



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



Institut international de
planification de l'éducation



Des évaluations simples, rapides et abordables

Améliorer l'apprentissage dans les pays en développement

Daniel A. Wagner



Une éducation de qualité pour tous

Des évaluations simples, rapides et abordables

Des évaluations simples, rapides et abordables

**Améliorer l'apprentissage
dans les pays en développement**

Daniel A. Wagner

Paris

UNESCO : Institut international de planification de l'éducation
Organisation internationale de la Francophonie (OIF)

Publié en 2015 par :
l'Institut international de la planification de l'éducation (IIPE)
de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO)
7-9 Rue Eugène Delacroix, 75116 Paris, France

et

l'Organisation internationale de la Francophonie
19/21 Avenue Bosquet
75007 Paris, France



© UNESCO 2015
ISBN : 978-92-803-2361-0



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 (www.unesco.org/open-access/terms-use-ccbysa-fr).

La présente licence s'applique exclusivement aux contenus textes de la publication.

Titre original :

Smaller, quicker, cheaper: Improving learning assessments for developing countries

Publié en 2011 par l'IIPE-UNESCO et le Partenariat mondial pour l'éducation (PME)



Les désignations employées dans cette publication et la présentation des données qui y figurent n'impliquent de la part de l'IIPE-UNESCO ou de l'OIF 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'IIPE-UNESCO ni de l'OIF et n'engagent en aucune façon ces Organisations.

Photo de couverture : UNESCO (Linda Shen)

Création graphique : IIPE

Graphisme de la couverture : IIPE

Illustrations : IIPE

Mise en page : Linéale Production

Imprimé par l'atelier d'impression de l'IIPE

REMERCIEMENTS DE L'AUTEUR

Le présent rapport* est une commande de l'IIPE-UNESCO, avec le soutien du Partenariat mondial pour l'éducation (PME), dans le cadre du Projet d'indicateurs de la qualité de l'apprentissage (QLIP). Je tiens à remercier l'IIPE et son personnel pour leur hospitalité tout au long du travail effectué en 2009 sur le QLIP et, en particulier, Mark Bray et Ken Ross pour leur soutien indéfectible, ainsi qu'Alcyone Vasconcelos qui a piloté le projet sous l'égide du PME. Mes plus vifs remerciements vont à plusieurs spécialistes, pour leur aide et leurs contributions à divers chapitres, dans le cadre du travail de fond commandé par l'IIPE : Nadir Altinok (*chapitre 4 et annexe A*), Scott Paris (*chapitres 5 et 6*), Liliane Sprenger-Charolles et Souhila Messaoud-Galusi (*chapitre 5 et annexes A et B*) et Andrew Babson (*chapitre 7*) – des parties de leur travail ont été adaptées afin d'être intégrées à cet ouvrage. Un grand merci également aux autres collègues qui ont apporté des idées et du contenu pour certaines parties de ce projet ou les versions préliminaires de cet ouvrage : Helen Abadzi, Gina Arnone, Samer Al-Samarrai, Aaron Benavot, Erik Bloom, Yuko Butler, Colette Chabott, Luis Crouch, Stéphanie Dolata, Peggy Dubeck, Cesar Guadalupe, Amber Gove, Vincent Greaney, Robin Horn, Matthew Jukes, Anil Kanjee, Ann Kennedy, Ines Kudo, Marlaine Lockheed, Ana Luisa Machado, Paul McDermott, Ina Mullis, Benjamin Piper, Bob Prouty, Jake Ross, Carlos Ruano, Andreas Schleicher, Kathleen Trong, Pierre Varly, Larry Wolff et d'autres. Je remercie vivement Koli Banik, Ryan Clennan, Carollyne Hutter, Amy Orr et Estelle Zadra pour leur aide précieuse lors de la révision du texte et de la préparation des documents, sans oublier Ruju Vyas et Rohen d'Aiglepierre pour leurs commentaires. Il va de soi que toutes les erreurs factuelles et d'interprétation éventuelles sont la responsabilité exclusive de l'auteur et ne sauraient en aucun cas représenter les points de vue des personnes susmentionnées, du PME, de l'IIPE, de l'UNESCO ou de toute autre agence ou organisation.

* Dans sa version originale en anglais.

TABLE DES MATIÈRES

Remerciements de l'auteur	5
Liste des tableaux	9
Liste des figures	10
Liste des abréviations et acronymes	12
Préface à l'Édition française	15
Préface	17
Résumé analytique	19
1. Introduction	27
Remarques préliminaires	27
L'histoire d'Aminata, village de Kahalé	28
Ce que signifie l'histoire d'Aminata	29
Structure de l'ouvrage	30
Limites	31
But et prochaines étapes	32
2. Acquis d'apprentissage et buts stratégiques	35
EPT et acquis d'apprentissage	35
La promesse d'une éducation de meilleure qualité	37
L'importance des mesures	42
Interrogations suscitées par les évaluations	42
3. Ce que les indicateurs d'apprentissage peuvent changer	45
Utilisations des indicateurs d'apprentissage	45
Définir et mesurer l'apprentissage	46
Apprendre à l'école et en dehors de l'école	47
Quelles sont mes options ? Le point de vue d'un Ministre de l'Éducation	52
4. Évaluation de l'apprentissage dans les pays en développement	55
Principales évaluations de l'apprentissage	55
Quelle doit être l'ampleur du champ des évaluations ?	64
Comparabilité des évaluations	76
Autres aspects du choix des évaluations	81
Crédibilité des évaluations	93
Choisir un type d'évaluation	96
5. Évaluer les compétences en lecture des enfants	99
Pourquoi la lecture ?	99
La science de l'apprentissage de la lecture	100
Évaluations des compétences en lecture	120
6. Tests de lecture : problèmes et perspectives	129
Quelques questions concernant l'évaluation des lecteurs débutants	129
Implications pédagogiques des évaluations	134

	Méthodes d'évaluation des compétences en lecture : quelques observations complémentaires	142
7.	Coût des évaluations	153
	Analyse coût-bénéfice des évaluations dans le domaine de l'éducation	154
	Calculer les coûts	155
	Catégories de coûts et comparaisons des coûts de quelques évaluations	157
	Comment envisager les coûts ?	165
	Additionner les coûts	168
8.	Évaluation du niveau d'alphabétisation des adultes	171
	Importance de l'alphabétisation des adultes aujourd'hui	171
	Évaluer l'alphabétisation des adultes	174
	Besoins des apprenants adultes	177
	Acquisition de la lecture chez l'enfant et chez l'adulte	178
	Perte d'alphabétisme	181
	Faire progresser l'évaluation des compétences en lecture des adultes	182
9.	Recommandations	183
	Le test de lecture « idéal » n'existe pas	183
	En cas de doute, opter pour des évaluations de taille modeste	186
	Les résultats les plus rapides sont les meilleurs	187
	Les résultats des évaluations ne sont pas toujours à la hauteur du prix payé	187
	Les évaluations des acquis scolaires doivent commencer le plus tôt possible (dans certaines limites)	188
	L'évaluation doit viser l'amélioration de l'enseignement	189
	La comparabilité transnationale ne garantit pas toujours une éducation de qualité pour tous	190
	Les « biais » culturels ne nuisent pas forcément à une évaluation	192
	Les nouvelles évaluations peuvent également favoriser l'alphabétisation des adultes	193
	La responsabilité de l'impact de l'apprentissage doit être largement partagée	194
	Les évaluations hybrides peuvent renforcer sensiblement l'impact d'une politique	195
10.	Conclusions	197
	Quelques points non résolus pour l'élaboration de nouvelles évaluations	198
	Utilisation des technologies de l'information et de la communication	201
	Aller de l'avant	202
	L'histoire d'Aminata revisitée	203
	Références	205
	Annexes	233
	Annexe A. Description des évaluations des compétences en lecture	235
	Annexe B. Exemples d'items utilisés dans les évaluations des compétences en lecture	247
	À propos de l'auteur	255

LISTE DES TABLEAUX

Tableau 2.1	Impact des compétences de base sur le revenu	40
Tableau 3.1	Temps annuel moyen d'instruction, par région et par année d'études, en heures, 2000	50
Tableau 4.1	Participation des pays du PME aux évaluations internationales, régionales et hybrides au cours de la dernière décennie	59
Tableau 4.2	Indicateurs de la participation à l'école primaire	80
Tableau 5.1	Estimation des taux d'analphabétisme et d'alphabétisme des adultes (population de plus de 15 ans) par région, 1990 et 2000-2004	104
Tableau 6.1	Pourcentage d'enfants, par classe, par niveau de lecture, toutes écoles confondues, 2010	149
Tableau 7.1	Catégories de coûts des évaluations employées dans une sélection d'études	158
Tableau 7.2	Études des coûts d'une sélection d'évaluations nationales, régionales et transnationales	159
Tableau 7.3	Coûts des évaluations nationales, régionales, internationales et EGRA	159
Tableau 7.4	Coûts par catégorie, en pourcentage des dépenses totales liées à l'évaluation	163
Tableau 9.1	Résumé des avantages et des limites de diverses évaluations	184

LISTE DES FIGURES

Figure 2.1	Comprendre la qualité de l'éducation	39
Figure 2.2	Alphabétisme des mères et scolarisation des enfants au Niger, en RDP lao et en Bolivie, 2000	39
Figure 3.1	Contexte d'alphabétisation et compétences en lecture dans l'enquête PIRLS, 2001	49
Figure 4.1	Progression de l'utilisation des évaluations nationales de l'apprentissage, 1995-2006	56
Figure 4.2	Continuum des évaluations : des évaluations SQC hybrides aux examens nationaux, en passant par les LSEA	66
Figure 4.3	PIRLS : pourcentage des élèves de 4 ^e année se situant dans le quartile inférieur de l'échelle internationale des capacités de lecture, 2001	68
Figure 4.4	PISA : pourcentage des élèves de 15 ans atteignant 5 niveaux de compétence en lecture, 2000-2002 (dans une sélection de pays)	70
Figure 4.5	SACMEQ : pourcentage d'élèves de 6 ^e année qui maîtrisent la lecture dans sept pays d'Afrique, 1995-1998	71
Figure 4.6	Disparités entre les genres en termes d'aptitudes en langue et en mathématiques en 6 ^e année, d'après les évaluations nationales de l'apprentissage	73
Figure 4.7	Pourcentage de différents groupes linguistiques se situant dans les 20 % des niveaux inférieurs de l'éducation dans une sélection de pays	74
Figure 4.8	Évolution des scores aux tests d'alphabétisme entre SACMEQ I et SACMEQ II	79
Figure 4.9	Taux de retour sur les investissements en capital humain en postulant un investissement de départ égal pour tous les âges	84
Figure 4.10	Écarts liés à la richesse : résultats aux tests par âge pour le décile le plus pauvre et pour le 4 ^e décile en Équateur, 2003-2004	85
Figure 4.11	Facteurs contextuels et compétences en lecture	87
Figure 4.12	Nombre de livres que des élèves de 6 ^e année ont déclaré posséder chez eux dans 15 systèmes éducatifs de pays africains du SACMEQ, 2000	88
Figure 4.13	Pourcentage d'élèves de 4 ^e année dans l'enquête PIRLS 2006	93
Figure 5.1	Approche des programmes d'éducation bilingue fondée sur une « théorie de la distance »	114
Figure 5.2	Gambie : Pourcentage d'élèves incapables de lire un seul mot, 2007 et 2009	127

Figure 6.1	Histogrammes des écarts types constatés chez les élèves Ashaninka (Pérou) lors d'un test de fluidité de la lecture orale et d'un test écrit administré en groupe (N = 40)	147
Figure 8.1	L'analphabétisme dans une sélection de pays en développement, par région	173
Figure 8.2	Adultes n'ayant pas dépassé le niveau d'éducation primaire et déclarant ne pas savoir lire	175
Figure 8.3	Pourcentage d'adultes par niveau de compétences de base en lecture dans l'évaluation nationale de l'alphabétisation des adultes réalisée aux États-Unis	190

LISTE DES ABRÉVIATIONS ET ACRONYMES

CONFEMEN	Conférence des ministres de l'Éducation des pays ayant le français en partage
ECR	essai comparatif randomisé
EGRA	évaluation des compétences fondamentales en lecture (<i>Early Grade Reading Assessment</i>)
EIAA	Enquête internationale sur l'alphabétisation des adultes
EPT	Éducation pour tous
HBES	Enquêtes sur l'éducation réalisées auprès des ménages (<i>Household-based Educational Surveys</i>)
IEA	Association internationale pour l'évaluation du rendement scolaire
IPE	Institut international de planification de l'éducation (UNESCO)
ILI	Institut international de l'alphabétisation
INDISSE	indicateurs dynamiques des savoirs essentiels en lecture
IRI	répertoires de lecture informels (<i>informal reading inventories</i>)
ISU	Institut de statistique de l'UNESCO
L1, L2	première langue (langue maternelle), seconde langue
LAMP	Programme d'évaluation et de suivi de l'alphabétisation (<i>Literacy Assessment and Monitoring Program</i>)
LAP	Programme d'évaluation de l'alphabétisation (<i>Literacy Assessment Project</i>)
LLECE	Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation
LSEA	évaluation à grande échelle de l'apprentissage (<i>Large-scale Educational Assessment</i>)
OCDE	Organisation de coopération et de développement économiques
NRC	coordinateurs nationaux de la recherche (<i>National Research Coordinators</i>)
OMD	Objectifs du Millénaire pour le développement
ONG	organisation non gouvernementale
ONU	Organisation des Nations Unies
ORF	fluidité de la lecture à haute voix (<i>Oral reading fluency</i>)
PASEC	Programme d'analyse des systèmes éducatifs des pays de la CONFEMEN
PME	Partenariat mondial pour l'éducation
PSE	environnement (d'alphabétisation) défavorable (<i>poorly-supported environment</i>)
PIAAC	Programme pour l'évaluation internationale des adultes
PIRLS	Programme international de recherche en lecture scolaire

PISA	Programme international pour le suivi des acquis des élèves
PMA	pays les moins avancés
QCM	question à choix multiples
QLIP	Projet d'indicateurs de la qualité de l'apprentissage
SACMEQ	Consortium de l'Afrique australe et orientale pour le pilotage de la qualité de l'éducation
SERCE	Seconde évaluation LLECE
SIMCE	système de mesure de la qualité de l'éducation (<i>Sistema de medición de la calidad de la educación</i>)
SISS	Deuxième étude internationale sur les mathématiques et les sciences
SQC	simples, rapides et abordables (<i>smaller, quicker, cheaper</i>)
TIC	Technologies de l'information et de la communication
TIMSS	Troisième étude internationale sur les sciences
TRI	théorie des réponses aux items
UNESCO	Organisation des Nations Unies pour l'éducation, la science et la culture
USAID	Agence des États-Unis pour le développement international
WSE	environnement (d'alphabétisation) favorable (<i>well-supported environment</i>)

PRÉFACE À L'ÉDITION FRANÇAISE

Cet ouvrage, initialement publié en anglais par l'Institut international de planification de l'éducation de l'UNESCO (IIPÉ-UNESCO) et le Partenariat mondial pour l'éducation (PME), rejoint les préoccupations de plusieurs États et gouvernements de la Francophonie, qui font face au défi de la qualité de l'éducation. À l'heure du bilan des Objectifs du Millénaire pour le développement et de ceux de l'Éducation pour tous – qui s'achèveront en 2015 – et malgré les remarquables progrès effectués depuis le Forum de Dakar en 2000, la scolarisation primaire universelle et la qualité de l'éducation dispensée demeurent des préoccupations majeures.

Une éducation de qualité favorise le développement des individus et répond aux attentes et aux besoins de la société à laquelle ils appartiennent. Ces dernières années, de nombreux travaux se sont intéressés à la question de la qualité de l'éducation, notamment celle de son évaluation. Ces travaux se complètent et aident à mieux appréhender ces problématiques dans toute leur complexité. Malheureusement, beaucoup de ces travaux sont introuvables en français.

En publiant cette édition française, l'Organisation internationale de la Francophonie (OIF) apporte une importante contribution, nécessaire à la circulation des savoirs. L'auteur, le professeur Daniel Wagner, occupe la Chaire UNESCO en apprentissages et alphabétisation à l'Université de Pennsylvanie. Il présente ici une méthodologie qui peut être utilisée pour des projets menés à une petite échelle. Les valeurs sur lesquelles se fonde la réflexion de l'auteur rejoignent celles que promeut l'OIF, à savoir le droit à une éducation de qualité pour tous, le respect des différences culturelles et linguistiques, la solidarité entre les peuples.

Nous savons aujourd'hui que les partenariats et les synergies entre toutes les parties prenantes sont indispensables pour appuyer efficacement les pays engagés sur le chemin d'une éducation de qualité pour tous. L'OIF se félicite donc de sa collaboration avec l'UNESCO et espère que cette édition – fruit de ce partenariat – répondra aux attentes des lecteurs francophones et contribuera à appuyer les acteurs concernés dans leurs efforts en faveur de l'éducation.

Ma-Umba Mabilia
Directeur de l'Éducation et de la Jeunesse
Organisation internationale de la Francophonie

PRÉFACE

Dans les pays en développement, les enfants sont de plus en plus nombreux à aller à l'école. Depuis le Sommet des Nations Unies sur l'Éducation pour tous (2000, Dakar), ce sont les nations les plus pauvres qui, de façon remarquable, ont le plus développé l'accès à l'éducation.

Ce succès s'accompagne de la prise de conscience qu'une progression rapide de la scolarisation ne suffit pas. Il faut un enseignement de bonne qualité pour tous les enfants, ce qui, à ce jour, n'est pas le cas pour un trop grand nombre d'entre eux. Il est probable que les efforts futurs de développement de l'éducation porteront en priorité sur l'amélioration de la qualité de l'apprentissage et de l'enseignement.

Si les évaluations des acquis scolaires peuvent jouer un rôle important dans de nombreux pays en stimulant les réformes scolaires, bien souvent, elles ne sont ni adaptées aux besoins des pays en développement ni financièrement viables. C'est pourquoi il est indispensable d'élaborer des outils adaptés, capables de mieux mesurer les acquis d'apprentissage auxquels les États souhaitent parvenir.

Le présent ouvrage, intitulé *Des évaluations simples, rapides et abordables. Améliorer l'apprentissage dans les pays en développement*, cherche à mieux comprendre comment les évaluations sont conçues, et leur rôle dans l'amélioration de l'apprentissage. Il s'agit d'un sujet de préoccupation (et de controverse) majeur pour des agences comme celles que nous représentons, ainsi que pour les responsables politiques, les spécialistes et le grand public.

Les conclusions de l'ouvrage confortent l'idée qu'un usage efficace des évaluations des performances scolaires est fondamental pour améliorer l'apprentissage et qu'il est utile d'adopter plusieurs approches, en adaptant chacune à une finalité particulière. Le document nous rappelle, en outre, que la valeur des évaluations des acquis scolaires dépend essentiellement des usages qui en sont faits. Améliorer les évaluations des acquis scolaires peut également aider à accorder davantage d'attention à ceux qui ont le plus de difficultés, ainsi qu'à améliorer l'enseignement en classe et la performance générale des écoles.

Cet ouvrage a donné à l'IPE et au PME une nouvelle occasion de travailler en partenariat pour atteindre un but commun : améliorer l'éducation

dans le monde. Ce livre est l'un des fruits du travail de l'auteur en tant qu'expert invité à l'IIPE, en particulier dans le cadre du pilotage d'un projet de recherche conjoint de nos agences. Nous remercions Dan Wagner d'avoir entrepris cette étude riche en réflexions, approfondie et précieuse pour tous ceux qui veulent améliorer la qualité de l'éducation et son évaluation dans les pays en développement.

Robert Prouty, responsable du PME*
Khalil Mahshi, directeur, IIPE-UNESCO**

* De 2010 à février 2013.

** De 2010 à février 2014.

RÉSUMÉ ANALYTIQUE

En 2015, date limite fixée pour la réalisation des objectifs du Millénaire pour le développement (OMD) des Nations Unies dans le secteur de l'éducation, ainsi que de l'Éducation pour tous (EPT), partout dans le monde, professionnels de l'éducation, responsables politiques et autres parties prenantes mettent tout en œuvre pour tenir leurs engagements. Si les enfants scolarisés sont plus nombreux que jamais, certains rapports montrent cependant que la *qualité* de l'éducation s'est dégradée ces dernières années dans de nombreux pays en développement. Pour comprendre ces tendances, les spécialistes ont besoin de mieux évaluer l'éducation. Le présent ouvrage passe en revue les évaluations des acquis scolaires, leur importance pour constituer une base de données empiriques et quelques idées nouvelles pour améliorer leur efficacité, en particulier pour les enfants qui ont le plus de difficultés.

La principale question abordée dans le présent rapport peut se résumer ainsi : *Les études existantes sur l'évaluation des acquis scolaires (en particulier l'apprentissage de la lecture) peuvent-elles aider à renforcer l'efficacité des efforts pour améliorer les résultats scolaires dans les pays en développement ?* Les principaux thèmes abordés pour répondre à cette vaste question sont exposés dans les dix chapitres qui constituent cet ouvrage.

- 1. Introduction.** Le premier chapitre présente le contexte général, avec la description succincte d'une école rurale en Afrique, où une fillette et son instituteur sont dans une situation d'apprentissage dysfonctionnelle, notamment en ce qui concerne la lecture. Il semble en ressortir que l'évaluation et le suivi peuvent mettre en évidence des besoins éducatifs critiques et suggérer des méthodes proactives pour remédier aux lacunes. Ce chapitre expose également la structure du rapport, ainsi que certaines de ses limites.
- 2. Résultats d'apprentissage et objectifs stratégiques.** Ce chapitre décrit les objectifs de l'EPT, ainsi que leurs liens avec les performances scolaires. Comment faut-il comprendre les modèles de qualité de l'éducation ? Comment une mère transmet-elle effectivement des savoir-faire, des attitudes et des valeurs à ses enfants, même en étant elle-même peu éduquée ? L'auteur expose les mérites de meilleures évaluations, ainsi que quelques préoccupations associées. La question

de la complexité des évaluations et de l'accès limité des parties prenantes à leurs résultats est également abordée.

- 3. Ce que les indicateurs d'apprentissage peuvent changer.** Ce chapitre s'intéresse aux utilisations des indicateurs d'apprentissage, depuis l'élaboration de la politique et la création de normes, jusqu'aux corrélats de l'apprentissage et la conception du matériel pédagogique. Les intrants peuvent être mesurés par rapport aux nombreuses expériences que les enfants apportent à l'école, ainsi que par l'adéquation ou non entre leur environnement d'apprentissage et leurs possibilités d'apprendre. Ce chapitre s'intéresse aux résultats d'apprentissage selon deux grands axes : la mesure des savoir-faire et des contenus qui sont directement enseignés dans les écoles (tests sur le contenu des programmes scolaires, par exemple) ; la mesure de ce que la société attend des apprenants en termes de connaissances et d'aptitudes concrètes (comme lire un journal).
- 4. Évaluation de l'apprentissage dans les pays en développement.** Ce chapitre décrit trois grands types d'évaluations : (a) les *évaluations à grande échelle de l'apprentissage* (Large-scale Educational Assessments – LSEA) sont de plus en plus utilisées par les agences nationales et internationales. Les progrès technologiques et méthodologiques, alliés à la pression politique pour améliorer les systèmes éducatifs, ont nourri cette tendance, y compris dans les pays les moins avancés (PMA). Néanmoins, la complexité croissante des LSEA a conduit certains à remettre en cause leur utilité dans les PMA ; (b) les *enquêtes pédagogiques auprès des ménages* (Household-based Educational Surveys – HBES) utilisent des méthodes d'échantillonnage pour réunir, à l'échelle des ménages, des types spécifiques d'informations sur des groupes de populations cibles, stratifiés en fonction de certains paramètres démographiques désirés. Enfin, (c) les *évaluations hybrides*, plus récentes, accordent une grande importance à un ensemble de facteurs tels que la diversité démographique, la diversité linguistique et orthographique, les différences d'apprentissage entre individus et l'opportunité de l'analyse. Cette approche hybride est qualifiée de « simple, rapide et abordable » (*smaller, quicker, cheaper* – SQC). Une évaluation hybride récente, qui suscite beaucoup d'intérêt dans les PMA, la *Early Grade Reading Assessment* (EGRA) ou évaluation des compétences fondamentales en lecture, est décrite dans ce chapitre, tout comme plusieurs LSEA régionales et internationales.

- 5. Évaluer les compétences en lecture des enfants.** Ce chapitre est consacré à la lecture, indicateur majeur de la qualité de l'éducation et volet essentiel des programmes scolaires partout dans le monde. Dans de nombreux PMA, les problèmes de lecture à l'école primaire font partie des facteurs prédictifs les plus sûrs de désavantages futurs et d'abandon de l'école. Certains enfants vivent dans des *environnements (d'alphabétisation) défavorables* (PSE), contrairement à d'autres, qui vivent dans des *environnements (d'alphabétisation) favorables* (WSE). Cette distinction est importante, car elle aide à mieux désagréger les facteurs qui favorisent l'acquisition de la lecture et l'ensemble des compétences relatives à cette acquisition. Parmi ces compétences figurent le principe alphabétique, la conscience phonémique, la fluidité de la lecture orale, le vocabulaire, la compréhension écrite et l'automatisme. D'autres facteurs, tels que la lecture dans les première et seconde langues, l'orthographe et le système d'écriture, sont également abordés. Diverses évaluations sont étudiées à la lumière des modèles actuels de lecture.
- 6. Tests de lecture : problèmes et perspectives.** Ce chapitre s'intéresse aux problèmes inhérents aux tests et aux évaluations. Par exemple, quelles compétences faut-il évaluer chez les enfants qui commencent à apprendre à lire ? En quoi l'orthographe (le système d'écriture) influence-t-elle l'évaluation ? Dans quelle(s) langue(s) faut-il tester l'enfant ? Telles sont les questions et autres problématiques connexes abordées dans ce chapitre. On s'est en outre aperçu que les « bons » tests (du point de vue empirique) ne sont pas toujours « bons » pour l'enfant du point de vue pédagogique. Comment les évaluations peuvent-elles résoudre ce défi ? Les dernières conclusions d'évaluations internationales et régionales, plusieurs études de terrain récentes ayant utilisé l'EGRA, et les nouveaux indicateurs de compétences du Partenariat mondial pour l'éducation (PME) sont examinés dans la perspective de l'élaboration de nouvelles évaluations visant à mieux mesurer la qualité de l'apprentissage dans les pays en développement.
- 7. Coût des évaluations.** Ce chapitre s'intéresse au poids fiscal des évaluations, question importante pour les responsables de la politique éducative. L'un des aspects majeurs à prendre en compte est le coût de l'expertise technique requise de l'organisme national ou international chargé des tests, ainsi que les capacités humaines nationales. Pour comparer le coût des LSEA et celui des évaluations de style SQC d'envergure plus modeste, il est essentiel d'examiner les questions d'échelle, d'opportunité et de rentabilité. Il faut

également trouver un compromis entre les considérations de temps et d'argent. Si, d'après les données disponibles, le coût par élève de l'EGRA semble du même ordre que celui des LSEA, le coût de l'EGRA devrait diminuer dans le futur, à mesure que les pays se familiariseront avec ses outils et que la formation des enquêteurs s'améliorera. De plus, il faut prendre en compte d'importants coûts d'opportunité : les LSEA évaluent généralement les enfants en 4^e année (voire plus tard), à un moment où ils sont susceptibles d'être très en retard dans l'apprentissage de la lecture, ce qui peut entraîner des coûts élevés de remédiation qu'une évaluation précoce pourrait éviter.

8. **Évaluation du niveau d'alphabétisme des adultes.** Ce chapitre s'intéresse au faible niveau d'alphabétisation et à l'analphabétisme des adultes, qui constituent actuellement un problème majeur à l'échelle internationale. En effet, faute d'évaluations exploitables, il est impossible de déterminer s'ils savent lire et écrire, quel est leur niveau de compétences et, par conséquent, comment concevoir des réponses politiques adaptées. Ce chapitre aborde également la question des besoins d'apprentissage des adultes (demande par rapport à l'offre), compare les modèles d'acquisition de la lecture chez les enfants et chez les adultes, et examine la notion de perte d'alphabétisme. L'alphabétisation des adultes est importante à la fois pour les droits de l'homme et pour la croissance économique. Elle constitue également un facteur prédictif majeur des compétences en lecture des enfants. Améliorer l'évaluation des compétences en lecture des adultes, en s'inspirant des outils employés pour évaluer les enfants, pourrait contribuer de façon significative à réaliser l'EPT.
9. **Recommandations.** Ce chapitre postule qu'il est possible de choisir entre divers outils de mesure et d'évaluation. Il faut calibrer les évaluations en fonction des objectifs stratégiques spécifiques, de l'opportunité et de leur coût, en adoptant ce que l'on appelle globalement l'approche SQC. Les questions abordées dans le présent ouvrage ont débouché sur un ensemble de recommandations stratégiques résumées ci-après.
 - i. ***Le test de lecture « idéal » n'existe pas.*** Un test de lecture, comme n'importe quel outil d'évaluation, n'est utile que dans la mesure où il répond à des besoins stratégiques particuliers. Les

décideurs doivent définir précisément leurs buts avant de choisir telle ou telle approche.

- ii. ***En cas de doute, opter pour des évaluations de taille modeste.*** Les évaluations SQC ont un avantage évident, lié à leur dimension modeste, à savoir que les ressources humaines nécessaires correspondent davantage aux capacités humaines réelles des sociétés à faibles revenus.
- iii. ***Les résultats les plus rapides sont les meilleurs.*** Les LSEA sont administrées tous les trois ou cinq ans, voire tous les dix ans, la complexité des comparaisons internationales exigeant plus de temps. À l'inverse, les évaluations hybrides ont des objectifs et des tailles d'échantillons plus ciblés et peuvent être plus fréquentes. Il est alors possible d'avoir une analyse en temps réel, avec des contreparties substantielles.
- iv. ***Les résultats des évaluations ne sont pas toujours à la hauteur du prix payé.*** Dans les processus de coût, il existe des compromis qui font que payer plus cher ne garantit pas nécessairement la réalisation des objectifs politiques visés. Les évaluations hybrides peuvent se révéler une méthode beaucoup moins onéreuse pour effectuer le travail d'évaluation.
- v. ***Les évaluations des acquis scolaires doivent commencer le plus tôt possible (dans certaines limites).*** Les compétences des enfants (ou des adultes) peuvent être mesurées à de nombreux stades, mais les bénéfices sont plus importants lorsqu'il existe un moyen concret d'évaluer les compétences au début d'une longue trajectoire d'apprentissage.
- vi. ***L'évaluation doit viser l'amélioration de l'enseignement.*** Les évaluations hybrides de la lecture peuvent être planifiées de façon à apporter des modifications au niveau de la classe (ou de l'individu), avant que l'enfant ne quitte le système scolaire. Les résultats de l'évaluation doivent guider les chefs d'établissement et les enseignants pour aider les enfants à apprendre.
- vii. ***La comparabilité transnationale ne garantit pas toujours une éducation de qualité dans les pays en développement.*** Le but des LSEA internationales est de permettre des comparaisons transnationales, ce qui n'est généralement pas le cas des évaluations hybrides, dont la conception fait qu'elles sont plutôt axées sur des comparaisons à l'échelle nationale. Elles offrent donc

certaines formes de comparabilité que les LSEA ne permettent pas. L'importance des types de comparabilité dépend des objectifs stratégiques visés.

- viii. ***Les biais culturels ne nuisent pas forcément à une évaluation.*** Si beaucoup d'experts estiment que les préjugés culturels sont « préjudiciables », l'aspect négatif dépend surtout du cadre de référence de chacun. Conçues pour mieux s'adapter à des contextes spécifiques, les évaluations hybrides de type SQC ont à cet égard un avantage relatif.
- ix. ***Les nouvelles évaluations peuvent également favoriser l'alphabétisation des adultes.*** Comme les enfants de parents analphabètes sont susceptibles d'avoir des problèmes ou des retards d'apprentissage de la lecture, de nouveaux moyens permettant d'améliorer la performance et l'efficacité des programmes d'alphabétisation des adultes pourraient favoriser une acquisition plus précoce de la lecture.
- x. ***La responsabilité de l'impact de l'apprentissage doit être largement partagée.*** Les spécialistes de l'éducation, les responsables politiques, les participants aux tables rondes intergouvernementales de haut niveau, les ministres de l'Éducation, les chefs de communauté des villages ruraux, les enseignants et les parents devraient tous être tenus responsables de ce que les enfants apprennent et de la façon dont ils apprennent. Les évaluations SQC peuvent ouvrir de nouvelles voies en matière de responsabilité et d'appropriation locale des résultats.
- xi. ***Les évaluations hybrides peuvent renforcer sensiblement l'impact d'une politique.*** Les évaluations SQC permettent de mieux suivre l'apprentissage au fil du temps, sont plus faciles à adapter au contexte linguistique local, et peuvent être conçues de façon à mieux comprendre les enfants qui se situent au bas des échelles d'apprentissage types. Dans les années à venir, elles auront un rôle important à jouer dans les politiques de développement de l'éducation.

10. Conclusions. Une utilisation efficace des évaluations de l'éducation est fondamentale pour améliorer l'apprentissage. Toutefois, cette utilisation ne dépend pas seulement de paramètres techniques ou de méthodologies statistiques. *La différence aujourd'hui – dans le*

contexte actuel d'exigence généralisée d'éducation – réside dans la nécessité de donner une plus grande priorité aux évaluations à court terme qui impliquent diverses parties prenantes, sont sensibles à la dimension culturelle et ont un fort impact local. Les évaluations de l'apprentissage, qu'il s'agisse d'enquêtes à grande échelle, auprès des ménages ou hybrides (SQC), ne valent qu'en proportion des usages qui en sont faits. Il faut intensifier la recherche et le développement, notamment dans le domaine en plein essor des technologies de l'information et de la communication.

Pour résumer, les évaluations hybrides SQC ont la capacité de renforcer la responsabilité pédagogique, d'accroître la transparence et de susciter un engagement plus grand des acteurs impliqués dans l'amélioration de l'apprentissage. Mais rien de tout cela ne sera possible si les stratégies et les évaluations ne ciblent pas de façon durable et substantielle les populations pauvres et marginalisées. L'effort actuel pour diversifier les méthodes d'évaluation des acquis scolaires dans les pays en développement est un moyen très important pour rendre possibles des améliorations réelles et durables dans le domaine de l'éducation.

1. INTRODUCTION

Remarques préliminaires

Chercher à réaliser l'éducation pour tous (EPT), c'est fondamentalement chercher à garantir qu'enfants, jeunes et adultes acquièrent les connaissances et les compétences dont ils ont besoin pour mener une vie meilleure et jouer un rôle dans l'édification de sociétés plus pacifiques et plus équitables. C'est pourquoi il est impératif de se concentrer sur la qualité pour réaliser l'EPT. Alors que beaucoup de sociétés s'efforcent d'universaliser l'éducation de base, elles sont confrontées au défi redoutable d'offrir des conditions dans lesquelles chaque apprenant puisse réellement apprendre¹.

Dans un domaine aussi complexe que la scolarisation et l'éducation dans le monde, il est difficile de savoir comment interpréter les recherches censées expliquer l'éducation. Comment prendre en compte la multitude de variables et de techniques qui ont constitué la matière première des chercheurs spécialisés en éducation : participation des élèves, fonds investis, heures d'enseignement direct, motivation, compétences métalinguistiques, capacité de résolution de problèmes et processus mental de haut niveau ? Des milliers de travaux de recherche en éducation ont été menés sur ces sujets et d'autres questions connexes.

La présente étude cherche à répondre à la question suivante : *Les études existantes sur l'évaluation des apprentissages (en particulier l'apprentissage de la lecture) peuvent-elles aider à renforcer l'efficacité des efforts pour améliorer les résultats scolaires dans les pays en développement ?* La réponse est évidemment « oui », mais le chemin pour y parvenir, dans un domaine comme l'apprentissage (ou la lecture), n'est pas si facile. Le but du présent ouvrage est d'aider les acteurs de l'éducation à s'engager sur de nouvelles voies prometteuses.

Le titre de l'ouvrage, *Des évaluations simples, rapides et abordables*, renvoie à une étude antérieure, publiée en 2003², et qui portait essentiellement sur la question de savoir pourquoi les chercheurs semblaient souvent faire exactement le contraire de ce que préconise le titre du présent ouvrage

-
1. UNESCO, 2004. *Rapport mondial de suivi sur l'éducation pour tous*, 2005, p. v.
 2. Wagner, 2003. Cette étude reposait également en partie sur des recherches antérieures menées sur le terrain au Maroc et au Zimbabwe, Wagner (1990, 1993).

pour se lancer dans des études trop vastes, trop lentes et trop chères pour être pertinentes, à une époque où accéder aux connaissances en temps réel peut avoir des conséquences réelles. Cette étude soutenait en outre que des évaluations fondées sur les besoins et les exigences de pays industrialisés et disposant d'importantes ressources n'étaient pas tout à fait adaptées au contexte de pays en développement, où la situation d'apprentissage varie de façon considérable et perceptible. Pourquoi les chercheurs ne commenceraient-ils pas par s'interroger sur les besoins réels d'apprentissage des enfants défavorisés, dans les écoles pauvres, afin de concevoir des évaluations à partir de ce point de vue ?

Les choses ont heureusement changé depuis la publication de cette étude. Aujourd'hui, on en sait plus sur ce qui est nécessaire pour évaluer et développer l'éducation. Le présent ouvrage revisite ces critiques antérieures et tente de combler les manques par de nouveaux constats et de nouvelles orientations.

L'histoire d'Aminata, village de Kahalé

C'est le petit matin au village de Kahalé, à 45 kilomètres environ de la capitale. Il a encore plu et l'eau a ruisselé du toit en tôle ondulé de l'école à classe unique qui occupe le centre du village. À cause de la pluie, Monsieur Mamadou, le professeur, a du mal à rejoindre son école, ce lundi matin : son taxi rural s'embourbe à plusieurs reprises, obligeant les six autres passagers à aider le chauffeur à remettre le véhicule sur la route qui mène au village. Arrivé à l'école, Monsieur Mamadou attend l'arrivée des élèves. À 9 heures, la classe n'est remplie qu'à moitié, ce qui n'est probablement pas une mauvaise chose, puisque la classe entière compte 65 enfants alors qu'il n'y a que 50 places assises.

Environ 35 élèves sont maintenant arrivés. Ceux qui portent des sandales dignes de ce nom et des chemises propres et boutonnées sont assis aux deux premiers rangs ; ceux qui n'ont pas de sandales et des chemises un peu moins propres s'assoient derrière. Les enfants, tous en 2^e année, ont entre 7 et 11 ans. Tout d'abord, Monsieur Mamadou s'adresse aux enfants en wolof pour les accueillir, les faire taire et réclamer leur attention. Il se met ensuite à écrire un texte en français au tableau, en prenant son temps pour bien écrire. C'est important, car seuls quelques enfants (tous assis au premier rang) ont un manuel devant eux. L'écriture du texte dure 15 minutes environ, pendant lesquelles les enfants bavardent, regardent par la fenêtre ou ferment les yeux, la tête appuyée sur leur bureau. Certains sont déjà fatigués et ont faim, parce qu'ils n'ont rien pris d'autre le matin qu'une tasse

de thé chaud et un morceau de pain rassis ou de la purée. Quand Monsieur Mamadou a fini d'écrire, il se retourne et s'adresse en français à la classe : « Vous allez maintenant copier ce texte dans vos cahiers. » Les enfants se mettent au travail pendant que Monsieur Mamadou sort fumer une cigarette.

Aminata, 9 ans, est assise au troisième rang. Elle sort son crayon et commence à écrire dans son cahier, s'appliquant à reproduire chaque mot écrit au tableau. Elle est heureuse d'avoir pu venir à l'école, car elle aurait dû ce jour-là rester à la maison pour s'occuper de sa petite sœur. Heureusement, sa tante étant venue rendre visite à la famille, elle a tout de même pu aller à l'école. S'il est mieux d'aller à l'école que de rester à la maison, Aminata a toutefois l'impression de ne pas faire un très bon usage de son temps. Elle peut recopier le texte, mais ne comprend pas ce qu'il dit. Aminata ne sait lire que quelques mots de français sur les pancartes et les panneaux publicitaires du village. Alors, même si elle est la seule à être « scolarisée » dans sa famille, elle n'est pas d'une grande aide pour sa mère, quand celle-ci veut savoir exactement ce qui est écrit sur le flacon de médicaments qu'elle doit prendre. Cette situation met Aminata mal à l'aise et elle se demande pourquoi ses camarades de classe du premier rang semblent connaître déjà un peu de français. Elle se demande aussi pourquoi Monsieur Mamadou ne semble faire venir au tableau que les élèves du premier rang et pas elle. Elle a entendu dire qu'il y avait une autre école après l'école primaire, mais, apparemment, seuls les élèves du premier rang peuvent s'y inscrire. À quoi cela sert-il d'étudier et de rester à l'école ? se demande-t-elle.

Ce que signifie l'histoire d'Aminata

L'histoire de Monsieur Mamadou ou d'Aminata n'a rien d'extraordinaire. Cette scène décrit des faits qui ne sont que trop familiers et qui se répètent dans le monde entier³. S'il existe des classes dysfonctionnelles dans tous les pays, leurs conséquences sont encore pires lorsque les ressources affectées à l'enseignement sont tellement limitées, comme c'est le cas dans les pays les plus pauvres d'Afrique. Cette anecdote évoque la pauvreté, l'inefficacité des systèmes éducatifs et les communautés qui ne voient pas ce qui ne fonctionne pas chez elles.

Ce récit est en fait une histoire d'évaluation de l'apprentissage, de ce qu'il faut faire dans ce domaine. Il parle de non-apprentissage, de non-lecture

3. Pour une autre description détaillée dans le contexte du Kenya, voir Commeyras et Inyega (2007).

et d'échec scolaire naissant. La plupart des enfants dans la même situation qu'Aminata ne seront pas correctement évalués avant d'abandonner l'école. De nombreux enfants comme Aminata n'existeront pas du point de vue des statistiques nationales. Ils n'iront ni à l'école secondaire, ni à l'université et ne trouveront pas un emploi dans l'économie mondiale. Cette année, ou la prochaine, sera probablement la dernière année d'école pour Aminata. Elle se mariera sans doute vers l'âge de la puberté et reproduira avec ses propres enfants le même schéma de non-éducation. Ce n'est certes pas le destin de tous les enfants, mais c'est celui de la plupart de ceux qui vivent dans les régions pauvres des pays pauvres. Il faut prendre cette situation trop banale à bras-le-corps et changer les choses.

Le présent ouvrage place l'histoire d'Aminata au cœur de la problématique de l'éducation dans les pays en développement les plus pauvres. L'évaluation et le suivi peuvent être considérés comme de simples exercices statistiques, mais ce serait une grave erreur. Menés de façon sérieuse, l'évaluation et le suivi permettent d'exprimer les besoins éducatifs illustrés par l'histoire d'Aminata ; correctement menés, ils peuvent non seulement renforcer la responsabilité au sein des systèmes éducatifs, mais aussi suggérer des méthodes proactives pour remédier aux problèmes. Le lecteur ne doit pas perdre de vue l'histoire d'Aminata : elle est la *raison d'être* de la quête de meilleurs indicateurs d'apprentissage.

Structure de l'ouvrage

Après l'introduction, le *chapitre 2* aborde la question de la prise en compte des objectifs stratégiques, en particulier l'amélioration de la qualité de l'éducation, lorsqu'on s'intéresse aux acquis de l'apprentissage. Il évoque également certaines inquiétudes suscitées par la façon dont les évaluations sont – ou ne sont pas – utilisées à l'heure actuelle.

Le *chapitre 3* décrit les nombreuses façons dont les indicateurs d'apprentissage peuvent être utilisés. Il définit l'apprentissage et indique quels sont les intrants et les extrants de l'éducation susceptibles d'être mieux compris grâce aux indicateurs d'apprentissage. Il suggère également diverses options aux responsables politiques qui réfléchissent aux moyens d'améliorer la qualité de l'éducation.

Le *chapitre 4* explore les principaux types d'évaluation des acquis scolaires, en privilégiant ceux qui sont les plus en vigueur dans les pays en développement, notamment les évaluations à grande échelle, les enquêtes auprès des ménages et les nouvelles évaluations hybrides (comme l'EGRA). Il aborde également les questions de compétences et d'échantillonnage de

la population, de comparabilité et de crédibilité des évaluations, ainsi que divers autres aspects liés aux mesures de l'apprentissage.

Le *chapitre 5* est consacré à l'évaluation des compétences en lecture. Il commence par expliquer pourquoi les tests de lecture sont si importants, puis s'intéresse aux modèles d'acquisition de la lecture, à la lecture dans la langue maternelle et dans une seconde langue, au rôle des systèmes d'écriture et aux types de test actuellement en usage.

Le *chapitre 6* aborde les implications pédagogiques de divers tests de lecture, et explique pourquoi certains tests ne sont pas si bons que cela pour les enfants.

Le *chapitre 7* traite du sujet fondamental des coûts. Quels sont les coûts totaux, combien coûtent réellement différentes évaluations et sont-elles vraiment rentables ?

Le *chapitre 8* décrit les évaluations de l'alphabétisation des adultes, notamment les efforts pour effectuer des enquêtes auprès des ménages dans les pays en développement.

Le *chapitre 9* évoque les principales conclusions et recommandations, avec des sous-sections sur les systèmes d'évaluation axés sur la détection précoce des difficultés, sur le problème de l'évaluation en temps réel et des questions annexes.

Le *chapitre 10*, qui sert de conclusion, revisite l'histoire d'Aminata à la lumière des innovations actuelles en matière d'évaluation.

Limites

La présente étude s'intéresse principalement à la matrice de décision qu'utiliserait un responsable politique (ministre de l'Éducation ou autre décideur de haut niveau) pour choisir tel ou tel type d'évaluation, dans le but d'améliorer la qualité de l'éducation. Toutefois, pour choisir entre différentes options, il faut tenir compte du fait que le domaine de l'évaluation est en constante évolution. De nouveaux outils (et des données permettant de tester leur utilité) sont constamment développés, de sorte que les évaluations sont en permanence corrigées et adaptées.

C'est pourquoi, une *première* limite de cette étude est qu'elle est nécessairement sélective et destinée avant tout à donner au lecteur une idée de ce qu'est le domaine de l'évaluation de la qualité de l'éducation. Il ne s'agit pas d'une synthèse cumulative définitive de ce qu'il faudra faire demain. Une *deuxième* limite concerne la substance ou le *contenu* de ce

qui est évalué. Il est impossible de couvrir les nombreuses questions portant sur le contenu des programmes scolaires élaborés par diverses agences gouvernementales et non gouvernementales ; il n'est pas non plus possible d'affirmer avec certitude ce que la « qualité de l'éducation » devrait ou ne devrait pas être. La présente étude part plutôt du constat que la lecture et les performances en lecture figurent dans toutes les listes des éléments de base de la qualité de l'apprentissage, ce que ni la recherche ni la politique n'ont sérieusement remis en question. *Troisièmement*, il existe des limites liées au *contexte* : cette étude est parrainée par l'Initiative de mise en œuvre accélérée de l'éducation pour tous et par l'UNESCO, qui s'intéressent toutes deux en priorité aux pays les plus pauvres. Or, les études sur la lecture ont été dans une large mesure entreprises dans des pays à hauts revenus de l'Organisation de coopération et de développement économiques (OCDE) et dans des langues européennes. Dans quelle mesure les études sur ces pays et ces cultures s'appliquent-elles au contexte de pays en développement, pauvres et non membres de l'OCDE ? Cette question classique constitue aussi une limite importante. *Quatrièmement*, il existe une diversité (filles et garçons, langues, appartenances ethniques, etc.) à l'intérieur même des pays, qu'ils soient membres de l'OCDE ou en développement, de sorte qu'il est difficile de considérer que des contextes aussi divers puissent être parfaitement compris par des statistiques à l'échelle nationale. *Cinquièmement*, il existe des limites liées à la *discipline* : quel crédit peut-on accorder à des explications internationales, nationales ou locales, à des études de cas ou à des enquêtes de grande envergure, ou encore à des conclusions fondées sur la neuropsychologie, des motivations individuelles ou des facteurs socioculturels ? Ces questions, qui sont débattues au sein des diverses sciences sociales et entre elles, ont tourmenté les spécialistes de la lecture et de l'éducation pendant de nombreuses années.

Enfin, il y a des limites dues au seul fait d'essayer de formuler un ensemble clair de recommandations générales, valables pour les pays pauvres, alors que chaque contexte est différent à de nombreux niveaux. Pourtant, sans cette analyse, de bonnes idées pour améliorer l'éducation pour tous risquent de rester lettre morte.

But et prochaines étapes

Le présent ouvrage est motivé par les efforts actuels pour promouvoir l'utilisation d'indicateurs d'apprentissage de meilleure qualité dans l'éducation, dans le cadre de l'Éducation pour tous et des objectifs du Millénaire pour le développement des Nations Unies. Il a pour but de fournir un cadre général et un contexte scientifiques et rigoureux, permettant

à ceux qui s'efforcent actuellement de créer, d'appliquer et d'utiliser ces indicateurs afin d'élaborer des stratégies et, à terme, d'améliorer l'apprentissage. L'accent est plus particulièrement mis sur les contextes pauvres et défavorisés des pays en développement.

Cette étude s'efforce d'être neutre, en ce sens qu'elle ne se fonde pas sur des idées préconçues sur ce qui fait qu'une évaluation est nécessairement meilleure qu'une autre. En effet, il n'existe pas d'évaluation qui soit meilleure que les autres dans l'absolu. Il est souhaitable de disposer d'un ensemble d'évaluations susceptible de s'appliquer à divers aspects stratégiques et contextes nationaux, de sorte que le choix des instruments d'évaluation soit fonction de leur adéquation à des buts stratégiques spécifiques.

Le travail sur les évaluations de toutes sortes, y compris le renouvellement futur de presque toutes les évaluations décrites dans cet ouvrage, se poursuit. Les résultats de ces évaluations seront débattus et ces discussions enrichiront le domaine de la qualité de l'éducation. La base de connaissances sur l'évaluation ne cessant de s'étendre, la prochaine étape devrait être l'élargissement et l'approfondissement de l'utilisation des indicateurs qui s'imposeront comme l'une des principales voies pour améliorer l'apprentissage et la scolarité dans le monde entier.

2. ACQUIS D'APPRENTISSAGE ET BUTS STRATÉGIQUES

L'UNESCO promeut l'accès à une éducation de qualité comme un droit humain et appuie une approche fondée sur les droits de toutes les activités d'éducation ... L'influence de cette approche sur l'apprentissage se fait ressentir à deux niveaux. Au niveau de l'apprenant, l'éducation doit chercher à déterminer et prendre en compte les connaissances déjà acquises par l'apprenant, valider les modes formels et informels, pratiquer la non-discrimination et offrir un environnement d'apprentissage sûr et propice ; au niveau du système d'apprentissage, une structure de soutien est requise pour mettre en œuvre les politiques, adopter des lois, répartir les ressources et mesurer les résultats d'apprentissage, de manière à avoir le meilleur impact possible sur l'apprentissage pour tous⁴.

EPT et acquis d'apprentissage

[L'éducation fondamentale doit donc être axée sur] l'acquisition effective et les résultats de l'apprentissage, et non pas sur le seul fait de s'inscrire à une formation, de la suivre jusqu'à son terme et d'obtenir le certificat qui la sanctionne⁵.

Beaucoup de spécialistes de l'éducation estiment que la Conférence mondiale sur l'éducation pour tous (Jomtien, Thaïlande, 1990) a marqué un tournant dans l'éducation et le développement à l'échelle internationale. Deux thèmes majeurs de cette réunion ont été particulièrement importants : premièrement, l'accent mis sur l'éducation des enfants (et des adultes) dans les pays pauvres, dans le cadre de plusieurs objectifs éducatifs ; deuxièmement, un effort transversal pour promouvoir la *qualité de l'apprentissage* dans l'éducation, en ne se contentant pas seulement de recenser les enfants scolarisés et ceux qui ne l'étaient pas. En 2000, lors d'une conférence sur l'Éducation pour tous (Dakar, Sénégal), l'accent a été mis sur ces deux enjeux dans une liste plus détaillée de six objectifs

4. UNESCO. 2004. *Rapport mondial de suivi sur l'EPT*, 2005, p. 32.

5. UNESCO. 1990. *Déclaration mondiale sur l'EPT*, article 4, p. 5.

concernant l'éducation⁶. Ils l'ont été de nouveau dans les objectifs du Millénaire pour le développement devant être concrétisés à l'horizon 2015⁷.

Ces objectifs et thèmes adoptés, les organisateurs de la conférence ont réalisé qu'il faudrait de meilleures méthodes pour mesurer les résultats d'apprentissage, en particulier dans le contexte des pays en développement les plus pauvres. On pensait qu'avec de meilleures méthodes d'évaluation et une plus grande capacité de collecte et d'analyse de données, il serait possible de répondre au besoin accru de données crédibles sur les résultats d'apprentissage d'un point de vue véritablement mondial. Dans les années qui ont suivi les conférences de Jomtien et de Dakar, diverses initiatives ont été prises pour que de nouvelles ressources substantielles soient consacrées aux compétences scolaires et à leur mesure⁸.

Or, la qualité de l'éducation n'est pas seulement une affaire d'engagement politique international, de financement suffisant, d'expertise technique et de ressources humaines. Il faut avant tout faire des choix importants concernant les informations (les données) à rechercher et à prendre en compte, ainsi que les acteurs concernés, lesquels peuvent se poser ce genre de questions :

- Au niveau international, un bailleur de fonds peut se demander : Comment pourrions-nous (la communauté internationale ou celle des donateurs) mieux apprécier la situation actuelle de l'apprentissage dans les différents pays ? Qui plus est, quels pays devrions-nous comparer ? Ou quel type d'apprentissage est suffisamment commun aux différents pays pour permettre une comparaison « impartiale » ?
- Au niveau national (du pays), un ministre de l'Éducation peut se demander : Comment améliorer l'apport de talents aux différents niveaux d'éducation ? Comment veiller à ce que tous les élèves atteignent un niveau minimal d'éducation, et permettre aux plus doués d'aller le plus loin possible dans le système éducatif ? Comment aider notre système à être plus performant ?

6. Les six objectifs du Cadre d'action de Dakar pour l'Éducation pour tous étaient les suivants : protection et éducation de la petite enfance (PEPE) ; enseignement primaire universel (EPU) ; répondre aux besoins éducatifs de tous les jeunes ; améliorer les niveaux d'alphabétisation des adultes ; éliminer les disparités entre les sexes ; améliorer la qualité de la mesure des résultats d'apprentissage. (Adapté de : UNESCO, 2004, p. 30).

7. Nations Unies (2000).

8. Notamment le projet de l'UNESCO relatif au Suivi permanent des acquis scolaires, la création de l'Institut de statistique de l'UNESCO et diverses évaluations internationales et régionales qui constituent le thème principal du présent ouvrage.

- Au niveau de l'apprenant (de l'individu), un élève peut se demander : Que va m'apporter le fait d'aller à l'école ou de suivre un programme d'enseignement non formel ? Que signifie pour moi le fait d'obtenir un certificat ou un diplôme ? Beaucoup de jeunes de mon âge ont des diplômes et pas d'emploi. À quoi sert réellement cette éducation ?

Ces interrogations varient non seulement en fonction des acteurs concernés, mais aussi selon le pays, le sexe, le groupe ethnique ou linguistique, et au sein des différentes régions d'un pays ou entre plusieurs pays. Cette variation commence à révéler les inégalités qui existent (et, surtout, dont les parties prenantes *pensent* qu'elles existent) entre les différents groupes. En d'autres termes, l'évaluation des apprentissages commence à permettre d'élaborer des stratégies susceptibles d'améliorer la qualité de l'éducation et de favoriser son évolution.

La promesse d'une éducation de meilleure qualité

[L]e niveau des capacités cognitives est une composante cruciale des perspectives de croissance à long terme. Ce qui a manqué, c'est de mettre l'accent sur la qualité plutôt que sur la quantité de l'éducation : faire en sorte que les élèves apprennent réellement ... Il est important de savoir qu'aller à l'école n'a de répercussions sur les résultats économiques que dans la mesure où cela permet vraiment aux élèves d'enrichir leurs acquis. Les résultats scolaires n'ont pas de lien significatif avec la croissance économique, même en prenant en considération les capacités cognitives⁹.

La *qualité* de l'éducation, thème du *Rapport mondial de suivi sur l'EPT 2005*, a à peu près autant de significations différentes qu'il y a de spécialistes et décideurs politiques ayant écrit sur le sujet. Néanmoins, un consensus semble se dégager sur plusieurs composantes majeures, notamment :

- **Ce que** les apprenants doivent savoir : les objectifs de tout système éducatif, perceptibles dans la formulation des missions et des valeurs, et développés dans les programmes scolaires et les niveaux de performance.
- Le **lieu** où a lieu l'apprentissage : le contexte dans lequel se déroule l'enseignement (la taille des classes, le degré de salubrité et de sécurité de l'environnement d'apprentissage, la disponibilité de ressources et d'installations pour faciliter l'apprentissage, telles que des salles de classe, des livres ou du matériel pédagogique).

9. Hanushek et Woessmann, 2009a. Voir également Hanushek et Woessmann, 2009b.

- **Comment** se déroule l'apprentissage : les caractéristiques des échanges entre l'élève et l'enseignant (tels que le rôle des élèves dans leur apprentissage, l'attitude des enseignants et des apprenants à l'égard de l'apprentissage, ainsi que d'autres pratiques pédagogiques).
- **Ce qui** est réellement acquis : les résultats de l'éducation (connaissances, savoir-faire, compétences, attitudes et valeurs que les apprenants acquièrent)¹⁰.

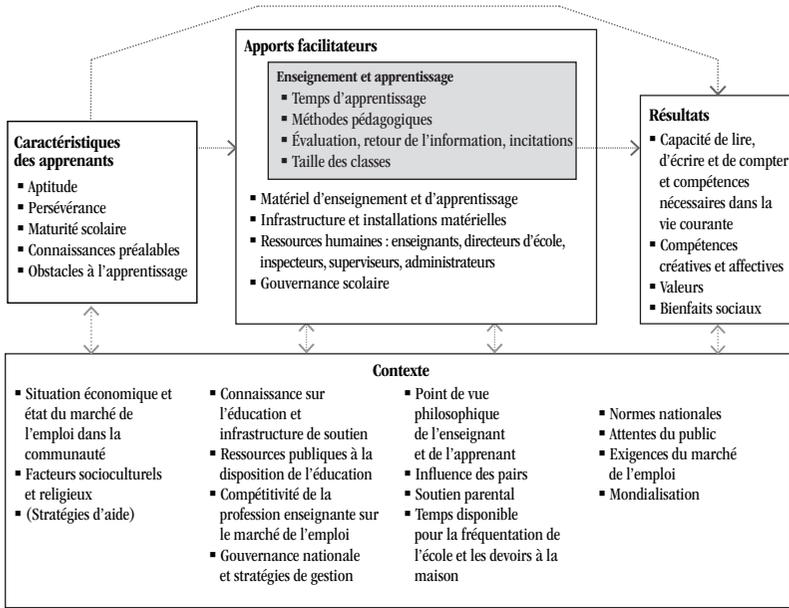
Une autre manière d'aborder la qualité de l'éducation consiste à utiliser des modèles d'intrants-extrants, comme celui illustré par la *figure 2.1*, extraite du *Rapport mondial de suivi* de l'UNESCO sur la qualité. Ce modèle prend en compte plusieurs caractéristiques clés de l'apprenant, et plus particulièrement ce qu'un enfant a appris à la maison avant d'être scolarisé. L'école fournit un ensemble d'intrants, dont du temps, des méthodes d'enseignement, les réactions des enseignants, du matériel pédagogique, etc. Ce processus peut notamment permettre à l'apprenant d'acquérir un ensemble de capacités cognitives (lire et écrire, par exemple), d'attitudes et de valeurs sociales. Ce modèle montre qu'il est important de mesurer divers résultats, mais n'indique pas quelles variables contextuelles intermédiaires déterminent quels résultats, ni comment les mesurer.

Comprendre les processus et tous leurs éléments permet de les améliorer. Par exemple, examinons le rôle du niveau d'éducation de la mère dans la réussite scolaire de ses enfants. Beaucoup de spécialistes soutiennent que l'éducation maternelle est l'un des déterminants les plus puissants de la scolarisation durable des enfants et de leur réussite scolaire (*figure 2.2*)¹¹. Cependant, comment un modèle de qualité de l'éducation agit-il systématiquement ? Comment une mère transmet-elle effectivement des savoir-faire, des attitudes et des valeurs à ses enfants, malgré son faible niveau d'éducation ? Comme on le verra plus en détail dans le chapitre suivant, de nouvelles recherches permettent de commencer à répondre à ces questions.

10. Adapté de Braun et Kanjee, 2006, p. 5. Outre la qualité, leur cadre aborde également les questions d'accès, d'équité et d'efficacité.

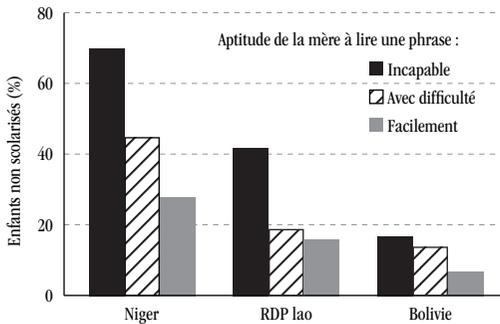
11. Voir Summers (1992) pour une prise de position importante de la Banque mondiale sur le genre et l'éducation.

Figure 2.1 Comprendre la qualité de l'éducation



Adapté de : UNESCO, 2004, p.39.

Figure 2.2 Alphabétisme des mères et scolarisation des enfants au Niger, en RDP lao et en Bolivie, 2000



Source: Calculs fondés sur la base de données MICS de l'UNICEF.

Adapté de : UNESCO, 2004, p. 145.

Une troisième façon d'étudier la promesse d'une éducation de meilleure qualité consiste à s'intéresser aux liens entre les acquis scolaires et le développement économique. De nombreuses études ont montré comment la façon de mesurer le retour sur investissement des investissements dans la scolarité (mesurés *via* les compétences de base acquises) peut être appliquée dans les pays en développement (*tableau 2.1*). Des agences internationales et gouvernementales utilisent cette méthode pour rationaliser les améliorations quantitatives et qualitatives de l'éducation.

Tableau 2.1 Impact des compétences de base sur le revenu

Étude	Pays	Effet estimé ¹	Remarques
Glewwe (1996)	Ghana	0,21** à 0,3** (public) 0,14 à 0,17 (privé)	D'autres méthodes d'estimation donnent des résultats quelque peu différents ; les effets des mathématiques sont en général plus importants que ceux de la lecture ; tous ces effets persistent même en prenant en compte les résultats au test d'aptitude de Raven.
Jolliffe (1998)	Ghana	0,05 à 0,07*	Le revenu des ménages est lié au score moyen en mathématiques, avec des variations relativement faibles selon la méthode d'estimation employée ; l'effet du revenu hors agriculture sur le revenu dans l'agriculture n'est pas lié aux compétences.
Vijverberg (1999)	Ghana	Incertain	Estimations du revenu du travail indépendant hors agriculture selon les niveaux de mathématiques et de lecture ; estimations très variables (effets aussi bien positifs que négatifs) mais, en général, effets statistiquement non significatifs.
Boissiere, Knight et Sabot (1985) ; Knight et Sabot (1990)	Kenya	0,19** à 0,22**	Estimations sur l'échantillon total : variations faibles entre les élèves sortant de l'école primaire et ceux sortant de l'école secondaire.
Angrist et Lavy (1997)	Maroc	Incertain	Conversion en scores standardisés impossible à cause de l'utilisation d'indices de performance ; le facteur « compétences en français écrit » semble être celui qui influence le plus les revenus mais les résultats dépendent de la méthode d'estimation utilisée.
Alderman <i>et al.</i> (1996)	Pakistan	0,12 à 0,28*	On constate des variations selon les méthodes employées et selon que les aptitudes et la santé ont été prises en compte ou non ; les effets sont plus importants et significatifs lorsqu'on ne tient pas compte des aptitudes et de la santé.

Étude	Pays	Effet estimé ¹	Remarques
Behrman, Ross et Sabot (à venir)	Pakistan	Incertain	Estimations d'un modèle structurel avec des scores combinés pour les compétences cognitives ; valeur de l'indicateur significative à 0,01 mais ne peut pas être traduite directement en taille estimée de l'effet.
Moll (1998)	Afrique du Sud	0,34** à 0,48**	Dépend de la méthode d'estimation, impact variable du calcul ; compréhension (non indiquée) généralement non significative.
Boissiere, Knight et Sabot (1985) ; Knight et Sabot (1990)	République-Unie de Tanzanie	0,07 à 0,13*	Estimations sur l'échantillon total : plus basses pour les élèves sortant de l'école primaire que pour ceux sortant de l'école secondaire.

Notes : * significatif à 0,05 ; ** significatif à 0,01.

1. Les estimations indiquent l'augmentation proportionnelle des revenus correspondant à une augmentation de 1 écart type des scores aux tests.

Source : Hanushek, 2004.

Adapté de : UNESCO, 2004, p. 46.

Enfin, les conséquences d'une meilleure qualité de l'éducation peuvent être considérées comme étant clairement un droit au développement personnel¹². Des enseignants qualifiés, des programmes scolaires et des manuels bien conçus, le soutien parental et des communautés motivées sont autant de facteurs qui peuvent avoir, et qui ont effectivement, un impact sur l'apprentissage des enfants. Le désir d'améliorer la qualité de l'apprentissage et la qualité globale de l'éducation est indubitable¹³. Ce qui est beaucoup moins évident, c'est comment se mettre d'accord sur la manière de déterminer *de façon empirique* ce qu'est la qualité, puis décider ce qu'il faut mettre en œuvre pour renforcer et améliorer la qualité. De meilleurs instruments de mesure jouent un rôle important dans ce processus.

12. Sen, 1999.

13. Voir l'important *Rapport mondial de suivi* consacré à la qualité (UNESCO, 2004) pour un point de vue politique sur cette question. Néanmoins, il faut également reconnaître que certains décideurs semblent parfois se soucier davantage des chiffres (résultats quantitatifs) que des résultats qualitatifs, plus difficiles à mesurer. De même, il ne serait pas surprenant de trouver des enseignants, des chefs d'établissement et des fonctionnaires tout à fait satisfaits du statu quo. Dépasser cette complaisance est, en grande partie, ce que les efforts actuels en faveur des approches SQC tentent de faire.

L'importance des mesures

La mesure des données éducatives et les variations culturelles sont deux mondes qui s'interpénètrent de manière souvent prévisible, mais également difficile à appréhender. Ce n'est pas seulement une affaire de comparabilité internationale. Disons plutôt que les variations culturelles sont endémiques dans tout contexte où on élève des enfants. Il peut également y avoir des variations importantes au sein de chaque ménage, en particulier si on considère les manières très différentes dont garçons et filles sont parfois traités dans de nombreuses cultures.

Si les mesures (y compris tous les tests et les évaluations) sont si difficiles et peuvent être remises en cause pour tant de raisons, comment peut-on les mettre en œuvre dans autant de pays et de manières aussi différentes ? La réponse est « avec précaution ». Les mesures font l'objet de nombreuses critiques, mais elles sont peut-être le meilleur moyen à notre disposition pour aborder des problèmes complexes, à partir d'une base de connaissances qui peuvent être comprises et débattues par des groupes ayant parfois des idées extrêmement différentes sur ce qui est le mieux pour les enfants.

Interrogations suscitées par les évaluations

Pour certains, l'évaluation est un moyen impartial et objectif de fixer et maintenir des normes, d'impulser une réforme tant au niveau politique que dans la pratique, et de poser les bases d'une redevabilité satisfaisante. Pour d'autres, c'est un instrument qui sert à maintenir le statu quo, totalement tendancieux et improductif sur le plan éducatif¹⁴. L'inaction face aux inégalités, à la stigmatisation et à la discrimination liées au revenu, au sexe, à l'origine ethnique, à la langue, au lieu de résidence et au handicap retarde les progrès vers l'éducation pour tous¹⁵.

Les évaluations dans le domaine de l'éducation n'ont jamais fait l'unanimité et restent, aujourd'hui encore, controversées. Chaque fois qu'une évaluation de l'éducation est mentionnée dans les médias, les critiques en remettent souvent les résultats en cause, en opposant des éléments contraires ou en prétendant qu'elle est entachée d'erreurs pour diverses raisons techniques. Ainsi, quand on a annoncé que les adultes français avaient de moins bons scores que les adultes des autres pays européens ayant participé à l'Enquête

14. Braun et Kanjee, 2006, p. 2.

15. UNESCO. 2010. *Rapport mondial de suivi sur l'EPT 2010*, p. 2.

internationale sur l'alphabétisation des adultes (EIAA ; voir le *chapitre 8*), les fonctionnaires français se sont retirés de l'étude en prétextant des défauts techniques. Des cas similaires se produisent dans presque tous les pays, quand les résultats sont mauvais. Bien sûr, cette attitude que l'on peut qualifier de « politique défensive » est l'autre face de la « sensibilité politique », et montre tout simplement que la mesure peut être une source importante de changement. Pourtant, comme dans la citation ci-dessus, certains considèrent les évaluations non comme un instrument de changement, mais plutôt comme un outil de renforcement du statu quo.

Les statistiques représentent un autre aspect sensible de l'évaluation. Les sciences statistiques appliquées à l'évaluation des compétences humaines ont une longue et riche histoire. En ce qui concerne les efforts actuels en faveur de l'éducation, de nombreuses techniques complexes ont été élaborées visant à apporter des corrections ou des ajustements, en fonction des différents types de population, du nombre de questions aux tests, de l'importance relative des différences entre groupes, etc. Mais il n'y a pas de consensus scientifique pour le choix des méthodes statistiques : le débat est animé entre spécialistes. Il est important de garder à l'esprit que, si certaines méthodologies ont été soumises à des tests préalables rigoureux (cas des évaluations internationales), d'autres évaluations à échelle plus modeste commencent seulement à le faire. En outre, scientifiquement parlant, les premières ne sont pas nécessairement meilleures que les secondes. La rigueur scientifique des évaluations et le degré de confiance que leur accordent le public et les spécialistes doivent être maintenus quel que soit le type d'évaluation choisi.

En fin de compte, l'évaluation est surtout ce que l'on en fait. Les évaluations ne valent que par leur qualité technique au regard de la population concernée (thème principal de la présente étude). Elles peuvent finir à la poubelle ou faire les gros titres des journaux. Aujourd'hui encore, elles sont le plus souvent utilisées par les rares spécialistes qui les comprennent le mieux, ainsi que par les premiers à trouver des ressources pour les mettre en pratique. Ainsi, l'un des principaux problèmes en matière d'évaluation est leur efficacité à l'usage, ce qui suppose de contribuer à rendre les décideurs, à différents niveaux, responsables de la qualité de l'éducation. Si aucun rapport sur les aspects techniques des évaluations des acquis scolaires ne peut garantir le bon usage de ces dernières, il appartient néanmoins aux concepteurs des évaluations d'aider les intéressés à en faire une utilisation efficace.

Enfin, il est important de considérer la question des inégalités ou des iniquités, évoquée dans la seconde citation ci-dessus. *Le Rapport mondial de suivi de l'EPT 2010 : Atteindre les marginalisés*, montre clairement les sérieux problèmes que pose la réalisation des objectifs éducatifs en faveur des populations les plus pauvres du monde. Il y a indéniablement des raisons politiques à cela, mais il existe également des causes techniques, dont certaines peuvent être résolues par les types d'évaluations évoqués dans le présent ouvrage.

3. CE QUE LES INDICATEURS D'APPRENTISSAGE PEUVENT CHANGER

Utilisations des indicateurs d'apprentissage

Plusieurs utilisations peuvent être faites des indicateurs d'apprentissage (et éducatifs), notamment¹⁶ :

- Guider la politique. Dans tous les pays, les ministères de l'Éducation dépensent une part conséquente du budget national. Pour les décideurs, les indicateurs sont un moyen important de déterminer si ces fonds sont utilement employés.
- Vérifier les critères et en créer de nouveaux. La plupart des pays incluent un ensemble d'objectifs ou de buts éducatifs dans les programmes scolaires. Souvent, ces buts reposent sur et sont contrôlés par des indicateurs d'apprentissage. Les indicateurs constituent une base importante pour les systèmes nationaux soucieux de changer les critères et les programmes scolaires.
- Identifier les corrélats de l'apprentissage. Quels sont les causes et les effets de l'apprentissage en classe ? Dans quelle mesure certains groupes (selon le sexe, la langue ou la région) parviennent-ils à maîtriser un programme scolaire spécifique ? Les indicateurs sont essentiels pour déterminer les niveaux de performance et pour comprendre le lien entre les facteurs clés.
- Promouvoir la responsabilité. Quels facteurs influencent l'évolution de l'éducation dans un pays, une communauté, une école, chez un enseignant, un parent ou un élève ? De nombreux acteurs (séparément ou collectivement) sont responsables des compétences scolaires.
- Sensibiliser davantage l'opinion publique. Comment amener les parents et les communautés à soutenir davantage l'éducation ? Dans la mesure où les indicateurs peuvent être compris par le grand public et diffusés par les médias, la mesure de l'apprentissage est un

16. Cette liste est pour l'essentiel adaptée de Greaney et Kellaghan, 1996. Bien sûr, l'usage qui est effectivement fait des indicateurs pour changer la politique éducative varie considérablement dans le monde (voir Kellaghan *et al.*, 2009, chapitre 1 ; également Abadzi, communication personnelle).

moyen de faire prendre conscience des résultats à ces consommateurs potentiels¹⁷.

- Éclairer le débat politique. L'éducation est nécessairement politique. Comme pour la responsabilité et les discussions en matière de stratégies, le fait de pouvoir s'appuyer sur des indicateurs et des résultats d'apprentissage permet une discussion plus rationnelle sur les résultats empiriques de toute intervention dans le domaine de l'éducation. Les indicateurs d'apprentissage peuvent jouer et jouent (quand ils existent) un rôle déterminant dans ces débats. Ils permettent également de commencer à déterminer qui peut être responsable de l'amélioration de l'apprentissage.

Définir et mesurer l'apprentissage

La façon dont les connaissances, les compétences et les valeurs sont transmises est une partie du programme scolaire aussi importante que ce qui est appris, car ce processus fait en réalité partie de « ce qui est appris »¹⁸.

L'apprentissage est souvent décrit comme l'activité la plus essentielle de l'être humain. Les recherches actuelles ont prouvé qu'une part importante de l'apprentissage commence dès la naissance (voire avant) et se poursuit tout au long de l'existence. En raison de ses multiples aspects, l'apprentissage se prête à de nombreuses définitions, selon les disciplines concernées. Pour les psychologues, par exemple, l'apprentissage c'est « tout changement mesurable du comportement », alors que les anthropologues le définissent comme une enculturation dans laquelle l'enfant est socialisé par d'autres individus qui lui inculquent les valeurs et comportements requis par la culture.

Dans les discussions sur l'apprentissage, les résultats des tests servent à mesurer la qualité de l'éducation. L'utilisation d'indicateurs d'apprentissage peut fournir des informations pertinentes à ceux qui veulent savoir dans quelle mesure les matières qui composent les programmes scolaires sont comprises comme étant un processus (comme dans la citation ci-dessus), une mesure formative pour les politiques d'enseignement et d'apprentissage, et un marqueur du degré de réussite des apprenants aux principaux niveaux

17. En fait, cette dimension est souvent sous-estimée, en particulier dans les pays en développement où l'éducation est la plupart du temps sous le contrôle de l'État. Les chances d'améliorer l'éducation semblent augmenter en fonction du degré d'appropriation par les parents et la communauté.

18. Pigozzi, 2006, p. 45.

de sortie du système scolaire. Ce dernier type d'évaluation sommative peut être une évaluation critérielle ou normative¹⁹ et être utilisé pour faciliter (et légitimer) l'accès aux hiérarchies sociales et économiques. Ainsi, les tests peuvent aider à vérifier si les programmes scolaires sont respectés et appris, mais ils peuvent aussi avoir des effets délétères, quand ils augmentent la pression pour réussir, ce qui se traduit par une importance excessive accordée à la réussite aux examens.

Apprendre à l'école et en dehors de l'école

Types d'intrants

Toutes choses égales par ailleurs, la réussite de l'enseignement et de l'apprentissage a des chances d'être fortement influencée par les ressources disponibles pour soutenir le processus et par les modes de gestion de ces ressources. Il est évident que des écoles sans enseignants, sans manuels ou sans matériels d'apprentissage ne sont pas en mesure de bien faire leur travail. En ce sens, les ressources sont importantes du point de vue de la qualité de l'éducation – bien qu'on n'ait pas encore déterminé comment et dans quelle mesure²⁰.

Dans la citation ci-dessus, les ressources dont il est question émanent du système scolaire proprement dit. Or, d'innombrables études dans le domaine des sciences sociales montrent qu'une grande partie (voire la majeure partie) de la variance statistique associée à la réussite ou à l'échec scolaire résulte de facteurs extérieurs à l'école, pour ne pas dire totalement étrangers²¹. Naturellement, comme l'impliquent les définitions psychologique et anthropologique de l'apprentissage mentionnées au début de ce chapitre, l'enfant apporte à l'école quantité d'expériences qui concernent non seulement des faits appris sur sa vie et sa communauté, mais aussi des attitudes et des valeurs, des structures d'appui qui englobent la langue, des processus culturels et plus encore. Pour certains enfants, ces intrants sont parfois reconnus quand ils entrent enfin à l'école primaire (comme la langue d'enseignement, si c'est celle qui est parlée à la maison).

19. Les évaluations critérielles (ou fondées sur des critères) sont celles qui permettent de juger des résultats par rapport à un certain niveau escompté. Les évaluations normatives sont celles qui donnent un niveau de compétence que l'on compare à la position relative des apprenants par rapport à leurs pairs ayant passé le même test.

20. UNESCO. 2004. *Rapport mondial de suivi de l'EPT 2005*, p. 40.

21. Beaucoup d'études se sont évidemment intéressées au rôle du statut socio-économique et aux facteurs internes aux écoles (manuels, formation des enseignants, gestion, utilisation des ressources, etc.) pour expliquer les résultats scolaires. Voir, par exemple, Heyneman et Loxley (1983), ainsi qu'une étude plus récente de Gamaron et Long (2006).

Le fait d'en apprendre plus sur la vie des enfants dans leur milieu familial permet de mieux comprendre une multitude de types d'intrants, ainsi que les problèmes d'adéquation entre l'école et les enfants. Il n'y a là rien de nouveau. Depuis que les écoles ont été inventées par les institutions religieuses, au fil des siècles, l'idée a été de modeler intentionnellement ce que les enfants apportent à l'école pour obtenir une sorte d'uniformité de connaissances. Parce que les problèmes d'éducation de masse et d'équité étaient moins prégnants à l'époque, les cas connus de fréquente inadéquation, de résistance et d'abandon n'étaient pas aussi importants.

Dans le monde actuel, où les OMD essaient de garantir l'éducation de base universelle, on ne peut plus ignorer les caractéristiques personnelles, sociales et ethnolinguistiques que les enfants apportent en classe. En outre, il est de plus en plus admis que pour toucher les populations les plus difficiles à atteindre (parfois dites « marginalisées ») et atteindre les objectifs de l'EPT, il faut leur accorder une attention particulière et mobiliser des moyens financiers dédiés²². S'il existe de nombreux exemples de ce genre d'intrants extérieurs à l'école, le degré d'alphabétisation de l'environnement familial de l'enfant est une caractéristique souvent citée. La *figure 3.1* montre que, selon l'évaluation internationale PIRLS²³, l'aptitude à lire des élèves de 4^e année est fortement liée au degré d'alphabétisation de leur famille, mais varie toutefois fortement d'un pays à l'autre.

Le contexte scolaire et les possibilités d'apprendre

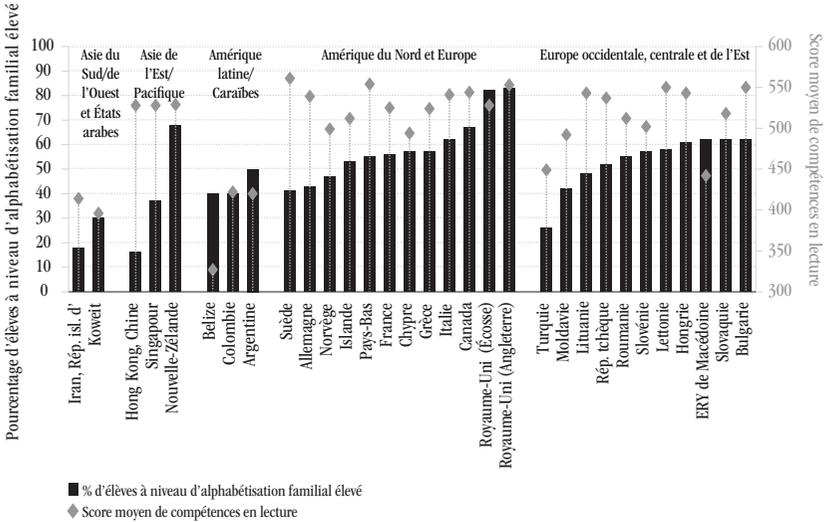
Si, au niveau international, le temps d'enseignement moyen prévu est d'environ 800 heures par an, avec peu de variation selon les régions, la durée de la scolarité obligatoire ou le niveau du revenu national, le nombre réel d'heures d'instruction dispensées peut varier considérablement. Les écoles peuvent être fermées pour des congés, des élections ou divers événements particuliers nationaux ou locaux non planifiés. ... Pour ces raisons et d'autres encore, le nombre réel d'heures d'enseignement peut être inférieur à 400 par an. Le temps d'apprentissage a été rarement étudié en profondeur dans les pays en développement, mais de nombreuses données officielles témoignent du grand gaspillage de temps²⁴.

22. Voir le *Rapport mondial de suivi* intitulé *Atteindre les marginalisés*, UNESCO, 2010.

23. Programme international de recherche en lecture scolaire.

24. Lockheed, 2004, p. 5.

Figure 3.1 Contexte d'alphabétisation et compétences en lecture dans l'enquête PIRLS, 2001



Note : L'indice d'activités familiales précoces d'alphabétisation utilisé dans le PIRLS est construit à partir des déclarations des parents concernant six activités : lire des livres, raconter des histoires, chanter des chansons, jouer à des jeux d'alphabet, jouer à des jeux de mots et lire à voix haute des enseignes et des étiquettes. Adapté de : UNESCO, 2004, p. 208.

Les écoles varient considérablement d'un pays à l'autre, d'une région à l'autre, et même d'une école à l'autre, y compris entre villages voisins. Cette distinction explique clairement pourquoi les résultats scolaires peuvent autant varier d'un enfant à l'autre et d'une école à l'autre.

Cette variation contextuelle de la scolarité peut être envisagée non seulement sous l'angle du temps d'enseignement (comme dans la citation ci-dessus), mais aussi par rapport aux possibilités d'apprendre. Il est de notoriété publique que le nombre réel d'heures d'enseignement est bien souvent inférieur à ce qui était prévu (voir dans le *tableau 3.1* les indicateurs du nombre d'heures d'enseignement au niveau mondial). En revanche, une étude de terrain récente, qui a utilisé la méthodologie EGRA, a constaté d'énormes pertes de possibilités d'apprendre dans un village rural, non seulement à cause des heures d'enseignement perdues (les écoles publiques étaient inopérantes pendant près de 25 % de l'année scolaire), mais aussi parce que les enseignants étaient occupés à faire autre chose (c'est-à-dire pas en train de travailler directement avec

les élèves) pendant plus de la moitié du temps²⁵. De ce fait, ainsi que l'a constaté l'étude, plus d'un tiers des élèves de 3^e année étaient incapables de lire le moindre mot. De même, dans le domaine de l'exposition à la langue, en dépit des politiques nationales, l'étude a constaté une grande variabilité dans l'utilisation effective de la langue d'enseignement par les enseignants en classe, d'où des différences importantes dans la maîtrise de la langue par les enfants selon les régions et l'enseignant²⁶. Ces résultats dramatiques ont suscité un regain d'attention pour la qualité de l'apprentissage précoce dans les pays les moins avancés (PMA).

Tableau 3.1 Temps annuel moyen d'instruction, par région et par année d'études, en heures, 2000

Régions EPT	1 ^{re} année	2 ^e année	3 ^e année	4 ^e année	5 ^e année	6 ^e année	7 ^e année	8 ^e année	9 ^e année	Nombre de pays
Afrique subsaharienne	755	775	812	847	872	871	951	946	965	16-18
États arabes	725	732	752	792	813	820	862	868	880	17
Asie centrale	533	575	620	647	740	754	798	812	830	9
Asie de l'Est et Pacifique	704	710	764	784	814	826	911	918	918	14
Asie du Sud et de l'Ouest	646	646	730	769	771	856	885	890	907	7-5
Amérique latine et Caraïbes	761	764	781	783	792	796	921	928	943	17-18

25. DeStefano et Elaheebocus (2009, p. 22) rapportent : « [N]otre étude indique que la majorité du temps disponible pour l'enseignement dans ces écoles [dans les zones rurales d'Éthiopie] est gaspillé. Des journées sont perdues quand l'école est fermée, quand les enseignants sont absents et quand les élèves sont absents. Toutefois, ces facteurs sont dérisoires au regard de la perte d'opportunité due au fait que les enseignants et les élèves sont trop souvent occupés à faire autre chose en 1^{re}, 2^e et 3^e années. Pendant 89 % du temps, les élèves ne sont pas occupés à apprendre, généralement parce que leur enseignant fait autre chose. » Ils indiquent également que « les élèves ayant indiqué avoir manqué l'école la semaine précédente avaient des taux de fluidité de lecture inférieurs de moitié à ceux des élèves ayant déclaré n'avoir pas manqué l'école. ... Le taux de présence indiqué par les élèves explique à lui seul 35 % de la variation du degré de fluidité moyen de lecture d'une école » (p. 13).

26. Voir Muthwii (2004), au Kenya et en Ouganda ; voir également Commeyras et Inyega (2007). Une étude de terrain récente comparant l'emploi de la langue d'enseignement au Kenya et en Ouganda a constaté d'importantes différences dans le respect effectif par les enseignants de la politique nationale en matière de langue d'enseignement, les enseignants ougandais attachant beaucoup plus d'attention que les enseignants kényans à l'utilisation de la langue maternelle en classe (Piper et Miksec, sous presse).

Régions EPT	1 ^{re} année	2 ^e année	3 ^e année	4 ^e année	5 ^e année	6 ^e année	7 ^e année	8 ^e année	9 ^e année	Nombre de pays
Amérique du Nord et Europe occidentale	743	748	790	799	845	847	894	906	933	23
Europe centrale et orientale	549	597	624	658	734	773	811	830	855	20
Total	689	705	742	766	804	819	883	891	908	122-125

Source : Benavot, 2004a.

Adapté de : UNESCO, 2004, p.171.

Les extrants de l'apprentissage

Il est possible d'utiliser d'autres mesures d'approximation des acquis d'apprentissage et des gains sociaux ou économiques d'ordre plus général, par exemple la réussite sur le marché de l'emploi. Il est utile de distinguer entre les acquis, la réussite scolaire et les autres mesures des résultats, qui peuvent inclure les bienfaits de l'éducation au sens large.²⁷

[S]i les acquis cognitifs des élèves sont acceptés comme critère légitime de la qualité de la scolarité, est-il raisonnable de fonder l'évaluation de cette qualité (et une possible attribution de responsabilité) sur une seule mesure de la performance des élèves au niveau d'une ou deux années ?²⁸

Si apprendre est l'activité essentielle de l'être humain, alors la scolarité semble être la meilleure façon de l'accomplir aux yeux de la plupart des nations. On demande aux écoles de résoudre de nombreux problèmes sociétaux, depuis la prise en charge des enfants quand les parents ne sont pas disponibles jusqu'au développement des compétences nécessaires à la croissance économique. Elles sont (de même que les programmes scolaires) la réponse nationale la plus répandue à la question de savoir ce que les enfants doivent apprendre. Et la recherche a démontré à de multiples reprises que les écoles peuvent avoir une influence considérable sur l'apprentissage, ce dernier recouvrant divers extrants : maîtrise de la langue ; savoir lire, écrire et compter ; comportement social et cohésion du groupe ; édification de la nation et solidarité politique ; compétences professionnelles et développement économique.

Quand les compétences de base enseignées principalement à l'école primaire (et dans les programmes informels et d'alphabétisation pour les

27. UNESCO. 2004. *Rapport mondial de suivi de l'EPT 2005*, p. 41.

28. Ladipo *et al.*, 2009, p. 8.

adultes) deviennent la priorité, il y a globalement deux manières d'envisager les extrants : (1) mesurer les compétences et contenus directement enseignés à l'école (par exemple, tests portant sur ce qui a été appris dans le cadre du programme scolaire) ; (2) mesurer ce que les apprenants sont censés savoir et être capables de faire selon la société (par exemple, lire un journal). Parmi les évaluations internationales, régionales et nationales décrites dans la présente étude, beaucoup privilégient la première dimension en essayant de déterminer dans quelle mesure les enfants ont acquis ce qu'on leur a enseigné à l'école. D'autres évaluations (principalement l'EGRA mais aussi certaines parties des enquêtes réalisées auprès des ménages comme l'EIAA) s'intéressent avant tout aux compétences générales que les apprenants (jeunes et moins jeunes) peuvent avoir besoin de posséder, en reléguant au second plan les matières du programme scolaire enseignées à l'école. Il n'y a pas de séparation nette entre ces deux extrants : chacun a ses mérites, tout dépend des buts de l'évaluation.

Lorsqu'on tient compte des extrants ou conséquences à plus long terme, n'importe quel outil d'évaluation peut être intégré dans une analyse plus globale. Par exemple, l'éducation maternelle est souvent considérée comme ayant une influence significative sur l'éducation de l'enfant et sur ses chances dans la vie, de même que sur sa santé et son bien-être. Le modèle de transmission de ces éléments a posé un défi majeur. Toutefois, des études récentes semblent étayer l'idée d'un solide modèle prédictif. Plusieurs facteurs entrent en jeu, comme le niveau d'alphabétisation de la mère et la façon dont elle maîtrise sa langue. Ces aptitudes entraînent une meilleure maîtrise de ces mêmes compétences chez leurs enfants, à condition que la communication verbale soit incluse dans le modèle statistique²⁹. Dans ce cas, comme dans d'autres modèles aussi complexes, il peut être difficile de différencier intrants et extrants. Mais le fait de pouvoir mesurer les composantes d'apprentissage des modèles permet d'espérer qu'une intervention (comme la scolarisation ou l'alphabétisation) pourra avoir une influence réelle sur l'élaboration des politiques.

Quelles sont mes options ? Le point de vue d'un Ministre de l'Éducation

[Toute] équipe d'évaluation devrait s'assurer que des systèmes et des stratégies sont en place pour communiquer ses conclusions aux institutions et acteurs qui auront un rôle à jouer dans la mise en œuvre de la politique ...³⁰

29. LeVine *et al.*, 2011. Également LeVine et LeVine, 2001.

30. Ladipo *et al.*, 2009, p. 70.

Les responsables politiques sont toujours amenés à faire des choix. Pour cela, tout responsable politique digne de ce nom veut s'appuyer sur les meilleures données disponibles. Cela implique de savoir ce qui est réellement « mieux » et quel en sera le « coût » (en termes de temps, d'argent, de ressources humaines et de coûts d'opportunité). Sachant que les ministres de l'Éducation (ou leur équivalent) sont à la fois investis d'une grande responsabilité et soumis à une forte pression face à certains des problèmes les plus difficiles de la société, il leur faut impérieusement et régulièrement disposer de données rapidement disponibles et stratégiquement pertinentes pour choisir des options.

Les questions qu'un ministre peut se poser et auxquelles les évaluations peuvent fournir des options stratégiques sont notamment les suivantes :

- Quelle est l'efficacité de notre système éducatif ? Comment pouvons-nous mesurer l'impact des changements de politique sur les acquis scolaires ? Par exemple, si nous décidons d'enseigner les compétences de base dans la langue maternelle pendant les premières années, en verrons-nous l'impact sur l'acquisition de la lecture dans la première ou la seconde langue dès la 3^e ou la 4^e année ?
- Où rencontrons-nous les problèmes les plus sérieux ? Si nous faisons une priorité des objectifs de l'EPT ou du Millénaire pour le développement, notamment l'éducation de base universelle ou l'égalité entre les sexes, comment utiliser la mesure des acquis scolaires pour améliorer notre réponse nationale ?
- Comment notre système éducatif se situe-t-il par rapport à celui de nos voisins ? Faisons-nous aussi bien avec des ressources similaires ? Que faudrait-il pour que la comparaison soit valable ?
- Où se situent les grandes disparités au sein de notre système éducatif national ? Pourquoi certaines régions, communautés ou écoles réussissent-elles très bien, alors que d'autres sont très loin derrière ? Comment encourager un système plus équitable et améliorer la qualité de l'apprentissage pour tous les enfants ?
- Quand serons-nous capables de régler certains de ces problèmes ? Les évaluations prennent du temps. Si nous essayons de résoudre des problèmes à long terme, par exemple en formant des enseignants qualifiés pour tous les élèves, il sera peut-être approprié de recourir à certaines méthodes de collecte de données capables de rassembler davantage d'informations. S'il faut prendre des décisions avant le début de la prochaine année scolaire, par exemple sur le choix des manuels ou des programmes scolaires, il vaut mieux envisager des évaluations réalisables à court terme.

- Combien cela coûtera-t-il de régler ces problèmes ? Nous avons de réelles contraintes budgétaires³¹.

Répondre à ce type de questions ne permet pas de réformer ou de restructurer un système éducatif, mais donne aux responsables politiques (notre ministre hypothétique, dans le cas présent) des pistes pour utiliser plus efficacement les divers outils dont ils disposent pour prendre des décisions stratégiques. Cependant, toutes les évaluations ne se valent pas. En fonction des buts poursuivis et des difficultés spécifiques à régler, tout ministre serait bien avisé de déterminer avec soin quelle évaluation répond le mieux à quel ensemble de questions.

Pour un responsable politique, le calendrier peut être une variable aussi importante que l'argent. Les évaluations sont très diverses : certaines exigent un temps de préparation très long, d'autres un temps d'analyse considérable, d'autres encore sont conçues pour recueillir des données plus ciblées en peu de temps. L'essentiel est d'essayer de définir le but poursuivi, puis de déterminer quelles options permettent de l'atteindre.

Les chapitres suivants décrivent ces divers types d'évaluation (en particulier dans le domaine de la lecture et de l'alphabétisation), en explicitant leurs avantages et leurs inconvénients à la lumière de l'expérience acquise à ce jour.

31. Les coûts récurrents sont ceux qui sont intégrés dans les budgets annuels, comme les salaires des enseignants. Les coûts non récurrents, comme les évaluations, doivent souvent être couverts par des postes extrabudgétaires plus limités.

4. ÉVALUATION DE L'APPRENTISSAGE DANS LES PAYS EN DÉVELOPPEMENT

Principales évaluations de l'apprentissage

Les évaluations de l'apprentissage peuvent prendre différentes formes, avec divers contenus et finalités. Elles existent au moins depuis le début des systèmes nationaux d'éducation publique qui firent leur apparition en France au XIX^e siècle³². Le Gouvernement français demanda à Alfred Binet (également connu pour être l'un des pères des tests d'intelligence) d'élaborer un outil d'évaluation qui permettrait de prédire quels élèves avaient le plus de chances de réussir à l'école publique. Ce facteur de prédiction de la réussite scolaire a marqué un tournant dans l'utilisation de tests pour élaborer les politiques. Au cours du siècle suivant, spécialistes de l'éducation et responsables politiques ont fait tout leur possible, à des époques et dans des endroits divers, pour prendre des décisions semblables, faisant ainsi progresser l'utilisation des instruments d'évaluation dans l'éducation.

Évaluations à grande échelle de l'apprentissage

À partir des années 1980, des agences nationales et internationales ont commencé à utiliser de plus en plus les évaluations à grande échelle de l'apprentissage (LSEA – *Large-scale Educational Assessment*). Jusque-là, seules quelques évaluations à grande échelle transnationales avaient été administrées, principalement par l'Association internationale d'évaluation du rendement scolaire (IEA)³³. Les progrès technologiques et méthodologiques dans le domaine des évaluations, conjugués à la pression politique pour améliorer les systèmes éducatifs, ont favorisé cette tendance, y compris dans les pays les moins avancés (PMA)³⁴. La Conférence de Jomtien (1990) sur l'éducation pour tous a exhorté à plus de responsabilisation et à une évaluation systémique dans les PMA et, de plus en plus, les LSEA sont

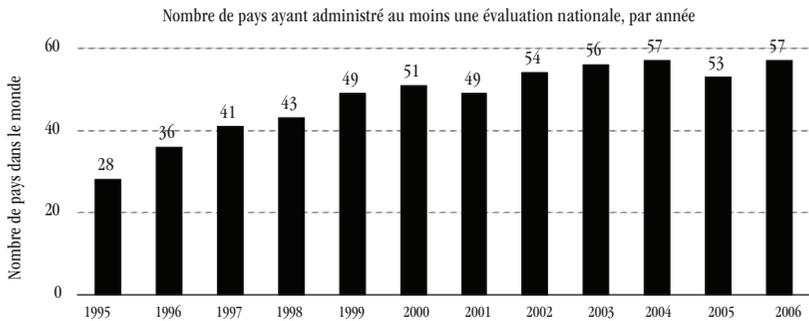
32. Les tests fondés sur les programmes scolaires viennent à l'origine de la Chine impériale. Toutefois, les examens chinois n'étaient pas axés sur l'éducation publique universelle (comme ce fut le cas dans la France post-révolutionnaire), mais plutôt sur une forme de sélection méritocratique pour l'administration publique.

33. Voir Chromy, 2002, p. 84 pour une liste des principales études, ainsi que Lockheed, 2008, p. 6.

34. Chromy, 2002 ; Kelleghan et Greaney, 2001, p. 32.

devenues un outil clé pour répondre à cette demande³⁵. En 2000, le Cadre d'action de Dakar (UNESCO) a appelé de ses vœux l'obtention de résultats d'apprentissage « quantifiables » et un « suivi systématique » des progrès³⁶. En conséquence, l'usage des outils d'évaluation a connu une très forte progression dans la planification de l'éducation (*figure 4.1*).

Figure 4.1 Progression de l'utilisation des évaluations nationales de l'apprentissage, 1995-2006



Adapté de : Benavot et Tanner, 2007, p. 6.

Malgré cette dynamique, la complexité et le coût croissants des LSEA ont conduit certaines personnes à remettre en question leur utilité dans les PMA³⁷. Si plusieurs agences ont administré des LSEA dans des pays de l'OCDE, il a fallu attendre les années 1990 pour que les PMA puissent davantage participer aux LSEA et entreprendre des évaluations nationales complexes³⁸. La complexité des intérêts des parties prenantes et les contraintes liées aux ressources ont limité la progression des LSEA dans les PMA. Quoi qu'il en soit, diverses agences, comme la Banque mondiale, se sont de plus en plus engagées dans le financement des LSEA, rendant ces dernières plus abordables et davantage envisageables, même quand les budgets nationaux étaient soumis à de fortes contraintes³⁹. Par ailleurs, la comparaison et la généralisation des données des tests sont d'autant plus

35. Lockheed et Verspoor, 1991.

36. UNESCO, 2000a, p. 21.

37. Braun et Kanjee, 2006, p. 8.

38. Greaney et Kelleghan, 2008, p. 8-9.

39. Selon une étude sur les décideurs nationaux, le financement de la Banque mondiale a été un facteur déterminant de l'adoption des LSEA par les pays à revenus faibles et moyens. Voir l'analyse dans Gilmore, 2005, p. 45.

difficiles dans un contexte et des circonstances politiques ou économiques en crise⁴⁰.

Évaluations internationales

[L]a valeur des études internationales tient peut-être davantage à leur capacité de générer des hypothèses sur les explications causales qu'à leur utilité pour vérifier des hypothèses⁴¹.

Dans le climat mondial actuel de compétitivité économique et de nouvelle ère de responsabilité pour tout, les gouvernements des pays développés tiennent plus que jamais à évaluer leurs systèmes éducatifs, en les comparant à ceux des autres, en termes d'extrants comme d'intrants et de processus. Une mesure comparative évidente des extrants est le niveau des élèves par matière, comme l'ont montré les résultats des tests obtenus dans des enquêtes internationales comme celles de l'IEA. En conséquence, au jourd'hui, les pays ne s'intéressent qu'à leur rang dans le classement international, lorsque les résultats des enquêtes de l'IEA sont publiés⁴².

Les évaluations internationales visent à évaluer les apprenants dans de nombreux pays. Elles ont également des objectifs multiples, notamment : (a) des comparaisons transnationales entre divers aspects des politiques éducatives ; (b) la production de tableaux de classement des scores de niveau par nation, région ou autres variables ; (c) la mesure des tendances dans le temps ; (d) des analyses nationales qui sont ensuite comparées à ce que font d'autres pays au niveau infranational. Les données recueillies concernent principalement les apprenants, les enseignants et les systèmes éducatifs, des paramètres qui constituent le meilleur moyen d'interpréter les résultats des tests.

Diverses organisations et agences internationales planifient et mettent en œuvre ce type d'études qui comportent le plus souvent des tests de lecture. L'Association internationale pour l'évaluation du rendement scolaire (IEA) pilote le *Programme international de recherche en lecture scolaire*⁴³ (PIRLS). L'Organisation de coopération et de développement économiques (OCDE) est responsable des études du *Programme international pour le*

40. Ross et Genevois, 2006.

41. Porter et Gamoran, 2002, p. 15 ; cité dans Braun et Kanjee, p. 33.

42. Johnson, 1999, p. 63.

43. Si l'accent est mis sur les études consacrées à la lecture, quelques références sont également faites aux études TIMSS et SISS sur les acquis en mathématiques, également pilotées par l'IEA.

suivi des acquis des élèves (PISA). Ces évaluations se caractérisent par la grande qualité des outils et la rigueur de la méthodologie employée sur le terrain, ainsi que par l'analyse très poussée des résultats. Chacune de ces évaluations internationales est actuellement en usage dans des douzaines de pays. Leur utilisation s'étend désormais bien au-delà du groupe des pays de l'OCDE qui furent les premiers à y participer⁴⁴.

Les évaluations internationales attirent souvent l'attention des médias, ce qui remet au cœur du débat le secteur de l'éducation et les résultats nationaux comparés à ceux des autres pays. Ces évaluations posent un certain nombre de problèmes qui doivent être pris en considération, en particulier l'âge des élèves évalués et la comparabilité entre pays⁴⁵.

Évaluations régionales

Soucieuses d'élargir l'usage des évaluations à grande échelle de l'apprentissage (LSEA) aux pays en développement, des organisations régionales et internationales ont uni leurs efforts pour mettre au point trois évaluations régionales majeures : le *Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation* (LLECE), le *Consortium de l'Afrique australe et orientale pour le pilotage de la qualité de l'éducation* (SACMEQ) et le *Programme d'analyse des systèmes éducatifs des pays de la CONFEMEN* [Afrique francophone] (PASEC). Ces trois évaluations sont décrites plus en détail dans l'*annexe A*.

Malgré leurs nombreux points communs avec les évaluations internationales, ces évaluations régionales présentent plusieurs différences majeures, notamment : la concordance relative entre le contenu des tests et celui des programmes scolaires ; des barèmes normatifs qui peuvent être ou ne pas être liés aux niveaux de compétence locaux (normés) ; la prise en compte des questions politiques locales (comme le rôle de la langue française dans le PASEC). Le chevauchement entre le travail des spécialistes travaillant aux niveaux régional et international a globalement conféré une

44. Dans une analyse récente, Kamens et McNeely (2010) soulignent que la mondialisation croissante a provoqué une augmentation considérable du nombre de pays qui participent aujourd'hui aux évaluations internationales et nationales. Ils affirment en outre que la mondialisation a favorisé l'apparition d'une « idéologie éducative mondiale » et d'une « hégémonie de la science », qui ont débouché sur une acceptation beaucoup plus grande des tests éducatifs.

45. Pour une analyse éclairante de ces contraintes et problèmes, voir Greaney et Kelleghan, 2008, p. 71-73.

grande crédibilité à ces tests régionaux. Un nombre croissant de pays en développement participent à ces évaluations régionales (*tableau 4.1*).

Tableau 4.1 Participation des pays du PME aux évaluations internationales, régionales et hybrides au cours de la dernière décennie

Pays	Évaluations internationales	Évaluations régionales	Évaluations hybrides
AFRIQUE			
Bénin			
Burkina Faso		PASEC	
Cameroun		PASEC	
Éthiopie			
Gambie			EGRA
Ghana	TIMSS 2003, SISS		EGRA
Guinée		PASEC	
Kenya		SACMEQ I et II	EGRA
Lesotho		SACMEQ II	
Libéria			EGRA
Madagascar		PASEC	
Mali			EGRA
Mozambique		SACMEQ II	
Niger		PASEC	EGRA
République centrafricaine		PASEC	
Rwanda			EGRA
São Tomé-et-Principe			
Sénégal		PASEC	EGRA
Sierra Leone			
ÉTATS ARABES			
Djibouti	TIMSS 2003, 2007	PASEC	
Mauritanie			
Yémen	TIMSS 2003, 2007		
ASIE ET PACIFIQUE			
Cambodge			EGRA
Mongolie	TIMSS 2007		
Tadjikistan			
Timor-Leste			EGRA
Vietnam			EGRA

Pays	Évaluations internationales	Évaluations régionales	Évaluations hybrides
AMÉRIQUE LATINE ET CARAÏBES			
Guyane			EGRA
Haïti		LLECE, SERCE	EGRA
Honduras	TIMSS 2007		EGRA
Nicaragua		LLECE	EGRA

Adapté de : M. Encinas-Martin, 2008, p. 30-31 ; et de RTI, 2009.

Évaluations nationales

Les évaluations nationales des acquis scolaires, très souvent laissées de côté dans les débats sur la qualité de l'éducation, peuvent être extrêmement utiles à deux titres. Premièrement, elles peuvent fournir des informations précieuses aux responsables des politiques éducatives sur les acquis scolaires dans les systèmes éducatifs nationaux, lesquels reflètent les orientations et les priorités des programmes scolaires nationaux. Deuxièmement, sachant que les cadres de suivi et d'évaluation sont une composante essentielle de la qualité de l'éducation, les évaluations nationales des acquis scolaires peuvent être un indicateur majeur de qualité et un point de départ pour améliorer la responsabilité et promouvoir la réforme. Les agences internationales et les organisations non gouvernementales doivent accorder plus de crédit aux évaluations nationales pour traiter les questions de qualité, même si ces évaluations sont une base insuffisante pour des comparaisons entre pays⁴⁶.

Les évaluations nationales (parfois appelées examens nationaux ou publics) visent à générer des informations sur le niveau des élèves dans un système éducatif donné. Presque tous les pays mènent une forme d'évaluation nationale pour déterminer si les buts éducatifs souhaités et planifiés ont été atteints⁴⁷. Les résultats peuvent être utilisés pour modifier les programmes scolaires, former les enseignants, réorganiser l'accès à l'école, ainsi que pour agir sur de nombreux autres aspects du système éducatif national. Ils peuvent aussi servir à la responsabilisation, pour prendre des décisions concernant l'affectation des ressources et sensibiliser davantage le public aux questions d'éducation. Ces évaluations sont souvent pratiquées sur une cohorte comprenant tous les élèves d'un niveau d'études donné (test fondé sur le nombre) ou sur un groupe choisi statistiquement (test sur un échantillon

46. Benavot et Tanner, 2007, p. 14.

47. Il s'agit souvent d'examens à enjeu élevé, comme le baccalauréat en France ; voir Greaney et Kellaghan, 2008.

de population). Elles peuvent aussi comprendre des questionnaires sur le milieu socioculturel destinés à différents participants (élèves, enseignants ou administrateurs), afin d'avoir des données contextuelles pour interpréter les résultats des tests. L'utilité des données obtenues dépend de la qualité et de la pertinence de l'évaluation, de l'exhaustivité du travail de terrain associé, ainsi que de la compétence des personnes chargées de l'analyse, de l'interprétation, de la rédaction des rapports et de la diffusion des résultats⁴⁸.

Enquêtes sur l'éducation réalisées auprès des ménages

Les enquêtes sur l'éducation réalisées auprès des ménages (HBES – *Household-based Educational Surveys*) sont utilisées depuis des décennies. Elles reposent souvent sur des méthodes d'échantillonnage qui permettent de recueillir, au sein d'un pays ou d'une région, des types spécifiques d'informations sur les groupes de population ciblés, et sont stratifiées selon des paramètres démographiques déterminés⁴⁹. En 2000, une initiative pluriannuelle a été engagée pour améliorer la collecte de données sur les taux d'alphabétisation dans les PMA⁵⁰. Une approche locale des enquêtes a été adoptée, la collecte de données visant davantage à répondre aux besoins locaux et nationaux, plutôt qu'à permettre des comparaisons internationales. Ont également été privilégiés des outils d'évaluation axés sur des programmes susceptibles d'être compris par tout un chacun, tout en exigeant moins de temps et d'efforts⁵¹. Le recours aux HBES se justifie

48. Les examens nationaux étant souvent élaborés dans le cadre de longs processus politiques (et internes) nationaux, ils sont généralement moins utiles pour les approches SQC décrites dans la présente étude. Ils ont en outre été analysés de façon approfondie dans d'autres rapports récents, tels que ceux de Greaney et Kellaghan (2008) et Kellaghan *et al.* (2011) ; il n'y a donc pas lieu d'entrer davantage dans les détails ici.

49. Dans le domaine de l'alphabétisation des adultes, l'une des premières enquêtes auprès des ménages a été entreprise (par l'auteur du présent rapport) au Zimbabwe, dans deux langues africaines locales (UNSO, 1989), suivie d'autres (par exemple, au Maroc, Lavy *et al.*, 1996 ; au Bangladesh, Greaney *et al.*, 1999 ; au Botswana, Commeyras et Chilisa, 2001). Pour une synthèse plus récente sur les HBES consacrées à l'alphabétisation des adultes, voir UNESCO (2008).

50. Cette initiative, appelée Programme d'évaluation et de suivi de l'alphabétisation (LAP), était un programme conjoint de l'Institut international de l'alphabétisation (ILI) et de l'UNESCO. Il a donné lieu à plusieurs rapports, parmi lesquels ILI-UNESCO 1998, 1999, 2002a, b. Voir aussi le *chapitre 8* du présent ouvrage.

51. On trouvera un résumé de cette approche dans Wagner (2003). Un effort notable, quoique de courte durée, dans ce domaine a été le projet Surveillance des acquis scolaires (UNESCO, 2000b ; Chinapah, 2003). Un autre effort de l'UNICEF a été l'approche ABC (Chowdhury et Ziegler, 1994). Merci à C. Chabbott pour cette remarque.

quand les personnes évaluées ne sont plus dans un contexte institutionnel, par exemple des adultes ou des jeunes non scolarisés. Toutefois, dans les écoles, il est beaucoup plus facile d'effectuer des évaluations car tous les élèves sont regroupés en un seul et même lieu. C'est pourquoi les évaluations HBES des compétences en lecture des enfants sont relativement rares⁵². On retrouve cependant certains aspects de leur méthodologie, en particulier l'échantillonnage ciblé, dans les évaluations hybrides abordées dans la section suivante.

Évaluations hybrides (y compris EGRA)

Améliorer l'évaluation des compétences en lecture dans un contexte comparatif peut avoir des conséquences contrastées sur les intérêts locaux, nationaux et internationaux. Les planificateurs impliqués dans les évaluations à grande échelle de l'apprentissage (LSEA) internationales peuvent considérer les intérêts nationaux et les considérations de politique intérieure (par exemple la démographie et la diversité ethnique) comme des problèmes épineux ou comme de simples contraintes. Dans d'autres cas, il peut s'avérer nécessaire de sacrifier les considérations nationales relatives à la diversité démographique, aux variations linguistiques, voire à la diversité orthographique (par exemple le rôle de l'écriture arabe au Mali ou au Sénégal), afin de pouvoir faire une comparaison internationale sur une base plus large. Pour toutes ces raisons et d'autres encore, les programmes locaux et les décideurs nationaux hésitent à sacrifier les intérêts locaux au profit d'institutions intéressées par des comparaisons régionales ou internationales, comme celles décrites antérieurement.

Une autre raison de privilégier le niveau local concerne les niveaux de compétences. Les LSEA internationales évaluent généralement le niveau d'un groupe d'élèves dans les écoles, ce qui suppose que ces derniers soient capables de passer un examen écrit de façon autonome. Dans les PMA les plus pauvres, surtout dans les petites classes, cette approche est extrêmement difficile à appliquer, même en simplifiant le contenu des tests (comme c'est le cas pour les pré-PIRLS⁵³). Si l'objectif est d'évaluer des enfants (ou des adultes) au niveau de l'*initiation* à la lecture (stade auquel

52. Des enquêtes sur la démographie et la santé ont été menées un peu partout dans le monde ; elles emploient parfois des mesures très simples de la performance en lecture (telles que « lisez cette phrase ») pour tenter de mettre en évidence les liens entre, par exemple, le fait de savoir lire et écrire et la santé. Merci à Luis Crouch pour cette remarque.

53. Voir l'analyse du pré-PIRLS dans l'*annexe A*.

restent de nombreux élèves dans les pays pauvres, même après deux ans de scolarité ou plus), il est pratiquement impossible d'obtenir une évaluation correcte avec la méthodologie des LSEA.

Depuis quelques années, une nouvelle approche de l'évaluation s'intéresse plus directement aux besoins des PMA. À l'origine, cette approche a été conceptualisée sous la dénomination de méthodes SQC (*smaller, quicker, cheaper* – simples, rapides et abordables) d'évaluation des compétences en lecture et écriture⁵⁴. L'idée était de voir si la méthodologie des LSEA et des HBES pouvait être remodelée pour élaborer des méthodes *hybrides*⁵⁵ juste de la bonne taille, permettant de collecter et d'analyser les données plus rapidement, et plus économiques en termes de temps et d'efforts. La méthodologie résultante serait suffisamment flexible pour être adaptée aux contextes locaux et, par conséquent, capable de traiter des problèmes tels que celui de la variation ethnolinguistique, auquel sont confrontés un grand nombre de pays pauvres dans le monde.

La méthodologie d'évaluation des compétences fondamentales en lecture (EGRA – *Early Grade Reading Assessment*) possède plusieurs des caractéristiques évoquées ci-dessus. À l'heure actuelle, il s'agit probablement de l'évaluation hybride des performances en lecture la plus connue. L'EGRA (traitée de façon approfondie dans le chapitre suivant) s'intéresse à l'apprentissage de la lecture et aux contextes locaux (plutôt qu'à la comparabilité des contextes), ainsi qu'aux caractéristiques linguistiques ou orthographiques locales en matière de lecture. Comme on le verra, le concept SQC ne facilite pas forcément l'évaluation ; il met simplement l'accent sur d'autres aspects. L'EGRA, en tant qu'évaluation hybride, a des objectifs différents de ceux mis en avant par les LSEA.

Une caractéristique supplémentaire des évaluations hybrides de type EGRA est la possibilité d'une plus grande transparence, et donc le caractère « partageable » des outils d'évaluation⁵⁶. Les outils développés pour les évaluations hybrides sont généralement plus flexibles et plus adaptables, car ils ne sont pas contraints de se conformer à un modèle reconnu à l'échelle internationale. Ils peuvent et doivent être partagés à divers niveaux

54. ILI/UNESCO, 1998 ; Wagner, 1990, 1997, 2003.

55. « Hybride » signifie une combinaison de deux choses ou plus. En l'occurrence, cela consiste à réunir quelques éléments des LSEA, des HBES et des évaluations des programmes scolaires, ainsi que des tests initialement destinés aux évaluations cognitives des compétences en lecture.

56. La notion de « partageabilité » a été pour la première fois mentionnée dans ILI/UNESCO, 1998 ; voir également le *chapitre 8* sur l'alphabétisation des adultes.

du système éducatif concerné. Cette capacité de partage peut déjà être constatée à travers les diverses utilisations qui sont d'ores et déjà faites des outils EGRA (*chapitre 6*). Des gains d'efficacité ainsi que des économies d'échelle peuvent être réalisés, si des outils d'évaluation identiques ou similaires sont employés pour réaliser aussi bien des enquêtes nationales que des évaluations locales.

Quelle doit être l'ampleur du champ des évaluations ?

Il faut du temps et de l'argent pour réaliser des évaluations. Si l'évaluation doit être représentative de l'ensemble de la population d'un pays, ou de plusieurs pays dans une démarche comparative, elle nécessitera probablement beaucoup plus de temps et d'argent. Il existe deux grands moyens de maîtriser les coûts : premièrement, limiter l'éventail des compétences à évaluer ; deuxièmement, limiter l'échantillon de population à tester. Ces deux formes d'échantillonnage doivent être comprises du point de vue des exigences techniques et statistiques, mais aussi des besoins stratégiques et des résultats.

Échantillonnage des compétences

L'ensemble d'items de test [IEA] résultant est ... délibérément conçu pour représenter les points communs des programmes scolaires nationaux dans la matière concernée. Compte tenu du degré de variété transnationale du contenu des programmes scolaires, cela réduit naturellement et inévitablement la capacité de cet ensemble d'items de représenter de façon exhaustive les programmes scolaires d'un pays, de sorte que, malgré tous les efforts bien intentionnés pour rendre le système impartial en termes de représentation des programmes scolaires, le résultat pourrait être le contraire⁵⁷.

La majorité des LSEA mettent généralement en œuvre des tests normalisés dans un domaine particulier, comme la lecture, les mathématiques ou les sciences. L'approche adoptée dans un domaine peut varier considérablement d'un test à l'autre, même si le même domaine est testé selon de nombreuses évaluations différentes. Les évaluations comme le PIRLS, le LLECE, le SACMEQ et le PASEC sont pour l'essentiel basées sur les programmes scolaires des pays concernés. Elles cherchent généralement à évaluer la concordance entre ce qui devrait avoir été enseigné (et appris) et ce que les élèves ont effectivement appris (comme le démontre l'évaluation). On trouvera ci-après un résumé des diverses approches de plusieurs LSEA

57. Johnson, 1999, p. 65.

majeures comportant des tests de lecture. Toutes, à l'exception de l'EGRA, sont administrées dans un contexte scolaire, sous forme de tests écrits ciblant des groupes d'élèves.

Le PIRLS évalue le niveau de compréhension écrite. En se référant à la *Reading Literacy Study* (étude des compétences en lecture), pour laquelle des données ont été collectées en 1990-1991, l'enquête PIRLS a été réalisée deux fois (en 2001 et 2006). Quatre processus de compréhension écrite ont été inclus pour mesurer les compétences suivantes : trouver et expliquer des informations spécifiques ; tirer des conclusions à partir de séquences logiques ou chronologiques et d'événements interdépendants ; interpréter et assimiler des idées et des informations ; examiner et évaluer le contenu, la langue et les éléments textuels.

Dans l'enquête PISA, le test de compétence en lecture part du postulat que les élèves savent lire et tente d'évaluer leur aptitude à comprendre et à réfléchir sur divers documents écrits. L'enquête PISA 2006 a testé les compétences suivantes : connaissances et compétences appliquées dans le contexte personnel, public, professionnel et éducatif ; contenu ou structure des textes (continus, sous forme de tableaux, de graphiques ou de formulaires) ; processus qui doivent être accomplis, comme l'extraction, la réflexion, l'évaluation et l'interprétation d'un texte écrit.

Le SACMEQ a adopté la définition des compétences en lecture employée dans l'étude sur l'apprentissage de la lecture (*Reading Literacy Study*) de l'IEA (1990) : « Aptitude à comprendre et à utiliser les formes de langage écrit requises par la société et/ou considérées comme importantes par l'individu. »⁵⁸. En outre, il a élaboré son test à partir des trois domaines identifiés dans cette étude : les documents (présentation structurée d'informations sous la forme de graphiques, tableaux, cartes, diagrammes, listes ou ensembles d'instructions) ; la prose narrative (texte continu, dans lequel l'auteur cherche à raconter une histoire, réelle ou fictive) ; la prose descriptive (texte continu visant à décrire, expliquer ou communiquer autrement une information factuelle ou une opinion).

Les tests de compétences en lecture du PASEC ont été construits en français, sur la base des éléments communs aux programmes scolaires des pays d'Afrique francophone. Au niveau de la 2^e année, les tests en français évaluent le vocabulaire visuel des élèves, la compréhension de phrases et de

58. Elley, 1992.

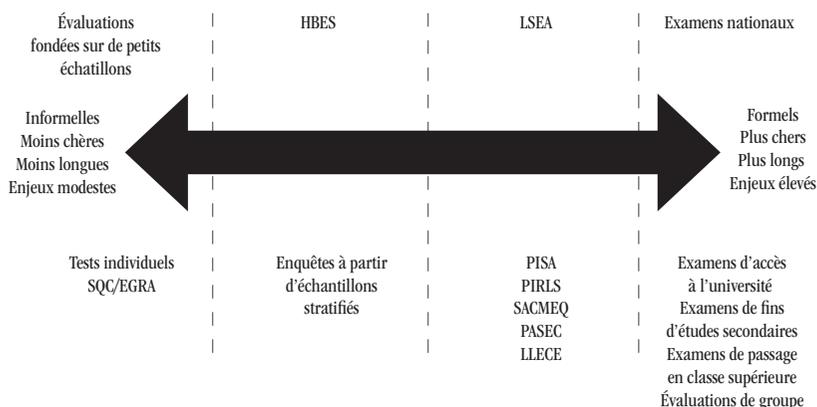
textes, ainsi que l'écriture. En 5^e année, en plus des items de la 2^e année, les tests évaluent l'orthographe et divers aspects de la grammaire⁵⁹.

Les tests du LLECE comprennent des questions à choix multiples (QCM) et des items à réponse non dirigée. Les composantes linguistiques sont notamment la compréhension écrite, la compétence métalinguistique et la rédaction d'un texte en espagnol. Au Brésil, les tests sont administrés en portugais⁶⁰.

L'EGRA comporte un ensemble de mesures administrées individuellement, essentiellement sur la base d'un certain nombre de compétences relatives à la fluidité de la lecture, qui ont été développées à l'origine afin de poser un diagnostic sur l'apprentissage de la lecture. On trouvera de plus amples détails sur l'EGRA au *chapitre 5*.

L'un des moyens d'étudier les divers types d'échantillonnage des compétences, ainsi que d'autres paramètres abordés plus loin, consiste à considérer ces évaluations comme un continuum allant de l'EGRA aux examens nationaux, comme illustré dans la *figure 4.2*.

Figure 4.2 Continuum des évaluations : des évaluations SQC hybrides aux examens nationaux, en passant par les LSEA



Adapté de : Kanjee, 2009.

59. CONFEMEN, 2008.

60. UNESCO-LLECE, 2008.

Échantillonnage de population

La représentativité de l'échantillon de la population est un aspect fondamental de toute évaluation, mais les procédures d'échantillonnage varient considérablement d'une évaluation à l'autre.

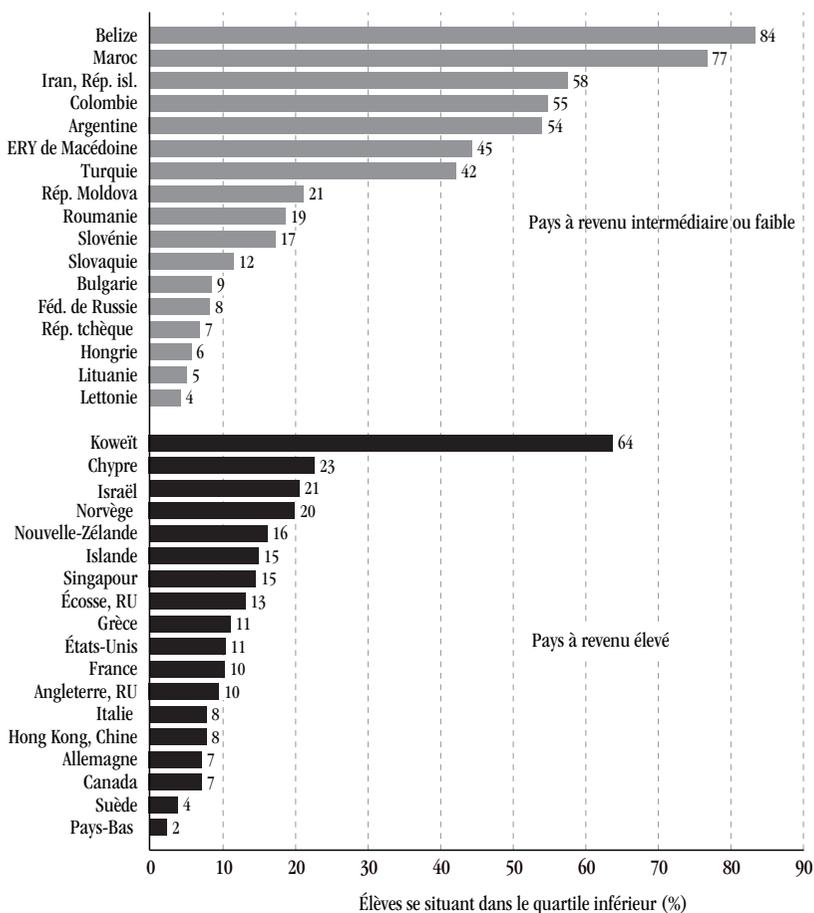
L'enquête PIRLS emploie une méthode d'échantillonnage en deux étapes. Un échantillon d'au moins 150 écoles, proportionnel au nombre d'élèves du niveau de scolarité considéré, est d'abord choisi. Dans un deuxième temps, on distingue dans chaque pays les élèves de 4^e année. Dans certains pays, l'âge de l'échantillon peut être très hétérogène, en particulier dans les pays en développement où la scolarisation tardive ou les redoublements sont fréquents. Deux autres critères sont importants : la localisation géographique de l'école et son statut (école publique, privée ou religieuse). Dans certains pays, ces statuts ne sont pas toujours clairs, ce qui pose des questions quant à la possibilité de comparer des sous-groupes de population à l'intérieur des pays (voir le *chapitre 6* sur la comparabilité).

En 1991 et 2001, l'enquête PIRLS a évalué les élèves de 4^e année (âge type de 9 ans environ), car on part du principe (du moins dans les pays de l'OCDE) que ces enfants savent lire et sont capables de passer un test écrit (*figure 4.3*)⁶¹. Depuis 2011, une évaluation « pré-PIRLS », qui est une version moins difficile de l'enquête PIRLS (vocabulaire plus facile, passages plus courts, etc.), est administrée au même niveau de scolarité, afin de recueillir une plus grande diversité de données à l'extrémité inférieure de l'échelle⁶².

61. Voir Olson *et al.*, 2008. Voir également le site Internet du PIRLS : <http://timssandpirls.bc.edu/isc/publications.html>.

62. Selon Mullis *et al.* (2009), le pré-PIRLS devrait également recueillir davantage d'informations contextuelles sur le foyer, les écoles, les salles de classe et les possibilités d'apprendre. Par ailleurs, les auteurs déclarent : « le pré-PIRLS ciblant les élèves à un stade plus précoce de l'apprentissage de la lecture, un pourcentage plus important d'items (50 % de l'évaluation) est consacré à la mesure de l'aptitude à se concentrer et à rechercher des informations explicitement énoncées – fondement premier de la compréhension du texte écrit » (p. 14). Les honoraires de participation (par pays) sont fixés à 30 000 dollars EU par an pour chacune des cinq années (soit un total de 150 000 dollars EU). Voir également le *chapitre 7* sur les coûts. En 2011, certains pays pouvaient choisir d'évaluer les enfants de 5^e ou de 6^e année, en particulier s'ils prévoient d'appliquer le pré-PIRLS en 4^e année.

Figure 4.3 PIRLS : pourcentage des élèves de 4^e année se situant dans le quartile inférieur de l'échelle internationale des capacités de lecture, 2001



Note : Le classement des pays par niveau de revenu est basé sur celui de la Banque mondiale, 2003b.

Source : Mullis *et al.*, 2003.

Adapté de : UNESCO, 2004, p. 137.

Dans l'enquête PISA, le principal critère de sélection des élèves est leur âge (15 ans), indépendamment de leur niveau de scolarité et du type d'établissement. Cela peut donner des différences importantes entre les pays, en termes d'expérience d'apprentissage. En France, par exemple, certains

élèves de 15 ans sont dans le second cycle du secondaire (« lycée » dans les pays francophones), tandis que d'autres sont dans le premier cycle (« collège »). Dans ce cas, contrairement à plusieurs autres pays, il faut choisir une certaine proportion d'élèves dans plusieurs niveaux de scolarité⁶³. L'enquête PISA utilise cinq niveaux de compétence en lecture (figure 4.4).

L'enquête LLECE prend en compte divers critères de stratification, notamment le type de zone géographique (métropole, milieu urbain, milieu rural) et le type d'école (publique ou privée). Environ 4 000 élèves sont sélectionnés (40 par école), répartis équitablement entre les deux niveaux de scolarité testés. L'enquête LLECE évalue les élèves de deux années consécutives (3^e et 4^e années) dans le cadre de la collecte de données. Selon les pays concernés, les élèves avaient 8 ou 9 ans. La seconde enquête LLECE⁶⁴ a évalué les élèves des 3^e et 6^e années⁶⁵.

L'enquête PASEC cible les enfants scolarisés à l'école primaire, en 2^e et 5^e années. Elle cherche à identifier les facteurs qui influent sur l'apprentissage des élèves en Afrique francophone⁶⁶. L'échantillonnage est effectué à deux niveaux. Premièrement, on sélectionne un échantillon d'écoles en fonction du nombre d'élèves dans chacun des deux niveaux de scolarité. Deuxièmement, les écoles sont choisies par stratification, de façon à ce qu'elles soient représentatives de l'ensemble du système éducatif national⁶⁷.

63. Selon Postlethwaite (2004, p. 3), on pourrait même parler de « pseudo-enseignants », dans la mesure où l'échantillon comprendra un groupe d'élèves de différents niveaux de scolarité, avec un grand nombre et une grande diversité d'enseignants. Néanmoins, un objectif important est de pouvoir évaluer les améliorations moyennes de la compétence d'une année à l'autre autour des niveaux de scolarité auxquels se situent le plus souvent les élèves de 15 ans. C'est l'un des très rares moyens existants pour se faire une idée de l'importance de ce que les enfants apprennent d'une année à l'autre, par rapport à une norme donnée de connaissances. Voir Filmer *et al.* (2006). Merci à Luis Crouch d'avoir attiré l'attention sur ce point.

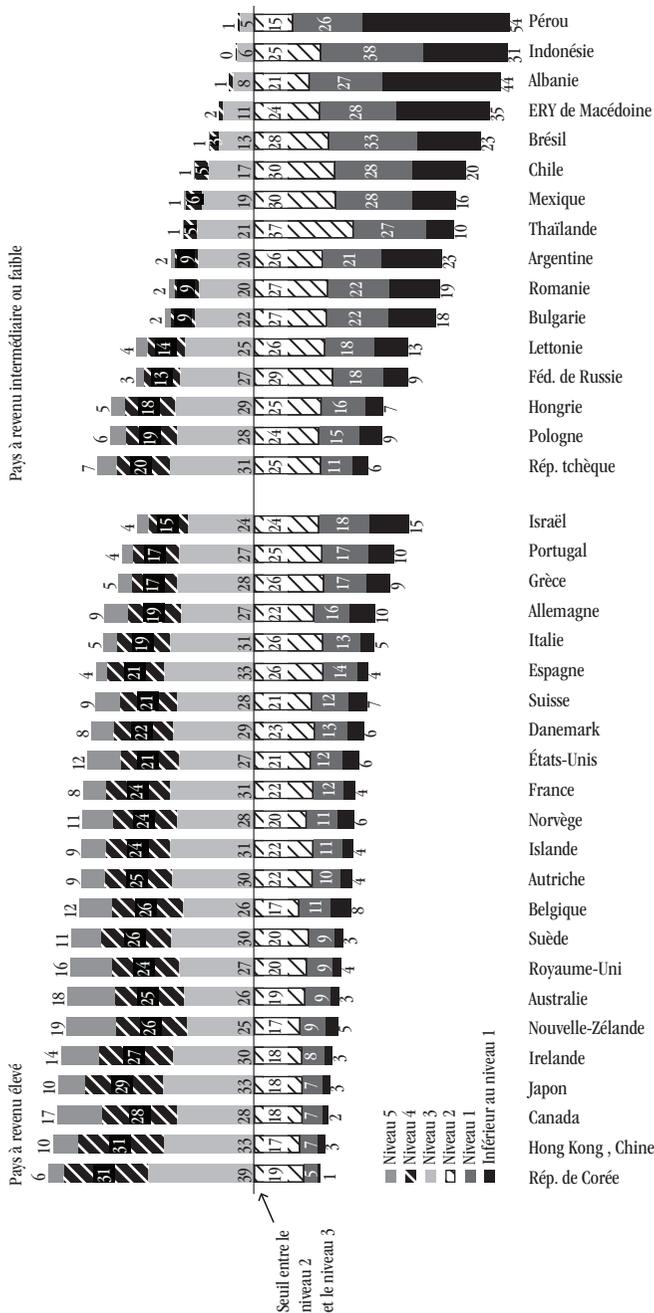
64. La seconde évaluation LLECE est appelée SERCE.

65. UNESCO-LLECE, 2008.

66. Pour une description plus complète de la scolarisation en Afrique francophone, voir Jarousse et Mingat, 1993.

67. La stratification est pratiquée par type d'école ou par type de zone géographique (par exemple urbaine ou rurale), mais sans différencier la zone géographique. Quand les écoles sont sélectionnées, le PASEC forme un nombre fixe de groupes d'élèves (15 élèves par groupe) pour chaque niveau évalué. Au total, il faut un minimum de 150 écoles.

Figure 4.4 PISA : pourcentage des élèves de 15 ans atteignant 5 niveaux de compétence en lecture, 2000-2002 (dans une sélection de pays)

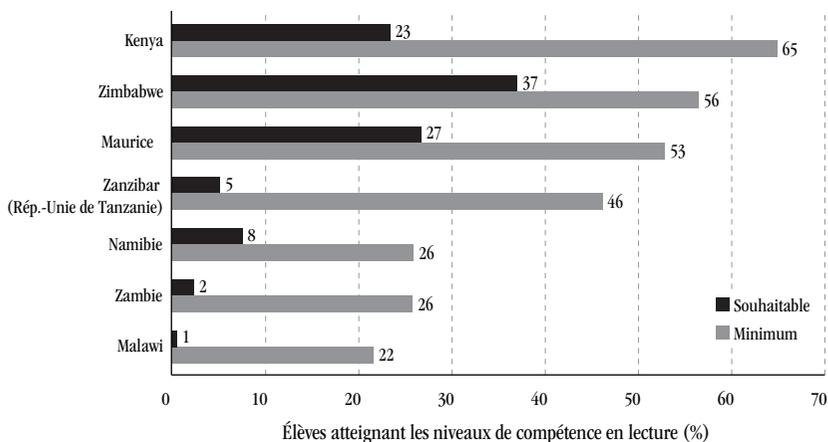


Adapté de : UNESCO, 2004, p.138.

Le PASEC évalue deux niveaux d'étude : le CP2 (2^e année du primaire) et le CM1 (4^e année du primaire). De plus, pour chacun de ces niveaux, les élèves sont évalués au début et à la fin de l'année scolaire. Cela permet d'analyser la variation dans le temps du niveau de performance de chaque élève. Le PASEC est la seule LSEA à procéder à ce type de mini-évaluation longitudinale.

L'enquête SACMEQ évalue les compétences en lecture des élèves de 6^e année (*figure 4.5*). Cette évaluation tardive est partiellement due au fait que les élèves des pays du SACMEQ passent de l'usage de la langue locale à celui de leur langue nationale au cours du cycle primaire. Cette transition linguistique intervient généralement autour de la 3^e (ou de la 4^e année), le postulat étant que la plupart ou la totalité des élèves auront acquis une maîtrise suffisante de la langue nationale en arrivant en 6^e année⁶⁸. La technique d'échantillonnage employée s'apparente à celle du PIRLS.

Figure 4.5 SACMEQ : pourcentage d'élèves de 6^e année qui maîtrisent la lecture dans sept pays d'Afrique, 1995-1998



Note : Les pays sont présentés en ordre décroissant de la proportion des élèves atteignant les niveaux minimums de compétence.

Source : Kulpoo (1998); Machingaidze, Pfkani and Shumba (1998); Milner *et al.* (2001); Nassor and Mohammed (1998); Nkamba and Kanyika (1998); Nzomo, Kariuki and Guantai (2001); Voigts (1998).

Adapté de : UNESCO, 2004, p. 136.

68. Voir Ross *et al.*, 2005, p. 39-41. Bien sûr, cette hypothèse varie fortement d'un endroit à un autre. C'est l'une des principales raisons de l'attractivité des évaluations EGRA dans les langues locales.

L'EGRA s'intéresse principalement à l'initiation à la lecture. C'est pourquoi ses évaluations sont en général orales et ont lieu de la 1^{re} à la 3^e année de scolarité. La taille moyenne de ses échantillons est généralement inférieure à celle des autres LSEA, mais avec une fourchette très large : de 800 enfants au Kenya jusqu'à environ 6 000 au Nicaragua.

Exclusions de populations

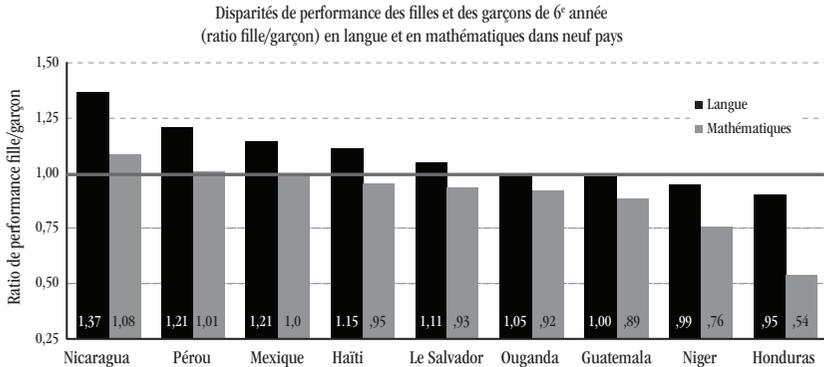
Paradoxalement et depuis toujours, force est de constater qu'une proportion importante des populations ayant le plus besoin d'une meilleure éducation est systématiquement exclue des mesures lors des évaluations à grande échelle de l'apprentissage (LSEA). Selon les spécialistes de l'évaluation : « Si vous n'êtes pas évalué, vous n'existez pas. » Cette situation semble être à la fois le résultat et une cause de l'exclusion des LSEA des populations les plus vulnérables. Les explications varient d'un test à l'autre et d'une politique nationale à l'autre, mais le résultat reste le même : ceux qui ont le moins de chances de réussir les tests et ceux qui sont les plus défavorisés représentent le groupe le plus souvent exclu de l'échantillon de population visée par l'évaluation. Pour comprendre pourquoi, il est utile d'explicitier ce que recouvre le terme « exclusion ».

Exclusion due au genre et à la zone géographique

Le genre est un des facteurs principaux de non-scolarisation dans les PMA, bien que des progrès significatifs aient été réalisés ces dernières décennies. Quoi qu'il en soit, dans les pays les plus pauvres, les filles continuent d'être moins scolarisées que les garçons, au niveau de l'entrée à l'école primaire et vers la 5^e année. Dans les PMA les plus pauvres, l'exclusion systématique des filles et la discrimination ont généralement pour effet une moindre fréquentation de l'école par les adolescentes, ainsi que des scores inférieurs à ceux des garçons lors des évaluations nationales (*figure 4.6*). Des tendances similaires font apparaître des différences importantes lors des évaluations nationales entre zones rurales et urbaines des PMA. Par ailleurs, la difficulté pour les autorités d'y localiser (au sens littéral) les enfants nomades peut rendre leur scolarisation onéreuse⁶⁹.

69. Par exemple, selon Greaney et Kellaghan (2008, p. 71), divers problèmes d'échantillonnage pour le TIMSS sont apparus en République du Yémen, où plusieurs écoles n'avaient pas de classes de 4^e année et où les enfants nomades ne pouvaient être localisés. Le fait qu'il y ait des exclusions inévitables ne signifie pas que les LSEA ignorent le problème. Ainsi, le PIRLS a présenté de façon explicite toutes ses prises de décision dans des rapports techniques comme le Rapport international sur le PIRLS 2006 (Annexe A) et le Rapport technique (chapitre 4, Annexe B). Communication personnelle, A. Kennedy et K. Trong, 2010.

Figure 4.6 Disparités entre les genres en termes d'aptitudes en langue et en mathématiques en 6^e année, d'après les évaluations nationales de l'apprentissage



Adapté de : Benavot et Tanner, 2007, p. 15.

Exclusion due à la langue et à l'ethnicité

[L]a formulation de la politique linguistique est le plus souvent examinée au niveau des États-nations pour ce qui est de la façon de structurer l'utilisation des langues à l'intérieur de leurs frontières. Cela a pour effet de conférer aux langues un certain statut, par exemple celui de langue nationale, de langue officielle, de langue provinciale ou une autre catégorie⁷⁰.

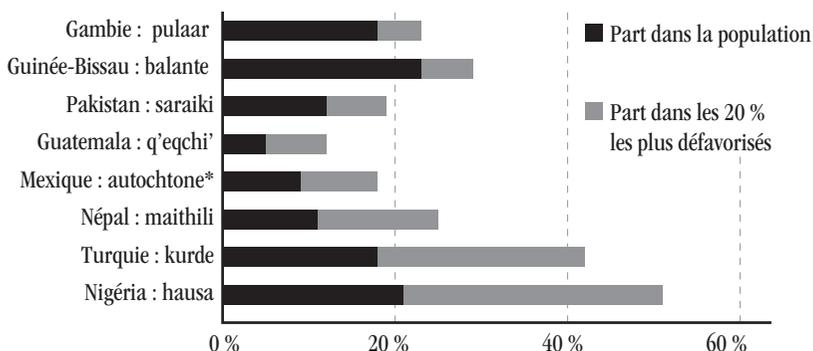
Dans presque tous les pays, il existe des variations linguistiques d'un groupe ethnique à l'autre, pour des raisons historiques et à cause de migrations récentes. Beaucoup de ces groupes, parfois appelés minorités ethnolinguistiques, sont bien intégrés dans le tissu national (comme en Suisse), mais leur présence peut parfois provoquer des conflits civils (comme au Rwanda). Souvent, les forces sociales et politiques s'efforcent d'atténuer les différences, notamment en prenant des décisions politiques qui débouchent sur une hiérarchie de langues dont l'usage est acceptable à l'école et dans les structures de gouvernance⁷¹. Dans ces contextes, que ce soit dans les pays de l'OCDE ou dans les PMA, il n'est pas rare que les

70. Robinson, 2004, p. 4.

71. Voir Hornberger, 2003. Bien sûr, tous les problèmes ethnolinguistiques, que ce soit dans le cadre de l'éducation ou à l'extérieur, ne sont pas aussi faciles à résoudre.

enfants qui parlent une langue minoritaire soient exclus des évaluations⁷². Ce phénomène peut être particulièrement accentué dans les régions en proie à un conflit civil ou à une crise économique qui provoquent d'importantes migrations transfrontières, où les immigrants (et leurs enfants) sont traités comme des personnes de passage, mais aussi pour les groupes dont les possibilités de scolarisation sont limitées ou nulles (figure 4.7).

Figure 4.7 Pourcentage de différents groupes linguistiques se situant dans les 20 % des niveaux inférieurs de l'éducation dans une sélection de pays



Note: Les « 20 % les plus défavorisés » correspondent aux 20 % de jeunes âgés de 17 à 22 ans ayant effectué le plus petit nombre d'années de scolarité.

* Au Mexique, sont classés dans la catégorie « autochtone » ceux qui utilisent des langues autochtones sans savoir parler espagnol.

Source : UNESCO-DME (2009).

Adapté de : UNESCO, 2010, p. 164.

Exclusion due à d'autres facteurs

L'enquête PISA insiste sur la nécessité pour les pays de préciser les raisons pour lesquelles certains élèves sont exclus des échantillons nationaux stratifiés. Faute de quoi, n'importe quel pays pourrait faire augmenter artificiellement sa moyenne nationale en ignorant certaines parties de sa population. Les règles d'exclusion du PISA sont les suivantes : élèves présentant un handicap mental ou physique ; élèves nés dans un autre pays (et pouvant donc avoir des problèmes de seconde langue) ; élèves

72. Aux États-Unis, par exemple, lors de l'évaluation nationale de l'alphabétisation des adultes en 2003, seules les compétences en anglais et en espagnol ont été évaluées, alors que des douzaines d'autres langues sont utilisées par les personnes fréquentant des cours pour adultes. Ministère américain de l'éducation, 2009.

diagnostiqués comme dyslexiques (avec déficiences en lecture)⁷³. À cause de l'importance des tableaux de classement, des pays continuent de se plaindre de ce que les échantillons de population d'autres pays ne sont pas véritablement représentatifs.

En ce qui concerne le SACMEQ, les élèves des « petites » écoles ont été exclus, même si la définition de ces écoles varie selon les pays participants. Au Lesotho, par exemple, si une école avait moins de 10 élèves en 6^e année, elle était écartée de l'échantillon de population. Aux Seychelles, au Botswana et en République-Unie de Tanzanie, les écoles ayant moins de 20 élèves ont été exclues. En Ouganda, on a exclu les élèves qui habitaient dans des zones de guerre civile⁷⁴. Comme on le voit, il y a en pratique beaucoup de raisons d'exclusion, en particulier pour les LSEA qui doivent évaluer efficacement des cohortes importantes d'élèves. Toutefois, si on privilégie les plus défavorisés, même les règles d'exclusion les plus impartiales ne seront pas équitables par rapport à l'objectif d'évaluer tous les élèves.

Enfin, dans les pays pauvres, de nombreux enfants ne sont pas scolarisés et ne seront donc pas évalués par les LSEA, qui ne sont généralement pas administrées avant la 4^e année. Les « taux de survie » (taux nets d'achèvement des études d'une cohorte) dans les pays pauvres sont généralement assez bas : pas plus de 20 à 30 % dans les PMA les plus pauvres. Les LSEA, comme le PIRLS et le PISA, font tout simplement l'impasse sur une grande partie des enfants les plus vulnérables des pays pauvres, parce que ces enfants ne vont plus à l'école quand ces évaluations internationales sont administrées. Le *Rapport mondial de suivi sur l'EPT 2010, Atteindre les marginalisés*, fait de ce constat le pivot de son argumentation. Ce fait pourrait constituer une importante raison d'adapter les outils SQC aux contextes non scolaires⁷⁵.

73. Ce type d'exclusion lié à la dyslexie est examiné au cas par cas par le consortium d'experts nationaux du PISA (OCDE, 2009b). Selon Wuttke (2008), le Danemark, l'Espagne, la Finlande, l'Irlande et la Pologne ont exclu des élèves pour ce motif de l'étude PISA de 2003. Le Danemark a exclu les élèves présentant des déficiences en mathématiques. Le Luxembourg a exclu les nouveaux immigrants.

74. Voir Ross *et al.*, 2005. Voir également le rapport SACMEQ II sur le Kenya (Onsumo *et al.*, 2005).

75. *Rapport mondial de suivi sur l'EPT* (UNESCO, 2010). Évaluer les enfants non scolarisés semble tout à fait réalisable avec des évaluations du type EGRA, bien que cela ne semble pas avoir été fait récemment ; dans les années 1980, Wagner (1993) a utilisé des évaluations hybrides similaires pour étudier les compétences en lecture d'enfants fréquentant des écoles islamiques au Maroc. Des méthodes semblables ont également été employées pour étudier les adultes peu alphabétisés ou analphabètes (Lavy *et al.*, 1995).

Comparabilité des évaluations

L'un des écueils potentiels de l'utilisation de données internationales dans ce but [comparatif] est le fait que, l'évaluation devant être administrée dans plusieurs pays, son contenu risque de ne représenter de façon adéquate les programmes scolaires d'aucun pays participant⁷⁶.

[L]es populations et les échantillons d'élèves inclus dans les évaluations internationales peuvent ne pas être strictement comparables. Par exemple, des différences de performances peuvent apparaître parce que les pays diffèrent dans le degré de maintien des élèves dans leur système éducatif, ou dans le cas où certaines catégories d'élèves sont mises à l'écart de la vie scolaire normale et risquent donc d'être exclues d'une évaluation⁷⁷.

L'EGRA *ne doit pas* être utilisée pour comparer les résultats dans différentes langues. Les langues ayant différents niveaux de transparence orthographique, il serait en effet déloyal de dire que le pays A (où tous les enfants lisent de façon automatique au plus tard en 2^e année) a de meilleurs résultats que le pays B (où les enfants n'atteignent ce niveau qu'en 3^e année), si la langue du pays A possède une orthographe beaucoup plus transparente que celle du pays B. Quoi qu'il en soit, déterminer en quelle classe les enfants de divers pays commencent généralement à « maîtriser » la lecture et comparer ces classes est un exercice stratégique et analytique utile, tant qu'il n'est pas utilisé à des fins de « classement » ou en vue d'établir une norme universelle unique, par exemple, de fluidité ou d'automaticité de la lecture⁷⁸.

La comparabilité des données est une préoccupation majeure des responsables politiques et des agences de planification. Si les définitions et les classifications varient, il peut être difficile, voire impossible, de comparer des données recueillies lors de différentes enquêtes et évaluations. La comparabilité et la stabilité sont forcément une marque distinctive des données recueillies par les Nations Unies, notamment par l'Institut de statistique de l'UNESCO (ISU). Toutefois, si la comparabilité devient le but premier et si l'on relègue au second plan la validité (locale et culturelle) des définitions et des classifications de l'apprentissage, alors ces données peuvent devenir moins significatives et potentiellement moins applicables sur le terrain. Il s'agit d'une tension naturelle et essentielle entre les

76. Ladipo *et al.*, 2009, p. 19.

77. Greaney et Kellaghan, 2008, p. 77.

78. RTI, 2009, p. 11, ce sont les auteurs du rapport original qui soulignent.

approches « émiques » (intraculturelles) et « étiques » (interculturelle) de l'évaluation⁷⁹.

Les études comparatives ne doivent pas nécessairement avoir les classements comme seule finalité. Elles peuvent aussi être un moyen de susciter le débat, quand une variation est constatée entre les pays. Par exemple, lors d'une enquête nationale auprès des ménages réalisée par la Banque mondiale au Bangladesh, on a constaté que cinq années d'école primaire n'aboutissaient qu'à l'équivalent d'une première année d'école primaire en termes d'acquis scolaires, et que trois années de scolarité avaient une valeur à peu près nulle en termes d'acquis scolaires⁸⁰. Cela a montré que les investissements réalisés par le Bangladesh dans le domaine des compétences de base étaient insuffisants pour atteindre les objectifs nationaux. Cette étude, ainsi que d'autres, a inspiré les évaluations hybrides de type SQC visant à détecter à un stade précoce les problèmes graves en matière d'éducation.

Dans les évaluations, est-il possible d'équilibrer de façon satisfaisante la comparabilité et la prise en compte du contexte ? Faut-il évaluer les pays ayant des scores moyens faibles selon les mêmes échelles que les pays qui ont des scores moyens beaucoup plus élevés ? S'il y a des pays (ou des groupes d'élèves) en bas de l'échelle, certains préconiseront d'abaisser l'échelle de difficulté. D'autres peuvent considérer que c'est l'échelle elle-même qui est imparfaite et qu'il existe différents types de compétences qui pourraient être mieux évalués, en particulier si les variables procèdent manifestement de la race, de l'ethnicité, de la langue et de variables associées, amenant à remettre en cause le test autant que le groupe testé. Pour certains, utiliser différentes échelles pour différents groupes (ou nations) est un compromis inconfortable par rapport aux normes générales.

Dans la mesure où la comparabilité est possible (et aucune évaluation ne revendique une comparabilité parfaite), les résultats permettent aux responsables politiques de considérer leur situation nationale (ou régionale)

79. Voir Wagner, 2004, par exemple. Les approches dites « émiques » sont celles qui sont délibérément axées sur la pertinence culturelle locale, comme les mots ou descripteurs locaux qualifiant une personne « intelligente ». Les approches dites « étiques » sont celles qui considèrent que l'« intelligence » est un concept universel et tentent de mesurer les individus de différentes cultures sur la base de ce seul concept ou définition. Certains voient aussi cette tension comme un moyen de penser la frontière entre les disciplines que sont l'anthropologie (émique) et la psychologie (étique). Pour une discussion antérieure, voir Harris, 1976.

80. Greaney *et al.*, 1999.

par rapport à celle des autres. Cela semble être surtout intéressant s'il y a des choix proximaux (par opposition à distaux) à faire. En Afrique, par exemple, si un pays voisin a adopté un programme éducatif bilingue particulier, qui semble mieux fonctionner à l'école primaire et si le ministre de l'Éducation pense que les deux pays ont une situation suffisamment proche, comparer les résultats de la performance en lecture au primaire, par exemple, présente alors un intérêt indéniable. Une comparaison plus distale pourrait consister à relever qu'un type donné de programme éducatif bilingue au Canada semble efficace, mais on peut avoir davantage de doutes sur son applicabilité en Afrique, où le contexte est très différent. La proximité n'est pas toujours la caractéristique la plus pertinente : il y a de nombreux cas (aux États-Unis et au Japon, par exemple) où les rivalités entre résultats scolaires et systèmes économiques donnent lieu à des discussions et des débats animés depuis de nombreuses années⁸¹. On peut aussi citer le cas où de hauts fonctionnaires du Botswana ont voulu savoir comment Singapour s'y prenait pour avoir les meilleurs résultats en mathématiques⁸². Les niveaux de compétence en lecture et en écriture évalués par le SACMEQ (*figure 4.8*), où la position relative des pays peut évoluer dans le temps, sont un bon exemple de comparaison régionale⁸³.

La question principale ici est celle de savoir dans quelle mesure il est nécessaire d'avoir une comparabilité totale, c'est-à-dire avec tous les individus et tous les groupes sur la même échelle de mesure. Ou, si le choix est fait de ne pas imposer les compromis nécessaires pour avoir une seule échelle unifiée, quels sont les gains et les pertes en termes de comparabilité ? On peut aussi se demander si les échelles doivent mesurer les mêmes compétences. Par exemple, l'EGRA valorise les compétences cognitives qui précèdent la lecture, telles que la conscience phonémique, alors que les évaluations à grande échelle de l'apprentissage (LSEA) internationales se concentrent sur la compréhension du texte écrit. Les statistiques internationales resteront-elles aussi stables et fiables si l'on opte pour des approches localisées plutôt que pour la comparabilité internationale ? Cela a conduit à des situations où certains PMA, bien que tentés de participer à des évaluations internationales, ont hésité à cause de la crainte de la publication de mauvais résultats ou du sentiment que la valeur ajoutée obtenue, en

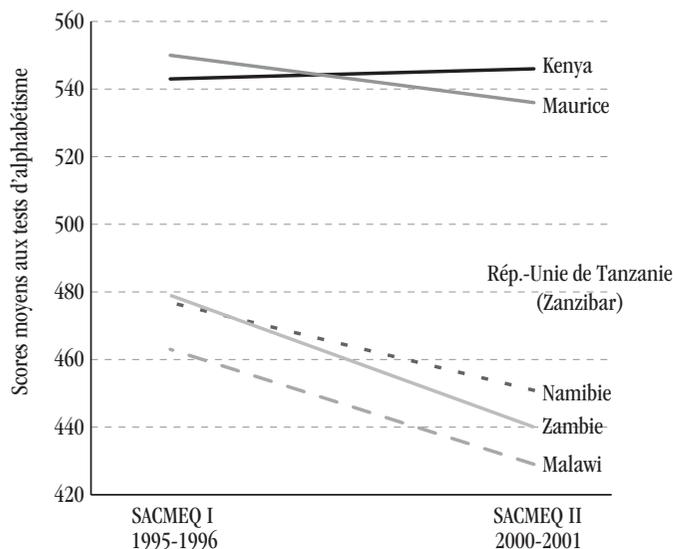
81. Stevenson et Stigler, 1982.

82. Gilmore, 2005, p. 26.

83. Le déclin des scores d'alphabétisation a été attribué à l'augmentation des taux de scolarisation dans ces pays, avec une dégradation concomitante de la qualité de l'instruction.

termes de prise de décision au niveau national, ne valait pas les dépenses qu'il faudrait engager⁸⁴. D'autres peuvent participer parce qu'ils ne veulent pas être considérés comme utilisant des critères inférieurs à ceux des pays de l'OCDE⁸⁵. Les LSEA (internationales et régionales) restent toutefois très utiles, car elles offrent des comparaisons fiables de diverses variables éducatives importantes, comme celles indiquées dans le *tableau 4.2*.

Figure 4.8 Évolution des scores aux tests d'alphabétisme entre SACMEQ I et SACMEQ II



Source : Postlethwaite (2004).

Adapté de : UNESCO, 2004, p. 52.

84. Voir Greaney et Kellaghan (1996) pour une utile présentation d'ensemble sur cette question.

85. Il convient de noter que les donateurs jouent souvent un rôle dans cette décision, en soutenant certaines évaluations dans le cadre d'un « package » d'appui au renforcement des capacités en matière d'évaluation.

Tableau 4.2 Indicateurs de la participation à l'école primaire

Étude	Pays	Cohorte	% d'enfants ayant été scolarisés à un moment quelconque (6-14 ans) ¹	% ayant survécu en 5 ^e année ²	% ayant atteint un niveau minimum de maîtrise ³	TSN dans le primaire pour la période ayant précédé le test ⁴
SACMEQ (1995)						
	Malawi	100	91	31 (34)	7 (22)	69
Test de lecture en 6 ^e année	Maurice	100	99	98 (99)	52 (53)	99
	Namibie	100	97	74 (76)	19 (26)	84
	Rép.-Unie de Tanzanie	100	87	70 (81)	18 (26)	54
PIRLS (2001)						
Test de lecture en 4 ^e année	Colombie	100	98	60 (61)	27 (45)	87
	Maroc	100	99	77 (78)	59 (77)	81
PASEC (milieu des années 1990)						
Test de français en 5 ^e année	Burkina Faso	100	35	25 (72)	21 (83)	28
	Cameroun	100	88	45 (51)	33 (73)	73
	Côte d'Ivoire	100	65	45 (70)	38 (84)	49
	Guinée	100	48	32 (66)	21 (65)	36
	Madagascar	100	78	31 (40)	20 (64)	63
	Sénégal	100	48	42 (87)	25 (59)	51
	Togo	100	82	49 (60)	40 (81)	66

Notes et sources :

1. Les données se rapportent à l'année la plus proche de l'année du test dans chaque pays. Banque mondiale, 2004b.
2. Le pourcentage de la cohorte qui a survécu en 5^e année est calculé en multipliant les taux de survie en 5^e année (entre parenthèses) par le pourcentage d'enfants qui ont été scolarisés à un moment quelconque. Les taux de survie sont tirés du CD-ROM du bilan de l'EPT à l'an 2000 pour SACMEQ I et le PASEC, pour l'année de test ou l'année la plus proche, et du tableau 7 de l'annexe statistique pour le PIRLS.
3. Le pourcentage des élèves ayant atteint le niveau de maîtrise est calculé en multipliant le pourcentage d'enfants soumis au test qui ont satisfait aux normes minimales (entre parenthèses) par le pourcentage d'enfants ayant survécu en 5^e année. Les critères utilisés pour déterminer si un élève a satisfait aux normes minimales diffèrent selon chaque étude et ne sont donc pas comparables (voir l'encadré 3.7). Pour les pays de SACMEQ I, les données sont tirées de Kulpoo (1998), Machingaidze *et al.* (1998), Milner *et al.* (2001), Nassor et Mohammed (1998), Nkamba et Kanyika (1998), Nzomo *et al.* (2001) et Voigts (1998). Pour les pays du PASEC et du PIRLS, les données sont tirées de Bernard (2003) et de Mullis *et al.* (2003), respectivement.
4. (TSN = taux de scolarisation net). Les moyennes ont été calculées pour chaque pays en utilisant les années disponibles. Pour les pays de SACMEQ I et du PASEC, les données sont tirées du CD-ROM du bilan de l'EPT à l'an 2000 et, pour les pays du PIRLS, du tableau 5 de l'annexe statistique.

Adapté d'UNESCO, 2004, p. 258.

Autres aspects du choix des évaluations

Évaluations à enjeux élevés ou faibles

Si certaines évaluations servent les intérêts des élèves, des enseignants, des parents et des responsables politiques en leur fournissant des informations utiles, d'autres mettent en évidence les efforts éducatifs en raison des conséquences attachées à la performance des apprenants. Ce double rôle aboutit au paradoxe des évaluations à « enjeux élevés » qui deviennent un outil de changement. En l'absence de conséquences sérieuses, une évaluation aura du mal à avoir beaucoup d'influence sur un système éducatif ; mais, si le résultat est lourd de conséquences, elle peut être à l'origine d'activités qui sont improductives sur le plan éducatif et qui peuvent même nuire à l'intégrité du système⁸⁶.

La psychologie des évaluations varie selon les acteurs concernés (élèves, enseignants, directeurs d'école, voire décideurs politiques nationaux), en fonction de la façon dont chacun d'entre eux perçoit les résultats. Pour les élèves, tout test peut sembler avoir un enjeu considérable (c'est-à-dire de haute importance). C'est le cas en particulier dans les PMA, où les évaluations sont peu fréquentes et où les tests nationaux ont souvent des conséquences majeures pour les individus⁸⁷. Cela peut induire un effort légitime chez certains élèves, ainsi que des pratiques discutables chez d'autres. Par exemple, l'anxiété et la tricherie redoublent souvent chez les élèves en proportion de l'enjeu du test. Le soutien parental et la préparation des tests (ainsi que l'assistance pédagogique de l'enseignant) augmentent également quand l'enjeu croît, mais pas de façon égale pour tous les élèves.

Enseignants, directeurs d'école ou décideurs nationaux peuvent voir dans les tests un moyen de les juger. Beaucoup de données ont été collectées sur ce sujet, pour tous les types d'évaluations⁸⁸. Il existe clairement des différences transnationales dans la perception de l'importance des enjeux. Dans l'étude PISA, par exemple, certaines données suggèrent que dans des pays comme la Norvège et le Danemark, de nombreux élèves ne sont plus motivés (ou stressés) par ces tests, tandis qu'à Taiwan et à Singapour,

86. Braun et Kanjee, 2006, p. 2.

87. La plupart des pays africains organisent des évaluations nationales à la fin des études secondaires, pilotées par des « jurys d'examen ». Généralement, elles sont correctement financées, en raison de leur enjeu capital pour l'accès à l'enseignement supérieur. Elles servent également de fondement au renforcement des capacités techniques. Merci à L. Wolff pour cette remarque.

88. Pour une analyse, voir Chapman et Snyder, 2000.

les élèves restent très motivés⁸⁹. On possède moins d'informations sur l'importance des enjeux lors des évaluations internationales dans les PMA même si, dans un contexte où les examens nationaux sont importants (le baccalauréat par exemple dans les pays d'Afrique francophone), il est probable que la perception d'un enjeu élevé jouera un rôle important⁹⁰. Selon la documentation de l'EGRA, « L'EGRA *ne doit pas* être utilisée dans les cas de responsabilité à enjeu élevé, qu'elle soit à visée punitive, interventionniste ou de récompense. »⁹¹ Pourtant, comme pour la plupart des autres formes d'évaluation, en particulier dans les environnements où l'évaluation est un événement relativement peu fréquent, il est difficile de faire en sorte que ces évaluations ne soient pas considérées comme ayant un enjeu élevé⁹².

Mesures directes ou indirectes

[D]u fait qu'on dispose de peu de données sur beaucoup de ces dimensions [c'est-à-dire de bonnes mesures des résultats d'apprentissage], les indicateurs indirects de qualité de l'éducation (par exemple, le taux d'enfants survivant à la 5^e année ou taux d'achèvement du cycle primaire) sont souvent devenus la base pour évaluer les progrès substantiels au niveau national (ou l'absence de progrès)⁹³.

Les variables indirectes d'apprentissage existent depuis longtemps, probablement depuis que la scolarisation a été introduite et que l'on a commencé à vouloir savoir ce qui se passait dans la salle de classe. Les taux d'alphabétisation ont joué un rôle important dans cette histoire. À l'époque coloniale, les gouvernements impérialistes n'hésitaient pas à parler de peuples prétendument illettrés et non civilisés qui n'étaient jamais allés à l'école. Quand l'UNESCO recueille des données sur l'alphabétisation dans les pays en développement, nombre d'entre eux déterminent les taux

89. Sjoberg, 2007.

90. La question n'est pas, ici, de savoir s'il vaut mieux des enjeux élevés ou faibles (bien que la plupart des spécialistes de l'évaluation préfèrent généralement les enjeux faibles). Il est plus important de savoir, à des fins de comparaison, qu'il n'y a pas de différences importantes et systématiques (qui pourraient fausser les données) quand les enjeux varient entre les pays ou les autres groupes comparés. Voir également la discussion antérieure sur les examens « publics » nationaux.

91. RTI, 2009, p. 10.

92. Même si ces tests sont considérés par leurs concepteurs comme étant à faible enjeu, ils peuvent ne pas être perçus comme tels sur le terrain. Par exemple, il n'est pas dénué de fondement, pour un enseignant, de penser que les mauvais résultats EGRA de ses élèves pourraient entraîner des répercussions négatives.

93. Benavot et Tanner, 2007, p. 4.

d'illettrisme sur la base du pourcentage d'adultes (de plus de 15 ans) qui ne sont pas allés à l'école⁹⁴. L'école peut donc être une mesure indirecte de l'alphabétisation et, de nos jours, c'est effectivement le cas dans de nombreux pays.

La scolarisation n'est qu'un indicateur de l'apprentissage parmi d'autres. Cherchant à mieux comprendre les types de variables qui influent sur l'apprentissage ou, plus spécifiquement, sur la lecture, les chercheurs se sont focalisés davantage sur les contextes plus généraux de l'apprentissage (comme les variables de milieu socioculturel, voir la section suivante) ainsi que sur la mesure des sous-composantes de la lecture, telles que celles que l'on trouve dans l'EGRA. Comme on le verra au *chapitre 5*, l'EGRA utilise des mesures telles que le nombre de lettres ou de mots qu'un enfant peut lire à haute voix en un temps donné. Certains pensent qu'il s'agit là d'une mesure approximative de la lecture, car peu de personnes, qu'elles soient débutantes ou confirmées, accomplissent réellement ce type de tâche dans leurs activités normales de lecture. Si ce type de mesure peut remplir une fonction importante dans l'EGRA, beaucoup d'items utilisés pour mesurer l'apprentissage de compétences spécifiques (comme nommer les lettres de l'alphabet) n'ont pas la même validité apparente pour mesurer le résultat ultime (comme lire un texte et le comprendre). Il s'agit de mesures indirectes⁹⁵.

L'importance du moment de l'évaluation

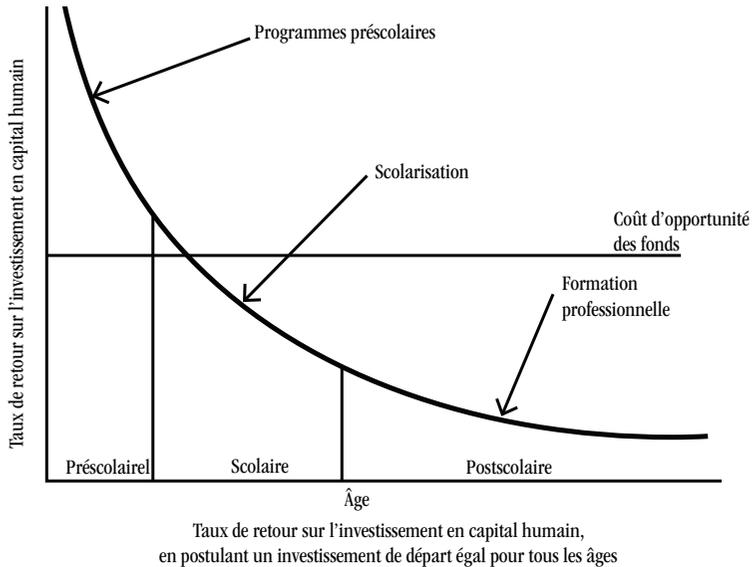
Les évaluations effectuées à l'école sont généralement administrées en se fondant sur deux paramètres clés. Premièrement, les années charnières où les élèves quittent un niveau d'éducation pour le suivant. Ainsi, de nombreux pays font passer des examens nationaux à la fin du cycle primaire, du premier cycle du secondaire et du second cycle du secondaire, afin de déterminer les élèves qui seront acceptés au cycle scolaire suivant. Deuxièmement, certains examens considèrent le niveau de compétence comme un élément cognitif qui mérite davantage d'être évalué chez les élèves. Comme indiqué précédemment, l'enquête PIRLS teste les enfants à la fin de la 4^e année (vers

94. Wagner, 1990, 2001.

95. Pour de nombreux spécialistes en sciences cognitives, une compétence comme nommer les lettres peut être considérée comme étant un élément de l'aptitude à lire et, par conséquent, plutôt comme un précurseur de compétences ultérieures qu'une mesure indirecte. Voir l'analyse détaillée au *chapitre 5*. On peut également dire que de nombreux tests (et éléments de test) sont des mesures indirectes, les tests d'intelligence ou de QI en étant un exemple majeur.

9 ans), stade auquel la plupart des enfants devraient avoir acquis les bases de la lecture, de l'écriture et des mathématiques ; de la même manière, le PASEC, le LLECE et le SACMEQ sont concentrés sur la seconde moitié de la scolarité primaire. En revanche, dans les pays de l'OCDE, le PISA est administré aux jeunes de 15 ans, afin d'évaluer leurs compétences à la fin de la scolarité de base obligatoire.

Figure 4.9 Taux de retour sur les investissements en capital humain, en postulant un investissement de départ égal pour tous les âges



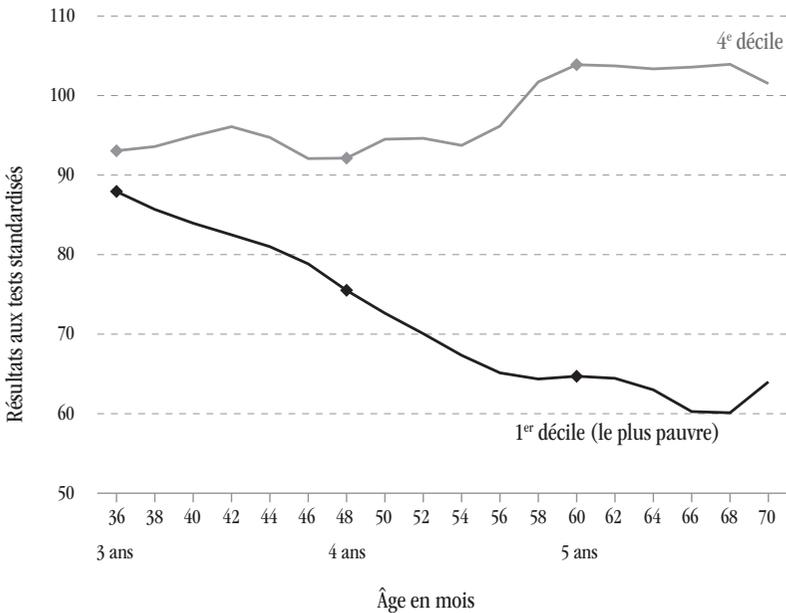
Original tiré de Carneiro et Heckman, 2003, p. 93.

Adapté de : Banque mondiale, 2011.

L'EGRA, pour sa part, se concentre davantage sur la période comprise entre la 1^{re} et la 3^e année, ce qui lui permet de détecter les graves problèmes de lecture beaucoup plus tôt que les autres tests évoqués dans la présente étude. Cette détection précoce est partiellement possible car les évaluations sont individuelles et essentiellement orales. Il y a également une raison politique très importante. Dans le domaine de l'éducation de la petite enfance, l'impact des interventions précoces est de plus en plus reconnu. Ainsi, un dollar investi dans les premières années portera des fruits plusieurs fois au cours

de la vie de l'individu (figure 4.9)⁹⁶. Des études complémentaires montrent, en outre, que les disparités fondées sur la richesse dans le développement cognitif des enfants s'accroissent dans le temps (figure 4.10)⁹⁷. La plupart des enseignants sont d'accord avec le constat que plus vite on détecte et remédie aux problèmes éducatifs (comme c'est aussi le cas dans le secteur de la santé), plus efficace et productive sera l'intervention.

Figure 4.10 Écarts liés à la richesse : résultats aux tests par âge pour le décile le plus pauvre et pour le 4^e décile en Équateur, 2003-2004



Notes : Les résultats des tests employés proviennent du Test de « Vocabulario en Imágenes Peabody », version espagnole du test de vocabulaire de Peabody. Le graphique présenté, version lissée du graphique original (qui apparaît dans le document source), a également été reproduit ailleurs (par exemple, dans Fiszbein *et al.*, 2009, et Banque mondiale, 2006j).

Source : Paxson et Schady, 2005b.

Adapté de : UNESCO, 2010, p. 55.

96. Carneiro et Heckman, 2003 ; Heckman, 2006.

97. UNESCO, 2010.

Variables contextuelles

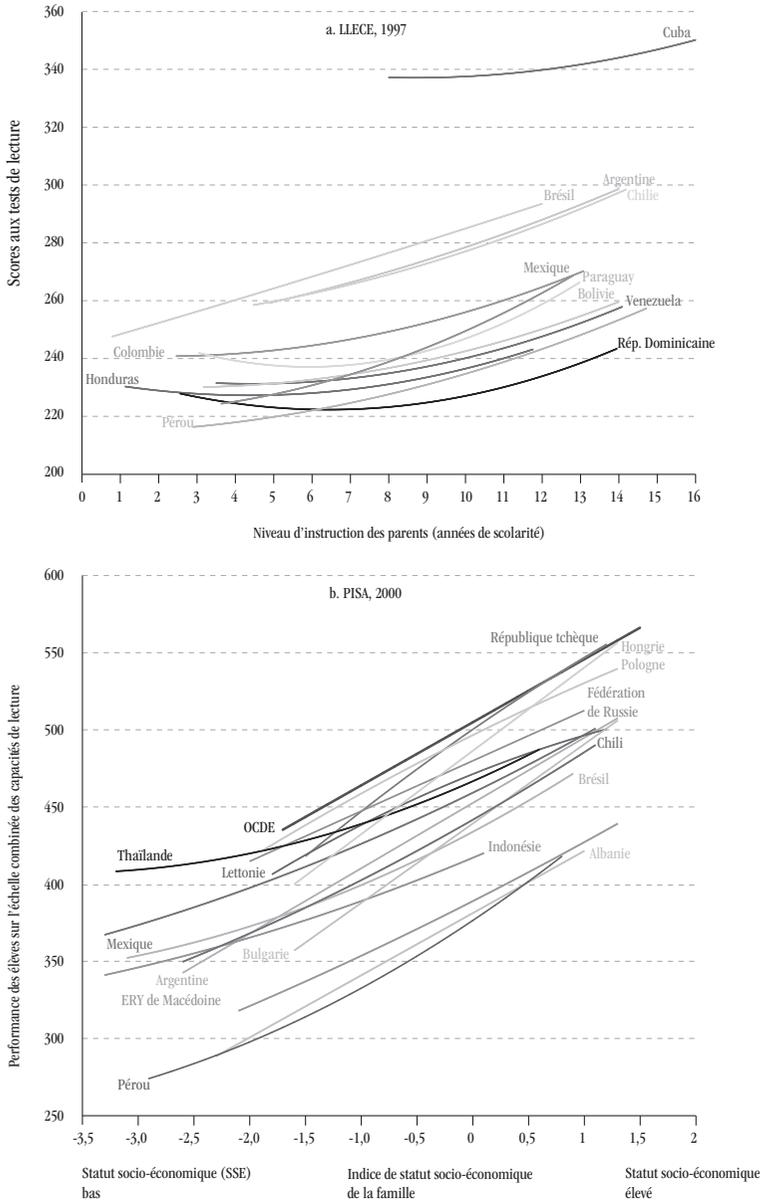
De nombreux facteurs peuvent contribuer à expliquer pourquoi un enfant n'est pas scolarisé, et donc pas inclus dans une évaluation, ou pourquoi les performances des enfants sont meilleures ou moins bonnes sur un ensemble de mesures. Ces variables contextuelles – âge, genre, statut socio-économique, origine ethnique, état de santé, langue maternelle et localisation géographique, entre autres – ont toujours été au centre des explications données par les sciences sociales concernant les extrants éducatifs. Naturellement, dans la plupart des cas, cette explication est en fait une forme de relation, dans la mesure où les données recueillies sur ces variables contextuelles sont considérées comme des statistiques corrélacionnelles. Le statut socio-économique est évidemment une variable de premier ordre et constitue une des explications de la variation des scores des LSEA, notamment dans le LLECE et le PISA (*figure 4.11*).

Il n'est pas facile de réunir des informations sur les variables contextuelles des LSEA. Dans le PIRLS et le PISA, des informations contextuelles (sur des variables telles que le niveau d'éducation des parents, leur emploi, la langue maternelle, la présence de livres à la maison, etc.) sont recueillies auprès des parents au moyen d'un questionnaire à compléter et à renvoyer à un organisme de collecte de données⁹⁸. Dans le cas des compétences en lecture, certains ont affirmé que le fait d'avoir accès à des livres à la maison était un facteur important pour les résultats des tests de lecture, comme illustré par le SACMEQ (*figure 4.12*). On ne peut toutefois invoquer la causalité, car ces données sont corrélacionnelles⁹⁹.

98. Contrairement au PISA, le PIRLS utilise une classification des emplois qui fait partie du questionnaire soumis aux parents.

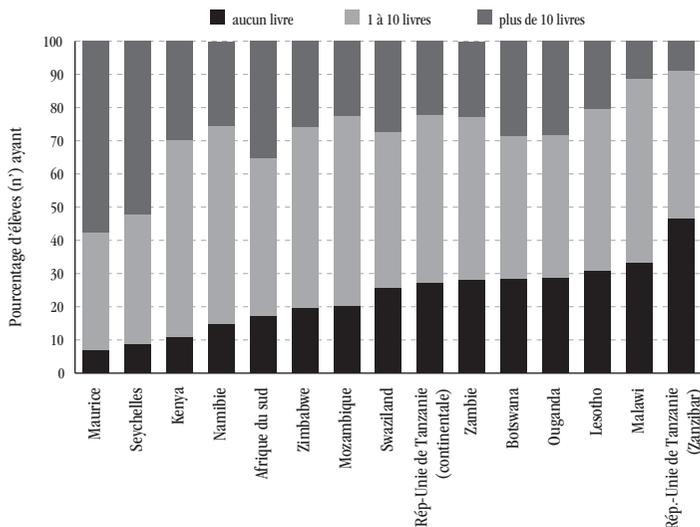
99. Comme l'ont fait remarquer Lapido *et al.* (2009, p. 42) : « ... bien que l'analyse puisse faire apparaître une corrélation positive entre l'apprentissage de l'élève et le nombre de livres à la maison, on ne saurait conclure – même en prenant en compte d'autres variables – qu'il y a un lien de cause à effet entre le nombre de livres et les résultats de l'élève. Bien que l'accès aux livres puisse être important, l'apprentissage de l'élève n'est probablement pas influencé directement par la présence de livres, mais par les caractéristiques d'un environnement qui aime les livres, comme celui où les parents accordent une grande valeur à la performance scolaire, donnent à l'enfant des conseils et une aide scolaires, incitent l'enfant à explorer des idées et à discuter d'événements, et placent très haut la barre et les attentes en matière de performance scolaire ».

Figure 4.11 Facteurs contextuels et compétences en lecture



Adapté de : UNESCO, 2004, p. 139.

Figure 4.12 Nombre de livres que des élèves de 6^e année ont déclaré posséder chez eux dans 15^e systèmes éducatifs de pays africains du SACMEQ, 2000



* Il y a 15 systèmes éducatifs pour 14 pays. Toutes les données proviennent des archives de SACMEQ II (2000), sauf celles du Zimbabwe qui proviennent de celles de SACMEQ I (1995).

Source : Ross *et al.*, 2004.

Adapté de : UNESCO, 2004. p. 208.

Méthodologies de collecte de données

L'une des principales difficultés des évaluations à grande échelle de l'apprentissage (LSEA) est de déterminer la quantité et le type de données qu'il faut recueillir. Il n'est pas si facile de réunir exactement la quantité nécessaire de données. Ce que certains qualifient de « bon dimensionnement » des données à recueillir a été plus récemment appelé « conception axée sur les données »¹⁰⁰. L'idée, essentielle pour l'approche SQC décrite précédemment, est de tenter de recueillir suffisamment de

100. « L'idée de base de la conception axée sur des données est que les concepteurs doivent « travailler en amont », en déterminant d'abord les demandes qu'ils voudraient que les utilisateurs soumettent à propos de l'évaluation, ainsi que les données requises pour étayer ces demandes. Ils peuvent ensuite créer les exercices (items, enquêtes, tests de performance, etc.) qui permettront d'obtenir des élèves les réponses voulues, les rubriques d'évaluation utilisées pour transformer ces réponses en données pertinentes, et les modèles de mesure qui cumulent ou synthétisent ces données » (Braun et Kanjee, 2006, p. 13).

données pour résoudre une difficulté concrète, mais sans risquer de rendre les informations collectées moins pertinentes à cause du coût en temps et en énergie qu'elles entraînent. Cette approche possède beaucoup de points communs avec le développement de tests dits de « format réduit », qui consiste à réduire la taille de tests plus longs, avec les divers risques statistiques que cela comporte en termes de validité et de fiabilité¹⁰¹. On trouvera ci-après une description succincte des méthodes les plus courantes.

Enquêtes

Chaque LSEA décrite auparavant est menée au niveau des écoles. D'autres enquêtes sont également réalisées dans le cadre des bureaux nationaux de recensement. L'accent est mis sur l'échantillonnage des ménages selon des paramètres démographiques, avec parfois des tests psychométriques et des techniques d'analyse. Les efforts pour rendre ces enquêtes comparables au niveau international sont au moins aussi complexes que pour les LSEA, voire nettement plus, à cause de la nécessité de se rendre dans chaque foyer. Un bon exemple dans le domaine de la lecture est l'Enquête internationale sur l'alphabétisation des adultes (EIAA), entreprise par l'OCDE en 1995¹⁰². À l'instar des LSEA, ces enquêtes exigent une extrême rigueur au niveau de la sélection des échantillons et du temps requis pour évaluer l'apprentissage.

Évaluation des programmes

L'évaluation des programmes est probablement le type de collecte de données le plus courant et le plus varié. De nombreuses agences procèdent à des évaluations au niveau local et peuvent également évaluer les programmes nationaux. Il y a un large choix de méthodologies, depuis celles qui déterminent si les fonds ont été correctement utilisés, jusqu'à celles qui mesurent les résultats d'apprentissage, ou qui évaluent l'implication de la communauté. Malheureusement, la plupart de ces évaluations sont isolées, dans le sens où les efforts visant à établir un lien entre les différentes évaluations des programmes sont faibles ou inexistantes. Par conséquent, il est difficile d'obtenir un résultat s'apparentant à une science cumulative et

101. Smith *et al.*, 2000. Cette étude décrit la façon dont divers tests connus ont donné lieu à des versions plus courtes, et présente des suggestions méthodologiques pour améliorer les versions abrégées.

102. OCDE/Statistique Canada, 1997. Voir le *chapitre 8* pour plus d'informations sur l'EIAA.

rare sont les évaluations des programmes effectuées dans les PMA ayant porté sur les compétences en lecture des enfants ou des adultes¹⁰³.

Psychométrie et théorie des réponses aux items

La psychométrie est la théorie et la technique des mesures psychologiques ou éducatives. Toutes les LSEA évoquées ici utilisent la psychométrie pour recueillir des données au moyen de tests de compétences et pour procéder à leur analyse. En outre, la plupart des LSEA emploient la technique statistique dite de « théorie des réponses aux items » (TRI)¹⁰⁴. L'approche TRI élargit la couverture globale des tests de compétence, en permettant d'inclure globalement un plus grand nombre d'items dans l'évaluation, tout en diminuant leur nombre par élève. Cela rend également possible l'utilisation de textes longs, un article de journal par exemple, pour évaluer la compréhension écrite. Dans les évaluations sans TRI (comme le PASEC ou l'EGRA), tous les élèves répondent à un ensemble complet d'items, ce qui permet une comparaison transparente entre des ensembles identiques d'items, mais limite en même temps l'étendue et la profondeur de l'évaluation. La TRI présente toutefois des inconvénients, en particulier pour les PMA qui se lancent dans un programme d'évaluation. La notation, l'échelonnement des scores et l'administration (notamment l'impression et la distribution) sont plus complexes. De même, les analyses portant sur certains élèves ou sur les données d'écoles peuvent s'avérer problématiques et nécessiter du personnel plus spécialisé¹⁰⁵. Comme toute technique statistique, la TRI, telle qu'elle est employée dans les évaluations internationales, n'est pas exempte de critiques^{106, 107}.

103. Dans le domaine de l'alphabétisation des adultes, il n'y a eu que quelques évaluations multiprogrammes significatives, comme Carron *et al.*, 1989 ; Okech *et al.*, 2001 ; Nordtveit, 2004.

104. Voir Hambleton *et al.*, 1991.

105. Voir Greaney et Kelleghan, 2008, p. 42.

106. Voir, par exemple, Goldstein, 2004 ; Goldstein *et al.*, 2007 ; Mislevy et Verhelst, 1990.

107. Les modèles statistiques psychométriques de la TRI reposent sur le postulat que la réponse de l'élève à un item ne dépend pas de ses réponses aux autres items de l'instrument d'évaluation. Cependant, comme toutes les LSEA contiennent des textes avec des questions multiples, si l'élève ne comprend pas un seul texte, cette incompréhension affectera de nombreuses autres réponses. Si cette approche est courante pour des raisons pratiques, elle empêche toute indépendance dans le traitement des résultats. Pour une analyse de ces questions, voir Dicks et Vrignaud, 1995 ; Wainer et Thissen, 1996.

Essais comparatifs randomisés

La randomisation comporte de nombreux pièges potentiels. Elle peut être coûteuse, exiger une surveillance substantielle et sérieuse pour en garantir l'intégrité, et nécessiter du temps pour produire des conclusions utiles. [...] Parce que la mise en œuvre et l'évaluation de certaines interventions majeures prend beaucoup de temps, chercheurs et responsables politiques doivent mettre en balance leur désir de résultats rapides et leur désir de solutions globales et importantes¹⁰⁸.

Depuis plusieurs décennies, les essais comparatifs randomisés (ECR) connaissent un succès croissant dans le domaine de la recherche en éducation. Ce succès semble être dû, d'une part, aux liens de plus en plus importants entre sciences sociales, sciences médicales et éducation et, d'autre part, à la frustration des responsables politiques devant la myriade de conclusions complexes (et parfois contradictoires) qui constituent la base de nombreuses questions d'éducation importantes. Les études avec ECR commencent à être menées depuis quelques années seulement dans les pays en développement, malgré leur fort potentiel lorsque des tests rapides et fréquents sont possibles. Dans une étude de ce type menée en Inde, on s'est aperçu que nommer un second enseignant dans les écoles qui n'en avaient qu'un faisait progresser la scolarisation des filles, mais avait un impact faible ou nul sur les résultats des élèves aux tests. Cette expérience a démontré de façon assez décisive que le soutien d'enseignants supplémentaires mal formés ne favorisait en rien l'amélioration des résultats des élèves dans le contexte étudié¹⁰⁹. Avec l'arrivée d'outils d'évaluation crédibles et relativement rapides comme l'EGRA, les études avec ECR se multiplieront probablement dans les années à venir (voir au *chapitre 6* l'exemple d'une récente étude de terrain de l'EGRA).

Fréquence des tests

Avec le suivi des progrès [en lecture], les élèves sont évalués au minimum trois fois par an, mais généralement plus souvent (par exemple toutes les semaines, tous les mois ou tous les trimestres) au moyen d'autres formes de test. Le but est d'estimer les taux de progression en lecture, de repérer les enfants qui ne progressent pas

108. Bettinger, 2006, p. 67.

109. Banerjee et Kremer, 2002. Généraliser cette expérience à d'autres situations et contextes soulèverait évidemment de nombreux débats, comme c'est le cas dans de nombreuses études avec ECR. Mais, dans la mesure où les variables (indépendantes et dépendantes) restent presque constantes ou contrôlées, la possibilité de généraliser devient plus importante.

comme il faut et qui ont besoin d'un soutien, et de comparer l'efficacité des différentes formes d'instruction pour un élève donné ...¹¹⁰.

La fréquence des évaluations nationales varie d'un pays à l'autre, depuis une fois par an jusqu'à une fois tous les 10 ans. On pourrait être tenté d'évaluer les résultats dans les mêmes matières et pour la même population tous les ans, mais cette fréquence est inutile et très dispendieuse ...¹¹¹.

La plupart des pays de l'OCDE, qui disposent de ressources fiscales et humaines relativement abondantes, évaluent régulièrement les élèves, comme la citation ci-dessus le décrit, s'agissant de l'apprentissage de la lecture. Pour pouvoir intervenir dans les cas où l'élève (ou l'école) montre des signes de retard d'acquisition de la lecture par rapport aux normes, des ressources supplémentaires sont mobilisées en temps opportun, souvent sous la forme d'évaluations très fréquentes. Dans les pays en développement, ainsi que dans les régions peu nanties de l'OCDE, de telles ressources ne sont généralement pas disponibles.

Si les évaluations internationales, régionales et nationales ne sont pas administrées tous les ans ou tous les deux ans, elles risquent d'avoir un impact politique plus limité. Si le but est de resserrer le lien entre les conclusions et les politiques qui peuvent être mises en œuvre pendant le cycle scolaire annuel, ou dans le cadre du mandat d'un ministre de l'Éducation, une plus grande fréquence s'impose. Pour cela, il faudra probablement recourir à des instruments tels que les outils SQC hybrides, dont le délai d'exécution et le coût plus modeste permettent une plus grande fréquence d'administration.

Disparités entre pays

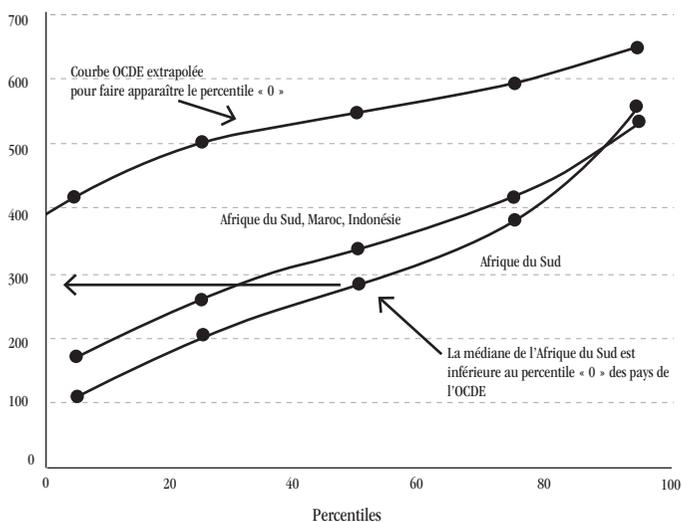
Les rapports statistiques internationaux sur l'éducation (comme ceux produits par l'ISU, l'UNICEF ou la Banque mondiale) reposent généralement sur les données des rapports nationaux, qui peuvent être recueillies de nombreuses manières différentes. Les LSEA, au contraire, et c'est un de leurs avantages, permettent d'établir un classement des nations (comme c'est le cas pour le PISA et le PIRLS). Cependant, comme il a déjà été évoqué, appliquer à des populations très différentes une échelle commune d'échantillonnage des compétences peut être problématique. Dans l'étude PIRLS 2006 sur les performances en lecture (*figure 4.13*), par exemple, le score médian des élèves de 4^e année en Afrique du Sud est inférieur au percentile « 0 » des nations à hauts revenus de l'OCDE.

110. Kame'enui *et al.*, 2006, p. 6.

111. Greaney et Kellaghan, 2008, p. 43.

Des disparités aussi grandes suscitent de vives inquiétudes quant au fossé que les PMA doivent combler pour rattraper les pays à hauts revenus. Dans l'étude pré-PIRLS 2011, des référentiels plus bas ont été utilisés afin d'avoir une efficacité statistique plus explicative dans le bas de l'échelle. Des effets planchers et plafonds sont possibles à tout moment, quand les résultats en termes de compétences varient de façon significative dans l'échantillon de population. Par exemple, les scores de l'EGRA en anglais des élèves de la Tanzanie rurale seront probablement beaucoup plus bas que ceux des élèves anglophones du même âge (ou de la même classe) qui vivent dans la banlieue de Washington, DC. Globalement, les disparités peuvent être un outil puissant de mise en évidence de différences, mais elles posent un problème constant de référentiels adaptés aux divers contextes.

Figure 4.13 Pourcentage d'élèves de 4^e année dans l'enquête PIRLS 2006



Adapté de : Crouch, 2009.

Crédibilité des évaluations

Il existe diverses façons de concevoir la crédibilité d'une évaluation. Généralement, les professionnels de l'évaluation définissent la crédibilité comme étant une combinaison de validité et de fiabilité. Il faut toutefois bien comprendre que la crédibilité, au sens non statistique, implique plus que les outils statistiques spécifiques dont disposent les concepteurs des tests. Ceci est dû en grande partie au fait que de nombreuses décisions difficiles

concernant la crédibilité sont prises *avant* d'employer les tests statistiques. Par exemple, une évaluation est-elle crédible si la plupart des enfants les plus pauvres en sont exclus ? Une évaluation est-elle crédible, si l'agent de recensement ne parle pas la langue de l'enfant ? Ce ne sont pas simplement des choix internes au test, mais plutôt des choix liés au contexte dans lequel l'évaluation a lieu. Partant du principe que la plupart des difficultés ci-dessus peuvent être raisonnablement réglées, il faut maintenant examiner les notions plus traditionnelles de validité et de fiabilité.

Validité

La validité est un jugement évaluatif intégré portant sur la mesure dans laquelle les données empiriques et les justifications théoriques confirment l'adéquation et la pertinence des conclusions et des initiatives fondées sur les résultats des tests ou autres modes d'évaluation¹¹².

La validité d'un instrument d'évaluation est la mesure dans laquelle les items d'un test peuvent être liés de façon crédible à la raison d'être conceptuelle de l'outil d'évaluation. Ainsi, les questions d'un test de type QCM ont-elles effectivement un rapport avec l'aptitude d'un enfant à lire, ou avec sa capacité à se souvenir de ce qu'il a lu avant ? La validité peut varier considérablement en fonction de la situation et de la population : un test valide à Londres peut ne pas l'être à Lahore¹¹³. Un test de lecture utilisé avec succès pour un groupe de locuteurs parlant leur langue maternelle peut être parfaitement inapproprié pour évaluer des enfants, dont cette même langue est leur seconde langue.

En ce qui concerne les évaluations à grande échelle de l'apprentissage (LSEA) internationales, la validité de leurs contenus a fait l'objet d'un certain nombre de critiques s'agissant du choix des items et de leur adéquation aux cultures et aux systèmes éducatifs locaux¹¹⁴. Il semble que les tests régionaux fassent un peu mieux dans ce domaine, car ils utilisent souvent la matière des programmes scolaires nationaux comme items des tests¹¹⁵. La traduction des LSEA internationales reste problématique, car on ne peut jamais être

112. Messick, 1989, p. 13 ; cité dans Braun et Kanjee, p. 15.

113. Braun et Kanjee, 2006, p. 15.

114. Sjoberg (2007) affirme que certains items de test s'écartent sensiblement du but déclaré du PISA, à savoir évaluer les compétences de la main-d'œuvre. Howie et Hugues (2000) ont constaté que le TIMSS ne couvrait qu'une toute petite fraction (18 %) du programme de sciences de la 7^e année en Afrique du Sud, mais au moins 50 % de la 8^e année.

115. Ross *et al.*, 2005.

certain qu'un item traduit aura les mêmes propriétés statistiques qu'un mot autochtone choisi de façon indépendante¹¹⁶. Cette question est devenue une évidence lors de l'Enquête internationale sur l'alphabétisation des adultes (EIAA) de l'OCDE mentionnée précédemment, quand la France s'est retirée de l'étude en dénonçant un parti pris dans la traduction¹¹⁷.

Fiabilité

La fiabilité est généralement mesurée de deux façons. Globalement, c'est la mesure dans laquelle le score d'un individu à un test reste à peu près constant chaque fois qu'il repasse ce test (ou un test équivalent). Une grande fiabilité implique généralement que le classement d'individus ayant passé un test donné serait sensiblement le même s'ils repassaient ce test. Dans la psychométrie de l'évaluation, il n'est pas rare d'obtenir un taux de fiabilité relativement élevé des LSEA entre un premier et un second test¹¹⁸. Ce résultat est dû en grande partie au fait que la fonction cognitive humaine est extrêmement stable, comme on le sait depuis la création des tout premiers tests d'intelligence. Une autre manière, plus facile, de mesurer la fiabilité consiste à considérer la fonction interne des items du test : les items de chaque partie de l'évaluation sont-ils fortement reliés les uns aux autres ? Il s'agit de ce qu'on appelle la fiabilité inter-items (également connue sous le nom de « coefficient alpha de Cronbach »).

La fiabilité renseigne peu sur la validité de l'instrument, qui consiste à parvenir à un consensus sur l'adéquation de l'instrument pour évaluer des résultats scolaires. Néanmoins, elle est cruciale pour mener toute LSEA, et l'incapacité d'obtenir un degré relativement élevé de fiabilité est souvent

116. Voir Hambleton et Kanjee (1995) pour une discussion sur les questions de traduction dans les évaluations internationales.

117. Blum *et al.*, 2001. La France a participé aux EIAA de 1995 et 1998. Apparemment, il y avait également des différences entre les traductions en français destinées à la Suisse romande et à la France. Voir également Carey (2000) et l'analyse plus poussée au *chapitre 8*.

118. Dans les évaluations psychométriques, il n'est pas difficile d'obtenir des indices de forte corrélation de la fiabilité pour un ensemble relativement différencié d'items, composé de 20 ou 30 items (voire plus) et adapté au niveau de l'apprenant. Les individus sont généralement sérieux quand ils passent des tests.

le signe de sérieux effets plafonds ou planchers¹¹⁹. Comme on le verra au *chapitre 6*, les tests de lecture (tels que l'EGRA) affichent souvent des effets plafonds et planchers importants, en particulier chez les jeunes enfants, dont les courbes d'apprentissage individuelles subissent de fortes variations. Ces effets peuvent causer de sérieux problèmes, tant pour l'élaboration des tests que pour l'interprétation des résultats.

Choisir un type d'évaluation

Toute initiative visant à améliorer la pratique des évaluations doit tenir compte des évaluations officielles en usage¹²⁰.

Aujourd'hui, tous les pays se prêtent à une évaluation de leur système éducatif. La plupart d'entre eux recourent essentiellement à des formes d'évaluation nationale. Est-il possible de les améliorer ? Faut-il ajouter d'autres évaluations à l'ensemble d'informations et de données que les agences s'efforcent d'utiliser ? Compte tenu de la multiplication des évaluations dans le monde, la réponse semble être affirmative, ainsi qu'en témoignent les faits sur le terrain.

Comment un organisme (local, national ou international) choisit-il de nouvelles évaluations ? De nombreuses réponses sont possibles. Une méthode, comme l'affirme la citation ci-dessus, consiste à s'appuyer sur les évaluations déjà en usage, ne serait-ce que pour des raisons de capacité en ressources humaines. Mais, comme on le verra dans les deux chapitres suivants, il faut aussi prendre en compte les objectifs spécifiques de l'évaluation. Par exemple, est-il suffisant de savoir qu'il y a des problèmes de lecture dans une population d'élèves de 5^e année dans un pays ou un groupe démographique donné ? Le ministère veut-il savoir *pourquoi*, ou quand ce problème est apparu, voire *comment* tenter d'y remédier ? Si c'est le cas, il faudra regarder de plus près les nouvelles évaluations des compétences en lecture. On trouvera une étude approfondie de cet aspect dans les *chapitres 5* et *6*. Qu'en est-il, par ailleurs, des questions de coût ? Le ministère peut-il se permettre d'engager des dépenses supplémentaires, alors qu'il est déjà soumis à des restrictions budgétaires ?

119. Des effets plafonds se produisent quand un test est tellement facile qu'un nombre important d'apprenants répondent correctement à la plupart ou à la totalité des items, ce qui réduit la variance des scores et fait donc chuter le coefficient de corrélation. De la même manière, des effets planchers se produisent s'il y a trop de scores à zéro ou près de zéro, ce qui fait également baisser le coefficient de corrélation.

120. Braun et Kanjee, 2006, p. 25.

Il est essentiel de prendre en considération l'ensemble des divers coûts d'un large éventail d'options d'évaluation, ce qui fait l'objet du *chapitre 7*. Faire des choix n'est jamais facile et la tentation de continuer à marcher sur les pas de ses prédécesseurs est réelle, parfois même efficace. Néanmoins, des méthodes alternatives peuvent l'emporter de loin sur le simple maintien d'un système d'évaluation utilisé précédemment, et se révéler à la fois plus riches en informations et moins coûteuses à mettre en œuvre.

5. ÉVALUER LES COMPÉTENCES EN LECTURE DES ENFANTS

L'évaluation des compétences en lecture est un facteur essentiel pour un apprentissage réussi. En effet, pour être efficace, l'enseignement doit être autant que possible calibré en fonction des connaissances, des compétences et des centres d'intérêt des élèves. Partout dans le monde, les enseignants qui obtiennent de bons résultats utilisent des tests de lecture, pour des finalités multiples. Ils peuvent recourir à des examens nationaux à enjeu élevé ou à des tests normalisés pour identifier les élèves susceptibles de bénéficier d'une promotion, d'une bourse ou d'une distinction. Ils peuvent utiliser les examens de fin d'année pour l'admission dans un établissement ou le passage dans une classe supérieure. Dans les écoles qui obtiennent de bons résultats (le plus souvent dans les pays riches), les enseignants peuvent soumettre les enfants à des évaluations informelles individualisées en début d'année scolaire pour vérifier s'ils savent lire et écrire et ce qu'ils peuvent lire et écrire. Ils utilisent aussi des exercices de lecture orale pour déterminer les aptitudes des élèves à décoder les mots et à lire avec aisance et expression. Ils peuvent faire passer des tests d'aptitude pour diagnostiquer les points forts et les points faibles de leurs élèves ; ils peuvent observer les stratégies de compréhension lors d'activités quotidiennes de lecture. Ils peuvent également mettre au point des auto-évaluations pour permettre aux élèves de suivre eux-mêmes leurs progrès. Certains enseignants créent un journal pour observer l'évolution de l'écriture des enfants, de leurs centres d'intérêt en matière de lecture et de leurs approximations phonétiques des mots.

Compte tenu des nombreux types d'évaluation utilisés à diverses fins, pour être efficaces, les enseignants actuels ont besoin, quel que soit le pays où ils exercent, de savoir quand et pourquoi se servir des divers outils et techniques d'évaluation¹²¹.

Pourquoi la lecture ?

Pouvoir lire et comprendre un texte simple est l'une des compétences les plus fondamentales qu'un enfant peut acquérir. Sans aptitudes de base en lecture et en écriture, un enfant a peu de chances de sortir du cycle intergénérationnel de la pauvreté. Or, dans de nombreux pays,

121. Shepard, 2000.

même en étant scolarisés pendant six ans parfois, certains élèves sont incapables de lire et de comprendre un texte simple¹²².

La lecture, indicateur principal de la qualité de l'éducation, est une composante essentielle et précoce des programmes scolaires dans le monde entier. Dans la plupart des pays, des difficultés de lecture à l'école primaire sont l'un des indicateurs les plus sûrs de futurs handicaps dans le domaine éducatif, social et économique, car savoir lire et écrire est la voie d'accès à l'emploi et à l'ascension sociale. Par exemple, l'analyse des qualifications requises aux États-Unis entre 1969 et 1998 fait apparaître un déclin du travail manuel non spécialisé et une augmentation équivalente des compétences liées à la résolution de problèmes et à la communication¹²³. L'évolution générale vers l'économie du savoir exige des travailleurs plus instruits, et la lecture est un fondement essentiel pour poursuivre des études, même si cette évolution peut être difficile à discerner dans les villages pauvres des PMA¹²⁴.

La science de l'apprentissage de la lecture

La science se construit avec des faits, tout comme une maison avec des pierres ; mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison¹²⁵.

La plupart des théories actuelles sur la lecture sont fondées sur l'étude de la langue anglaise dans les pays de l'OCDE¹²⁶. Depuis quelques décennies, les débats scientifiques concernant la lecture portent avant tout sur l'acquisition de la lecture en langue anglaise. Très peu de travaux ont été menés sur les autres langues (principalement européennes), avec une pénurie de recherche sur les langues non européennes des pays en développement. C'est pourquoi les théories actuelles sur la lecture, qui reposent principalement sur des études effectuées dans les pays de l'OCDE, ne sont peut-être pas vraiment adaptées ou applicables au contexte des PMA qui ont des langues et des systèmes d'orthographe spécifiques, souvent multiples. Si certains aspects des théories sur la lecture peuvent avoir un caractère universel, beaucoup de chercheurs et d'enseignants pensent qu'il faudrait les adapter pour tenir

122. RTI, 2009, p. 1.

123. Levy et Murnane, 2004.

124. OCDE, 1997.

125. Henri Poincaré, vers 1905.

126. Share, 2008.

compte de la diversité des cultures, des langues, des systèmes d'écriture, des expériences, des situations familiales et des contextes scolaires locaux¹²⁷.

Dans l'analyse qui suit, cette diversité culturelle et linguistique doit être prise en compte. En même temps, il est malgré tout possible de tirer des informations utiles de la recherche mondiale en matière de lecture, même en prenant acte du fait qu'elle est encore balbutiante et incomplète.

Environnements d'alphabétisation favorables et défavorables

Qu'il vive dans un pays en développement ou riche, un enfant peut grandir dans des contextes extrêmement différents, y compris à l'intérieur des mêmes frontières nationales. Il y a des enfants vivant à Denver, à Düsseldorf, à Dhaka et à Delhi, qui ont des parents peu ou pas instruits, qui savent à peine lire et écrire, dont les enseignants ne savent pas vraiment comment enseigner à lire, ou qui parlent à la maison une langue différente de celle qui est enseignée à l'école. Ces enfants vivent dans des *environnements (d'alphabétisation) défavorables* (PSE – *poorly-supported (literacy) environments*)¹²⁸. La plupart d'entre eux vivent dans des pays en développement ou dans la pauvreté.

Dans ces mêmes villes et pays, on peut aussi trouver des enfants qui sont élevés dans des familles instruites, fréquentent de bonnes écoles où enseignent des professeurs expérimentés, et qui ont accès chez eux à divers textes imprimés (et sous forme électronique). Ces enfants vivent dans des *environnements (d'alphabétisation) favorables* (WSE – *well-supported (literacy) environments*). La plupart d'entre eux habitent dans des pays à revenus élevés, comme les pays de l'OCDE. Il existe de grandes variations entre PSE et WSE, mais cette distinction est importante car l'opposition entre pays moins développés et pays industrialisés ne rend pas compte

127. Pour une étude instructive sur les questions d'universalité de la lecture, voir Perfetti (2003).

128. L'expression « environnement d'alphabétisation » est employée de diverses façons ; voir Easton (2010) pour une analyse récente. Comme il le fait remarquer, la plupart de ces désignations sont très mal définies, voire pas du tout. Dans le contexte actuel, les expressions « environnements favorables/défavorables à l'apprentissage de la lecture et de l'écriture » sont employées principalement pour distinguer tout un éventail de facteurs présents entre différents pays et au sein d'un même pays. Bien que ces environnements soient présentés de façon dichotomique, il faut bien comprendre qu'il s'agit non seulement d'une variable continue, mais aussi d'une variable qui doit être un terme composite multicritères. Cette expression est utilisée ici comme estimation approximative des divers types de contextes dans lesquels vivent les enfants dans le monde.

de façon adéquate de la variété des environnements d'alphabétisation (et d'éducation) dans lesquels les enfants grandissent. La distinction entre PSE et WSE peut être utile pour mieux comprendre pourquoi certains enfants apprennent à lire correctement et d'autres n'y parviennent pas. Le point important, c'est que certaines divisions classiques dans les comparaisons *entre* pays pauvres et pays riches ont tendance à masquer les très grandes variations (et inégalités) des performances en lecture qui peuvent exister à l'intérieur des pays.

Le contexte de l'apprentissage précoce

[En] Guinée francophone, seul un élève sur 10 connaissait l'alphabet en entier à la fin de la 2^e année de scolarité et, en moyenne, les élèves étaient capables de lire 4 mots sur les 20 présentés ... Au Pérou, seuls 25 % des enfants de 1^{re} année et 45 % de ceux de 2^e année échantillonnés étaient capables de lire un seul mot ... Sans maîtrise de la lecture, les élèves ne peuvent ni apprendre dans les manuels, ni répondre de façon fiable à des tests de performance normalisés. Les enfants les plus pauvres prennent du retard en 1^{re} année et ne réussissent généralement pas à le rattraper¹²⁹.

Dans le monde industrialisé et, plus généralement, dans les environnements d'alphabétisation favorables (WSE), les enfants de trois et quatre ans écoutent généralement des histoires que leur lisent ou racontent des adultes. Les parents et les enfants posent souvent des questions pendant qu'ils construisent ensemble du sens à partir des livres. Les enfants reconnaissent également les textes écrits présents dans leur environnement, commencent à griffonner et à écrire, font rimer les mots, jouent à des jeux de langage et reconnaissent des mots familiers comme leur nom – le tout vers quatre ou cinq ans¹³⁰.

129. Abadzi, 2008, p. 4.

130. Snow *et al.*, 1998.

Ces précurseurs de l’alphabétisme peuvent être considérés comme des concepts cognitifs (ou métacognitifs¹³¹) et des compétences (ou activités d’alphabétisme « émergentes ») qui se rapprochent de la lecture et de l’écriture conventionnelles, mais ne sont généralement pas enseignés directement aux enfants d’âge préscolaire¹³². Dans ces contextes, les premières approches de la lecture et de l’écriture sont intégrées dans les échanges parent-enfant, qui sont principalement des expériences sociales et affectives. La lecture au moment du coucher est courante dans certains WSE, mais avec des différences culturelles importantes. Par exemple, dans de nombreux pays asiatiques, les parents apprennent à lire à leurs enfants de façon plus didactique¹³³. Dans les pays les plus pauvres de la planète, où il y a moins de livres à la maison et où les deux parents sont souvent analphabètes ou peu alphabétisés (*tableau 5.1*), les jeunes enfants sont souvent privés de ces possibilités d’apprendre à utiliser le texte¹³⁴.

-
131. Les stratégies métacognitives précèdent généralement les compétences en lecture, parce qu’elles peuvent engendrer chez les enfants la compréhension de la lecture et du texte. Mais il existe parfois une confusion entre compétences et stratégies (métacognitives). Dans une étude relativement récente, Afllerbach *et al.* (2008) écrivent : « Les stratégies de lecture sont des tentatives délibérées et orientées vers un objectif afin de contrôler et modifier les efforts du lecteur pour décoder le texte, comprendre les mots et construire les significations du texte. Les compétences en lecture sont des actions automatiques qui ont pour résultat le décodage et la compréhension avec rapidité, efficacité et fluidité, et qui interviennent généralement sans que le sujet soit conscient des composantes ou du contrôle qui entrent en jeu » (p. 368).
132. Il y a, bien entendu, de nombreuses autres compétences cognitives qui se développent naturellement chez un jeune enfant normal. L’une des plus couramment citées est la mémoire de reconnaissance qui apparaît chez les enfants à un âge précoce (Wagner, 1980) et qui est relativement invariante d’une culture à l’autre.
133. Mee et Gan (1998) ont constaté que seuls 31 % des parents singapouriens font la lecture à leurs enfants, mais que 69 % essaient d’apprendre à lire à leurs enfants. Par ailleurs, 66 % des parents dont les enfants apprennent à lire achètent des cahiers d’examen blancs pour les faire travailler.
134. Cela ne veut évidemment pas dire que les compétences métacognitives font défaut chez les enfants et dans les familles qui vivent dans des environnements d’alphabétisation défavorables (PSE). Raconter des histoires oralement (sans texte écrit) est une longue tradition dans presque toutes les cultures et tient (et a tenu) une place particulièrement importante dans les sociétés de tradition orale (voir Vansina, 1965, pour une référence ancienne). Il y a eu beaucoup de recherches aux États-Unis sur le rôle des histoires que l’on raconte en s’appuyant sur un livre (par exemple, Heath, 1982), mais moins sur le rôle des histoires de tradition orale dans le développement de la lecture dans les PMA (ou les PSE).

Tableau 5.1 Estimation des taux d'analphabétisme et d'alphabétisme des adultes (population de plus de 15 ans) par région, 1990 et 2000-2004

	Évolution entre 1990 et 2000-2004						
	Nombre d'analphabètes (milliers)		Taux d'alphabétisme (%)		Nombre d'analphabètes (milliers)		Taux d'alphabétisme (points de pourcentage)
	1990	2000-2004	1990	2000-2004	(milliers)	(%)	
Monde	871 750	771 129	75,4	81,9	-100 621	-12	6,4
Pays en développement	855 127	759 199	67,0	76,4	-95 928	-11	9,4
Pays développés	14 864	10 498	98,0	98,7	-4 365	-29	0,7
Pays en transition	1 759	1 431	99,2	99,4	-328	-19	0,2
Afrique subsaharienne	128 980	140 544	49,9	59,7	+11 564	+9	9,8
États arabes	63 023	65 128	50,0	62,7	+2 105	+3	12,6
Asie centrale	572	404	98,7	99,2	-168	-29	0,5
Asie de l'Est et Pacifique	232 255	129 922	81,8	91,4	-102 333	-44	9,6
Asie du Sud et de l'Ouest	382 353	381 116	47,5	58,6	-1 237	-0,3	11,2
Amérique latine et Caraïbes	41 742	37 901	85,0	89,7	-3 841	-9	4,7
Europe centrale et orientale	11 500	8 374	96,2	97,4	-3 126	-27	1,2
Amérique du Nord et Europe de l'Ouest	11 326	7 740	97,9	98,7	-3 585	-32	0,8

Note : Les totaux peuvent ne pas tomber juste à cause des arrondis.

Source : Annexe statistique, tableau 2A.

Adapté de : UNESCO, 2005, p. 63.

Quoi qu'il en soit, c'est la fréquence et la qualité des interactions autour de l'alphabétisation qui ouvrent différentes voies aux enfants pendant les années préscolaires – et elles peuvent varier considérablement d'une famille à l'autre, à travers le monde. Ainsi, quand les enfants entrent à l'école, ils présentent une remarquable diversité en termes de compétences et d'expériences, qu'ils vivent dans des pays riches ou pauvres, dans des PSE ou des WSE. Dans la plupart des pays industrialisés, avant même d'entrer à l'école, certains enfants connaissent l'alphabet, sont capables d'écrire des mots familiers et de réciter le texte des livres qu'ils connaissent bien, alors que d'autres en sont incapables. Des études ont montré que les enfants ont besoin d'une diversité de compétences, de concepts et d'expériences pour pouvoir commencer à lire. Les enfants qui sont davantage familiarisés avec

la lecture et l'écriture quand ils sont petits ont généralement une longueur d'avance sur leurs pairs pour apprendre à lire¹³⁵.

Les composantes de l'apprentissage de la lecture

Le processus du développement précoce de la lecture a fait l'objet de nombreuses théories, études et affirmations qui ont produit un mélange confus d'opinions et de politiques un peu partout dans le monde. Aucune théorie sur la lecture ne peut être considérée comme étant dominante dans ce domaine (qui donne parfois lieu à de vifs débats), mais il existe une convergence de conclusions, en particulier concernant les premières étapes vers l'apprentissage de la lecture. Au cours de la dernière décennie, des méta-analyses (entreprises principalement aux États-Unis¹³⁶) ont mis en évidence cinq composantes essentielles de l'apprentissage de la lecture : le principe alphabétique, la conscience phonémique, la maîtrise de la lecture à haute voix, le vocabulaire et la compréhension¹³⁷. Chacune est décrite de façon succincte, sachant qu'un consensus se dessine parmi les spécialistes de la lecture sur le fait que ces composantes (y compris au niveau mondial¹³⁸) constituent les fondements de l'enseignement, des interventions et des réformes des pratiques pédagogiques dans les programmes scolaires, l'enseignement et les évaluations.

135. Voir Adams, 1990, pour une étude instructive.

136. Aux États-Unis, sous les auspices de l'Académie nationale des sciences, l'US National Reading Panel (2000) a entrepris d'analyser de façon exhaustive de nombreuses études sur les compétences et les expériences qui influencent l'apprentissage de la lecture. Les auteurs ont identifié trois obstacles à la maîtrise de la lecture qui influent sur les jeunes enfants : les difficultés pour utiliser et comprendre le principe de l'alphabet, l'incapacité de transférer vers la lecture les capacités de compréhension du langage parlé, ainsi que le manque de motivation pour la lecture. Ce rapport national a fortement influencé les politiques éducatives des États-Unis dans le domaine de la lecture.

137. Ces cinq composantes peuvent être considérées comme nécessaires mais pas suffisantes. Il est bien évident, comme nous l'avons fait remarquer précédemment, qu'il y a de nombreuses variables environnementales différentes (notamment liées à l'instruction formelle et informelle) qui ont un impact sur l'apprentissage de la lecture. Merci à A. Gove (communication personnelle) d'avoir attiré l'attention sur ce point.

138. Il y a quelques exceptions. Comme on le montrera plus loin dans ce chapitre, les systèmes d'écriture peuvent mettre l'accent de façons très différentes sur le rôle de ces différentes composantes. Voir également August et Shanahan (2006), pour une étude sur l'apprentissage de la lecture dans une langue non maternelle.

Le principe alphabétique : connaissance du nom et du son des lettres

Dans les pays de l'OCDE, la plupart des enfants apprennent le nom et le son de quelques lettres de l'alphabet avant d'entrer à l'école. Vers le milieu de la première année de scolarité, ces enfants connaissent déjà l'alphabet en entier¹³⁹. Ils acquièrent ces compétences par discrimination visuelle des symboles, mémorisant le nom et le son des lettres et coordonnant les associations audiovisuelles. Ces compétences sont intégrées dans le principe alphabétique, qui constitue la base de l'apprentissage de la lecture. La connaissance par les enfants des lettres et des relations entre les lettres et les sons constitue souvent un présage de leur capacité à lire¹⁴⁰. Cependant, la plupart des données sur le pouvoir prédictif de la reconnaissance du nom et du son des lettres sont corrélationnelles plutôt qu'expérimentales, de sorte que la maîtrise précoce du principe alphabétique indique également de nombreuses autres opportunités d'apprendre. Les différences initiales peuvent être temporaires et, en raison d'un apprentissage initial plus rapide, elles ne semblent pas avoir d'effet direct sur la compréhension ultérieure du texte écrit¹⁴¹.

La conscience phonémique

La conscience phonémique est la capacité de reconnaître et de manipuler des phonèmes pour fabriquer des syllabes et des mots *parlés*. Comprendre les relations entre les sons (*phonèmes*) et les lettres (*graphèmes*) dans un texte écrit est une compétence de *décodage*¹⁴², et relève clairement de la conscience phonémique. Savoir quels sons sont associés aux lettres aide les enfants à identifier les différents phonèmes associés au texte écrit. Vers cinq ans, la plupart des enfants anglophones des pays de l'OCDE peuvent identifier des modèles phonologiques – tels que *c-at*, *h-at* et *f-at* – qui forment la base des premières rimes. Plus tard, ils développent la capacité de segmenter

139. Morris *et al.* (2003), par exemple, ont utilisé un exercice consistant à identifier 15 lettres majuscules et minuscules et ont constaté que les enfants en connaissaient à peu près la moitié à leur entrée au jardin d'enfants et la totalité vers la fin de l'année.

140. Lonigan *et al.* (2000) écrivent : « ... la connaissance de l'alphabet (c'est-à-dire connaître le nom des lettres et les sons auxquels elles correspondent) lors de l'entrée à l'école est l'un des facteurs prédictifs les plus sûrs de succès à court et long termes dans l'apprentissage de la lecture ... » (p. 597).

141. Voir Paris (2005) pour une analyse des études principalement consacrées aux pays de l'OCDE. Inversement, le fait de ne pas réussir à assimiler le principe alphabétique (et le nom des lettres) restera un obstacle majeur pour les enfants, notamment dans les PSE, où l'alphabet peut ne pas être appris dans sa totalité.

142. La capacité de décoder est parfois appelée « syllabique ».

les mots en phonèmes et de mélanger des sons différents pour former des mots. Les mêmes compétences peuvent être appliquées aux mots écrits ou prononcés. Il s'agit là des aspects analytiques et synthétiques de base du décodage qui font suite à la conscience phonémique. De nombreux travaux de recherche ont mis en évidence d'importantes corrélations simultanées et prédictives entre la conscience phonémique et l'acquisition de la lecture¹⁴³. La conscience que les mots peuvent être divisés en phonèmes (autrement dit, la conscience phonémique) est cruciale pour le lecteur débutant. Cependant, le rôle causal direct de la conscience phonémique pour améliorer la lecture a été récemment remis en question¹⁴⁴.

La fluidité en lecture orale

La fluidité en lecture orale (ORF – *oral reading fluency*), qui résulte de la coordination de plusieurs compétences de décodage automatique¹⁴⁵, se développe par la pratique. La fluidité consiste à lire un texte rapidement,

143. Voir, par exemple, Bradley et Bryant, 1983 ; Juel *et al.*, 1986 ; Rayner *et al.*, 2001.

144. Certains chercheurs suggèrent que le lien avec la conscience phonémique est facilité par la connaissance des lettres (Blaklock, 2004), tandis que d'autres soutiennent qu'aucun lien causal n'a été démontré dans les études antérieures (Castles et Coltheart, 2004). L'US National Reading Panel (NRP, 2000) a trouvé près de 2 000 citations concernant la conscience phonémique, mais n'a effectué sa méta-analyse que sur 52 études répondant à ses critères. Ces études ont montré que l'entraînement à la connaissance phonémique améliore la lecture et l'orthographe des enfants. Le NRP a conclu en outre que toutes les formes d'entraînement systématique à la syllabique, y compris la syllabique par analogie, la syllabique analytique, la syllabique incrustée, la syllabique par l'orthographe et la syllabique synthétique, produisent des résultats significatifs chez les élèves des classes élémentaires qui ont du mal à lire. Le NRP a préconisé d'intégrer l'enseignement des phonèmes dans un programme de lecture global qui insiste également sur les quatre autres composantes fondamentales. Georgiou *et al.* (2008), estiment au contraire que « la conscience phonologique peut ne pas être un facteur prédictif majeur de l'aptitude à lire ou n'être importante que pendant la première ou les deux premières années d'école ». De même, dans une importante analyse, Slavin *et al.* (2010) ont affirmé que, dans le contexte des États-Unis, la conscience phonémique n'est pas « la lecture » à proprement parler, et peut être un artefact sans lien avec l'apprentissage de la lecture. D'autres (par exemple Berninger *et al.*, 2010a) font valoir qu'il existe de multiples formes de conscience linguistique (incluant notamment les formes orthographiques et morphologiques en plus de la forme phonologique) et qu'on peut faire fausse route en se focalisant sur cette dernière. Néanmoins, comme pour le principe alphabétique, dans les PSE et avec des écritures plus transparentes, les enfants qui ne maîtrisent pas la conscience phonémique au début de leur scolarité peuvent se retrouver durablement handicapés.

145. Voir plus loin l'analyse sur l'automatisme, dans la section « La compréhension écrite ».

avec précision et une bonne intonation. Mesurer la précision de la lecture orale des enfants est une pratique concrète ancienne dans les répertoires de lecture informels qui répertorient les fautes ou les fiches d'observation individualisées¹⁴⁶. La vitesse de lecture étant un indicateur de décodage automatique, les enfants qui lisent plus vite identifient souvent les mots de façon plus précise et ont plus de ressources cognitives disponibles pour lire de façon expressive et en comprenant le texte.

L'emploi de la vitesse de lecture comme mesure de fluidité est également une pratique ancienne dans l'éducation spécialisée, connue sous la dénomination d'évaluation axée sur le programme scolaire¹⁴⁷. L'ORF, définie généralement comme étant le nombre de mots lus correctement en une minute dans des textes tirés du programme scolaire de l'élève, est une mesure centrale dans l'évaluation axée sur le programme scolaire. Elle s'appuie sur un texte extrait du programme scolaire normal pour obtenir des indicateurs de résultat qui peuvent être contrôlés au fil du temps à des fins de diagnostic et de contrôle de l'action menée. L'ORF est une composante fondamentale des *indicateurs dynamiques des savoirs essentiels en lecture* (INDISSE), ensemble d'évaluations précoces des compétences en lecture¹⁴⁸. Avec la mesure de l'ORF, il apparaît relativement facile d'évaluer combien de mots les enfants lisent correctement en une minute et de comparer leur vitesse de lecture aux normes correspondant à leur niveau de scolarité. Toutefois, il faut être prudent quand on interprète les données issues de l'ORF, parce que la vitesse n'est qu'un indice parmi d'autres de maîtrise de la lecture. De plus, l'évaluation elle-même pose un certain nombre de problèmes majeurs quand ceux qui administrent les tests manquent d'expérience.

Le vocabulaire

Les compétences en vocabulaire incluent la compréhension de mots parlés ou écrits ; la connaissance de ces deux formes améliore la compréhension écrite, quelle que soit la langue. Le vocabulaire étant lié, du point de vue du développement, aux compétences en langage oral, l'enrichissement du vocabulaire pendant les années préscolaires prépare la voie à l'acquisition

146. Clay, 1991 ; Paris et Carpenter, 2003. Clay (2000) a également fait un travail précurseur sur une mesure de la lecture appelée « Concepts associés à l'écrit » (*Concepts about Print – CAP*), qui a été utilisée dans certaines études EGRA.

147. Deno *et al.*, 1982 ; Fuchs et Fuchs, 1999.

148. Dans INDISSE, voir Good et Kaminski, 2002 ; la fluidité de la lecture orale fait également partie de la Trousse d'outils EGRA (RTI, 2009).

de la lecture quand l'enfant entre à l'école¹⁴⁹. De plus, le vocabulaire (tant réceptif, par la compréhension, qu'expressif, par la production du discours) est un présage de compétences précoces en lecture¹⁵⁰ ; le nombre de mots différents qu'un enfant comprend et emploie facilite le décodage des mots et peut soutenir le développement de sa conscience phonémique¹⁵¹.

Dans les environnements d'alphabetisation favorables (WSE), la répétition, l'exposition fréquente aux mots, l'informatique et les expériences d'apprentissage contribuent à renforcer l'acquisition de vocabulaire. Inversement, l'absence de ces expériences limite considérablement les compétences en vocabulaire d'un enfant dans n'importe quelle langue. Les recherches montrent que l'enseignement initial du vocabulaire et du contenu conceptuel associé peut faciliter ultérieurement la compréhension écrite. De même, l'enseignement axé sur l'étude des mots peut développer la compréhension par l'enfant du système d'écriture, de l'orthographe et du vocabulaire¹⁵².

La compréhension écrite

Le développement des facultés de compréhension est un long processus de développement qui dépend de la richesse des expériences vécues dès la petite enfance en matière de mots, de textes et de langage. Ce processus dépend aussi de l'apprentissage du décodage, de la fluidité du décodage, en partie, par le développement d'un large répertoire de mots mémorisés de façon globale, de l'apprentissage des significations des mots couramment rencontrés dans les textes, ainsi que de l'apprentissage de la façon d'extraire le sens d'un texte en utilisant les processus de compréhension employés par les lecteurs confirmés¹⁵³.

Comprendre le sens des mots et des textes écrits, fonction première de l'alphabetisation, permet aux individus de communiquer des messages à travers l'espace et le temps, et de s'exprimer au-delà de la gestuelle. Apprendre à comprendre les mots écrits et à communiquer par le biais de textes, avec une pensée interprétative, constructive et critique, est sans

149. Hart et Risley, 2005.

150. Storch et Whitehurst, 2002.

151. Dickinson *et al.*, 2003. L'US National Reading Panel (2000) a également passé en revue 50 études tirées d'un ensemble potentiel de 20 000 citations concernant le vocabulaire et a conclu que l'enseignement du vocabulaire facilite directement la compréhension écrite.

152. Beck *et al.*, 2002 ; Bear *et al.*, 2004. Il n'y a pratiquement pas d'études sur cette notion dans les PMA.

153. Pressley, 2000, p. 556.

doute la tâche principale de l'éducation formelle partout dans le monde. Sans compréhension, lire des mots se réduit à imiter les sons de la langue, répéter du texte n'est rien de plus qu'une mémorisation, et écrire des lettres et des caractères est juste de la copie ou du griffonnage¹⁵⁴.

La compréhension comprend de nombreux niveaux différents, ce qui la rend difficile à définir et à mesurer de façon fiable. Ce processus complexe est influencé par la richesse du vocabulaire et l'instruction, l'interaction riche en réflexion entre le lecteur et le texte, ainsi que par la capacité des enseignants d'inculquer aux élèves des stratégies de lecture appropriées. L'utilisation efficace des stratégies de lecture devient plus importante, à mesure que les textes se complexifient et que les objectifs des jeunes lecteurs se diversifient. L'instruction dans les petites classes aide les enfants à apprendre à lire, principalement en décodant des mots écrits mais, en 2^e année (dans les WSE, plus tard dans les PSE), on leur enseigne à lire afin d'apprendre des choses très diverses¹⁵⁵. Les chercheurs ont identifié sept types d'instruction favorisant la compréhension écrite, notamment s'ils sont utilisés de façon combinée, en tant qu'approches stratégiques multiples : suivi de la compréhension, apprentissage coopératif, utilisation d'organiseurs graphiques et sémantiques, réponse à des questions, production de questions, structure du récit et résumé¹⁵⁶. En utilisant ces stratégies et techniques, les enfants peuvent comprendre et mémoriser plus facilement le sens de ce qu'ils lisent.

L'automatisme de la lecture

La plupart des théories sur l'acquisition de la lecture établissent une distinction entre les processus liés au décodage mot écrit/son et les processus liés au déchiffrement du sens du texte (compréhension). L'un des liens entre ces deux classes générales de processus est la notion d'automatisme. Plus simplement, elle indique quand le lecteur est capable de reconnaître

154. Pourtant, dans certains environnements d'alphabétisation défavorables (PSE), en particulier dans les pays en développement, des enseignants acceptent parfois cette façon de lire comme indicateur suffisant de développement de la lecture. C'est pourquoi le développement de nouveaux outils d'évaluation hybrides, comme l'EGRA, est important, précisément parce qu'ils peuvent et doivent faire la distinction entre l'imitation et la compréhension au niveau individuel. Dubeck (communication personnelle) fait à juste titre valoir que l'on peut avoir des scripteurs qui, tout en ne comprenant pas, sont très forts pour transcrire avec précision un texte oral. L'éducation islamique, où l'arabe est parfois transcrit sans le comprendre, en est un exemple (Wagner, 1993).

155. Voir Chall (1967, 1996), sur les étapes de la lecture.

156. US National Reading Panel, 2000.

automatiquement (c'est-à-dire rapidement et sans effort) le son des lettres et des mots, ainsi que les mots familiers, ce qui lui permet d'améliorer la vitesse et la précision de sa lecture¹⁵⁷. Cette fluidité de la lecture orale est importante, car lorsque les processus de base (comme le décodage et la reconnaissance des mots) sont devenus automatiques, le lecteur a davantage de capacité mentale disponible (mémoire opérationnelle notamment) et a le temps d'accorder plus d'attention au sens du texte. Lorsqu'un lecteur débutant passe trop de temps et d'énergie à déchiffrer les sons et les mots d'un texte, il lui arrive d'oublier ou de ne plus se souvenir des mots qu'il vient juste de lire¹⁵⁸. Le lecteur qui lit couramment remarque généralement qu'il lit presque sans effort et n'est pas conscient des divers processus qui lui permettent d'extraire le sens d'un texte écrit.

-
157. Un processus automatique est généralement jugé en fonction de la vitesse avec laquelle il se produit. Ainsi, pour déterminer si des mots écrits sont automatiquement identifiés dans un test de lecture, il est important de tenir compte de la rapidité du processus. C'est le cas, en particulier, dans des études portant sur des langues ayant des orthographe relativement transparentes comme l'espagnol, l'italien, l'allemand ou le français, plutôt que sur l'anglais. L'identification automatique des mots écrits est un processus très rapide, qui ne prend que quelques millisecondes à un lecteur confirmé. Dans les études qui prennent en compte la latence des réponses vocales (c'est-à-dire le temps qui s'écoule entre l'apparition du mot sur l'écran d'un ordinateur et le début de sa prononciation par le participant), les différences entre les bons et les mauvais lecteurs sont de l'ordre de 200 millisecondes par mot. Cela correspond à une différence d'une minute environ entre les mauvais et les bons lecteurs pour la lecture d'un texte de 300 mots. Voir Sprenger-Charolles *et al.* (2006) qui soutiennent que le manque d'automatisme est l'un des principaux freins à la facilité de lecture.
158. Bien sûr, la notion de « temps excessif » est relative. En effet, c'est un domaine où les études menées dans des pays de l'OCDE, et principalement en anglais, peuvent avoir un impact non seulement sur les théories relatives à la lecture, mais aussi sur la façon dont les capacités en lecture doivent être évaluées. D'après les données d'études EGRA récentes, il apparaît que même les lecteurs « lents » peuvent lire avec un niveau élevé de compréhension, et que les lecteurs « rapides » peuvent le faire avec un faible niveau de compréhension. C'est surtout évident quand de jeunes lecteurs apprennent à lire dans une seconde langue (A. Gove, communication personnelle, 2009).

L'automatisme résulte principalement d'une longue pratique de la lecture, souvent sur plusieurs années, et n'est donc ni immédiate ni rapide¹⁵⁹. Elle peut toutefois découler de l'application de règles et stratégies simples de décodage. En anglais, les jeunes lecteurs capables de reconnaître des modèles phonologiques, comme *d-og*, *l-og*, *f-og*, arrivent facilement à prononcer des mots peu familiers comme *c-og*, voire des mots qui n'ont pas de sens comme *m-og*. Les évaluations qui mesurent la rapidité avec laquelle les lecteurs débutants sont capables d'identifier les sons des lettres, les mots charnières et les mots dépourvus de sens mesurent, par conséquent, la capacité de décodage automatique. En général, les mesures de la maîtrise relative et de l'automatisme du décodage chez les lecteurs débutants sont de bons facteurs prédictifs d'un apprentissage rapide de la lecture¹⁶⁰.

La lecture dans la première et dans la seconde langue

[...] (L)e développement de l'alphabétisation est conditionné par de nombreux facteurs : la maîtrise de la lecture et de l'écriture dans les dernières classes de l'école élémentaire et dans le secondaire n'est pas possible sans un niveau élevé d'aptitudes linguistiques, l'accès à d'importants fonds de connaissances et la maîtrise des normes culturelles locales de communication¹⁶¹.

159. Certains ont avancé que lire de façon automatique, c'est comme faire du vélo (métaphoriquement). La comparaison est pertinente en ce sens que, lorsqu'on sait faire du vélo, on peut faire attention à d'autres aspects (au vent ou au paysage, par exemple). Mais cette métaphore donne l'impression que l'automatisme vient facilement et rapidement, comme quand on apprend à faire du vélo. C'est là qu'elle est moins adaptée. Comme le font remarquer Fuchs *et al.* (2001, p. 240) : « La fluidité de la lecture orale se développe progressivement pendant les années d'école élémentaire ... », avec des implications similaires en termes d'automatisme.

160. Voir Stanovich (2000) à propos de l'importance de l'automatisme dans la lecture. Même dans ce cas, des facteurs prédictifs de la capacité de lire comme l'automatisme peuvent dépendre du contexte et de la langue. Plus généralement, les études sur la cognition et l'automatisme donnent un tableau mitigé : la conception « tout ou rien » de l'automatisme a été remise en question par des études montrant l'absence de co-occurrence entre les caractéristiques centrales des processus automatiques. Par exemple, des chercheurs spécialisés dans la lecture (comme Stanovich) ont largement utilisé la mesure d'interférence de Stroop pour mesurer l'automatisme dans des expériences réalisées en laboratoire. Des études récentes semblent indiquer que l'automatisme n'est, en réalité, pas aussi automatique, mais soumise à de nombreux facteurs contextuels. Voir Moors et De Houwer, 2006.

161. Snow et Kang, 2006, p. 76.

Durant de nombreuses décennies, enseignants et autres acteurs ont débattu de la nature et de l'acquisition de la lecture dans une seconde langue, à l'instar des discussions sur le bilinguisme. Ces débats vont de la « confusion linguistique » qu'entraîne l'apprentissage pendant l'enfance, lorsque les enfants parlent plus d'une langue à la maison, à une controverse similaire concernant le stade, plus ou moins précoce, auquel l'apprentissage de la lecture dans une seconde langue devrait intervenir dans les programmes scolaires. Les politiques liées au choix des langues, dans les pays où ces questions peuvent provoquer de vives controverses, ajoutent à la tension. Même si ces débats dépassent le champ de la présente étude, les évaluations de la lecture dans la première langue (L1) et dans la seconde langue (L2) ont une place importante dans la présente discussion, dans la mesure où beaucoup d'enfants dans les PMA sont confrontés à cette réalité multilingue, bien plus fréquemment que dans la plupart des pays de l'OCDE¹⁶².

Au-delà des décisions politiques sur le choix de la langue et l'enseignement bilingue, une science englobant les dimensions sociales et cognitives de l'apprentissage (dont l'apprentissage de la lecture) dans plusieurs langues est en train d'émerger. Entre sociolinguistes et historiens, la conversation tourne souvent autour du respect et de la résistance. Les enfants (et leurs parents) se sentent respectés quand leur langue maternelle (L1) est employée à l'école, et une résistance s'exprime (pour certains) quand tel n'est pas le cas¹⁶³. Du point de vue de l'apprentissage cognitif, la conversation tourne largement autour du concept de transfert. Dans quelle mesure les compétences acquises lors de l'apprentissage de la première langue (oralement ou sous forme écrite) s'appliquent-elles à une seconde langue ? Au fil des ans, les conclusions de la majorité des études tendent à corroborer la notion « additive » ou « interdépendante », selon laquelle la maîtrise de la langue et de la lecture dans la L1 renforce généralement ces compétences dans la L2¹⁶⁴. Il reste toutefois de nombreuses questions non

162. Le sous-domaine de l'acquisition comparative de la lecture n'est pas une nouveauté. Il y a près de quatre décennies, Downing (1973) a publié une étude majeure sur l'acquisition de l'aptitude à lire dans de nombreuses langues et systèmes d'écriture.

163. Pour une analyse, voir Hornberger (2003) et Snow et Kang (2006). Pour les points de vue ethnographiques, voir Wagner, 2004.

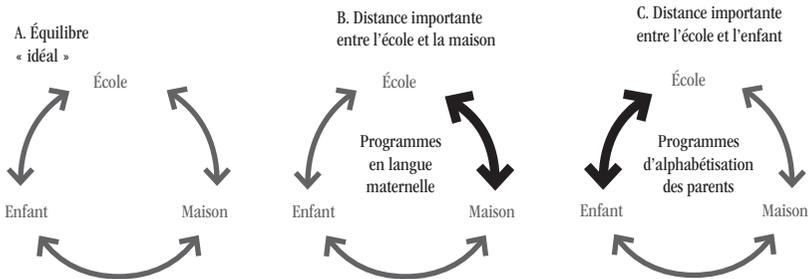
164. Voir Cummins *et al.* (1994) pour les premiers travaux dans ce domaine et, plus récemment, Bialystok *et al.*, 2005. Voir également Koda et Reddy (2008) pour une approche plus nuancée, qui différencie le transfert en s'intéressant à des composantes spécifiques qui semblent les plus susceptibles d'être transférées.

résolues. Quand l'enseignement de la lecture dans la L2 doit-il commencer ?
Doit-il s'accompagner de l'enseignement d'autres matières en L2¹⁶⁵ ?

Une manière d'envisager la question du transfert consiste à étudier la distance entre les deux langues et leurs contextes d'apprentissage¹⁶⁶. La *figure 5.1* montre trois types généraux de contextes. Le modèle A (« idéal ») évoque une harmonie et un équilibre entre la maison, l'enfant et l'école. Le modèle B fait apparaître une distance relativement plus grande entre la maison et l'école, suggérant qu'une politique d'éducation appropriée serait peut-être de réduire cette distance par un programme d'éducation non formelle, d'éducation des parents ou d'alphabétisation des adultes. Le modèle C décrit une situation où c'est entre l'enfant et l'école que la distance est la plus grande, donnant à penser qu'il faudrait privilégier un programme d'éducation dans la langue maternelle (ou bilingue). Bien que ces modèles ne soient pas nécessairement aussi faciles à distinguer les uns des autres, ils constituent un moyen d'engager la discussion sur la façon d'aborder les questions de L1 et L2 dans les contextes sociopolitiques.

Figure 5.1 Approche des programmes d'éducation bilingue fondée sur une « théorie de la distance »

Théorie de la distance : l'approche de la « réduction »



Tiré de Wagner, 2009a.

La mesure des capacités en lecture en L1 et en L2 pose des problèmes spécifiques. La question du transfert est complexe et recouvre (au

165. Le débat actuel en Afrique subsaharienne porte sur la question de savoir si les élèves ont besoin de trois ou six années (ou une autre durée) d'école primaire en L1 pour bénéficier du transfert vers la L2. Voir Alidou *et al.*, 2006. Il y a peu de consensus sur cette question pour le moment.

166. Wagner, 2008.

minimum) les éléments suivants : expression orale et aptitude à la lecture en L1, expression orale et aptitude à la lecture en L2, acquis antérieurs dans les deux langues, stratégies de traitement et vocabulaire. De plus, les mots apparentés (et non apparentés) de L1 et L2 influencent les processus d'acquisition linguistique et de la lecture dans les deux langues, processus très complexe, en effet ! Dans la plupart des cas, il semble que les enfants soient capables de s'appuyer sur leurs compétences (orales et écrites) en L1, mais qu'il y ait aussi une grande variabilité liée aux variables sociales et environnementales¹⁶⁷.

Il ressort de certaines études que l'importance des différences dépend de la capacité de l'enfant à identifier automatiquement des mots et à comprendre ce qu'il lit : les élèves qui apprennent à lire en L2 peuvent rattraper ceux dont c'est la langue maternelle pour la lecture des mots, mais probablement pas en ce qui concerne la compréhension écrite¹⁶⁸. Le succès de cette dernière dépend pour une large part des compétences en langue orale (ainsi que des connaissances et concepts liés), c'est pourquoi il est difficile pour les apprenants en L2 dotés d'aptitudes modestes à l'oral dans cette même langue de rattraper les enfants qui lisent en L1. À l'inverse, des capacités bien développées à l'oral en L2 favorisent généralement une bonne compréhension écrite en L2¹⁶⁹.

Dans l'ensemble, les données disponibles à ce jour corroborent généralement la notion selon laquelle le bilinguisme et la bialphabétisation peuvent être additifs si l'apprenant peut (grâce à un bon programme scolaire, à un enseignant bien formé et à un environnement propice) acquérir des compétences en L2 sur la base de ses acquis en L1¹⁷⁰. Dans les environnements d'alphabétisation défavorables (PSE), ces conditions favorables sont souvent insuffisamment satisfaites. Sachant que, dans les PMA, les enfants sont souvent confrontés à cette situation d'apprentissage complexe, il faut donc que les évaluations mesurent non seulement les compétences en lecture dans la langue maternelle, mais aussi dans la deuxième (voire la troisième) langue.

167. Voir la discussion dans Bernhardt, 2005.

168. Voir Lesaux et Geva, 2006 ; Lesaux *et al.*, 2006a ; Wang et Koda, 2007.

169. Voir Crosson *et al.*, 2008 ; Droop et Verhoeven, 1998. Voir également Crosson et Lesaux (2010) à propos de l'importance de la fluidité de la lecture pour les apprenants de minorités linguistiques (seconde langue) qui lisent en anglais, suggérant des limites possibles à l'application des mesures de fluidité aux apprenants d'une seconde langue.

170. De bons programmes d'enseignement bilingue peuvent également être rentables du point de vue des coûts (voir, par exemple, Patrinos et Velez, 2009).

Systèmes d'écriture, lecture et orthographe

Les études internationales actuelles sur la lecture commencent à s'intéresser de plus près aux questions de « transparence » et d'« opacité » du système d'écriture d'une langue par rapport à la version orale de celle-ci¹⁷¹. Par exemple, l'espagnol est extrêmement transparent du fait de l'homogénéité considérable des liens entre graphèmes et phonèmes. Le graphème « a » correspond systématiquement au phonème /a/. L'anglais, en revanche, est relativement opaque, en raison du manque considérable d'uniformité dans les liens entre graphèmes et phonèmes. En anglais, le graphème « a », par exemple, correspond à deux phonèmes différents dans les mots *cat* et *date*. Il est intéressant de noter que beaucoup de langues des pays en développement (sinon la plupart) utilisent un système d'écriture plus transparent que l'anglais¹⁷². Cet aspect est important, notamment parce que les études montrent que les compétences en lecture sont acquises plus rapidement dans les langues dont le système d'écriture est transparent¹⁷³.

Les correspondances entre graphèmes et phonèmes (GPC) qui servent à lire ne sont pas toujours comparables aux correspondances entre phonèmes et graphèmes (PGC) utilisées pour écrire sans faute. En français, par exemple, les GPC sont plus homogènes que les PGC : le « o » – comme dans « do » – se prononcera toujours comme la voyelle /o/ ; il existe toutefois de nombreuses façons d'écrire /o/ : « eau », « au » et « o ». La relation entre GPC et PGC est asymétrique, dans le sens où il est plus facile de lire un mot que de l'orthographier. En conséquence, apprendre à lire le français est relativement plus facile qu'apprendre à l'orthographier¹⁷⁴. En Afrique

171. Voir Sprenger-Charolles, 2003 ; Sprenger-Charolles *et al.*, 2006 ; Ziegler et Goswami, 2006.

172. Cela peut être dû au caractère récent de l'apparition des systèmes d'écriture pour la plupart des langues d'Afrique subsaharienne (par exemple le swahili ou le wolof) ; Institut international des langues et cultures africaines, 1930 ; voir également Sebba, 2007.

173. Seymour *et al.* (2003) ont comparé les débuts de l'acquisition de la lecture dans 13 langues (allemand, anglais, danois, espagnol, français et portugais, entre autres) et ont constaté qu'une plus grande transparence permettait une acquisition plus rapide de l'aptitude à décoder chez les apprenants en L1. Voir également une étude récente auprès d'élèves hispanophones en Espagne, où il est montré que la maîtrise précoce du décodage permet une progression régulière de la vitesse et de la précision de lecture de la maternelle jusqu'à la 4^e année (Cuetos et Suarez-Coalla, 2009). Des conclusions similaires ont été tirées pour les apprenants en L2 (Geva et Siegel, 2000).

174. Par ailleurs, les mots ordinaires sont lus de façon plus correcte que les pseudo-mots, sans être pour autant plus faciles à écrire, en particulier en français. Voir Alegria et Mousty, 1996 ; Sprenger-Charolles *et al.*, 2003.

subsaharienne, où l'anglais ou le français sont souvent utilisés comme L2 à l'école, cette variation de transparence (GPC ou PGC) peut poser des difficultés pour l'acquisition précoce de la lecture et de l'orthographe (en anglais), de l'orthographe (en français).

Pour résumer, les codes orthographiques jouent différemment pour les enfants qui apprennent à lire, en particulier au début du décodage des compétences en lecture et de l'acquisition de l'orthographe. Ces différences – d'après des études menées à ce jour essentiellement dans des pays de l'OCDE – sont importantes. Ce qui manque, c'est un nombre suffisant d'études montrant l'impact des systèmes d'écriture des langues indigènes dans les PMA¹⁷⁵, ainsi que des données attestant que les avantages des systèmes d'écriture transparents en début d'apprentissage ont un impact à long terme. Comme il a déjà été signalé, même si la connaissance précoce de l'alphabet est un facteur prédictif utile de lecture précoce, elle a peu d'effet direct sur l'apprentissage de la lecture¹⁷⁶. De même, on s'est aperçu que les variations de la complexité apparente des différentes langues avaient peu d'impact sur la rapidité avec laquelle les enfants maîtrisent leur langue, à l'oral, dans le monde entier¹⁷⁷. Enfin, dans les pays à langues et systèmes

175. Voir, cependant, une étude récente (Kim *et al.*, 2008) menée au Kenya sur l'acquisition de la lecture en swahili (dont le système d'écriture est relativement transparent), qui a utilisé plusieurs sous-tests de lecture (semblables à l'EGRA) pour déterminer le rôle de la fluidité de la lecture et d'autres compétences dans la compréhension écrite. Dans l'ensemble, les résultats corroborent la théorie de l'importance de la fluidité de la lecture (et du décodage) pour prédire l'aptitude à la compréhension écrite.

176. Ces différences d'évolution de la trajectoire d'apprentissage fondées sur les différences entre les systèmes d'écriture sont l'une des raisons pour lesquelles il ne sert à rien de créer des référentiels pour les premiers stades d'acquisition de la lecture dans différentes langues et systèmes d'écriture.

177. Voir Slobin, 1986 ; Snow, 2006. Une étude récente de Feng *et al.* (2009) montre que s'il y a quelques différences initiales de compétences entre les apprentis lecteurs d'anglais et de chinois (qui ont des systèmes d'écriture distincts et extrêmement différents), aucune différence de vitesse de lecture n'a été observée entre les lecteurs confirmés (plus âgés). Feng et ses collègues ont utilisé des techniques sophistiquées de suivi des mouvements oculaires (saccades oculaires) pour étudier le développement de la lecture et sa rapidité.

d'écriture multiples, on ne sait pas toujours très clairement quels systèmes d'écriture sont utilisés au quotidien¹⁷⁸.

Trajectoires de développement : prévoir l'échec et le succès

D'après certaines études, il semble que les trajectoires de développement des composantes de la lecture suivent des schémas différents avec des durées variables, certaines compétences ayant un champ et un temps d'acquisition plus limités que d'autres¹⁷⁹. Par rapport au vocabulaire et à la compréhension, par exemple, le principe alphabétique est acquis rapidement¹⁸⁰. Apprendre le nom des lettres et les sons associés dans n'importe quel alphabet représente

178. Au Sénégal, la Constitution de 2001 reconnaît six langues nationales (wolof, seereer, pulaar, mandinka, soninké et joola) en plus du français (qui est la langue officielle), alors que l'on estime à une trentaine le nombre de langues en usage dans le pays. Le ministère de l'Éducation a entrepris d'élaborer des programmes scolaires dans ces langues nationales et de former des enseignants pilotes pour des projets expérimentaux en langue locale au cycle primaire. Le français reste toutefois la langue d'instruction dans les écoles sénégalaises à tous les niveaux. Le wolof a le statut de *lingua franca* au Sénégal (il est parlé par plus de 80 % des Sénégalais, alors que seulement 44 % environ appartiennent à l'ethnie wofof ; voir McLaughlin, 2001). L'exemple des enfants sénégalais en primaire illustre bien les problèmes que posent un environnement sociolinguistique complexe et le manque de normalisation des systèmes d'écriture pour évaluer l'alphabétisme en wolof. Entre 1968 et 1977, le gouvernement a pris une série de décrets pour demander une codification écrite normalisée du wolof fondée sur le système d'écriture latin, l'*ijjib wolof*, qui diffère de diverses façons des règles orthographiques du français. Mais cette décision du gouvernement a suscité de nombreux débats, à la fois chez les utilisateurs de l'alphabet *wolofal* (fondé sur l'écriture arabe du wolof ; voir Prinz, 1996) et chez les universitaires élaborant d'autres alphabets wolof indigènes. Aujourd'hui encore, le système d'écriture *ijjib wolof* élaboré à la demande du gouvernement est rarement utilisé dans la culture populaire sénégalaise et dans les pratiques quotidiennes de la lecture et de l'écriture (journaux, enseignes de restaurants, brochures religieuses, bandes dessinées et paroles de chansons), certaines personnes préférant écrire en *wolofal* et d'autres en utilisant le système d'écriture français du wolof. La coexistence au Sénégal de plusieurs systèmes d'écriture (normalisé et français) du wolof, en plus de l'écriture arabe, rend complexe l'élaboration d'évaluations des compétences en lecture dans les langues locales. Merci à Cecile Evers (communication personnelle) pour cette analyse précieuse.

179. Paris, 2005. Les compétences limitées sont également circonscrites en variance, dans la mesure où la limite supérieure est fixée et atteinte relativement vite dans les environnements d'alphabétisation favorables (WSE).

180. Cette vitesse relative vaut probablement pour les WSE et les PSE, bien qu'il faille attendre les prochaines conclusions de l'EGRA pour le confirmer. Cependant, l'apprentissage du principe alphabétique (et de l'alphabet) peut être beaucoup plus lent dans les régions et milieux pauvres des pays en développement, comme l'ont montré de nombreuses études EGRA (RTI, 2009).

un petit univers de connaissances, comparé aux nouveaux mots que l'on apprend tout au long de l'existence¹⁸¹. La conscience phonémique a une base de connaissances plus large, mais la plupart des enfants qui vivent dans un environnement d'alphabétisation favorable (WSE) apprennent les principales techniques de rime, de segmentation et de mélange des phonèmes au cours des premières années de scolarité.

La fluidité de la lecture à haute voix (ORF) est un champ moins limité que la connaissance de l'alphabet, mais davantage que le développement du vocabulaire, parce que la plupart des enfants atteignent leur niveau asymptotique de lecture orale précise en 4^e ou 5^e année dans les WSE. Ainsi, les compétences qui composent la lecture varient selon la sphère des connaissances acquises et la durée d'apprentissage¹⁸².

Qu'est-ce qui permet, alors, de prévoir la trajectoire de maîtrise de la lecture d'un enfant ? Pour les lecteurs débutants (entre quatre et six ans) vivant dans un WSE, les meilleurs facteurs prédictifs sont la connaissance du nom des lettres et des sons associés, ainsi que la capacité de faire rimer, segmenter et mélanger des phonèmes. À mesure que les enfants apprennent à décoder un texte, leur vitesse de lecture orale (c'est-à-dire de reconnaissance automatique des mots) devient un bon facteur prédictif de maîtrise ultérieure de la lecture. Chez les enfants un peu plus âgés, les scores de compréhension écrite permettent de prévoir les scores des tests de compétences¹⁸³. Le pouvoir prédictif des différentes compétences qui composent la lecture change donc en fonction des mesures des résultats et du moment où elles sont effectuées. De plus, les facteurs prédictifs de compétence en lecture dans les petites classes (connaissance de l'alphabet et fluidité de la lecture orale, par exemple) sont des mesures indirectes de nombreuses expériences d'apprentissage, comme l'exposition aux textes écrits et à la langue, l'instruction précoce des parents et des enseignants, ainsi qu'un vocabulaire riche. Par conséquent, des enfants de cinq ans qui vivent dans un WSE et qui connaissent mieux l'alphabet que leurs pairs ont

181. Scarborough (1998) a analysé 61 études sur le début de l'apprentissage de la lecture et de l'écriture (essentiellement aux États-Unis) et a constaté que le facteur prédictif le plus fort des différences individuelles dans l'aptitude à lire, entre la 1^{re} et la 3^e année de scolarité, était la connaissance du nom des lettres en maternelle, sur les 24 exemples d'études ayant mesuré cette variable. Mais, en 1^{re} année, le fait de connaître le nom des lettres n'était plus le facteur prédictif le plus fiable des compétences futures en lecture : il était le plus souvent éclipsé par les compétences phonologiques, comme la connaissance du son des lettres et les tâches de synthèse et d'analyse des phonèmes.

182. Fuchs et Fuchs, 1999.

183. Aux États-Unis, par exemple, voir Pearson et Hamm, 2005.

des chances de savoir mieux lire qu'eux à sept ou huit ans, mais la raison ne tient pas seulement à leur connaissance de l'alphabet, mais au fait qu'ils ont une longueur d'avance grâce à leurs nombreuses expériences de lecture et d'écriture¹⁸⁴. De même, les enfants de huit ans (ayant deux années de scolarité) qui vivent dans un PSE dans les pays en développement en sont peut-être encore à apprendre l'alphabet, rendant ainsi difficile leur maîtrise des autres compétences inhérentes à la lecture.

Évaluations des compétences en lecture

Il existe deux types principaux d'évaluation des compétences en lecture : celles qui peuvent être effectuées à l'écrit (LSEA, par exemple) et qui concernent uniquement les élèves qui ont déjà une maîtrise suffisante de la lecture pour lire et comprendre les instructions et remplir un formulaire de test écrit, et celles destinées aux lecteurs débutants qui ne peuvent passer des

184. Il semble y avoir un hiatus de développement entre la fluidité de la lecture orale (ORF) et la compréhension, parce que ce lien est plus fort pour les lecteurs débutants et les lecteurs ayant des difficultés (Paris, 2005). À mesure qu'augmentent l'âge et la compétence en lecture, on observe chez les enfants plus âgés et les lecteurs confirmés des relations plus variables entre l'ORF et la compréhension. Par exemple, Fuchs *et al.* (2001) ont observé un déclin de l'importance de l'ORF à mesure que l'âge augmente. Certains chercheurs affirment toutefois qu'il y a une forte relation entre l'ORF et la compréhension (par exemple Abadzi, 2008 ; RTI, 2009 ; Good *et al.*, 2001). Il se pourrait que les fortes corrélations entre l'ORF et les scores aux tests de lecture reflètent des différences générales de développement entre bons et mauvais lecteurs, plutôt qu'un lien causal entre ORF et compréhension. Selon ce point de vue, la vitesse de la lecture orale peut être une mesure indirecte très utile pour de nombreuses différences concomitantes de développement, notamment la reconnaissance automatique des mots, la richesse du vocabulaire, la connaissance du contenu, la motivation, l'habileté à passer des tests, l'intelligence, les facteurs contextuels, etc. Les lecteurs lents en 1^{re} et 2^e années (en particulier dans les PSE) peuvent se distinguer des lecteurs rapides à de nombreux égards, et leur vitesse de lecture orale n'est qu'une variable indirecte des différences qui influencent en fait la compréhension écrite. Le déchiffrement laborieux chez les lecteurs débutants (comme l'EGRA permet de l'observer dans les PSE) peut aussi surcharger la mémoire opérationnelle (Abadzi, 2008) et entraver des stratégies de compréhension efficaces. Les mesures de fluidité de la lecture orale doivent être utilisées, dans ce cas, comme indicateurs de décodage automatique plutôt que comme mesures de la compréhension ou des objectifs pédagogiques. En effet, certains ont soutenu que les problèmes de fluidité de la lecture orale étaient également causés par des problèmes de compréhension : « les problèmes de fluidité [peuvent] simplement refléter ... la lenteur à exécuter une tâche ou des problèmes de synchronisation pour coordonner des processus multiples dans l'esprit de l'apprenant ; la non-fluidité peut être le *résultat*, et non la cause, d'un problème de lecture » (Beringer *et al.*, 2010b).

tests écrits et doivent donc être testés à l'oral. Chacun de ces types de test est décrit ci-après et illustré par des exemples à l'*annexe B*.

Évaluations écrites des compétences en lecture

Comme décrit au *chapitre 4*, les tests normalisés développés par des éditeurs commerciaux, ainsi que par des agences nationales ou internationales, sont les indicateurs scientifiques et objectifs les plus reconnus des compétences en lecture des élèves¹⁸⁵. Dans les WSE, les élèves qui ont des difficultés bénéficient parfois d'un soutien spécial pendant l'année scolaire ou lors de cours d'été. Dans les PSE, les enfants qui ont des difficultés redoublent le plus souvent leur année, quittent l'école ou n'obtiennent pas leur certificat de fin d'études primaires. Ainsi, les tests de lecture peuvent avoir un impact important sur les élèves, car le redoublement et la non-obtention du certificat de fin d'études primaires ont souvent des répercussions négatives qui durent toute la vie. Inversement, ceux qui obtiennent de bons scores aux tests peuvent poursuivre leurs études après l'école primaire, se voient éventuellement offrir une place dans des classes de niveau supérieur, voire bénéficient d'une bourse.

Évaluations des compétences en lecture orale

Il y a plus de 50 ans, des spécialistes de l'éducation ont élaboré des répertoires de lecture informels (IRI) afin d'évaluer divers aspects des compétences des jeunes enfants en lecture orale dans des situations naturelles, c'est-à-dire des enfants lisant des textes avec leur enseignant en classe. L'un de ces répertoires, appelé relevé, consiste à faire noter par l'enseignant la précision de la lecture orale de l'enfant¹⁸⁶. L'analyse des erreurs¹⁸⁷ est une méthode similaire. Dans les deux cas, l'enseignant analyse les diverses difficultés rencontrées par les enfants lors de la lecture orale. À l'heure actuelle, on trouve dans le monde entier de nombreux IRI produits par des éditeurs commerciaux et des ministères de l'Éducation. Pendant que les enfants lisent un texte, l'enseignement observe leurs points forts et leurs points faibles, leur pose des questions pour tester leur compréhension et leurs

185. Aux États-Unis, l'importance accordée aux tests s'étend à tous les tests de performance scolaire utilisés pour prendre des décisions capitales en matière d'évaluation des élèves du primaire et du secondaire. Ces tests sont administrés dans tous les États, dont beaucoup ont créé leurs propres tests de performance. Il s'agit presque toujours de tests critériels, par opposition aux tests normatifs commerciaux.

186. Clay, 1991.

187. Goodman et Burke, 1972.

connaissances, et note des informations quantitatives et qualitatives. Ces évaluations sont informelles et diagnostiques, car l'administration de l'IRI est adaptée à chaque élève et les observations ne mettent pas l'accent sur les données uniformes ou comparatives. Les IRI contiennent généralement des évaluations sur la précision de la lecture à haute voix¹⁸⁸, des listes de mots adaptés au niveau de l'enfant (vocabulaire visuel), des questions de compréhension, des rubriques pour reformuler ce qui a été lu, et des passages qui vont du niveau préprimaire au collège. Certains incluent des procédures d'évaluation des préacquis, de la compréhension orale, des lectures répétées ou de la lecture silencieuse. Certaines tâches de l'EGRA sont similaires à cette tradition ancienne, bien que l'EGRA n'ait pas une vocation de diagnostic au niveau de l'apprenant individuel^{189, 190}.

Dans les pays industrialisés, le principal motif d'utilisation des IRI avec les lecteurs débutants est la détection des difficultés des enfants, afin d'axer les cours de soutien sur ces compétences. Trop souvent, les difficultés rencontrées par les enfants au début de l'apprentissage de la lecture ne sont pas détectées avant la 2^e ou la 3^e année d'école, une situation exacerbée historiquement par les effectifs chargés des classes du primaire, le peu de temps disponible pour évaluer les enfants individuellement et l'insuffisance des outils d'évaluation à la disposition des enseignants. Plus vite les difficultés sont détectées, plus tôt on peut proposer des cours de soutien (à condition que ces ressources soient disponibles) pour améliorer diverses compétences en lecture. Une deuxième raison importante pour utiliser ces répertoires est de recueillir des données sur la progression des enfants en lecture. Les IRI sont rapides, flexibles, contrôlés par l'enseignant et centrés sur les élèves, autant de caractéristiques positives des évaluations en classe. Ils peuvent fournir des informations utiles aux élèves sur leurs progrès,

188. La précision de lecture est un facteur plus saillant dans les systèmes d'écriture moins transparents, comme celui de l'anglais. Utiliser la précision de lecture avec des langues transparentes (comme l'espagnol ou le swahili) serait moins fiable, du fait que les effets plafonds se produisent plus tôt en raison des avantages qu'offrent ces langues en termes de déchiffrage.

189. Quelques études récentes au Libéria et au Kenya utilisent des outils EGRA pour des interventions de diagnostic au niveau des écoles et des élèves (Gove, 2010).

190. On peut également noter qu'il y a une distinction importante entre les tests où « les instructions sont données oralement et ceux où elles sont données par écrit, plutôt qu'entre les tests où la réponse est orale ou écrite. Nous utilisons avec succès plusieurs tests de groupe avec des enfants de 1^{re} année [en Afrique de l'Est] où on leur donne des réponses écrites simples (qu'ils doivent cocher ou barrer) après avoir écouté des instructions orales » (M. Jukes, communication personnelle, 2010).

aux parents sur les performances et sur les compétences à améliorer, et aux enseignants sur la pédagogie et les textes les plus appropriés, autant de conséquences positives pour les acteurs concernés. Une troisième raison de recourir aux IRI – qui est une raison d’être initiale de l’EGRA – est la possibilité d’informer très tôt les responsables politiques d’éventuels problèmes de performance en lecture au niveau des écoles (voire du système éducatif).

Évaluation des compétences fondamentales en lecture (EGRA)

Présentation générale

L’EGRA ou *Early Grade Reading Assessment* (évaluation des compétences fondamentales en lecture) a été conçue pour fournir une batterie de tests destinés à évaluer les compétences de base en lecture, utilisables à l’échelle internationale par les pays en développement pour suivre le niveau de lecture des petites classes du primaire. Comme l’indiquent ses auteurs, les outils EGRA sont destinés à évaluer l’état du système à partir d’un échantillon d’élèves, le but principal étant de mesurer la « performance des élèves en lecture dans les petites classes, afin d’informer les ministères et les donateurs des besoins du système en vue d’améliorer l’enseignement »¹⁹¹.

L’objet, la théorie et la méthodologie des outils EGRA reposent globalement sur les mêmes recherches que les fameuses études américaines à grande échelle sur la lecture évoquées précédemment. Ces outils rappellent l’importance de cinq compétences essentielles pour commencer à lire¹⁹². Les outils d’évaluation EGRA s’inspiraient à l’origine de ceux développés pour les évaluations des indicateurs dynamiques des savoirs essentiels en lecture (INDISSE), d’usage très répandu aux États-Unis. L’accent mis par l’EGRA sur les éléments d’aptitude, tels que la connaissance de l’alphabet et le décodage, a permis d’adapter cette évaluation à diverses langues dans quelques-uns des pays les plus pauvres du monde.

L’EGRA est actuellement employée dans plus de 35 pays en développement¹⁹³. Elle comprend de nombreux sous-tests d’aptitudes et de connaissances qui ont démontré des capacités de prédiction des compétences

191. RTI, 2009, p. 6.

192. RTI, 2009 ; National Reading Report, 2000 ; Snow *et al.*, 1998. Jukes *et al.* (2006) ont également contribué à l’élaboration d’évaluations s’inspirant de l’EGRA.

193. Si l’EGRA a été expérimentée dans 35 pays, elle a été adoptée pour des finalités nationales plus importantes dans cinq pays en 2010, date de rédaction de la présente étude. L. Crouch, communication personnelle.

en lecture des élèves dans les classes primaires¹⁹⁴. La batterie de l'EGRA semble particulièrement adaptée aux lecteurs débutants (pour une fourchette d'âge qui reste à préciser) possédant diverses compétences et peut fournir des informations initiales précieuses aux responsables politiques qui veulent comparer les élèves et les écoles de différentes régions ou niveaux.

Trois critiques à l'encontre de l'EGRA méritent d'être notées. L'une concerne le manque relatif d'attention accordé par l'EGRA aux compétences linguistiques précoces (notamment le vocabulaire) et aux possibilités d'apprendre la langue, la lecture et l'écriture pendant la petite enfance. Une deuxième concerne la place primordiale faite aux aptitudes au décodage. Les compétences liées à la connaissance de l'alphabet, les concepts relatifs au texte imprimé, la conscience phonémique et la fluidité de la lecture orale (ORF) sont généralement les principales compétences évaluées chez les enfants qui apprennent à lire, le vocabulaire et la compréhension étant quelque peu négligés. Une troisième critique concerne le manque de différenciation des trajectoires de développement des diverses compétences associées à la lecture. La plupart des compétences liées au décodage sont apprises plus rapidement et maîtrisées au même niveau, contrairement au vocabulaire et à la compréhension, dont le développement est plus lent. Ainsi, les changements à courte échéance interviennent plus rapidement, grâce à l'apprentissage ou à des interventions sur certaines aptitudes liées au décodage, plutôt que sur des aptitudes spontanées telles que le vocabulaire. De ce fait, certaines études mettent en évidence d'importants effets plafonds et planchers dans les résultats de l'EGRA et des changements plus importants dans les aptitudes à décoder, suite de brèves interventions¹⁹⁵.

Un problème général avec l'EGRA, comme avec d'autres batteries d'évaluation des lecteurs débutants, est que les prescriptions concernant les politiques et les interventions découlent d'évaluations de précurseurs de la lecture extrêmement basiques et nécessaires. Les enfants doivent apprendre l'alphabet, les phonèmes et à décoder de façon fluide, mais ces aptitudes

194. Les données concernant la prévision de la performance ultérieure en lecture proviennent principalement d'études sur l'INDISSE entreprises aux États-Unis et en anglais (Roehrig *et al.*, 2007). La valeur de la fluidité de la lecture orale, en particulier, a été confirmée par des analyses statistiques. Mais des données concernant les élèves dont l'anglais est la seconde langue montrent également que la fluidité de la lecture orale n'est pas un facteur prédictif fiable (Riedel, 2007).

195. Voir le *chapitre 6* pour une analyse plus poussée des effets plafonds et planchers des études EGRA. Abadzi *et al.* (2005) est la première étude EGRA au Pérou ayant abordé ces questions.

seules ne suffisent pas pour maîtriser la lecture. Si ce sont les principaux référentiels utilisés pour évaluer la lecture dans une école, une région ou une nation, ils produiront quelques informations sur les compétences minimales nécessaires pour décoder le texte en sons, ce qui est important. Toutefois, les prescriptions éducatives subséquentes demandent aux enseignants d’enseigner ces compétences en premier (ou principalement), ce qui pourrait, à terme, défavoriser les environnements riches sur le plan linguistique et de l’alphabétisme, l’aide des adultes et les possibilités d’apprendre et de pratiquer la lecture et l’écriture. Il est important de prendre conscience de l’éventail limité des compétences évaluées par l’EGRA et de les compléter par des mesures plus larges, en particulier au fur et à mesure que les enfants avancent en âge et améliorent leurs compétences.

Dans certains pays, les utilisateurs de l’EGRA ont commencé à introduire un plus grand nombre de mesures, en ajoutant des tâches telles qu’un « repérage dans le texte imprimé », qui consiste à poser à l’enfant un certain nombre de questions liées à la compréhension des règles élémentaires régissant le texte imprimé, notamment où on doit commencer à lire sur la page, où continuer et où aller quand on arrive à la fin d’une ligne¹⁹⁶. Presque tous les enfants apprennent ces règles assez rapidement après avoir commencé l’école, de sorte que cette variabilité n’apparaît que pendant un temps limité chez les lecteurs débutants¹⁹⁷. D’autres études sont actuellement menées sur les contextes offrant diverses possibilités d’apprendre¹⁹⁸.

Objet et utilisations de l’EGRA

L’utilisation de l’EGRA au cours du cycle d’apprentissage des jeunes enfants est logique et analogue à l’utilisation de l’INDISSE dans d’autres pays (en particulier pour l’anglais aux États-Unis). Toutefois, l’éventail limité des compétences évaluées, ainsi que des prescriptions pédagogiques qui en découlent, constitue un problème potentiel. Par exemple, on a tendance à mettre l’accent sur l’ORF en tant que référentiel clé de l’apprentissage de la lecture. Si cette fluidité peut fournir des informations utiles quand les enfants apprennent à décoder, une fois que le décodage correct atteint un

196. RTI, 2009, p. 21.

197. Des instruments similaires ont été développés pour les jeunes enfants marocains où l’orientation dans le texte imprimé (suivi des règles du système d’écriture de droite à gauche) s’est avérée un facteur prédictif précoce de la lecture en arabe (Wagner, 1993). Voir également Clay, 2000.

198. Voir au *chapitre 6* les travaux de DeStefano et Elaheebocus, 2009.

rythme de 80 à 100 mots par minute, cette mesure devient moins utile pour distinguer les enfants qui comprennent bien de ceux qui comprennent moins bien¹⁹⁹. Actuellement, l'EGRA est surtout utile pour la détection précoce des aptitudes, mais relativement moins efficace pour les interventions aux stades ultérieurs de leur développement. De plus, trop insister sur la vitesse de lecture lors des évaluations et de l'enseignement peut envoyer aux lecteurs débutants (ainsi qu'aux enseignants et aux ministres de l'Éducation) un message erroné selon lequel l'objectif principal vise la vitesse en lecture²⁰⁰.

Utiliser l'EGRA pour détecter les besoins du système

Dans une étude réalisée aux États-Unis²⁰¹, l'EGRA fait apparaître une progression rapide de la fluidité de la lecture orale (ORF), les enfants passant de 20 à 100 mots lus par minute entre la 1^{re} et la 2^e année d'école. C'est la période pendant laquelle l'ORF est la plus névralgique. Cependant, la validité et l'utilité de l'EGRA risquent de diminuer avant et après cette période de progression rapide du décodage automatique ; elles peuvent aussi varier en fonction du contexte et de la langue. De plus, si la lenteur de la lecture ou du décodage peut effectivement freiner la compréhension, rien n'indique que la rapidité en lecture la stimule. L'EGRA est correctement décrit comme inadapté pour les évaluations à forts enjeux ou pour les comparaisons entre les langues. À ce jour, il doit être utilisé avec prudence pour les langues autres que l'anglais, dont les systèmes d'écriture transparents permettent l'acquisition rapide de l'aptitude au décodage et à

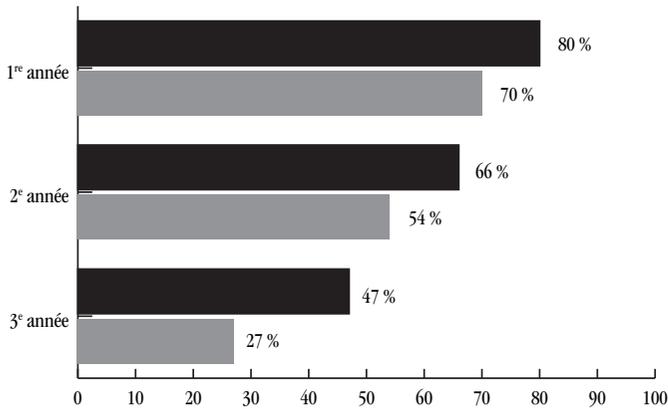
199. Fuchs *et al.* (2000, cités dans Fuchs, 2001, p. 247). Voir également Paris *et al.*, 2005 ; Paris et Hamilton, 2009.

200. Certains s'inquiètent de ce que l'impératif de fluidité de la lecture orale (ORF) puisse conduire à « aboyer devant le texte imprimé » (Samuels, 2007), les enfants auxquels on demande de lire le plus rapidement possible pouvant le faire sans comprendre ce qu'ils lisent. Kudo et Bazan (2009), dans une étude de terrain portant sur l'EGRA, au Pérou, adressent également une mise en garde : « [La fluidité de la lecture orale] peut avoir un effet négatif sur les enfants, créant trop de pression pour lire vite, ce qui leur enlève le plaisir de lire. ... Et ... la preuve de l'existence d'un lien de *cause à effet* entre l'entraînement à la fluidité et la compréhension n'est pas suffisamment convaincante, que ce soit en s'appuyant sur cette étude ou sur d'autres publications, pour promouvoir l'entraînement à la fluidité de la lecture en tant que stratégie unique pour améliorer le niveau de lecture » (p. 9, ce sont les auteurs qui soulignent). Le problème de l'« aboiement » ou celui de l'« appel de mots sans signification » n'est pas nouveau ; voir, par exemple, Allington (1983, p. 556) sur l'appel de mots sans signification et les mauvais lecteurs aux États-Unis.

201. RTI, 2009, Exhibit 3, p. 8.

une identification relativement plus précise des mots²⁰². Comme le montre la *figure 5.2*, l'EGRA est particulièrement efficace pour détecter les écarts importants aux premiers stades de l'apprentissage de la lecture, par exemple le nombre important d'enfants en Gambie incapables de lire un seul mot, même en 3^e année.

Figure 5.2 Gambie : Pourcentage d'élèves incapables de lire un seul mot, 2007 et 2009



Sources : Sprenger-Charolles, 2008 ; ministère de l'Éducation de base et secondaire, 2009.
Adapté de : Gove et Cvelich, 2010, p. 7.

Autres utilisations de l'EGRA

Selon la Trousse d'outils EGRA, cette forme d'évaluation pourrait également être utilisée (avec des modifications) à des fins de sélection et de suivi des progrès. Actuellement, cette approche semble poser des problèmes statistiques. Par exemple, utiliser l'EGRA dans le but d'évaluer les interventions pour favoriser l'aptitude au décodage en 1^{re} et 2^e année, à un stade où ces aptitudes évoluent rapidement, fera apparaître de larges variances d'aptitudes normalement distribuées ; mais, évaluer les mêmes enfants pour les mêmes aptitudes, avant ou après la période de progression rapide, produira des distributions non normales et des variances inégales, en raison du risque accru d'effets planchers et plafonds. Cela signifie que

202. Les préoccupations concernant l'ORF et d'autres sous-compétences de la lecture sont essentiellement prospectives, en particulier dans les PMA, dans la mesure où il n'y a pas encore de solide base de recherche.

les effets des interventions dépendront totalement du degré de maîtrise mis en évidence dans l'échantillon spécifique mesuré. Ainsi, avant d'utiliser l'EGRA à des fins plus larges de sélection ou de suivi des progrès, il sera judicieux de définir des référentiels de développement des aptitudes, en fonction des différentes années de scolarité et dans divers contextes²⁰³.

L'EGRA et sa théorie de la lecture précoce

La Trousse d'outils EGRA postule une théorie de la lecture précoce et de ses phases qui met l'accent sur les processus orthographiques et logographiques liés à la conversion de symboles en sons et en significations des mots²⁰⁴. Selon cette théorie, la lecture passe par l'identification de lettres et de phonèmes, le décodage et la prononciation. Entendre des mots ne fait pas appel à ces processus, mais écouter et lire nécessite de se souvenir de la signification et du son des mots. Ce modèle, comme de nombreux autres, décrit plusieurs relations clés, mais n'aborde pas la causalité, la séquence, le développement, le poids relatif des processus ou leurs différences. Par ailleurs, bien que l'évaluation de la compréhension des lecteurs débutants soit plus difficile, cet aspect mérite d'être davantage pris en compte par ceux qui utilisent l'EGRA ou des instruments similaires²⁰⁵.

203. Ceux qui travaillent sur l'EGRA semblent être conscients de ce problème et sont en train de rassembler des données pour élargir l'échantillonnage. Cela peut aussi passer par l'utilisation d'une base critérielle plutôt que normative pour comparer la progression des aptitudes.

204. RTI, 2009, Exhibit 6, p. 12-13.

205. Les évaluations de la compréhension orale ou des images suscitent beaucoup de débats, mais cela ne veut pas dire qu'il ne faut pas évaluer ces compétences. En fait, des chercheurs ont montré que la compréhension par les enfants de livres d'images dépourvus de mots (Paris et Paris, 2003) et d'épisodes de séries télévisées (van den Broek *et al.*, 2005) est un facteur prédictif des scores de compréhension écrite quelques années plus tard.

6. TESTS DE LECTURE : PROBLÈMES ET PERSPECTIVES

Quelques questions concernant l'évaluation des lecteurs débutants

Évaluer des lecteurs débutants ayant différents niveaux de compétence, de développement du langage et d'expériences dans des environnements d'alphabetisation favorables (WSE) et défavorables (PSE) suscite un certain nombre d'importantes questions.

- a. Quelles compétences faut-il mesurer chez les enfants qui commencent à apprendre à lire ? La plupart des évaluations précoces portent sur les aptitudes liées au décodage, telles que la connaissance du nom et du son des lettres, la conscience phonémique et des concepts relatifs au texte imprimé. Ces compétences peuvent être évaluées relativement facilement et rapidement, mais les résultats doivent être interprétés avec prudence, parce que (comme il a déjà été évoqué) tous les lecteurs confirmés les acquièrent à des niveaux asymptotiques en quelques années. Elles indiquent donc un avantage *relatif* du point de vue du développement au début de l'apprentissage, plutôt qu'une différence stable et pérenne des aptitudes individuelles à la lecture²⁰⁶. Des études récentes menées aux États-Unis viennent corroborer ce point de vue, en montrant que la fluidité de la lecture orale (ORF) en 2^e année n'apporte qu'une contribution modeste aux mesures ultérieures de la compréhension écrite en 6^e année ; l'ORF (mesures

206. Les données concernant les interventions sur la lecture qui se concentrent principalement sur le décodage sont mitigées. En Allemagne, par exemple, Landerl et Wimmer (2008, p. 159) indiquent que : « [L]es conclusions de l'étude longitudinale actuelle, qui a suivi l'évolution de la fluidité de la lecture et de l'orthographe d'enfants germanophones depuis le début de la 1^{re} année jusqu'à la 8^e année, confirment une fois de plus que pour les enfants qui s'engagent dans l'apprentissage de la lecture et de l'écriture avec certains facteurs de risque faciles à identifier, le pronostic à long terme est extraordinairement pauvre. » De même, Paris et Paris (2006, p. 55) constatent que « l'enseignement explicite des phonèmes permet une meilleure connaissance de l'alphabet et une meilleure reconnaissance des mots, ce qui concerne généralement uniquement les enfants qui ont les compétences alphabétiques les moins développées. Une meilleure reconnaissance des mots peut permettre à ces enfants de comprendre ce qu'ils lisent, mais ce n'est ni un lien suffisant ni un lien de cause à effet ». Bien sûr, dans les pays de mise en œuvre accélérée de l'EPT et dans d'autres contextes de PSE, un enfant qui n'apprend pas les rudiments du décodage et qui n'a pas accès à un enseignement ou à des interventions de soutien aura peu de chances de progresser jusqu'à la compréhension écrite.

INDISSE) peut en outre créer des faux positifs (enfants étiquetés « à risque » alors que les tests ultérieurs montrent qu'ils ne l'étaient pas) et des faux négatifs (élèves qui *ne* sont *pas* jugés « à risque » alors que, selon un test ultérieur, ils l'étaient)²⁰⁷. En général, l'EGRA comprend des mesures de l'initiation des enfants au texte imprimé, de leur connaissance des lettres, de leur conscience phonémique et de leur ORF, mesures qui font toutes apparaître un apprentissage rapide et des asymptotes similaires des progrès satisfaisants des enfants dans l'apprentissage de la lecture. Les autres compétences non liées au décodage, mais qui pourraient être incluses dans les évaluations futures, sont l'expressivité du langage, le vocabulaire, les erreurs de lecture orale et la reformulation du texte.

- b. Comment évaluer la compréhension ? Pour les enfants incapables de décoder les mots, il est possible de tester leur compréhension orale (comme dans l'EGRA) au lieu de poser des questions après leur avoir fait lire un texte. Lecteurs et non-lecteurs peuvent raconter le contenu du texte ou répondre à des questions standard. Ces questions doivent reposer sur des informations textuelles explicites et implicites, et être formulées avec soin, en veillant à ce qu'elles fassent appel à des informations données par le texte et non à des connaissances préalables. Cette mesure risque d'avoir peu de valeur si elle ne comporte que quelques questions sur un seul passage²⁰⁸.
- c. Le système d'écriture de la langue influe-t-il l'évaluation ? Dans la mesure où l'orthographe a une forte incidence sur l'apprentissage de la lecture et de l'écriture, il faut tenir compte des caractéristiques linguistiques (telles que l'homogénéité de la correspondance entre graphèmes et phonèmes [GPC] et entre phonèmes et graphèmes

207. Voir Valencia *et al.*, 2010. Cette intéressante étude sur le développement, effectuée aux États-Unis, a mis en évidence les handicaps des tests de lecture, comme l'INDISSE, axés sur une fourchette relativement étroite d'outils d'évaluation. La question des faux positifs et des faux négatifs est importante. Il est bien évident qu'aucun test ne permet de faire des prévisions parfaites ; mais l'étude de Valencia a fait apparaître des erreurs d'appréciation concernant jusqu'à 25 % des enfants de la 2^e à la 6^e année en utilisant des tests INDISSE, ce qui signifie que près d'un enfant sur quatre a été victime d'un diagnostic erroné. Une étude antérieure ayant constaté des faux négatifs similaires a été entreprise par Schilling *et al.*, 2007.

208. Dubeck (communication personnelle) considère que, même avec une ORF de 45 mots par minute, de nombreux enfants des pays pauvres ne pourront pas comprendre un texte s'ils n'ont pas des capacités de compréhension orale suffisantes dans la langue cible.

[PGC]²⁰⁹) lors de la conception des évaluations des apprenants qui commencent à lire et à écrire. De plus, il faut utiliser avec beaucoup de prudence les mesures du niveau de lecture axées sur des réponses écrites (comme les LSEA), compte tenu de l'asymétrie entre GPC (utilisées pour lire) et PGC (utilisées pour écrire). Ces questions sont particulièrement importantes pour les élèves qui apprennent à lire et à écrire en L2²¹⁰.

- d. Quels textes faut-il utiliser pour évaluer la lecture à haute voix ? Les textes retenus doivent être nombreux et inclure divers genres²¹¹ et niveaux de difficulté, afin de pouvoir déterminer la fluidité de la lecture orale de l'enfant (ORF). De nombreux passages, avec un contenu et du vocabulaire familiers, conviennent tout particulièrement. Des mesures de vitesse et de précision peuvent être effectuées pendant la lecture orale, mais il faut également envisager d'inclure des tests de narration et de compréhension. Si les enfants ne sont pas capables de décoder, les mesures d'ORF peuvent ne pas être adaptées²¹².
- e. Dans quelle(s) langue(s) faut-il tester l'enfant ? Comme indiqué auparavant, certains pays ont une politique pédagogique officielle, qui prescrit une seule langue nationale, différente de celle(s) que l'enfant pratique chez lui, tandis que d'autres ont instauré un mélange de langues maternelle, locale et nationale dans les programmes scolaires. À un niveau purement politique, certaines évaluations (comme le SACMEQ) appliquent simplement la politique linguistique du gouvernement. Cependant, afin de davantage prendre en compte le point de vue de l'enfant, les évaluations (ainsi que les enseignants et les parents) doivent s'appuyer sur ses forces linguistiques et autres, pour des raisons à la fois cognitives et affectives. Les évaluations hybrides étant particulièrement efficaces pour comprendre l'apprentissage de la

209. Voir à ce sujet l'analyse faite au *chapitre 5*.

210. Voir Genesee *et al.*, 2006.

211. L'EGRA a employé jusqu'à présent des textes narratifs, mais a l'intention d'utiliser aussi des textes non narratifs dans le futur (Gove, 2010).

212. L'EGRA (RTI, 2009, p. 21) a pris acte de ce sérieux problème d'effets planchers (scores de zéro ou proches de zéro) des tests d'ORF. Dans les PMA, les effets planchers sont très fréquents dans les écoles pauvres, de sorte que les ORF risquent de donner des distributions anormales, faussant par conséquent l'analyse des données.

- lecture dans un contexte local, il convient d'évaluer les enfants dans toute langue ou système d'écriture pertinent pour l'apprentissage²¹³.
- f. Qui évalue les enfants ? Il faut des évaluateurs qualifiés pour administrer les évaluations du niveau de lecture, car ils doivent juger les compétences en décodage (par exemple les bonnes rimes, le mélange de syllabes et la précision de la lecture orale), ainsi que la compréhension. Même dans les pays de l'OCDE, où il est généralement possible de trouver des évaluateurs qualifiés justifiant de diplômes universitaires, il reste difficile d'obtenir une grande fiabilité interévaluateurs. Dans les PMA, recueillir des données de bonne qualité reste une gageure à cause des coûts de formation²¹⁴.
- g. Les techniques d'évaluation sont-elles adaptées à la diversité des apprenants et des évaluateurs ? La plupart des évaluations (LSEA ou EGRA) utilisent des documents qui sont conçus et imprimés pour être utilisés par des évaluateurs locaux. C'est pourquoi les instructions doivent être claires et suivies de façon aussi rigoureuse que possible. Dans le cas des répertoires de lecture informels (IRI), les évaluateurs doivent être suffisamment formés pour piloter une tâche limitée dans le temps (avec un chronomètre), consistant notamment (pour l'ORF) à compter le nombre de mots qu'un enfant peut lire et à

213. Du point de vue empirique, il est probablement prématuré de définir le terme « pertinent ». La pertinence variera probablement d'un contexte à l'autre, mais cela ne signifie pas qu'elle ne peut être définie par les communautés ou les régions, voire au niveau national. Quand de nouvelles études hybrides du type EGRA auront été compilées, il sera plus aisé de comprendre quelles langues il faut enseigner et à quel niveau scolaire.

214. On peut s'interroger sur les coûts relatifs des formations utilisant différentes évaluations comme les LSEA, l'EGRA et d'autres. On a relativement peu d'informations sur le coût réel des formations, voir toutefois le *chapitre 7* ; les données disponibles, très limitées, montrent que les coûts de formation pour le PASEC (en pourcentage du coût total) sont approximativement les mêmes que dans les deux études EGRA mentionnées dans l'analyse des coûts. En outre, on en sait relativement peu sur la question de la fiabilité interévaluateurs. L. Crouch (communication personnelle) fait remarquer que cette fiabilité interévaluateurs est grande dans la plupart des études EGRA réalisées à ce jour, mais que la formation est plutôt conséquente. On peut aussi être tenté, en particulier dans les PMA, d'utiliser les enseignants comme évaluateurs. Mais ces derniers connaissent trop bien leurs élèves pour les évaluer objectivement, et peuvent en outre être sujets à des partis pris sociopolitiques locaux. Par ailleurs, parce qu'ils sont généralement plus éduqués que leur milieu local, les enseignants passent pour de bons évaluateurs, alors qu'ils ne le sont pas forcément. Ils peuvent en outre avoir, à l'égard de certains élèves, des attentes pouvant fortement influencer les performances de ces derniers. Globalement, des efforts sont encore nécessaires pour mieux comprendre les besoins en formation des évaluateurs.

déterminer s'il sait lire avec précision dans une langue qui n'est pas nécessairement sa langue maternelle ou celle de l'évaluateur. La clarté de l'articulation et la fluidité de l'expression orale de l'évaluateur sont également importantes pour la compréhension orale et la dictée, quand l'évaluateur lit à haute voix à l'enfant²¹⁵. Il existe également un problème sérieux qui n'est pas encore très bien compris : celui de savoir si la contrainte horaire peut introduire un parti pris dans les résultats²¹⁶, voire avoir des conséquences pédagogiques négatives (voir la section suivante).

215. On a longtemps objecté que les évaluateurs (comme dans le cas de l'EGRA) et les autres personnes administrant des évaluations transculturelles n'avaient pas nécessairement les compétences requises pour évaluer de façon impartiale et sans parti pris les compétences des enfants (pour un exemple récent, voir Jukes et Girgorenko, 2010). Se pose également la question des interférences : des enfants nommant des lettres anglaises au lieu de lettres swahili ou orthographiant des mots anglais avec une terminaison swahilie. Un autre problème se pose : comment évaluer des compétences en plein développement dans des régions où les enfants passent d'une langue à une autre à un stade donné de leur scolarité (par exemple du swahili à l'anglais) ? Merci à M. Jukes (communication personnelle) pour ces réflexions.
216. Il est surprenant que la question des tests de lecture chronométrés ait relativement peu retenu l'attention des chercheurs, surtout compte tenu des inquiétudes notoires et majeures suscitées par les techniques d'évaluation transculturelles en général (voir par exemple Brislin *et al.*, 1971). Une étude (Lesaux *et al.*, 2006) portant sur des lecteurs adultes s'est intéressée à l'effet que produit sur la compréhension écrite, avec ou sans déficience de lecture (lecteurs moyens et supérieurs contre mauvais et très mauvais lecteurs), le fait d'accorder plus de temps. Il a été demandé aux participants de lire à leur rythme (sans contrainte horaire, mais pas plus de 40 minutes) ou avec contrainte horaire (20 minutes). Tous les élèves présentant des difficultés de lecture ont eu du temps supplémentaire, mais les lecteurs affichant des performances normales ont eu des résultats similaires, avec ou sans contrainte horaire. De plus, les très mauvais lecteurs ont amélioré leur score lors du test non chronométré, mais cela n'a pas suffi pour avoir les mêmes résultats que les lecteurs moyens, contrairement aux mauvais lecteurs. S'il n'y a pas grand-chose dans cette étude qui puisse être appliqué directement aux PMA et aux PSE, il semble utile de regarder de plus près la question des tests chronométrés dans les PMA. Toutes les évaluations (LSEA et EGRA, par exemple) sont minutées, mais c'est l'EGRA qui insiste le plus sur l'aspect temporel, les évaluateurs utilisant systématiquement un chronomètre pour tester chaque enfant. Les instructions de l'EGRA précisent que l'évaluateur doit dire à l'enfant : « Répond le plus rapidement et le mieux que tu peux. » Comme dans tous ces types de tests (en particulier avec des enfants, mais aussi avec des adultes), les instructions peuvent donner lieu à des interprétations très différentes. Compte tenu de la présence importante du chronomètre, il ne serait pas surprenant que l'enfant se dise que le temps est l'élément principal de ce qui lui est demandé.

- h. Comment faut-il analyser les données de l'évaluation ? La validité et la fiabilité interévaluateurs des données d'évaluation des compétences en lecture dépendent de l'homogénéité des procédures d'administration et d'attribution des scores. Celles-ci découlent, à leur tour, de la qualification et de la formation des évaluateurs. La manière habituelle d'analyser les données provenant de batteries de tests consiste à utiliser des statistiques paramétriques, comme les corrélations de Pearson, les analyses factorielles, les analyses de régression, les analyses de variance et la modélisation à plusieurs niveaux axée sur les modèles linéaires hiérarchiques, mais la validité de ces méthodes avec des compétences en lecture limitées a été remise en question²¹⁷. Des analyses non paramétriques des données, ou peut-être des systèmes de référentiels fondés sur des critères spécifiques comme ceux décrits dans l'EGRA (*annexe B*), seraient peut-être plus adaptées.
- i. Comment faut-il interpréter les données sur les compétences en lecture précoce ? Les conceptions qui prévalent dans ce domaine considèrent ces compétences comme normatives dans un échantillon de population et ne font donc des déductions que sur ces populations d'enfants. Par exemple, l'affirmation fréquente selon laquelle la connaissance du nom et des sons des lettres dès la maternelle est le « meilleur facteur prédictif » du fait de savoir lire en 2^e année renforce la perception que la connaissance des lettres a un lien stable et durable avec le fait de savoir lire, ce qui n'est pas le cas à long terme²¹⁸. Cette corrélation semble également susciter des interprétations causales qui ont conduit à préconiser un enseignement explicite du son et du nom des lettres, ce qui n'est peut-être pas le moyen le plus productif d'enseigner la lecture à ce stade.

Implications pédagogiques des évaluations

Les évaluations des compétences en lecture ont de nombreuses vocations différentes, notamment des fonctions formatives et sommatives. Les tests de lecture sommatifs sont utilisés pour comparer des groupes d'élèves de plusieurs écoles, régions ou nations, parce qu'ils concernent les niveaux moyens de performance à des tests normatifs ou critériels. En revanche, la raison d'être fondamentale de l'utilisation d'évaluations formatives pour évaluer la performance en lecture des élèves est de leur dispenser un enseignement adapté. Par exemple, des tests de sélection peuvent être

217. Paris, 2005.

218. Ibid.

utilisés pour placer les élèves dans les niveaux de scolarité appropriés, les répartir dans des groupes de lecture, ou pour leur fournir le matériel pédagogique adéquat. Des tests de suivi des progrès peuvent être administrés pour déterminer les réponses des élèves à l'enseignement et aux programmes scolaires, qui peuvent ainsi être ajustés le cas échéant. Les tests de diagnostic s'intéressent à des compétences ou difficultés de lecture spécifiques, et peuvent servir de base à un « enseignement différencié » pour certains élèves. Le recours aux évaluations pour guider la pédagogie est souvent appelé « évaluation *pour* l'apprentissage »²¹⁹ (par opposition avec l'évaluation *de* l'apprentissage) et connaît un engouement croissant. L'évaluation pour l'apprentissage est particulièrement pertinente dans les pays en développement, où les lecteurs débutants ont des expériences de la lecture et de l'écriture très diverses, parce qu'elle associe l'évaluation et l'enseignement dans des cycles en temps réel d'apprentissage, sujet qui est abordé au *chapitre 9*. Bien entendu, l'enseignement différencié exige des enseignants qualifiés et, à terme, davantage d'investissement dans la formation²²⁰.

Pourquoi les « bons » tests ne sont pas toujours bons pour l'enfant

Un test est généralement considéré comme « bon » s'il est statistiquement fiable et valide, comme on l'a vu au *chapitre 4*. Toutefois, en dépit de la fiabilité des statistiques, de sérieuses questions continuent de se poser sur la pédagogie la mieux adaptée pour les enfants.

- a. Les tests doivent être adaptés du point de vue du développement à tout échantillon d'enfants. Dans les PMA, où les compétences sont très hétérogènes, un test unique peut ne pas convenir à tous les élèves, même à niveau de scolarité égal. Les professionnels de l'éducation peuvent avoir besoin d'un plus grand choix d'évaluations pour les élèves qui ne savent pas lire, ceux qui commencent à lire ou ceux qui savent déjà lire couramment, afin que les évaluations soient adaptées au développement des élèves et à la finalité du test.

219. Black et Wiliam, 1998.

220. Il convient de noter ici que les approches SQC ne peuvent tout couvrir et il est évident que des outils de diagnostic complexes sont peu susceptibles de rentrer dans la catégorie de ce qu'on peut mettre en œuvre lors d'une version allégée de l'évaluation. C'est pourquoi la référence ici à des évaluations plus complexes sert surtout à montrer les conditions limites de ce que certaines personnes veulent faire, plutôt qu'à plaider pour ce que l'on devrait essayer de faire dans les PSE, par exemple dans les pays de l'Initiative pour la mise en œuvre accélérée de l'éducation pour tous, du moins à court terme.

- b. Dans le droit fil du point précédent, les tests de lecture devraient porter à la fois sur les aptitudes au décodage et sur la compréhension à différents niveaux de compétences en lecture, afin que les enseignants et les élèves soient conscients que ces deux compétences sont importantes et devraient bénéficier d'un temps d'enseignement convenable.
- c. Les tests doivent avoir une bonne validité corrélative²²¹ pour l'enfant, l'enseignant et les parents (ainsi que pour le système), afin que les informations résultant de l'évaluation soient directement utiles. Une évaluation peut avoir des conséquences négatives à long terme, si les tests réduisent l'étendue des programmes scolaires, si leur préparation occupe une part importante du temps d'enseignement ou s'ils n'évaluent qu'un éventail limité de compétences. Inversement, les tests qui sont équilibrés et correctement mis en œuvre ont des implications importantes concernant la responsabilité des résultats entre les divers acteurs (voir également le *chapitre 9*).
- d. Les éducateurs ont besoin à la fois d'évaluations formatives et sommatives du niveau de lecture pour des élèves dotés de différentes capacités et pour divers objectifs. Les décideurs doivent concevoir des systèmes d'évaluation flexibles afin que les enseignants aient le contrôle des tâches d'évaluation dans leurs classes et aient accès aux données, afin que les évaluations puissent guider et améliorer l'enseignement. Quand ces éléments sont en place, on peut obtenir au niveau de la classe un diagnostic extrêmement utile pour améliorer les performances²²². Les implications pédagogiques de l'évaluation sont peut-être l'aspect le plus important pour les décideurs, parce que des évaluations utiles encouragent l'apprentissage et motivent autant les enseignants que les élèves, alors que des évaluations médiocres restreignent les programmes scolaires, déqualifient et démotivent les enseignants, et frustrent les élèves.
- e. Tests chronométrés et rapidité de lecture. Dans l'EGRA, dans l'ORF et dans plusieurs autres sous-tests de fluidité, l'évaluateur utilise un chronomètre pour limiter le temps de réponse des élèves à 60 secondes maximum. Cela permet d'uniformiser la durée du test et d'accélérer celui-ci, mais cela peut aussi envoyer le signal à l'élève, à l'enseignant, voire aux parents, que la vitesse de lecture est un référentiel essentiel d'une lecture correcte. En effet, certains décideurs qui utilisent l'EGRA se sont plaints de ce qu'une vitesse de

221. Linn, 2000.

222. Merci à B. Prouty pour cette remarque.

lecture donnée (par exemple 60 mots corrects à la minute) constitue un seuil minimal pour un niveau correct de lecture dans les PMA²²³. Bien que l'on puisse débattre des mérites de cette conclusion²²⁴, beaucoup partagent l'opinion selon laquelle les enfants auxquels on demande de lire au-delà d'une vitesse confortable pour la compréhension (en particulier dans une seconde langue) se trouvent dans une situation où ils peuvent répéter ou dire des mots d'un texte sans nécessairement les comprendre²²⁵. Une étude de terrain en cours au Kenya corrobore ce sujet de préoccupation, en ce sens que les enfants sont capables de

223. Abadzi *et al.*, 2005 ; Abadzi, 2008 ; RTI, 2009.

224. Certains soutiennent depuis longtemps qu'il existe des arbitrages (en particulier entre la vitesse et la précision) dans le processus cognitif en général et dans la lecture en particulier (pour la lecture, voir Stanovich, 1980). Pousser un élève à lire plus vite peut l'amener à adopter des stratégies inadéquates pouvant conduire à un déclin de son aptitude à lire. « Une autre explication plus plausible du résultat est que les lecteurs plus chevronnés adoptent une stratégie de lecture plus risquée que les jeunes lecteurs. En fait, pour tenter d'accélérer leur rythme de lecture, les adultes plus âgés peuvent s'appuyer davantage sur des informations parafovéales partielles. Cela peut leur permettre de parcourir le texte plus rapidement, mais avec pour conséquence possible d'avoir à revenir plus souvent sur des passages antérieurs pour clarifier du texte qui n'a pas été correctement traité avant » (Rayner *et al.*, 2006, p. 457).

225. Voir le renvoi précédent à Samuels (2007). Plusieurs études ont en outre porté sur la fluidité de la lecture orale (ORF) et l'impact de la pression en faveur des mesures de la vitesse de lecture. Hudson *et al.* (2009, p. 15) indiquent que « l'effort mental au service de la compréhension post-lexicale est un ingrédient majeur de fluidité de la lecture, ce qui signifie que plus le lecteur passe du temps à traiter le sens du texte, moins la lecture du texte en question est fluide ». De plus, Colon et Kranzler (2006) se sont intéressés aux conséquences de l'injonction adressée aux enfants de « lire plus vite » et ont constaté que : « Quand on leur demande de lire le plus vite possible, en moyenne, les élèves lisent correctement beaucoup plus de mots par minute et font beaucoup plus d'erreurs que lorsqu'on leur demande de faire de leur mieux ou simplement de lire à voix haute ». Ils concluent en outre que les différences dans les instructions ou dans l'interprétation des instructions (un problème sérieux dans les PSE) peuvent se traduire par d'importantes variations des scores. Naturellement, limiter les tests dans le temps semble inévitable dans la plupart des évaluations effectuées à l'école, ne serait-ce que pour des raisons d'efficacité. Le problème, ici, est que si un test chronométré spécifique (comme l'ORF) devient fortement associé à une intervention pédagogique, il existe alors un risque qu'un enseignement inapproprié reçoive le soutien des diverses parties prenantes, y compris des enseignants.

lire plus vite (meilleure ORF) en anglais (L2) qu'en swahili (L1), mais leurs scores de compréhension sont plus élevés en swahili²²⁶.

Il y a aussi un risque que les élèves cherchent à faire plaisir à l'enseignant ou à l'évaluateur en lisant vite, quitte à ignorer le sens du texte, ou à utiliser des stratégies inadéquates pour lire efficacement²²⁷. Cette question mérite que l'on y prête davantage d'attention, compte tenu du développement de l'ORF.

Réponse aux interventions et formation des enseignants

Une approche récente pour relier évaluation et enseignement est en train de gagner rapidement du terrain : elle est appelée *Réponse aux interventions*²²⁸. Dans cette approche à trois niveaux de l'enseignement, les élèves qui ont des difficultés avec l'enseignement normal en classe (1^{er} niveau) reçoivent davantage de tutorat, de temps d'instruction, ainsi que des cours de soutien dans leur salle de classe, tous les jours ou une fois par semaine (2^e niveau). Ceux qui n'arrivent pas à progresser avec les interventions de niveau 2 passent au niveau 3, qui est généralement un programme de tutorat intensif. En théorie, la réponse aux interventions est un bon moyen de relier évaluation et enseignement, mais tout dépend des évaluations spécifiques utilisées pour mesurer les progrès. Avec l'INDISSE et l'EGRAs, des tâches comme la fluidité de la lecture orale (ORF) peuvent devenir le référentiel de réussite et, par conséquent, la cible pédagogique. Dans les PMA, si des techniques aussi individualisées sont déployées, la question

226. Piper *et al.*, 2011. Les auteurs ont également constaté que les aptitudes à identifier les mots et à décoder des pseudo-mots étaient davantage prédictives de la fluidité de la lecture orale et des scores de compréhension dans la langue maternelle (swahili) qu'en anglais, ce qui donne à penser que des capacités limitées en vocabulaire à l'oral pourraient être responsables des faibles niveaux de compréhension.

227. Bien entendu, certains affirment que c'est précisément la lenteur de la lecture (dans les PSE) qui empêche les enfants de saisir le sens du texte, comme dans le cas des élèves des écoles maliennes (M. Keita-Diarra, communication personnelle). Ainsi qu'il a été indiqué auparavant, la vitesse de lecture peut également dépendre de la langue d'enseignement, selon qu'il s'agit de la L1 ou de la L2 de l'enfant.

228. Justice, 2006.

de la formation des enseignants prendra encore davantage d'importance, et toute amélioration de la conception pédagogique ira dans ce sens²²⁹.

Y a-t-il des limites cérébrales à l'apprentissage de la lecture ?

On a avancé que l'une des principales raisons pour intervenir tôt dans l'apprentissage de la lecture était qu'un ensemble de contraintes neurologiques pèsent sur l'enfant en développement et que, s'il ne lit pas à un âge donné, il risque de ne pas recevoir suffisamment d'apports²³⁰. Les données disponibles ne semblent pas montrer que les enfants (ou adultes) normaux sont pénalisés de manière significative par des limites cérébrales (ou biologiques) s'ils apprennent à lire (ou à faire à peu près n'importe quelle autre tâche cognitive) avec plusieurs mois ou années de retard²³¹. Cela ne veut pas dire pour autant qu'il n'y a pas, chez les êtres humains, de contraintes cognitives à l'apprentissage : il en existe bien. Mais un retard

229. Cette analyse n'est pas spécialement axée sur la formation des enseignants, bien que les implications des évaluations des compétences en lecture soient importantes pour cette formation. Selon M. Jukes (communication personnelle), ce qu'il faut, c'est comprendre comment changer de façon durable les pratiques pédagogiques dans l'ensemble d'un système éducatif, quand celui-ci est confronté à des problèmes de classes surchargées, d'absentéisme des enseignants et autres difficultés. Certes, une intervention bien pensée peut améliorer la qualité de l'enseignement et favoriser ainsi l'apprentissage chez les enfants, mais une fois le projet terminé, les choses reprennent leur cours normal.

230. Abadzi, 2008 ; Shawitz, 2003. Plus récemment, Shawitz et Shawitz (2008) ont fait valoir que « des agents pharmacothérapeutiques » peuvent se révéler très utiles pour renforcer la capacité d'attention et, à terme, la fluidité de la lecture chez les lecteurs dyslexiques. Il a également été avancé que les adultes analphabètes doivent faire face à des limites neurologiques lorsqu'ils apprennent à lire (Abadzi, 2006).

231. Voir Hruby et Hynd (2007) pour une analyse particulièrement pénétrante de la littérature neuroscientifique sur la lecture. Selon eux : « Il ne faut pas oublier que les modèles de facultés mentales de la lecture sont pour l'essentiel des analogies descriptives visualisables ... Ils éclairent une notion mal comprise par une autre qui l'est mieux. Mais cela implique toujours l'identité du phénomène étudié *et de quelque chose qu'il n'est pas*. Si de telles analogies peuvent être une source d'inspiration pour élaborer une théorie, elles sont fondamentalement fausses » (p. 551 ; les italiques sont dans le texte original). Pour un argument similaire sur les programmes d'intervention en matière d'éducation de la petite enfance, voir Hirsch-Pasek et Bruer (2007). Pour une désagrégation plus poussée des théories fondées sur l'inné (le cerveau) et l'acquis (théories sociales/expérientielles) du comportement des jeunes enfants, voir Kagan (2008) qui conclut qu'« un concept qui prétend expliquer qu'une compétence psychologique est un processus inhérent à l'organisation du cerveau du nourrisson n'a pas besoin d'avoir la même signification que celui qui explique un processus exigeant une grande expérience » (p. 1619).

dans la trajectoire d'apprentissage, s'il n'est pas souhaitable (en particulier pour les enfants d'âge scolaire), n'implique pas qu'un rattrapage est peu probable, voire impossible²³². Les vifs arguments postulant une fonction cérébrale compromise chez des enfants par ailleurs normaux pour apprendre à lire, parler ou exécuter d'autres activités cognitives, semblent être, en l'état actuel des choses, exagérés et non étayés par la littérature scientifique. Ce que l'on peut dire, c'est que l'apprentissage précoce est précieux car il laisse plus de temps à l'enfant pour acquérir une compétence, et lui permet de progresser conformément aux normes typiques d'un environnement d'alphabétisation favorable (WSE)²³³.

Une autre façon d'aborder l'argument cérébral consiste à considérer la valeur ajoutée des neurosciences par rapport à celle des sciences sociales et comportementales. Quand on travaille avec des enfants normaux, il est tout simplement plus plausible de penser que les apports sociaux et cognitifs (comme la présence de livres à la maison, le fait que les parents lisent des histoires à leurs enfants et l'accès à la langue d'enseignement) expliquent une variance plus grande dans le processus d'apprentissage que les différences minimales liées aux liaisons neuronales²³⁴.

232. Naturellement, cela ne veut pas dire que les retards d'apprentissage soient bons pour les enfants. Les enfants vivant dans des environnements d'alphabétisation défavorables (PSE), qui mettent trois ans pour apprendre l'alphabet et cinq pour parvenir à une relative fluidité de la lecture passent à côté d'une grande partie de ce qui est enseigné dans les programmes scolaires et accumulent tellement de retard que l'échec scolaire est probable. Ces retards d'apprentissage, ainsi qu'il est affirmé ici, sont liés à des facteurs sociaux et n'impliquent ni explications ni limites neuroscientifiques.

233. Les limites cérébrales au cours de l'évolution individuelle (ontogénie) n'impliquent pas qu'il n'y a pas de limites à la façon dont le cerveau humain traite les informations. Comme nous l'avons montré à propos de l'« automaticité » de la lecture, il y a des normes qui décrivent à quelle vitesse un lecteur chevronné peut lire, ainsi que les problèmes que les lecteurs dyslexiques rencontrent pour lire. Ces constatations montrent simplement qu'il existe des normes pour la fonction cérébrale et qu'elles peuvent être définies dans divers échantillons d'individus. Dans le domaine de la dyslexie (par opposition avec le terme « normal » employé plus haut), il existe des preuves écrasantes d'anomalies neurologiques ; voir, par exemple, Eden et Moats, 2002.

234. Pour une étude utile sur l'accumulation de retards dans l'acquisition de compétences cognitives, voir Stanovich (1986). Pour un argument axé sur les neurosciences, voir OCDE (2002) ; Dumont *et al.* (2010) est une illustration d'une approche plus récente et plus générale de l'apprentissage adoptée par l'OCDE.

Évaluer dans des contextes de grande pauvreté

Les observateurs qui travaillent dans les environnements d’alphabétisation défavorables (PSE), typiques des pays pauvres, ont tous connu ce qui est décrit dans l’histoire d’Aminata au *chapitre 1*. L’enfant moyen (celui qui est assis derrière les deux fameux premiers rangs de la classe), outre qu’il retient peu l’attention de l’enseignant, est souvent mal préparé à rester assis en classe pour diverses raisons : préparation inadéquate à la maison, mauvaise alimentation, matériel scolaire médiocre et enseignant qui a tendance à surtout s’intéresser (quand il n’est pas absent pour diverses raisons) aux enfants qui ont les meilleures chances d’apprendre et à ceux dont les parents comptent au sein de la communauté²³⁵. De nombreux problèmes concourent à ce que ce soit les élèves pauvres qui tirent le moins de profit de l’éducation. Bien connu sur le plan socioculturel, ce problème pourrait aussi être considéré du point de vue cognitif individuel. Dans les pays en développement, par exemple, de nombreux enfants vivant dans des PSE font preuve de lassitude, de léthargie ou d’inattention en classe. Ce comportement est dû en partie à la malnutrition, à des maladies chroniques, à la fatigue, à la longueur des trajets pour se rendre à l’école, ainsi qu’à d’autres facteurs. On a moins parlé des conséquences de ce type de difficultés physiologiques sur l’acquisition de la lecture. Des études récentes semblent montrer que, dans ce contexte, les enfants peuvent « décrocher », au sens où ils ne se concentrent plus sur la tâche de lire et deviennent inattentifs en classe. Cela a notamment pour conséquence que les enfants donnent l’impression de lire, alors que leur compréhension du texte est très limitée, voire nulle²³⁶. Plusieurs études EGRA menées sur le terrain, comme celles décrites plus en détail au *chapitre 5* et plus loin, font état d’enfants scolarisés qui sont incapables de lire un seul mot.

235. Voir Samoff (2003) pour une étude plus générale des implications politiques des PSE et des limites matérielles dans les PMA.

236. Bien que les conséquences négatives de l’inattention ou de ce qu’on appelle parfois le « décrochage » sur la construction d’un modèle paraissent simples et directes, les processus cognitifs ou perceptifs qui commandent le comportement quand l’esprit vagabonde restent peu clairs. Il y a deux explications au « décrochage » à l’origine de ce que l’on appelle la « lecture machinale ». Selon Smallwood *et al.* (2008) : « La lecture machinale trahit une rupture dans le contrôle descendant de la compréhension ; la lecture s’effectue alors soit par simple contrôle moteur soit par référence à des caractéristiques relativement élémentaires du seul texte. » En d’autres termes, dans les PSE, il n’est pas rare, en particulier pour les mauvais lecteurs, de se surprendre à lire machinalement plutôt que de faire l’effort d’essayer de comprendre. C’est naturellement l’une des raisons d’être des évaluations du type SQC : étayer par des données les signes précoces d’échec.

Méthodes d'évaluation des compétences en lecture : quelques observations complémentaires

Évaluations internationales

Le Programme international de recherche en lecture scolaire (PIRLS) et le Programme international pour le suivi des acquis des élèves (PISA) restent les évaluations internationales les plus connues et les plus utilisées dans le domaine de la lecture. Comme nous l'avons déjà expliqué, chacun de ces tests a été conçu de façon rigoureuse à partir de bases théoriques solides, par des spécialistes hautement qualifiés dans le domaine technique et statistique. Comme nous l'avons également fait remarquer, les enquêtes PIRLS et PISA diffèrent par la population ciblée (selon l'âge ou les niveaux de scolarité), ce qui a suscité de nombreux débats au fil des années. De plus, chaque test a été soumis à un examen détaillé, en vue de leur utilisation dans les pays en développement, pour au moins trois raisons liées à la comparabilité. Premièrement, dans le cas du PISA, les proportions d'élèves scolarisés dans les classes post-primaires varient considérablement entre la plupart des pays de l'OCDE et les PMA, ce qui donne des échantillons faussés et non comparables d'élèves²³⁷. Deuxièmement, chaque test ne peut être adapté à la culture que de façon approximative, parce qu'il est impossible d'obtenir une base linguistique vraiment comparable pour les tests, à cause de la trop grande variabilité qui existe dans le monde. La situation est encore plus problématique dans les PMA, où de nombreux enfants doivent passer le test dans leur deuxième ou troisième langue. Troisièmement, dans les PMA, les scores des apprenants peuvent être si bas qu'ils ne sont plus statistiquement fiables pour les populations échantillons, un grand nombre d'enfants se situant dans le bas de l'échelle statistique²³⁸.

Il est en outre difficile, dans ce type de test normalisé, de déterminer pourquoi un élève ne répond pas correctement à une question. Est-ce parce qu'il n'a pas pu lire la totalité du texte ? Parce qu'il a été incapable de se souvenir du contenu du texte au moment de répondre aux questions ? Ou bien, a-t-il été capable de lire le texte, mais a-t-il simplement décidé de ne pas le faire ? Par ailleurs, les enquêtes PIRLS et PISA ne comportent pas de mesures pour évaluer le niveau de compréhension de la langue orale, de

237. Voir UNESCO, 2004, p. 48.

238. Le développement des pré-PIRLS montre que le problème du changement d'échelle a été reconnu et sera probablement traité. La question contestée de la comparabilité linguistique et culturelle risque, en revanche, de demeurer.

sorte que les résultats des autres sous-tests de lecture (en particulier dans les contextes multilingues) sont plus difficiles à interpréter.

Une autre limite sérieuse à l'utilisation des enquêtes PIRLS et PISA à des fins pédagogiques dans les pays en développement est que ces tests arrivent trop tard dans le cycle d'acquisition de la lecture par l'enfant. Dans les deux cas, ils sont administrés vers la fin (PIRLS) ou juste après la fin de l'école primaire (PISA), alors que les problèmes identifiés dans les environnements (d'alphabetisation) défavorables (PSE) montrent que les enfants échouent à un stade beaucoup plus précoce du cycle scolaire. Savoir que les enfants lisent mal en 6^e année (pour ceux qui arrivent jusque-là, et ils sont souvent peu nombreux dans beaucoup de pays concernés par le PME) offre relativement peu d'intérêt sur le plan pédagogique, car les problèmes surviennent beaucoup plus tôt. En outre, dans les PMA, beaucoup d'enfants abandonnent l'école avant la 6^e année. Et même si ces enfants sont détectés, les méthodes de diagnostic et d'intervention précoce fonctionnent beaucoup mieux avec de jeunes enfants, souvent au stade du décodage.

Évaluations régionales

L'enquête du SACMEQ a beaucoup de points communs avec les évaluations des compétences en lecture du PIRLS et du PISA. Conçu pour les sociétés et cultures complexes de l'Afrique orientale et australe, le SACMEQ adhère largement à une préoccupation partagée par tous les ministères de l'Éducation : l'importance des performances scolaires. C'est pourquoi seules les langues prescrites par les ministères nationaux sont utilisées (principalement l'anglais, mais aussi le portugais au Mozambique et le swahili en République-Unie de Tanzanie et à Zanzibar). Il convient également de noter que le SACMEQ a fait un effort important pour considérer les évaluations comme une occasion de renforcer les capacités. Comme l'a fait observer une étude récente : « Une caractéristique particulière de cette approche est la formation "sur le tas" des planificateurs qu'elle cherche à impliquer directement dans la conduite des études²³⁹ ». Le renforcement des capacités locales confère de la valeur ajoutée à la région. Comme pour le PIRLS et le PISA, la mise en œuvre, les analyses et la production d'un rapport complet sur l'étude peuvent prendre jusqu'à cinq ans²⁴⁰.

239. Ladipo *et al.*, 2009, p. 87.

240. Voir Ross *et al.* (2005) pour une étude approfondie des 14 phases principales du travail du SAQMEC.

Le PASEC vise à mesurer les compétences en lecture, ainsi que d'autres compétences linguistiques, telles que la grammaire et la conjugaison en français (*annexe A*). Le PASEC semble être la seule évaluation requérant une bonne connaissance de la grammaire, perçue comme importante dans la tradition pédagogique francophone, même si les données prouvant l'importance de ces compétences pour la compréhension écrite globale du français sont rares ou inexistantes. D'autre part, le PASEC ne semble pas évaluer directement les correspondances graphèmes-phonèmes (GPC), ni la compréhension orale. Le PASEC comprend des tâches d'appariement de mots ou de phrases et d'illustrations, ainsi que des textes à trous utilisés pour évaluer la compréhension des mots, des phrases et des textes. Ces tests présentent plusieurs avantages potentiels, comme l'administration à des groupes et leur facilité d'emploi avec de jeunes enfants, ainsi que la possibilité d'utiliser la même image pour plusieurs langues²⁴¹.

Évaluations EGRA : études de terrain

[U]ne approche de la lecture directe, simple, explicite et intense ... paraît être considérée favorablement par les enseignants, les chefs d'établissement et les parents et semble capable, pour certaines mesures, d'améliorer les résultats assez rapidement²⁴².

Contrairement aux évaluations comparatives internationales et régionales, l'évaluation des compétences fondamentales en lecture (EGRA) vise à mesurer les performances en lecture des enfants qui commencent à apprendre à lire. Comme il a été dit plus tôt, l'EGRA est critiquable à plusieurs points de vue, et les études de terrain actuellement en cours devraient permettre de la perfectionner et de l'améliorer. Les principaux atouts de sa mise en œuvre dans les PSE sont les suivants : (a) l'EGRA est conçue pour prendre en compte les effets planchers (les niveaux les plus bas des élèves) ; (b) elle s'adapte à n'importe quelle langue ou système d'écriture, car il n'y a pas de contraintes de comparabilité stricte ; (c) les échantillons de population peuvent être plus petits puisque l'EGRA se veut (du moins à l'heure actuelle) un instrument de suivi plutôt qu'une évaluation nationale représentative à enjeu élevé ; (d) le temps nécessaire entre la conception des

241. Voir Wagner (1993, p. 88-92), où des enfants marocains ont été testés en arabe et en français. Voir également l'analyse du test de vocabulaire par l'image de Peabody (PPVT), dans RTI (2009, p. 35) ; des questions de droits d'auteur semblent constituer l'une des principales raisons qui font hésiter à utiliser ce test dans le cadre de l'EGRA. Pourtant, le PPVT est l'une des meilleures mesures (et l'une des plus simples) du développement du vocabulaire, quelle que soit la langue.

242. Crouch *et al.*, 2009, p. 2.

tests et la production d'un rapport complet peut être considérablement plus court que pour les autres évaluations majeures, grâce à sa taille globalement plus modeste.

Toutefois, le dimensionnement optimal des tests EGRA en fonction des nombreux contextes, langues et systèmes d'écriture dans lesquels ils sont actuellement administrés reste problématique. Cela signifie que plusieurs sous-tests poseront des problèmes d'effets planchers et plafonds, et qu'il faudra procéder à des ajustements au fur et à mesure de l'application de l'EGRA²⁴³. De plus, pour certains tests (comme la conscience phonémique), une administration correcte est difficile et pose des problèmes de formation des évaluateurs dans les endroits où existent de nombreuses difficultés spécifiques – autant de facteurs qui peuvent limiter sérieusement la fiabilité interévaluateurs.

Plusieurs études de terrain récentes²⁴⁴ illustrent à la fois les promesses et les difficultés de l'approche EGRA. Les brefs aperçus ci-dessous donnent des directions utiles pour poursuivre le travail avec les outils EGRA :

- Éthiopie²⁴⁵. Cette étude a été entreprise dans un contexte rural pauvre, sur un échantillon de 24 écoles et environ 450 enfants en 3^e année, répartis à peu près également entre filles et garçons. Elle portait principalement sur les variables d'opportunité d'apprendre, comme le nombre de jours d'école, la présence des enseignants, le lien entre les tâches des enseignants et celles des élèves et, en dernière analyse, la grande fréquence (36 %) d'élèves se situant au bas de l'échelle de l'ORF (c'est-à-dire zéro mot correct par minute). Cette étude décrit un environnement d'apprentissage similaire à celui dépeint dans l'histoire d'Aminata exposée dans l'*introduction*.
- Pérou²⁴⁶. Une étude a été entreprise auprès de 475 enfants, notamment pour déterminer s'il existe des différences entre l'EGRA administrée individuellement et les tests de compréhension écrite administrés en groupe. Les résultats montrent des corrélations mutuelles relativement

243. Par exemple, des élèves sénégalais et gambiens testés dans le cadre de l'EGRA n'ont pu lire qu'un petit nombre de mots et les scores d'une grande proportion d'élèves de 1^{re} année étaient au bas de l'échelle dans le test de lecture de mots. Les niveaux de lecture de mots étaient à zéro (bas de l'échelle) pour 80 %, 50 % et 71 % des élèves respectivement de L2 anglais, L2 français et L1 wolof. Voir Sprenger-Charolles, 2008*a, b*.

244. De nombreuses autres études sont actuellement au stade de projet ou en cours, mais ne sont pas examinées dans le présent ouvrage.

245. DeStefano et Elaheebocus, 2009.

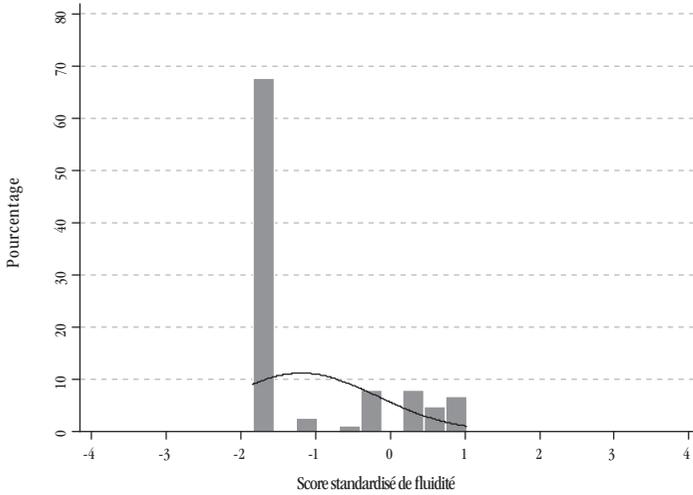
246. Kudo et Bazan, 2009.

importantes entre les deux formes d'évaluation. Les auteurs concluent que ces deux types de tests sont plus « complémentaires » qu'« irréciliables ». Pourtant, ces résultats ne semblent pas résister aussi bien à l'analyse dans le cas des enfants indigènes appartenant à la minorité ethnolinguistique des Ashaninka (*figure 6.1*), car les courbes de discrimination sont très différentes, laissant entendre qu'au moins l'un des tests pourrait être faussé avec ce type de populations²⁴⁷.

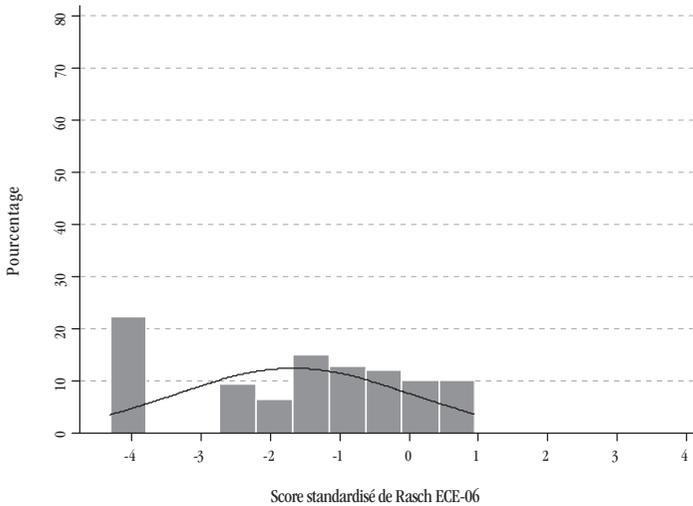
247. Chez les Ashaninka, en ce qui concerne cette étude réalisée au Pérou, il est impossible de déterminer si la méthodologie d'évaluation ou les items testés sont à l'origine des courbes de réponse contrastées de la *figure 6.1*. Les auteurs font une analyse intéressante de la faible corrélation entre la fluidité de la lecture orale (ORF) et la compréhension écrite, même si l'ORF est le facteur prédictif le plus fiable de tous les sous-tests ; ils constatent que, dans la population évaluée, la relation entre fluidité et compréhension est « non linéaire » (p. 47). Cette dernière observation n'est pas en contradiction avec d'autres parties de la présente étude. Il n'y a rien de surprenant, en tout cas, à ce qu'il y ait des corrélations positives entre les sous-tests de l'étude de terrain effectuée au Pérou. Depuis le début de la recherche sur les tests cognitifs, quel que soit le test, les individus qui obtiennent globalement de bons scores à un type de test cognitif ont souvent de bons résultats aux autres tests cognitifs. Fait intéressant, les auteurs indiquent également (p. 50) : « nous avons constaté que la relation [fluidité – compréhension] était plus forte chez les élèves Ashaninka ; une augmentation de la fluidité de la lecture se traduit par une augmentation de la compréhension plus importante que chez les élèves non-Ashaninka, même quand les autres facteurs restent constants. Cela pourrait impliquer que travailler sur la fluidité de bonne heure est encore plus important pour les enfants les plus défavorisés – dont les Ashaninka font partie – que pour les élèves avantagés, car cela semble avoir un impact plus grand sur leur capacité de comprendre ce qu'ils lisent ». En d'autres termes, il y a de multiples façons d'interpréter les résultats, en particulier pour les sous-groupes de population, et il faudra les étudier attentivement lors d'études ultérieures. On retrouve ici les mêmes différences à tous les niveaux d'études dans les intercorrélations entre ORF et compréhension observées par Valencia *et al.*, 2010, comme il a été évoqué auparavant.

Figure 6.1 Histogrammes des écarts types constatés chez les élèves Ashaninka (Pérou), lors d'un test de fluidité de la lecture orale et d'un test écrit administré en groupe (N = 40)

a. Test de fluidité orale (écarts types)



b. Test écrit administré en groupe (écarts types des scores de Rasch)



Adapté de : Kudo et Bazan, 2009, p. 33.

- Inde²⁴⁸. Un peu à l'instar de l'étude réalisée au Pérou, des chercheurs (Bihar et Uttarakhand) ont mené en Inde plusieurs études comparant l'utilisation de la mesure de la fluidité orale de l'EGRA et les évaluations READ INDIA (écrites) en hindi utilisées auparavant, de la 1^{re} à la 8^e année. Si les sous-études varient en taille et en contenu d'évaluation, les chercheurs ont évalué plus de 15 000 enfants au total. Comme pour celle du Pérou, l'étude indienne constate les hauts niveaux de fiabilité entre les divers instruments²⁴⁹. Les auteurs soutiennent, à juste titre, qu'il n'y a pas eu suffisamment d'évaluations des compétences en lecture en hindi et que, par conséquent, des analyses détaillées – comme celles de la présente étude – devraient ouvrir la voie à de nouvelles façons de mesurer la trajectoire d'apprentissage des élèves du primaire, ainsi que les niveaux de performance attendus. Une dimension importante du travail réalisé par READ INDIA est la participation de la société civile à la promotion des politiques de sensibilisation concernant les premiers stades de l'apprentissage de la lecture, aux niveaux national, régional et local. Dans une étude nationale récente portant sur les zones urbaines des États indiens, il a été constaté que 30 % des enfants de 3^e année étaient incapables de lire un seul mot dans leur langue maternelle (*tableau 6.1*).
- Kenya²⁵⁰. Il s'agit d'une étude expérimentale pour laquelle 40 écoles (sur 120) ont été sélectionnées, en vue d'une intervention comprenant une formation des enseignants de 1^{re} et 2^e années, les enseignants dits « de contrôle » ne bénéficiant pas de cette formation. Cette formation, qui s'est étalée sur plus d'une année, comprenait des cours hebdomadaires sur la conscience phonologique, le principe alphabétique, le vocabulaire, la fluidité et la compréhension. Des

248. ASER, 2009 ; Abdul Latif Jameel Poverty Action Lab (J-PAL) *et al.*, 2009.

249. Le rapport affirme également qu'il « corrobore la validité discriminante-convergente [car] nous avons constaté entre la Fluency Battery et le test de lecture ASER [READ] des corrélations plus fortes que chacune de leurs corrélations avec les tests de mathématiques ». Toutefois, d'autres tests de mathématiques et de lecture ont montré des corrélations plus fortes qu'entre certains tests de lecture. Ces observations viennent corroborer une conclusion fondamentale du présent ouvrage, à savoir que de nombreux tests utiles (quand les effets planchers et plafonds sont éliminés ou réduits) montrent de fortes intercorrélations. Ce n'est pas un problème en soi ; mais c'est un problème quand des affirmations sont faites sur la valeur de sous-tests spécifiques des évaluations des compétences en lecture (et pour une éventuelle mise en œuvre), dans la mesure où tant d'instruments d'évaluation sont fortement liés les uns aux autres.

250. Crouch *et al.*, 2009.

outils EGRA (en anglais et en swahili) ont été utilisés pour mesurer les progrès en lecture pendant la même période, dans les écoles concernées par l'intervention ainsi que dans les écoles témoins²⁵¹. En termes de conséquences directes sur les compétences en lecture, l'étude a constaté que l'intervention permettait de réduire le nombre d'enfants incapables de lire en swahili (mais pas celui des enfants incapables de lire en anglais).

Tableau 6.1 Pourcentage d'enfants, par classe, par niveau de lecture, toutes écoles confondues, 2010

Année	Rien	Lettres	Mots	Texte niveau I	Texte niveau II	Total
1	34,0	41,1	17,0	4,4	3,4	100
2	12,1	32,4	32,4	13,9	9,1	100
3	6,0	18,8	29,6	25,7	20,0	100
4	3,1	10,1	19,4	29,3	38,1	100
5	2,2	6,7	12,7	25,1	53,4	100
6	1,3	4,0	7,6	19,7	67,5	100
7	1,0	2,7	5,2	15,0	76,2	100
8	0,7	1,9	3,2	11,3	82,9	100
Total	8,3	15,9	16,8	18,2	40,9	100

Comment lire le tableau : chaque cellule indique le niveau de lecture le plus haut atteint par un enfant. Par exemple en 3^e année, 6 % des enfants ne savent même pas lire les lettres ; 18,8 % peuvent lire les lettres mais pas plus ; 29,6 % peuvent lire des mots mais pas un texte de niveau I ou supérieur ; 25,7 % peuvent lire un texte de niveau I mais pas de niveau II, et 20 % peuvent lire un texte de niveau II. Pour chaque classe, le total de toutes ces catégories exclusives est de 100 %.

Adapté d'ASER, 2009.

251. L'étude effectuée au Kenya fait un constat inattendu, à savoir des améliorations significatives (et indistingables) de la lecture à la fois dans le groupe d'intervention/expérimental et dans le groupe de contrôle. L'explication donnée est l'existence d'effets de « propagation » entre le groupe expérimental et le groupe de contrôle : les enseignants des écoles de l'échantillon de contrôle ont trouvé le moyen de reprendre à leur compte des connaissances des enseignants du groupe expérimental. Cette explication semble plausible, mais il est aussi possible qu'il y ait un effet « Hawthorne » dans l'étude et que la simple présence de chercheurs et les forts enjeux associés au fait de participer (ou non) à l'étude aient amené les enseignants à accorder plus d'attention à l'enseignement de la lecture dans ces localités.

- Libéria²⁵². Dans cette étude, une expérience d'essai comparatif randomisé a été entreprise en 2^e et 3^e années pour comparer trois niveaux d'intervention dans 60 écoles : témoin, « léger » (la communauté est informée de l'importance de la lecture) et « complet », les enseignants et les parents participant à une intervention dans le domaine de la lecture, inspirée du modèle EGRA de lecture (avec l'anglais comme langue cible)²⁵³. Les enseignants avaient environ trois mois pour prendre part à l'intervention avant la production d'un rapport sur une première série de conclusions. Ces conclusions à mi-parcours comparent une évaluation de base et trois mois d'aide à l'intervention. Elles révèlent que la mesure de la fluidité de la lecture orale (ORF) est plus sensible à l'intervention complète qu'à l'intervention légère mais que les deux interventions ont eu un impact positif sur cette mesure²⁵⁴. Toutefois, les différences d'impact sur la compréhension écrite semblent beaucoup plus faibles, ce qui pose la question de savoir si les interventions aux stades précoces de l'apprentissage de la lecture sont étroitement liées aux compétences générales en lecture.

Globalement, ces études et d'autres toujours en cours²⁵⁵ donnent une idée de l'énergie que suscite l'approche EGRA dans de nombreux pays. Certes, il y aura toujours des divergences d'opinion sur les meilleurs outils d'évaluation, mais les connaissances sur les débuts de l'apprentissage de la lecture dans les PMA progressent à un rythme productif et rapide.

252. Piper et Korda, 2009. L'étude (ébauche) à mi-parcours ne dit rien sur l'importance de la lecture en anglais ; on suppose que beaucoup d'élèves de l'échantillon, sinon la plupart, ont probablement une langue africaine comme langue maternelle.

253. L'intervention dite « complète » est décrite ainsi : les enseignants sont formés « sur la façon d'évaluer en continu les performances des élèves ; ils reçoivent une aide pédagogique fréquente à l'école, du matériel pédagogique et des livres ; qui plus est, les parents et les communautés sont informés des performances des élèves » (Piper et Korda, 2009, p. 5).

254. Dans ce rapport intérimaire du Libéria, il semble également y avoir quelques interactions importantes par genre et par sous-test que la collecte de plus amples données et une analyse plus poussée permettront probablement d'expliquer. Il apparaît en outre qu'il y a quelques différences élémentaires entre le groupe de contrôle et le groupe ciblé par l'intervention dès le début de l'étude ; on ne sait pas très bien, pour le moment, si ces différences auront une incidence sur l'interprétation en aval.

255. Abadzi (2010) a récemment compilé une synthèse actualisée des études de terrain EGRA, où l'accent est mis sur les mesures de fluidité de la lecture. Selon ce rapport, en septembre 2010, 29 des 71 pays éligibles au PME avaient administré une ou plusieurs enquêtes EGRA, ou d'autres études similaires sur les premiers stades de l'apprentissage de la lecture.

Indicateurs de compétence en lecture du PME

En octobre 2009, le PME a adopté deux indicateurs de compétence en lecture destinés à évaluer la qualité des écoles dans les pays participants²⁵⁶ :

- Proportion d'élèves qui, après deux années de scolarité, ont une fluidité de lecture et une compréhension suffisantes pour être capables de « lire pour apprendre » ;
- Proportion d'élèves capables de lire en comprenant ce qu'ils lisent, en fonction des objectifs pédagogiques de leur pays, avant la fin de l'école primaire.

Ces deux indicateurs d'apprentissage proposés appellent plusieurs commentaires. Premièrement, ils semblent dérivés du modèle EGRA, dont l'objectif principal est la détection précoce des problèmes de lecture, avant la fin de l'école primaire. Deuxièmement, ils apportent également une composante « compréhension » dans la combinaison, à un stade plus précoce que certains spécialistes de l'apprentissage de la lecture, mais plus en accord avec le point de vue adopté dans le présent ouvrage. Troisièmement, la notion de « lire pour apprendre » met davantage l'accent sur la qualité de l'apprentissage en classe, s'opposant implicitement à la mémorisation par cœur couramment observée dans les classes des PMA les plus pauvres. Enfin, ce que ces deux indicateurs *ne disent pas* est également intéressant : par exemple, les instruments qui pourraient être utilisés pour les établir et, notamment, l'importance des questions de première et seconde langues à l'école primaire²⁵⁷.

Perspectives et évolution des mesures d'évaluation de la lecture

D'ici à 2015, date d'échéance des Objectifs du Millénaire pour le développement (OMD), l'une des avancées du point de vue des objectifs de l'éducation de base sera une meilleure capacité de fixer des objectifs réalistes et d'en suivre la réalisation. Au cours des prochaines années, les gouvernements et les agences non gouvernementales disposeront d'un éventail plus large d'outils d'évaluation pour étudier les premiers stades et l'évolution de l'apprentissage de la lecture (ainsi que de l'écriture et des mathématiques) dans le but d'atteindre les objectifs internationaux et nationaux, ainsi que pour informer les parents, les élèves et les communautés.

256. www.educationfasttrack.org/themes/learning-outcomes/ (consulté le 31 janvier 2011).

257. Le présent ouvrage a été en grande partie rédigé avant que ces indicateurs ne soient annoncés par le PME ; il n'y a pour le moment pas suffisamment d'informations sur les résultats pour formuler de plus amples commentaires.

Le développement des évaluations des compétences en lecture présente cependant un certain paradoxe : plus le modèle d'évaluation est complexe (comme c'est le cas des LSEA internationales et régionales), moins les résultats sont transparents et se prêtent à une utilisation efficace dans les écoles, ce qui rend la réponse pédagogique plus difficile. Mais, plus l'outil d'évaluation est simple (un test d'ORF, par exemple), plus le danger d'une réponse pédagogique simpliste est grand, comme si l'outil de mesure impliquait un lien de cause à effet pour l'apprentissage.

On voit se multiplier d'autres types d'évaluations (et de SQC), ainsi qu'un intérêt accru des décideurs pour leur mise en œuvre. En 2011, l'Agence des États-Unis pour le développement international (USAID) a inscrit en tête de ses objectifs stratégiques pour les cinq prochaines années : « Améliorer les *compétences en lecture* de cent millions d'enfants des classes primaires d'ici à 2015 », avec une part substantielle de ses efforts consacrée à l'amélioration des techniques de mesure²⁵⁸. L'aide britannique au développement met également l'accent sur les évaluations des acquis scolaires²⁵⁹. Les études de terrain sur l'apprentissage des compétences de base se sont considérablement développées dans les pays en développement, avec près de 70 pays menant en 2010 des études sur la fluidité de la lecture²⁶⁰. D'autres formes d'évaluation précoce de la lecture sont également apparues au cours des dernières années, notamment l'apprentissage des mathématiques²⁶¹, le recours à des évaluateurs bénévoles²⁶², des collaborations transnationales²⁶³, ainsi que l'approfondissement du travail sur les interventions volontaristes pour améliorer les performances en lecture²⁶⁴.

En résumé, les méthodes SQC, sous de nombreuses formes différentes, sont de plus en plus utilisées. Elles peuvent être employées à un stade plus précoce du développement (et de la scolarité) de l'enfant, tout en offrant de nouvelles possibilités pour détecter les problèmes d'apprentissage de la lecture. Déterminer quel est le meilleur type de programme d'intervention, en s'inspirant au moins en partie des instruments de diagnostic, reste un problème majeur à résoudre dans le contexte des PMA.

258. USAID, 2011. Voir également Gove et Cvelich, 2010.

259. DIFD, 2011.

260. Abadzi, 2010, p. 18.

261. Rubens et Crouch, 2009.

262. En Inde, Banerji et Wadhwa, 2006 ; ASER, 2010.

263. Voir le travail d'Uwezo (2010) au Kenya et en Ouganda, qui s'appuie sur le travail de Pratham/ASER.

264. Dowd *et al.*, 2010 ; Dubeck *et al.*, 2010.

7. COÛT DES ÉVALUATIONS

[P]articiper activement [aux études de l'IEA] coût cher, trop cher pour de nombreux pays en développement. Et, lorsque certains ont pu financer leur participation, on peut se demander si le coût de cette participation est totalement justifié, par rapport à ce qui en a été tiré et aux autres utilisations qui auraient pu être faites des fonds investis. Quel intérêt cela présente-t-il pour la Thaïlande, l'Afrique du Sud et la Colombie de se retrouver en bas ou presque du classement international en science, alors que la Corée et le Japon caracolent en tête et que les pays européens se répartissent entre ces deux extrêmes ?²⁶⁵

Avec la multiplication des LSEA sur la scène mondiale et les coûts généralement élevés de participation et de gestion qu'elles entraînent, la question du poids fiscal des évaluations retient de plus en plus l'attention. Sachant que ces coûts sont généralement pris en charge par des agences extérieures (gouvernements et donateurs), le coût fiscal de ces investissements dans le savoir a été considéré comme minime comparé aux sommes considérables investies dans l'éducation proprement dite²⁶⁶.

L'opinion selon laquelle ce type d'évaluations (LSEA et autres) représente des décisions fiscales plutôt modestes est une position relative, du moins tant que les coûts réels ne sont pas pris en compte²⁶⁷. Quelques études montrent que les LSEA consomment une très petite part des budgets nationaux consacrés à l'éducation²⁶⁸. Cependant, ces études concernent souvent des pays relativement plus riches, dotés de budgets plus importants, et qui ne semblent pas rendre compte de la modestie des fonds discrétionnaires dont disposent généralement les ministres de l'Éducation pour ces activités dans les pays à faibles revenus.

265. Johnson, 1999, p. 70.

266. Lockheed et Hanushek, 1988 ; Porter et Gamoran, 2002 ; Wolff, 2008.

267. Postlethwaite (2004, p. 17), l'un des experts en LSEA les plus connus, fait remarquer : « [I]l convient également de signaler qu'il y a de grandes différences entre les divers projets, en termes de moyens financiers nécessaires pour les mener à bien. Ces sommes varient entre une dépense internationale annuelle de 200 000 dollars EU pour le SACMEQ, environ 3,6 millions de dollars EU pour l'enquête PISA et environ 7 millions de dollars EU pour l'IEA. Sans compter les sommes que les pays doivent payer pour leur propre personnel affecté aux projets et les frais de collecte de données. »

268. Voir l'analyse de Hoxby, 2002. Voir également Wolff (2008, p. 14), qui affirme que « les évaluations en Amérique latine, comme aux États-Unis, ne représentent pas une charge financière significative, soit 0,3 % ou moins du budget total du niveau d'études testé ».

Pour être réussie, une évaluation internationale exige des compétences de haut niveau en conception, planification et gestion, compétences trop rares dans le monde, en particulier dans les PMA²⁶⁹. Outre le coût et la complexité relativement importants d'une LSEA, se pose également le problème du choix parmi une foule de possibilités. Un ministre de l'Éducation doit non seulement décider de participer ou non aux LSEA, mais aussi choisir les tests à administrer en fonction des élèves, des langues et des systèmes éducatifs²⁷⁰.

Dans la matrice décisionnelle du rapport coût-bénéfice, il faut aussi prendre en compte les évaluations hybrides. Elles peuvent être adaptées aux objectifs stratégiques plus spécifiques d'un pays, voire d'un nombre limité d'écoles, avec des tailles d'échantillons et un délai de mise en œuvre réduits. Elles offrent donc la possibilité d'ajuster une étude aux paramètres à la fois stratégiques et budgétaires, en reléguant au second plan les objectifs comparatifs qui guident souvent la conception des LSEA au profit d'autres paramètres.

Analyse coût-bénéfice des évaluations dans le domaine de l'éducation

Globalement, on peut affirmer que, pour les systèmes éducatifs de création récente, des évaluations moins fréquentes devraient suffire, après une évaluation de départ, dans la mesure où la plupart des problèmes à régler sont connus et où une amélioration substantielle prendra plusieurs années. Il vaudra mieux consacrer les modestes ressources disponibles à des évaluations visant à améliorer l'apprentissage et l'enseignement, où le retour sur investissement a toutes les chances d'être plus important²⁷¹.

Au début des années 1990, quelques études se sont intéressées aux coûts et, dans une moindre mesure, aux avantages des LSEA²⁷². Les résultats ont confirmé la valeur des LSEA pour deux raisons principales : les coûts déclarés (explicitement budgétés et comptabilisés) relativement bas par rapport au budget global de l'éducation²⁷³ et les avantages potentiels considérables des

269. Lockheed, 2008, p. 10.

270. Comme le font remarquer Braun et Kanjee (2006, p. 24), « dans les systèmes éducatifs qui n'ont même pas les ressources élémentaires, la décision de financer une évaluation nationale est extrêmement difficile à prendre ».

271. Braun et Kanjee, 2006, p. 8.

272. Ilon, 1992, 1996 ; Koeffler, 1991 ; Loxley, 1992 ; pour un point de vue historique, voir Lockheed, 2008, p. 3. Costrell et Peyser, 2004 ; Hoxby, 2002.

273. Costrell et Peyser, 2004 ; Hoxby, 2002.

LSEA pour obtenir des résultats sur la base desquels on pourra agir²⁷⁴. Ces premières études ont également mis en évidence la complexité financière et contextuelle du calcul du coût des évaluations. En effet, les ministères de l'Éducation des PMA doivent prendre en compte l'expertise technique de l'organisme qui réalise les tests, ainsi que l'expertise disponible au niveau national. Les initiatives pour élaborer des LSEA adaptées aux besoins nationaux peuvent utiliser l'expertise de l'organisme chargé de la mise en œuvre, mais aussi épuiser (ou dépasser) l'expertise nationale dont dispose le ministère²⁷⁵. Ainsi, si des analyses de coût ne sont pas effectuées en amont, on s'expose fortement à du gaspillage (ou à un échec) ultérieurement²⁷⁶.

Bien que des études plus récentes semblent corroborer la thèse du faible coût fiscal des évaluations, le coût des LSEA est de plus en plus considéré comme étant un sérieux obstacle pour les PMA²⁷⁷. L'expression « faible coût » doit être pensée par rapport aux ressources disponibles²⁷⁸. Les études montrent que le coût moyen des LSEA est plutôt modeste (généralement moins de 1 % du budget national de l'éducation, voire seulement 0,3 % dans certains cas), mais ces chiffres peuvent ne pas refléter le pourcentage de fonds discrétionnaires disponibles dans le budget d'un ministère²⁷⁹.

Calculer les coûts

Pour choisir une évaluation en fonction de son coût, il faut prendre en compte à la fois les coûts explicites et les coûts cachés que comporte toute évaluation²⁸⁰. Cet aspect est développé ci-après.

Coûts explicites

Les coûts explicites sont ceux qui sont habituellement prévus à l'avance et inclus dans les mécanismes comptables de l'organisme qui pilote la LSEA : frais de personnel pilotant l'évaluation (notamment conception et mise en œuvre des tests), de formation, de déplacements, de fournitures et

274. Braun et Kanjee, 2006, p. 12 ; Hanushek et Woesmann, 2005.

275. Greaney et Kelleghan, 2008, p. 49 ; Ilon, 1996 ; Wolff, 2007.

276. Wolff, 2008, p. 5.

277. Siniscalco, 2006 ; Ravela *et al.*, 2008 ; Wolff, 1998.

278. Wolff, 2008.

279. Coombs et Hallak, 1987, p. 50 ; Ilon, 1996, p. 86.

280. Greaney et Kellaghan, 2008, p. 49-50.

d'équipements²⁸¹. Ces coûts peuvent varier en fonction de la localisation : coûts *domestiques* (par exemple, déploiement et gestion du processus d'évaluation dans le pays) ; coûts en nature (par exemple, contributions non pécuniaires, telles que la mise à disposition de personnels du ministère, de spécialistes, de chefs d'établissement et d'enseignants) ; coûts *internationaux* (par exemple, frais généraux de l'agence internationale, experts internationaux et déplacements).

Coûts cachés

Si les coûts explicites sont évidents dans la conception d'un projet, d'autres coûts peuvent échapper à l'attention des autorités qui élaborent des plans financiers pour les évaluations. Ce sont notamment :

- Les coûts indirects (ou frais généraux). Ce sont les agences qui prennent en charge ces coûts en pilotant le programme. S'ils sont souvent pris en compte dans les pays riches, ils échappent parfois à l'attention des ministères et autres agences des PMA. Les exemples les plus flagrants sont notamment le coût d'utilisation des infrastructures (bâtiments, réseaux, entretien des ordinateurs, etc.). Des coûts moins évidents, mais non négligeables, peuvent être associés au personnel détaché des ministères et aux personnes qui interviennent sur le terrain, comme les inspecteurs scolaires ou les enseignants²⁸².
- Les coûts d'opportunité. Ces coûts concernent les autres stratégies qui auraient pu être adoptées à la place de celle choisie. Par exemple, en ne procédant pas à une évaluation une année donnée, le ministère pourrait avoir plus de ressources pour l'effectuer une autre année. Ou, le choix d'un type d'évaluation peut exclure la possibilité d'un choix

281. Lockheed (2008, p. 9) soutient que : « Les évaluations nationales des acquis scolaires dans les pays en développement ou en transition emploient rarement des instruments de mesure complexes, parce que ces pays ont rarement la capacité nationale requise ou les moyens d'acheter cette expertise à l'étranger ». Topol *et al.* (2010), pour leur part, ont étudié récemment les efforts de mesure du coût des évaluations complexes aux États-Unis ; ils avancent, entre autres choses, qu'une meilleure technologie permet de réduire les coûts des efforts supplémentaires en R&D. Mais, dans la mesure où les PMA sont pour le moment handicapés par des contraintes technologiques, les coûts plus élevés de R&D devraient finalement devenir une nouvelle ligne de dépense.

282. Ilon, 1992.

différent²⁸³. Toutefois, le coût de la *non*-participation à une évaluation, c'est-à-dire le fait de renoncer aux avantages potentiels (en termes de développement du personnel, de résultats potentiels, etc.) de cette participation, doit également être considéré comme un autre type de coût d'opportunité.

Catégories de coûts et comparaisons des coûts de quelques évaluations

Le *tableau 7.1* résume les catégories de coûts évoquées ci-dessus. À des fins de comparaison, plusieurs organismes d'évaluation réputés ont été contactés pour leur demander des chiffres récents concernant leurs dépenses (certaines sont des estimations). Les études concernées figurent dans le *tableau 7.2*. Les données nationales recueillies sur chacune des études retenues sont reportées dans le *tableau 7.3* ; celui-ci montre la variabilité des coûts connus, par évaluation et en fonction du contexte national, pour 13 évaluations récentes. Le *tableau 7.4* propose une synthèse du pourcentage moyen des dépenses totales que représentent les six catégories principales de coûts²⁸⁴.

283. Par exemple, l'Afrique du Sud a fait ce type de choix en décidant de ne pas participer à l'enquête TIMSS, en invoquant le coût global en termes de temps et de ressources (Greaney et Kelleghan, 2008, p. 75). Toujours au sujet de l'Afrique du Sud, voir Braun et Kanjee, 2006, p. 19.

284. Ces données ont été recueillies pour les besoins du présent ouvrage, et nous tenons à remercier les divers organismes et leurs représentants d'avoir bien voulu nous les fournir, certaines étant des estimations, comme il est précisé dans le *tableau 7.3*. Les pourcentages sont arrondis à l'entier le plus proche.

Tableau 7.1 Catégories de coûts des évaluations employées dans une sélection d'études

1. Préparation de l'évaluation <ul style="list-style-type: none">a. Création des items de test et contrôle rédactionnelb. Essai pilotec. Formation
2. Administration de l'évaluation <ul style="list-style-type: none">a. Conception et contrôle rédactionnel du testb. Impression des testsc. Impression d'autres documentsd. Distribution aux examinateurse. Évaluation sur le terrainf. Contrôle et supervision
3. Traitement et analyse <ul style="list-style-type: none">a. Codage et saisie informatiqueb. Notation des questions à réponse librec. Analyses complémentaires
4. Diffusion <ul style="list-style-type: none">a. Compte rendu à chaque écoleb. Production et distribution du rapportc. Relations publiques
5. Coûts institutionnels <ul style="list-style-type: none">a. Personnel – budget du projetb. Personnel – mis à disposition (par exemple consultants)c. Infrastructures – budget du projet (locaux pour le personnel)d. Infrastructures – mises à dispositione. Équipements – budget du projet (par exemple ordinateurs et équipements d'évaluation associés)f. Équipements – mis à dispositiong. Autres (par exemple télécommunications, électricité et fournitures de bureau)h. Frais d'inscription
6. Ventilation des coûts <ul style="list-style-type: none">a. Coût par élève évaluéb. Coût d'éducation d'un élève (au niveau de scolarité testé)c. Coût de l'évaluation en pourcentage du budget total pour un niveau de scolaritéd. Coût de l'évaluation en pourcentage du budget total de l'enseignement secondaire

Tableau 7.2 Études des coûts d'une sélection d'évaluations nationales, régionales et transnationales

<ul style="list-style-type: none"> • Évaluations nationales : <ul style="list-style-type: none"> ▫ SIMCE/LLECE 2004 ▫ Évaluation nationale 2002 en Uruguay ▫ Évaluation nationale 2002 au Honduras 	<ul style="list-style-type: none"> • Évaluations régionales : <ul style="list-style-type: none"> ▫ SACMEQ II <ul style="list-style-type: none"> – Swaziland 2006 – République-Unie de Tanzanie 2006 – Zambie 2006 ▫ PASEC 2010
<ul style="list-style-type: none"> • Évaluations internationales : <ul style="list-style-type: none"> ▫ PISA <ul style="list-style-type: none"> – PISA Chili 2009 – PISA Mexique 2009 – PISA Panama 2009 – PISA Pérou 2000 – PISA Pérou 2009 – PISA Uruguay 2003 ▫ PIRLS 	<ul style="list-style-type: none"> • Évaluations hybrides : <ul style="list-style-type: none"> ▫ EGRA <ul style="list-style-type: none"> – Libéria 2008 – Nicaragua 2008

Tableau 7.3 Coûts des évaluations nationales, régionales, internationales et EGRA*

Coût monétaire des évaluations (dollars EU)	Évaluations nationales			Évaluations régionales		
	SIMCE 2004 ^a	Honduras 2004 ^b	Uruguay 2003 ^c	PASEC 2010 ^d	SACMEQ III Swaziland 2007 ^e	SACMEQ III Tanzanie 2007 ^f
Préparation de l'évaluation	258 236	174 275	21 528	34 164	12 561	12 666
Création et contrôle rédactionnel des items de test	184 515			7 895		1 000
Essai pilote	73 721			15 749	12 561	11 666
Formation				10 520		
Administration de l'évaluation	1 163 764	435 717	57 289	91 705	170 732	89 900
Conception et contrôle rédactionnel du test	29 403			7 415		2 000
Impression des tests	324 712			9 744	15 488	12 000
Impression d'autres documents	236 076				3 049	4 200
Distribution aux examinateurs	103 124			6 455	73 171	2 000
Évaluation sur le terrain	406 103			68 091	79 024	56 700
Contrôle et supervision	64 346					13 000

Des évaluations simples, rapides et abordables
Améliorer l'apprentissage dans les pays en développement

Coût monétaire des évaluations (dollars EU)	Évaluations nationales			Évaluations régionales		
	SIMCE 2004 ^a	Honduras 2004 ^b	Uruguay 2003 ^c	PASEC 2010 ^d	SACMEQ III Swaziland 2007 ^e	SACMEQ III Tanzanie 2007 ^f
Traitement et analyse	382 239	130 721	26 272	12 624	454	33 300
Codage et saisie informatique	216 048			12 624		33 300
Notation des questions à réponse libre	166 191				454	
Analyses complémentaires						
Diffusion	100 567	130 721	531	32 193	4 195	2 000
Compte rendu à chaque école	100 567				4 195	2 000
Production et distribution du rapport						
Relations publiques						
Sous-total	1 904 806	871 434	105 620	170 686	187 942	137 866
Coûts institutionnels	938 766			12 481	24 878	25 500
Personnel – budget du projet	796 864			2 737	17 561	10 000
Personnel – mis à disposition						
Infrastructures – budget du projet	35 369					5 000
Infrastructures – mises à disposition						
Équipements – budget du projet	106 533			9 744	7 317	10 500
Équipements – mis à disposition						
Frais d'inscription						
Autres	20 028			2 043		
TOTAL	2 863 600	871 434	105 620	185 210	212 820	163 366
Nombre total d'élèves	300 000	45 657	12 993	5 400	4 155	3 000 ^m
Nombre total d'écoles						
Ventilation des coûts						
Coût par élève évalué	10	19	8	34	51	55
Coût d'éducation d'un élève	767	130	484		66	
Coût de l'évaluation en % du budget total pour un niveau de scolarité	0,83	2,63				
Coût de l'évaluation en % du budget total de l'enseignement secondaire	0,17	0,33	0,07			

Tableau 7.3 Coûts des évaluations nationales, régionales, internationales et EGRA* (suite)

Coût monétaire des évaluations (dollars EU)	Évaluations internationales					Évaluations EGRA	
	PISA	PISA	PISA	PISA	PISA	EGRA	EGRA
	Chili 2009 ^e	Mexique 2009 ^h	Panama 2009 ⁱ	Pérou 2009 ^j	Uruguay 2003 ^k	Libéria 2008	Nicaragua 2008
Préparation de l'évaluation	26 448	100 301	61 475	47 956	12 357	29 345	10 882
Création et contrôle rédactionnel des items de test	26 448	3 802	13 661				
Essai pilote		96 499	47 814			16 031	4 756
Formation						13 314	6 126
Administration de l'évaluation	597 958	891 501	187 157	212 486	29 707	82 260	68 683
Conception et contrôle rédactionnel du test	8 976		13 661	2 590		8 800	
Impression des tests		254 899	54 644	7 196		5 600	1 395
Impression d'autres documents		116 156	6 831				
Distribution aux examinateurs		123 845	6 831				
Évaluation sur le terrain	462 705	394 235	98 359	198 261		67 860	67 288
Contrôle et supervision	126 277	2 366	6 831	4 439			
Traitement et analyse		167 782	128 414		22 838	13 533	5 734
Codage et saisie informatique		56 899	114 753			13 533	5 734
Notation des questions à réponse libre		110 883	13 661				
Analyses complémentaires							
Diffusion	49 912		34 153	3 865	14 092	1 850	
Compte rendu à chaque école			34 153	3 865		1 500	
Production et distribution du rapport	49 912					350	
Relations publiques							
Sous-total	674 318	1 159 584	411 199	264 307	78 994	126 988	85 299
Coûts institutionnels	179 233	490 203	94 261	20 473		103 520	87 157
Personnel – budget du projet	179 233	321 246	73 769	9 324		101 858	83 675
Personnel – mis à disposition		107 286		11 149		1 403	2 500
Infrastructures – budget du projet		2 743	6 831				
Infrastructures – mises à disposition							
Équipements – budget du projet		58 928	13 661			259	982

Des évaluations simples, rapides et abordables
Améliorer l'apprentissage dans les pays en développement

Coût monétaire des évaluations (dollars EU)	Évaluations internationales				Évaluations EGRA		
	PISA Chili 2009 ^e	PISA Mexique 2009 ^h	PISA Panama 2009 ⁱ	PISA Pérou 2009 ^j	PISA Uruguay 2003 ^k	EGRA Libéria 2008	EGRA Nicaragua 2008
Équipements – mis à disposition							
Frais d'inscription	49 863	118 599			43 197		
Autres	72 494		13 661	2 000		10 619	6 958
TOTAL	975 908	1 768 386	519 121	286 780	122 191	241 127	179 414
Nombre total d'élèves	5 700	45 079	42 000	7 967	5 797	3 770	5 760
Nombre total d'écoles						240	120
Ventilation des coûts							
Coût par élève évalué	171	39	12	36	21	64	31
Coût d'éducation d'un élève		9 439	1 023	396	479		
Coût de l'évaluation en % du budget total pour un niveau de scolarité			1,20838				
Coût de l'évaluation en % du budget total de l'enseignement secondaire		0,001767	0,04419		0,08		

* Les sources du *tableau 7.3* sont les suivantes :

- Source* : Wolff, 2007, p. 6 (pour le test SIMCE 2004). Les chiffres originaux pour toutes les évaluations nationales ci-dessus (à savoir SIMCE 2004, Honduras 2004 et Uruguay 2003) et PISA Uruguay 2003 ont été publiés dans Wolff, 2007 en devises locales.
- Source* : Wolff, 2007, p. 13 ; en 2004, 1 dollar EU = 17,68 lempiras honduriens.
- Source* : Wolff, 2007, p. 11 ; en 2003, 1 dollar EU = 28,24279 pesos uruguayens.
- Source* : Rapport technique PASEC 2010 (communication personnelle, P. Varly, mai 2009). Euros convertis en dollars EU, taux de change annuel 2009.
- Source* : Communication personnelle, A. Mrutu, août 2009.
- Source* : Communication personnelle, J. Shabalala, août 2009.
- Source* : Communication personnelle, E. Lagos, septembre et octobre 2009.
- Source* : Communication personnelle, M. A. Diaz, septembre 2009.
- Source* : Communication personnelle, Z. Castillo, septembre 2009.
- Source* : Communication personnelle, L. Molina, septembre 2009.
- Source* : Wolff, 2007, p. 14 ; en 2003, 1 dollar EU = 28,24279 pesos uruguayens.
- Source* : Communication personnelle, A. Gove, août 2009.
- Estimation fondée sur un échantillon de SACMEQ II de 2 854 élèves.
- Estimation fondée sur un courriel de E. Lagos, octobre 2009.

Tableau 7.4 Coûts par catégorie, en pourcentage des dépenses totales liées à l'évaluation

Catégorie de coût	Moyenne	Le plus bas	Le plus élevé
Préparation de l'évaluation	11 %	3 % (PISA Chili, 2009)	20 % (Évaluation nationale Uruguay, 2003)
Administration de l'évaluation	50 %	24 % (PISA Uruguay, 2003)	80 % (SACMEQ III, Swaziland)
Traitement et analyse	13 %	1 % (SACMEQ III, Swaziland)	25 % (Évaluation nationale Uruguay, 2003)
Diffusion	6 %	1 % (Évaluation nationale Uruguay, 2003)	17 % (PASEC, 2010)
Coûts institutionnels	23 %	7 % (PASEC, 2010)	49 % (Évaluation nationale Uruguay, 2003)
Droits d'inscription	16 %	5 % (PISA Chili, 2009)	35 % (PISA Uruguay, 2003)
Autres	3 %	1 % (PISA Pérou, 2009)	7 % (PISA Chili, 2009)

Note : Les calculs ci-dessus sont basés sur des données provenant de 13 évaluations (voir le *tableau 7.3* pour les coûts inclus dans chaque catégorie et pour chaque évaluation).

Le *tableau 7.3* appelle un certain nombre d'observations. Premièrement, les populations d'élèves vont de 3 770 (chiffre plutôt modeste) pour l'ÉGRA-Libéria à environ 300 000 pour le SIMCE (Chili)²⁸⁵. Deuxièmement, le total des coûts explicites (listés) d'administration des évaluations est compris entre un minimum d'environ 122 000 dollars pour le PISA (en Uruguay) et un maximum de 2,8 millions de dollars pour le SIMCE (Chili). Troisièmement, on peut calculer le « coût par élève » en prenant ces deux premiers paramètres, moyen utile d'examiner les coûts, quelle que soit l'envergure de l'initiative. Les résultats montrent que ce paramètre va d'environ 8 dollars pour l'évaluation nationale en Uruguay à environ 51 dollars pour l'étude SACMEQ III au Swaziland, et environ 171 dollars pour le PISA au Chili. La moyenne, pour cet échantillon d'études, est d'environ 42 dollars par élève évalué. De plus (*tableau 7.2*), certains coûts sont plus importants que d'autres, comme les coûts d'application des tests (50 %) et les coûts institutionnels (23 %), alors que les coûts du

285. Les données sur la taille des échantillons des évaluations internationales compilées pour les différents pays peuvent donner des chiffres de population totale beaucoup plus importants, et le nombre de pays participants continue d'augmenter. Par exemple, plus de 400 000 élèves de 57 pays ont participé à l'enquête PISA 2006. Pour des informations sur le SIMCE, voir Meckes et Carrasco, 2010.

traitement et de l'analyse (13 %) et de la préparation des tests (11 %) sont nettement plus faibles²⁸⁶.

Les données montrent que, sur le terrain, les niveaux moyens de coût par apprenant ne sont pas extrêmement différents d'une évaluation à l'autre. Certaines évaluations coûtent clairement plus cher, mais les études nationales et internationales de plus grande envergure permettent des économies d'échelle qui réduisent les coûts d'évaluation unitaires. Actuellement, les études EGRA de plus petite envergure ne sont pas moins onéreuses sur le terrain. De plus, certains pays peuvent avoir beaucoup plus de ressources (financières, intellectuelles et infrastructurelles) dans leurs services d'évaluation, ce qui a probablement une incidence sur plusieurs variables de coût, comme les honoraires de consultants internes ou externes et les frais de voyage. Par ailleurs, les évaluations hybrides en sont encore au stade de la recherche (avec les coûts de tâtonnement que cela suppose) ; on peut donc s'attendre à ce que leurs coûts en aval diminuent considérablement à mesure qu'elles se développeront. Qui plus est, les exigences et les besoins nationaux spécifiques (par exemple la logistique dans un environnement difficile) peuvent aussi jouer un rôle majeur dans le choix du type d'évaluation et avoir une incidence sur le montant qu'elle coûtera au final.

L'exactitude des estimations et la prise en compte ou pas de la totalité des coûts cachés ont une incidence majeure sur ces données. Toutes les équipes ne recueillent pas et ne stockent pas de données sur les coûts. Et quand elles le font, les données sont parfois incomplètes ou insuffisamment détaillées pour permettre des analyses comparatives. Les inexacitudes et les divergences sont souvent le résultat de l'insuffisance des fonds²⁸⁷. C'est pourquoi ces données doivent être considérées comme étant un aperçu

286. Il convient de noter que les données n'étaient pas toutes complètes pour chaque catégorie, ni représentatives de la totalité des coûts réels. Par exemple, les seules données disponibles pour le PASEC étaient les coûts prévisionnels des évaluations de 2010 ; seules trois sources ont fourni des données sur le *prix des tests* ; plusieurs sources n'ont pas communiqué de données pour les catégories *traitement et analyse ou diffusion*. En outre, à propos des fourchettes ci-dessus, certaines catégories ont affiché plus de variabilité que d'autres. Par exemple, la catégorie *traitement et analyse* inclut des dépenses moyennes de 0,02 % à 24,8 %, alors qu'à l'exception de trois évaluations (évaluation nationale 2004 au Honduras, PASEC 2010 et PISA Uruguay 2003), les dépenses de diffusion affichaient une moyenne de 5,9 %. De plus, l'analyse devrait également tenir compte des coûts cachés, voire des coûts non spécifiés déjà évoqués ; par exemple les coûts de la catégorie *autres* pour l'enquête PISA Chili 2009 étaient supérieurs à 7 %.

287. Lockheed, 2008, p. 16.

préliminaire d'une comparaison des coûts. Des efforts restent à faire en ce qui concerne la mise en place d'audits complets et fiables.

Comment envisager les coûts ?

Dans les pays en développement, les responsables de l'éducation se retrouveront avec plus de choix que de ressources disponibles. Il n'existe pas de vision complète des coûts et des bénéfices, parce que l'on n'a pas recueilli suffisamment de données fiables sur les coûts d'évaluation par rapport à la qualité de l'information obtenue. De plus, le dynamisme scientifique, technologique et politique actuel dans le domaine de l'amélioration de l'éducation laisse supposer que les modèles d'évaluation évolueront très certainement à la faveur des progrès des tests et de l'accroissement de la demande. La nécessité d'indicateurs exploitables et d'un choix clair de tests devrait se faire sentir de plus en plus.

Les innovations récentes dans le domaine des évaluations laissent augurer une dynamique en faveur de nouveaux modèles privilégiant une approche « bien dosée », centrée sur les besoins²⁸⁸. Cela signifie que les innovations (les outils EGRA en sont un bon exemple) peuvent contribuer à élargir le champ d'application des tests, à diminuer les coûts initiaux explicites, tels que la traduction et la préparation des tests, et à réduire le temps de traitement, de sorte que les ministres disposent plus rapidement de données exploitables, avec moins de personnel et moins de frais généraux²⁸⁹. Les trois paramètres clés ci-après résument les questions de coût des évaluations qui doivent être prises en compte, en particulier dans les PMA aux ressources limitées.

Échelle

Les ministères de l'Éducation des pays en développement peuvent avoir besoin de déterminer quelle évaluation leur permettra d'obtenir les données dont ils ont besoin sur une population spécifique (par exemple les filles vivant en zone rurale ou certains groupes ethnolinguistiques), un groupe d'écoles ou une matière particulière dans une classe donnée. Les LSEA n'ont généralement pas la flexibilité requise pour répondre à ce type de demande, à cause du travail considérable de préparation en amont et de

288. Wagner, 2003.

289. D'après Braun et Kanjee, si les ressources sont limitées, les ministères de l'Éducation ont plutôt intérêt à envisager une participation partielle à une évaluation régionale ou internationale (2006, p. 36). Cette étape intermédiaire peut aider à mieux discerner les bénéfices réels que peut procurer une LSEA.

pré-évaluation, qui rend difficiles les changements à brève échéance et garantit les paramètres de comparabilité. De plus, la plupart des LSEA ne sont pas destinées à produire des indicateurs au niveau des classes mais plutôt des indicateurs systémiques²⁹⁰.

En revanche, les enquêtes menées auprès d'échantillons limités de ménages ou les évaluations hybrides du type EGRA permettent d'économiser de l'argent, avec moins de personnes à évaluer pour obtenir des réponses à un ensemble plus spécifique de questions stratégiques. Ces enquêtes peuvent aussi être mises en œuvre et ajustées plus fréquemment. Cependant, des innovations introduites récemment par les LSEA en matière d'échantillonnage (comme le PIRLS) permettent de penser que ces études génèrent des données à plusieurs niveaux²⁹¹ et que prendre un échantillon plus grand peut valoir les coûts marginaux induits, du fait des économies d'échelle réalisées²⁹². En d'autres termes, le coût inférieur par apprenant est une valeur relative.

Agir au bon moment

Deux actions doivent absolument être menées au bon moment pour profiter des avantages potentiels offerts par les évaluations : l'organisation en temps opportun du cycle d'évaluation, depuis la planification jusqu'à la diffusion des résultats, en passant par le lancement, la collecte et l'analyse des données (sans oublier les débats politiques qui s'ensuivent). La question de l'opportunité peut aussi concerner la disponibilité et l'utilisation des informations. Par exemple, avant de prendre d'importantes décisions financières concernant un groupe d'écoles, la disponibilité en temps voulu d'informations à jour les concernant peut être un facteur essentiel. Ou encore, une population d'élèves peut avoir besoin d'aide pour acquérir le niveau voulu de compétence en lecture, et les données recueillies peuvent confirmer, infirmer ou guider les processus décisionnels. De plus, comme il a déjà été mentionné, il faut tenir compte des mérites d'une intervention précoce dans la trajectoire d'apprentissage des élèves, à l'instar des arguments invoqués dans le domaine médical pour justifier les systèmes de

290. Voir Volante, 2006, p. 7.

291. Porter et Gamoran, 2002 ; RTI, 2009, p. 76.

292. Wolff (2008, p. 19) indique : « ... les échantillons de grande taille peuvent être appliqués aux évaluations censitaires à un faible coût marginal, du fait que les coûts fixes de développement des items et des essais pilotes peuvent être amortis avec une population plus large ».

dépistage précoce²⁹³. En résumé, il faut collecter le plus rapidement possible des données d'évaluation crédibles pour pouvoir élaborer efficacement des stratégies ; mais il faut aussi que ces données soient disponibles en temps voulu. S'efforcer d'agir au bon moment peut permettre de réduire les coûts globaux d'évaluation et d'intervention.

Rentabilité

Comme il a été dit précédemment, certaines évaluations sont relativement onéreuses en termes de mises de fonds initiales, car elles nécessitent de faire appel à des spécialistes et à des consultants qui coûtent cher, ainsi qu'à des enquêteurs de terrain qualifiés. Ces coûts, ainsi que d'autres, peuvent être envisagés sous l'angle du coût total ou du coût par apprenant. Dans un cas comme dans l'autre, les limites budgétaires qui pèsent sur les fonds discrétionnaires dans les PMA obligent à examiner de très près les options avant de choisir une évaluation. En raison de la rareté actuelle de données crédibles sur les coûts dans les PMA, il est difficile de décider d'une marche à suivre adaptée à des contextes divers. Il faut absolument déterminer de façon plus précise les dépenses susceptibles de produire des types particuliers de résultats, pertinents du point de vue stratégique. Par exemple : dans quelle mesure une meilleure formation des enquêteurs produira-t-elle une plus grande fiabilité interévaluateurs ? Ou, comme on l'a vu dans une initiative récente en Inde, des bénévoles peuvent-ils devenir des enquêteurs fiables, durables et peu onéreux avec relativement peu de

293. La médecine préventive insiste sur la nécessité d'une bonne information en temps voulu. L'actualité de l'information peut avoir des conséquences sur la vie ou la mort, la propagation d'une épidémie ou son enragement. Les mesures proactives coûtent moins cher et aident à éviter le pire. La médecine préventive est encouragée non seulement pour prévenir la maladie, mais aussi pour réduire les coûts de diagnostic et de traitement de la maladie (Szucs, 1997). Des arguments similaires peuvent être employés dans le domaine de l'éducation. Par exemple, l'absentéisme et le décrochage scolaire sont des problèmes courants dans les PMA, et ont un coût financier et social considérable. Deux mesures se sont avérées particulièrement efficaces pour résoudre ces problèmes : la diminution des redoublements (Ndaruhutse, 2008) et le développement de l'enseignement bilingue (Grin, 2005 ; Heugh, 2006). Si des tests pouvaient aider à détecter et « diagnostiquer » plus tôt les difficultés scolaires (depuis les problèmes cognitifs jusqu'aux difficultés sociocomportementales), ils pourraient contribuer à prévenir des problèmes coûteux tels que les taux de décrochage scolaire. En d'autres termes, même si les outils de diagnostic de type SQC ne permettent pas de déterminer facilement le « meilleur » plan d'action (qui peut être varié et complexe) pour remédier à un problème, sa détection précoce est presque systématiquement plus rentable à long terme.

formation²⁹⁴ ? Il faut poursuivre les recherches pour clarifier les avantages des différentes évaluations en termes de coût.

Additionner les coûts

Les coûts font partie intégrante de toute intervention sociale ; l'évaluation de l'apprentissage (ainsi que ses conséquences stratégiques) en est un exemple flagrant. La question est de savoir ce qu'un ministère (ou un donateur) obtient en contrepartie de ses investissements. Il n'est pas facile de réunir des données sur les coûts permettant de *comparer* des évaluations. Toutefois, il existe maintenant quelques points de repère qui peuvent être pris en compte. Le plus important est peut-être le compromis entre le temps et l'argent. Prenons, par exemple, le cas d'un ministre de l'Éducation qui dispose de cinq années pour choisir une politique et l'appliquer. Des LSEA régionales ou internationales, comme le SACMEQ ou le PASEC, peuvent alors apporter quelques réponses valables sur des points clés et offrir une possibilité de comparaison entre pays. Compte tenu des économies d'échelle actuelles dans les pays qui réitèrent des évaluations internationales, le coût réel par apprenant de ces LSEA n'est pas très différent de celui de l'EGRA et des évaluations hybrides fondées sur des échantillons beaucoup plus petits de populations.

D'un autre côté, si un ministre n'a pas trois à cinq ans devant lui et si son but est avant tout d'aider les programmes, les écoles et les districts régionaux à améliorer leurs résultats scolaires à court terme, alors une évaluation axée sur un petit échantillon comme l'EGRA apparaît beaucoup moins onéreuse²⁹⁵. Si le coût par apprenant de l'EGRA semble actuellement similaire à celui des grandes évaluations internationales, les coûts futurs devraient diminuer, à mesure que les pays se familiariseront avec les outils EGRA et que les enquêteurs seront mieux formés. Le délai plus court pour aboutir à l'analyse des données et à la diffusion des résultats devrait, selon toute vraisemblance, réduire les coûts récurrents en ressources humaines.

Enfin, il faut prendre en compte les coûts d'opportunité. Pour évaluer les enfants, les LSEA attendent qu'ils soient en 4^e année (au plus tôt), à

294. Banerji, 2006.

295. Voir également Chabott (2006, p. 24) qui soutient : « De bons programmes d'apprentissage de la lecture coûteront plus cher par élève que les manuels actuels de lecture et l'enseignement des professeurs. Ils prendront probablement aussi plus de temps que ce que leur allouent actuellement les programmes scolaires. Mais de bons programmes d'apprentissage de la lecture peuvent être plus rentables que des programmes médiocres. »

un stade où ils peuvent être très en retard dans leur progression en lecture. Cela peut entraîner des coûts élevés d'intervention qu'une évaluation plus précoce aurait évités. Rattraper le retard est difficile et coûteux, et risque de conduire à l'échec scolaire, qui constitue le coût le plus important que les responsables politiques cherchent à éviter.

En résumé, évaluer l'apprentissage est fondamental pour faire évoluer l'éducation dans n'importe quel pays. Mais ces évaluations ont des coûts qui doivent être évalués et comparés. Le temps est loin où les services ministériels pouvaient prêter leur personnel à d'autres administrations, et où les donateurs extérieurs prenaient totalement à leur charge les frais d'évaluations à grande échelle. L'époque est dorénavant celle des contraintes fiscales, celle aussi où il faut trouver un compromis entre la compréhension de l'apprentissage et ce qui est appris, dans quel but *et à quel coût*. Il est probable que le calcul du coût des évaluations deviendra, dans les années à venir, une question à laquelle il faudra s'intéresser davantage.

8. ÉVALUATION DU NIVEAU D'ALPHABÉTISATION DES ADULTES

Là où les systèmes d'éducation formelle s'accompagnent de programmes d'apprentissage précoce et de programmes d'alphabétisation et de développement des compétences, il en résulte des bénéfices supplémentaires pour l'individu, la communauté, la société et l'éducation formelle elle-même. Les enfants qui ont bénéficié de possibilités d'apprentissage précoce apprennent mieux dans l'enseignement formel, et les adultes éduqués font de plus gros efforts en tant que parents pour scolariser leurs enfants et les aider quand ils sont à l'école²⁹⁶.

Depuis quelques années, on voit se multiplier les études universitaires sur l'alphabétisation des adultes²⁹⁷. La recherche historique indique que, traditionnellement, l'apprentissage de la lecture et de l'écriture se déroulait souvent hors du système scolaire et, en ce sens, constituait davantage une transmission sociale qu'un processus éducatif au sens strict. Cela ouvre une perspective importante : à savoir que l'alphabétisation, phénomène *culturel*, est pratiquée dans des environnements et des contextes variés et complexes. Si la plupart des enfants apprennent actuellement à lire en classe, les niveaux d'acquisition de compétences des enfants et des adultes peuvent être influencés de manière significative par des déterminants *extérieurs à l'école*.

Les chapitres précédents étaient essentiellement consacrés à l'évaluation des compétences en lecture des enfants scolarisés. En revanche, le présent chapitre sur l'évaluation du niveau d'alphabétisation des adultes s'intéresse davantage à l'utilisation de la lecture et de l'écriture hors du contexte scolaire et à la façon dont cette utilisation est mesurée.

Importance de l'alphabétisation des adultes aujourd'hui

De tous les objectifs de l'éducation, l'alphabétisation demeure l'un des plus négligés. Les progrès en direction de l'objectif consistant à diminuer de moitié l'analphabétisme d'ici à 2015 [objectif 4 de l'EPT] ont été beaucoup trop lents et inégaux²⁹⁸.

296. UNESCO. 2004. *Rapport mondial de suivi sur l'EPT, 2005*, p 67.

297. Voir Wagner *et al.* 1999, pour une présentation générale.

298. UNESCO. 2010. *Rapport mondial de suivi sur l'EPT 2010*, p. 103.

Lors de sa création en 1946, l'UNESCO a inscrit l'alphabétisation au premier rang de ses priorités en matière d'éducation et de droits humains. Des progrès considérables ont été enregistrés au cours du dernier demi-siècle. Les niveaux de faible alphabétisme et d'analphabétisme sont considérés comme étant un problème majeur du XXI^e siècle partout dans le monde (*tableau 5.1*), mais surtout dans les pays en développement (*figure 8.1*)²⁹⁹. Au fil du temps, de nombreuses raisons ont été invoquées pour justifier les investissements dans l'alphabétisation des adultes : économie (un plus haut niveau de compétence favorise la croissance économique) ; développement social (autonomisation des femmes) ; science politique (développement de la démocratie ; identité nationale) ; éducation (les parents alphabétisés encouragent l'alphabétisation des enfants)³⁰⁰.

Pour les besoins de la présente étude, l'alphabétisation des adultes est considérée à la fois comme un objectif de l'EPT (objectif 4) de plein droit, et comme une solide preuve de l'importance du rôle de l'alphabétisation des parents dans celle des enfants : ils poussent ces derniers à apprendre à lire et à écrire. De plus, la science de l'alphabétisation peut offrir des perspectives qui se renforcent mutuellement. Par exemple, l'acquisition de la lecture et de l'écriture chez les adultes a beaucoup de points communs avec l'assimilation de ces compétences par les enfants, de même que l'apprentissage d'une seconde langue par les adultes présente des similitudes fondamentales avec les processus correspondants chez les enfants. C'est pourquoi il est important d'envisager l'acquisition de l'alphabétisme et son évaluation comme des questions *pertinentes tout au long de la vie* dans le domaine de l'éducation et du développement³⁰¹.

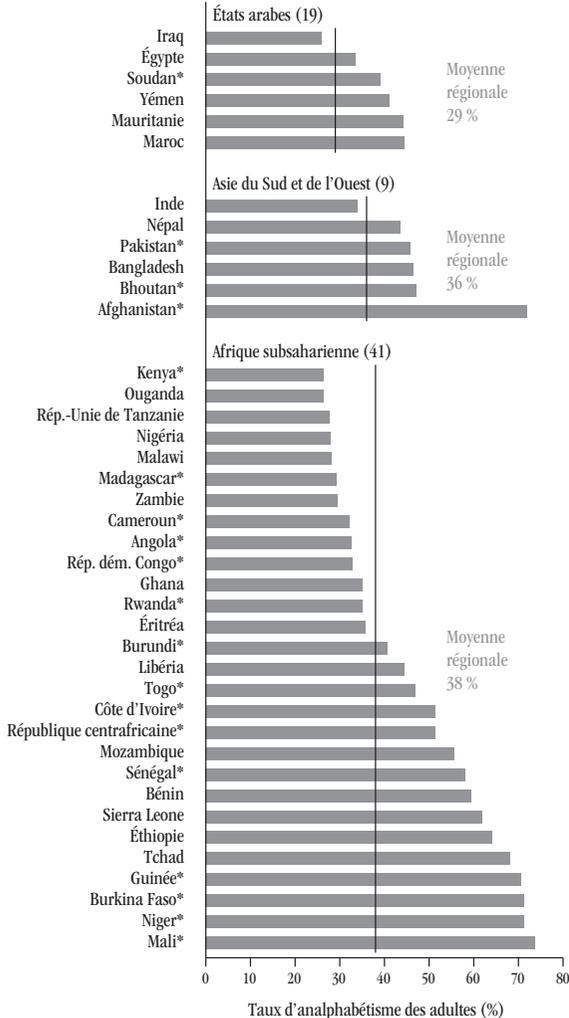
299. UNESCO, 2005. *Le Rapport mondial de suivi sur l'EPT 2006, L'alphabétisation, un enjeu vital* reprend les questions d'alphabétisation des enfants et des adultes, ainsi que les statistiques et l'évaluation de l'alphabétisation.

300. Voir Wagner (1995) pour une analyse approfondie de chacune de ces raisons.

301. Pour en savoir plus sur une approche de l'alphabétisation tout au long de la vie, voir Wagner, 2010.

Figure 8.1 L'analphabétisme dans une sélection de pays en développement, par région

Taux d'analphabétisme des adultes (à partir de 15 ans) dans les pays où ce taux atteint au moins 25 %, dans une sélection de régions, 2000-2007



Notes : L'astérisque indique que les données nationales sur l'alphabétisation ont été utilisées. Pour tous les autres, ce sont les estimations de l'ISU. Ces estimations ont été obtenues à l'aide du modèle mondial de projections de l'alphabétisation par âge (Global Age-specific Literacy Projections) de l'ISU. Les chiffres indiqués entre parenthèses à la suite du nom des régions indiquent le nombre de pays de la région qui disposent de données publiables.

Adapté de : UNESCO, 2010, p. 96.

Évaluer l'alphabétisation des adultes

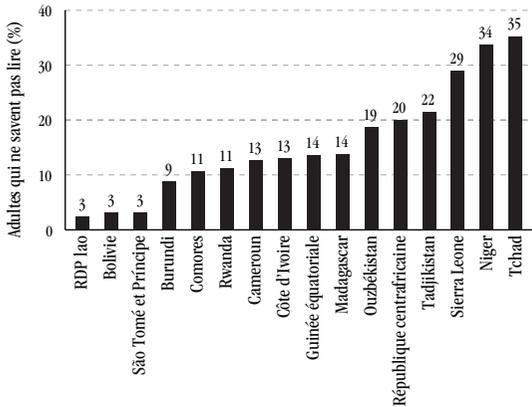
Jusqu'au milieu des années 1990, la plupart des pays mesuraient l'alphabétisation des adultes en leur posant uniquement cette simple question : « Savez-vous lire ou pas ? » Généralement, cette question était posée lors d'un recensement et non sous forme d'une évaluation directe des compétences (*figure 8.2*)³⁰². Comme il a déjà été mentionné, l'UNESCO sollicite partout dans le monde des données sur l'alphabétisme et, dans de nombreux pays (en particulier les pays en développement), ces données sont fournies par rapport au nombre de personnes « alphabétisées » et « analphabètes »³⁰³. Pour la plupart des pays, ce type de classification dichotomique est relativement peu onéreux à réaliser et pose peu de problèmes pratiques (ou techniques), tout en fournissant aux agences internationales un cadre transnational et chronologique pour analyser l'alphabétisation par régions géographiques ou économiques mondiales³⁰⁴.

302. À propos de l'étude sur l'alphabétisation des adultes préparée pour les réunions sur l'EPT à Dakar, voir Wagner, 2000.

303. En effet, certaines données pour les PMA sur l'alphabétisation des adultes dans le *Rapport mondial de suivi de l'EPT* consacré à l'alphabétisation (UNESCO, 2005) restent des sources de données dichotomiques de ce type. Mais il y a eu une évolution au cours des dernières décennies. Selon Smyth (2005, p. 12), les données sur l'alphabétisation dans UNESCO (1978) reposaient sur ce qui suit : « Dans l'exercice de prévision, on a utilisé les taux de scolarisation du groupe des 6-11 ans pour estimer les taux d'analphabétisme futurs du groupe des 15-19 ans, lesquels ... ont été utilisés pour estimer les taux d'analphabétisme futurs de la population à partir de 15 ans ». En outre (Smyth, p. 21), de nombreux pays ont réuni des informations par le biais du recensement national ; mais « à n'importe quel moment des années 1980 et 1990, seul un nombre limité de pays disposaient de données actualisées ; ... pour les autres, les données les plus récentes dataient de 10, 15 ou 20 ans, voire plus ... ». En d'autres termes, même s'il existait des données de recensement, elles reposaient presque toujours sur un « analphabétisme » autodéclaré, sans information sur la langue lue et écrite ; par ailleurs, près de la moitié des pays utilisaient des données périmées depuis déjà au moins une ou plusieurs décennies.

304. Un rapport de l'ISU (Carr-Hill, 2008, p. 18) affirme que la définition dichotomique utilisée lors des recensements devrait être conservée : « Le problème avec les méthodes multidimensionnelles est que les différentes dimensions et les relations qui existent entre elles risquent davantage d'être comprises différemment selon les époques et les cultures, contrairement à la simple variable dichotomique, « Savez-vous lire et écrire ? », qui a plus de chances d'être perçue et comprise de la même façon à toute époque et quelle que soit la culture. ... La variable dichotomique est souvent utilisée comme variable explicative et l'évolution du niveau d'alphabétisation mesuré de cette façon sert, par exemple, de facteur prédictif des taux de fertilité probables ». Pour une analyse critique de ces « dichotomies » dans les statistiques sur l'alphabétisation, voir Wagner, 2001.

Figure 8.2 Adultes n'ayant pas dépassé le niveau d'éducation primaire et déclarant ne pas savoir lire



Source : Calculs fondés sur des chiffres de la base de données MICS de l'UNICEF. Adapté de UNESCO, 2005, p. 128.

Cette dichotomie autodéclarée étant un instrument de mesure peu précis, de valeur stratégique et individuelle limitée, la communauté internationale a fait des efforts substantiels au cours des deux dernières décennies pour se faire une idée plus précise des niveaux spécifiques d'alphabétisation des adultes³⁰⁵. Le premier effort international majeur d'évaluation des compétences a été, en 1995, l'*Enquête internationale sur l'alphabétisation des adultes* (EIAA), entreprise principalement dans les pays industrialisés³⁰⁶. Cette enquête utilise une méthode de catégorisation à cinq niveaux de l'alphabétisation selon trois échelles différentes : lecture courante, compréhension de textes schématisés et aptitude au calcul. Des critiques ont attiré l'attention sur les problèmes que posent ces échelles en

305. Comme déjà mentionné dans la note 49 (*chapitre 4*), l'enquête sur l'alphabétisation des adultes effectuée auprès des ménages au Zimbabwe, dans deux langues africaines locales (UNSO, 1989), a été l'une des premières tentatives pour aller au-delà de la simple collecte de données de type dichotomique.

306. OCDE/Statistique Canada, 1995, 1997, 2000. La méthodologie EIAA repose en grande partie sur plusieurs antécédents nationaux, tels que l'enquête américaine sur l'alphabétisation des adultes de 1993, *U.S. National Adult Literacy Survey* (Kirsch *et al.*, 1993), qui a investi d'importantes ressources pour améliorer les propriétés techniques et psychométriques des instruments d'évaluation de l'alphabétisation en utilisant diverses techniques, entre autres des méthodes pour élargir la fourchette d'items utilisés dans les enquêtes, notamment la théorie de la réponse aux items.

termes de comparabilité internationale, de différences d'échantillonnage des populations dans les pays concernés et de comparabilité des items^{307, 308}.

À peu près à la même époque, l'Institut international de l'alphabétisation (ILI)³⁰⁹ a collaboré avec l'UNESCO au *Programme d'évaluation de l'alphabétisation* (LAP), qui privilégiait les initiatives à plus petite échelle et plus flexibles pour évaluer l'alphabétisation des adultes dans les PMA. Le but était à la fois d'obtenir des données plus précises sur les taux d'alphabétisation, de rendre ces données plus transparentes et accessibles aux principaux utilisateurs finaux (les agences et les programmes d'alphabétisation des adultes), tout en accélérant la vitesse de publication des résultats. Les rapports du LAP se sont intéressés à plusieurs types d'outils d'évaluation, et plusieurs initiatives pilotes ont été financées dans les pays en développement³¹⁰. Le LAP a également insisté sur la notion déjà évoquée ici de « partageabilité » entre agences internationales et programmes locaux, avec des initiatives visant à mettre en place des méthodes et des outils de stockage des données faciles à utiliser. L'idée était de remédier à cette situation problématique (mise en évidence dans l'EIAA et d'autres LSEA), où seul un groupe restreint de spécialistes était capable de comprendre (et donc de remettre en cause ou de réutiliser) les données recueillies. Le travail du LAP sur les évaluations plus simples et plus rapides a contribué à l'élaboration du modèle SQC d'évaluation hybride et a préfiguré le travail ultérieur de l'EGRA auprès des enfants³¹¹.

À l'occasion du lancement, en 2003, de la Décennie des Nations Unies pour l'alphabétisation, l'ISU a initié le *Programme d'évaluation et de suivi de l'alphabétisation* (LAMP), qui s'appuie sur quelques-uns des outils développés dans le cadre des EIAA, mais avec un recentrage sur l'évaluation

307. Voir Kalton *et al.*, 1998 ; Levine, 1998.

308. En 2011, l'OCDE a lancé une étude de suivi de l'alphabétisation des adultes (après l'EIAA), *Programme pour l'évaluation internationale des adultes* (PIAAC) ; voir www.oecd.org/dataoecd/13/45/41690983.pdf

309. L'Institut international de l'alphabétisation (Université de Pennsylvanie) a été cofondé par l'UNESCO en 1994.

310. Voir ILI-UNESCO (1998, 1999, 2002) pour plus d'informations sur le LAP. Les documents correspondants peuvent être téléchargés sur le site Web de l'ILI : www.literacy.org

311. On retrouve dans le LAP la plupart des principes SQC (Wagner, 2003) sur lesquels se fonde l'approche EGRA, notamment la détection précoce, la validité avec une moindre dépendance à l'égard de la collecte chronophage de grandes quantités de données, la rigueur empirique par rapport à certaines méthodologies locales antérieures et l'importance de l'administration en temps opportun.

de l’alphabétisation des adultes dans les pays en développement³¹². Au cours des six dernières années, le LAMP a procédé à des essais pilotes pour tester divers instruments destinés à améliorer l’information sur les taux d’alphabétisation dans les PMA. Il a tenté d’appliquer une approche orientée vers la comparabilité internationale, tout en tenant compte de la diversité culturelle, avec les difficultés inhérentes (déjà évoquées auparavant) à un tel parti pris³¹³. À l’instar du SACMEQ, le LAMP a fait du renforcement des capacités dans les pays en développement l’une de ses missions premières³¹⁴.

Besoins des apprenants adultes

L’un des principaux traits distinctifs des programmes d’éducation non formelle et d’éducation pour les adultes est que la participation y est presque toujours facultative (contrairement à la scolarisation)³¹⁵. Cela signifie que, dans les contextes d’apprentissage pour adultes, les apprenants sont amenés à faire des choix importants. Ils peuvent par exemple se demander : que vais-je retirer de ma participation à ce programme d’alphabétisation, en particulier par rapport à mes besoins professionnels et personnels, au sein de ma famille ou de ma communauté ?³¹⁶ Les taux d’abandon des programmes d’éducation des adultes dans le monde tournent souvent autour de 50 % au cours de la première moitié des programmes d’étude substantiels. C’est un

312. Voir ISU, 2009. Le LAMP a débuté en 2003, grâce à un financement important de la Banque mondiale et de l’UNESCO. Depuis lors, l’accent a surtout été mis sur la planification et la conception, avec des essais pilotes menés en El Salvador, au Maroc, en Mongolie, au Niger et dans les Territoires palestiniens autonomes.

313. Dans le LAMP, la comparabilité est décrite comme suit : « La composante de comparaison est liée à la nécessité d’une identité de vues correspondant à l’universalité du droit à l’éducation, ce qui exclut l’adoption de procédures différenciées (ou de doubles normes) susceptibles d’introduire une discrimination. L’existence d’ensembles différenciés de définitions de l’alphabétisation pour les pauvres et les riches, les femmes et les hommes, les populations indigènes et non indigènes serait la porte ouverte à une pratique potentiellement très discriminatoire, pouvant entraîner des disparités dans l’octroi du droit à l’éducation » (LAMP, 2009, p. 24). Cette explication ne semble pas pouvoir donner une orientation de fond pour résoudre la question de savoir comment les différences culturelles peuvent satisfaire la norme de comparaison internationale.

314. Pour une étude récente des initiatives de l’UNESCO en matière d’évaluation de l’alphabétisation, voir Wagner, 2011.

315. Certaines campagnes d’alphabétisation lancées peu après la Seconde Guerre mondiale ont été, en fait, obligatoires (Arnove et Graff, 1987).

316. Diverses études ethnographiques des programmes d’alphabétisation des adultes corroborent les doutes concernant l’intérêt de nombreux apprenants adultes pour ces programmes et leur volonté d’y consacrer du temps. Voir, par exemple, Puchner, 2001 et Robinson-Pant, 2004.

indicateur souvent cité que les apprenants adultes « votent avec leurs pieds » quand les programmes se révèlent peu adaptés à leurs centres d'intérêt³¹⁷.

Les pays soucieux d'abolir les inégalités devront recueillir des informations plus précises sur les trajectoires d'apprentissage des apprenants adultes, ainsi que sur leurs dispositions psychologiques vis-à-vis de leur participation, leurs objectifs particuliers et leurs langues spécifiques³¹⁸. Améliorer les mesures d'alphabétisation et la collecte de données peut apporter de meilleures réponses aux programmes pour apprenants adultes, ainsi qu'aux adultes eux-mêmes. Dans le futur, on devrait voir se multiplier les évaluations hybrides des apprenants adultes (avec des méthodes du type SQC et EGRA), notamment parce qu'il est important de fournir un retour d'information en temps opportun aux agences et à ceux qui participent aux programmes.

Acquisition de la lecture chez l'enfant et chez l'adulte

Par rapport à la recherche sur les processus de base chez les enfants, ... la recherche fondamentale sur les processus de lecture chez les adultes peu alphabétisés s'appauvrit³¹⁹.

Quels sont les points communs et les différences entre les débuts de l'apprentissage de la lecture chez les enfants et chez les adultes ? Les adultes ayant un répertoire beaucoup plus complet de compétences cognitives et linguistiques (ainsi que de connaissances générales) que la plupart des

317. Ce point de vue va aussi dans le sens des approches ethnographiques de l'alphabétisation des adultes axées sur leurs usages et pratiques en matière de lecture et d'écriture. Voir, par exemple, Papen, 2005 et Street, 2001.

318. Robinson (2004, p. 15-16) indique : « Le préjugé de certaines élites et de certains groupes est tel que, dans certaines situations, une langue est définie en tant que telle [uniquement] parce qu'elle est écrite, ce qui condamne les langues non écrites à un statut inférieur, souvent celui de dialecte et non de langue. ... Le mythe persiste dans certains milieux que l'acquisition de l'alphabétisation dans une langue diminue les chances de le faire de façon satisfaisante dans une autre : ainsi, apprendre à lire et à écrire dans la langue maternelle peut être considéré comme un frein à l'apprentissage de la lecture et de l'écriture dans une langue d'usage plus répandu. » De plus, en ce qui concerne le domaine du choix linguistique des adultes, il est important de ne pas oublier que beaucoup d'individus vivent dans des contextes de diglossie, où plus d'une langue est utilisée pour communiquer. Dans de nombreux pays arabophones, par exemple, les femmes (en particulier) parlent l'arabe vernaculaire, alors que les hommes parlent généralement l'arabe vernaculaire et l'arabe littéraire (Khachan, 2009). Cela peut poser des problèmes considérables pour concevoir et promouvoir des programmes d'alphabétisation des adultes en arabe.

319. Venezky et Sabatini, 2002, p. 217.

lecteurs débutants à l'école primaire, on pourrait penser que l'acquisition de la lecture sera plus rapide chez les adultes, mais peut-être en suivant les mêmes schémas que pour les enfants (*chapitres 5 et 6*). Bien qu'il s'agisse d'un argument de poids sur le plan conceptuel, peu de recherches empiriques ont été entreprises sur ce sujet, que ce soit dans les pays industrialisés ou dans les pays en développement. D'après les travaux disponibles à ce jour, il semble qu'il y ait quelques points communs et certaines différences entre l'acquisition de la lecture par les adultes et les enfants³²⁰. La diversité des langues à laquelle sont confrontés aussi bien les adultes que les enfants dans des sociétés multilingues est indéniablement un point commun. Le lien entre diversité linguistique et alphabétisation des adultes est frappant.

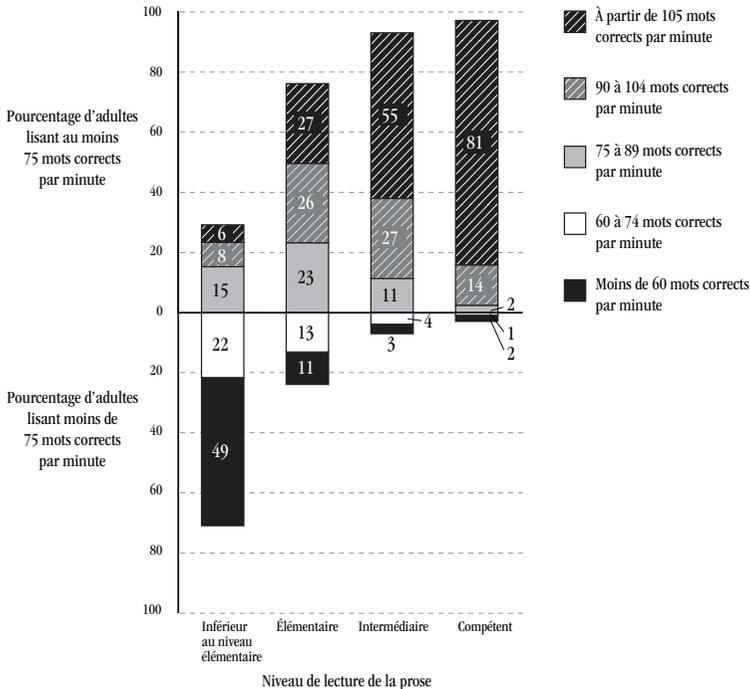
Un rapport récent, inspiré d'une évaluation antérieure de l'alphabétisation des adultes aux États-Unis, s'est intéressé aux profils de ceux que l'on a qualifiés d'« adultes les moins alphabétisés », c'est-à-dire ceux se situant au niveau le plus bas de l'échelle de l'alphabétisation³²¹. Ce rapport constate que près de 60 % de ces adultes vivent en dessous du seuil de pauvreté des États-Unis, et que l'on avait auparavant diagnostiqué chez un nombre significatif d'entre eux des difficultés d'apprentissage. Fait intéressant à noter du point de vue du présent ouvrage, cette étude compare la fluidité de la lecture aux différents niveaux de maîtrise de la lecture chez les

320. Durgunolu et Öney (2002) ont testé un échantillon de femmes turques participant à un programme d'alphabétisation pour adultes avant le début du programme et après 90 heures d'enseignement, en évaluant plusieurs compétences cognitives. Ils ont comparé les résultats avec les données obtenues lors d'études antérieures concernant des enfants apprenant à lire en turque. Ils concluent que les compétences clés de développement de la lecture et de l'écriture sont les mêmes pour les enfants et les adultes, mais que la conscience phonologique, en tant que facteur prédictif de l'aptitude à lire, fonctionne différemment après la première année d'école, dans le cas des enfants, ou les 90 heures de cours, dans le cas des adultes. Par ailleurs, en comparant ces résultats aux compétences observées aux États-Unis chez des enfants anglophones, ils se sont aperçus que le rôle joué par la reconnaissance des lettres, la conscience phonologique et la compréhension orale était fortement lié à la langue cible. Dans une autre étude réalisée aux États-Unis, Greenberg *et al.* (2002) ont apparié des enfants anglophones et des apprenants anglophones adultes de niveau équivalent en lecture (de la 3^e à la 5^e année) en fonction du genre, de la race et du lieu de résidence. Ils ont ensuite étudié les erreurs de reconnaissance de mots dans les deux échantillons. Ils ont constaté que les adultes utilisent généralement des stratégies plus visuelles et orthographiques, quand ils rencontrent des problèmes de reconnaissance de mots, alors que les enfants de niveau équivalent en lecture utilisent davantage le décodage et d'autres stratégies phonétiques, que ce soit pour la reconnaissance des mots ou pour l'orthographe.

321. US Department of Education, NCES, 2009.

adultes, depuis « inférieur au niveau élémentaire » jusqu'à « compétent ». Comme le montre la *figure 8.3*, le risque est beaucoup plus grand pour les adultes qui ont le niveau le plus bas d'avoir une faible fluidité de la lecture (moins de 75 mots corrects par minute) que pour les adultes se situant aux trois niveaux les plus élevés. Les compétences de base en décodage sont un problème pour ces adultes, de façon similaire aux enfants qui commencent à lire. Si l'on prend le modèle cognitif évoqué au *chapitre 5*, cette similitude semble évidente.

Figure 8.3 Pourcentage d'adultes par niveau de compétences de base en lecture dans l'évaluation nationale de l'alphabétisation des adultes réalisée aux États-Unis



Note : La somme des chiffres peut ne pas correspondre aux totaux à cause des arrondis. Sont inclus dans la catégorie des adultes les personnes de 16 ans et plus, qui vivent chez elles ou qui sont en prison. Les adultes n'ayant pu être interrogés pour des raisons de langue ou de déficiences cognitives ou mentales (3 % en 2003) ne sont pas pris en compte dans ces chiffres.

Source : Ministère de l'Éducation américain, Institute of Education Sciences, Centre national de statistiques de l'éducation (NCES), Évaluation nationale de l'alphabétisme des adultes (NAAL) 2003.

Adapté de : Ministère de l'Éducation américain, 2009, p. 23.

Perte d'alphabétisme

Les enfants qui n'atteignent pas un certain niveau de fluidité de lecture entre [la 1^{re} et la 3^e année] risquent probablement de retomber dans l'analphabétisme s'ils abandonnent l'école en [4^e ou 5^e année]³²².

Comme il a été mentionné au *chapitre 1*, l'une des questions les plus importantes de la politique d'éducation mondiale est la durée de scolarité nécessaire pour que les enfants (ou les adultes) atteignent un taux suffisant d'alphabétisme et d'autres compétences pour avoir une influence sur la vie sociale et économique. En effet, d'aucuns soutiennent parfois que, dans les pays en développement, un minimum de quatre à six années de scolarisation des enfants constitue la base des ressources humaines intellectuelles indispensables pour assurer la croissance économique nationale³²³. Une thèse similaire pourrait également être avancée à propos de la durée d'instruction requise pour que les apprenants conservent les compétences acquises. L'argument est qu'un nombre minimal d'années d'éducation est nécessaire pour que des compétences en lecture plus ou moins permanentes soient acquises par un enfant d'âge scolaire, un adolescent ou un adulte.

Dans ce raisonnement, le concept de rétention de l'alphabétisation (ou, à l'inverse, de ce qui est souvent qualifié de « perte » d'alphabétisme) est central, car on considère que ce que les enfants apprennent et retiennent de leurs années d'école – ou des programmes d'alphabétisation et d'éducation non formelle dans le cas des adultes – est ce qui peut être utilisé dans des activités économiques productives plus tard, au cours de la vie. L'oubli de ce qui a été enseigné dans un programme éducatif constitue du gaspillage pédagogique. Ces individus (enfants ou adultes) n'atteindront pas le seuil présumé d'apprentissage minimal requis pour conserver les acquis et maintenir l'apprentissage autonome. Des arguments similaires ont été avancés dans d'autres domaines cognitifs, tels que l'acquisition d'une langue étrangère (et d'une seconde langue).

Très peu de recherches empiriques ont traité directement de la rétention de l'alphabétisation dans les pays en développement³²⁴. Le manque d'outils d'évaluation adaptés a jusqu'à présent limité les possibilités de

322. Chabbott, 2006, p. 25.

323. Comme il a été dit au *chapitre 5*, des arguments similaires sont avancés concernant le passage de L1 à L2, pour atteindre un niveau d'acquisition de compétences.

324. La Banque mondiale a récemment accordé plus d'attention à cette question dans une étude des évaluations (Abadzi, 2003). Les études antérieures sont celles de Comings, 1995 ; Hartley et Swanson, 1986 ; Roy et Kapoor, 1975.

tirer des conclusions fiables dans les pays en développement. De plus, des études longitudinales sont nécessaires pour pouvoir comparer les résultats d'un élève ou d'un adulte suivant un programme d'alphabétisation et ses performances quelques années après avoir quitté le contexte éducatif. La recherche empirique limitée dans ce domaine ne vient pas vraiment corroborer la notion de perte d'alphabétisme³²⁵. Il est manifeste qu'il faut développer les études, ce que devrait favoriser l'apparition de nouveaux outils d'évaluation hybride.

Faire progresser l'évaluation des compétences en lecture des adultes

La plupart des évaluations de l'acquisition de compétences en lecture concernaient les enfants, et non les jeunes ou les adultes. Comme déjà mentionné auparavant, l'essentiel de la recherche a porté sur quelques langues seulement. Par conséquent, il reste encore beaucoup à faire pour obtenir un tableau plus complet de l'acquisition de la lecture chez les adultes qui soit l'équivalent de celui de l'acquisition chez les enfants, en particulier en raison de l'importance constante de l'alphabétisation des adultes dans les PMA. Les rares études qui existent semblent indiquer que l'apprentissage des compétences suit globalement le même schéma chez les enfants et les adultes. Pourtant, il existe des différences importantes. Les adultes ont un vocabulaire actif plus étendu que celui des jeunes enfants : quel rôle cela joue-t-il dans l'acquisition de la lecture chez les adultes ? D'autres questions se posent à propos de l'alphabétisation des adultes : quel rôle l'attitude et la motivation jouent-elles ? Pourquoi les adultes *devraient-ils* avoir envie d'apprendre à lire ? Comment les responsables politiques et les spécialistes de l'éducation peuvent-ils influencer sur la réponse à cette dernière question ? De meilleurs outils d'évaluation du travail d'alphabétisation des adultes dans les PMA devraient permettre d'apporter des réponses à ces questions dans un proche avenir.

325. Dans l'une des rares études longitudinales menées, Wagner *et al.* (1989) se sont intéressés à la rétention de l'alphabétisation chez des adolescents marocains, qui avaient tous abandonné l'école avant la fin de la 5^e année et qui ont été observés dans leur vie quotidienne pendant les deux années qui ont suivi. Les résultats montrent que, deux ans après la fin de la scolarité, les compétences en lecture et écriture en arabe n'avaient pas été perdues. En effet, selon la nature de l'expérience postscolaire (par exemple travail hors de la maison par opposition aux tâches ménagères accomplies essentiellement au sein de la famille), de nombreux adolescents avaient en fait développé leurs compétences en lecture et en écriture.

9. RECOMMANDATIONS

Si vous ne savez pas où vous allez, n'importe quelle route vous y conduira³²⁶.

Les évaluations ont encore un bel avenir et seront de plus en plus utilisées à l'échelle mondiale et locale pour diverses finalités stratégiques et concrètes. Le présent ouvrage a présenté quelques-uns des avantages et des difficultés liés à l'utilisation des évaluations des acquis scolaires, en s'intéressant plus particulièrement aux pays pauvres et en développement. Il n'y a pas qu'une seule manière d'évaluer les compétences, et la finalité des évaluations peut être très différente d'un pays à l'autre. Il n'y a pas non plus d'évaluation idéale, mais plutôt diverses approches scientifiques qui peuvent fournir et fourniront des moyens fiables et crédibles d'améliorer la qualité de l'éducation. Il n'existe pas de panacée.

Il existe une multitude d'outils de mesure et d'évaluation. Il faut de toute urgence calibrer les évaluations en fonction d'objectifs stratégiques spécifiques, des impératifs temporels et de leurs coûts. Ces considérations, ainsi que d'autres, ont débouché sur un ensemble de recommandations stratégiques qui sont résumées ci-après. Le *tableau 9.1* donne une vue synthétique des avantages et des inconvénients de divers outils d'évaluation.

Le test de lecture « idéal » n'existe pas

Un test de lecture, comme n'importe quel outil d'évaluation, n'est utile que dans la mesure où il répond à des besoins stratégiques particuliers. À l'une des extrémités du spectre, on trouve des LSEA d'envergure relativement large (comme les enquêtes PIRLS et PISA), dont la préoccupation première est d'assurer la qualité et la comparabilité internationale, et qui demandent plusieurs années pour produire des résultats aux niveaux international et national. Les évaluations régionales (SACMEQ, PASEC et LLECE, par exemple) fonctionnent un peu comme les LSEA de grande envergure, mais prennent en compte de façon significative les aspects régionaux de l'enseignement.

Les évaluations hybrides, comme l'EGRA (dans plusieurs variantes actuelles), se caractérisent par une moindre normalisation et une comparabilité transnationale limitée. Elles mettent l'accent sur les

326. Carroll, 1865.

composantes cognitives qui sont à la base de l'apprentissage de la lecture chez les jeunes enfants. L'EGRA peut aussi servir de guide pour améliorer l'enseignement, et plusieurs études sur les interventions dans ce sens sont actuellement en cours.

Tableau 9.1 Résumé des avantages et des limites de diverses évaluations

Type d'évaluation	Avantages	Limites
LSEA/évaluations internationales	<ul style="list-style-type: none"> • Crédibilité mondiale • Comparabilité entre nations • Affirmations généralisées concernant les liens entre variables • Technicité statistique • Développement des capacités • Évaluation fondée sur des échantillons • Analyses secondaires possibles à partir des bases de données stockées 	<ul style="list-style-type: none"> • Effort considérable pour assurer la comparabilité • Les compromis nécessaires pour obtenir la comparabilité technique peuvent aboutir à une perte de validité au niveau local • Les compromis pour obtenir des échantillons internationaux peuvent laisser de côté des groupes importants (par exemple par langue, ethnicité, citoyenneté) • Le délai nécessaire pour obtenir des résultats est généralement de 3 à 5 ans minimum • L'âge d'évaluation, pour les tests de groupes, ne commence pas avant la 4^e année, au plus tôt • Elles exigent, en règle générale, des compétences de haut niveau en statistique pour le traitement des données (par exemple TRI, HLM) • Les données sont souvent trop complexes ou trop tardives pour permettre des analyses locales/nationales • Les coûts généraux sont importants, en particulier quand les frais de personnel sont inclus en totalité
Évaluations régionales	<ul style="list-style-type: none"> • La dimension régionale permet aux pays participants de travailler ensemble pour atteindre des objectifs communs • Les évaluations peuvent contenir des éléments des évaluations nationales, ce qui confère une plus grande validité par rapport aux systèmes scolaires • Le renforcement des capacités est un aspect important des efforts, en raison du lien avec les ministères nationaux de l'Éducation et leur personnel 	<ul style="list-style-type: none"> • L'enjeu régional ne coïncide ni avec les évaluations internationales ni avec les évaluations du type SQC/EGRA • Le délai pour les mener à bien semble aussi long (voire plus long) que les LSEA, mieux financées • Variance régionale par premières années testées et fréquence de l'évaluation incertaine

Type d'évaluation	Avantages	Limites
Évaluations nationales	<ul style="list-style-type: none"> • Soutien direct à la mission du ministère de l'Éducation • Utilise le personnel du ministère de l'Éducation • L'évaluation couvre tous les enfants scolarisés (base censitaire) • Grande validité concomitante avec le contenu des programmes scolaires 	<ul style="list-style-type: none"> • Peu de rapports avec la collecte de données dans d'autres pays • Ne couvrent pas les enfants non scolarisés • Les données ne sont disponibles qu'à la fin de l'année scolaire, voire après • Les données recueillies peuvent ne pas avoir de lien avec les variables d'enseignement fondamentales • Effectifs importants de personnel mobilisés pour recueillir les données
SQC/EGRA	<ul style="list-style-type: none"> • Localisation de la conception et du contenu des items de test, notamment dans la langue maternelle • Évaluations fondées sur des échantillons • Les données peuvent être recueillies par les enseignants • Possibilité de « cibler » des populations particulières (par exemple par langue, ethnicité, citoyenneté, jeunes non scolarisés) • Valeur placée sur les compétences cognitives de base nécessaires pour acquérir la maîtrise de la lecture • L'évaluation peut commencer à un âge précoce (1^{re} année), ce qui permet de détecter tôt les problèmes • Possibilité d'agir sur l'enseignement au niveau individuel, grâce aux tests individualisés • Peut étayer un développement professionnel dirigé • Une démarche individualisée peut être adaptée à d'autres âges (alphabétisation des adultes notamment) • Possibilité d'avoir un impact stratégique non seulement au niveau national, mais aussi au niveau des provinces, des écoles et des enseignants • Le temps nécessaire ainsi que le coût par élève évalué seront probablement inférieurs à celui des autres évaluations • Peuvent être entreprises par des ONG en collaboration avec le gouvernement – lancement et temps d'exécution rapides 	<ul style="list-style-type: none"> • En 2011, la crédibilité globale reste modeste, bien qu'en progrès • Le renforcement des capacités locales exige davantage d'attention • S'intéressent en général seulement aux trois premières années de scolarité • Validité concomitante limitée avec le contenu des programmes scolaires • Si administrées principalement par une ONG, peuvent être ignorées par le ministère de l'Éducation • Analyses secondaires peu probables

Les normes constituent un autre problème. Si le but est de mieux comprendre comment les meilleurs élèves apprennent (afin de faire le poids face aux pays de l'OCDE les plus performants), il est alors pertinent de s'inspirer de ces pays pour établir des normes d'apprentissage ou des référentiels. Si, en revanche, l'objectif est de faire acquérir à tous les enfants des compétences de base (ainsi que la maîtrise de la lecture et de l'écriture) en tant que norme de résultat, même dans les contextes les plus pauvres, alors les questions de multilinguisme, de contenu local et de déficiences systémiques devront être privilégiées.

Enfin, ces différents types de test sont complémentaires. Le test idéal n'existe pas, mais les responsables politiques doivent préciser leurs objectifs avant d'opter pour une approche ou une autre.

En cas de doute, opter pour des évaluations de taille modeste

Les LSEA sont généralement de grande envergure : le nombre de pays participants est important, les échantillons de population sont grands, les outils de test doivent être validés par des experts, le coût global se chiffre souvent en millions de dollars, et, enfin, il faut souvent plusieurs années pour les clôturer. Ce modèle est en accord avec l'objectif général de la plupart des LSEA, à savoir fournir une méthodologie de référentiel extrêmement crédible, permettant à un ministre de l'Éducation de suivre l'évolution des niveaux nationaux de performance par rapport à d'autres pays (ainsi que d'autres résultats).

Les évaluations hybrides plus modestes s'appuient au contraire sur un ensemble d'acteurs nationaux et locaux, comprenant les directeurs et les enseignants, et éventuellement les employeurs, les communautés et les parents. Les apprenants peuvent être considérés comme des parties prenantes, dans la mesure où ils sont directement concernés par la qualité du programme qu'ils suivent. Les évaluations hybrides peuvent tirer parti de leur taille modeste pour explorer plus avant les facteurs multiples (et souvent conditionnés par le contexte) qui influent sur les résultats d'apprentissage, comme la langue d'enseignement, la langue d'évaluation et les possibilités d'apprendre. L'engagement précoce et l'implication de cet ensemble hétérogène d'intervenants peuvent également être pris en considération. C'est ce que l'EGRA et d'autres évaluations du même type s'efforcent de faire. Dans l'ensemble, les évaluations SQC ont un avantage spécifique lié à leur dimension modeste, en ce sens que les besoins en ressources humaines peuvent être mieux adaptés aux réalités des capacités humaines

des sociétés à faibles revenus. Ces évaluations doivent être soigneusement dimensionnées pour être « juste à la bonne taille »³²⁷.

Les résultats les plus rapides sont les meilleurs

Certaines évaluations, en particulier aux niveaux international ou régional, sont administrées tous les 3 ou 5 ans, voire tous les 10 ans. Il y a des coûts évidents de temps, d'argent et d'opportunités perdues associés aux évaluations qui ont un délai d'exécution relativement long. Comme nous l'avons vu, le délai de clôture de la plupart des LSEA peut être une limite extrêmement sérieuse.

Inversement, avec l'apparition des évaluations hybrides plus rapides, dont les objectifs et les tailles d'échantillon sont plus modestes et plus spécifiques, les délais d'administration deviennent plus réalistes, avec un budget plus limité. La fréquence devient moins importante si l'on ne vise pas à tout prix des interventions à court terme. Mais, si la détection précoce des problèmes est un but important, afin de mettre en œuvre de nouvelles politiques qui seront jugées sur leur impact à court terme, alors la fréquence devient un moyen clé d'obtenir des résultats en rapport avec le but visé. De plus, avec les évaluations SQC hybrides, il devient possible de fournir les résultats presque en temps réel, en quelques mois le cas échéant et, généralement, dans un délai qui permet au décideur ayant autorisé l'étude d'en voir les résultats³²⁸. Il va sans dire que les évaluations menées en temps réel peuvent avoir d'énormes avantages pour les enseignants, les écoles et les élèves, dont la vie peut être impactée positivement par les résultats.

Les résultats des évaluations ne sont pas toujours à la hauteur du prix payé

[Nous devons] cultiver chez tous les acteurs concernés (classe politique, décideurs politiques, personnel éducatif, chefs d'établissement,

327. Abraham Lincoln, le président américain, est souvent cité pour avoir tenu ces propos : À la question « Quelle doit être la longueur des jambes d'un homme par rapport à son corps ? », il aurait répondu : « Je ne me suis pas beaucoup penché sur la question mais, à première vue, il me semble qu'elles doivent être suffisamment longues pour aller de son corps jusqu'au sol. » C'était, semble-t-il, pendant sa campagne présidentielle (1858), alors qu'on lui avait demandé combien il mesurait. La réponse sous-entendait : je suis juste assez grand.

328. Dans les tests ASER, effectués par Pratham en Inde, l'un des objectifs était de fournir immédiatement à la communauté un retour d'information sur l'apprentissage précoce de la lecture. Pour le moment, les détails sur ces délais exceptionnels ne sont pas encore disponibles.

enseignants et parents) une meilleure appréciation du pouvoir et de la rentabilité des évaluations. Cela nécessite la mise en place d'un cadre général pour structurer la discussion sur l'évaluation et la capacité d'évaluer, ainsi que des études de cas permettant d'étayer les retours sur investissement (favorables) issus d'investissements bien planifiés dans l'évaluation, souvent dans le cadre d'une initiative de réforme générale de l'éducation³²⁹.

Selon un vieux dicton : si vous trouvez que le savoir coûte cher, essayez l'ignorance. Si les politiques que l'on tente d'appliquer (telles que l'éducation de base universelle, avec comme composante principale l'alphabetisation des enfants) échouent, alors le coût de ces échecs doit être mis en balance avec ce que coûtent les outils de remédiation aux stratégies inefficaces. Curieusement, le coût des évaluations relativement onéreuses par rapport à celles qui le sont moins n'a que rarement été un sujet de préoccupation majeure chez les responsables politiques.

Comme il a été décrit au *chapitre 7*, les coûts des divers types d'évaluation sont très variables. Cependant, si l'on considère le coût par personne, on observe des similitudes entre les LSEA et les évaluations hybrides, en raison des économies d'échelle des premières. Le coût total des LSEA peut toutefois être considérable, compte tenu du nombre de pays participants, de la nécessité d'une expertise professionnelle de haut niveau et des exigences techniques de la collecte et de l'analyse des données.

Il existe évidemment des arbitrages dans les processus de détermination des coûts : limitation de la taille des échantillons, longueur des tests créés, niveau de qualification du personnel requis. La plupart de ces arbitrages sont plus faciles à effectuer dans les évaluations hybrides, du fait que le degré de comparabilité est moindre et qu'il est possible de mettre l'accent sur un ensemble limité d'objectifs stratégiques locaux ou nationaux. Les évaluations hybrides, en particulier à une époque où la fréquence et la ponctualité prennent davantage d'importance, contribueront probablement à faire baisser le coût des activités d'évaluation.

Les évaluations des acquis scolaires doivent commencer le plus tôt possible (dans certaines limites)

Comme pour beaucoup d'autres formes d'investissement, l'apprentissage peut être envisagé comme étant une fonction de croissance du capital. Le *chapitre 4* montre que le retour sur investissement dans l'éducation est

329. Braun et Kanjee, 2006, p. 36.

nettement plus important quand les investissements interviennent à un stade plus précoce de l'apprentissage. Dans le domaine de l'évaluation, les coûts de traitement seront moins élevés (et le retour sur investissement concomitant plus élevé) si la détection des compétences en lecture – et d'autres – intervient tôt que plus tard. Les évaluations hybrides (comme l'EGRA) peuvent être administrées à de jeunes enfants dès la première année d'école³³⁰, bien avant qu'ils ne soient capables de passer des tests écrits en groupe (comme dans les LSEA). C'est un moyen de détecter beaucoup plus tôt les problèmes d'apprentissage au niveau individuel (ainsi qu'au niveau des écoles). Il y a beaucoup de stades auxquels on peut utilement évaluer les compétences des enfants (ou des adultes), mais le bénéfice est plus grand s'il est possible de les mesurer au début d'une longue trajectoire d'apprentissage³³¹.

L'évaluation doit viser l'amélioration de l'enseignement

La tendance qui consiste à évaluer *pour* apprendre³³² se développe ; cela signifie que les enseignants utilisent les résultats de l'évaluation pour aider les enfants à apprendre. Cependant, l'efficacité des LSEA du point de vue pédagogique est limitée par le fait qu'elles interviennent *après* que l'apprentissage a eu lieu et souvent après que l'élève a terminé ses études, autrement dit beaucoup trop tard pour l'aider. Pourtant, de nombreuses données montrent que les évaluations des compétences en lecture, au niveau individuel ou du groupe, peuvent être utilisées pour améliorer l'acquisition de la lecture. Dans les PMA, le manque d'outils localisés et les ressources humaines limitées, comme le faible nombre de spécialistes de la lecture dûment qualifiés, ont rendu cette utilisation difficile. L'apparition des évaluations hybrides des compétences en lecture rend désormais possible

330. Comme il a déjà été mentionné, dans un environnement d'alphabétisation défavorable (PSE), de nombreux enfants de 2^e et 3^e années (et postérieures) peuvent se situer à l'échelon le plus bas des mesures EGRA. Par conséquent, il faut soit trouver des mesures plus simples, soit reporter les évaluations à la 2^e ou 3^e année.

331. En ce qui concerne les enfants scolarisés, l'EGRA et les instruments similaires ont ciblé jusqu'à présent les élèves des 1^{re}, 2^e et 3^e années. La raison en a été décrite auparavant, et elle se justifie. Il y aura inévitablement des pressions pour tester les enfants avant la 1^{re} année de scolarité. En effet, dans certains pays à revenus élevés, des tentatives sont faites pour étudier les compétences antérieures à l'acquisition de la lecture lors des premières étapes (les plus faciles) de l'EGRA, avec notamment plus de tests de langage et de vocabulaire. Cependant, compte tenu des conditions actuelles de l'éducation dans les pays à faibles revenus, la 1^{re} année d'école est probablement le meilleur moment pour lancer le processus d'évaluation.

332. Voir le *chapitre 6* sur la Réponse aux interventions et les évaluations pour l'apprentissage.

l'organisation d'évaluations formatives dans un délai qui permet d'apporter des changements au niveau de la classe (ou au niveau individuel), avant que l'enfant ne quitte le système scolaire. Les informations obtenues grâce à ce nouveau type d'évaluation ne seront pas faciles à exploiter pour former les enseignants, mais elles peuvent les aider à revoir leurs méthodes pédagogiques. À mesure que ce domaine évoluera, il faudra veiller à tirer davantage parti des évaluations formatives pour concevoir l'enseignement et le perfectionnement professionnel des enseignants.

La comparabilité transnationale ne garantit pas toujours une éducation de qualité pour tous

L'un des points forts, souvent cité, des évaluations internationales et régionales est leur capacité de fournir des points de comparaison entre nations, régions et continents, en utilisant les meilleurs outils méthodologiques disponibles pour générer des scores cumulatifs nationaux, dans le cadre de tests internationaux. Comme il a déjà été dit au *chapitre 4*, des évaluations comparatives comme les enquêtes PISA et PIRLS constituent une base substantielle de discussion et de débat au sein des systèmes éducatifs et entre eux. Les comparaisons internationales ont été l'élément déclencheur de débats dans de nombreux pays riches et pauvres, ouvrant ainsi de nouvelles voies dans le domaine de l'élaboration des politiques, de la recherche et de la sensibilisation du grand public. Par ailleurs, les statistiques de l'UNESCO sur l'éducation (même si elles ne s'appuient pas *stricto sensu* sur les LSEA) seraient sans valeur, si la comparabilité n'était pas un but important³³³. Pourtant, les évaluations internationales ont dû faire divers compromis pour parvenir à un consensus transnational, comme limiter l'échantillonnage de population (en excluant les groupes et langues marginalisés)³³⁴. S'ils sont intéressants pour certaines nations, les « classements » des LSEA peuvent se révéler moins utiles pour les PMA qui ont des scores tellement faibles que la comparaison avec les pays de l'OCDE a peu de valeur stratégique.

333. Comme il est mentionné au *chapitre 8* sur l'alphabétisation des adultes, la comparabilité ne suffit évidemment pas, en particulier si l'unité d'analyse est, comme dans le cas de l'alphabétisation, la simple question : « Savez-vous lire ? » Bien que comparable, il s'agit d'une mesure très inadéquate. Toutefois, les statistiques de l'Institut de statistique de l'UNESCO et des LSEA internationales sont au cœur de nombreuses discussions sur l'éducation internationale, comme le montrent les nombreuses utilisations de ces statistiques dans le présent rapport ; merci à C. Guadelupe (communication personnelle, 2010) pour cette observation précieuse.

334. Pour une analyse plus poussée des questions relatives aux populations marginalisées et à l'exclusion, voir UNESCO, 2010.

En d'autres termes, la comparabilité internationale, si elle n'a pour but que de désigner des vainqueurs, pourrait être d'une utilité limitée pour les pays à faibles revenus.

Les évaluations hybrides, au contraire, favorisent deux autres types de comparabilité. *Premièrement*, en se focalisant sur les évaluations au niveau de la classe et du contexte, elles peuvent favoriser une compréhension beaucoup plus nuancée des variables qui sont en jeu au niveau individuel ou de la classe. Ces relations peuvent ensuite être comparées (ou opposées) à d'autres contextes similaires ou différents. *Deuxièmement*, ces évaluations permettent de s'intéresser à des référentiels généraux plutôt qu'aux scores cumulatifs totaux obtenus à un test international. Par exemple, comme il a été mentionné auparavant, les indicateurs récemment préconisés par le PME³³⁵ (fondés globalement sur l'approche EGRA) suggèrent un référentiel axé sur l'école, à savoir la proportion d'élèves qui, après deux années de scolarité, sont capables de « lire pour apprendre ». « Être capable de lire un court texte dans sa langue maternelle » pourrait également servir de référentiel. Divers indicateurs fiables (à forte validité apparente et indirecte) peuvent être dérivés des évaluations hybrides ou inclus dans celles-ci ; ils pourraient aider à surmonter certaines difficultés posées par la comparabilité transnationale des LSEA³³⁶. Quoiqu'il en soit, diverses sortes de comparaisons doivent faire partie de toute évaluation hybride digne de ce nom, comme la comparabilité des élèves d'un échantillon donné, dans un contexte linguistique spécifique et dans le temps (c'est-à-dire lors d'études longitudinales)³³⁷.

En fin de compte, toutes les évaluations cherchent la comparabilité, mais de différentes manières. Les LSEA internationales et régionales visent la comparabilité transnationale, tandis que les évaluations hybrides sont davantage axées sur les contextes locaux et une validité accrue. Les évaluations hybrides offrent des formes de comparabilité que n'offrent pas

335. Voir le *chapitre 6*.

336. Comme il a été évoqué au *chapitre 4*, les résultats d'apprentissage sont extrêmement différents dans le monde, tant entre pays qu'à l'intérieur des pays. Cela amène inévitablement à se demander si les référentiels établis par des évaluations hybrides telles que l'EGRA ou READ INDIA ne sont pas « trop bas » ; les enfants ne pourront soutenir la concurrence, si on ne relève pas les normes au-dessus des niveaux modestes de fluidité de la lecture. Les réponses à cette question sont nombreuses, mais la principale justification est de fournir quelques mécanismes concrets pour aider tous les enfants à atteindre un niveau où ils deviennent capables de lire pour apprendre et c'est ce que ces évaluations s'efforcent de faire.

337. Voir Wagner, 2010.

les LSEA, notamment dans le cas de populations marginalisées ou d'enfants plus jeunes. L'importance des types de comparabilité dépend des objectifs stratégiques visés, ainsi que des considérations de calendrier et de coût. Comme c'est généralement le cas dans l'éducation comparative, le contexte culturel déterminera si et quand les interprétations empiriques sont jugées crédibles³³⁸. Globalement, les évaluations hybrides visent la validité locale plutôt que la comparabilité internationale.

Les « biais » culturels ne nuisent pas forcément à une évaluation

Les évaluations de l'éducation sont souvent accusées de biais culturels, et c'est probablement inévitable. Les évaluations sont destinées à comparer des individus et des groupes d'individus. Les chercheurs en sciences sociales, notamment les spécialistes de l'évaluation, s'efforcent de limiter les biais culturels, en concentrant les efforts sur de meilleures rétrotraductions, en mettant leur veto aux items qui semblent aberrants dans une langue ou une culture donnée, en posant des objectifs communs, etc. Le principal moyen employé pour réduire le parti pris culturel est le compromis, car on ne connaît aucune méthode permettant de le supprimer totalement. C'est pourquoi, si le préjugé culturel est considéré par de nombreux experts comme une mauvaise chose, le degré d'inquiétude qu'il suscite est fonction du cadre de référence choisi.

Il s'ensuit que plus l'échantillon général est large et diversifié, plus il faut faire de compromis. Inversement, si l'objectif est d'étudier les compétences de base en lecture au sein d'un seul groupe ethnolinguistique – comme l'a fait, par exemple, Pratham à Bihar, en Inde (*chapitre 6*) –, malgré une part de diversité inévitable à Bihar, la comparaison des résultats de Bihar avec ceux d'Andhra Pradesh ou de Los Angeles perd de l'importance. L'idée, ici, est que le parti pris est omniprésent. En étant moins exigeant sur la comparabilité (souvent suscitée par des forces extérieures, telles que les agences intergouvernementales), il devient alors moins nécessaire, et donc moins coûteux, de s'employer à supprimer la variation culturelle. Les évaluations hybrides du type SQC ont un avantage relatif dans ce domaine, car elles sont par nature plus adaptables à des contextes spécifiques. En d'autres termes, quand le préjugé est non connu ou non voulu, tout doit être fait pour le réduire. En revanche, si l'on veut axer les efforts sur un contexte

338. Voir Steiner-Khamsi, 2010.

ou un groupe ethnolinguistique particulier, cette forme de parti pris peut alors être souhaitable³³⁹.

Les nouvelles évaluations peuvent également favoriser l’alphabétisation des adultes

L’alphabétisation des adultes a fait l’objet de beaucoup moins de recherche et de financements que l’apprentissage de la lecture à l’école primaire. L’une des conséquences est que les chiffres les plus cités concernant l’alphabétisation des adultes dans les pays en développement (comme indiqué au *chapitre 8*) risquent d’être fortement inexacts³⁴⁰. En outre, l’alphabétisation des adultes se heurte aux mêmes problèmes que ceux que l’on rencontre dans les écoles primaires, notamment des enseignants mal formés, du matériel pédagogique inadapté, des pédagogies ne reposant pas sur des compétences linguistiques spécifiques dans des sociétés multilingues, et une mauvaise conception de l’enseignement. Pour faire progresser l’évaluation et l’enseignement de la lecture, il est important de réduire les barrières entre ceux qui travaillent sur l’apprentissage de la lecture chez les enfants et ceux qui s’occupent des adultes. Une synergie évidente existe quand les parents sont alphabétisés et qu’ils peuvent inciter leurs enfants à suivre leur exemple, surveiller leur travail à l’école et avoir des attentes plus grandes quant à leur réussite future. Comme il y a plus de probabilités pour que les enfants de parents analphabètes aient des problèmes ou des retards d’acquisition de la lecture, il faut trouver de nouvelles manières de renforcer la responsabilisation et l’efficacité des programmes d’alphabétisation des adultes, afin de favoriser une acquisition plus précoce de la lecture chez les enfants.

339. Le biais dans les évaluations n’est pas toujours d’ordre culturel. Par exemple, il y a des paramètres d’échantillonnage de la variation qui révèlent systématiquement les différences entre groupes et contextes. L’une des distinctions importantes décrites dans le présent ouvrage est celle qui existe entre WSE et PSE. Les nombreuses données disponibles montrent que ces environnements d’alphabétisation peuvent varier considérablement d’un pays à l’autre et à l’intérieur des frontières nationales, et qu’ils peuvent inclure des variables au niveau des composantes (présence de livres à la maison, alphabétisme maternel, qualification des professeurs de lecture, etc. Si ces paramètres ne sont pas mesurés correctement, ils peuvent donner lieu à des résultats biaisés.

340. Voir les *Rapports mondiaux de suivi de l’EPT, 2005 et 2010* de l’UNESCO, ainsi que la discussion sur les statistiques sur l’alphabétisation au *chapitre 8*.

La responsabilité de l'impact de l'apprentissage doit être largement partagée

Spécialistes de l'éducation, décideurs, participants aux tables rondes intergouvernementales de haut niveau, ministres de l'Éducation, chefs locaux des villages à la campagne, enseignants et parents doivent partager la responsabilité du contenu enseigné aux enfants et des méthodes pédagogiques employées. Tous sont des consommateurs de connaissances sur l'apprentissage. Jusqu'à aujourd'hui, les spécialistes de l'éducation et les statisticiens étaient, dans la plupart des pays (et en particulier dans les PMA), les principaux gardiens des résultats des évaluations de l'apprentissage. Cet accès limité aux informations sur les acquis d'apprentissage est dû, au moins en partie, à la complexité des évaluations à grande échelle, mais peut-être aussi à la réticence de certains responsables politiques, inquiets du fait de rendre publiques les différences entre les groupes d'enfants évalués (par exemple entre groupes ethnolinguistiques ou entre écoles privées et publiques).

Aujourd'hui, l'importance d'impliquer des acteurs multiples à la prise de décision dans le domaine éducatif est davantage admise. L'intérêt du consommateur pour l'apprentissage des enfants a pris une importance centrale, que ce soit du fait d'une plus grande transparence des gouvernements, de l'influence des agences internationales, des efforts des ONG, de l'activisme accru des communautés ou de la participation des parents. Pour répondre à cet intérêt croissant, il faut produire des données en temps réel, mieux ciblées, compréhensibles, transparentes, et qui répondent aux intérêts locaux des décideurs, des communautés, des parents et même des enfants. L'implication de divers acteurs permettra de mieux faire connaître les atouts et les faiblesses de l'école. Comme il a été mentionné au *chapitre 6* (dans les études de terrain EGRA au Kenya et au Libéria), il a été impossible d'empêcher la participation de la communauté, même quand les enquêteurs espéraient pouvoir faire des comparaisons expérimentales régulières. La « divulgation » de « secrets » professionnels peut être un problème lors de ces expériences scientifiques, mais c'est également le signe d'une plus grande participation des communautés au progrès social.

Ces formes d'échange d'informations à divers niveaux sont une autre façon de parler de responsabilisation et d'attentes. Qui est responsable si un enfant, un enseignant, une école, une région ou une nation n'atteint pas un niveau d'apprentissage donné ? Et d'ailleurs, comment ces attentes sont-elles suscitées ? Lesquelles faut-il satisfaire en priorité ? Les évaluations

SQC, bien qu'elles commencent seulement à apparaître³⁴¹, peuvent ouvrir des perspectives nouvelles, en termes de responsabilisation et d'adhésion locale, en grande partie grâce à leur objectif stratégique clair de fournir en temps voulu des informations pertinentes sur des groupes spécifiques, afin que le changement soit possible, négociable et prévisible.

Les évaluations hybrides peuvent renforcer sensiblement l'impact d'une politique

Les évaluations à grande échelle de l'apprentissage (LSEA), comme les enquêtes PISA, PIRLS, ainsi que SACMEQ, LLECE et PASEC intéressent tout particulièrement les médias et la classe politique, ce qui a parfois donné lieu à des vifs débats et à de grands changements d'orientation de la politique éducative aux niveaux international et national. C'est une force importante que les spécialistes de l'évaluation doivent protéger et sur laquelle ils doivent s'appuyer. Les sociétés du savoir actuelles ont besoin d'études sérieuses et crédibles, qui inspirent confiance aux citoyens et à la presse. Beaucoup de donateurs et d'agences intergouvernementales ou régionales ont financé des LSEA. Toutefois, ces études très influentes ne représentent pas la totalité du spectre des évaluations utiles possibles.

Comme il est avancé ici, l'éventail des objectifs aux niveaux international, national et local implique de nouveaux types d'évaluations hybrides. Ces évaluations peuvent apporter une contribution substantielle à l'histoire d'Aminata, aux populations pauvres et marginalisées, ainsi qu'à ceux qui se trouvent à l'extrémité inférieure du spectre de l'éducation. Les évaluations SQC permettent de mieux suivre l'apprentissage dans le temps, s'adaptent mieux aux contextes linguistiques locaux, et peuvent être conçues pour mieux comprendre les enfants qui sont au bas des échelles d'apprentissage. Ces enfants ne sont pas moins capables d'apprendre, mais rencontrent plus d'obstacles sur leur chemin. Si ces enfants ne sont pas pleinement intégrés dans les méthodes d'évaluation, ils auront peu d'influence sur l'élaboration future des politiques. Pour répondre à ces obstacles et les surmonter, il faut des outils de mesure formatifs (évaluations *pour apprendre*) qui soient aussi sérieux et crédibles que les LSEA, qui respectent les exigences scientifiques de validité et de fiabilité, et dont les buts politiques soient axés sur les besoins locaux, nationaux et internationaux. Nul doute que les évaluations SQC joueront un rôle important dans les

341. Voir Clarke (2010) pour une discussion instructive sur les phases d'émergence, d'établissement et de maturité du développement des systèmes d'évaluation.

politiques de développement de l'éducation au cours des prochaines années, tout simplement parce qu'elles sont plus efficaces dans certains domaines.

En fin de compte, les évaluations éducatives peuvent avoir des objets et des buts très divers. Elles sont par nature complémentaires les unes par rapport aux autres, car chacune recueille des données différentes sur les performances cognitives et le contexte social, et chacune exige des investissements et des catégories de dépenses différents. Ce qu'il faut, c'est clarifier les intentions et proposer des options d'évaluation réellement capables d'induire un changement dans l'éducation et d'améliorer l'apprentissage.

10. CONCLUSIONS

Dans la lutte contre la marginalisation, le suivi et l'évaluation revêtent une importance capitale et devraient être considérés comme faisant partie intégrante des stratégies visant à identifier les groupes sociaux et les régions mises en marge, à leur donner une meilleure visibilité et à déterminer des modes d'intervention politique efficaces. Un suivi actif et la ventilation des données sont également indispensables pour mesurer les progrès accomplis vers les cibles axées sur l'équité. Trop souvent, les enquêtes statistiques nationales ne parviennent pas à rendre compte adéquatement des spécificités des laissés-pour-compte, ce qui accentue la marginalisation de ces groupes. Il est encore plus difficile d'obtenir des données actualisées qui permettraient de mesurer les inégalités dans l'éducation³⁴².

Le présent ouvrage s'ouvre sur une question : *Les études disponibles sur l'évaluation de l'apprentissage (et en particulier sur l'apprentissage de la lecture) peuvent-elles contribuer plus efficacement à l'amélioration des résultats éducatifs dans les pays en développement ?* Il a été largement tenu pour acquis que la réponse était positive. Mais personne n'a indiqué clairement quels types d'évaluation utiliser ni dans quels buts. Jusqu'à une époque assez récente, la plupart des responsables politiques nationaux et internationaux recouraient à des études comparatives internationales extrêmement techniques pour se tenir au courant de la situation. Les responsables locaux, les chefs d'établissement et les enseignants, quant à eux, se fiaient souvent aux examens nationaux de fin de scolarité pour déterminer comment « leurs » élèves apprenaient. Comme le présent ouvrage l'a montré, ces deux approches font l'impasse sur *l'élève en tant qu'apprenant*. Au lieu de s'inquiéter de la *façon* dont un enfant apprend à lire, la plupart des responsables politiques ne connaissent que le score individuel cumulé ou moyen, et ont très peu d'informations sur ce qui a permis d'atteindre ce score, et sur ce qu'il faudrait faire pour améliorer la conception pédagogique. Avec l'apparition des évaluations hybrides SQC, il est désormais possible de mettre en évidence la nature des difficultés rencontrées par les enfants, avec une précision et dans des délais qui permettent éventuellement d'intervenir avant qu'il ne soit trop tard.

Ce passage d'une vision « macro » à une vision « micro » n'est pas anodin. Il est également incomplet. Il reste beaucoup à faire pour inscrire les

342. UNESCO, 2010. *Rapport mondial de suivi sur l'EPT, 2010*, p. 292.

nouvelles approches SQC dans la prise de décision future dans le domaine de l'éducation.

Quelques points non résolus pour l'élaboration de nouvelles évaluations

Le choix des évaluations et leurs résultats susciteront probablement des débats pendant des années, et ces discussions viendront enrichir le domaine de la qualité de l'éducation. À mesure que se développe la base de connaissances sur l'évaluation, les prochaines étapes consisteront probablement à élargir et approfondir l'utilisation des indicateurs, qui est l'une des voies essentielles pour améliorer l'apprentissage, l'enseignement et la scolarité partout dans le monde, en particulier pour les groupes les plus en difficulté dans les pays en développement.

Le présent tour d'horizon a évoqué de nombreux problèmes. Certains semblent pouvoir être réglés en s'appuyant sur des études relativement fiables. Par exemple, les évaluations hybrides permettent d'aborder efficacement les problèmes stratégiques liés à l'apprentissage des compétences de base, avec des méthodes plus ciblées, tenant compte des contraintes temporelles et globalement moins onéreuses. Elles peuvent également être adaptées aux variations linguistiques et orthographiques, sans rien perdre de leur validité et de leur fiabilité. Elles peuvent en outre être conçues pour servir efficacement les intérêts d'un groupe plus large d'intervenants.

En même temps, on commence tout juste à comprendre les évaluations hybrides SQC du point de vue conceptuel, empirique et pratique, en partie parce qu'on sait encore peu de choses sur leur utilisation par rapport aux LSEA. Pour perfectionner les évaluations hybrides et en élargir l'usage, il faut approfondir la recherche sur un certain nombre de questions clés, en s'appuyant sur les expérimentations effectuées sur le terrain dans les PSE des pays à faibles revenus. Ces essais sont notamment :

- a. **Compréhension écrite.** Les éléments d'aptitude ont été jusqu'à présent le sujet principal des études hybrides. On a besoin d'en savoir davantage sur les liens de cause à effet entre ces éléments et la compréhension écrite.
- b. **Études longitudinales.** Pour améliorer le pouvoir prédictif des études hybrides, il est crucial de suivre les élèves de la 1^{re} année jusqu'à la fin du cycle primaire, ou par le biais d'études longitudinales à court terme se recoupant partiellement. C'est particulièrement important pour les études sur les interventions (voir le point suivant).

- c. **Études sur les interventions.** Des recherches ont débuté récemment pour déterminer comment introduire les composantes de la lecture dans les programmes scolaires (voir au *chapitre 5* le projet Liberia Plus). Il faudra davantage d'interventions axées sur un plus grand nombre de variables, pour mieux comprendre la diversité des interventions qui peuvent être efficaces dans différents contextes et pour différentes langues.
- d. **Conception pédagogique.** Le but ultime de la recherche sur l'apprentissage des élèves est d'élaborer de meilleures méthodes pédagogiques, où l'enseignement englobe l'ensemble des facteurs qui ont trait à l'apprentissage (comme la préparation des enseignants, l'élaboration du matériel pédagogique, les compétences préscolaires et le soutien des parents). Il faut que les évaluations hybrides soient capables de guider et d'améliorer la conception pédagogique, sans surcharger les enseignants.
- e. **Tests chronométrés.** Les taux de fluidité de la lecture orale (ORF) peuvent être étudiés en termes de mots par minute, et de nombreux taux ont été considérés comme nécessaires pour la compréhension en lecture. Toutefois, certaines questions ont été soulevées au sujet des conséquences pédagogiques des tests chronométrés dans les PSE. L'approche chronométrée intensive pose-t-elle un problème ? Y a-t-il des alternatives efficaces à la stricte pression temporelle ? Ces questions appellent de plus amples recherches.
- f. **Tests individuels ou tests de groupe ?** À ce jour, presque tous les tests de type EGRA ont été administrés individuellement. Il y a de bonnes raisons à cela, en particulier quand il s'agit de très jeunes enfants. Cependant, compte tenu du coût en temps et en ressources des tests individualisés, des recherches plus poussées doivent étudier les moyens d'obtenir des données d'aussi grande qualité en administrant des tests de groupe, le cas échéant.
- g. **Lecture en L1 et L2.** La majeure partie de la base de connaissances sur l'alphabétisation en L1 et L2 provient d'études menées dans des pays de l'OCDE, en particulier avec l'anglais en L2. La diversité des langues et des systèmes d'écriture en usage dans les PMA devrait fournir des opportunités pour mieux comprendre l'acquisition de la lecture dans les première et seconde langues, ainsi que des manières de rendre la transition entre les deux plus faciles et plus efficaces, à l'oral comme à l'écrit.

- h. **Acquisition de la lecture chez l'enfant et l'adulte.** La plupart des études sur l'apprentissage de la lecture effectuées à ce jour traitent différemment les adultes et les enfants. Une compréhension de l'acquisition de la lecture pendant toute la vie est nécessaire, de même que des évaluations axées sur ce modèle. Les programmes d'enseignement de la lecture pour enfants et adultes auraient tout à gagner d'une meilleure compréhension et d'une plus grande interaction.
- i. **Lecture, mathématiques et autres compétences cognitives.** Des progrès ont été faits ces dernières années dans le domaine des évaluations hybrides de la lecture, et des travaux ont également été engagés concernant les mathématiques³⁴³. Les futures méthodes d'évaluations hybrides administrées au cours du cycle primaire pourraient procurer une valeur ajoutée similaire à celle des évaluations hybrides de la lecture.
- j. **L'EGRA à l'échelle d'un système éducatif.** Quelle doit être l'envergure des données recueillies par les systèmes éducatifs sur les compétences précoces en alphabétisation ? À l'heure actuelle, une évaluation systémique à grande échelle utilisant des outils EGRA ou des outils similaires est administrée uniquement en Inde (par Pratham), la plupart des autres pays participant à l'EGRA ayant privilégié des échantillons plus modestes dans des régions limitées. Il est important de mieux comprendre quels sont les avantages et les inconvénients des évaluations hybrides pour des objectifs éducatifs systémiques plus larges³⁴⁴.
- k. **Diagnostics individualisés.** Dans les pays de l'OCDE, les écoles ont généralement un spécialiste de la lecture qui peut aider les enfants ayant des problèmes d'apprentissage. De tels spécialistes sont rares dans les PSE des pays en développement. Les évaluations hybrides peuvent être un moyen, pour les enseignants ou les assistants, d'apporter aux enfants un soutien spécial à l'apprentissage de la lecture, dont ils ne bénéficieraient pas autrement. La recherche devra clarifier les possibilités d'interventions individualisées.

343. Voir le travail du Research Triangle Institute sur l'EGMA (*Early Grade Math Assessment* ou évaluation précoce des performances en mathématiques). Voir www.eddataglobal.org/documents/index.cfm?fuseaction=showdiretruid=5etstatusID=3 (consulté le 14 mars 2010).

344. Merci à M. Jukes pour cette observation utile.

1. **Psychométrie des évaluations SQC.** Quelles sont les caractéristiques statistiques de la mise en œuvre des évaluations SQC ? De quelle taille doivent être les échantillons d'items et de population pour être statistiquement fiables ? Quelle part dépend de l'échantillon de population concerné et du contexte particulier dans lequel se déroule l'évaluation ? À mesure que l'usage des évaluations hybrides se répand, il devient nécessaire de se pencher plus sérieusement sur ces points, ainsi que sur les problèmes empiriques et psychométriques associés.

Utilisation des technologies de l'information et de la communication

Étant donné la grande complexité d'élaborer des évaluations adéquates, en particulier dans les contextes de grande pauvreté, il peut paraître exagéré de vouloir ajouter une dimension technologique aux aspects à prendre en compte. Pourtant, la technologie change rapidement la vie de tout le monde, en cette époque de mondialisation croissante, et elle doit également être prise en compte dans le présent ouvrage, le cas échéant. Les technologies de l'information et de la communication (TIC) peuvent apporter et apporteront un appui aux recommandations formulées plus tôt, qu'elles concernent l'enseignement multilingue, la collecte de données sur le terrain ou l'utilisation des télécommunications pour une mise en œuvre en temps réel des interventions³⁴⁵.

Selon diverses estimations, seule une petite fraction (moins de 5 %) des investissements dans les TIC réalisés dans le monde privilégie les populations pauvres et peu alphabétisées³⁴⁶. Beaucoup d'initiatives actuelles pour appliquer les TIC à l'éducation (ICT4E), même celles qui ont eu un impact global positif, n'ont pas été suffisamment *axées sur les populations les plus pauvres*. Par exemple, la grande majorité des logiciels/ contenus du Web (essentiellement dans les langues principales comme l'anglais, le chinois, le français, l'espagnol) a peu d'utilité pour les millions de personnes marginalisées pour des raisons d'alphabétisation, de langue ou de culture. Il est de plus en plus évident que des produits multilingues et conviviaux axés sur les TIC peuvent répondre aux besoins des populations pauvres au-delà de ce que l'on avait imaginé jusqu'à présent. Mettre à disposition ces outils et développer les capacités des ressources humaines

345. Pour un aperçu général de l'application des TIC au suivi et à l'évaluation dans le domaine de l'éducation, voir Wagner, 2005 ; à propos de l'appui au travail sur l'alphabétisation, voir Wagner et Kozma, 2005.

346. Wagner et Kozma, 2005.

pour soutenir le développement et la distribution de contenus pertinents au niveau local est un moyen important de contribuer à créer une spirale positive de développement durable.

Comment les évaluations SQC peuvent-elles y contribuer ? Tout d'abord, il existe de nouvelles possibilités de créer des environnements pédagogiques fondés sur les TIC, en tirant parti de ce que nous apprennent les évaluations de la lecture. Pouvons-nous, par exemple, dispenser un enseignement individualisé dans plusieurs langues, qui soit adapté aux niveaux de compétence des enfants dans chaque langue ? Les données d'études menées récemment en Inde et en Afrique du Sud donnent des raisons d'être optimiste³⁴⁷. Deuxièmement, les TIC peuvent également être utilisées pour recueillir des données en recourant aux outils des évaluations hybrides. Correctement appliqués, non seulement ces outils (très probablement implémentés sur des téléphones mobiles) permettront d'obtenir des données plus fiables sur le lieu de collecte, mais aussi de réduire très significativement le temps nécessaire pour le transfert et l'analyse des données, une grande priorité des approches SQC³⁴⁸.

Aller de l'avant

Pour améliorer l'apprentissage, une utilisation efficace des évaluations de l'éducation est fondamentale. Toutefois, la notion d'utilisation efficace ne fait pas référence uniquement aux paramètres techniques de fiabilité et de validité, par exemple. *Ce qui est différent aujourd'hui, c'est que l'on accorde une plus grande priorité aux évaluations à court terme, impliquant divers intervenants, sensibles aux particularités culturelles et ayant un fort impact local.* Les évaluations de l'apprentissage, qu'il s'agisse d'études à grande échelle, d'enquêtes auprès des ménages ou d'évaluations hybrides

347. Dans le cadre d'un projet entrepris dans l'État d'Andhra Pradesh (Inde), Wagner (2009b) et Wagner *et al.* (2010) ont constaté les résultats positifs d'une intervention à court terme utilisant le multimédia dans la langue locale (le telugu) pour aider à apprendre à lire et à écrire à des enfants à l'école primaire et à des jeunes non scolarisés, grâce à des outils d'évaluation du type SQC. Cette étude a également montré la capacité des TIC à soutenir les environnements d'apprentissage multilingues pour les populations très pauvres, peu ou pas familiarisées avec les ordinateurs. En anticipant sur l'avenir, l'importante progression des évaluations SQC s'appuiera probablement sur les TIC (en particulier les terminaux mobiles) pour fournir des données encore plus crédibles dans des délais plus courts.

348. L'Institut international de l'alphabétisation vient de mettre au point un outil de ce type, sur la base d'un système d'exploitation Android et destiné à recueillir des données EGRA sur le terrain.

(SQC – simples, rapides et abordables), ne valent que par les usages que l'on peut en faire. La plupart d'entre elles sont en permanence améliorées et perfectionnées.

La mondialisation et les efforts des pays les plus riches pour se disputer la dernière série de « compétences globales » se poursuivront, et il ne fait aucun doute que ces pays en tireront profit en fonction de leurs investissements. Mais la mondialisation des évaluations, si celles-ci sont définies exclusivement autour des compétences utilisées dans les nations industrialisées, maintiendra nécessairement les enfants les plus pauvres au bas de n'importe quelle échelle mondiale, de sorte qu'il sera difficile de comprendre les facteurs qui conduisent les enfants pauvres à rester mal servis. Les initiatives actuelles pour élargir l'administration des évaluations dans les pays en développement deviendront une composante importante des solutions visant à améliorer la qualité de l'éducation.

Les évaluations de l'apprentissage peuvent ouvrir de nouvelles voies à la responsabilité pédagogique, principalement en fixant comme but stratégique de produire en temps voulu des informations pertinentes pour des groupes spécifiques, de sorte que le changement devienne prévisible. Les efforts actuels pour élargir l'administration des évaluations dans les pays en développement renforceront la responsabilisation, et c'est le seul moyen de rendre possible un changement réel et durable. Enfin, et surtout, il faut que les politiques et les évaluations les plus importantes restent durablement focalisées sur les populations pauvres et marginalisées, celles qui sont la cible principale des OMD et de l'EPT. Il s'agit d'enfants pauvres vivant dans des milieux pauvres. Ceux qui se trouvent au bas de l'échelle de l'apprentissage.

L'histoire d'Aminata revisitée

Deux ans ont passé. Aminata a maintenant 11 ans et continue d'aller à l'école. Son instituteur, Monsieur Mamadou, a pu suivre une formation sur les moyens d'améliorer son école et d'aider ses élèves à mieux lire – une première pour l'école. Au lieu d'être tenue à l'écart de l'enseignement et de l'apprentissage, Aminata a été interrogée en classe, au même titre que tous les enfants, même ceux des derniers rangs. Il règne aujourd'hui un esprit communautaire, surtout depuis que Monsieur Mamadou a déclaré que si tous les enfants apprennent à lire, l'école remporterait un prix. Aminata n'y a pas vraiment pensé, jusqu'à ce que toute la classe reçoive de nouveaux manuels, un pour chaque enfant. Chaque manuel est écrit dans la langue maternelle de l'enfant, la dernière partie étant en français. Le manuel comporte des

dessins colorés et beaucoup d'exercices amusants pour apprendre à prononcer les lettres, les syllabes et les mots. Aminata s'entraîne à faire ces exercices en dehors de la classe avec son cousin et a fini par découvrir le code, ce qui lui permet désormais de lire beaucoup de mots. Elle peut aussi aider sa mère à comprendre ses ordonnances, ainsi que sa sœur, qui vient d'entrer à l'école en 1^{re} année.

Cette histoire revisitée peut-elle se produire ? Pour nombre de ceux qui travaillent dans le domaine de l'éducation internationale, un changement aussi radical paraît peu probable. Par ailleurs, il semble difficile d'inclure un tel récit dans une étude sur les aspects techniques des évaluations de l'apprentissage. Néanmoins, c'est probablement le seul moyen pour que cette histoire se concrétise. L'histoire d'Aminata ne sera pas modifiée par une volonté accrue, des enseignants plus gentils, des bienfaiteurs étrangers, plus de manuels, un meilleur éclairage et davantage de toilettes, bien que tous ces facteurs puissent jouer, et jouent effectivement, un rôle dans l'apprentissage et la scolarité. Seules une révision des attentes et la responsabilisation concomitante des multiples acteurs concernés permettront à l'histoire d'Aminata revisitée de devenir réalité.

RÉFÉRENCES

- Abadzi, H. 2003. *Adult Literacy: A Review of Implementation Experience*. Operations Evaluation Department, World Bank. Washington, DC : Banque mondiale.
- . 2006. ‘Adult Illiteracy, Brain Architecture, and Empowerment of the Poor.’ *Adult Education and Development*, 65, 19-34.
- . 2008. ‘Efficient Learning for the Poor: New Insight into Literacy Acquisition for Children’. *International Review of Education*, 54, 5, 581-605.
- . 2010. *Reading Fluency Measurements in FTI Countries: Outcomes and Improvement Prospects*. Projet. Série des documents de travail. Washington, DC : PME.
- Abadzi, H. ; Crouch, L. ; Echegaray, M. ; Pasco, C. ; Sampe, J. 2005. ‘Monitoring Basic Skills Acquisition through Rapid Learning Assessments: A Case Study from Peru.’ *Prospects*, Vol. 35, 2, 137-156.
- Abdul Latif Jameel Poverty Action Lab (J-PAL) ; Pratham ; ASER. 2009. *Evaluating READ INDIA: The Development of Tools for Assessing Hindi Reading and Writing Ability and Math Skills of Rural Indian Children in Grades 1-5*. Non publié. Projet. Chennai, Inde : J-PAL.
- Adams, M.J. 1990. *Beginning to Read: Thinking and Learning about Print*. Cambridge, MA : MIT Press.
- Afflerbach, P. ; Pearson, P.D. ; Paris, S.G. 2008. ‘Clarifying differences between Reading Skills and Reading Strategies.’ *The Reading Teacher*, 61, 5, 364-373.
- Alegria, J. ; Mousty, P. 1996. ‘The Development of Spelling Procedures in French-Speaking, Normal and Reading-Disabled Children: Effects of Frequency and Lexicality.’ *Journal of Experimental Child Psychology*, 63(2), 312-338.
- Alidou, H. ; Boly, A. ; Brock-Utne, B. ; Diallo, Y.S. ; Heugh, K. ; Wolff, H.E. 2006. *Optimising Learning and Education in Africa: The Language Factor*. Paris : ADEA, GTZ, Commonwealth Secretariat.

Références

- Allington, R.L. 1983. 'The Reading Instruction Provided Readers of Differing Reading Abilities.' *The Elementary School Journal*, 83, 548-559.
- Altinok, N. 2009. *A Technical Analysis of Skills and Tests of Performance in the Context of the Quality Learning Indicators Project*. Document de travail informatif non publié. Paris : Project QLIP.
- Arnové, R.F. ; Graff, H.J. (dir. publ.). 1987. *National Literacy Campaigns: Historical and Comparative Perspectives*. New York : Plenum.
- ASER. 2009. *Evaluating the Reliability and Validity of the ASER Testing Tools*. Projet non publié. New Delhi : (www.asercentre.org).
- . 2010. *Enrollment and Learning Report Card: India Rural*. New Delhi (www.asercentre.org).
- August, D. ; Shanahan, T. (dir. publ.) 2006. *Developing Literacy in Second Language Learners. Report of the National Reading Panel on Language Minority and Youth*. Mahwah (NJ) : Lawrence Erlbaum Associates.
- Babson, A. 2009. *Costs of Large Scale Educational Assessments*. Document de travail informatif non publié. Paris : Project QLIP.
- Banerji, R. ; Wadhwa, W. 2006. 'Between Outlays and Outcomes.' *Times of India*. Consulté le 28 février 2011 : http://images2.asercentre.org/Resources/Articles/Between_outlays_-_outcomes.pdf.
- Banque mondiale. 2011. Why Invest in Early Child Development (ECD). Web-based document. Consulté le 10 juin 2011 : <http://web.worldbank.org/WEBSITE/EXTERNAL/TOPICS/EXTCY/EXTECD/0,,contentMDK:20207747~menuPK:527098~pagePK:148956~piPK:216618~theSitePK:344939,00.html>
- Bear, D.R. ; Invernizzi, M. ; Templeton, S. ; Johnston, F. 2004. *Words their Way*. Upper Saddle River, NJ : Pearson.
- Beck, I.L. ; McKeown, M.G. ; Kucan, L. 2002. *Bringing Words to Life: Robust Vocabulary Instruction*. New York : Guilford.
- Benavot, A. ; Tanner, E. 2007. *The Growth of National Learning Assessments in the World, 1995-2006*. Document de référence préparé en vue du Rapport mondial de suivi sur l'EPT 2008. Paris : UNESCO.

- Bernhardt, E. 2005. 'Progress and Procrastination in Second Language Reading.' *Annual Review of Applied Linguistics*, 25, 133-150.
- Berninger, V.W. ; Abbott, R.D. ; Nagy, W ; Carlisle, J. 2010. 'Growth in Phonological, Orthographic, and Morphological Awareness in Grades 1 to 6.' *Journal of Psycholinguistic Research*, 39,141-163.
- Berninger, V.W. ; Abbott, R.D. ; Trivedi, P. ; Olson, E. ; Gould, L. ; Hiramatsu, S. ; Holsinger, M. ; McShane, M. ; Murphy, H. ; Norton, J. ; Boyd, A.S. ; Westhaggen, S.Y. 2010b. 'Applying the Multiple Dimensions of Reading Fluency to Assessment and Instruction.' *Journal of Psychoeducational Assessment*, 28, 3-18.
- Bettinger, E. 2006. 'Evaluating Educational Interventions in Developing Countries. Using Assessment to Improve Education in Developing Nations.' Dans : Braun *et al.*, (dir. publ.), *Improving education through assessment, innovation and evaluation*, p. 1-46. Cambridge, MA : American Academy of Arts and Sciences.
- Bialystok, E. ; Luk G. ; Kwan, E. 2005. 'Bilingualism, Biliteracy, and Learning to Read: Interactions among Languages and Writing Systems.' *Scientific Studies of Reading*, 9(1), 43-61.
- Black, P. ; Wiliam, D. 1998. 'Assessment and Classroom Learning.' *Assessment in Education*, 5(1), 7-74.
- Blaiklock, K.E. 2004. 'The Importance of Letter Knowledge in the Relationship between Phonological Awareness and Reading.' *Journal of Research in Reading*, 27(1), 36-57.
- Blum, A. ; Goldstein, H. ; Guérin-Pace, F. 2001. 'International Adult Literacy Survey (IALS): An Analysis of Adult Literacy', *Assessment in Education*, Vol. 8, n° 2, p. 225-246.
- Bradley, L. ; Bryant, P.E. 1983. 'Categorizing Sounds and Learning to Read: A Causal Connection.' *Nature*, 301(5899), 419-421.
- Braun, H. ; Kanjee, A. 2006. 'Using Assessment to Improve Education in Developing Nations.' Dans : Cohen, J.E., Bloom, D.E., Malin, M. (dir. publ.). *Improving Education through Assessment, Innovation, and Evaluation*. Cambridge, MA : American Academy of Arts and Sciences, p. 1-46.
- Brislin, R.W. ; Lonner, W.J. ; Thorndike, R.M. 1973. *Cross-cultural Research Methods*. NY : J. Wiley.

- Carey, S. (dir. publ.). 2000. *Measuring Adult Literacy – the International Adult Literacy Survey in the European Context*. Londres : Office for National Statistics.
- Carneiro, P. ; Heckman, J.J. 2003. *Human Capital Policy*. Cambridge, MA : Document de travail 9495 du NBER.
- Carr-Hill, R. 2008. *International Literacy Statistics: A Review of Concepts, Methodology and Current Data*. Montréal : Institut de statistique de l'UNESCO.
- Carroll, L. 1865/2006. *Alice in Wonderland*. NY : Firefly Books.
- Carron, G. ; Mwiria, K. ; Righa, G. 1989. *The Functioning and Effects of the Kenyan Literacy Program. Rapport de recherche de l'IIEP, n° 76*. Paris : IIEP-UNESCO.
- Castles, A. ; Coltheart, M. 2004. 'Is There a Causal Link from Phonological Awareness to Success in Learning to Read?' *Cognition*, 91, 77-111.
- Chabbott, C. 2006. *Accelerating Early Grades Reading in High Priority EFA Countries: A Desk Review*.
www.equip123.net/docs/E1EGRinEFACountriesDeskStudy.pdf
- Chall, J.S. 1967. *Learning to Read: The Great Debate*. New York : McGraw Hill.
- . 1996. *Stages of Reading Development, 2nd edition*. Orlando, FL : Harcourt Brace & Company.
- Chapman, D.W. ; Snyder, C.W. 2000. 'Can High Stakes National Testing Improve Instruction: Reexamining Conventional Wisdom.' *International Journal of Educational Development*, 20, 457-474.
- Chinapah, V. 2003. *Monitoring Learning Achievement (MLA) Project in Africa*. Association for the Development of Education in Africa (ADEA). Document de travail. Paris : ADEA.
- Chowdhury, A.M.R. ; Zieghan, L. 1994. 'Assessing Basic Competencies: A Practical Methodology.' *International Review of Education*, 40, 437-454.
- Chromy, J.R. 2002. 'Sampling Issues in Design, Conduct, and Interpretation of International Comparative Studies of School Achievement.' Dans : A.C. Porter, A. Gamoran (dir. publ.), *Methodological advances in cross-national surveys of educational achievement*, p. 80-116. Washington, DC : The National Academies Press.

- Clarke, M. 2010. *Roadmap for Building an Effective Assessment System*. Projet non publié. Washington, DC : Banque mondiale.
- Clay, M.M. 1991. *Becoming Literate: The Construction of Inner Control*. Auckland, NZ : Heinemann.
- . 2000. *Concepts about Print: What Have Children Learned About The Way We Print Language?* Portsmouth, NH : Heinemann.
- Colón, E. ; Kranzler, J.H. 2006. ‘Effect of Instructions on Curriculum-Based Measurement of Reading.’ *Journal of Psychoeducational Assessment*, 24, 318-328.
- Comings, J. 1995. ‘Literacy Skill Retention in Adult Students in Developing Countries.’ *International Journal of Educational Development*, 15, 37-46.
- Commeyras, M. ; Chilisa, B. 2001. ‘Assessing Botswana’s First National Survey on Literacy with Wagner’s Proposed Schema for Surveying Literacy in the ‘Third World’.’ *International Journal of Educational Development*, 21, 433-446.
- Commeyras, M. ; Inyega, H.N. 2007. ‘An Integrative Review of Teaching Reading in Kenyan Primary Schools.’ *Reading Research Quarterly*, Vol. 42, n° 2, 258-281.
- CONFEMEN. 2008. *Vers la scolarisation universelle de qualité pour 2015. Evaluation diagnostique*. GABON. Programme d’analyse des systèmes éducatifs de la CONFEMEN (PASEC). Dakar : CONFEMEN.
- Coombs, P.H. ; Hallak, J. 1972. *Managing Educational Costs*. New York : Oxford University Press.
- Crosson, A.C. ; Lesaux, N.K. ; Martiniello, M. 2008. ‘Factors that Influence Comprehension of Connectives Among Language Minority Children from Spanish-Speaking Backgrounds.’ *Applied Psycholinguistics*, 29, 603-625.
- Crosson, A.C. ; Lesaux, N.K. 2010. ‘Revisiting Assumptions about the Relationship of Fluent Reading to Comprehension: Spanish-Speakers’ Text-Reading Fluency in English.’ *Reading and Writing*, 23, 475-494.
- Crouch, L. 2009. *Literacy, Quality Education, and Socioeconomic Development*. Présentation Powerpoint, Washington, DC : USAID.

Références

- Crouch, L. ; Korda, M. ; Mumo, D. 2009. *Improvements in Reading Skills in Kenya: An Experiment in the Malindi District*. Rapport préparé pour l'USAID. Research Triangle Institute/Aga Khan Foundation.
- Crouch, L. ; Winkler, D. 2008. *Governance, Management and Financing of Education for All: Basic Frameworks and Case Studies*. Document préparé en vue du *Rapport mondial de suivi sur l'EPT 2009*, intitulé ' *Vaincre l'inégalité : l'importance de la gouvernance* '. Paris : UNESCO.
- Cuetos, F. ; Suarez-Coalla, P. 2009. 'From Grapheme to Word in Reading Acquisition in Spanish.' *Applied Psycholinguistics*, 30, 583-601.
- Cummins, J. ; Swain, M. ; Nakajima, K. ; Handscombe, J. ; Green, D. ; Tran, C. 1984. 'Linguistic Interdependence among Japanese and Vietnamese Immigrant Students.' Dans : *Communicative competence approaches to language proficiency assessment: Research and application*, C. Rivera (dir. publ.), 60-81. Clevedon, Royaume-Uni : Multilingual Matters.
- Deno, S.L. ; Mirkin, P. ; Chiang, B. 1982. Identifying valid measures of reading. *Exceptional Children*, 49, 36-45.
- DeStefano, J. ; Elaheebocus, N. 2009. *School Effectiveness in Woliso, Ethiopia: Measuring Opportunity to Learn and Early Grade Reading Fluency*. Projet non publié. Save The Children.
- DFID (Department for International Development). 2011. *National and International Assessments of Student Achievement*. Note d'orientation. Londres : DFID.
- Dickes, P. ; Vrignaud, P. 1995. *Rapport sur les traitements des données françaises de l'enquête internationale sur la littéracie*. Rapport pour le ministère de l'Éducation nationale. Direction de l'Évaluation et de la Prospective.
- Dickinson, D.K. ; McCabe, A. ; Anastasopoulos, L. 2003. 'The Comprehensive Language Approach to Early Literacy: The Interrelationships among Vocabulary, Phonological Sensitivity, and Print Knowledge among Preschool-Aged Children.' *Journal of Educational Psychology*, 95(3), 465-481.
- Dowd, A.J. ; Wiener, K. ; Mabeti, F. 2010. *Malawi Literacy Boost. Annual Report, 2009*. Westport, CT : Save the Children.
- Downing, J. 1973. *Comparative Reading*. New York : Macmillan.

- Droop, M. ; Verhoeven, L. 1998. Background Knowledge, Linguistic Complexity, and Second-Language Reading Comprehension. *Journal of Literacy Research*, 30, 253-271.
- Dubeck, M.M. ; Jukes, M.C.H. ; Okello, G. 2010. *Early Primary Literacy Instruction in Kenya*. Manuscrit non publié. Cambridge, MA : Harvard University, Graduate School of Education.
- Dumont, H. ; Istance, D. ; Benavides, F. (dir. publ.) 2010. *The Nature of Learning. Using Research to Inspire Practice*. Paris : OCDE.
- Durgunolu, A.Y. ; Öney, B. 2002. 'Phonological Awareness in Literacy Acquisition: It's Not Only for Children.' *Scientific Studies of Reading*, 6, 245-266.
- Easton, P. 2010. *Defining Literate Environments*. Manuscrit non publié. Tallahassee : Florida State University.
- Eden, G.F. ; Moats, L. 2002. 'The Role of Neuroscience in the Remediation of Students with Dyslexia.' *Nature Neuroscience*, 5, 1080-1084.
- Ehri, L. 2005. 'Learning to Read Words: Theory, Findings, and Issues.' *Scientific Studies of Reading*, 9(2), 167-188.
- Elley, W.1992. *How in the World Do Students Read?* The International Association for the Evaluation of Educational Achievement. La Haye : IEA.
- Encinas-Martin, M. 2008. *Overview of Approaches to Understanding, Assessing and Improving the Quality of Learning for All*. Paris : UNESCO.
- Feng, G. ; Miller, K. ; Shu, H. ; Zhang, H. 2009. 'Orthography and the Development of Reading Processes: An Eye-Movement Study of Chinese and English.' *Child Development*, Vol. 80, n° 3, p. 736-749.
- Filmer, D. ; Hasan, A. ; Pritchett, L. 2006. *A Millenium Learning Goal: Measuring Real Progress in Education*. Document de travail n° 97. Washington, DC : Center for Global Development.
- Fuchs, L.S. ; Fuchs, D. 1999. 'Monitoring Student Progress toward the Development of Reading. Competence: A Review of Three Forms of Classroom-Based Assessment.' *School Psychology Review*, 28, 659-671.

- Fuchs, L.S. ; Fuchs, D. ; Eaton, S. ; Hamlett, C.L. 2000. 'Relation between Reading Fluency and Reading Comprehension as a Function of Silent versus Oral Reading Mode.' Données non publiées. Nashville : Vanderbilt University.
- Fuchs, L. ; Fuchs, D. ; Hosp, M.K. ; Jenkins, J. 2001. 'Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis.' *Scientific Studies of Reading*, 5(3), 239-256.
- Gameron, A. ; Long, D.A. 2006. *Equality of Educational Opportunity: A 40-Year Retrospective*. Document de travail n° 2006-9 du WCER. Madison, WI : WCER.
- Genesee, F. ; Geva, E. ; Dresler, C. ; Kamil, M.L. 2006. 'Synthesis: Cross Linguistic Relationships.' Dans : D. August, T. Shanahan (dir. publ.) *Developing literacy in second language learners. Report of the National Reading Panel on language minority and youth*, p. 153-173. Lawrence Erlbaum associates.
- Georgiou, G.K. ; Parrila, R. ; Papadopoulos, T.C. 2008. 'Predictors of Word Decoding and Reading Fluency across Languages Varying in Orthographic Consistency.' *Journal of Educational Psychology*, Vol. 100(3), 566-580.
- Geva, E. ; Siegel, L.S. 2000. 'Orthographic and Cognitive Factors in the Concurrent Development of Basic Reading Skills in Two Languages.' *Reading and Writing: An Interdisciplinary Journal*, 12, 1-30.
- Gilmore, A. 2005. *The Impact of PIRLS (2001) and TIMMS (2003) in Low- and Middle-Income Countries: An Evaluation of the Value of World Bank Support for International Surveys of Reading Literacy (PIRLS) and Mathematics and Science (TIMSS)*. Nouvelle-Zélande : IEA.
- Goldstein, H. 2004. 'International Comparisons of Student Attainment: Some Issues Arising from the PISA Study.' *Assessment in Education*, 11, 3, 319-330.
- Goldstein, H. ; Bonnet, G. ; Rocher, T. 2007. 'Multilevel Structural Equation Models for the Analysis of Comparative Data on Educational Performance.' *Journal of Educational and behavioral Statistics*, 32, 3, 252-286.
- Good, R.H. ; Kaminski, R.A. (dir. publ.). 2002. *Dynamic Indicators of Basic Early Literacy Skills* (6^e édition.). Eugene, OR : Institute for the Development of Educational Achievement.

- Good, R.H. ; Simmons, D.C. ; Kame'enui, E.J. 2001. 'The Importance and Decision-Making Utility of a Continuum of Fluency-Based Indicators of Foundational Reading Skills for Third Grade High-Stakes Outcomes.' *Scientific Study of Reading*, 5, 257-288.
- Goodman, Y.M. ; Burke, C.L. 1972. *Reading Miscue Inventory*. New York : Robert C. Owen.
- Gove, A. 2010. *Early Grade Reading Assessments: Evolution and Implementation to Date*. Réunions annuelles de la Société d'éducation comparative internationale, Chicago, mars.
- Gove, A. ; Cvelich, P. 2010. *Early Reading: Igniting Education for All*. A report by the Early Grade Learning Community of Practice. Washington, DC : RTI.
- Greaney, V. ; Khandker, S.R. ; Alam, M. 1999. *Bangladesh: Assessing Basic Learning Skills*. Washington, DC/Dhaka : Banque mondiale.
- Greaney, V. ; Kellaghan, T. 1996. *Monitoring the Learning Outcomes of Education Systems*. Washington, DC : Banque mondiale.
- . 2008. *Assessing National Achievement Levels in Education*. Dans la série National Assessments of Educational Achievement, Vol. 1. Washington, DC : Banque mondiale.
- Greenberg, D. ; Ehri, L.C. ; Perin, D. 2002. 'Do Adult Literacy Students Make the Same Word-Reading and Spelling Errors as Children Matched for Word-Reading Age?' *Scientific Studies of Reading*, 6, 221-243.
- Grin, F. 2005. 'The Economics of Language Policy Implementation: Identifying and Measuring Costs.' Dans : *Mother Tongue-Based Bilingual Education in Southern Africa: the Dynamics of Implementation*. Actes d'un Colloque tenu à l'Université du Cap, 16-19 octobre 2003. Neville Alexander (dir. publ.). Le Cap : Volkswagen Foundation & PRAESA.
- Hambleton, R.K. ; Kanjee, A. 1995. 'Increasing the Validity of Cross-Cultural Assessments: Use of Improved Methods for Test Adaptation.' *European Journal of Psychological Assessment*, Vol. 11, n° 3, 147-157.
- Hambleton, R.K. ; Swaminathan, R. ; Rogers, H.J. 1991. *Fundamentals of Item Response Theory*. Newbury Park, CA : Sage.

Références

- Hanushek, E. ; Woessmann, L. 2009a. 'Poor Student Learning Explains the Latin American Growth Puzzle.' *VOX: Research-based policy analysis and commentary from leading economists*. Consulté le 21 novembre 2009 :
www.voxeu.org/index.php?q=node/3869
- Hanushek, E. ; Woessmann, L. 2009b. *Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation*. Document de travail 14633 du National Bureau of Economic Research. Washington, DC : NBER.
- Harris, M. 1976. 'History and Significance of the Emic/Etic Distinction.' *Annual Review of Anthropology*, 5, 329-350.
- Hart, B. ; Risley, T.R. 2003. 'The Early Catastrophe: The 30 Million Word Gap by Age 3.' *American Educator*, Printemps, 4-9.
- Hartley, M.J. ; Swanson, E.V. 1986. *Retention of Basic Skills among Dropouts from Egyptian Primary Schools* (Education and Training Series, Rapport n° EDT40). Washington, DC : Banque mondiale.
- Hasbrouck, J. ; Tindal, G.A. 2006. Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers. *The Reading Teacher*, 59, 636-644.
- Heath, S.B. 1982. 'What No Bedtime Story Means: Narrative Skills at Home and School.' *Language and Society*, 11, 49-76.
- Heckman, J.J. 2006. 'Skill Formation and the Economics of Investing in Disadvantaged Children.' *Science*, Vol. 312, n° 5782, p. 1900-1902.
- Heugh, K. 2006b. 'Cost Implications of the Provision of Mother Tongue and Strong Bilingual Models of Education in Africa.' Dans : H. Alidou, A. Boly, B. Brock-Utne, Y. Diallo, K. Heugh, H. Wolff . 2006. *Optimizing Learning and Education in Africa –The Language Factor: A Stock-Taking Research on Mother Tongue and Bilingual Education in Sub-Saharan Africa*. Paris : IPE-ADEA.
- Heyneman, S.P. ; Loxley, W.A. 1983. 'The Effect of Primary-School Quality on Academic Achievement across Twenty-Nine High- and Low-Income Countries.' *American Journal of Sociology*, Vol. 88, n° 6, p. 1162-1194.
- Hirsh-Pasek, K. ; Bruer, J.T. 2007. 'The Brain/Education Barrier.' *Science*, Vol. 317, 5843, p. 1293.

- Hornberger, N.H. 2003 (dir. publ.). *Continua of Biliteracy: An Ecological Framework for Educational Policy, Research and Practice in Multilingual Settings*. Clevedon, Royaume-Uni : Multilingual Matters.
- Howie, S. ; Hughes, C. 2000. 'South Africa.' Dans : D. Robitaille, A. Beaton, Plomb, T. (dir. publ.). *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science*, p. 139-145. Vancouver, BC : Pacific Educational Press.
- Hoxby, C. 2002. *The Cost of Accountability*. Document de travail 88553. Cambridge, MA. : National Board of Economic Research.
- Hruby, G.G. ; Hynd, G.W. 2006. 'Decoding Shaywitz: The Modular Brain and its Discontents.' *Reading Research Quarterly*, Vol. 41, n° 4, 544-556.
- Hudson, R.F. ; Pullen, P.C. ; Lane, H.B. ; Torgesen, J.K. 2009. 'The Complex Nature of Reading Fluency: A Multidimensional View.' *Reading & Writing Quarterly*, 25, 1, 4-32.
- ILI/UNESCO. 1998. *Literacy Assessment for Out-of-school Youth and Adults*. (ILI/UNESCO Technical Report from Expert Seminar, Paris, juin 1998). Philadelphie : Institut international d'alphabétisation. Université de Pennsylvanie.
- . 1999. *Assessing Basic Learning Competencies in Youth and Adults in Developing Countries: Analytic Survey Framework and Implementation Guidelines*. Rapport technique ILI/UNESCO. Philadelphie : Institut international d'alphabétisation. Université de Pennsylvanie.
- . 2002a. *Towards Guidelines for the Improvement of Literacy Assessment in Developing Countries: Conceptual Dimensions Based on the LAP Project*. Philadelphie : Institut international d'alphabétisation. Université de Pennsylvanie.
- . 2002b. *Analytic Review of Four LAP Country Case Studies*. Philadelphie : Institut international d'alphabétisation. Université de Pennsylvanie.
- Ilon, L. 1992. *A Framework for Costing Tests in Third World Countries*. PHREE/92/65. Washington, DC : Banque mondiale.

- . 1996. 'Considerations for Costing National Assessments'. Dans : P. Murphy *et al.* (dir. publ.) *National Assessments: Testing the System*, p. 69-88, Washington, DC : Banque mondiale.
- International Institute of African Languages and Cultures (IILAC). 1930. *Practical Orthography of African Languages*. Londres : Oxford University Press.
- ISU (Institut de statistique de l'UNESCO). 2009. *The Next Generation of Literacy Statistics: Implementing the Literacy Assessment and Monitoring Programme (LAMP)*. Rapport technique n° 1. Montréal : ISU.
- Jesson, D. ; Mayston, D. ; Smith, P. 1987. 'Performance Assessment in the Education Sector: Educational and Economic Perspectives.' *Oxford Review of Education*, 13(3), 249-266.
- Jarousse, J.P. ; Mingat, A. 1993. *L'école primaire en Afrique*. Paris : L'Harmattan.
- Johansson, E. 1987. 'Literacy Campaigns in Sweden.' Dans : R.F. Arnone, H.J., Graff (dir. publ.). *National Literacy Campaigns*. New York : Plenum.
- Johnson, S. 1999. 'International Association for the Evaluation of Educational Achievement Science Assessment in Developing Countries.' *Assessment in Education* 6(1), 57-73.
- Juel, C. ; Griffith, P.L. ; Gough, P.B. 1986. 'Acquisition of Literacy: A Longitudinal Study of Children in First and Second Grade.' *Journal of Educational Psychology*, 78, 243-255.
- Jukes, M.C.H. ; Vagh, S.B. ; Kim, Y.S. 2006. *Development of Assessments of Reading Ability and Classroom Behavior*. Rapport non publié. Cambridge : Harvard Graduate School of Education.
- Jukes, M.C.H. ; Grigorenko, E.L. 2010. 'Assessment of Cognitive Abilities in Multiethnic Countries: The Case of the Wolof and Mandinka in the Gambia.' *British Journal of Educational Psychology*, 80, 77-97.
- Justice, L. 'Evidence-based Practice, Response to Intervention, and the Prevention of Reading Difficulties.' *Language, Speech and Hearing Services in Schools*, 37, 284-297.
- Kagan, J. 2008. 'In Defense of Qualitative Changes in Development.' *Child Development*, 79, 1606-1624.

- Kalton, G. ; Lyberg, L. ; Rempp, J.-M. 1998. Review of Methodology. Dans : T.S. Murray, I.S. Kirsch, L.B. Jenkins (dir. publ.), *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey. Report NCES 98-053*. Washington, DC : Ministère de l'Éducation américain (Annexe A).
- Kame'enui, E.J. ; Fuchs, L. ; Francis, D.J. ; Good, R.H. ; III ; O'Connor, R.E. ; Simmons, D.C. *et al.* 2006. The Adequacy of Tools for Assessing Reading Competence: A Framework and Review. *Educational Researcher*, 35(4), 3-11.
- Kamens, D.H. ; McNeely, C.L. 2010. Globalization and the Growth of International Educational Testing and National Assessment. *Comparative Education Review*, 54, 1, p. 5-25.
- Kanjee, A. 2009. *Assessment Overview*. Présentation à la première Conférence mondiale READ. Moscou, octobre 2009.
- Kellaghan, T. ; Bethell, G. ; Ross, J. 2011. *National and International Assessments of Student Achievement*. Note d'orientation : Document sur les meilleures pratiques du DFID. Londres : DFID.
- Kellaghan, T. ; Greaney, V. 2001. *L'évaluation pour améliorer la qualité de l'enseignement*. Principes de la planification de l'éducation n° 71. Paris : IIPÉ-UNESCO.
- Kellaghan, T. ; Greaney, V. ; Murray, T.S. 2009. *National Assessments of Educational Achievement, Volume 5: Using the Results of a National Assessment*. Washington, DC : Banque mondiale.
- Khachan, V.A. 2009. 'Diglossic Needs of Illiterate Adult Women in Egypt: A Needs Assessment.' *International Journal of Lifelong Education*, 28, 5, 649-660.
- Kim, Y.-S. ; Vagh, S.B. ; Jukes, M. 2008. *The Relationship between Fluency in Various Reading Sublevels and Reading Comprehension in the Transparent Orthography of Swahili*. Manuscrit non publié. Tallahassee, FL : Florida State University.
- Kirsch, I.S. ; Jungeblut, A. ; Jenkins, L. ; Kolstad, A. 1993. *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey*. Washington, DC : Centre national de statistiques éducatives, Ministère de l'Éducation américain.
- Koda, K. ; Reddy, P. 2008. 'Cross-linguistic Transfer in Second Language Reading.' *Language Teacher*, 41, 4, 497-508.

Références

- Kudo, I. ; Bazan, J. 2009. *Measuring Beginner Reading Skills. An Empirical Evaluation of Alternative Instruments and Their Potential Use for Policymaking and Accountability in Peru. Document de recherches sur les politiques de la Banque mondiale* n° 4812. Washington, DC : Banque mondiale.
- LaBerge, D. ; Samuels, S.J., 1974. 'Toward a Theory of Automatic Information Processing in Reading.' *Cognitive Psychology*, 6, 293-323.
- Landerl, K. ; Wimmer, H. 2008. 'Development of Word Reading Fluency and Spelling in a Consistent Orthography: An 8-Year follow-up.' *Journal of Educational Psychology*, 100, 1, 150-161.
- Ladipo, O. ; Murray, T.S ; Greaney, V. 2009. *Using the Results of a National Assessment of Educational Achievement, Volume 5*. Washington, DC : Banque mondiale.
- Lavy, V. ; Spratt, J. ; Leboucher, N. 1995. *Changing Patterns of Illiteracy in Morocco: Assessment Methods Compared*. LSMS Paper 115. Washington, DC : Banque mondiale.
- Lesaux, N.K. ; Geva, E. 2006a. 'Synthesis: Development of Literacy in Language Minority Students.' Dans : D. August, T. Shanahan (dir. publ.) *Developing Literacy in Second Language Learners. Report of the National Reading Panel on Language Minority and Youth* (chapitre 3, p. 53-74). Lawrence Erlbaum associates.
- Lesaux, N.K. ; Pearson, M.R. ; Siegel, L.S. 2006b. 'The Effects of Timed and Untimed Testing Conditions on the Reading Comprehension Performance of Adults with Reading Disabilities.' *Reading and Writing*, 19(1), 21-48.
- Levine, K. 1998. 'Definitional and Methodological Problems in the Cross-National Measurement of Adult Literacy: The Case of the IALS.' *Written Language and Literacy*, 1(1), 41-61.
- Levine, R. ; Lloyd, C. ; Greene, M. ; Grown, C. 2008. *Girls Count: A Global Investment and Action Agenda*. Washington, DC : Centre pour le développement mondial.
- LeVine, R.A. ; LeVine, S.E. 2001. 'The Schooling of Women: Maternal Behavior and Child Environments.' *Ethos*, 29, 259-270.

- LeVine, R.A. ; LeVine, S.E. ; Schnell-Anzola, B. ; Rowe, M.L. ; Dexter, E. 2011. *Literacy and Mothering: How Women's Schooling Changes the Lives of the World's Children*. Oxford University Press.
- Levy, F. ; Murnane, R.J. 2004. 'Education and the Changing Job Market.' *Educational Leadership*, 62(2), 82.
- Lewis, M. ; Lockheed, M. (dir. publ.). March 2007. *Inexcusable Absence: Why 60 Million Girls Still Aren't in School and What to Do About It. CGD Brief*. Washington, DC : Centre pour le développement mondial.
- Linn, R.L. 2000. 'Assessments and Accountability.' *Educational Researcher*, 29(2), 4-16.
- Lockheed, M. 2004. *The Purpose of Good Quality Education. Paper commissioned for the EFA Global Monitoring Report 2005, The Quality Imperative*. Paris : UNESCO.
- . 2008. *Measuring Progress with Tests of Learning: Pros and Cons for 'Cash on Delivery Aid' in Education*. Document de travail n° 147. Washington, DC : Centre pour le développement mondial.
- Lockheed, M. ; Hanushek, E. 1988. 'Improving Educational Efficiency in Developing Countries: What Do We Know?' *Compare*, 18(1), 21-38.
- Lockheed, M. ; Verspoor, A. 1991. *Improving Primary Education in Developing Countries*. Oxford : Oxford University Press.
- Lonigan, C.J. ; Burgess, S.R. ; Anthony, J.L. 2000. 'Development of Emergent Literacy and Early Reading Skills in Preschool Children: Evidence from a Latent-Variable Longitudinal Study.' *Developmental Psychology*, 36(5), 596-613.
- McCloughlin, F. 2001. 'Dakar Wolof and the Configuration of an Urban Identity. *Journal of African Cultural Studies*,' Vol. 14, 2, p. 153-172.
- Meckes, L. ; Carrasco, R. 2010. 'Two Decades of SIMCE: An Overview of the National Assessment System in Chile.' *Assessment in Education: Principles, Policy & Practice*, 17, 2, p. 233-248.
- Mee, C.Y. ; Gan, L. 1998. 'Reading Practices in Singapore Homes.' *Early Child Development and Care*, 144(1), 13-20.
- Messick, S.J. 1989. 'Validity.' Dans : R.L. Linn (dir. publ.). *Educational Measurement*, 3^e édition. New York : American Council on Education & Macmillan, p. 13-103.

Références

- Mislevy, R.J. 2003. 'On the Structure of Educational Assessments.' *Measurement: Interdisciplinary Research and Perspectives*. 1, 3-62.
- Mislevy, R.J. ; Verhelst, N. 1990. 'Modeling Item Responses when Different Subjects Employ Different Solution Strategies.' *Psychometrika*, 55, 2, 195-215.
- Moors, A. ; De Houwer, J. 2006. 'Automaticity: A Theoretical and Conceptual Analysis.' *Psychological Bulletin*, Vol. 132(2), 297-326.
- Morris, D. ; Bloodgood, J.W. ; Lomax, R.G. ; Perney, J. 2003. 'Developmental Steps in Learning to Read: A Longitudinal Study in Kindergarten and First Grade.' *Reading Research Quarterly*, 38(3), 302-328.
- Mullis, I. ; Martin, M. ; Foy, P. 2008. *TIMSS 2007. International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, Mass., Boston College, Lynch School of Education, Centre d'études internationales TIMSS & PIRLS.
- Mullis, I.V.S. ; Martin, M.O. ; Kennedy, A.M. ; Trong, K.L. ; Sains, M. 2009. *PIRLS 2011 Assessment Framework*. Boston : Boston College, Centre d'études internationales TIMSS & PIRLS.
- Muthwii, M. 2004. 'Language of Instruction: A Qualitative Analysis of the Perception of Parents, Pupils, and Teachers among the Kalenjin in Kenya.' *Language, Culture, and Curriculum*, 17, 15-32.
- National Reading Panel 2000. *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction*. Bethesda, MD : NICHD.
- Nations Unies. 2000. *United Nations Millennium Declaration*. Résolution adoptée par l'Assemblée générale (United Nations A/RES/55/2). Consulté le 23 octobre 2010 : www.un.org/millennium/declaration/ares552e.htm
- Ndaruhutse, S. 2008. *Grade Repetition in Primary Schools in Sub-Saharan Africa: An Evidence Base for Change*. Londres : CfBT Education Trust.

Nordveit, B.H. 2004. *Managing Public-Private Partnership. Lessons from Literacy Education in Senegal. Africa Region Human Development*, Document de travail n° 72. Washington, DC : Banque mondiale.

OCDE (Organisation de coopération et de développement économiques). 2002. *Understanding the Brain – Towards a New Learning Science*. Paris : OCDE.

———. 2006. PISA 2006, *Executive Summary*. Paris : OCDE.

———. 2009a. PISA 2006, *Science Competencies for Tomorrow's World, Volume 2*. Paris : OCDE.

———. 2009b. PISA 2006, *Technical Report*. Paris : OCDE.

———. 2009c. PISA 2006. *Take the Test: Sample Questions from OECD's PISA Assessments*. Paris : OCDE.

OCDE/Statistics Canada. 1995. *Literacy, Economy and Society*. Paris : OCDE.

———. 1997. *Literacy Skills for the Knowledge Society: Further Results from the International Adult Literacy Survey*. Paris : OCDE.

———. 2000. *Literacy in the Information Age*. Paris : OCDE.

Olson, J.F. ; Martin, M.O. ; Mullis, I.V.S. 2008. *TIMSS 2007 Technical Report*. Association internationale pour l'évaluation du rendement scolaire (IEA), Centre d'études internationales TIMSS & PIRLS, Boston College.

Okech, A. ; Carr-Hill, R.A. ; Kataboire, A.R. ; Kakooza, T. ; Ndidde, A.N. 1999. *Evaluation of the Functional Literacy Program in Uganda*. Kampala : Ministère de l'Égalité des sexes, du Travail et du Développement social/Banque mondiale.

Onsumu, E. ; Nzomo, J. ; Obiero, C. 2005. *The SACMEQ II Project in Kenya: A study of the Conditions of Schooling and the Quality of Education*. Harare : SACMEQ.

Papen, U. 2005. Literacy and Development: What Works for Whom? Or, How Relevant is the Social Practices View of Literacy for Literacy Education in Developing Countries? *International Journal of Educational Development*, 25, 5-17.

Références

- Paris, A.H. ; Paris, S.G. 2003. 'Assessing Narrative Comprehension in Young Children.' *Reading Research Quarterly*, 38(1), 36-76.
- Paris, S.G. 2002. 'Measuring Children's Reading Development Using Leveled Texts.' *The Reading Teacher*, 56(2), 168-170.
- . 2005. 'Reinterpreting the Development of Reading Skills.' *Reading Research Quarterly*, Vol. 40, n° 2, 184-202.
- Paris, S.G. ; Carpenter, R.D. 2003. 'FAQs about IRIs.' *The Reading Teacher*, 56(6), 578-580.
- Paris, S.G. ; Carpenter, R.D. ; Paris, A.H. ; Hamilton, E.E. 2005. 'Spurious and Genuine Correlates of Children's Reading Comprehension.' Dans : S.G. Paris, S.A. Stahl (dir. publ.), *Children's Reading Comprehension and Assessment*, p. 131-160. Mahwah, NJ : Lawrence Erlbaum Associates.
- Paris, S.G. ; Hamilton, E.E. 2009. 'The Development of Reading Comprehension.' Dans : S. Israel, G. Duffy (dir. publ.). *Handbook of Reading Comprehension* (p. 32-53). Routledge : NY.
- Paris, S.G. ; Paris, A.H. 2006. 'The Influence of Developmental Skill Trajectories on Assessments of Children's early Reading.' Dans : W. Damon, R. Lerner, K.A. Renninger, I.E. Siegel (dir. publ.), *Handbook of Child Psychology: Vol. 4. Child Psychology in Practice* (6^e édition), p. 48-74. Hoboken, NJ : Wiley.
- Paris, S.G. ; Paris, A.H. ; Carpenter, R.D. 2002. 'Effective Practices for Assessing Young Readers.' Dans : B. Taylor, P.D. Pearson (dir. publ.), *Teaching Reading: Effective Schools and Accomplished Teachers*, p. 141-160. Mahwah, NJ : Lawrence Erlbaum Associates.
- Paris, S.G. ; Morrison, F.J. ; Miller, K.F. 2006. 'Academic Pathways from Preschool through Elementary School.' Dans : P. Alexander, P. Winne (dir. publ.), *Handbook of Research in Educational Psychology*, 2^e édition, p. 61-85. Mahwah, NJ : Lawrence Erlbaum Associates.
- Patrinos, H.A. ; Velez, E. 2009. 'Costs and Benefits of Bilingual Education in Guatemala: A Partial Analysis.' *International Journal of Educational Development*, 29, 594-598.

- Pearson, P.D. ; Hamm, D.N. 2005. 'The Assessment of Reading Comprehension: A Review of Practices – Past, Present, and Future', p. 13-69. Dans : S. Paris, S.Stahl (dir. publ.), *Children's Reading Comprehension and Assessment*. Mahwah, NJ : Lawrence Erlbaum Associates.
- Perfetti, C.A. 2003. 'The Universal Grammar of Reading.' *Scientific Studies of Reading*, 7, 3-24.
- Perfetti, C.A. ; Landi, N. ; Oakhill, J. 2005. 'The Acquisition of Reading Comprehension Skill.' Dans : M.J. Snowling, C. Hulme (dir. publ.), *The Science of Reading: A Handbook*, p. 227-247. Oxford : Blackwell.
- Pigozzi, M.J. 2006. 'Cross-national Studies of the Quality of Education.' Dans : K.N. Ross, I.J. Genevois (dir. publ.). *Cross-national Studies of the Quality of Education: Planning Their Design and Managing Their Impact*. Paris : IIPÉ-UNESCO.
- Piper, B. ; Korda, M. 2009. *EGRA Plus: Liberia. Data Analytic Report*. Rapport technique non publié. Washington, DC : RTI & Liberian Education Trust.
- Piper, B. ; Miksic, E. 2011. 'Mother Tongue and Reading: Using Early Grade Reading Assessments to Investigate Language-of-Instruction Policy in East Africa.' Dans : A. Gove, A. Wetterberg (dir. publ.). *The Early Grade Reading Assessment: Application and intervention to Improve Basic Literacy*. Research Triangle Park, NC : RTI Press.
- Piper, B. ; Schroeder, L. ; Trudell, B. 2011. *Oral Reading Fluency and Comprehension in Kenya: Reading Acquisition in a Multilingual Environment*. Document non publié.
- Porter, A.C. ; Gamoran, A. 2002. 'Progress and Challenges for Large-Scale Studies.' Dans : A.C. Porter, A. Gamoran (dir. publ.). *Methodological Advances in Cross-national Surveys of Educational Achievement*, p. 3-23. Conseil sur les études comparatives internationales en éducation. Washington, DC : National Academies Press.
- Postlethwaite, T.N. 2004. *What Do International Assessment Studies Tell Us About the Quality of School Systems?* Document de référence préparé en vue du *Rapport mondial de suivi sur l'EPT 2005 'L'exigence de qualité'*. 2005/ED/EFA/MRT/PI/40. Paris : UNESCO.

Références

- Pressley, M. 2000. 'What Should Comprehension Instruction Be the Instruction of?' Dans : M. Kamil, P. Mosenthal, P.D. Pearson, R. Barr (dir. publ.), *Handbook of Reading Research*, Vol. III, p. 545-561. Mahwah, NJ : Lawrence Erlbaum Associates.
- Prinz, M. 1996. *L'alphabétisation au Sénégal*. Paris : L'Harmattan.
- Puchner, L. 2001. 'Researching Women's Literacy in Mali: A Case Study of Dialogue among Researchers, Practitioners, and Policy Makers.' *Comparative Education Review*, 45(2), 242-256.
- Ravela, P. ; Arregui, P. ; Valverde, G. ; Wolfe, R. ; Ferrer, G. ; Martínez, F. ; Aylwin, M. ; Wolff, L. 2008. *The Educational Assessment Latin America Needs* (Document de travail n° 40). Washington, DC : PREAL.
- Rayner, K. ; Foorman, B.R. ; Perfetti, C.A. ; Pesetsky, D. ; Seidenberg, M.S. 2001. 'How Psychological Science Informs the Teaching of Reading.' *Psychological Science in the Public Interest*, 2, 31-74.
- Rayner, K. ; Reichle, E.D. ; Stroud, M.J. ; Williams, C.C. ; Pollatsek, A. 2006. 'The Effect of Word Frequency, Word Predictability, and Font Difficulty on the Eye Movements of Young and Older Readers.' *Psychology and Ageing*, Vol. 21, n° 3, p. 448-465.
- Research Triangle Institute (RTI). 2009. *Early Grade Reading Assessment Toolkit*. Washington, DC : RTI International.
- Riedel, B. 2007. 'The Relation between DIBELS, Reading Comprehension, and Vocabulary in Urban First-Grade Students.' *Reading Research Quarterly*. 42(4), 546-567.
- Robinson-Pant, A. (dir. publ.) 2004. *Women, Literacy and Development: Alternative Perspectives*. New York : Routledge.
- Robinson, C. 2005. *Languages and Literacies*. Document de référence préparé en vue du *Rapport mondial de suivi sur l'EPT 2006* « *L'alphabétisation, un enjeu vital* ». Paris : UNESCO.
- Roehrig, A.D. ; Petscher, Y. ; Nettles, S.M. ; Hudson, R.F. ; Torgesen, J.K. 2007. 'Accuracy of the DIBELS Oral Reading Fluency Measure for Predicting Third Grade Reading Comprehension Outcomes.' *Journal of School Psychology*, 46, 343-366.

- Ross, K.N. ; Genevois, I.J. 2006. *Cross-national Studies of the Quality of Education: Planning Their Design and Managing their Impact*. Paris : IIPÉ-UNESCO.
- Ross, K.N. ; Postlethwaite, T.N. 1991. *Indicators of the Quality of Education: A Study of Zimbabwean Primary Schools*. Harare : Ministère de l'Éducation et de la Culture ; Paris : IIPÉ-UNESCO.
- Ross, K.R. ; Saito, M. ; Dolata, S. ; Ikeda, M. ; Zuze, L. ; Murimba, S. ; Postlethwaite, T.N. ; Griffin, P. 2005. 'The Conduct of the SACMEQ II Project.' Dans : E. Onsomu, J. Nzomo, C. Obiero (dir. publ.) *The SACMEQ II Project in Kenya: A Study of the Conditions of Schooling and the Quality of Education*. Paris : SACMEQ-IIPÉ.
- Roy, P. ; Kapoor, J.M. 1975. *The Retention of Literacy*. Delhi : Macmillan of India.
- Rubens, A. ; Crouch, L. 2009. *Early Grade Mathematics Assessment (EGMA): A Conceptual Framework Based on Mathematics Skills Development in Children*. Rapport technique EdDataII. Washington, DC : USAID.
- Samoff, J. 2003. 'No Teacher Guide, No Textbooks, No Chairs: Contending with Crisis in African Education.' Dans : R.F. Arno, C.A. Torres (dir. publ.). *Comparative Education: The Dialectic of the Global and the Local*, p. 509-545. Boulder : Rowman & Littlefield.
- Samuels, S.J. 2007. 'The DIBELS tests: Is Speed of Barking at Print What We Mean by Reading Fluency?' *Reading Research Quarterly*, 42, 546-567.
- Scarborough, H.S. 1998. 'Early Identification of Children at Risk for Reading Disabilities: Phonological Awareness and Some Other Promising Predictors.' Dans : P. Accardo, A. Capute, B. Shapiro (dir. publ.), *Specific Reading Disability: A View of the Spectrum*. Timonium, MD : York Press.
- Schilling, S.G. ; Carlisle, J.F. ; Scott, S.E. ; Zeng, J. 2007. 'Are Fluency Measures Accurate Predictors of Reading Achievement?' *The Elementary School Journal*, 107, 5, p. 429-448.
- Scribner, S. ; Cole, M. 1981. *The Psychology of Literacy*. Cambridge : Harvard University Press.
- Sebba, M. 2007. *Spelling and Society: The Culture and Politics of Orthography around the World*. Cambridge University Press.

- Sen, A. 1999. *Development as Freedom*. NY : Anchor books.
- Share, D.L. 2008. 'On the Anglocentricities of Current Reading Research and Practice: The Perils of Overreliance on an 'Outlier' Orthography.' *Psychological Bulletin*, Vol. 134, n° 4, 584-615.
- Shaywitz, S. 2003. *Overcoming Dyslexia: A New and Complete Science-Based Program for Reading Problems at Any Level*. NY : A. Knopf.
- Shaywitz, S. ; Shawitz, B. 2008. 'Paying Attention to Reading: The Neurobiology of Reading and Dyslexia.' *Development and Psychopathology*, 20, 1329-1349
- Shepard, L.A. 2000. 'The Role of Assessment in a Learning Culture.' *Educational Researcher*, 29(7), 4-14.
- Sjoberg, S. 2007. 'PISA and 'Real Life Challenges': Mission Impossible?' Dans : S.T. Hopmann, G. Brinek, M. Retzl (dir. publ.), *PISA According to PISA. Does PISA Keep What It Promises?* Vienna : LIT Verlag. Consulté le 23 octobre 2010 : <http://folk.uio.no/sveinsj/Sjoberg-PISA-book-2007.pdf>
- Siniscalco, M.T. 2006. 'What are the National Costs for a Cross-National Study?' Dans : K. Ross, I.J. Genevois, (dir. publ.), *Cross-national Studies of the Quality of Education: Planning Their Design and Managing Their Impact*, p. 185-209. Paris : IIPPE-UNESCO.
- Slavin, R.E. ; Lake, C. ; Chambers, B. ; Cheung, A. ; Davis, S. 2009. *Effective Beginning Reading Programs: A Best-Evidence Synthesis. Best Evidence Encyclopedia*. Baltimore : Johns Hopkins University.
- Slobin, D.I. 1986. *The Cross-Linguistic Study of Language Acquisition*. Hillsdale, NJ : L. Erlbaum.
- Smallwood, J. ; McSpadden, M. ; Schooler, J.W. 2008. 'When Attention Matters: The Curious Incident of the Wandering Mind.' *Memory & Cognition*, 36, 1144-1151.
- Smith, G.T. ; McCarthy, D.M. ; Anderson, K.G. 2000. 'On the Sins of Short-Form Development.' *Psychological Assessment*, 12, 102-111.
- Smyth, J.A. 2005. *UNESCO's International Literacy Statistics 1950-2000*. Document de référence préparé en vue du *Rapport mondial de suivi sur l'EPT 2006 'L'alphabétisation, un enjeu vital'*. 2006/ED/EFA/MRT/PI/90. Paris : UNESCO.

- Snow, C. 2006. Cross Cutting Themes and Future Research Directions. Dans : D. August, T. Shanahan (dir. publ.) *Developing Literacy in Second Language Learners. Report of the National Reading Panel on Language Minority and Youth*, chapitre 21, p. 631-651. Hillsdale, NJ : L. Erlbaum.
- Snow, C.E. ; Burns, M.S. ; Griffin, P. 1998. *Preventing Reading Difficulties in Young Children*. Washington, DC : National Academy Press.
- Snow, C.E. ; Kang, J.Y. 2006. Becoming Bilingual, Biliterate, and Bicultural. Dans : W. Damon, R.M. Lerner, A. Renninger, I.E. Sigel (dir. publ.), *Handbook of Child Psychology, Volume 4, Child Psychology in Practice*, p. 75-102. Hoboken, NJ : John Wiley & Sons.
- Sprenger-Charolles, L. 2003. 'Linguistic Processes in Reading and Spelling. The Case of Alphabetic Writing Systems: English, French, German and Spanish.' Dans : T. Nunes, P. Bryant (dir. publ.). *Handbook of Children's Literacy*, p. 43-65. Dordrecht : Kluwer Academic.
- . 2008a. *EGRA (Early Grade Reading Assessment): Results from Primary School Students Learning to Read in French and in Wolof*. http://pdf.usaid.gov/pdf_docs/PNADL691.pdf
- . 2008b. *EGRA (Early Grade Reading Assessment): Results of 1200 Gambian Children Learning to Read in English*. http://pdf.usaid.gov/pdf_docs/PNADL690.pdf
- Sprenger-Charolles, L. ; Colé, P. ; Serniclaes, W. 2006. *Reading Acquisition and Developmental Dyslexia*. New York, NY : Psychology Press.
- Sprenger-Charolles, L. ; Messaoud-Galusi, S. 2009. *Review of Research on Reading Acquisition and Analyses of the Main International Reading Assessment Tools*. Document de travail informatif non publié. Paris : Project QLIP.
- Stanovich, K.E. 1980. 'Toward an Interactive-Compensatory Model of Individual Differences in the Development of Reading Fluency.' *Reading Research Quarterly*, 16, 32-71.
- . 1986. 'Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy.' *Reading Research Quarterly*, 21(4), 360-407.
- . 2000. *Progress in Understanding Reading: Scientific Foundations and New Frontiers*. New York : Guilford.

Références

- Stevenson, H.W. ; Stigler, J.W. 1982. *The Learning Gap: Why Our Schools Are Failing and What We Can Learn From Japanese and Chinese Education*. NY : Summit.
- Steiner-Khamsi, G. 2010. 'The Politics and Economics of Comparison.' *Comparative Education Review*, 54, 323-342.
- Storch, S.A. ; Whitehurst, G.J. 2002. 'Oral Language and Code-Related Precursors to Reading: Evidence from a Longitudinal Structural Model.' *Developmental Psychology*, 38(6), 934-947.
- Street, B.V. 2001. *Literacy and Development: Ethnographic Perspectives*. Londres : Routledge.
- Summers, L.H. 1992. 'The Most Influential Investment.' *Scientific American*, p. 132.
- Szucs, T. ; Belisari, A. ; Mantovani, L. 1997. 'Is Preventive Medical Care Worth the Cost?' *Biologicals*, 25, 247-252.
- Topol, B. ; Olson, J. ; Roeber, E. 2010. *The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments*. Stanford, CA : Stanford University, Stanford Center for Opportunity Policy in Education.
- UNESCO. 1978. *Towards a Methodology for Projecting Rates of Literacy and Educational Attainment*. Current Surveys and Research in Statistics, n° 28. Paris : UNESCO.
- . 1990. *Rapport final de la Conférence mondiale sur l'Éducation pour tous : Répondre aux besoins éducatifs fondamentaux. Jomtien, Thaïlande*. Paris : UNESCO.
- . 2000a. *Cadre d'action de Dakar. Éducation pour tous : tenir nos engagements collectifs*. Dakar/Paris : UNESCO.
- . 2000b. *L'évaluation des acquis scolaires*. Forum mondial sur l'éducation, Dakar, Sénégal : UNESCO.
- . 2004. *Rapport mondial sur l'EPT 2005. L'exigence de qualité*. Paris : UNESCO.
- . 2005. *Rapport mondial sur l'EPT 2006. L'alphabétisation, un enjeu vital*. Paris : UNESCO.
- . 2008. *Using a Literacy Module in Household Surveys: A Guidebook*. Bangkok : UNESCO.

- . 2010. *Rapport mondial sur l'EPT 2010. Atteindre les marginalisés*. Paris : UNESCO.
- UNESCO-LLECE. 2008. *Student Achievement in Latin America and the Caribbean. Results of the Second Regional Comparative and Explanatory Study (SERCE)*. Santiago, Chili : Bureau régional de l'UNESCO pour l'éducation en Amérique latine et dans les Caraïbes (LLECE). Consulté le 23 octobre 2010 : <http://unesdoc.unesco.org/images/0016/001610/161045e.pdf>
- UNSO (Bureau de statistique des Nations Unies – Wagner, D.A. ; Srivastava, A.B.L., auteurs principaux). 1989. *Measuring Literacy through Household Surveys*. Doc. n° DP/UN/INT-88-X01/10E. New York : UNSO.
- US Department of Education, NCES. 2009. *Basic Reading Skills and the Literacy of America's Least Literate Adults: Results from the 2003 National Assessment of Adult Literacy (NAAL) Supplemental Studies*. Report NCES 2009-481. Washington, DC : Ministère de l'Éducation américain.
- USAID. 2011. *Education: Opportunity through Learning*. USAID Education Strategy, 2011-2015. Washington, DC : USAID.
- Valencia, S.W. ; Smith, A.T. ; Reece, A.M. ; Li, M. ; Wixson, K.K. ; Newman, H. 2010. Oral Reading Fluency Assessment: Issues of Construct, Criterion, and Consequential Validity. *Reading Research Quarterly*, 45, 270-295.
- Uwezo. 2010. *Are Our Children Learning: Annual Learning Assessment Report, Uganda 2010*. Kampala. www.Uwezo.net
- van den Broek, P. ; Kendeou, P. ; Kremer, K. ; Lynch, J. ; Butler, J. ; White, M.J. ; Lorch, E.P. 2005. Assessment of Comprehension Abilities in Young Children. Dans : S.G. Paris, S.A. Stahl (dir. publ.), *Current Issues in Reading Comprehension and Assessment*. Mahwah, NJ : Lawrence Erlbaum Associates.
- Vansina, J. 1961. ' *De la tradition orale, essai de méthode historique* '. Musée royal de l'Afrique centrale. *Annales sciences humaines*, n° 16. Tervuren.
- Venezky, R.L. ; Sabatini, J.P. 2002. Introduction to this Special Issue: Reading Development in Adults. *Scientific Studies of Reading*, 6(3), 217-220.

- Volante, L. 2006. An Alternative for Large-Scale Assessment in Canada. *Journal of Learning and Teaching*, 4(1), 1-14.
- Wagner, D.A. 1980. Culture and Memory Development. Dans : H. Triandis, A. Heron (dir. publ.), *Handbook of Cross-Cultural Psychology*, Vol. 4, New York : Allyn & Bacon.
- . 1990. Literacy Assessment in the Third World: An Overview and Proposed Schema for Survey Use. *Comparative Education Review*, 33, 1, 112-138.
- . 1993. *Literacy, Culture and Development: Becoming Literate in Morocco*. New York : Cambridge University Press.
- . 1994. *Use it or Lose it? The Problem of Adult Literacy Skill Retention*. Rapport technique TR94-07 du NCAL, Philadelphie : Université de Pennsylvanie.
- . 1995. Literacy and Development: Rationales, Myths, Innovations, and Future Directions. *International Journal of Educational Development*, 15, 341-362.
- . 1997. Adult Literacy Assessment in Comparative Contexts. Dans : A. Tuijnman, I. Kirsch, D.A. Wagner, D.A. (dir. publ.). *Adult Basic Skills: Innovations in Measurement and Policy Analysis*. Cresskill, NJ : Hampton Press.
- . 1998. 'Literacy Retention: Comparisons across Age, Time and Culture.' Dans : S.G. Paris, H. Wellman, (dir. publ.). *Global Prospects for Education: Development, Culture and Schooling*, p. 229-251. Washington, DC : American Psychological Association.
- . 2000. *Literacy and Adult Education*. Étude thématique mondiale préparée en vue du Forum mondial sur l'éducation des Nations Unies. Dakar, Sénégal. Paris : UNESCO.
- . 2001. 'Conceptual Dichotomies and the Future of Literacy Work across Cultures.' Dans : C. Snow, L. Verhoeven (dir. publ.). *Literacy and Motivation: Reading Engagement in Individuals and Groups*. NJ : L. Erlbaum.
- . 2004. 'Literacy(ies), Culture(s) and Development(s): The Ethnographic Challenge.' *Reading Research Quarterly*, 39, 2, 234-241.

- . (dir. publ.). 2005. *Monitoring and Evaluation of ICT in Education Projects: A Handbook for Developing Countries*. Washington, DC : Banque mondiale.
- . 2008 (avril). *Educational Equity in a Multi-Lingual World*. Document présenté aux Réunions annuelles de la Société d'éducation comparative internationale. New York.
- . 2009a. *Mother Tongue and Other Tongue: A Fundamental Problem of the Home-School Connection*. Document présenté aux Réunions annuelles de la Société d'éducation comparative internationale. Charleston, SC.
- . 2009b. 'Pro-Poor Approaches to Using Technology for Human Development: Monitoring and Evaluation Perspectives.' Dans : S. Bekman, A. Aksu-Koç (dir. publ.). *Perspectives on human development, family and culture: Essays in honor of Cigdem Kagiticibasi*. Londres : Cambridge University Press.
- . 2010. 'Literacy.' Dans : M. Bornstein (dir. publ.). *Handbook of Cultural Developmental Science*, p. 161-173. NY : Taylor & Francis.
- . 2010. Quality of Education, Comparability, and Assessment Choice in Developing Countries. *COMPARE: A Journal of Comparative and International Education*, 40, 6, 741-760.
- . 2011. What Happened to Literacy? Historical and Conceptual Perspectives on Literacy in UNESCO. *International Journal of Educational Development*. 31, 319-323.
- Wagner, D.A. ; Daswani, C.J. ; Karnati, R. 2010. 'Technology and Mother-Tongue Literacy in Southern India: Impact Studies among Young Children and Out-of-School Youth.' *Information Technology and International Development*, 6, 4, 23-43.
- Wagner, D.A. ; Kozma, R. 2005. *New Technologies for Literacy and Adult Education: A Global Perspective*. Paris : UNESCO.
- Wagner, D.A. ; Spratt, J.E. ; Ezzaki, A. 1989. 'Does Learning to Read a Second Language Always Put the Child at a Disadvantage? Some Counter-Evidence from Morocco.' *Applied Psycholinguistics*, 10, 31-48.
- Wagner, D.A. ; Spratt, J.E. ; Klein, G. ; Ezzaki, A. 1989. 'The Myth of Literacy Relapse: Literacy Retention among Fifth-Grade Moroccan

Références

- School Leavers.' *International Journal of Educational Development*, 9, 307-315.
- Wagner, D.A. ; Venezky, R.L. ; Street, B.V. (dir. publ.). 1999. *Literacy: An International Handbook*. Boulder, CO : Westview Press.
- Wang, M. ; Koda, K. 2007. 'Commonalities and Differences in Word Identification Skills among Learners of English as a Second Language'. *Language Learning*, 5, Supplément 1, 201-222.
- Wainer, H. ; Thissen, D. 1996. 'How is Reliability Related to the Quality of Test Scores? What is the Effect of Local Dependence on Reliability?' *Educational Measurement: Issues and Practice*, 15, p. 22-29.
- Wolff, L. 2007. *The Costs of Student Assessments in Latin America*. Série des documents de travail n° 38. Washington, DC : PREAL.
- . 2008. *Costs and Financial Implications for Implementing National Student Assessments: Lessons for South Africa*. Document non publié.
- Wolff, L. ; Gurria, M. 2005. *Money Counts?* Banque interaméricaine de développement, Document de travail de l'ISU n° 3.
- Wuttke, J. 2008. 'Uncertainties and Bias in PISA.' Dans : S.T. Hopmann, G. Brinek, M. Retzl (dir. publ.), *PISA According to PISA. Does PISA Keep What it Promises?* p. 241-264.
www.univie.ac.at/pisaaccordingtopisa/pisazufolgepisa.pdf
- Ziegler, J.C. ; Goswami, U. 2006. 'Becoming Literate in Different Languages: Similar Problems, Different Solutions.' *Developmental Science*, 9, 429-436.

ANNEXES

ANNEXE A. DESCRIPTION DES ÉVALUATIONS DES COMPÉTENCES EN LECTURE

Remarque préliminaire : La présente annexe décrit succinctement les principaux outils d'évaluation évoqués dans ce rapport, en s'intéressant plus particulièrement à la lecture. Pour chaque outil, on trouvera un bref résumé contextuel, la population ciblée, les méthodes et le contenu de base, ainsi que la conception du matériel pédagogique. Bien que la plupart des organisations qui pilotent ces évaluations s'intéressent à l'apprentissage en général et non spécifiquement à la lecture, c'est essentiellement cet aspect qui est pris en compte dans les descriptions qui suivent.

Programme international de recherche en lecture scolaire (PIRLS)

a. **Contexte.** L'Association internationale pour l'évaluation du rendement scolaire (IEA), créée au début des années 1960, est le premier organisme à avoir mesuré les performances scolaires individuelles à des fins de comparaison internationale. Le Programme international de recherche en lecture scolaire (PIRLS) est la principale LSEA sur la lecture à l'école primaire. Il a été administré deux fois, en 2001 et 2006, et une nouvelle campagne a eu lieu en 2011³⁴⁹. Ce sont généralement les élèves de 4^e année, âgés de neuf ans en moyenne, qui sont évalués. Le cycle d'évaluation de 2006 a été administré dans 35 pays et la participation a été plus massive en 2011.

L'enquête PIRLS repose sur un modèle théorique de lecture, qui s'intéresse principalement aux processus de compréhension écrite³⁵⁰. Elle n'évalue pas les activités de décodage ou d'identification de mots, ni les liens entre compréhension de la langue écrite et de la langue parlée. Elle a pour objet de mesurer la compréhension écrite, définie par quatre compétences :

- se concentrer sur l'information clairement énoncée et la récupérer ;
- faire de simples inférences à partir d'événements logiques et reliés entre eux ;
- interpréter et intégrer des idées et de l'information ;
- examiner et évaluer le contenu, le langage et les éléments textuels.

349. Pour une description de l'enquête PIRLS, voir Mullis *et al.*, 2009.

350. Voir Kintsch et Van Dijk, 1978 ; Spiro, Bruce et Brewer, 1980.

Ces quatre processus interviennent dans la lecture pour accéder aux textes narratifs, ainsi que pour acquérir et utiliser l'information, objectifs qui sont censés résumer l'activité de lecture caractéristique des élèves de 4^e année dans le monde.

- b. **Population cible.** La 4^e année a été choisie parce qu'elle constitue une étape importante de l'acquisition de la lecture, étape à laquelle les élèves sont censés avoir acquis des compétences de base en matière de décodage, ainsi qu'être capables de passer un test, dont les instructions et les réponses sont formulées par écrit.
- c. **Méthode d'évaluation et contenu des tests.** La batterie de tests PIRLS est administrée collectivement en temps limité (80 minutes). Le PIRLS consiste en un ensemble de dix passages de textes, narratifs et informatifs. Toutefois, les élèves ne sont évalués que sur un passage de chaque type (un récit et un texte informatif en moyenne). La compréhension écrite de chaque passage est évaluée à l'aide d'une douzaine de questions, dont une moitié de questions à choix multiples (QCM) et une moitié de questions à réponse construite. Les élèves lisent les passages en silence, puis répondent individuellement aux questions par écrit. Quinze à trente minutes supplémentaires leur sont accordées pour remplir un questionnaire personnel.
- d. **Conception du matériel d'évaluation.** Des coordinateurs nationaux de la recherche (NRC), réunissant des représentants de chaque pays participant, soumettent des textes à un groupe de développement de la lecture, qui les approuve selon le principe de l'unanimité. Ces passages doivent respecter les consignes suivantes :
 - convenir à des élèves de 4^e année, en termes de contenu, d'intérêt et de capacité de lecture ;
 - être bien écrits, en termes d'intensité et de complexité, afin de permettre la formulation de questions, conformément aux processus et stratégies définis dans le cadre du PIRLS 2006 ;
 - tenir compte des groupes culturels pour éviter, dans la mesure du possible, des références culturelles spécifiques.

Les outils du PIRLS ont été rédigés en anglais, puis traduits dans 45 langues, à l'issue d'un processus minutieux de vérification. Une certaine liberté a été accordée à chaque pays pour traduire les passages quand cela s'avérait nécessaire, pour tenir compte des spécificités culturelles et linguistiques. Les NRC ont également rédigé un ensemble de questions pour chaque texte. Pour ce faire, ils ont tout particulièrement veillé à poser des questions en rapport avec l'objet du passage, et couvrant les différents

processus qui composent le PIRLS, tout en tenant compte du temps imparti, des sources potentielles de biais et de la facilité de traduction.

- e. **Pré-PIRLS.** Une version plus facile du PIRLS, destinée à évaluer la compréhension écrite des enfants encore en train d'apprendre à lire, a été élaboré et achevée en 2011. L'enquête pré-PIRLS repose sur les mêmes principes que le PIRLS et utilise une méthodologie similaire ; toutefois, les textes sont plus courts (environ 400 mots au lieu de 800 dans le PIRLS), avec un vocabulaire et une syntaxe plus simples. Qui plus est, l'accent est davantage mis sur les processus d'extraction d'informations et d'inférences directes, et moins sur l'assimilation des idées et l'évaluation du contenu. La méthodologie d'évaluation de la compréhension repose sur des questions, contrairement à l'enquête PIRLS (dans laquelle les questions sont posées après lecture du passage). Certaines questions du pré-PIRLS sont également éparpillées dans le texte : les élèves ont ainsi moins de texte à mémoriser pour trouver les réponses et peuvent donc répondre à certains items, même s'ils n'arrivent pas à lire tout le passage.

Programme international pour le suivi des acquis des élèves (PISA)

- a. **Contexte.** L'OCDE a lancé le Programme international pour le suivi des acquis des élèves (PISA) en 1997 pour répondre au besoin spécifique de disposer de données sur les performances des élèves, qui soient facilement comparables au niveau international. Le PISA a également été conçu pour recueillir des informations stratégiques pertinentes, afin d'aider les responsables politiques à expliquer les différences de performance entre écoles et pays³⁵¹. Depuis 2000, le PISA évalue tous les trois ans les compétences d'élèves de 15 ans, d'abord principalement dans des pays de l'OCDE, et aujourd'hui dans 57 pays au total. Le PISA porte sur trois domaines clés : les mathématiques, les sciences et la lecture. Chaque cycle PISA met l'accent sur l'un de ces domaines.

Le test de lecture PISA s'appuie sur des postulats théoriques similaires, mais non identiques à ceux du PIRLS. Le PISA veut aller plus loin que le simple décodage et l'interprétation littérale de l'information écrite, en évaluant les compétences en lecture et écriture dans des situations de la vie quotidienne. L'évaluation de la lecture du PISA définit cinq processus associés à une bonne compréhension d'un texte :

³⁵¹. Postlethwaite, 2004, p. 3.

- extraire de l'information ;
 - construire une compréhension générale ;
 - élaborer une interprétation ;
 - réfléchir sur et évaluer le contenu d'un texte ;
 - réfléchir sur et évaluer la forme d'un texte.
- b. **Population cible.** Le sous-test de lecture du PISA a pour but d'évaluer les capacités des élèves de 15 ans (quel que soit leur niveau scolaire), qui arrivent en fin de scolarité obligatoire, afin de déterminer dans quelle mesure ils sont préparés à relever les défis de la société actuelle, en mesurant ce qu'ils sont capables de faire par rapport à ce qu'ils ont appris à l'école. À ce titre, le PISA est davantage une évaluation de l'employabilité potentielle qu'une évaluation des processus scolaires.
- c. **Méthode d'évaluation et contenu des tests.** Le PISA cherchant à évaluer les formes de lecture qui se produisent en classe et en dehors, les textes sélectionnés relèvent de quatre contextes de lecture : pour l'usage privé, pour l'usage public, pour le travail, pour l'éducation. Les passages à lire sont également composés de textes continus (notamment des récits et rapports) et non continus (graphiques, cartes, publicités, etc.). Près de la moitié des questions destinées à évaluer la compréhension écrite des extraits sont des questions ouvertes (qui demandent de rédiger une réponse), tandis que les autres sont des questions fermées (réponses par oui/non ou QCM)³⁵². Chaque élève est évalué pendant plus de deux heures, dont entre 60 et 90 minutes sont dédiées au test de lecture. Différentes combinaisons de textes sont regroupées dans neuf livrets d'évaluation différents, afin de veiller à ce qu'un échantillon représentatif d'élèves soit testé sur chaque texte.
- d. **Conception du matériel d'évaluation.** Les pays participants ont répondu à un appel à soumission de textes. Ils avaient reçu des consignes décrivant l'objet du projet et précisant un certain nombre de variables, telles que les types et formats de textes, ainsi que le format

352. Les membres du groupe d'experts en lecture et les développeurs des tests ont par ailleurs identifié les processus pouvant avoir une incidence sur la difficulté d'un test de lecture. Ces processus sont notamment : établir un lien simple entre des informations ; faire des hypothèses à propos du texte ; décider de la quantité d'informations à extraire ; sélectionner le nombre de critères auxquels les informations doivent satisfaire ; choisir l'ordre des informations à extraire ; sélectionner la quantité de texte à assimiler ; spécifier les connaissances qui doivent être tirées du texte ; sélectionner l'importance de l'information (le degré d'explicitation avec lequel le lecteur est conduit vers cette information).

des réponses et le contexte. La soumission de textes authentiques a été encouragée, de préférence des articles de presse ou des textes originaux publiés. Après avoir passé en revue les textes reçus, les concepteurs du test ont validé un ensemble d'items pour l'évaluation. Ils ont également élaboré plusieurs textes et items totalement nouveaux. Après sélection, les items retenus ont été fournis en français ou en anglais aux équipes de traducteurs, qui ont dû résoudre les problèmes particuliers qui se posaient et ont veillé à la pertinence des traductions dans chaque pays³⁵³.

Consortium de l'Afrique australe et orientale pour le pilotage de la qualité de l'éducation (SACMEQ)

- a. **Contexte.** Le Consortium de l'Afrique australe et orientale pour le pilotage de la qualité de l'éducation (SACMEQ) est le fruit d'une vaste enquête nationale sur la qualité de l'enseignement primaire entreprise au Zimbabwe en 1991, avec le soutien de l'Institut international de planification de l'éducation (IIPÉ) de l'UNESCO³⁵⁴. La première étude, SACMEQ I, s'est déroulée entre 1995 et 1999 et a couvert sept pays. Elle a évalué la performance en lecture des élèves en 6^e année. Les pays participants étaient le Kenya, le Malawi, Maurice, la Namibie, la République-Unie de Tanzanie (Zanzibar), la Zambie et le Zimbabwe. La deuxième étude, SACMEQ II, s'est déroulée entre 2000 et 2002 dans 14 pays et un territoire (Zanzibar). Elle a évalué la performance en lecture et en mathématiques. La troisième étude,

353. Les problèmes de traduction entre langues ne peuvent être totalement résolus, même par des équipes multinationales d'experts de bonne volonté. Comme le disent Greaney et Kellaghan (2008, p. 42) : « Si l'on veut comparer les performances évaluées dans des langues différentes, l'analyse doit prendre en compte la possibilité que les écarts potentiels puissent être imputables à des différences de difficulté des tâches d'évaluation liées aux langues. Le problème est en partie réglé en changeant des mots. Par exemple, dans une évaluation internationale effectuée en Afrique du Sud, des mots comme « *gasoline* » (« *petrol* ») et « *flashlight* » (« *torch* ») ont été changés. Au Ghana, le mot « *neige* » a été remplacé par « *pluie* ». Si les différences linguistiques covarient avec des facteurs culturels et économiques, le problème est aggravé par le fait qu'il est parfois difficile de trouver un équivalent à la façon dont les questions sont rédigées, et de faire en sorte que le contenu d'un test soit approprié sur le plan culturel dans toutes les versions linguistiques. Par exemple, du matériel adapté au contexte d'élèves vivant en zone rurale (couvrant la chasse, le marché local, les activités agricoles et les jeux locaux) peut être peu familier pour des élèves vivant en milieu urbain. »

354. Voir Ross et Postlethwaite, 1991.

SACMEQ III, mise en œuvre en 2007, couvrait les mêmes pays qu'en 2002. Elle s'est achevée en 2011.

Les enquêtes SACMEQ II et III ont mesuré les niveaux de performance en lecture et en mathématiques des élèves ainsi que des enseignants³⁵⁵. SACMEQ II a défini le fait de savoir lire comme « la capacité de comprendre et d'utiliser les formes de langage écrit requises par la société et/ou auxquelles l'individu accorde de l'importance », comme dans l'enquête PIRLS.

- b. **Population cible.** Les études SACMEQ visait la population des élèves de 6^e année.
- c. **Méthode d'évaluation et contenu des tests.** Dans le SACMEQ, une analyse détaillée des programmes scolaires a tout d'abord été effectuée dans tous les pays, afin de définir les compétences en lecture considérées comme les plus importantes par chaque pays. Cette analyse a été réalisée après une discussion approfondie sur les compétences les plus importantes contenues dans les programmes de lecture de 6^e année. Il a été décidé d'adopter les grands domaines constitutifs de l'aptitude à lire utilisés dans le PIRLS. Un examen approfondi des programmes scolaires a également été effectué pour identifier des niveaux de compétences descriptifs, qui définissent une dimension reconnaissable et significative :
- Niveau 1 : les élèves doivent être capables d'associer des mots et des images d'objets courants « concrets ».
 - Niveau 2 : les élèves doivent être capables d'associer des mots à des concepts plus abstraits, comme des prépositions de lieu et de direction et, éventuellement, des idées et des concepts tels que les comparatifs et les superlatifs (le plus heureux, le plus gros, en dessous, etc.).
 - Niveau 3 : les élèves doivent être capables d'associer un ensemble de mots (comme une expression ou une courte phrase) à un autre, avec une concordance entre les deux ensembles.
 - Niveau 4 : les élèves doivent être capables de traiter des textes plus longs, contenant une séquence d'idées et de contenu, dont la compréhension découle des informations accumulées au fil de la lecture.
 - Niveau 5 : les élèves doivent être capables de lire un texte d'un bout à l'autre pour confirmer leur compréhension, associer de

355. L'évaluation des enseignants, si elle n'est pas rare dans les pays de l'OCDE, est inhabituelle dans les LSEA administrées dans les pays en développement.

nouvelles informations à un élément lu auparavant, relier les idées de passages distincts d'un texte, ou se montrer capables d'inférer l'intention de l'auteur.

Ces dimensions, en combinaison avec les trois domaines de la lecture, constituent un cadre (ou un exemple) pour élaborer des items de test adaptés.

- d. **Conception du matériel d'évaluation.** Une analyse détaillée des programmes scolaires a tout d'abord été entreprise dans tous les pays participants, afin de définir les compétences en lecture que ces derniers considéraient comme étant les plus importantes en 6^e année.

Programme d'analyse des systèmes éducatifs des pays de la CONFEMEN (PASEC)

- a. **Contexte.** Des enquêtes pour le Programme d'analyse des systèmes éducatifs des pays de la CONFEMEN³⁵⁶ (ou PASEC) ont été menées dans les pays francophones d'Afrique subsaharienne. En 1990, lors de la 42^e session de la CONFEMEN, à Bamako, l'Afrique francophone a décidé de relever le défi de l'EPT, lancé la même année à Jomtien. Les ministres ont décidé d'entreprendre un programme d'évaluation commun et le PASEC a été adopté en 1991, lors de la 43^e session de la CONFEMEN, à Djibouti. Le PASEC a pour objet de mesurer le niveau d'instruction de base en lecture (en français) et en mathématiques des élèves des écoles primaires des pays de l'Afrique francophone.
- b. **Population cible.** La population cible inclut les élèves de 2^e et 5^e années (avec un pré-test au début de chaque année et un post-test à la fin de chaque année).
- c. **Méthode d'évaluation et contenu des tests.** Contrairement aux autres LSEA, le PASEC fait une large part à la grammaire, en particulier pour les élèves de 5^e année, avec 10 sous-tests. La compréhension écrite est évaluée au niveau des mots, des phrases et des textes avec des textes à trous et des tâches d'appariement de mots/phrase-images (cinq sous-tests à la fin de la 2^e année). Les autres tests sont notamment la discrimination phonémique (trois sous-tests à la fin de 2^e année).
- d. **Conception du matériel d'évaluation.** Le contenu du test est basé sur le tronc commun des programmes scolaires des pays participants.

356. Conférence des ministres de l'Éducation des pays ayant le français en partage.

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

- a. **Contexte.** Le réseau des systèmes nationaux d'éducation des pays d'Amérique latine et des Caraïbes, connu sous la dénomination de Laboratoire latino-américain pour l'évaluation de la qualité de l'éducation (LLECE), a été créé en 1994 ; il est coordonné par le Bureau régional de l'UNESCO pour l'éducation en Amérique latine et dans les Caraïbes. Les évaluations du LLECE portent principalement sur les acquis scolaires en lecture et en mathématiques des élèves de 3^e et 4^e années, dans 13 pays du sous-continent, à savoir : Argentine, Bolivie, Brésil, Chili, Colombie, Costa Rica, Cuba, Honduras, Mexique, Paraguay, Pérou, République bolivarienne du Venezuela et République dominicaine. Le LLECE vise à produire des informations utiles pour la formulation et l'exécution des politiques éducatives dans les pays, en évaluant les acquis de populations d'élèves des écoles primaires³⁵⁷.
- b. **Population cible.** Dans chaque pays participant, des échantillons d'environ 4 000 élèves en 3^e année (ayant entre huit et neuf ans) et en 4^e année (entre neuf et dix ans) ont été évalués.
- c. **Méthode d'évaluation et contenu des tests.** Des tests de performance (deux formes) en langue (lecture) et en mathématiques ont été élaborés, en tenant compte des programmes scolaires de chaque pays participant. Les tests de langue comprenaient des QCM et des questions ouvertes. Les compétences linguistiques comprenaient la compréhension écrite, la pratique métalinguistique et la production d'un texte en espagnol (sauf au Brésil, où les élèves ont été testés en portugais).
- d. **Conception du matériel d'évaluation.** De nombreuses informations ont été collectées au moyen de questionnaires (remplis par les élèves, les enseignants, les chefs d'établissement et les parents ou tuteurs) sur les facteurs considérés comme potentiellement liés aux performances des élèves (par exemple la localisation de l'école, le type d'établissement, le niveau d'éducation des parents ou des tuteurs, le sentiment des enseignants et des élèves sur la disponibilité des ressources d'apprentissage à l'école).

357. Cette description est adaptée de Greeney et Kellaghan, 2008.

Évaluation des compétences fondamentales en lecture (EGRA)

- a. **Contexte.** L'évaluation des compétences fondamentales en lecture (EGRA) a pour objet de mesurer les performances en lecture des élèves du primaire, qui commencent à apprendre à lire, dans les pays en développement. Les sous-tests EGRA sont similaires à ceux des batteries de tests existantes, comme l'INDISSE (d'usage très répandu aux États-Unis), le but étant dans les deux cas d'évaluer les compétences émergentes en lecture-écriture qui sont, comme on le sait, liées à la maîtrise de la lecture.
- b. **Population cible.** L'objet principal de l'EGRA est d'évaluer les compétences en lecture au début de son enseignement, essentiellement de la 1^{re} à la 4^e année, dans les pays en développement.
- c. **Méthode d'évaluation et contenu des tests.** La plupart des sous-tests consistent à demander aux élèves de lire à voix haute et nécessitent, par conséquent, l'intervention d'un enquêteur. Les tâches de lecture à voix haute permettent de mesurer la fluidité (c'est-à-dire la précision et la vitesse) en termes de nombre d'items correctement traités en une minute. Les différentes sous-tâches³⁵⁸ sont :
 1. L'approche et le rapport au texte imprimé. Indiquer où commencer à lire et le sens de lecture sur une ligne et une page.
 2. La connaissance du nom des lettres (test d'une minute). Donner le nom (et parfois le son) de lettres en majuscule et en minuscule rangées dans un ordre aléatoire.
 3. La conscience phonémique. Segmenter les mots en phonèmes (prononciation des différents phonèmes d'un mot contenant deux à cinq phonèmes), en identifiant les sons initiaux dans différents mots.
 4. La lecture de mots familiers (test d'une minute). Lire des mots simples et courants d'une à deux syllabes.
 5. La lecture de mots inexistantes (ou pseudo-mots) non familiers (test d'une minute). Utilisation des correspondances graphème-phonème pour lire des mots simples sans signification.
 6. La fluidité de la lecture orale (ORF) en lisant un texte (test d'une minute). Lire un texte court avec précision.
 7. La compréhension écrite. Répondre correctement à différents types de questions (littérales et inférentielles) sur le texte qui a été lu auparavant.

358. Adapté de la Trousse d'outils EGRA (2009), p. 21-22.

8. La compréhension orale. Répondre à différents types de questions (semblables à celles utilisées pour évaluer la compréhension écrite) sur une histoire racontée par un enquêteur adulte.
 9. Dictée. Écrire, orthographier et utiliser correctement la grammaire lors d'un exercice de dictée.
- d. **Conception du matériel d'évaluation.** Le test peut être adapté en suivant les directives de la Trousse d'outils EGRA. Ces dernières conseillent d'élaborer des sous-tests respectant les spécificités de la langue : fréquence des lettres, des graphèmes et des mots, structure syllabique et position des lettres. La Trousse d'outils suggère également que les sous-tests de compréhension soient conçus en utilisant des récits provenant des manuels des élèves, en cohérence avec la culture locale, et que les questions reposent sur des faits et fassent appel à l'inférence (en évitant les réponses oui/non). L'EGRA ne se fonde pas sur une simple traduction, mais cherche plutôt à mettre en exergue la spécificité linguistique locale et les contraintes du langage écrit.

READ India (Pratham)

- a. **Contexte.** READ INDIA³⁵⁹ a pour objet de promouvoir un niveau élémentaire de compétence en lecture et en mathématiques. Une évaluation récente des outils de test a été effectuée, à partir de données de référence portant sur 15 000 enfants initialement testés³⁶⁰.
- b. **Population cible.** Ce programme cible les élèves de la 1^{re} à la 5^e année d'école. La langue d'enseignement est l'hindi.
- c. **Méthode d'évaluation et contenu des tests.** La campagne READ INDIA a été menée dans 350 districts, dans toute l'Inde. Elle comporte deux volets : évaluation des compétences de base en lecture et en mathématiques ; évaluation des compétences de plus haut niveau en lecture, en écriture et en mathématiques. Elle s'appuie sur un test préexistant (www.asercentre.org), dont le contenu est aligné sur les manuels officiels des 1^{re} et 2^e années d'enseignement de la langue. Les

359. READ INDIA est un projet de Pratham, une ONG indienne (voir www.pratham.org). « READ INDIA, en collaboration avec les gouvernements des États, veille à ce que tous les enfants indiens de la 1^{re} à la 5^e année sachent lire et faire des opérations mathématiques de base dans un délai de trois ans. ... Au cours de l'année scolaire 2008-2009, la campagne READ INDIA a été menée dans 350 districts, aux quatre coins de l'Inde. L'évaluation du programme READ INDIA est en cours dans deux districts des États de Bihar et d'Uttarakhand » (Abdul Latif Jameel Poverty Action Lab *et al.*, 2009, p. 1).

360. Abdul Latif Jameel Poverty Action Lab *et al.* (2009).

tests évaluent chaque année les compétences élémentaires en lecture et en arithmétique. Chaque année, de nouvelles matières/compétences sont également évaluées, telles que l'anglais, la compréhension et la résolution de problèmes. En ce qui concerne la lecture, on évalue si les élèves sont capables :

- d'identifier correctement quatre lettres sur cinq sélectionnées au hasard ;
- de lire correctement quatre mots courants sur cinq sélectionnés au hasard ;
- de lire, en 1^{re} année, des passages courts de quatre phrases d'environ 19 mots, que les élèves doivent lire « comme s'ils lisaient une phrase et non une suite de mots » ;
- de lire, en 2^e année, une histoire de sept à dix phrases d'environ 60 mots « avec fluidité et aisance » ;
- de répondre oralement à deux questions après avoir lu un texte.

READ INDIA utilise par ailleurs les sous-tests suivants issus de la batterie EGRA :

- facilité à nommer les caractères reconnus ;
- fluidité de la lecture des mots ;
- fluidité de la lecture de pseudo-mots ;
- fluidité de la lecture d'un texte ;
- compréhension écrite.

Des tests de langue écrits ont également été développés pour évaluer :

- la connaissance des lettres : dictée de lettres ;
- la connaissance des mots : associer une image et un mot, choisir un antonyme, donner un nom à une image ;
- la compréhension de phrases (tâche de décision lexicale) ;
- une phrase de closure : choisir le bon mot pour compléter une phrase ;
- la compréhension d'un passage (factuelle et inférentielle) : lire deux passages et répondre à des questions ;
- l'aptitude à écrire : dictée de mots (orthographe) ; donner un nom à des images ; construire une phrase ; lire des passages et répondre à des questions de compréhension.

- d. **Conception du matériel d'évaluation.** Le contenu des tests a été créé en s'inspirant de certaines approches des procédures de l'EGRA. Une

attention particulière a été portée au système d'écriture hindi³⁶¹. Les tests ont été conçus selon les principes suivants³⁶² :

- (a) le test doit comporter des contenus variés, afin que les tests contiennent des items adaptés aux élèves de la 1^{re} à la 5^e année ;
- (b) le contenu du test doit être adapté au contexte, à la langue et au programme scolaire de la population cible, et concorder avec les aptitudes et les compétences visées par le programme d'intervention ;
- (c) les tests doivent s'appuyer sur la recherche dans le domaine de la lecture pour évaluer les compétences identifiées comme importantes pour l'aptitude à lire ;
- (d) les formats des tests doivent avoir démontré leur faisabilité dans le cadre des évaluations à grande échelle ;
- (e) les tests doivent rendre compte des divers degrés d'aptitude, afin de mettre en évidence toute la palette des niveaux de performance ;
- (f) les items des tests doivent distinguer si les enfants se situent en haut ou en bas de l'échelle des aptitudes ;
- (g) le format des tests doit être facile à comprendre et être connu de la population cible ;
- (h) les tests doivent comprendre une combinaison de tests, oraux et écrits (crayon-papier) pour rendre compte des diverses compétences ;
- (i) les tests écrits doivent comprendre un mélange de questions à choix multiples et de questions ouvertes, afin de mettre en évidence les diverses compétences ;
- (j) les tests doivent être faciles à administrer et à noter, afin de pouvoir normaliser l'administration et la notation.

361. Selon Abdul Latif Jameel Poverty Action Lab *et al.* (2009, p. 2) : « L'hindi a un système d'écriture relativement simple. Toutefois, la transparence de la représentation lettre-son s'accompagne de la difficulté d'apprendre un grand nombre de caractères : formes primaires et secondaires des voyelles, unités consonne-voyelle, consonnes conjointes et groupes de consonnes. L'hindi n'a pas d'*aksbars* en minuscule et en majuscule, et une chaîne d'*aksbars* formant un mot est reliée par un trait supérieur. Ces spécificités de l'écriture hindi ont été prises en compte pour concevoir les outils d'évaluation ».

362. Abdul Latif Jameel Poverty Action Lab *et al.*, 2009, p. 2.

ANNEXE B. EXEMPLES D'ITEMS UTILISÉS DANS LES ÉVALUATIONS DES COMPÉTENCES EN LECTURE

PISA, exemple d'item³⁶³

Programme de vaccination volontaire contre la grippe d'ACOL

Vous savez sans doute qu'en hiver, la grippe peut se déclarer et se propager rapidement. Elle peut clouer ses victimes au lit pendant des semaines. Le meilleur moyen de lutter contre le virus est d'être en bonne santé. Il est fortement recommandé de faire chaque jour de l'exercice et d'adopter un régime alimentaire riche en fruits et légumes, pour aider le système immunitaire à lutter contre ce virus envahissant.

ACOL a décidé de donner au personnel la possibilité de se faire vacciner contre la grippe, comme mesure complémentaire, afin d'empêcher ce virus insidieux de se propager parmi nous. L'entreprise a demandé à une infirmière de venir faire la vaccination dans ses locaux, lors d'une demi-journée de la semaine du 17 mai, pendant les heures de travail. Ce programme est gratuit et destiné à tous les membres du personnel.

La participation est volontaire. Les personnes intéressées devront signer un formulaire de consentement dans lequel elles déclareront ne pas souffrir d'allergie et être informées du risque d'effets secondaires mineurs. Selon les médecins, la vaccination ne provoque pas la grippe. Mais elle peut avoir des effets secondaires tels que fatigue, légère fièvre et sensibilité du bras.

Qui doit se faire vacciner ?

Toute personne désirant se protéger contre le virus. Cette vaccination est particulièrement recommandée pour les plus de 65 ans. Mais, indépendamment de l'âge, elle est conseillée à toute personne atteinte d'une affection chronique débilante, touchant en particulier le cœur, les poumons ou les bronches, ou souffrant de diabète. Dans un environnement de bureau, tous les membres du personnel sont exposés au risque de contracter la grippe.

363. Synthèse adaptée de l'enquête PISA (2009c, p. 19-20).

Qui ne doit pas se faire vacciner ?

Les individus hypersensibles aux œufs, les personnes souffrant d'une maladie fébrile aiguë et les femmes enceintes. Si vous prenez des médicaments ou avez déjà fait une réaction à un vaccin contre la grippe, prenez conseil auprès de votre médecin. Si vous souhaitez vous faire vacciner la semaine du 17 mai, inscrivez-vous auprès de la responsable du personnel, Fiona McSweeney, au plus tard le vendredi 7 mai. La date et l'heure seront fixées en fonction de la disponibilité de l'infirmière, du nombre de participants et de l'heure la plus commode pour la plupart des membres du personnel. Si vous voulez vous faire vacciner pour cet hiver mais ne pouvez être présent à la date fixée, informez-en Fiona. Une autre séance pourra être organisée, s'il y a un nombre suffisant de personnes intéressées.

Pour toute information complémentaire, contactez Fiona, poste 5577.

Questions

Question 2.1

Laquelle des affirmations suivantes décrit une particularité du programme de vaccination contre la grippe proposé par ACOL ?

- A. Des séances d'exercice physique seront organisées tous les jours pendant l'hiver.
- B. La vaccination sera effectuée pendant les heures de travail.
- C. Une petite prime sera accordée aux participants.
- D. La vaccination sera effectuée par un médecin.

Question 2.2

Nous pouvons parler du contenu d'un texte (ce qu'il dit). Nous pouvons parler de son style (la façon dont il est présenté). Fiona voulait que le style de cette note d'information soit convivial et incitatif. Selon vous, a-t-elle réussi ? Expliquez votre réponse en faisant référence en détail à la présentation, au style d'écriture, aux illustrations ou autres éléments graphiques.

Question 2.3

Cette note d'information laisse entendre que si vous voulez vous protéger contre le virus de la grippe, le vaccin contre la grippe est :

- A. plus efficace que l'exercice physique et un régime alimentaire sain, mais plus risqué.

- B. une bonne idée, mais ne remplace pas l'exercice physique et un régime alimentaire sain.
- C. aussi efficace que l'exercice physique et un régime alimentaire sain, et moins difficile.
- D. superflu si vous faites beaucoup d'exercice physique et avez un régime alimentaire sain.

SACMEQ, conception du test de lecture

Niveau de compétence	Récit	Texte descriptif	Document	
Niveau 1	Associer un mot ou une image avec des prépositions de position ou de direction nécessitant de relier une image à une position ou à une direction pour répondre à la question	Associer un mot ou une image avec des prépositions de position ou de direction nécessitant de relier une image à une position ou à une direction pour répondre à la question	Associer un mot ou une image avec des prépositions de position ou de direction nécessitant de relier une image à une position ou à une direction pour répondre à la question	
Items	2	2	2	6
Niveau 2	Reconnaître le sens d'un mot et être capable de trouver son synonyme pour répondre à la question	Reconnaître le sens d'un mot et être capable de trouver son synonyme pour répondre à la question	Relier une information simple à un item ou à une instruction	
Items	7	6	9	22
Niveau 3	Relier des informations présentées par séquences d'idées et de contenu en continuant à lire	Relier des informations présentées par séquences d'idées et de contenu en continuant à lire	Rechercher systématiquement des informations en continuant à lire	
Items	8	10	8	26
Niveau 4	Chercher et confirmer des informations en relisant un texte à l'envers	Chercher et confirmer des informations en lisant un texte à l'envers	Relier plusieurs informations à différentes parties d'un document	
Items	9	5	4	18
Niveau 5	Relier les idées de différentes parties d'un texte. Faire des inférences à partir du texte ou au-delà, pour inférer les valeurs et opinions de l'auteur	Relier les idées de différentes parties d'un texte. Faire des inférences à partir du texte ou au-delà	Utiliser des listes intégrées, voire des publicités subtiles, où le message n'est pas explicitement exprimé	
Items	6	3	2	11
Nombre total d'items	32	26	25	83

PASEC, buts et items**Tableau synthétique Début 2^e année**

Tableau synthétique Début 2 ^e année		
Exercices	Domaines	Objectifs
5	Compréhension de mots (vocabulaire)	Identifier parmi 3 mots celui qui correspond à l'image
2	Compréhension de phrase	Écrire une phrase à partir de 4-5 mots donnés dans le désordre
8-9		Identifier la phrase (parmi 3) qui correspond à l'image (2 sous-tests)
1-6	Lecture/déchiffrement	Identifier une syllabe dans une série de mots (« pi » dans « épine, pipe, pilon ») Reconnaître un mot identique au mot test parmi 4 mots proches visuellement ou se prononçant de la même façon (« sot » : « saut, seau, pot, sot »)
7		Copie : Écrire le mot qui manque dans une phrase incomplète, la phrase complète étant présentée au-dessus de celle qui est à compléter
3-4	Écriture	Écrire une syllabe (3) ou un mot (4) à partir d'une lettre (2 sous-tests)

Tableau synthétique Fin 2 ^e année		
Exercices	Domaines	Objectifs
1	Compréhension de mots (vocabulaire)	Identifier parmi 4 images celle qui correspond au mot écrit présenté
4	Compréhension de phrases	Identifier le mot qui donne du sens à la phrase (« Il prend le train à la ... » [gare-oiseau-école])
6		Écrire une phrase à partir de 4-5 mots donnés dans le désordre
9		À l'aide d'une image, identifier la préposition donnant du sens à la phrase (« Sidi est [à-de-dans] la voiture »)
10	Compréhension de textes	Compléter un texte comportant des mots qui manquent (donnés, mais dans le désordre)
2-3-8	Lecture-écriture : discrimination de sons proches (t-d ; f-v ; br-pr)	Écrire après écoute la lettre (ou le groupe de lettres) qui manque (par exemple : « t ou d » dans « maXame » et « paXate » ; « f ou v » dans « Xarine » et « Xie » ; « pr ou br » dans « XXépare » et « XXanche », 3 sous-tests)
5	Grammaire (conjugaison)	Identifier le pronom personnel qui va avec le verbe conjugué (« ... parles trop » [tu-nous-vous])
7	Grammaire	Distinguer le singulier et le pluriel des noms (« Il porte des [cahiers, livre, mètre] »)

Tableau synthétique Début 5 ^e année		
Exercices	Domaines	Objectifs
1	Compréhension de mots et de phrases	Identifier le sens d'un mot dans une phrase : « la grande sœur a discuté avec son frère » signifie : « elle a travaillé avec lui », « elle a joué avec lui », « elle a parlé avec lui », « elle a mangé avec lui »
2		Identifier la préposition correcte (« le cheval trotte [contre-sous-dans] la rue »)
15-16	Compréhension de textes	Répondre à des questions dont la réponse se trouve explicitement dans le texte (lecture d'une notice de médicament) Lire un texte à trous et le compléter avec des mots donnés, dont 1 en trop
3	Grammaire 1	Accorder le participe passé : « Ma mère prépare mon plat... [préféré-préférée-préférés-préférer] »
4		Accorder le verbe avec le sujet : « Mon père et moi [allons-va-vont] à la foire »
5-6-7	Grammaire 2 (conjugaison)	Identifier le temps d'un verbe (indicatif présent, imparfait, passé composé, et futur simple) Identifier une phrase écrite sans erreur orthographique dans le verbe
8	Grammaire 3 (forme de la phrase)	Transformer une phrase affirmative en une phrase interrogative
9-10-11-13	Grammaire 4	Entourer le complément d'objet indirect ou le sujet d'une phrase (2 sous-tests) Entourer le pronom qui peut remplacer le groupe souligné (par exemple, « la fête aura lieu dimanche » [elles-vous-elle]) Compléter la phrase : « C'est l'école de Mady et de Kassi : c'est..... école »
12-14	Orthographe	Identifier le nom qui se termine par « x » au pluriel (bleu, chapeau, jupe) Orthographier correctement des homophones (« il [s'est-ces-ses-c'est] blessé »)

EGRA : Études de terrain et sous-tests utilisés

Langue(s) de l'évaluation, classe(s) testée(s), nombre d'enfants évalués ; les sous-tests sont numérotés comme dans la liste présentée auparavant en *Annexe A*. Un « X » dans la colonne « intervention » signifie que l'EGRA a été administrée pour suivre les progrès d'un programme d'intervention dans le domaine de la lecture. La taille des échantillons est indiquée entre parenthèses. La liste de pays n'est pas exhaustive, mais représentative.

364. Lorsque les données sur la taille de l'échantillon ne sont pas indiquées, cela signifie qu'elles n'ont pas été communiquées.

365. L'information sur les autres langues n'est pas disponible.

366. Une tâche supplémentaire de compréhension écrite a été ajoutée (texte à trous).

Annexes

Pays	Langue de l'évaluation	Classe testée (nombre d'élèves évalués)	Sous-tests employés	Intervention
Libéria		2 ^e année (429) 3 ^e année (407)	1, 2, 3, 4, 5, 6, 7, 8	X
Kenya	Anglais, swahili	2 ^e année ³⁶⁴	2, 3 (en anglais seulement), 5, 6, 7	X
Gambie	Anglais	1 ^{re} , 2 ^e et 3 ^e années (1 200)	1, 2, 3, 4, 5, 6, 7, 8, 9	
Sénégal	Français Wolof	Français : 1 ^{re} à 3 ^e année (502) Wolof : 1 ^{re} et 3 ^e années (186)	1, 2, 3, 4, 5, 6, 7, 8, 9	
Égypte	Arabe	(100)	1, 2, 3, 4, 5, 6, 7, 8, 9	X
Guatemala	Espagnol et langue maternelle ³⁶⁵ Espagnol, mam, k'iche, ixil	2 ^e et 3 ^e années 3 ^e année	2, 3, 4, 5, 6, 7, 8, 9 1, 6, 7	
Haïti	Créole haïtien, français	De la 2 ^e à la 4 ^e année (3 000)	2, 3, 4, 5, 6, 7, 8, 9 + une tâche de vocabulaire supplémentaire	
Honduras	Espagnol	De la 2 ^e à la 4 ^e année (2 226)	1, 6, 7	
Mali	Français, arabe Bamanankan, bomu, songhoi, fulfulde	Français pour les 2 ^e , 4 ^e et 6 ^e années ; arabe pour les 2 ^e et 4 ^e années, De la 1 ^{re} à la 3 ^e année dans les 4 autres langues ³⁶⁶	1, 2, 3, 4, 5, 6, 7, 8, 9	
Éthiopie	Ofo aromo	3 ^e année	1, 6, 7	
Guyane	Anglais	De la 1 ^{re} à la 3 ^e année (2 699)	2, 3, 4, 5, 6, 7, 8, 9	
Ouganda	Anglais, luganda, lango		2, 3, 4, 5, 6, 7	

READ India (Pratham) : contenu du test de lecture/écriture³⁶⁷

Synthèse du contenu des tests et description des items du test de lecture/écriture par écrit pour les 1 ^{re} -2 ^e années et les 3 ^e -5 ^e années			
Domaine	Description de l'item	1 ^{re} -2 ^e années	3 ^e -5 ^e années
Connaissance de l'akshar	Dictée en akshar	Oui	Oui
Lire du vocabulaire	Associer une image et un mot	Oui	Oui
	Choisir l'antonyme (opposé)	Oui	Oui
Mot et phrase	Compléter par un mot correspondant à l'image	Oui	Oui
	Écrire le mot qui décrit l'image	Oui	Oui
	Écouter et écrire un mot	Oui	Oui
	Utiliser des mots dans des phrases	Oui	Oui
Compréhension de phrases (tâche de décision lexicale)	Choisir parmi les options le mot qui convient pour compléter une phrase (tâche de labyrinthe)	Oui	Oui
	Trouver le mot qui convient pour compléter une phrase (tâche de closure)	Non	Oui
Compréhension d'un passage	Lire le passage 1 et répondre à des questions littérales (pour comprendre ce qui est lu)	Oui (QCM)	Oui (questions ouvertes)
	Lire un passage et répondre à des questions qui obligent à synthétiser l'information et à interpréter les idées	Non	Oui

Note : Cette synthèse repose sur la version finale du texte élaboré après le deuxième cycle de pilotage.

367. Adapté de Abdul Latif Jameel Poverty Action Lab *et al.*, 2009, p. 10.

À PROPOS DE L'AUTEUR

Professeur en sciences de l'éducation à l'Université de Pennsylvanie, Dan Wagner est titulaire de la chaire UNESCO d'apprentissage et d'alphabétisation. Il est directeur de l'Institut international d'alphabétisation, cofondé par l'UNESCO et l'Université de Pennsylvanie (www.literacy.org). Il dirige aussi son Centre national d'alphabétisation des adultes. Il est également directeur du Programme international de développement de l'éducation, un programme d'études supérieures de l'Université de Pennsylvanie. Après avoir obtenu une licence d'ingénierie à l'Université Cornell et s'être porté volontaire dans les Corps de la Paix au Maroc, il obtient son doctorat en psychologie à l'Université du Michigan, puis fait un stage postdoctoral de deux ans à l'Université de Harvard. Il a été chercheur invité (à deux reprises) à l'Institut international de planification de l'éducation (IIPÉ-UNESCO) à Paris, professeur invité à l'Université de Genève (Suisse) et titulaire d'une bourse d'études Fulbright à l'Université de Paris-V. Dan Wagner, qui possède une grande expérience dans le domaine des questions d'éducation aux niveaux national et international, a été conseiller auprès de l'UNESCO, de l'UNICEF, de la Banque mondiale, de l'USAID, du Department for International Development (DfID, Royaume-Uni) et d'autres agences travaillant sur les questions de développement international. Ses projets pluriannuels les plus récents concernent l'Inde, l'Afrique du Sud et le Maroc. Outre de nombreuses publications professionnelles, Dan Wagner a écrit ou dirigé la publication de plus de 20 ouvrages, parmi lesquels : *Literacy: Developing the future* (publié en cinq langues) ; *Literacy: An international handbook* ; *Learning to bridge the digital divide* ; *New technologies for literacy and adult education: A global review* ; *Monitoring and evaluation of ICT for education in developing countries*.

Publications et documents de l’IIEPE

Plus de 1 500 ouvrages sur la planification de l’éducation ont été publiés par l’Institut international de planification de l’éducation. Un catalogue détaillé est disponible ; il aborde les sujets suivants :

Planification de l’éducation

Généralités – contexte du développement

Administration et gestion de l’éducation

Décentralisation – participation – enseignement à distance –
carte scolaire – enseignants

Économie de l’éducation

Coûts et financement – emploi – coopération internationale

Qualité de l’éducation

Évaluation – innovations – inspection

Différents niveaux d’éducation formelle

De l’enseignement primaire au supérieur

Stratégies alternatives pour l’éducation

Éducation permanente – éducation non formelle – groupes défavorisés –
éducation des filles

Pour obtenir le catalogue, s’adresser à :
IIEPE, Unité de la communication et des publications
info@iiep.unesco.org

Les titres et les résumés des nouvelles publications peuvent être consultés
sur le site web de l’IIEPE, à l’adresse suivante :
www.iiep.unesco.org

L'Institut international de planification de l'éducation

L'Institut international de planification de l'éducation (IPE) est un centre international, créé par l'UNESCO en 1963, pour la formation et la recherche dans le domaine de la planification de l'éducation. Le financement de l'Institut est assuré par l'UNESCO et les contributions volontaires des États membres. Au cours des dernières années, l'Institut a reçu des contributions volontaires des États membres suivants : Australie, Danemark, Espagne, Inde, Irlande, Norvège, Pays-Bas, Suède et Suisse.

L'Institut a pour but de contribuer au développement de l'éducation à travers le monde par l'accroissement aussi bien des connaissances que du nombre d'experts compétents en matière de planification de l'éducation. Pour atteindre ce but, l'Institut apporte sa collaboration aux organisations dans les États membres qui s'intéressent à cet aspect de la formation et de la recherche. Le Conseil d'administration de l'IPE, qui donne son accord au programme et au budget de l'Institut, se compose d'un maximum de huit membres élus et de quatre membres désignés par l'Organisation des Nations Unies et par certains de ses institutions et instituts spécialisés.

Président :

Birger Fredriksen (Norvège)

Conseiller sur le développement de l'éducation pour la Banque mondiale (Ancien conseiller principal en éducation pour la région Afrique, Banque mondiale), Washington, DC.

Membres désignés :

Tiziana Bonapace (Italie)

Chef de la Section Développement et TIC (IDD), Commission économique et sociale des Nations Unies pour l'Asie et le Pacifique (CESAP), Bangkok, Thaïlande.

Carlos Lopes (Guinée-Bissau)

Secrétaire général adjoint des Nations Unies et Secrétaire exécutif de la Commission économique des Nations Unies pour l'Afrique (CEA), Addis-Abeba, Éthiopie.

Juan Manuel Moreno (Espagne)

Spécialiste principal en éducation, Département Moyen-Orient et Afrique du Nord, Banque mondiale, Washington, DC, États-Unis.

Moujahed Achouri (Tunisie)

Directeur, Division de la mise en valeur des terres et des eaux, Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO).

Membres élus :

Madiha Al-Shaibani (Sultanat d'Oman)

Ministre de l'Éducation, Mascate, Oman.

Ricardo Henriques (Brésil)

Secrétaire d'État au Travail social et aux Droits de l'homme (ancien Conseiller spécial du Président, Banque nationale de développement économique et social), Gouvernement d'État de Rio de Janeiro.

Valérie Liehti (Suisse)

Conseillère en matière de politique éducative, Coordinatrice pour l'éducation, Agence suisse pour le développement et la coopération (DDC), Département fédéral des affaires étrangères (DFAE), Section de l'Afrique de l'Ouest, Berne.

Dzingai Mutumbuka (Zimbabwe)

Président de l'Association pour le développement de l'éducation en Afrique (ADEA), Tunis, Tunisie.

Jean-Jacques Paul (France)

Professeur d'économie de l'éducation, Département économie et gestion des affaires, Université de Bourgogne, Dijon.

Hyunsook Yu (République de Corée)

Chargé de recherche principal, Bureau de recherche sur l'enseignement supérieur et l'éducation tout au long de la vie, Institut coréen du développement de l'éducation (KEDI), Séoul.

Zhang Xinsheng (Chine)

Président, Association éducative de la Chine pour l'échange international, Pékin.

Pour obtenir des renseignements sur l'Institut, s'adresser au :
Secrétariat du Directeur, Institut international de planification de l'éducation,
7-9, rue Eugène Delacroix, 75116 Paris, France



Pour améliorer l'apprentissage, il est fondamental d'utiliser efficacement les évaluations des acquis scolaires. Mais cette notion d'utilisation efficace ne fait pas uniquement référence aux paramètres techniques ou aux méthodes statistiques. Les évaluations de l'apprentissage en usage actuellement – qu'il s'agisse d'évaluations à grande échelle, d'enquêtes auprès des ménages ou d'évaluations hybrides (simples, rapides et abordables, « *smaller, quicker, cheaper* » ou SQC) – ont diverses finalités et applications. Le présent ouvrage donne une vue d'ensemble des évaluations de l'apprentissage et de leur importance pour la constitution d'une base de données empiriques ; il propose quelques idées nouvelles pour les rendre plus efficaces, en particulier pour les enfants les plus en difficulté.

Selon l'auteur, les évaluations hybrides de l'apprentissage peuvent renforcer la responsabilité éducative, favoriser une plus grande transparence et susciter un plus grand engagement des parties prenantes soucieuses d'améliorer l'apprentissage. Les pays ont, par ailleurs, besoin d'une politique à long terme pour choisir les évaluations les plus pertinentes, en privilégiant notamment les populations pauvres et marginalisées. Les initiatives actuelles pour élargir l'administration des évaluations des acquis scolaires dans les pays en développement sont déterminantes pour permettre des améliorations de l'éducation réelles et durables.

L'auteur

Professeur en sciences de l'éducation et directeur du Programme international de développement de l'éducation de l'Université de Pennsylvanie, Daniel A. Wagner est titulaire de la chaire UNESCO d'apprentissage et d'alphabétisation. Il est directeur de l'Institut international d'alphabétisation, cofondé par l'UNESCO et l'Université de Pennsylvanie. Il dirige également le Centre national d'alphabétisation des adultes. Sa grande expérience des questions éducatives aux niveaux national et international l'a amené à être conseiller auprès de l'UNESCO, de l'UNICEF, de la Banque mondiale, de l'USAID et d'autres organisations.