techniques, tandis que d'autres aspects, notamment la collaboration en ligne et la résolution de problèmes avec l'aide des TIC, sont moins fréquemment pris en considération (Siddiq *et al.*, 2016). Là encore, ces tendances donnent à penser que, de par leur conception, les LSLA ne peuvent aider à évaluer qu'un éventail relativement étroit de compétences. Cela pose problème, notamment parce que la maîtrise de compétences numériques de plus en plus complexes et sophistiquées est une condition préalable à un apprentissage plus poussé. En d'autres termes, l'apprentissage dans divers domaines est grande partie subordonné à la capacité de l'apprenant de réaliser des opérations techniques en rapport avec les TIC, mais aussi de s'attaquer à des problèmes complexes, de porter un regard critique sur les informations, etc. (UNESCO, 2018). Certaines initiatives nouvelles tentent précisément de prendre en compte un ensemble plus complet de compétences liées aux TIC malgré les difficultés qu'il y a à les mesurer.



Tel est le cas du projet *Digital Kids Asia Pacific* lancé par le Bureau de l'UNESCO à Bangkok, qui a récemment élaboré un cadre régional d'évaluation des compétences numériques nécessaires à l'exercice de la citoyenneté qui couvre toute une série de domaines, dont la sécurité numérique, l'intelligence affective et la créativité. Ce cadre d'évaluation a fait l'objet d'un essai pilote dans quatre pays et devrait fournir des éléments factuels permettant de comprendre comment, au-delà de leurs compétences de base, les enfants utilisent les TIC<sup>32</sup>.

### **Une marge de manœuvre : appréhender d'autres domaines du savoir**

Pour résumer, et comme le suggèrent les exemples ci-dessus, des organisations et des gouvernements tentent d'appréhender d'autres domaines du savoir, comme les compétences transversales, transférables ou dites du XXI<sup>e</sup> siècle, y compris les compétences et les savoir-faire « non cognitifs » liés au développement durable, à la citoyenneté dans le contexte de la mondialisation et à la santé, ainsi que ceux qui relèvent de la maîtrise du numérique. Ces acquis étant considérés comme essentiels pour être un travailleur et un citoyen efficace dans les sociétés du savoir, l'OCDE, l'IEA, ainsi que les organismes gérant des évaluations régionales ont lancé diverses initiatives visant à les mesurer. Cette tendance témoigne de l'acceptation d'une vision intégrée de l'apprentissage dans laquelle ce ne sont pas seulement les connaissances générales et les « compétences fondamentales », mais aussi des compétences plus spécifiques en rapport avec la citoyenneté mondiale, le développement durable et la santé, qui sont considérées comme essentielles et pertinentes dans le monde d'aujourd'hui. On s'intéresse de même à toute une série de connaissances, compétences, attitudes et valeurs fondamentales nécessaires pour communiquer, collaborer, résoudre des problèmes et nouer des relations sociales de manière efficace.

Dans le même temps, la meilleure façon de prendre en compte la conscience et la compétence dans des domaines liés à la socialisation et à l'acquisition de valeurs qui semblent dépendre du contexte culturel est sans doute de recourir à des formes d'évaluation non traditionnelles qui font partie intégrante du processus d'enseignement et d'apprentissage. Malgré l'intérêt que suscitent ces compétences, l'utilité de mesures réalisées à l'échelle mondiale demeure donc discutable. De plus, ces efforts pourraient se traduire par une plus grande fréquence des tests et un accroissement considérable des ressources financières, et du temps, consacrés à l'évaluation.

En somme, on perçoit une certaine contradiction entre la nécessité d'élargir et de diversifier le champ des LSLA et l'impossibilité et l'inutilité d'y recourir pour évaluer tous les domaines d'apprentissage. Le corollaire de ce constat est que l'on ne peut remédier aux effets occultant des

On ne peut remédier aux effets occultant des LSLA en cherchant exclusivement à diversifier les épreuves. La solution serait plutôt d'explorer les moyens d'intégrer les LSLA dans des systèmes d'évaluation plus larges, en les combinant à d'autres formes d'évaluation de telle façon qu'elles ne monopolisent pas l'attention.

---

32 Pour plus d'informations concernant cette initiative, voir : [https://teams.unesco.org/ORG/fu/bangkok/public\\_events/Shared%20Documents/EISD/2017/Oct2017%20-%20KFIT%203%20Launch%20-%20Dig%20Citizenship/DKAP-project-brief.pdf](https://teams.unesco.org/ORG/fu/bangkok/public_events/Shared%20Documents/EISD/2017/Oct2017%20-%20KFIT%203%20Launch%20-%20Dig%20Citizenship/DKAP-project-brief.pdf).

LSLA en cherchant exclusivement à diversifier les épreuves. La solution serait plutôt d'explorer les moyens d'intégrer les LSLA dans des systèmes d'évaluation plus larges, en les combinant à d'autres formes d'évaluation de telle façon qu'elles ne monopolisent pas l'attention.

## Limitation des catégories d'apprenants qui retiennent notre attention

La présente section s'interroge sur les limites imposées par les LSLA en termes d'inclusion. Après une analyse du potentiel des LSLA en tant que vecteurs d'équité, elle examine les différentes façons dont certains groupes sont exclus des exercices d'évaluation – ou y participent dans des conditions relativement défavorables. Enfin, elle passe en revue diverses tentatives récentes visant à rendre les évaluations plus inclusives, ainsi que les difficultés et les risques qu'elles comportent.

### L'évaluation au service de l'équité

L'agenda Éducation 2030 engage sans ambiguïté à œuvrer en faveur de l'équité et de l'inclusion – considérées l'une et l'autre comme les pierres angulaires d'un programme véritablement transformateur. L'équité implique que le processus éducatif soit ouvert à tous, et que tout soit fait pour que chacun, quelles que soient ses origines, acquière des connaissances et des compétences. Les appels à l'équité *dans l'apprentissage* (Wagner, 2018) représentent précisément la volonté de répondre à la nécessité de relever les niveaux d'instruction dans les couches les plus désavantagées de la population.

Cette approche a des conséquences fondamentales sur le plan de la mesure des apprentissages. Les évaluations sont censées prendre en considération les priorités politiques en matière d'équité dès le début du processus et collecter les informations permettant de mettre en relation les résultats des tests avec les politiques de lutte contre les inégalités. Plus précisément, cette vision nouvelle de l'éducation exige que les données soient ventilées selon un éventail de catégories aussi large que possible (UNESCO, 2016c). Les LSLA peuvent effectivement nous aider à comprendre les différentes dimensions de l'équité au sein du système éducatif, y compris l'équité entre régions, l'équité entre établissements, l'égalité des genres, l'équité sur le plan linguistique, l'équité par rapport à la situation socioéconomique et l'équité à l'égard de certains sous-groupes (par exemple élèves handicapés, ou immigrants/réfugiés de fraîche date) (Brink, 2018).

On attend donc des évaluations qu'elles soient inclusives, c'est-à-dire pertinentes pour un nombre aussi grand que possible de membres de la population cible, et qu'elles répondent de manière efficace aux besoins de tous les apprenants.

On attend donc des évaluations qu'elles soient inclusives, c'est-à-dire pertinentes pour un nombre aussi grand que possible de membres de la population cible, et qu'elles répondent de manière efficace aux besoins de tous les apprenants (UNESCO, 2017b). En d'autres termes, l'échantillon d'apprenants ne doit pas délibérément exclure certains groupes – par exemple

les personnes souffrant d'un handicap<sup>33</sup>. Il faut en outre que les élèves passant les épreuves dans toutes les régions du pays ou dans différents pays puissent comprendre les questions et y répondre de la même manière, quels que soient leurs origines ethniques, leur statut par rapport à l'immigration, leur sexe ou leur lieu de résidence.

Le souci de respecter ces principes d'équité et d'inclusion oblige à être très rigoureux en matière de mesure. Il faut par exemple traduire les matériels de tests et les adapter à différents groupes de population, ou aux enfants ayant des besoins spéciaux. Il faut aussi faire en sorte que les groupes vulnérables ou minoritaires ne soient pas sous-représentés dans les échantillons et veiller à ce que le nombre d'élèves issus de ces groupes soit suffisant pour que l'analyse des résultats ait une réelle portée. Il peut être en outre nécessaire d'adapter les instruments et d'ajuster les conditions dans lesquelles les épreuves sont organisées de telle sorte que les enfants appartenant à ces groupes aient des chances égales de les passer. De même, il convient d'ajuster les questionnaires de contexte pour s'assurer que toutes les informations pertinentes seront collectées.

### **Qui est exclu des évaluations ?**

L'impératif d'inclusion n'est pas toujours respecté. Dans les faits, bon nombre de groupes vulnérables et marginalisés (élèves ou établissements) sont exclus, pour diverses raisons, dès la conception de l'évaluation et la constitution de l'échantillon. Pour commencer, les enfants non scolarisés (ou inscrits dans des établissements non enregistrés) ne sont généralement pas pris en compte dans l'échantillon, car les LSLA ont pour objet le système éducatif formel et n'évaluent que des enfants inscrits dans des écoles homologuées et officiellement reconnues<sup>34</sup>.

Des établissements entiers sont parfois exclus parce qu'ils sont situés dans des lieux reculés et difficiles d'accès ou parce qu'ils se consacrent à l'éducation de groupes d'enfants ne pouvant participer aux tests (par exemple les écoles pour aveugles). L'exclusion frappe aussi les élèves individuellement. C'est le cas bien souvent des enfants atteints d'un handicap physique permanent plus ou moins grave, de ceux qui présentent des troubles cognitifs, comportementaux ou affectifs, ainsi que de ceux qui ne maîtrisent pas suffisamment la langue d'enseignement. Pour compliquer encore les choses, on connaît assez mal, faute d'informations suffisantes, les mesures prises pour lutter contre l'exclusion et prendre en compte la totalité des apprenants dans bon nombre des évaluations nationales. Bien que ces questions soient cruciales, les conseils techniques sont limités et souvent réservés aux pays à revenu élevé.

La dynamique de l'exclusion et de la sous-représentation est particulièrement manifeste dans les pays caractérisés par une forte diversité linguistique. Dans de tels contextes, la langue d'instruction pose problème alors qu'elle est souvent la clé pour rendre l'éducation plus accessible aux groupes défavorisés (UNESCO, 2016b). Il est difficile de déterminer la proportion

---

33 Voir Rowe, E. 2017. *Australie Country Case Study*. Document de travail sur la pratique des échantillonnages sélectifs en Australie rédigé pour le *Rapport mondial de suivi sur l'éducation 2017/2018*.

34 On prend de mieux en mieux conscience de la nécessité d'inclure ces groupes de population dans les LSLA. Parmi les tentatives récentes pour prendre en compte les enfants et les jeunes non scolarisés figurent les enquêtes PISA pour le développement, ainsi que les Évaluations des compétences en lecture dans les petites classes (EGRA) et certaines évaluations d'origine citoyenne (voir Yasunaga, 2014).

d'élèves à qui l'enseignement est dispensé dans la langue parlée chez eux (UNESCO, 2017d), mais selon une estimation, jusqu'à 40 % des élèves ne sont pas instruits dans une langue qu'ils parlent ou comprennent suffisamment (Walter et Benson, 2012). Les tests des LSLA, comme d'autres formes d'évaluation, sont généralement administrés dans la langue d'instruction, que les enfants la maîtrisent ou non. Cela peut avoir pour effet d'exclure par négligence les groupes marginalisés ou d'accroître leur non-participation. De ce fait, le degré d'apprentissage de bon nombre des groupes les plus marginalisés demeurent invisibles dans les données relatives à l'éducation collectées aux niveaux mondial et national (ISU, 2018a). Cette absence de données occulte peut-être les difficultés auxquelles certains pays font face en matière d'équité et le déni de leur droit à l'éducation dont sont victimes des groupes vulnérables. Bien sûr, les évaluations peuvent aussi attirer l'attention sur de tels problèmes. Quand une forte proportion d'enfants ne sont pas instruits dans leur langue maternelle, les faibles taux de réussite font mieux prendre conscience aux responsables gouvernementaux et aux professionnels du développement qu'il faudrait à tout le moins que l'initiation à la lecture et à l'écriture se fasse dans la langue de l'élève (Benson et Wong, 2015). L'interprétation des résultats demeure donc un élément déterminant de l'impact final des LSLA – et de leur capacité de contribuer au succès des efforts en faveur de l'équité.

### **Les risques liés à une trop grande accessibilité des tests**

Des tentatives ont été récemment faites pour étendre certaines évaluations à des groupes marginalisés. Dans le cadre de l'enquête ASER Pakistan, les tests ont été soumis à des enfants affectés de troubles modérés à sévères de la vision, de l'audition, de la motricité, de l'affect, de la compréhension ou de la mémoire (Singal et Sabates, 2016). De même, RTI international a adapté les outils EGRA pour les rendre accessibles aux enfants malvoyants et malentendants, et améliorer leur vécu en salle de classe. L'enquête PISA 2015 et les suivantes ont cherché à élargir l'accès aux tests. Comme l'illustre toutefois, l'encadré 5, accroître l'équité et l'inclusion est une bonne intention qui peut être elle-même source de difficultés.

#### **Encadré 5. L'aménagement des tests – une arme à double tranchant ?**

En 2011, le conseil d'administration du programme PISA a marqué son intérêt pour un élargissement de l'accès des tests aux élèves intellectuellement déficients, fonctionnellement (physiquement) handicapés ou ne maîtrisant pas suffisamment la langue des tests. Jusqu'alors, ces aménagements ne concernaient que les tests en petits groupes ou individualisés. D'autres formes d'aménagements (par exemple modifications graphiques, présentation audio, dictionnaires bilingues, possibilité de dicter ses réponses, systèmes de traitement de texte, mobilier ou outils adaptatifs, etc.) n'ont pas été largement autorisés dans le PISA. Sans ces sortes d'aménagements, des élèves qui auraient pu, autrement, y participer pleinement doivent être exclus de l'évaluation. Or, on constate que l'aménagement le plus fréquemment demandé par les élèves ayant des besoins éducatifs spéciaux est qu'il leur soit accordé un temps plus long ou illimité, chose qui n'est pas autorisée pour les élèves passant les épreuves standardisées contenues dans les brochures du PISA.

Toutefois, prévoir des aménagements afin d'inclure les groupes marginalisés ou vulnérables dans le processus de constitution des échantillons ne va pas sans risques. Tout d'abord, il est à craindre que les besoins des élèves varient à tel point qu'il devient difficile d'assurer la comparabilité au sein d'un même pays et entre pays. Deuxièmement, certains aménagements pourraient sembler placer certains élèves dans une position relativement avantagés par rapport à ceux qui n'en bénéficient pas. Enfin, cela pourrait avoir pour effet pervers d'inciter à choisir des élèves ayant des besoins spéciaux ne serait-ce que pour améliorer les scores.

Source : Basé sur OCDE, 2011.

Les instruments d'évaluation sont de plus en plus souvent traduits et adaptés à des contextes multilingues. De même, toutes les initiatives transnationales donnent lieu à des directives détaillées concernant la traduction et l'adaptation des tests, et tentent, dans toute la mesure du possible, de faire en sorte que les résultats soient comparables dans l'ensemble des pays et des langues. Dans les pays en développement, très peu de LSLA sont réalisées dans la plupart ou la totalité des langues d'instruction, mais des efforts louables sont néanmoins faits pour accroître le nombre de langues utilisées. Les tests de l'étude PASEC 2019 seront par exemple proposés aussi dans les langues maternelles des élèves des petites classes alors qu'ils ne l'avaient été que dans la langue d'instruction principale lors de l'édition 2013-2014. De même, l'Afrique du Sud a organisé son évaluation nationale annuelle dans 11 autres langues maternelles avant d'y mettre fin en 2014 (ISU, 2019a).

### **L'effet de plancher**

On appelle «effet de plancher» une autre limite liée à la conception des enquêtes qui peut affaiblir la capacité des évaluations de faire avancer le mouvement en faveur de l'équité. L'effet de plancher se produit lorsqu'une majorité des participants à une évaluation donnée obtient des scores proches de la valeur la plus basse. Comme l'ont fait valoir Wagner *et al.* (2018), attacher une importance excessive à la comparabilité au détriment de la validité locale (comme c'est le cas des LSLA) risque d'aller à l'encontre d'une bonne compréhension de la situation des populations défavorisées – lesquelles vont très probablement figurer dans la partie basse de l'échelle des notes. Dans le même ordre d'idées, Winthrop et Anderson (2013) font observer que, même si les évaluations présentant un fort effet de plancher peuvent attirer l'attention sur certaines questions pertinentes, leur potentiel analytique risque d'en être affecté. Bref, une validité locale limitée a un impact négatif particulier sur la capacité des LSLA de soutenir la campagne pour l'équité.

Attacher une importance excessive à la comparabilité au détriment de la validité locale risque d'aller à l'encontre d'une bonne compréhension de la situation des populations défavorisées.

### **Les limites dans la conception se traduisent-elles par des limites dans l'utilisation ?**

Globalement, les tendances décrites dans ce chapitre témoignent de la nature nécessairement partielle et imparfaite des données produites par les LSLA. Il ne faut pas y voir une déficience ou une limitation propres aux tests standardisés, car toute forme de collecte de données privilégiera probablement certaines compétences plutôt que d'autres et s'appuiera inévitablement sur une définition sélective de la qualité de l'éducation. Néanmoins, ces tendances poussent à mettre en garde contre les risques qu'il y a à accorder un crédit ou une attention peu équilibrés et excessifs aux données issues des LSLA. Si les résultats de ces évaluations ne sont pas convenablement combinés à d'autres sources de données (qualitatives ou quantitatives), les limites intrinsèques des LSLA ouvrent en effet la voie aux interprétations erronées. Par conséquent, puisque les données issues des évaluations sont par définition partielles et imparfaites, une attention trop exclusive aux LSLA aussi bien que leur non-intégration dans d'autres composantes du système, risquent d'altérer leur capacité de contribuer à la quête pour plus de qualité et d'équité dans l'éducation.

## 4. Au-delà des données considérées comme une fin en soi : les limites de l'utilisation des évaluations à grande échelle des apprentissages

Nous avons vu dans le précédent chapitre que certaines des conséquences involontaires des évaluations à grande échelle des apprentissages (LSLA) tiennent à leur conception même. Toutefois, la plupart des limites de ces outils découlent directement de l'utilisation – ou de l'absence d'utilisation – des données qu'elles produisent. Le présent chapitre commence par apporter quelques clarifications d'ordre conceptuel en ce qui concerne les finalités des LSLA, avant d'examiner les trois grands types de problèmes que soulève l'utilisation de ces données, à savoir leur sous-exploitation, leur surexploitation, et la combinaison de ces données avec des cadres de réédition des comptes (ou leur subordination à ces cadres).

### Les utilisations des évaluations à grande échelle des apprentissages : termes et concepts

Toute analyse des conséquences indésirables des LSLA nécessite un examen préalable des finalités de ces outils et des utilisations que l'on compte en faire. Nous allons donc, dans cette section, essayer de distinguer clairement les différents types d'informations générés par les LSLA et, sur cette base, passer en revue les multiples utilisations et finalités des LSLA, et les différentes parties prenantes auxquelles elles peuvent être utiles.

#### « Résultats » et « éléments factuels »

En matière d'utilisation des données issues des LSLA, un « résultat » est une constatation ressortant d'une analyse simple. Par exemple, les scores moyens des garçons et des filles dans une matière ou un domaine d'apprentissage donné ; ou la distribution d'un groupe donné entre différents niveaux de maîtrise. Les « éléments factuels », en revanche, sont un ensemble de constatations utilisé pour apporter une réponse éclairée par les faits à une question relative à une politique (ou à une pratique), qui consiste généralement à décrire des tendances, à dégager un tableau caractéristique, à indiquer les facteurs clés expliquant certaines régularités et à mettre en évidence les domaines d'intervention possibles. La politique ou la pratique qui sera mise en œuvre doit être fondée sur des éléments factuels ou des résultats provenant de plusieurs analyses ou glanés auprès de multiples sources de données et, par-dessus tout, comporter une part de jugement ou de réflexion (Brink, 2018).

Si, par exemple, la politique vise à assurer l'équité entre garçons et filles de façon à réduire l'écart de performance entre eux, on pourra recueillir des éléments factuels en analysant la façon dont l'écart s'est creusé ou réduit au fil du temps et/ou l'éventuelle persistance de la tendance. Si l'on combine cette analyse avec un examen des politiques mises en place dans le passé pour résoudre le problème, ainsi que des recherches menées sur les facteurs clés qui déterminent la performance en fonction des politiques, on pourra utiliser les informations obtenues pour infléchir la conception et la formulation des politiques (Brink, 2018). Les résultats ne parlent donc pas d'eux-mêmes et les conseils fondés sur les éléments factuels ne découlent pas *spontanément* des données brutes.

## Analyse et utilisation

Les données livrées par les LSLA – en un instant donné ou au fil de séries d'évaluations – apportent aux gouvernements des éléments factuels qui, en fournissant des réponses à des questions cruciales, telles que de savoir qui apprend quoi et qui n'y parvient pas, où, quand et pourquoi, leur permettent de corriger les déficiences du système (Montoya, 2016). Les experts et les chercheurs<sup>35</sup> utilisent généralement les scores mesurant les acquis et les informations recueillies au moyen des questionnaires de contexte pour décrire les connaissances et les compétences de la population cible. Cela nécessite en général plusieurs types d'analyse. Il faut tout d'abord comprendre les facteurs qui influent sur les résultats de l'apprentissage (par exemple, l'environnement et les pratiques constatées au domicile de l'élève et à l'école), et déterminer s'ils évoluent dans le temps. Deuxièmement, il faut dégager les tendances générales en matière d'acquis et évaluer les progrès par rapport à des cibles spécifiques à l'aide d'une série d'indicateurs. Enfin, il faut mettre en lumière les disparités sur le plan des capacités cognitives entre sous-populations d'apprenants, en fonction des paramètres pertinents, notamment la situation socioéconomique, la région de résidence, le sexe, le statut en matière de migration et la langue maternelle.

Les décideurs utilisent ensuite les résultats ou éléments factuels issus des LSLA à des fins variées<sup>36</sup> :

**Suivi et évaluation** – suivre l'évolution d'une série de questions et de tendances par rapport aux objectifs et cibles nationaux, régionaux ou mondiaux, et évaluer les politiques de l'éducation.

**Aide à la formulation de politiques** – orienter la formulation, la conception et l'application des politiques, ou améliorer l'efficacité des mesures déjà en place (voir **tableau 5**).

**Définition des priorités** – sensibiliser à certains enjeux, encourager le débat au sein de la société civile et dans les milieux gouvernementaux, identifier les problèmes prioritaires et exercer des « pressions positives » pour inciter les pouvoirs publics à entreprendre des réformes

---

35 Les analystes appartiennent en général à des organismes gouvernementaux, à des instituts d'évaluation publics, à des départements et services de recherche d'université, à des organismes privés spécialisés, à des groupements réalisant des évaluations transnationales, ou à des organisations internationales. Leur affiliation précise varie toutefois selon les traditions administratives, éducatives et culturelles.

36 Les utilisations et objectifs des LSLA ont été étudiés par un large éventail de chercheurs. Pour une analyse des activités énumérées ci-dessus, voir Tobin *et al.* (2015), Lingard et Grek (2007), Goldstein et Thomas (2008), Fullan (2010) et Kellaghan *et al.* (2011).

ou à poursuivre celles qui sont en cours – en particulier lorsque les résultats de l'évaluation sont surprenants ou préoccupants.

**Contrôle** – vérifier les performances des acteurs immédiats de l'éducation (administrateurs, chefs d'établissement, enseignants, conseils scolaires, cours privés) ainsi que des parties pouvant être considérées comme extérieures au système éducatif (organismes publics, collectivités locales, etc.).

**Analyse approfondie** – approfondir et mettre en perspective l'analyse des systèmes éducatifs, et inviter les responsables de l'élaboration des politiques et les praticiens à mener de nouvelles études ciblées qui puissent mieux rendre compte de la situation et des réalités locales.

Si les utilisations potentielles des données issues des LSLA par les responsables de l'élaboration des politiques, les experts et les chercheurs sont plus ou moins connues, d'autres groupes d'acteurs peuvent aussi être intéressés par ces données et par les conclusions des évaluations. C'est le cas notamment des enseignants et des directeurs d'école, des parents et des élèves, de groupes d'intérêt tels que des ONG, des concepteurs de programmes d'enseignement, des formateurs d'enseignants ou des syndicats d'enseignants, et des chercheurs et universitaires réalisant des analyses plus poussées. Le grand public peut aussi consulter avec profit les résultats des évaluations d'apprentissages dès lors qu'ils ont été adaptés aux profanes, sous la forme par exemple de graphiques simples diffusés dans les médias sociaux (pour un tour d'horizon des différents utilisateurs potentiels et groupes cibles, voir ISU 2018c).

Chacun de ces groupes peut de fait exploiter et mettre à profit les LSLA de diverses façons. Les enseignants, par exemple, pourront s'appuyer sur les résultats des évaluations pour adapter leurs méthodes pédagogiques et encadrer convenablement leurs élèves, tandis que les parents et le grand public s'en serviront pour demander des comptes aux écoles ou aux administrations (ISU, 2018c). Nos connaissances sont toutefois limitées s'agissant de savoir comment et dans quelles circonstances les différentes parties prenantes de l'éducation peuvent faire l'usage le plus efficace et le plus approprié des données disponibles, et quel est l'impact final de leur action et les risques et possibles difficultés qui pourraient en résulter.

**Tableau 5. Exemples d'initiatives politiques que les évaluations à grande échelle des apprentissages peuvent guider, par niveau d'intervention**

SYSTÈME ÉDUCATIF	PROGRAMMES D'ENSEIGNEMENT	ÉTABLISSEMENTS SCOLAIRES	ENSEIGNANTS	FAMILLES
<ul style="list-style-type: none"> <li>• Allocation de ressources</li> <li>• Réformes d'ordre programmatique</li> <li>• Définition des objectifs assignés aux élèves</li> </ul>	<ul style="list-style-type: none"> <li>• Élaboration des programmes</li> <li>• Contenus et méthodes d'enseignement</li> <li>• Conception des programmes</li> </ul>	<ul style="list-style-type: none"> <li>• Définition des priorités du corps enseignant</li> <li>• Amélioration des services d'accompagnement des élèves</li> <li>• Enrichissement de l'environnement scolaire</li> </ul>	<ul style="list-style-type: none"> <li>• Fourniture des ressources nécessaires au perfectionnement ou à l'amélioration des pratiques pédagogiques</li> <li>• Révision des cours et des attributions</li> </ul>	<ul style="list-style-type: none"> <li>• Conception des programmes visant à encourager la participation des parents</li> </ul>

Source : Les auteurs.



## Sous-exploitation des données d'évaluation

Les éléments factuels mis en évidence par les LSLA sont censés être des outils essentiels pour l'amélioration des systèmes éducatifs et la réussite des élèves. Toutefois, ces effets ne peuvent se produire que si les acteurs clés agissent sur la base des données disponibles, en analysant ces données, en communiquant les résultats aux groupes cibles et en s'en servant pour guider l'adoption de mesures concrètes. Nous allons à présent examiner des facteurs moins communément reconnus qui font souvent obstacle à une utilisation efficace et suffisante des données – en particulier, l'appropriation limitée du financement et de la conception de la collecte des données, de la gestion et de la diffusion.

### La collecte de données n'améliore pas la qualité

On l'a vu, les LSLA sont aujourd'hui des instruments essentiels qui livrent de précieuses indications sur de multiples dimensions des systèmes éducatifs. Aussi est-il nécessaire de comprendre les résultats ainsi que leur évolution et les raisons de celle-ci afin d'adopter et ajuster les politiques et stratégies qui ont pu être mises en place pour améliorer l'apprentissage. Si les données recueillies ne sont pas soigneusement analysées et diffusées auprès des acteurs concernés, pour que ces derniers puissent prendre de nouvelles mesures, le processus ne sert à rien (Kellaghan et Greaney, 2001).

Toutefois, on ne peut tenir pour acquis que les résultats seront effectivement utilisés et diffusés (Tobin *et al.*, 2015 ; ISU, 2018c). Le processus d'évaluation de l'apprentissage aboutit en général à l'établissement d'un rapport qui récapitule les conclusions de l'enquête, et dont le lancement fait parfois l'objet d'un événement spécial, mais l'examen des résultats tend ensuite à s'essouffler. Les gouvernements, les organisations régionales ou internationales et les autres utilisateurs se contentent trop souvent d'expliquer l'état et les tendances de l'éducation, ou ne se font l'écho que d'un ensemble prédéfini d'indicateurs sans chercher à savoir comment utiliser au mieux les conclusions pour améliorer le contexte et le processus qui rendent l'apprentissage possible. Autre obstacle à l'exploitation productive des résultats, les LSLA ne sont pas toujours perçues comme utiles par les enseignants et les éducateurs, de sorte qu'elles n'ont finalement que peu d'impact sur l'enseignement et l'apprentissage. Si la publication d'un rapport est certes nécessaire (car elle suppose la compréhension des résultats et des relations complexes entre l'apprentissage et les variables qui influent sur lui), le simple fait de recueillir des données n'entraîne pas automatiquement une amélioration de l'apprentissage. Lorsque l'évaluation n'est pas utilisée pour susciter un changement, la mesure des résultats de l'apprentissage risque de devenir une fin en soi.

On ne peut tenir pour acquis que les résultats seront effectivement utilisés et diffusés. Le simple fait de recueillir des données n'entraîne pas automatiquement une amélioration de l'apprentissage.

Comme il a été dit précédemment, l'élaboration et l'administration des LSLA demandent beaucoup de ressources – fonds, équipements et capacités. Le coût total peut-être lourd pour

des pays à revenu faible ou moyen-faible<sup>37</sup>. Il est donc crucial que les données obtenues soient utilisées pour rentabiliser au maximum l'investissement et pour que l'évaluation ne soit pas perçue comme un gaspillage de ressources grevant sans raison les finances publiques<sup>38</sup>.

## Reconnaître les obstacles pour permettre une exploitation efficace

Malgré les nombreuses fins que les LSLA sont censées servir, on ne sait pas encore assez bien comment et dans quelles conditions les résultats de ces évaluations peuvent être exploités de manière efficace. Certaines de ces conditions ont été abondamment étudiées et des préoccupations ont été formulées dans des travaux fondateurs sur les LSLA, notamment *Using assessment to improve the quality of education* de Kellaghan et Greaney (2001) (voir aussi ISU, 2018c, et Raudonyte, 2019, pour une étude plus récente sur les contraintes limitant l'utilisation des données des LSLA). Parmi les nombreuses contraintes fréquemment reconnues figurent les obstacles d'ordre financier et technique (par exemple le caractère limité ou dispersé des compétences dont disposent les parties concernées ou la qualité médiocre des évaluations en place) ; l'absence d'engagement ou de volonté politique ; l'intégration insuffisante dans le processus d'élaboration des politiques ; l'adéquation limitée entre les évaluations et d'autres éléments du système éducatif ; la faible fiabilité ou pertinence des informations ; le caractère trop superficiel de l'analyse ; la présentation inadéquate des données ; et leur diffusion insuffisante ou inappropriée (ISU, 2018a, 2018c ; Lockheed, 2015 ; Tobin *et al.*, 2015).

Cela étant, d'autres contraintes demeurent insuffisamment théorisées ou explorées et, chose plus grave, les analyses des nombreux obstacles limitant l'utilisation des données font relativement abstraction du contexte. D'où le risque de méconnaître la dynamique particulière qui affecte les pays à faible revenu ou tardivement convertis aux LSLA. La présente section se penche sur une catégorie particulière de contraintes propres aux pays en développement, à savoir, l'appropriation insuffisante des évaluations. L'exploitation limitée de leurs résultats ne peut jamais être exclusivement attribuée à un manque de volonté politique. Dans tous les cas, un facteur important est le caractère insuffisant des *possibilités*, du *soutien*, des *ressources* ou des *capacités* dont auraient besoin les principaux acteurs de l'éducation pour accéder aux données disponibles, les interpréter et agir en conséquence. Cela est tout

L'exploitation limitée de leurs résultats ne peut jamais être exclusivement attribuée à un manque de volonté politique. Dans tous les cas, un facteur important est le caractère insuffisant des *possibilités*, du *soutien*, des *ressources* ou des *capacités* dont auraient besoin les principaux acteurs de l'éducation pour accéder aux données disponibles, les interpréter et agir en conséquence.

37 Les coûts financiers directs comprennent le personnel, la formation et le renforcement des capacités, les voyages et les équipements, les rapports et leur diffusion (Wagner, 2011 ; Lockheed, 2015). Des frais de participation doivent en outre être acquittés pour certaines initiatives transnationales. Le montant total des coûts directs va de 400 000 à 2 millions de dollars des États-Unis (Clarke, 2016).

38 La question de savoir si les avantages des évaluations l'emportent sur leur coût a fait débat. Certains auteurs ont fait valoir que les évaluations figurent parmi les innovations les moins coûteuses en matière de réforme de l'éducation (Clarke, 2012) ou que leur coût est relativement faible par rapport à celui de la mise en œuvre d'une politique, ou de la prise de décisions mal informées ou incorrectes (ISU, 2016a). D'autres sont toutefois d'avis que toute analyse du rapport coût-avantage de tels exercices comporte une part de jugement normatif et que ces questions n'ont pas été suffisamment traitées dans le discours public (Engel et Rutkowski, 2018).

particulièrement vrai des pays les moins fortunés, dont les capacités actuelles d'intégrer les initiatives nationales et transnationales dans la gestion de leur système éducatif tendent à accuser un certain retard du fait d'une institutionnalisation limitée (ISU, 2018c).

### **Conséquences de l'appropriation insuffisante des évaluations**

Dans bien des pays, le décalage entre production et utilisation des données est aussi une conséquence directe d'une prise en main insuffisante des données issues des évaluations. Il est permis de penser que, lorsqu'un gouvernement n'a pas le sentiment d'avoir un certain pouvoir de décision sur le processus d'évaluation, il sera moins enclin à s'appuyer sur les données pour prendre des mesures concrètes. Cette absence d'appropriation s'explique elle-même par une faible participation aux différents éléments du processus d'évaluation, notamment son financement, sa conception, la gestion des données et la diffusion des résultats, et par une influence limitée.

**Évaluations financées par des sources extérieures.** On note que les évaluations tant nationales que transnationales de l'apprentissage sont généralement financées par des sources extérieures, en particulier dans les pays en développement. C'est ainsi que les deux tiers au moins des évaluations nationales sont financées (en partie ou en totalité) par des donateurs étrangers ou extérieurs (ISU, 2019a) tels que le GPE, l'UNICEF, l'USAID et la Banque mondiale<sup>39</sup>. Depuis 2014, les évaluations transnationales du PASEC sont financées pour moitié par la CONFEMEN, les gouvernements qui souhaitent y participer devant prendre à leur charge l'autre moitié. Ces financements extérieurs pourraient expliquer le caractère irrégulier et sporadique des évaluations dans bon nombre de pays. Plus préoccupante encore est la possibilité que les évaluations ainsi financées contribuent au manque de coordination entre les évaluations nationales et les autres éléments du système éducatif, qui a pour corollaire une faible appropriation des résultats au niveau national.

**Contrôle du processus d'évaluation, de conception et de gestion.** Lorsque les *évaluations* sont sous-traitées auprès d'organisations non gouvernementales ou d'organismes internationaux, la responsabilité des pays est limitée et ils n'ont que peu de contrôle sur la conception des instruments d'évaluation et la gestion des données obtenues – du fait en particulier que les représentants et experts nationaux ne contribuent guère à l'élaboration desdits instruments. C'est le cas des *évaluations* transnationales, dans lesquelles l'organisme qui pilote l'exercice établit avec un contractant spécialisé les séries de données des pays participants sous une forme standardisée, avant de les communiquer aux *équipes* nationales, qui procèdent à leurs propres analyses. Si ces arrangements facilitent la standardisation des données, l'organisme pilote peut être perçu comme « contrôlant » les données et comme en ayant la « propriété ».

**Un calendrier de diffusion strict.** Il est également courant que, dans les initiatives transnationales, les pays n'aient guère leur mot à dire sur les données ou les analyses les

---

39 Dans l'Amérique du Nord et l'Europe occidentale, aucune évaluation n'est financée par des sources extérieures. Peu de pays d'Amérique latine reçoivent une aide extérieure pour leurs évaluations (11 %) mais les pourcentages sont supérieurs dans d'autres régions : Asie et Pacifique (30 %), Europe orientale (30 %), Afrique subsaharienne (36 %) et États arabes (50 %) (Clarke, 2017).

concernant qui sont publiées, ni sur les niveaux d'apprentissage de la population cible. Ils ne sont parfois autorisés à diffuser leur rapport national qu'après divulgation de l'étude comparative internationale (OCDE, 2009 ; IEA, 2019).

Des conventions sont signées entre gouvernements et organismes pilotant une évaluation transnationale, en vertu desquelles la décision de diffuser les résultats et leur interprétation (que les pays les jugent satisfaisants ou non) appartient à l'organisme, pour éviter tout conflit d'intérêts<sup>40</sup>.

L'appropriation des données pose moins de problèmes dans les évaluations nationales, dont les pays contrôlent en général la stratégie, la conception et les analyses. Il arrive toutefois que même la conception et la mise au point des tests soit réalisées (ou fortement influencées) par les donateurs finançant l'exercice, dans le cas en particulier de pays à revenu faible ou moyen. On connaît des exemples d'évaluations nationales dont les résultats ne sont pas validés ou diffusés de manière adéquate par les autorités, ce qui en amoindrit la valeur pour la formulation des politiques (Union européenne, 2014). Quel que soit le type d'évaluation, il convient de noter qu'une appropriation insuffisante risque non seulement d'aboutir à une sous-exploitation des données et à l'inertie des pouvoirs publics, mais aussi d'inciter le gouvernement, les autorités éducatives et d'autres acteurs à faire de ces évaluations une interprétation ou une utilisation malavisée. Cela peut se traduire par une distorsion des priorités et des politiques erronées.

Une appropriation insuffisante risque non seulement d'aboutir à une sous-exploitation des données et à l'inertie des pouvoirs publics, mais aussi d'inciter le gouvernement, les autorités éducatives et d'autres acteurs à faire de ces évaluations une interprétation ou une utilisation malavisée.

## Surexploitation des données d'évaluation

Nous allons nous pencher dans cette section sur les risques liés à la surexploitation et à la surinterprétation des données d'évaluation – c'est-à-dire les cas dans lesquels il est prêté un intérêt, une attention ou un crédit excessif aux résultats d'une évaluation. Une éventuelle surexploitation de ces résultats peut créer une dynamique aboutissant à des effets de distorsion ou de diversion sur les décisions et les comportements d'un large éventail de parties prenantes et d'utilisateurs des LSLA – tout particulièrement les responsables de l'élaboration des politiques, les enseignants et les chercheurs. Après avoir examiné les différents formats, canaux et agents de diffusion des données, nous passerons en revue les quatre différents mécanismes par lesquels les données d'évaluation encouragent cette dynamique : inciter à prendre des mesures pour manipuler les chiffres, encourager à copier des politiques sans discernement, conduire à postuler à tort des liens de causalité, et promouvoir les systèmes de financement en fonction des résultats.

40 Certains pays se voient offrir la possibilité de supprimer certaines variables avant la diffusion ou la publication des données, mais seulement s'il est démontré que ces données « présentent un risque pour l'anonymat de l'élève, de l'école, et/ou de l'enseignant, ou des erreurs techniques qui n'auraient pu être corrigées par les fournisseurs de données » (OCDE, 2018b, p. 1).

## Des formes de diffusion multiples

Une fois les données produites par les LSLA analysées, les résultats ou conclusions de cette analyse sont récapitulés dans un ou plusieurs produits de diffusion – rapports analytiques, notes d'orientation, rapports dans les médias, communiqués de presse, bases de données d'évaluation, ou même blogs et messages dans les médias sociaux visant à toucher le public profane. Ces produits sont ensuite diffusés de diverses façons. Bien entendu, il n'est pas nécessaire de les utiliser tous.

Dans le cas des évaluations transnationales des apprentissages, la diffusion et la communication de l'information s'opère d'une manière systématique et relativement homogène. La date à laquelle les résultats seront rendus publics est fixée et les conclusions sont placées sous embargo jusque-là. Puis, de nouvelles conclusions et rapports sommatifs et thématiques paraissent périodiquement, une base de données est mise en ligne et des articles sont publiés dans des journaux nationaux et des blogs internationaux. En revanche, la diffusion des résultats des évaluations nationales a un caractère plus hétérogène. Une évaluation sur quatre seulement donnera lieu à la mise en ligne d'une base de données et une sur trois sera commentée dans les médias (ISU, 2019a).

## Le rôle des médias

Quels que soient les modes de diffusion choisis par les autorités compétentes, l'influence des LSLA passe aussi par leur impact sur l'opinion et les sentiments du public (Feuer, 2012 ; Addey, 2018 ; Yasukawa *et al.*, 2017). Cet impact dépend en grande partie de l'attention que les médias vont accorder à telle ou telle évaluation. Par « médias », il faut entendre ici la presse traditionnelle, imprimée et sur les ondes, et les contenus numériques et interactifs. Une large couverture et un fort écho dans les médias permettent aux LSLA d'influer sur la vision commune de ce qui mérite d'être promu dans l'éducation (Hamilton, 2017 ; voir des exemples de cette dynamique dans Dixon *et al.*, 2013 ; Steiner-Khamsi *et al.*, 2017 ; Waldow *et al.*, 2014). De ce fait, c'est le discours dominant tenu dans les évaluations transnationales des apprentissages les plus vastes et les plus largement diffusées, c'est-à-dire les enquêtes PISA, TIMSS et PIRLS, qui décide de ce qui sera considéré comme utile (et financé) dans un système éducatif. Si cela n'a, en soi, rien de problématique, cela risque de conforter le recours excessif aux données d'évaluation, en détournant l'attention d'autres questions pertinentes, et d'alimenter des hypothèses et des affirmations faisant état de liens de causalité erronés.

Une large couverture et un fort écho dans les médias permettent aux LSLA d'influer sur la vision commune de ce qui mérite d'être promu dans l'éducation.

Un examen plus poussé est toutefois nécessaire pour comprendre comment différents médias contribuent à la formation du discours public sur les LSLA (Hamilton, 2017), étant donné en particulier la très grande diversité des médias, qu'il s'agisse de leur public cible, de leurs partenaires, de l'exactitude de leurs comptes rendus, de leur façon de jouer sur l'émotion, etc.

## Encourager la réforme des politiques – en mettant l'accent sur les chiffres

Nous l'avons dit, les LSLA sont censées contribuer au discours relatif aux politiques de l'éducation en éclairant les responsables de l'élaboration de ces politiques sur les priorités et les grands objectifs. Toutefois, cette influence présente des risques importants si l'action politique vise exclusivement à modifier et améliorer les chiffres plutôt qu'à entreprendre des changements de fond, de caractère structurel.

Une telle dynamique risque en particulier de se manifester dans le cas d'évaluations transnationales. Cela tient en grande partie au fait que les résultats de ces exercices sont souvent diffusés sous la forme de classements. Ces derniers sont devenus dans les médias un moyen habituel de comparer la situation et les performances des services éducatifs, et d'illustrer les progrès ou les échecs dans les différents pays. La présentation des résultats, les pays les plus méritants figurant en tête du classement et les moins performants tout en bas, permet aux responsables de l'élaboration des politiques d'apprécier la performance des enfants inscrits dans leurs écoles par rapport aux résultats des enfants d'autres pays. Le classement donne donc une image du progrès national. Cette tendance à chercher à améliorer les chiffres est illustrée, par exemple, par l'ambition affichée par l'Australie de figurer parmi les cinq pays en tête du classement du PISA d'ici 2025, la volonté de la Thaïlande d'atteindre d'ici 2021 le niveau de performance moyen établi par l'OCDE dans le cadre du PISA (Breakspear, 2014) ou l'objectif des États du Golfe d'améliorer leur rang dans le classement du PISA (UNESCO, 2017f).

En mettant en relief les champions et ceux qui font moins bien, les données comparatives de ce type peuvent intensifier les pressions incitant les gouvernements à adopter des normes plus élevées dans les diverses dimensions auxquelles sont associés les résultats de l'apprentissage – même parmi les pays les moins performants. Les comparaisons entre pays peuvent donc encourager des initiatives politiques visant à améliorer les acquis des élèves. Les exemples empiriques d'une telle dynamique sont multiples. C'est ainsi que la publication des résultats de l'enquête TIMSS (1999) a poussé les États-Unis et le Royaume-Uni à entreprendre des réformes (Gorard, 2001), et que les résultats d'une enquête PISA (2000) ont fait apparaître de grandes disparités en Allemagne, suscitant un intense débat public sur la nécessité d'une réforme.

Si ces tendances ne constituent pas, en soi, un problème, les analyses tendent fâcheusement à accorder une importance excessive au classement et aux scores moyens, plutôt qu'aux facteurs qui expliquent les différences dans la performance des pays (UNESCO, 2017a ; Rutkowski et Rutkowski, 2016 ; Torney-Purta et Amadeo, 2013). Cet accent mis sur le classement et les scores peut avoir un effet de diversion ou induire en erreur. C'est en particulier le cas des réformes dictées par des tentatives (malavisées) de reproduire des interventions sur la base d'une compréhension très superficielle et décontextualisée de leur impact et de leur capacité de faire progresser les scores.

Les palmarès encouragent les gouvernements à définir des cibles en fonction des résultats mesurables atteints par d'autres systèmes éducatifs, ce qui risque de les pousser à prendre modèle sur les politiques et pratiques appliquées ailleurs ou à en tirer les leçons.

## **Incitation à reproduire uniformément certaines politiques**

On observe depuis quelques années qu'un nombre croissant de pays mettent en œuvre des réformes de l'éducation ou des initiatives politiques au vu des évaluations (Fischman et Topper, 2017 ; Baird *et al.*, 2016)<sup>41</sup>. Les palmarès encouragent les gouvernements à définir des cibles en fonction des résultats mesurables atteints par d'autres systèmes éducatifs, ce qui risque de les pousser à prendre modèle sur les politiques et pratiques appliquées ailleurs ou à en tirer les leçons (OCDE, 2016a). En ce sens, les classements et palmarès sont devenus d'importants outils politiques pour accélérer le changement et l'innovation (Steiner-Khamsi, 2003). Ainsi, les LSLA pourraient non seulement stimuler des réformes, mais aussi encourager des pays à copier les pratiques et les politiques des mieux classés ou des « champions », comme la Finlande ou la ville de Shanghai, ce qui se traduit par une sorte de « pèlerinage éducatif » dans lequel les lieux les plus performants deviennent des destinations privilégiées pour les responsables de l'élaboration des politiques de l'éducation, les chercheurs en pédagogie et autres acteurs du système éducatif (Sahlberg, 2007 ; Addey, 2014).

C'est pourquoi, même si les évaluations sont un moteur essentiel du développement de l'éducation, des craintes ont été émises concernant le risque que les systèmes éducatifs et leurs partenaires fondent par trop leurs réformes sur les tests. Des chercheurs ont décrit les erreurs auxquelles cette dynamique peut conduire les décideurs. À terme, elle peut inciter les responsables à emprunter telle ou telle politique à des pays obtenant des scores élevés sans guère se soucier du contexte culturel ou de la cohérence globale de leur action (Breakspear, 2014), ou encore à adopter des solutions miracles et politiquement séduisantes (Lingard et Sellar, 2013) plutôt que de s'engager dans une réflexion rigoureuse et à long terme, prenant mûrement en considération les visées du système scolaire. Gorur et Wu (2014) ont montré par exemple comment l'ambition de l'Australie de figurer parmi les cinq pays arrivant en tête du classement du PISA l'a poussée à prendre exemple sur les « bons élèves » de l'Asie de l'Est, alors même que la connaissance des spécificités de ces systèmes est limitée et ne va guère plus loin que leurs scores moyens et leur place dans le palmarès.

Autre facteur aggravant, le recours limité et irrégulier aux « *big data* » (mégadonnées) relatives à l'éducation pour éclairer la formulation des politiques, dû en partie à la capacité institutionnelle insuffisante de certains pays s'agissant d'analyser ces données et d'adapter les politiques en fonction des résultats. Ces pays risquent de ce fait de s'appuyer sur les conclusions de recherches menées dans d'autres pays et régions, qui ne sont pas nécessairement pertinentes pour leur propre système éducatif (UNESCO, 2017e). À cet égard, il est réconfortant de constater que certains organismes, comme la CONFEMEN, ont récemment commencé à présenter les scores des pays dans l'ordre alphabétique de ces derniers, et non pas selon leur performance – une petite innovation qui pourrait contribuer à neutraliser cette tendance. Voir à la **Figure 2** un exemple d'une telle présentation alphabétique des conclusions concernant les niveaux de compétences.

---

41 Il apparaît aussi de plus en plus clairement que les gouvernements utilisent leur participation aux initiatives pour légitimer ou accélérer des réformes, y compris des programmes préétablis et répondant à des visées idéologiques qu'il serait autrement difficile de justifier d'un point de vue purement technique (voir Raudonyte, 2019).

## Recherche et établissement de liens de causalité

L'intérêt de la communauté des chercheurs pour les LSLA est allé croissant. En témoigne le nombre grandissant de commentaires subjectifs et d'articles savants vérifiés par les pairs qui leur ont été consacrés. De caractères variés, ces publications traitent des incidences des LSLA et de l'utilisation des données aux fins de l'amélioration des politiques nationales ou présentent des analyses plus poussées fondées sur des recherches empiriques (Fischman et Topper, 2017).

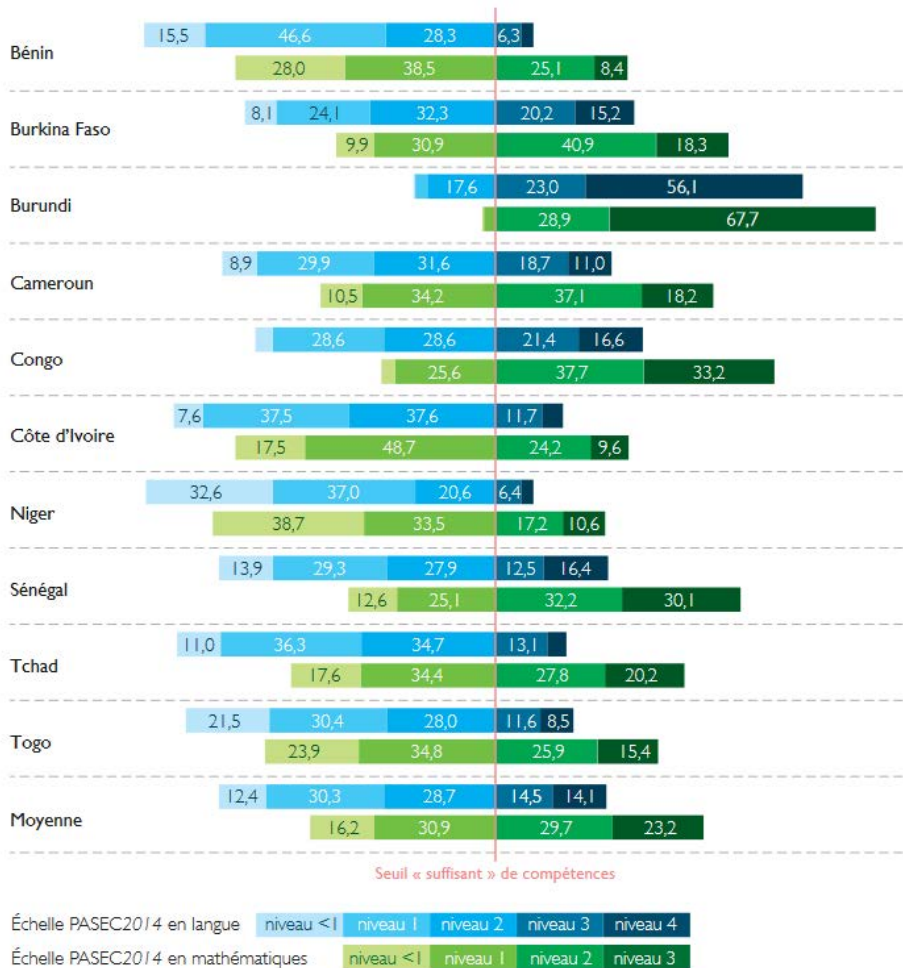
Un sujet de recherche qui a pris récemment une plus grande place dans le discours international sur les politiques de l'éducation est l'impact sur la croissance économique de la qualité de l'éducation, telle que mesurée par les scores moyens obtenus dans les évaluations transnationales. Il est au centre de débats animés entre universitaires et organisations internationales. Hanushek et Woessmann (2015) ont fait valoir par exemple que l'amélioration des compétences d'une population, mesurée par la progression des scores entre deux éditions d'une évaluation transnationale, contribue à une croissance sensible du PIB. Ce message a de quoi séduire et pourrait inciter à croire que des scores plus élevés se traduiraient par des gains sur le plan économique. D'autres analystes ont noté toutefois qu'un tel message peut être qualifié de simpliste pour deux raisons. Premièrement, l'apprentissage ne s'arrête pas à la fin de la scolarité : c'est un processus qui se poursuit en tous lieux et tout au long de la vie (OCDE, 2010). Deuxièmement, les performances économiques d'un pays dépendent de très nombreuses variables sans rapport avec le niveau d'instruction et la réussite scolaire (Feuer, 2012). Ainsi, l'affirmation d'une relation de cause à effet entre réussite scolaire et performance économique peut être trompeuse. Des chercheurs tels que Komatsu et Rappleye (2017) ont démontré que la corrélation entre ces deux paramètres était faible, ce qui semble exclure plus ou moins tout lien de causalité. De même, Feuer (2012) a montré dans le passé qu'il n'y avait pas de données empiriques suffisantes pour conforter la thèse d'une corrélation entre résultats éducatifs et résultats économiques. Ces auteurs notent aussi que les scores obtenus dans les évaluations ne sont pas des données suffisamment fiables pour prédire le bien-être économique global d'une nation.

L'un des principaux obstacles auxquels se heurtent les recherches sur la valeur prédictive des LSLA tient à la difficulté de tenir compte d'un certain nombre de variables de confusion. Par exemple, les résultats des LSLA peuvent être considérés comme un bon critère pour prédire les taux d'inscription et de réussite dans l'enseignement supérieur – comme l'indiquent des enquêtes longitudinales menées dans différents pays, la performance aux épreuves du PISA est un facteur prédictif solide et constant des études ultérieures (OCDE, 2010). Toutefois, il se peut aussi que de meilleures performances économiques se traduisent par une participation et une réussite accrues à de telles études. Il est également possible qu'un fort taux de niveau d'instruction supérieure se traduise par de meilleures performances économiques (Earle, 2010).

En résumé, il existe de bonnes raisons de n'invoquer des liens de causalité qu'avec plus de prudence. Comme l'ont souligné Komatsu et Rappleye (2017), il conviendrait en outre de faire preuve de plus de rigueur dans les débats mondiaux sur les politiques de l'éducation et de



**Figure 2. Pourcentage d'élèves selon le niveau de compétences atteint en langue et mathématiques – Début de scolarité**



Source : PASEC, 2015.

se montrer plus attentif à la manière dont les statistiques et autres données sont utilisées à l'appui de la formulation des politiques. De même, l'importance relative des données issues des évaluations par rapport à d'autres sources d'information est un autre domaine où une réflexion plus poussée permettrait d'asseoir les recherches sur un ensemble plus diversifié d'observations factuelles.

## Financements sur la base des résultats

Les évaluations des apprentissages peuvent aussi être surexploitées au sein de la communauté des acteurs du développement. C'est le cas en particulier lorsque le montant des fonds reçus au niveau national de donateurs extérieurs dépend de la qualité et des résultats des LSLA. C'est ainsi que le plan stratégique pour 2016-2020 du Partenariat mondial pour l'éducation (GPE), qui vise à améliorer l'apprentissage et l'équité en renforçant les systèmes éducatifs, encourage le développement des systèmes d'évaluation nationaux. Le GPE demeure en même temps centré sur les données et les résultats et utilise un cadre axé sur les résultats pour mesurer les progrès accomplis et vérifier les performances de tous les membres du partenariat (GPE, 2016). Il peut donc geler les fonds lorsqu'il n'est pas procédé à des évaluations ou lorsque les acquis des élèves semblent ne pas s'améliorer. Si de tels programmes offrent de réelles possibilités d'encourager des réformes des politiques, ils présentent aussi des risques importants. En particulier, les mécanismes de financement fondés sur les résultats peuvent inciter à « manipuler le système » (UNESCO, 2017d). Ces pratiques visant à s'assurer des aides financières consistent notamment à procéder à un échantillonnage sélectif, à composer la série d'épreuves sur mesure ou à choisir les élèves les plus performants.

Les mécanismes de financement fondés sur les résultats peuvent inciter à « manipuler le système », [par exemple] à procéder à un échantillonnage sélectif, à composer la série d'épreuves sur mesure ou à choisir les élèves les plus performants.

## Couplage des évaluations avec des cadres de réédition des comptes

Un certain nombre de risques généralement associés aux LSLA sont en fait imputables à leur combinaison avec des cadres de réédition des comptes présentant des enjeux élevés (ou leur subordination à ces cadres). La présente section examine la nature des dispositifs combinant les deux outils ainsi que les principaux risques et obstacles qu'ils comportent. Elle se conclut sur une mise en garde concernant l'importance relative des cadres de responsabilité à forts enjeux au niveau mondial, suivie d'une analyse insistant sur la nécessité de chercher à mieux comprendre d'autres formes de LSLA – tout particulièrement les programmes reposant sur des échantillons et ne comportant pas d'enjeux élevés.

## Les LSLA comme outil de contrôle des politiques

Les LSLA sont généralement considérées comme n'ayant pas de grandes conséquences pour les élèves qui passent les tests. Elles diffèrent à cet égard des examens nationaux qui sont utilisés à des fins de sélection, de répartition dans des classes homogènes ou de certification en fin de cycle d'études (ISU, 2019b). Mais si ces évaluations sont sans grandes conséquences pour les élèves, il en va tout autrement pour d'autres parties prenantes, notamment les enseignants, les chefs d'établissement, les écoles, ou les autorités responsables de l'éducation.

Tel est le cas des cadres de réédition des comptes visant à contrôler ou vérifier par des tests les compétences dites « managériales ». Cette forme de contrôle particulière se caractérise par l'exploitation de données issues de tests standardisés et une insistance claire sur les résultats,

généralement assimilés aux acquis de l'apprentissage. Dans les cadres de réédition des comptes fondés sur les tests, certaines parties prenantes, en particulier les enseignants, les chefs d'établissement et les écoles, sont tenus responsables des performances des élèves par rapport à un ensemble prédéfini de normes et aux scores attribués à l'issue des LSLA (Verger et Parcerisa, 2017 ; Anderson, 2005 ; Hamilton *et al.*, 2002). Elles peuvent être récompensées quand les résultats sont satisfaisants, mais aussi pénalisées ou sanctionnées dans le cas contraire. Les conséquences varient considérablement. Le fait de ne pas avoir atteint les normes minimales acceptables peut entraîner la fermeture d'une école, la nomination d'une nouvelle direction et le licenciement ou la mutation des enseignants et du chef d'établissement. Les bonnes performances sont en revanche encouragées par un certain nombre d'incitations matérielles ou organisationnelles, telles que des primes, des possibilités de promotion, ou une autonomie accrue accordée à l'établissement.

### **Des incitations perverses lorsque les scores sont au centre du dispositif**

Les incidences et effets non voulus des cadres de réédition des comptes couplés à l'évaluation de l'apprentissage, notamment lorsque les résultats des évaluations ont des conséquences claires et directes pour les enseignants, les chefs d'établissement et les écoles, ont été abondamment analysés dans la littérature spécialisée.

Comme indiqué plus haut, en cas de sanctions et de récompenses dûment annoncées, les scores peuvent devenir l'objectif central de l'enseignement scolaire, au point que les heures de cours sont consacrées principalement à la préparation aux tests, et non plus à l'étude de contenus standardisés (Weiss *et al.*, 2002). Cela peut avoir pour effet pervers d'inciter les écoles et les enseignants à adopter des pratiques discutables afin d'obtenir des scores maximaux. Les écoles vont par exemple sélectionner un échantillon de « bons élèves » pour biaiser les résultats aux épreuves, restreindre l'éventail des questions traitées dans les programmes, encourager un enseignement axé sur les tests, ne s'intéresser qu'aux élèves ayant le plus de chances de réussir les épreuves, voire encourager les élèves à tricher purement et simplement (UNESCO, 2017d).

En cas de sanctions et de récompenses dûment annoncées, les scores peuvent devenir l'objectif central de l'enseignement scolaire, au point que les heures de cours sont consacrées principalement à la préparation aux tests, et non plus à l'étude de contenus standardisés. Cela peut avoir pour effet pervers d'inciter les écoles et les enseignants à adopter des pratiques discutables afin d'obtenir des scores maximaux.

Ces pratiques font problème pour un certain nombre de raisons. Tout d'abord, elles constituent une forme de manipulation du système qui peut tromper les efforts de suivi et produire aussi des scores artificiellement élevés qui ne reflètent pas les acquis réels des élèves (Koretz, 2017). Deuxièmement, elles risquent de compromettre à terme la qualité et l'équité en matière d'éducation. De fait, même leurs effets sur les résultats de l'apprentissage demeurent sujets à débat. Les conclusions de l'édition 2017/2018 du *Rapport mondial de suivi sur l'éducation* montrent par exemple que les résultats de l'apprentissage, mesurés au regard des scores obtenus au PISA entre 2003 et 2015, ne se sont pas améliorés dans les pays appliquant

un système de contrôle redditionnel fondé sur les tests. Enfin, elles peuvent aboutir à un appauvrissement de fait ou informel des programmes d'enseignement en raison de l'attention moindre qui est portée aux connaissances, aux savoir-faire, aux compétences ainsi qu'aux attitudes et aux valeurs ne faisant pas l'objet de tests<sup>42</sup>.

### **Le cas de la rémunération au mérite**

Les effets du système de sanctions et de récompenses sont particulièrement visibles dans les cas où la rémunération est calculée en fonction des performances – une option qui gagne effectivement du terrain depuis une dizaine d'années. Dans bon nombre de pays, les scores des élèves sont de plus en plus souvent pris en compte dans la notation des enseignants (UNESCO, 2017d) et influent par conséquent sur leur rémunération, leurs promotions et leurs conditions d'emploi (Larsen, 2005). Différents programmes de rémunération au mérite ont été, par exemple, expérimentés aux États-Unis, au Portugal, au Chili et en Angleterre (UNESCO, 2014).

Si cela incite sans doute les enseignants à faire tout leur possible pour améliorer les résultats de leurs élèves, les sanctions auxquelles ils sont exposés en fonction des scores ou des évaluations peuvent avoir de multiples conséquences négatives sur les niveaux d'instruction, l'apprentissage et l'équité (UNESCO, 2017d). Selon l'édition 2017/2018 du *Rapport mondial de suivi sur l'éducation*, la rémunération au vu des performances peut créer un climat de compétition malsain, démotiver les enseignants et encourager un enseignement conçu en fonction des tests (UNESCO, 2017d). Il convient aussi de noter que la qualité de l'enseignement ne devrait pas être définie par les seuls résultats aux tests, mais aussi par les valeurs et principes défendus par les enseignants, par leur professionnalisme, l'utilisation qu'ils font de l'équipement de la salle de classe et des autres matériels disponibles, ou par leur capacité de créer un environnement propice à l'apprentissage et d'adapter leur pédagogie aux besoins de chaque élève. Par conséquent, lier la rémunération des enseignants aux performances des élèves peut conduire à restreindre considérablement les qualités que l'on attend d'un « bon enseignant ». Mécontents et découragés, des enseignants risquent alors de décider de quitter leur emploi, et il deviendra plus difficile de recruter et garder des éléments de talent. Cet enchaînement a été décrit aux États-Unis, où le système de responsabilité des résultats aux tests est particulièrement bien établi. Comme le montrent Ryan *et al.* (2017), les évaluations lourdes de conséquences ont un effet négatif sur la fidélisation des enseignants, victimes de stress et d'épuisement professionnel.

### **Un phénomène mondial en plein essor ?**

L'utilisation des LSLA comme une sorte de cadre redditionnel s'observe dans divers pays du globe et constitue une tendance répandue et solidement établie. Certains des exemples les mieux connus et les mieux étudiés en sont l'Australie, où les LSLA servent à comparer publiquement les *écoles* et à récompenser certains enseignants (Hardy et Boyle, 2011), les *États-Unis*, où les résultats aux tests sont invoqués pour fermer des *écoles* dont les *élèves* affichent des résultats inférieurs aux normes de compétence fixés par l'État (Darling-Hammond, 2004),

42 Les preuves avancées sur la question demeurent fortement controversées. Certains auteurs se fondent sur une analyse du contenu des manuels pour affirmer que rien n'indique un quelconque appauvrissement des programmes d'enseignement (Ramirez *et al.*, 2018). D'autres ont fait observer que les effets de cette nouvelle culture des tests peuvent se faire sentir de manière moins manifeste, notamment dans la répartition des heures de cours entre les différents domaines de compétences (Smith, 2019).

le Chili, où le financement des *écoles* dépend des scores obtenus lors des *évaluation des* apprentissage (UNESCO, 2017d) et, plus récemment, la Tanzanie, où un projet pilote a pour objet la rémunération des enseignants et le financement des écoles sur la base des résultats aux tests (Mbiti *et al.*, 2018).

Les données disponibles suggèrent toutefois que, dans les pays en développement, la plupart des LSLA portent sur des échantillons (ISU, 2019a) et sont donc peu adaptées à la mise en œuvre de cadres de réédition des comptes présentant des enjeux élevés. De plus, ces dispositifs ne semblent pas se propager de manière linéaire ou irréversible (Verger *et al.*, 2019). On constate que des pays ont récemment fait un effort pour dissocier les contrôles des responsabilités des LSLA. Tel est le cas du Danemark, qui a pris des mesures pour protéger les *élèves*, les enseignants et les *écoles* des sanctions en cas de mauvaises performances dans des *évaluations* (voir **Encadré 6**). Ensemble, ces tendances donnent à penser que si le phénomène tend à se répandre dans le monde, il ne se *généralise* pas.

---

### **Encadré 6. Protéger les élèves, les enseignants et les écoles des sanctions liées aux performances**

Le Danemark a mis en place une évaluation à grande échelle fondé sur le recensement pour mesurer les compétences dans diverses matières du niveau 2 au niveau 8. Ce test, informatisé et adaptatif (de telle sorte qu'aucun élève ne passe le même test que les autres) fait également partie de la batterie d'outils pédagogiques mise à la disposition des enseignants pour établir des diagnostics, aider à évaluer les performances des élèves et concevoir des plans d'études individualisés.

Les résultats sont analysés au niveau de chaque élève, de groupes d'élèves (agrégés par sexe, âge, situation socioéconomique ou autre), ainsi que de la salle de classe, de l'école, de la municipalité et de la région. Ils ne sont rendus publics à l'échelon national que par sous-catégories d'élèves. Les données demeurent confidentielles et protégées par une loi nationale. En garantissant l'anonymat des élèves et la confidentialité d'autres données d'identification, celle-ci protège les écoles, les enseignants et les élèves contre d'éventuelles mesures punitives liées aux résultats de l'évaluation.

*Source* : Basé sur Shewbridge *et al.*, 2011 ; OCDE, 2014.

---

Ces observations ont d'importantes conséquences pour le débat concernant les LSLA. Dans un certain nombre de pays à revenu élevé, la diffusion et la consolidation des tests standardisés sont allées de pair avec l'adoption de sévères instruments de contrôle (ou l'ont précédée). Dans de pareils cas, le bien-fondé des LSLA et les risques qu'elles comportent ont généralement été considérés comme un aspect des cadres de réédition des comptes. Une telle approche n'est toutefois pas appropriée quand on s'interroge sur le rôle et l'impact des LSLA dans les pays à revenu faible ou moyen. De plus, subordonner le débat sur l'évaluation au débat sur les contrôles présentant des enjeux élevés n'est pas sans danger s'agissant des pays riches. Cela d'autant plus que l'on peut facilement méconnaître certaines des conséquences de l'adoption d'instruments d'évaluation – et non de leur utilisation à des fins de contrôle des responsabilités.

De plus, subordonner le débat sur l'évaluation au débat sur les contrôles présentant des enjeux élevés n'est pas sans danger s'agissant des pays riches.

## Évaluations à faibles enjeux et enseignement en fonction des tests : un nouveau débat

L'attention accrue qui est portée aux cadres de réédition des comptes à enjeux élevés met en évidence notre compréhension et notre reconnaissance limitées des conséquences des évaluations sans grands enjeux. C'est ainsi que l'enseignement en fonction des tests et l'appauvrissement des programmes d'enseignement ont été étudiés en général comme une conséquence involontaire alimentée par les examens à enjeux élevés et non par les LSLA ayant pour objet premier de produire des diagnostics sur la santé générale du système éducatif. Certains chercheurs concèdent que les inconvénients de l'enseignement en fonction des tests inhérents à l'influence des LSLA sur les politiques sont peut-être exagérés (Ritzen, 2013). D'autres font toutefois valoir que même les évaluations nationales qui sont sans grandes conséquences pour l'apprenant peuvent pousser les enseignants à adapter leur enseignement en raison des conséquences symboliques des LSLA et du tort qui pourrait être causé à leur réputation et à celle de leur établissement. Nous l'avons vu, les pays utilisent de plus en plus leur classement global dans les palmarès d'évaluations internationales telles que le PISA comme un indicateur national des progrès et améliorations réalisées – ou pas<sup>43</sup>. Les autorités responsables de l'éducation peuvent alors être tentées de remanier les programmes d'enseignement ou d'adapter les pratiques pédagogiques en fonction du contenu ou du style des évaluations transnationales afin d'améliorer les scores des élèves. De même, les évaluations nationales non assorties formellement d'incitations matérielles pour les enseignants et les écoles sont parfois utilisées à dessein pour orienter les choix des écoles et encourager la compétition entre elles. Cela peut se produire en particulier lorsque les résultats aux tests sont ventilés par établissement et mis à la disposition du grand public, parfois sous la forme de palmarès et de classements<sup>44</sup>. De la sorte, même les évaluations conçues sans grands enjeux peuvent finir par en acquérir sur le plan symbolique, encourageant les écoles à entrer en compétition les unes avec les autres ou les poussant à tenter de remanier leurs ensembles d'épreuves<sup>45</sup>. Un nouveau débat se dessine alors sur la question de savoir si les évaluations transnationales fondées sur un échantillon ou les évaluations nationales à faibles enjeux pourraient elles aussi influencer sur les stratégies d'enseignement et d'apprentissage.

Même les évaluations conçues sans grands enjeux peuvent finir par en acquérir sur le plan symbolique, encourageant les écoles à entrer en compétition les unes avec les autres ou les poussant à tenter de remanier leurs ensembles d'épreuves.

43 Voir par exemple Takayama (2008), Niemann *et al.* (2017) ou Baroutsis et Lingard (2017) pour une analyse des cas du Japon, de l'Allemagne et de l'Australie, respectivement.

44 De fait, même en l'absence de tels dispositifs, il arrive que des écoles mettent en avant leurs résultats à des fins de promotion.

45 Cette dynamique se manifeste plus souvent dans les systèmes scolaires où le choix entre établissements est très large et obéit aux lois du marché. Dans de tels contextes, les écoles sont davantage incitées à concentrer leurs efforts sur l'amélioration de leur performance dans les tests standardisés car la réputation ainsi acquise aura probablement un effet déterminant sur le nombre d'inscriptions (voir par exemple Falabella, 2014 en ce qui concerne le Chili).

# Réflexions en guise de conclusion

Un nombre croissant de travaux spécialisés ont solidement mis en évidence le potentiel des évaluations à grande échelle des apprentissages (LSLA) en tant que moteurs du progrès de l'éducation. Au cours des dernières décennies, l'intérêt pour les LSLA organisées aux niveaux national, régional et international n'a fait que croître – atteignant sans doute un point de non-retour. Quelle que soit la forme que prendront les systèmes éducatifs de demain, les LSLA ne sont pas près de disparaître. Non seulement le nombre de pays qui entreprennent de telles évaluations va croissant, mais le marché de l'évaluation a connu lui aussi une profonde transformation. Toutes sortes d'initiatives visant à soutenir la mise en œuvre d'évaluations ont vu le jour, avec le concours d'organismes de développement internationaux et régionaux. Le tableau général a ainsi gagné en complexité et s'est diversifié, et ces transformations ont créé de nouveaux défis.

Comme nous l'avons fait valoir ici, les LSLA sont riches de promesses, mais les avantages doivent être mis en balance avec les risques potentiels. Certains de ces risques résultent directement de la conception des LSLA, tandis que d'autres sont la conséquence d'utilisations inappropriées des résultats. Si ces déficiences ne rendent pas les évaluations inutiles ou intrinsèquement problématiques, une plus grande vigilance s'impose néanmoins pour éviter les risques d'abus. Il importe de concevoir les outils de manière plus rigoureuse et équilibrée et d'être plus prudent dans l'utilisation des résultats obtenus de façon à éclairer convenablement les décisions politiques.

Il apparaît en outre que les LSLA sont de plus en plus considérées comme allant de soi dans la plupart des systèmes éducatifs. Or, on commence à comprendre que l'impact final des LSLA dépend essentiellement de *la manière* dont nous les utilisons – au moins autant que de leur conception et des méthodes de collecte des données. Pourtant, un certain nombre de questions concernant l'exploitation des données ne sont encore pas suffisamment étudiées ni débattues. Voici quelques aspects auxquels il conviendrait de consacrer plus d'attention, de réflexion et de recherches.

## Pour des évaluations qui tiennent leurs promesses

**Comprendre les évaluations dans leur contexte.** Les fins que les LSLA sont censées servir et les utilisations que l'on entend en faire sont multiples. Toutefois, nous ne savons trop *comment* et dans *quelles conditions* elles peuvent véritablement tenir leurs promesses, c'est-à-dire contribuer efficacement à des réformes des politiques et/ou des pratiques qui permettent à terme d'améliorer la qualité et l'équité en matière d'éducation. Il est donc nécessaire d'affiner notre

compréhension de l'impact des facteurs contextuels – économiques, institutionnels et politiques – sur les LSLA.

Cela est essentiel compte tenu des éléments nouveaux qui font apparaître de possibles effets pervers et conséquences involontaires des évaluations. Des recherches toujours plus nombreuses montrent comment les LSLA peuvent encourager des comportements indésirables ou même opportunistes de la part de diverses parties prenantes. Il convient donc de mener des recherches additionnelles pour comprendre quelles sont les formes d'évaluation et les conditions qui ont des chances d'améliorer la qualité et l'équité en matière d'éducation, et quelles sont celles qui risquent de produire des effets contraires.

**Au-delà des évaluations à grande échelle des apprentissages.** Comme nous avons tenté de le démontrer tout au long de cette étude, des ajustements pourraient être nécessaires pour que les LSLA fassent avancer les efforts en faveur de la qualité et de l'équité. Cela étant, les LSLA ne peuvent naturellement pas servir *tous* les buts et objectifs auxquels un système d'évaluation est censé répondre. Une réflexion plus poussée devra être menée sur le rôle que peuvent jouer d'autres modalités d'évaluation, y compris les évaluations réalisées en salle de classe à des fins formatives ou de diagnostic, les évaluations sur dossier, l'observation ou les évaluations par les pairs. Ces types d'évaluation peuvent en effet se révéler mieux adaptés pour mesurer les acquis pertinents dans des domaines ne se prêtant pas à des comparaisons à l'échelle du système éducatif ou entre pays. Il convient donc de s'intéresser davantage à ces autres formes d'évaluation afin de déterminer quels sont les moyens les plus appropriés de rendre compte de toutes les compétences auxquelles nous attachons de la valeur.

De même, si les évaluations organisées par les gouvernements ou au niveau intergouvernemental ont fait l'objet d'une attention considérable au cours des dernières décennies, d'autres modalités nouvelles sont moins bien comprises. C'est le cas en particulier des évaluations d'initiative citoyenne et des enquêtes auprès des ménages qui comportent une part d'évaluation. L'intérêt pour ces questions demeure en grande partie circonscrit aux communautés de pratiques qui s'y consacrent. De ce fait, les possibilités et les risques que présentent ces évaluations sont rarement étudiés dans les recherches dominantes. Pourtant, l'étude de ces autres formes d'évaluation pourrait mettre en évidence de nouvelles façons de lever certaines limites caractéristiques des LSLA classiques ou de compenser ces limites. Plus précisément, des modalités nouvelles pourraient assurer des niveaux d'équité et d'inclusion plus élevés en prenant en compte des langues additionnelles et les populations peu accessibles ou les enfants non scolarisés ou déscolarisés. Une réflexion éclairée sur leur potentiel et leurs limites pourrait faciliter la conception de cadres d'évaluation efficaces.

## L'avenir des évaluations

Les possibles déficiences relevées plus haut doivent être corrigées avant que les LSLA puissent servir au mieux la cause de la qualité et de l'équité en matière d'éducation. Toutefois, la plupart des difficultés liées aux LSLA ne se révéleront que beaucoup plus tard. Voici quelques questions qui appellent une réflexion plus poussée :



**Le rôle du secteur privé.** Le monde de l'évaluation est peuplé par un nombre d'organisations de plus en plus diversifiées – dont certaines sont peu connues et rarement appelées à rendre des comptes. Toutes sortes d'acteurs invisibles participent en effet aux processus d'évaluation. Parmi eux figurent en particulier la plupart des contractants qui effectuent une large part du travail d'évaluation, tout en étant relativement peu visibles et peu exposés à l'examen du public. Étant donné le statut privé de nombre d'entre eux, ces intermédiaires restent fréquemment dans l'ombre. Il convient donc d'explorer ce nouveau territoire en prêtant une attention particulière au rôle du secteur privé.

Un changement de l'environnement d'apprentissage. On sait encore peu de choses sur la manière dont la nature de l'apprentissage a évolué au cours de ces dernières années, et en particulier des effets des mutations technologiques sur l'apprentissage. Les questionnaires de contexte qui accompagnent habituellement les évaluations n'ont prêté qu'une attention limitée et inégale aux pratiques d'apprentissage des enfants et des jeunes, ainsi qu'à leurs attitudes à l'égard des activités d'apprentissage se déroulant à l'intérieur et à l'extérieur du cadre scolaire. De plus amples efforts sont requis pour intégrer ces dimensions dans les questionnaires, de façon que les évaluations permettent de mieux comprendre les environnements nouveaux et en constante mutation dans lesquels évoluent les apprenants du XXI<sup>e</sup> siècle.

**L'éthique de l'évaluation.** L'existence d'un ensemble toujours plus vaste de données d'évaluation soulève toute une série de questions éthiques concernant le consentement et la transparence, la propriété des données, et la sécurité et l'anonymat. On note un intérêt croissant pour la formulation de normes d'éthique relatives à la collecte et à l'utilisation des données, mais ces préoccupations restent encore timides. Des recherches et une réflexion plus poussées sont nécessaires pour mieux comprendre la nature des risques et des défis et pour clarifier les différents niveaux de responsabilité à cet égard. Il importe aussi de forger une vision partagée des bonnes pratiques en matière de gestion des données qui permette de relever adéquatement les nouveaux défis créés par l'émergence des « mégadonnées » et de l'intelligence artificielle.

**L'exploitation des données d'évaluation comme « mégadonnées ».** Le concept de « mégadonnées » est de fait l'un des plus surexploités et les moins compris de notre époque. Malgré la place éminente qui lui est faite dans les débats publics et spécialisé, on ne sait toujours pas clairement si les données d'évaluation entrent ou non dans cette catégorie. L'essor des évaluations informatisées et des tests adaptatifs, ainsi que le recours croissant aux algorithmes pour analyser les données d'évaluation et en tirer des conclusions ne se sont pas accompagnés d'une réflexion ou d'un débat concernant les conséquences de ces tendances pour les apprenants évalués ou les systèmes éducatifs en général – ainsi que pour notre conception de la qualité et de l'équité dans l'éducation. Il est donc essentiel d'avoir une idée plus claire des rapports entre mégadonnées et données d'évaluation afin que les données demeurent un outil au service des parties prenantes de l'éducation – et non l'inverse.

Il est d'autant plus urgent d'élargir et d'approfondir notre compréhension de ces défis nouveaux que l'éducation et l'apprentissage sont en pleine mutation. Nous devons nous engager dans la réflexion critique et le dialogue de fond qui s'imposent si nous voulons tenir notre engagement collectif de garantir à tous un apprentissage efficace et pertinent.

## Références

- ACER-ISU. 2017. *The Principles of Good Practice in Learning Assessment*. Melbourne, ACER.
- Addey, C. 2014. *Why do countries join international literacy assessments? An Actor-Network Theory analysis with case studies from Lao PDR and Mongolia*. United Kingdom School of Education and Lifelong Learning, Université d'East Anglia.
- Addey, C. 2018. The assessment culture of international organizations: from philosophical doubt to statistical certainty through the appearance and growth of international large-scale assessments in C. Alarcón and M. Lawn (dir. publ.), *Assessment Cultures: Historical perspectives*. Francfort-sur-le-Main, Peter Lang.
- Addey, C., and Sellar, S. 2019. Cela en vaut-il la peine ? Raisons de participation (ou non) aux évaluations internationales à grande échelle des apprentissages. *Recherche et Prospective en Éducation*, n° 24. Paris, UNESCO.
- Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B. et Verger, T. 2017. The rise of international large-scale assessments and rationales for participation. *Compare : A Journal of Comparative and International Education*.
- Anderson, J. A. 2005. *Accountability in education*. Paris: Institut international de l'UNESCO pour la planification de l'éducation.
- Agence suédoise de coopération internationale pour le développement (ASDI). 2018. Page Web sur l'éducation au sida. <https://www.sida.se/English/how-we-work/our-fields-of-work/education/> (consulte en juillet 2018).
- Baird, J. A., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G. et Yu, G. 2016. On the supranational spell of PISA in policy. *Educational Research*, Vol. 58, No. 2, pages 121-138.
- Banque mondiale. 2011. *Learning for All: Investing in People Knowledge et Skills to Promote Development*. Washington D.C.: Banque mondiale.
- Banque mondiale. 2018. *Rapport sur le développement dans le monde 2018 : La promesse de l'éducation*. Washington, D.C.: Banque mondiale.
- Barnes, S., Lyonette, C., Atfield, G. et Owen, D. 2016. *Teachers' Pay and Equality : A Literature Review – Longitudinal Research into the Impact of Changes to Teachers' Pay on Equality in Schools in England*. Warwickshire, Royaume-Uni, Warwick Institute for Employment Research.
- Baroutsis, A. et Lingard, B. 2017. Counting and comparing school performance: an analysis of media coverage of PISA in Australia, 2000–2014. *Journal of Education Policy*, Vol. 32, No. 4, pages 432-449.
- Barrett, A. M. 2016. Measuring Learning Outcomes and Education for Sustainable Development: the new education development goal, dans W. C. Smith, et W. C. Smith (dir. publ.), *The Global Testing shaping education policy, perceptions, and practice*. Royaume-Uni, Symposium Books.
- Beggs, C. 2016. Overview of EGRA and EGMA. Global Alliance to Monitor Learning. [http://gaml.isu.unesco.org/wp-content/uploads/sites/2/2018/08/4\\_USAID\\_20160511.pdf](http://gaml.isu.unesco.org/wp-content/uploads/sites/2/2018/08/4_USAID_20160511.pdf) (consulté en février 2019).
- Benavot, A. et Köseleci, N. 2015. *Seeking Quality in Education: The Growth of National Learning Assessments, 1990-2013*. Étude commandée pour le Rapport mondial de suivi sur l'EPT 2015.

- Benson, C. et Wong, K. 2015. Development discourse on language of instruction and literacy: Sound policy and Ubuntu, or lip service? *Reconsidering Development*, Vol. 4, No. 1, pages 1-16.
- Best, M., Knight, P, Lietz, P, Lockwood, C., Nugroho, D. et Tobin, M. 2013. *The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries*. Londres, EPPI-Centre, Social Science Research Unit, Institute of Education, Université de Londres.
- Biesta, G. 2009. Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, Vol. 21, No. 1, pages 33-46.
- Breakspear, S. 2014. *How does PISA shape education policy making? Why how we measure learning determines what counts in education*. Australie, Center for Strategic Education.
- Brese, F., Jung, M., Mirazchiyski, P., Schulz, W. et Zuehlke, O. 2011. *ICCS 2009 User Guide for the International Database – Supplement 5*. Amsterdam, Secrétariat de l'IEA.
- Brink, S. 2018. Learning from promising practices in ensuring effective use of large-scale learning assessment data for education policy and practice: Developing evidence from national and international LSLA for policy and practice. Présentation faite lors de l'atelier TALENT, novembre 2018, Dakar, Sénégal. [http://www.education2030-africa.org/images/Developing\\_evidence\\_from\\_national\\_and\\_international\\_LSLA\\_for\\_politique\\_and\\_pratique.pdf](http://www.education2030-africa.org/images/Developing_evidence_from_national_and_international_LSLA_for_politique_and_pratique.pdf) (consulté en février 2019).
- Colclough, C. 2012. Education, poverty and development—mapping their interconnections. *Comparative Education*, Vol. 48, No. 2, pages 135-148.
- Clarke, M. 2012. *What Matters Most for Student Assessment Systems: A Framework Paper*. SABER Working Paper No. 1. Washington D.C., Groupe de la Banque mondiale.
- Clarke, M. 2016. Value for money: What do developing countries really get from participating in international assessment programs? Présentation faite le 13 janvier 2016, Washington D.C. [http://siteresources.worldbank.org/1818SOCIETY/Resources/Value\\_for\\_Money\\_Clarke\\_13Jan2016.pdf](http://siteresources.worldbank.org/1818SOCIETY/Resources/Value_for_Money_Clarke_13Jan2016.pdf) (consulté en mai 2019).
- Clarke, P. 2017. Making use of assessments for creating stronger education systems and improving teaching and learning. Document de travail pour le Rapport mondial de suivi sur l'éducation 2017/18.
- Cresswell, J., Schwantner, U. et Waters, C. 2015. *A Review of International Large-Scale Assessments in Education: Assessing Component Skills and Collecting Contextual Data*, PISA, Banque mondiale, Washington, D.C./OCDE Publishing, Paris, OCDE.
- Darling-Hammond, L. 2004. Standards, accountability and school reform. *Teachers College Record*, Vol. 106, No. 6, pages 1047-1085.
- De Coster, I. et Sigalas, E. 2017. *L'éducation à la citoyenneté à l'école en Europe*. Commission européenne : Agence exécutive « Éducation, audiovisuel et culture ».
- DFAT Australia. 2015. *Strategy for Australia's aid investments in education 2015–2020*. Australie, Ministère des affaires étrangères et du commerce.
- DFAT Canada. 2014. *Canada's International Education Strategy: Harnessing our knowledge advantage to drive innovation and prosperity*. Ottawa, Ministère des affaires étrangères, du commerce et du développement du Canada.
- DFID UK. 2018. *DFID Education Policy: Get Children Learning*. Londres, Département du développement international.

- Diaz, M. et Flores, G. 2008. SERCE, Resultados Nacionales [SERCE, résultats nationaux]. Mexico D.F., Instituto Nacional para la Evaluación de la Educación.
- Dixon, R., Arndt, C., Mullers, M., Vakkuri, J., Engblom-Pelkkala, K. et Hood, C. 2013. A Lever for Improvement or a Magnet for Blame? Press and Political Responses to International Educational Rankings in Four EU Countries. *Public Administration*, Vol. 91, No. 2, pages 484-505.
- Dolata, S. 2011. Are there any disparities between girls and boys in the response of the education sector of HIV and AIDS? Assessment of educational HIV/AIDS prevention programmes applied by SACMEQ III countries. Communication présentée au Forum politique de l'IIPE sur l'égalité des genres dans l'éducation, 3-4 octobre, 2011, Paris, France.
- Earle, D. 2010. *How can tertiary education deliver better value to the economy?* Tertiary education occasional paper 2010/08. Nouvelle-Zélande, Gouvernement néo-zélandais.
- Engel, L. C. et Rutkowski, D. 2018. *Pay to play: what does PISA participation cost in the US?* Discourse, Studies in the Cultural Politics of Education. Routledge.
- Falabella, A. 2014. The Performing School: The Effects of Market & Accountability Policies. *Education Policy Analysis Archives*, Vol. 22, No. 70.
- Fenwick, T. et Edwards, R. 2014. Networks of knowledge, matters of learning, and criticality in higher education. *Higher Education*, Vol. 67, No. 1, pages 35-50.
- Ferrer, G. et Fiszbein, A. 2015. *What has happened with learning assessment systems in Latin America? Lessons from the Last Decade of Experience.* The Dialogue: Leadership for the Americas. Commission for quality education for all: background paper. Washington D.C., Banque mondiale.
- Feuer, M. 2012. *No country left behind: rhetoric and reality of international large-scale assessment.* William Angoff Memorial Lecture Series. New Jersey, Educational Testing Service.
- Fischman, G. et Topper, A. M. 2017. *An Examination of the Influence of International Large Scale Assessments and Global Learning Metrics on National School Reform Policies.* Mary Lou Foulton Teachers College, Arizona States University.
- Fiske, E.B. 2000. *L'évaluation des acquis scolaires. Éducation pour tous – Situation et tendances 2000.* Paris, UNESCO, Paris. Secrétariat du Forum sur l'Éducation pour tous.
- Friedman, T., Schwantner, U., Spink, J., Tabata, N. et Waters, C. 2016. *Improving Quality Education and Children's Learning Outcomes and Effective Practices in the Eastern and Southern Africa Region: Main Report.* Report for UNICEF ESARO. Camberwell, Victoria, ACER Press.
- Fullan, M. 2010. Positive pressures. Dans A. Hargreaves, A. Lieberman, M. Fullan et D. Hopkins (dir. publ.). *Second international handbook of educational change* (pages 119-130). Springer, Dordrecht.
- Goldstein, H. et Thomas, S. 2008. Reflections on the international comparative surveys debate. *Assessment in Education: Principles, Policy & Practice*, Vol. 15, No. 3, pages 215-222.
- Gorard, S. 2001. International Comparisons of School Effectiveness: the second component of the 'crisis account' in England? *Comparative Education*, Vol. 37, No. 3, pages 279-296.
- Gorur, R., et Wu, M. 2014. Leaning too far? PISA, politics and Australia "top five" ambitions. *Discourse: Studies in the Cultural Politics of Education*, Vol. 36, No. 5, pages 647-664.
- Greeney, V. 2008. *Assessing National Achievement Levels in Education.* Washington DC, Banque mondiale.
- Hamilton, L.S., Stecher, B.M. et Klein, S.P. 2002. *Making Sense of TBA in Education.* Santa Monica, CA, Rand Corporation.

- Hamilton, M. 2017. How International Large-Scale Skills Assessments engage with national actors : mobilising networks through policy, media and public knowledge. *Critical Studies in Education*, Vol. 58, No. 3, pages 280-294.
- Hannon, V., Patton, A. et Temperley, J. 2011. Developing an Innovation - Ecosystem for Education. Indianapolis, CISCO.
- Hanushek, E. A., et Woessmann, L. 2015. *The Knowledge Capital of Nations: Education and the Economics of Growth*. Cambridge, MA, MIT Press.
- Hardy, I. et Boyle, C. 2011. My school? Critiquing the abstraction and quantification of education. *Asia Pacific Journal of Teacher Education*, Vol. 39, No. 3, pages 211-222.
- IEA. 2015. TIMSS 2015 Context Questionnaires. <https://timssandpirls.bc.edu/timss2015/questionnaires/index.html> (consulté en mai 2019).
- IEA. 2016. PIRLS 2016 Context Questionnaires. <https://timssandpirls.bc.edu/pirls2016/questionnaires/index.html> (consulté en mai 2019).
- IEA. 2019. National Study Reports. [http://pub.iea.nl/national\\_reports.html](http://pub.iea.nl/national_reports.html). (consulté en mai 2019).
- Institut international de l'UNESCO pour la planification de l'éducation. 2017. *Planifier l'éducation, préparer le futur : 10<sup>e</sup> Stratégie à moyen terme, 2018-2021*. Paris, UNESCO IIEP.
- Institut de statistique de l'UNESCO. 2016a. *The cost of not assessing learning outcomes*. Information Paper No. 26. Montréal, Institut de statistique de l'UNESCO.
- Institut de statistique de l'UNESCO. 2016b. *Comprendre ce qui fonctionne dans les évaluations orales de la lecture*. Montréal, Institut de statistique de l'UNESCO.
- Institut de statistique de l'UNESCO. 2018a. The impact of large-scale learning assessments. Montréal, Institut de statistique de l'UNESCO.
- Institut de statistique de l'UNESCO. 2018b. *Costs and Benefits of Different Approaches to Measuring the Learning Proficiency of Students (SDG Indicator 4.1.1)*. Information Paper No. 53. Montréal, Institut de statistique de l'UNESCO.
- Institut de statistique de l'UNESCO. 2018c. *SDG4 Data Digest. Data to Nurture Learning*. Montréal, Institut de statistique de l'UNESCO.
- Institut de statistique de l'UNESCO. 2019a. Base de données. <http://ISU.unesco.org/> (consulté en janvier 2019).
- Institut de statistique de l'UNESCO. 2019b. Glossaire. <http://uis.unesco.org/fr/glossary> (consulté en janvier 2019).
- Kamens, D. et McNeely, C. 2010. Globalization and the Growth of International Educational Testing and National Assessment. *Comparative Education Review*, Vol. 54, No. 1, pages 5-25.
- Kellaghan, T. et Greaney, V. 2001. *L'évaluation pour améliorer la qualité de l'enseignement*. Paris, UNESCO-IIEP.
- Kellaghan, T., Bethell, G. et Ross, J. 2011. *National and international assessments of student achievement. Guidance Note*. A DFID Practice Paper. Londres, Département du développement international (DFID).
- King, K. 2017. Lost in translation? The challenge of translating the global education goal and targets into global indicators. *Compare : A Journal of Comparative and International Education*, Vol. 47, No. 6, pages 801-817.
- Komatsu, H. et Rapple, J. 2017. A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth. *Comparative Education*, Vol. 53, No. 2, pages 166-191.

- Koretz, D. 2017. *The testing charade. Pretending to make schools better*. Chicago et Londres, The University of Chicago Press.
- Larsen, M. A. 2005. A Critical Analysis of Teacher Evaluation Policy Trends. *Australian Journal of Education*, Vol. 49, No. 3, pages 292-305.
- Lingard, B. et Grek, S. 2007. The OECD, indicators and PISA: an exploration of events and theoretical perspectives. ESRC/ESF Research Project on Fabricating Quality in Education Working Paper 2. Université d'Édimbourg.
- Lingard, B., et Sellar, S. 2013. 'Catalyst data': Perverse systemic effects of audit and accountability in Australian schooling. *Journal of Education Policy*, Vol. 28, No. 5, pages 634-656.
- LMTF (Learning Metrics Task Force). 2013. *Vers l'apprentissage universel: recommandations du Comité de réflexion sur la métrique de l'apprentissage*. Montréal et Washington, D. C., Institut de statistique de l'UNESCO et le Centre pour l'éducation universelle à la Brookings Institution.
- Lockheed, M., 2015. *Why do countries participate in international large-scale assessments? The Case of PISA*. Policy Research Working Paper. Banque mondiale.
- Mbiti, I., Romero, M. et Schipper, Y. 2018. *The Challenge of Designing Effective Teacher Performance Pay Programs: Experimental Evidence from Tanzania*. Working Paper. [http://barrett.dyson.cornell.edu/NEUDC/paper\\_613.pdf](http://barrett.dyson.cornell.edu/NEUDC/paper_613.pdf)
- Ministre de l'éducation, Finlande. 2006. *Education Strategy for Finland's Development Cooperation*. Finlande, Ministère de l'éducation.
- Mitra, D., Mann, B. et Hlavacik, M. 2016. Opting Out : Parents Creating Contested Spaces to Challenge Standardized Tests. *Education Policy Analysis Archives*, Vol. 24, No. 31.
- Montoya, S. 2016. *The Cost of Ignorance Revisited: A Reply*. <https://norrag.wordpress.com/2016/02/08/measuring-learning-the-cost-of-ignorance/> (consulté en juin 2017).
- Niemann, D., Martens, K. et Teltemann, J. 2017. PISA and its consequences: Shaping education policies through international comparisons. *European Journal of Education*, Vol. 52, No. 2, pages 175-183.
- Nikel, J. et Lowe, L. 2010. Talking of fabric: a multi- dimensional model of quality in education. *Compare : A Journal of Comparative and International Education*, Vol. 40, No. 5, pp. 589-605.
- Norwegian Agency for Development Cooperation (Agence norvégienne de coopération pour le développement - NORAD). 2018. Education Webpage. <https://norad.no/en/front/thematic-areas/education/> (consulté en juillet 2018).
- OCDE. 2009. *PISA Technical Report*. Paris, OCDE.
- OCDE. 2010. *The high cost of low educational performance : The long-run economic impact of improving PISA outcomes*. Paris, OCDE.
- OCDE. 2011. *Improving access to PISA for students with disabilities and other special education needs*. Paris, OCDE.
- OCDE. 2013. *Synergies for Better Learning. An international perspective on evaluation and assessment. OECD Reviews of Evaluation and Assessment in Education*. Paris, OCDE.
- OCDE. 2014. *Education Policy Outlook: Denmark*. Paris, OCDE.
- OCDE. 2015. Courir par temps chaud – Introduction. PISA 2015 test item. [https://www.oecd.org/pisa/test/161117\\_PISA%202015%20MS%20-%20Released%20Item%20Descriptions\\_FINAL\\_French.pdf](https://www.oecd.org/pisa/test/161117_PISA%202015%20MS%20-%20Released%20Item%20Descriptions_FINAL_French.pdf) (consulté en décembre 2018).
- OCDE. 2016a. *PISA Results in Focus*. Paris, OCDE.

- OCDE. 2016b. Brochure PISA pour le développement. [http://www.oecd.org/pisa/pisa-for-development/PISA-D\\_brochure\\_2016\\_FRE.pdf](http://www.oecd.org/pisa/pisa-for-development/PISA-D_brochure_2016_FRE.pdf) (consulté en août 2017).
- OCDE. 2017. *PISA 2015 Collaborative Problem-Solving Framework*. Paris, OCDE.
- OCDE. 2018a. *Le futur de l'éducation et des compétences : Projet Éducation 2030*. Paris, OCDE.
- OCDE. 2018b. Data Management Procedures. Dans *PISA for Development Technical Report*. Paris, OCDE. <http://www.oecd.org/pisa/pisa-for-development/pisaforddevelopment2018technicalreport/PISA-D%20TR%20Chapter%2010%20-%20DataManagement%20-%20final.pdf> (consulté en mai 2019).
- Parker, R. et Fraillon, J. 2016. *Southeast Asia Primary Learning Metrics (SEA-PLM) - Global Citizenship Domain Assessment Framework*. Melbourne, ACER.
- Partenariat mondial pour l'éducation. 2016. *GPE 2020 Strategic Plan*. Washington D.C., GPE.
- Partenariat mondial pour l'éducation. 2018. *GPE 2020 Strategic Plan*. Washington D.C., GPE.
- PASEC. 2015. PASEC 2014. *Performances des systèmes éducatifs en Afrique subsaharienne francophone : Compétences et facteurs de réussite au primaire*. Dakar: Programme d'analyse des systèmes éducatifs de la CONFEMEN.
- People's Action for Learning (PAL) Network. 2019. Our growth. <https://palnetwork.org/our-growth/> (consulté en février 2019).
- Pizmony-Levy, O. 2013. *Back to the Future in International Assessments*. [http://blogs.edweek.org/edweek/assessing\\_the\\_assessments/2014/04/back\\_to\\_the\\_future\\_how\\_international\\_large-scale\\_assessments\\_came\\_about.html](http://blogs.edweek.org/edweek/assessing_the_assessments/2014/04/back_to_the_future_how_international_large-scale_assessments_came_about.html) (consulté en août 2017).
- Pizmony-Levy, O., Harvey, J., Schmidt, W.H., Noonan, R., Engel, L., Feuer, M.J., Braun, H., Santorno, Rotberg, I.C., Ash, P., Chatterji, M. et Torney-Purta, J. 2014. On the merits of, and myths about, international assessments. *Quality Assurance in Education*, Vol. 22, No. 4, pages 319-338.
- Pizmony-Levy, O. et Green Sarisky, N. 2016. *Who opts out and why? Results from a national survey on opting out of standardized test*. New York, Teachers College – Université Columbia.
- Ramirez, F. O., Schofer, E. et Meyer, J.W. 2018. International Tests, National Assessments, and Educational Development (1970–2012). *Comparative Education Review*, Vol. 62, No. 3, pages 344-364.
- Raudonyte, I. 2019. *Use of learning assessment data in education policy-making*. Document de travail de l'IIPE-UNESCO. Paris, IIPE-UNESCO.
- Ritzen J. 2013. International large-scale assessments as change agents. von Davier M, Gonzalez E, Kirsch I, et al. (dir. publ.). *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Dordrecht, Pays-Bas, Springer, pages 13-24.
- Rowe, E. 2017. Australia Country Case Study. Document de travail sur la pratique des échantillonnages sélectifs en Australie établi pour le Rapport mondial de suivi sur l'éducation 2017/18.
- Rutkowski, L. et Rutkowski, D. 2016. A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher*, Vol. 45, No. 4, pages 252-257.
- Ryan, S.V., von der Embse, N.P., Pendergast, L.L., Saeki, E., Segool, N. et Schwing, S. 2017. Leaving the teaching profession: The role of teacher stress and educational accountability policies on turnover intent. *Teaching and Teacher Education*, Vol. 66, pages 1-11.
- Sahlberg, P. 2007. Education policies for raising student learning: The Finnish approach. *Journal of education policy*, Vol. 22, No. 2, pages 147-171.
- Sayed, Y., Ahmed, R. et Mogliacci, R. 2018. The 2030 Global Education Agenda and the SDGs: Process, Policy and Prospects. Dans A. Verger, M. Novelli, & H. K. Altinyelken (dir. publ.). *Global Education Policy and International Development (2e édition)*. Londres/New York, Bloomsbury, pages 185-208.

- Shewbridge, C., Jang, E., Matthews, P. et Santiago, P. 2011. *OECD Reviews of Evaluation and Assessment in Education*. Paris, OCDE.
- Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I. et Scherer, R. 2016. Taking a future perspective by learning from the past – A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review*, Vol. 19, pages 58-84.
- Singal et Sabates. 2016. *Access and learning are equally important for children with disabilities*. Site Web du Partenariat mondial pour l'éducation. <https://www.globalpartnership.org/blog/access-and-learning-are-equally-important-children-disabilities> (consulté en mai 2019).
- Smith, W.C. 2016. An Introduction to the Global Testing Culture. Dans W. C. Smith (dir. publ.), *The Global Testing Culture; shaping educational policy, perceptions and practice*. Oxford Studies in Comparative Education Series, Symposium Books.
- Smith, W.C. 2019. . Evaluating the Global Testing Culture: A Response to Ramirez and Colleagues. *Comparative Education Review*, Vol. 63, No. 1, pages 121-126.
- Southeast Asia Primary Learning Metrics (SEA-PLM), SEAMEO, UNICEF et ACER. 2017. Assessment Framework. Revised draft document incorporating feedback from the Domain Technical Review Teams.
- Steiner-Khamsi, G. 2003. The politics of league tables. *Journal of Social Science Education*. OJSSE 1-2003, Civic Education.
- Steiner-Khamsi, G., Appleton, M. et Vellani, S. 2017. Understanding business interests in international large-scale student assessments: a media analysis of *The Economist*, *Financial Times*, and *Wall Street Journal*. *Oxford Review of Education*, Vol. 44, No. 2, 190-203.
- Tawil, S., Akkari, A. et Macedo, B. 2012. Au-delà du labyrinthe conceptuel: la notion de qualité en éducation. *Recherche et prospective en éducation n° 02*. Paris, UNESCO.
- The Education Commission (Commission internationale pour le financement de possibilités d'éducation dans le monde). 2016. The Learning Generation – Investing in education for a changing world. *Comparative Education Review*, Vol. 61, No. 2, pages 214-217.
- Takayama, K. 2008. The politics of international league tables: PISA in Japan's achievement crisis debate. *Comparative Education*, Vol. 44, No. 4, pages 387-407.
- Tikly, I. et Barrett, A.M. 2007. *Education Quality; Research priorities and approaches in the global era*. Bristol, University of Bristol.
- Tobin, M., Lietz, P., Nugroho, D., Vivekanandan, R. et Nyamkhuu, T. 2015. *Using large-scale assessments of students' learning to inform education policy Insights from the Asia-Pacific region*. ACER et UNESCO Bangkok, ACER.
- Torney-Purta, J. et Amadeo, J.A. 2013. International large-scale assessments: Challenges in reporting and potentials for secondary analysis. *Research in Comparative and International Education*, Vol. 8, No. 3, pages 248-258.
- Initiative mondiale du Secrétaire général de l'ONU pour l'éducation. 2012 -2016. *Initiative mondiale pour l'éducation avant tout*.
- UNESCO, UNICEF, l'Institution Brookings et Banque mondiale. 2017. Aperçu MELQO: *Initiative sur la mesure de la qualité et des acquis d'apprentissage préscolaire*. Paris, UNESCO.
- UNESCO, Bureau international d'éducation (UNESCO-IBE). 2013. *L'apprentissage dans l'agenda pour l'éducation et le développement post-2015*.



- UNESCO-Santiago (Bureau régional de l'UNESCO pour l'éducation en Amérique latine et dans les Caraïbes - OREALC/UNESCO Santiago). 2007. *Quality Education For All : A human rights issue. Educational policies within the framework of the II Intergovernmental Meeting of the Regional Project in Education for Latin America and the Caribbean (EFA/PRELAC)*. Document de travail. Santiago di Chili, Bureau régional de l'UNESCO pour l'éducation en Amérique latine et dans les Caraïbes.
- UNESCO. 1974. *Recommandation sur l'éducation pour la compréhension, la coopération et la paix internationales et l'éducation relative aux droits de l'homme et aux libertés fondamentales*. Dans Actes de la Conférence générale. Dix-huitième session. Paris, 17 octobre-23 novembre 1974, Vol. 1. Résolutions. Paris, UNESCO.
- UNESCO. 1996. *L'Éducation : un trésor est caché dedans, rapport à l'UNESCO de la Commission internationale sur l'éducation pour le vingt et unième siècle (extraits)*. Président : J. Delors. Paris, UNESCO.
- UNESCO. 2014. *Rapport mondial de suivi sur l'EPT, 2013-2014. Enseigner et apprendre : atteindre la qualité pour tous*. Paris, UNESCO.
- UNESCO. 2015a. *Repenser l'éducation: vers un bien commun mondial ?* Paris, UNESCO.
- UNESCO. 2015b. *Rapport mondial de suivi sur l'EPT 2015. Éducation pour tous 2000-2015 : progrès et enjeux*. Paris, UNESCO.
- UNESCO. 2016a. *Comprendre l'Objectif de développement durable 4: Education 2030*. Paris, UNESCO.
- UNESCO. 2016b. *Rapport mondial de suivi sur l'éducation (GEM) 2016 : l'éducation pour les peuples et la planète*. Paris, UNESCO.
- UNESCO. 2016c. *Déclaration d'Incheon : Éducation 2030: vers une éducation inclusive et équitable de qualité et un apprentissage tout au long de la vie pour tous*. Paris, UNESCO.
- UNESCO. 2017a. *Analyzing and utilizing assessment data for better learning outcomes*. Paris et Bangkok, UNESCO.
- UNESCO. 2017b. *Un guide pour assurer l'inclusion et l'équité dans l'éducation*. Paris, UNESCO.
- UNESCO. 2017c. *Évaluation de l'apprentissage à l'UNESCO : Assurer un apprentissage efficace et pertinent pour tous*. Paris, UNESCO. [https://unesdoc.unesco.org/ark:/48223/pf0000260325\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000260325_fre)
- UNESCO. 2017d. *Rendre des comptes en matière d'éducation : tenir nos engagements*. Rapport mondial de suivi sur l'éducation (GEM). Paris, UNESCO.
- UNESCO. 2017e. *Large-scale Assessment Data and Learning Outcomes: Linking Assessments to Evidence-based Policy Making and Improved Learning*. Bangkok, UNESCO.
- UNESCO. 2017f. *SDG 4 Framework for Action: Symposium on Policy implications of learning assessments to enhance educational quality in the Gulf Cooperation Council States*. Doha, UNESCO.
- UNESCO. 2018. *Building tomorrow compétences numériques - what conclusions can we draw from international comparative indicateurs?* Working Papers of Education Policy No. 6. Paris, UNESCO.
- Ungerleider, C. 2006. Reflections on the use of large-scale student assessment for improving student success. *Canadian Journal of Education*, Vol. 29, No. 3, pages 873-883.
- Unterhalter, E. 2019. The Many Meanings of Quality Education: Politics of Targets and Indicators in SDG 4. *Global Policy*, Vol. 10, pages 39-51.
- Unterhalter, E. et Brighouse, H. 2007. Distribution of what for social justice in education ? The case of Education for All by 2015. M. Walker et E. Unterhalter (dir. publ.). *Amartya Sen's capability approach and social justice in education* Palgrave Macmillan, New York, pages 67-86.
- Union européenne. 2014. *Learning Assessments*. DEVCO B4 Education discussion paper. Bruxelles, Union européenne.

- USAID. 2011. *Education Opportunity Through Learning*. USAID Education Strategy. Washington D.C., USAID.
- Verger, A. et Parcerisa, L. 2017. Accountability and education in the post-2015 scenario: International trends, enactment dynamics and socio-educational effects. Document de travail pour le *Rapport mondial de suivi 2017/2018*.
- Verger, A., Parcerisa, L. et Fontdevila, C. 2018. The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms. *Educational Review*, Vol. 71, No. 1, pages 5-30.
- Verger, A., Fontdevila, C. et Parcerisa, L. 2019. Reforming governance through politique instruments: how et to what extent normes, tests et redditionnel in éducation spread worldwide. *Discourse: Studies in the Cultural Politics of Éducation*, pp. 1-23.
- Wagner, D.A. 2011. *Des évaluations simples, rapides et abordables: améliorer l'apprentissage dans les pays en développement*. Paris: UNESCO-IIPE.
- Wagner, D., Lockheed, M., Mullis, I., Martin, M., Kanjee, A., Gove, A. et Dowd, A. 2012. The debate on learning assessments in developing countries. *Compare: A Journal of Comparative and International Education*, Vol. 42, No. 3, pp. 509-545.
- Wagner, D.A., Wolf, S. et Boruch, R.F. (eds.) 2018. *Learning at the bottom of the pyramid: Science, measurement, and policy in low-income countries*. Paris: UNESCO-IIPE.
- Waldow, F., Takayama, K. et Sung, Y.K. 2014. Rethinking the Pattern of External Policy Referencing: Media Discourses Over the 'Asian Tigers'. PISA Success in Australia, Germany and South Korea. *Comparative Education*, Vol. 50, No. 3, pp. 302-321.
- Walter, S. et Benson, C. 2012. Language policy and medium of instruction in formal education. B. Spolsky (ed.). *The Cambridge Handbook of Language Policy*. Cambridge: Cambridge University Press, pp. 278-300.
- Weiss, I., Knapp, M., Hollweg, K. et Burrill, G. 2002. *Investigating the influence of Standards: A framework for research in mathematics, science and technology*. Washington D.C.: National Academy Press. Washington D.C.: National Academy Press.
- Winthrop, R. et Anderson, K. 2013. Can international large-scale assessments inform a global learning goal? Insights from the Learning Metrics Task Force. *Research in Comparative and International Education*, Vol. 8, No. 3, pp. 279-295.
- Yasukawa, K., Hamilton, M. et Evans, J. 2017. A comparative analysis of national media responses to the OECD Survey of Adult Skills: policy making from the global to the local? *Compare: a journal of comparative and international education*, Vol. 47, No. 2, pp. 271-285.
- Yasunaga, M. 2014. *Non-formal education as a means to meet learning needs of out-of-school children and adolescents*. Online publication: UNESCO/UNICEF.
- Zadja, J. 2014. Understanding the quality debate in education. Conference paper. Faculty of Education and Arts, Australian Catholic University (Melbourne Campus).

# Annexe 1 :

# Évaluations transnationales à grande échelle des apprentissages

ÉTUDES INTERNATIONALES						
Évaluation	Domaines évalués	Groupes cibles	Organisme d'exécution	Années de mise en œuvre	Liste actualisée des pays participants	Recherches et analyses consolidées – gérées par les organismes d'exécution
TIMSS TIMSS Numeracy TIMSS Advanced	Mathématiques, sciences	Niveaux 4 et 8 Niveaux 4 à 6  Élèves en dernière année du deuxième cycle de l'enseignement secondaire	Association internationale pour l'évaluation du rendement scolaire (IEA)	1995, 1999, 2003, 2008, 2007, 2011, 2015  Prochaine étude : 2019	<a href="http://timss2015.org/timss-2015/about-timss-2015/">http://timss2015.org/timss-2015/about-timss-2015/</a>  <a href="http://timssandpirls.bc.edu/timss2015/international-results/advanced/timss-advanced-2015/about-timss-advanced/">http://timssandpirls.bc.edu/timss2015/international-results/advanced/timss-advanced-2015/about-timss-advanced/</a>	<a href="https://timssandpirls.bc.edu/isc/publications.html">https://timssandpirls.bc.edu/isc/publications.html</a>  <a href="https://iccs.iea.nl/resources/publications.html">https://iccs.iea.nl/resources/publications.html</a>  <a href="https://icils.acer.org/">https://icils.acer.org/</a>  <a href="http://pub.iea.nl/national_reports.html">http://pub.iea.nl/national_reports.html</a>
PIRLS PIRLS Literacy	Maîtrise de la lecture	Niveau 4 Niveaux 4 à 6		2001, 2006, 2011, 2016  Prochaine étude : 2021	<a href="http://timssandpirls.bc.edu/pirls2016/international-results/pirls/about-pirls-2016/pirls-2016-countries/">http://timssandpirls.bc.edu/pirls2016/international-results/pirls/about-pirls-2016/pirls-2016-countries/</a>	
ICCS	Instruction civique et éducation à la citoyenneté	Niveau 8		2009, 2016	<a href="https://iccs.iea.nl/cycles/2016/countries.html">https://iccs.iea.nl/cycles/2016/countries.html</a>	
ICILS	Initiation à l'informatique et à l'information	Niveau 8		2013  Prochaine étude : 2018	<a href="https://www.iea.nl/icils">https://www.iea.nl/icils</a>	
LaNA	Maîtrise de la lecture et du calcul	Niveaux 4 à 6		Étude pilote	<i>Destinée aux pays en développement qui ne sont pas prêts pour les enquêtes TIMSS ou PIRLS</i>	

ÉTUDES INTERNATIONALES						
Évaluation	Domaines évalués	Groupes cibles	Organisme d'exécution	Années de mise en œuvre	Liste actualisée des pays participants	Recherches et analyses consolidées – gérées par les organismes d'exécution
PISA	Lecture, mathématiques et sciences  Domaines ouverts à l'innovation  Initiation aux questions financières (en option)	Élèves âgés de 15 ans	Organisation de coopération et de développement économique (OCDE)	2000, 2003, 2006, 2009, 2012, 2015, 2018	<a href="https://www.oecd.org/pisa/aboutpisa/pisa-2018-participants.htm">https://www.oecd.org/pisa/aboutpisa/pisa-2018-participants.htm</a>	<a href="http://www.oecd.org/pisa/publications/">http://www.oecd.org/pisa/publications/</a>  <a href="https://www.oecd.org/pisa/pisa-for-development/">https://www.oecd.org/pisa/pisa-for-development/</a>
PISA-D	Lecture, mathématiques et sciences	Élèves âgés de 15 ans		2017	<a href="https://www.oecd.org/pisa/aboutpisa/pisa-for-development-participating-countries.htm">https://www.oecd.org/pisa/aboutpisa/pisa-for-development-participating-countries.htm</a>	

ÉTUDES RÉGIONALES						
Évaluation	Domaines évalués	Groupes cibles	Organisme d'exécution	Années de mise en oeuvre	Liste actualisée des pays participants	Recherches et analyses consolidées – gérées par les organismes d'exécution
PASEC <sup>46</sup>	Mathématiques et compétences en lecture	Niveaux 2 et 6	Conférence des ministres de l'éducation des États et gouvernements de la Francophonie (CONFEMEN)	2014 Prochaine étude : 2019	<a href="http://www.pasec.confemen.org/wp-content/uploads/2015/12/Resume_Pasec2014_FR_BD1.pdf">http://www.pasec.confemen.org/wp-content/uploads/2015/12/Resume_Pasec2014_FR_BD1.pdf</a>  (voir la section Contexte, objectifs et méthodologie de l'enquête PASEC 2014)	<a href="http://www.pasec.confemen.org/pasec2014-3/">http://www.pasec.confemen.org/pasec2014-3/</a>
PILNA	Compétences en lecture, écriture et calcul	Niveaux 4 et 6	Educational Quality and Evaluation Programme (EQAP) de la Communauté du Pacifique (CPS)	2012 et 2015	<a href="https://www.eqap.org.fj/getattachment/Resources/Final-Regional-Report.pdf.aspx">https://www.eqap.org.fj/getattachment/Resources/Final-Regional-Report.pdf.aspx</a>  (voir Executive Summary)	<a href="https://www.eqap.org.fj/work/Assessment.aspx">https://www.eqap.org.fj/work/Assessment.aspx</a>
PERCE	Mathématiques, sciences et lecture	Niveaux 3 et 6	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation) (LLECE)	1997	<a href="https://unesdoc.unesco.org/ark:/48223/pf0000243532">https://unesdoc.unesco.org/ark:/48223/pf0000243532</a>  (voir Introducción general)	<a href="http://www.unesco.org/new/es/santiago/education/education-assessment-llece/first-regional-comparative-and-explanatory-study/">http://www.unesco.org/new/es/santiago/education/education-assessment-llece/first-regional-comparative-and-explanatory-study/</a>
SERCE				2006		<a href="http://www.unesco.org/new/es/santiago/education/education-assessment-llece/second-regional-comparative-and-explanatory-study-serce/">http://www.unesco.org/new/es/santiago/education/education-assessment-llece/second-regional-comparative-and-explanatory-study-serce/</a>
TERCE				2013		<a href="http://www.unesco.org/new/es/santiago/education/education-assessment-llece/terce/documents/">http://www.unesco.org/new/es/santiago/education/education-assessment-llece/terce/documents/</a>

46 Avant 2014, le PASEC a mené à bien 35 évaluations nationales dans plus de 20 pays francophones d'Afrique, d'Asie et du Moyen-Orient.

ÉTUDES RÉGIONALES						
Évaluation	Domaines évalués	Groupes cibles	Organisme d'exécution	Années de mise en oeuvre	Liste actualisée des pays participants	Recherches et analyses consolidées – gérées par les organismes d'exécution
SACMEQ ou SEACMEQ	Lecture (1995-1999) Lecture et mathématiques (autres éditions) Connaissances relatives au VIH-sida en 2006-2010	Niveau 6	Consortium de l'Afrique australe et orientale pour le pilotage de la qualité de l'éducation (SACMEQ/SEACMEQ)	1995-1999, 2000-2004, 2006-2010, 2012-2014	<a href="http://www.sacmeq.org/?q=sacmeq-projects/sacmeq-iv">http://www.sacmeq.org/?q=sacmeq-projects/sacmeq-iv</a>	<a href="http://www.sacmeq.org/?q=publications">http://www.sacmeq.org/?q=publications</a>
SEA-PLM	Lecture, écriture, mathématiques et citoyenneté mondiale	Niveau 5	Organisation des ministres de l'éducation des pays du Sud-Est asiatique (SEAMEO), en collaboration avec l'UNICEF et le Conseil australien de recherche pédagogique (ACER)	2018-2019	<a href="http://www.seaplms.org/seaplms/index.php/about/countries">http://www.seaplms.org/seaplms/index.php/about/countries</a>	<a href="http://www.seaplms.org/seaplms/index.php/capacity-building-2">http://www.seaplms.org/seaplms/index.php/capacity-building-2</a>

*Note* : Pour les besoins de la présente étude, l'EGRA et l'EGMA (voir annexe 2) ne sont *pas* considérées comme des évaluations transnationales dans la mesure où, même si les modèles d'évaluation appliqués dans différents pays reposent sur un cadre méthodologique commun, ils ne sont *pas* conçus pour permettre les comparaisons entre pays.

## Annexe 2 : Évaluations nationales utilisant un cadre méthodologique commun

**Évaluation de la lecture dans les petites classes (EGRA) et Évaluation des compétences en mathématiques dans les petites classes (EGMA).** Il s'agit d'épreuves orales, qui se déroulent en face à face et généralement dans des écoles. Les résultats sont souvent utilisés pour plaider en faveur de réformes du système éducatif. Les épreuves, de conception simple et d'un faible coût, visent à mesurer les compétences en lecture et en mathématiques. Elles ont été mises au point par le Research Triangle Institute (RTI International) – en 2006 pour l'EGRA et quelques années plus tard pour l'EGMA. L'EGRA (lecture) évalue la reconnaissance des caractères, la distinction des phonèmes, la capacité de lire des mots simples et la compréhension orale. L'EGMA (mathématiques) évalue la reconnaissance des chiffres, la capacité de procéder à des comparaisons et la capacité de classer dans l'ordre des séries d'objets. L'EGRA a été utilisée par plus de 30 organisations, dans plus de 70 pays et adaptée en plus de 120 langues ; l'EGMA a été organisée dans 22 pays et adaptée en 24 langues (Beggs, 2016).

**Évaluations d'initiative citoyenne.** Elles sont généralement pilotées par des organismes citoyens ou des organisations de la société civile, plutôt que par des gouvernements. Les épreuves se déroulent aux domiciles des enfants plutôt que dans des écoles et mesurent les compétences de base en lecture et en calcul. Les évaluations d'initiatives citoyennes produisent des estimations récurrentes des niveaux d'instruction élémentaire des enfants et (à ce jour) tendent à être conçues et administrées de façons similaires. De telles évaluations sont conduites dans 14 pays dans le cadre du réseau People's Action for Learning. Ce mouvement a été lancé en Inde (2005), puis repris au Pakistan (2006), au Kenya et en Tanzanie (2007), en Ouganda (2009), au Mali (2010), au Sénégal (2010), au Mexique (2013), au Nigéria (2014), au Bangladesh, au Cameroun, au Ghana et en Mozambique (2016), et au Népal (2017). Au total, 7,5 millions d'enfants, vivant dans des conditions de très grande vulnérabilité ont passé les tests lors de 48 évaluations réalisées dans 30 langues, avec l'aide de 600 000 volontaires (PAL Network, 2019).







Organisation  
des Nations Unies  
pour l'éducation,  
la science et la culture

Secteur de  
l'éducation

## La promesse de l'évaluation

### à grande échelle des apprentissages

#### Reconnaître les limites pour libérer les potentialités

Le Programme de développement durable à l'horizon 2030 et l'agenda Éducation 2030 attachent une importance nouvelle à la nécessité d'offrir à tous des possibilités d'apprentissage équitables, efficaces et pertinents. Des efforts accrus sont donc faits pour suivre les progrès accomplis dans la réalisation de cet objectif. Les évaluations de l'apprentissage à grande échelle (*Large-scale learning assessments – LSLA*), qui se sont imposées comme un outil stratégique décisif pour vérifier les résultats des processus éducatifs, occupent une place centrale dans cette entreprise. De fait, le principe s'en est répandu dans le monde entier au cours des deux dernières décennies et elles sont aujourd'hui pratiquées dans plus de la moitié des pays du globe. Leur utilité comme moteur du progrès dans le domaine de l'éducation est confirmée par un nombre croissant d'études spécialisées.

Mais si les utilisations potentielles de cet outil d'évaluation ne cessent de se multiplier, leurs limites nous sont moins connues. Divers acteurs et chercheurs du monde de l'éducation se sont inquiétés d'éventuelles conséquences indésirables des LSLA. Conséquences qui pourraient résulter de leur conception même et du mauvais usage qui risque d'être fait des données d'évaluation pour influencer les politiques et les pratiques dans le domaine de l'éducation. Idéalement, il conviendrait que la compréhension que l'on a des effets bénéfiques des LSLA pour la réalisation de l'agenda de l'éducation soit mise en balance avec une égale compréhension des écueils en cas d'utilisation incorrecte. La présente publication entend contribuer au débat mondial en cours en passant en revue ce que l'on sait concernant les aspects positifs des LSLA tout en clarifiant leurs possibles conséquences négatives.

