

L'ÉVALUATION DE LA LECTURE-COMPRÉHENSION DANS LES ENQUÊTES INTERNATIONALES : ENJEUX ET PERSPECTIVES

Dominique Lafontaine, Patricia Schillings

Pub. linguistiques | « *Revue française de linguistique appliquée* »

2015/2 Vol. XX | pages 9 à 20

ISSN 1386-1204

Article disponible en ligne à l'adresse :

<http://www.cairn.info/revue-francaise-de-linguistique-appliquee-2015-2-page-9.htm>

!Pour citer cet article :

Dominique Lafontaine, Patricia Schillings, « L'évaluation de la lecture-compréhension dans les enquêtes internationales : enjeux et perspectives », *Revue française de linguistique appliquée* 2015/2 (Vol. XX), p. 9-20.

Distribution électronique Cairn.info pour Pub. linguistiques.

© Pub. linguistiques. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

L'évaluation de la lecture-compréhension dans les enquêtes internationales : enjeux et perspectives

Dominique Lafontaine & Patricia Schillings, Université de Liège

Résumé : Cet article vise à mieux connaître la manière dont la lecture-compréhension est évaluée dans les deux principales évaluations internationales contemporaines : PIRLS (Progress in International Reading Literacy Study) et PISA (Programme international pour le suivi des acquis des élèves). La manière dont sont élaborées les épreuves internationales d'évaluation de la lecture (fondements théoriques, choix des textes, développement des épreuves, conditions de passation, contrôles de qualité en vigueur, etc.) est d'abord décrite de façon générale. Ensuite, la manière dont sont évaluées les compétences en lecture compréhension dans PIRLS et dans PISA est abordée de manière plus approfondie, en insistant sur les développements prévus pour PISA 2018. L'article dresse enfin un portrait synthétique des évolutions qu'ont connues les grandes évaluations internationales de la lecture en près de cinquante ans d'existence, en discutant certaines de leurs limites et les enjeux pour le futur de ce type d'évaluation.

Abstract: This paper sets out to present the way reading literacy is assessed in the two main comparative studies: PIRLS (Progress in International Reading Literacy Study) and PISA (the Programme for International Student Assessment). The way cognitive tests are developed is first presented in general terms (framework, selection of texts, test development, administration, quality controls, etc.). The way reading literacy is assessed in PIRLS and PISA is then described in greater detail. Finally, the paper provides an overview of the main trends in reading literacy comparative studies in nearly fifty years of existence; limitations, issues and challenges for the future are discussed.

Mots clés : enquêtes internationales, lecture-compréhension, modèles de référence

Keywords: international investigations, reading comprehension, reference models

1. Introduction

On ne connaît souvent des grandes enquêtes internationales portant sur la lecture telles que PISA ou PIRLS que le classement obtenu au test par les différents pays ou systèmes éducatifs. En dépit des efforts déployés par les organismes internationaux (IEA¹ ou OCDE) et les centres de recherche internationaux et nationaux en charge de ces évaluations pour produire des analyses qui dépassent le spectaculaire du classement et tentent d'expliquer le niveau de performances entre pays et à l'intérieur des pays, les informations pertinentes que l'on peut tirer des évaluations internationales sont souvent mal connues et sous-exploitées. Par ailleurs, ce que mesurent véritablement ces évaluations est tout aussi méconnu : nombre d'idées

¹ IEA est l'acronyme désignant l'Association internationale pour l'évaluation du rendement scolaire (voir <<http://www.iea.nl>>). L'IEA a conduit des enquêtes comparatives bien avant l'OCDE, dès la fin des années 1960, auxquelles la France et la Belgique francophone ont régulièrement pris part. Parmi les enquêtes les plus connues, on peut citer PIRLS (Progress in International Reading Literacy Study) et TIMSS (Third Study in Mathematics and Science).

erronées circulent à leur propos, donnant à penser par exemple que ces épreuves n'utilisent que des questions fermées, à choix multiple, portant sur des textes simplistes, pour la plupart d'origine anglo-saxonne, ou n'évaluent que des compétences de bas niveau. Dans cet article, nous tenterons de faire mieux connaître la manière dont la lecture (*reading literacy* en anglais) est évaluée dans les deux principales évaluations internationales de la lecture contemporaines : PIRLS (*Progress in International Reading Literacy Study*) et PISA (Programme international pour le suivi des acquis des élèves). Dans un premier temps, nous décrirons de façon générale la manière dont sont élaborées les épreuves internationales d'évaluation de la lecture (fondements théoriques, choix des textes, développement des épreuves, conditions de passation, contrôles de qualité en vigueur, etc.). Ensuite, nous aborderons de manière plus approfondie la manière dont sont évaluées les compétences en lecture compréhension dans PIRLS et dans PISA, en insistant sur les développements prévus pour PISA 2018. Nous terminerons en dressant un portrait synthétique des évolutions qu'ont connues les grandes évaluations internationales de la lecture en près de cinquante ans d'existence, en situant certaines de leurs limites et les enjeux pour le futur de ce type d'évaluation.

La première étude, PIRLS <<http://timssandpirls.bc.edu/>>, menée sous les auspices de l'Association internationale pour l'évaluation du rendement scolaire, concerne les élèves âgés de 9-10 ans (4^e année de l'enseignement primaire) ; elle a débuté en 2001 et se répète tous les cinq ans (2006, 2011, 2016). En 2011, 58 systèmes éducatifs ont pris part à PIRLS, dont 31 membres de l'Union européenne ou de l'OCDE. La seconde enquête, PISA <<http://www.oecd.org/pisa/>>, menée à l'initiative de l'OCDE, concerne les élèves âgés de 15 ans ; elle a débuté en 2000 et se répète tous les trois ans. Lors de chaque cycle, les élèves sont évalués en lecture, en mathématiques et en sciences, mais un des trois domaines, dit « majeur », est évalué plus en profondeur. La lecture a été domaine majeur en 2000 et en 2009, et le sera à nouveau en 2018. Tous les pays de l'OCDE prennent part à l'enquête PISA ; un certain nombre de pays hors OCDE y participent aussi, portant à 65 le nombre de participants à PISA 2012.

2. Qu'est-ce qu'une enquête internationale : aperçu en quelques points

Une enquête internationale a pour visée non seulement de comparer les performances d'échantillons représentatifs d'élèves dans les pays prenant part à l'enquête, mais aussi d'éclairer les différences de performances entre pays, ainsi qu'entre élèves et écoles à l'intérieur des pays, à l'aide de données de contexte recueillies auprès des élèves, des chefs d'établissement, parfois des enseignants et des parents.

- Les élèves de tous les pays participants passent *des épreuves identiques* qui ont été *traduites dans les différentes langues* au départ d'une version anglaise pour PIRLS, de deux versions sources, l'une en anglais, l'autre en français pour PISA. Les contrôles portant sur la qualité et l'équivalence des traductions sont extrêmement rigoureux et bien plus sévères qu'ils ne l'étaient par le passé (Grisay & Halleux 2007 ; Lafontaine 2001).
- Les *procédures de passation* des tests sont *standardisées* (les consignes sont les mêmes dans les différents pays) et des contrôles indépendants du pays sont effectués afin de vérifier si les procédures standard prévues sont respectées. Les épreuves d'évaluation sont administrées soit par du personnel extérieur à l'établissement, soit, le cas échéant, par une personne de l'établissement dûment formée pour cette tâche, mais qui ne peut être le professeur d'aucun des élèves évalués.
- Aussi bien PISA que PIRLS comportent *des questions à choix multiple et une proportion importante de questions à réponse ouverte où l'élève doit élaborer sa propre réponse*. A ces questions ouvertes, il existe souvent plusieurs possibilités de réponses, partiellement

ou entièrement correctes. Les questions ouvertes sont corrigées sur la base de grilles de correction critériées et standardisées. Les correcteurs sont spécifiquement formés pour cette tâche. La concordance entre correcteurs est contrôlée, sur le plan national et international, par une procédure de correction multiple des questions ouvertes (plusieurs correcteurs différents corrigeant un même ensemble de questions). Ces procédures veillent à éviter les divergences d'appréciation entre pays (sévérité ou laxisme excessif dans certains pays).

- Des *contrôles de qualité stricts* sont effectués à toutes les étapes du processus (traduction des tests, échantillonnage, administration et correction des tests).
- Les résultats sont exprimés en scores standardisés ; la moyenne est de 500 et l'écart type de 100. Les résultats au-dessus de 500 sont supérieurs à la moyenne et les résultats en dessous de 500 sont inférieurs à la moyenne. Il n'y a pas de maximum ni de minimum sur ces échelles ; les scores ne correspondent en rien à des pourcentages de réussite. Le principe est d'exprimer les scores en écarts par rapport à une moyenne fixée arbitrairement à 500, afin de faciliter les comparaisons d'une échelle et d'un cycle étude à l'autre.
- Les résultats en lecture sont présentés sur des échelles de compétences, comportant plusieurs niveaux de compétences hiérarchisés. A chaque niveau correspondent des tâches de lecture. Grâce à ces échelles, basées sur ce que l'on appelle des analyses IRT (*Item response theory*), on peut tout à la fois savoir quel pourcentage d'élèves a atteint un niveau rudimentaire, bon, voire excellent, et pour chacun des niveaux, appréhender le type de tâches de lecture que les élèves sont capables d'accomplir avec une certaine probabilité de réussite et donc leur niveau de compétence en lecture. Par exemple, les résultats de l'étude PIRLS 2011 indiquent qu'en Fédération Wallonie-Bruxelles, 45 % de nos élèves maîtrisent les compétences qui caractérisent le deuxième niveau de l'échelle (niveau intermédiaire). Les processus de compréhension mis en œuvre à cette étape consistent à identifier, à localiser plusieurs éléments inclus dans un texte, à faire des inférences directes et à localiser les différentes parties d'un document en recourant notamment aux sous-titres et aux encarts. Ces élèves éprouvent en revanche des difficultés majeures pour effectuer des inférences, pour expliquer le lien entre les intentions, les actions, les événements d'un récit, pour justifier une réponse en se référant au texte ou encore exprimer une préférence et la justifier (Schillings & al. 2012 ; Schillings & Lafontaine 2013).
- L'une des richesses de PIRLS et PISA est le caractère cyclique de ces évaluations, qui permet d'évaluer d'une manière rigoureuse l'évolution des performances dans un même système éducatif. Le principe sur lequel repose cette analyse est simple : l'essentiel des unités (textes et questions) font l'objet d'un embargo strict ; seules quelques unités sont rendues publiques après chaque cycle à titre illustratif. Une partie des épreuves gardées secrètes peut être réutilisée à l'identique lors de cycles ultérieurs, à côté de nouvelles unités.

3. Comment est élaborée une enquête internationale ?

Contrairement à ce que l'on pense parfois, les grandes enquêtes internationales contemporaines portant sur la lecture comme PISA ou PIRLS ne bâtissent pas sur du sable. Des théories ou modèles de la lecture sont en effet mobilisés pour l'élaboration et la mise en œuvre de ces enquêtes. Ceux-ci sont explicités dans le cadre de référence.

3.1. Le cadre de référence

De nos jours, préalablement à toute épreuve internationale, un cadre d'évaluation (*framework*) est défini (Mullis & al. 2009 ; OCDE 2012). Il n'en a pas toujours été ainsi. Les études internationales les plus anciennes, menées sous les auspices de l'IEA² entre les années 1968 et 1995, ne comportaient pas de cadre de référence formalisé³. A cet égard, l'enquête PISA (OCDE 1999) marque un véritable tournant. Depuis lors, toute épreuve internationale comporte ce document « fondateur » qu'est le cadre de référence pour l'évaluation. Il revient à un groupe d'experts⁴ internationaux du domaine évalué de produire une première version de ce cadre de référence ; celle-ci est largement mise en débat auprès de représentants des systèmes éducatifs ou des gestionnaires nationaux en charge de l'enquête, ensuite révisée afin d'arriver à une version définitive qui a reçu l'approbation des pays participant à l'étude.

Typiquement, le cadre de référence envisage les aspects suivants :

- Quel est l'objet de l'évaluation ? Quelle définition se donne-t-on de l'objet ? A quels modèles théoriques fait-on référence pour évaluer ce domaine ?
- Comment se structure le domaine que l'on veut évaluer ? Quels aspects, quelles compétences, quels processus l'évaluation doit-elle prendre en compte ?
- Quelles modalités d'évaluation seront utilisées ?
- Comment présenter et interpréter les résultats de l'évaluation ? Existe-t-il une compétence représentée par un score global sur une échelle, ou les compétences sont-elles multidimensionnelles, appelant plusieurs scores distincts ou plusieurs sous-échelles ?

3.2. Modèles de la lecture

Tant l'enquête PIRLS que PISA font référence à des modèles qui envisagent la lecture comme un processus interactif de construction de sens entre un texte, un lecteur et un contexte. « Les lecteurs sont vus comme construisant activement la signification (du texte) et connaissant des stratégies de lecture efficaces et comment réfléchir sur la lecture » (Mullis & al. 2009, 11). Dans PISA 2000 et 2009, par exemple, à côté de l'échelle globale définie tous textes et toutes questions confondus, le cadre de référence prévoit de présenter les résultats d'une part par aspect ou processus de lecture – *retrouver l'information, interpréter le texte et réfléchir sur le texte* –, d'autre part par type de texte en distinguant les textes continus (prose) des textes dits non continus comportant du texte, mais aussi des tableaux, des graphiques, des schémas ou de images. Il est intéressant de procéder de la sorte, pour affiner le diagnostic posé. En dépit de corrélations élevées⁵ entre les processus, certains systèmes éducatifs peuvent présenter des points forts ou des faiblesses spécifiques pour certains des processus ou des types de textes ainsi distingués par sous-échelle. Ainsi, dans PISA, on observe généralement que les pays du sud, de tradition latine (Espagne, Portugal, Grèce, pays d'Amérique du sud), sont relativement plus performants sur l'échelle *Réfléchir et évaluer* que sur l'échelle *Retrouver de l'information*. Ces points forts et faiblesses peuvent, à l'échelle des pays, être mis en relation avec les curriculums et les occasions d'apprendre (*opportunity to*

² Voir <<http://www.iea.nl>>.

³ Elles comportaient toutefois une analyse approfondie des curriculums nationaux. Les études IEA continuent à produire cette étude du curriculum.

⁴ Une des auteures du présent article, Dominique Lafontaine, est membre du 'Reading experts group' de PISA depuis la création du programme en 1998, et a aussi été membre du groupe PIRLS pour le cycle 2006. De son côté, Patricia Schillings est responsable nationale de l'enquête PIRLS pour la Belgique francophone et fait partie de l'équipe de développeurs de tests en lecture pour PISA 2018.

⁵ Les corrélations latentes entre les sous-échelles sont égales ou supérieures à .90.

learn) et déboucher sur des pistes d'amélioration plus spécifiques que ne le permettrait une échelle globale de compréhension en lecture.

3.3. Types/genres de textes et format de questions

Les évaluations internationales, souhaitant refléter la diversité des usages et des contextes, ont depuis toujours utilisé divers genres et types de textes. Les caractéristiques des textes qui servent de support dans le domaine de la lecture sont également définis dans le cadre de référence, qui précise les équilibres à respecter : combien de textes continus (prose) et non continus (mêlant de la prose et des graphiques, tableaux, listes...), combien de textes narratifs, argumentatifs, informatifs seront inclus dans l'épreuve de lecture.

Par ailleurs, le format des questions (questions à choix multiple, questions ouvertes à réponse courte et à réponse élaborée) figure également dans le cadre de référence, dans une table de spécifications où sont croisées les différentes dimensions à prendre en considération et les équilibres à respecter pour éviter tout biais de construction. Si par exemple, un processus de lecture était systématiquement associé à un format de question, il en résulterait une confusion peu opportune entre le processus et le format. La question de l'impact du format de question sur les performances a notamment été étudiée par Lafontaine et Monseur (2003). Ces auteurs ont ainsi pu montrer qu'une évaluation de la compréhension en lecture comportant davantage de questions ouvertes est de nature à augmenter significativement l'écart de performances entre les garçons et les filles, en faveur des filles.

Le cadre de référence définit enfin le type de support – papier-crayon ou support électronique – *via* lequel les épreuves seront administrées. De plus en plus, les évaluations internationales évoluent vers une administration des épreuves en ligne. Alors que l'évaluation PIRLS, qui concerne les élèves de 4^e année primaire, est toujours administrée en version papier, l'évaluation PISA 2018 dans laquelle la lecture sera pour la troisième fois le domaine majeur, ne sera plus administrée que sur ordinateur dans la majorité des pays.

3.4. Développement des unités de lecture (textes et questions/items portant sur le texte)

Le processus de développement des épreuves s'appuie sur un ensemble de lignes directrices préparées au début de projet : le cadre d'évaluation, une table spécifiant les caractéristiques des items, une présentation des aspects influençant la difficulté des items et des exemples d'items.

Première phase : développement local

Les équipes membres du consortium en charge de l'enquête pour un cycle développent un ensemble d'items (accompagnés de leur guide de correction) dans leur langue nationale (français, anglais, allemand, japonais...). Ces items font l'objet de laboratoires cognitifs individuels ou par petits groupes (les étudiants sont invités à réfléchir à voix haute en répondant aux items ou sont interviewés ...). Des ensembles d'items plus larges sont ensuite prétestés dans différentes classes d'élèves de 15 ans dans des établissements du pays où ils ont été développés. A chacune des étapes, les items sont révisés, voire éliminés s'ils posent des problèmes.

Deuxième phase : développement international

A ce stade, chaque unité (texte + items) est examinée par une des équipes du consortium autre que celle responsable du développement initial. On examine ici avec un soin particulier comment les items se comportent dans des cultures et contextes nationaux différents. Ceci conduit à écarter certaines unités ou certains items.

Proposition d'unités par les pays

En vue d'assurer une diversité culturelle et contextuelle, la contribution des pays à l'élaboration du test est sollicitée. Ceux-ci reçoivent un ensemble de lignes directrices, comme au début du processus, et sont invités à proposer des unités de lecture.

Révision des unités par les pays

A ce stade, chaque centre national reçoit un ensemble d'unités comportant souvent plusieurs centaines d'items qu'ils doivent évaluer – en s'entourant de panels d'experts nationaux – sur différents critères, tels que leur pertinence et leur intérêt pour des élèves de 15 ans, leur lien avec le curriculum national ou leur authenticité. Les gestionnaires nationaux doivent veiller à identifier tout problème d'ordre culturel ou linguistique (possibles difficultés de traduction par exemple).

Les unités et items sont aussi évalués, en parallèle, par les groupes internationaux d'experts qui se réunissent régulièrement. A l'issue de ces divers cycles de développement, consultations et révisions, deux versions – dites sources (une en anglais et une en français) – des unités sont préparées et distribuées aux pays pour traduction et adaptation locale. Ce processus de traduction/adaptation est étroitement cadré. Les unités ainsi traduites sont ensuite prétestées à large échelle dans tous les pays participants lors de *l'essai de terrain*.

Le panorama dressé à larges traits ci-dessus illustre bien en quoi l'élaboration d'un test international est une entreprise fondée sur la coopération. Plusieurs types d'acteurs y sont impliqués – développeurs de tests membres du consortium et nationaux, experts du domaine évalué, représentants des pays participants, gestionnaires nationaux des projets PIRLS ou PISA et panels d'experts nationaux – s'associent pour contribuer à produire un test qui soit culturellement diversifié, conforme à la fois aux attentes du cadre de référence et aux exigences de qualité et qui ne heurte pas les sensibilités culturelles ou nationales d'autres pays participants.

Lors de l'essai de terrain et du test définitif, il est systématiquement vérifié, via des procédures statistiques appropriées, que les items ne comportent pas de biais susceptibles d'avantager ou de désavantager certains groupes d'élèves – en l'occurrence les filles ou les garçons, les élèves de milieu socialement peu favorisé, issus de différentes cultures, ou encore des pays ou des groupes de langues. Si un item comporte ce que l'on appelle un DIF (*differential item functioning*), il est éliminé. Toutes ces procédures sont documentées en détails dans les rapports techniques produits à propos de chaque étude (Martin & al. 2007 ; OCDE 2012).

4. Comment sont évaluées les compétences en lecture-compréhension dans l'enquête internationale PIRLS ?

4.1. Evaluer des processus de lecture complexes

La littératie (*reading literacy*), objet d'étude de PIRLS, est définie comme « la capacité de comprendre et d'utiliser les formes du langage écrit requises par la société ou valorisées par l'individu » (Mullis & al. 2009). Les questions qui accompagnent les textes de l'épreuve sont élaborées de manière à couvrir les quatre processus suivants :

- retrouver et prélever des informations explicites (l'élève doit retrouver une information donnée dans le texte, par exemple quand et où se passe une histoire) ;
- faire des inférences simples (l'élève doit construire ou déduire une information qui n'est pas donnée telle quelle dans le texte, par exemple si le texte dit que les arbres n'ont pas de feuilles, déduire que l'histoire se passe en hiver ou identifier le référent d'un pronom) ;

- interpréter et intégrer des idées et des informations (l'élève doit utiliser ses propres connaissances du monde pour interpréter ce qui est dit dans le texte, par exemple identifier les sentiments ou les motivations d'un personnage) ;
- examiner et évaluer le contenu, la langue et les éléments des textes (dans ce cas, il s'agit de porter un jugement sur le texte, par exemple estimer si la fin d'un récit est attendue ou surprenante, et expliquer pourquoi).

Chacun de ces processus de compréhension est évalué dans des textes visant deux grands objectifs de lecture :

- lire pour l'expérience littéraire (textes narratifs) ;
- lire pour acquérir et utiliser de l'information (textes informatifs, argumentatifs ou injonctifs).

4.2. Cibler des textes d'envergure

L'épreuve de PIRLS 2011 comporte un total de dix unités (une unité se compose d'un texte et des questions qui s'y rapportent) – cinq unités littéraires et cinq unités informatives. Lors de la passation, chaque élève reçoit un carnet constitué de deux unités et dispose de deux pages de 40 minutes pour lire les textes et répondre aux questions. Tous les élèves ne passent donc pas le même carnet ; il existe un système de rotation. Les textes sur lesquels repose l'épreuve PIRLS sont tous relativement longs ; ceux-ci comportent en moyenne entre 800 et 1000 mots.

L'élève en relative difficulté ou peu motivé par la lecture peut certes être rebuté par la longueur des textes, et n'est pas placé dans une situation lui permettant de faire preuve de ses compétences, fussent-elles lacunaires, face à un texte plus court. C'est l'une des critiques que l'on peut adresser à l'enquête PIRLS, l'option d'avoir des textes longs uniquement ne reposant pas sur une argumentation solidement étayée dans le Cadre de référence théorique. Cependant, les élèves de tous les pays participants ayant été amenés à lire les mêmes textes, la comparaison entre pays reste pertinente. Les résultats peuvent être examinés à la lumière des informations relatives au contexte scolaire (recueillies via le questionnaire adressé aux enseignants et aux directions d'établissement). S'il s'avère, par exemple, que dans certains pays, les élèves sont rarement confrontés en classe à la lecture de textes relativement longs, ceci peut en partie expliquer pourquoi les élèves de ces pays éprouvent davantage de difficultés à réussir l'épreuve.

5. Comment les compétences en lecture-compréhension ont été et seront à l'avenir évaluées dans PISA ?

Dans cette section, après un bref rappel des caractéristiques de l'évaluation en 2000 et 2009, l'accent sera particulièrement mis sur les nouveaux développements envisagés pour PISA 2018.

Pour rappel, la définition de la lecture-compréhension adoptée dans PISA 2000 est la suivante.

Comprendre l'écrit, c'est « comprendre, utiliser des textes écrits, mais aussi réfléchir à leur propos, pour réaliser ses objectifs personnels, développer ses connaissances, ses potentialités, et participer à la vie en société⁶ » (OCDE 1999, 20).

Le cadre de référence précise en outre que « la compréhension de l'écrit va au-delà du simple décodage et de la compréhension littérale, elle implique aussi des facultés d'interprétation, de

⁶ En anglais : "Understanding, using and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and potential and to participate in society".

réflexion et la capacité d'utiliser la lecture pour accomplir des objectifs personnels. Comme l'enquête PISA donne la priorité à la lecture pour apprendre, plutôt qu'à l'apprentissage de la lecture, les compétences en lecture les plus élémentaires ne sont pas évaluées » (OCDE 2012, 23).

En 2009, le terme d'engagement a été ajouté à la définition pour souligner l'importance des aspects motivationnels de la lecture.

PISA évalue trois « aspects » ou processus de lecture :

1. accéder à et retrouver de l'information,
2. intégrer et interpréter les éléments d'un texte,
3. réfléchir à partir d'un texte, et évaluer son contenu ou sa forme.

Les textes qui servent de support aux questions peuvent être continus (prose) ou non continus (textes comprenant des tableaux, schémas, images, graphiques). Différents genres/types de textes sont représentés (narratifs, descriptifs, argumentatifs, injonctifs, transactionnels).

Conscients de l'importance croissante de la lecture électronique⁷ dans nos sociétés, et soucieux d'évaluer la capacité des jeunes de 15 ans face à ce type d'écrits, les responsables de PISA ont décidé, dès PISA 2009, d'une part, d'intégrer pleinement la lecture électronique dans le cadre de référence de l'évaluation de la lecture, et, d'autre part, de proposer une composante optionnelle : la lecture électronique (*Electronic Reading Assessment – ERA*)⁸. Les résultats des pays qui ont pris part à cette option en 2009 et en 2012 sont présentés dans les rapports produits par l'OCDE (OCDE 2011 ; OECD 2015).

Ces développements sont à nouveau au centre des travaux, en cours en cette année 2015, de préparation du Cadre de référence pour PISA 2018. Pour cette évaluation où la lecture se déclinera à nouveau en mode majeur, la lecture de textes en ligne ne sera plus une option, mais sera au cœur de l'évaluation même.

Il ne fait en effet aucun doute, pour les experts du domaine, que la production et la réception de l'écrit ont été profondément bouleversées par l'arrivée de la lecture en ligne ou sur support digital (ordinateur, tablette, smartphone). La différence entre la lecture sur papier et la lecture sur écran n'est pas seulement une différence de support : la différence de support introduit des modifications tant dans la forme et la structure que dans le contenu des textes. Si elle ne modifie pas en profondeur la nature des processus de lecture, certains d'entre eux y prennent une nouvelle importance – nous pensons ici à l'identification des sources que l'on trouve sur la toile et à l'évaluation de leur crédibilité. Le lecteur ne disposant pas de compétences affûtées à cet égard ne peut que se perdre dans le dédale qu'est Internet, au risque d'en subir des conséquences potentiellement graves. On imagine aisément, dans nos sociétés de l'information, les risques et dérives qu'entraîne tant au niveau des individus que de la société l'incapacité à faire la différence entre information étayée et propagande.

Le cadre de référence pour PISA 2018 sera donc revisité en profondeur pour tenir compte des spécificités de la lecture électronique et des avancées récentes de la recherche à ce propos (Afflerbach & Cho 2010 ; Coiro & Dobler 2007 ; Rouet & Britt 2011 ; Rouet & Britt 2014). S'il existe de nombreuses similarités entre les processus de lecture mobilisés face à un texte classique (papier) et un texte électronique, ce qui permet d'assurer la nécessaire continuité entre les cycles successifs de PISA, tant le cadre de référence que l'évaluation en 2018 prendront en compte les complexités supplémentaires qu'implique la lecture électronique, en

⁷ Divers termes sont utilisés de manière synonyme, y compris à l'intérieur des documents PISA : lecture électronique, lecture digitale ou lecture en ligne.

⁸ L'enquête PIRLS comportera aussi en 2016 une option de lecture électronique, intitulée e-PIRLS.

particulier le fait que le texte n'est plus donné, mais à construire, au gré des recherches, des clics sur les hyperliens ou les menus, ou de l'utilisation de la barre de défilement. Pour ce faire, le lecteur compétent a besoin de connaissances préalables, non seulement sur le contenu des textes et leur genre, mais aussi sur les moteurs de recherche et l'organisation des pages sur la toile. Selon Coiro & Dobler (2007), le lecteur doit faire davantage d'inférences « vers l'avant » pour prédire ce qui se cache derrière un lien et juger s'il est pertinent de visiter telle page plutôt que telle autre. Les chemins pour accéder à l'information étant multiples et les possibilités de choix bien plus nombreuses font que les stratégies de régulation (se fixer des buts, éviter de s'égarer) prennent une importance accrue dans la lecture en ligne. Enfin, comme mentionné ci-avant, la capacité à identifier les sources de l'information, la qualité de celle-ci et le degré de crédit que l'on peut lui accorder deviennent plus cruciales dans un univers où les sources sont souvent masquées et où n'importe qui peut désormais produire de l'information et la diffuser au monde entier, sans aucun contrôle éditorial.

6. Bilan et perspectives

La toute première évaluation internationale de la lecture a vu le jour au début des années 1970 (IEA Reading Comprehension 1971 ; Thorndike 1973). Au début des années 2000, nous nous étions livrées à une analyse approfondie des évolutions majeures qu'avait connues le champ en trente ans (Lafontaine 2001). Le bilan tiré alors reste largement d'actualité ; nous le revisitions dans les paragraphes qui suivent, en l'actualisant et en indiquant, en guise de conclusion, les enjeux et les perspectives qui se dessinent pour le futur.

Depuis le début des années 1970 :

- *Les conceptions de ce qu'est la compréhension en lecture, puis la littératie (reading literacy), se sont considérablement étendues et enrichies*, parallèlement au développement des recherches cognitives sur les processus de lecture. D'une conception de la compréhension centrée sur la restitution d'un sens univoque déposé dans le texte⁹, on a progressivement glissé à une conception de la littératie réconciliant les versants cognitifs et sociocognitifs, où est pleinement reconnu que le processus d'élaboration du ou des sens d'un texte résulte d'une interaction entre ce texte, un lecteur et un contexte (McNamara & Magliano 2009 ; Rumelhart 1985).
- *Le travail préalable d'élaboration théorique* s'est sensiblement développé sous la double influence des modèles théoriques (processus cognitifs, typologies de textes, recherches sur les facteurs prédictifs de la difficulté des tâches) et des modèles psychométriques (échelles IRT, Laveault & Grégoire 1997 ; Rasch 1980). La démarche empirique caractéristique des premières études a progressivement cédé le pas à une approche structurée, planifiée, où les équilibres entre différentes contraintes et critères sont définis a priori.
- La proportion de questions à choix multiple a régressé tandis que se développait l'usage de *questions à réponse construite* et de guides destinés à assurer une correction aussi objective que possible de ces réponses complexes, mieux en accord avec les conceptions contemporaines de la lecture. Un effort a été fait pour inclure dans les épreuves des textes ou des documents plus authentiques, que l'on n'écrit ou ne réécrit pas artificiellement pour les besoins de l'évaluation.
- L'apparition de nouveaux modèles psychométriques (modèles de la réponse à l'item) a dynamisé, ou accompagné, une approche plus diagnostique, davantage centrée sur les

⁹ Ce qui dépasse la restitution du sens était dans les années 1970 renvoyé à la sphère du littéraire, à laquelle une évaluation internationale a été consacrée par l'IEA (Purves 1973).

compétences. Les enquêtes internationales ne se limitent plus aujourd'hui à produire un score global de rendement, elles visent à établir quelles proportions de lecteurs sont capables de mobiliser certains types de compétences face à des tâches de lecture diversifiées. Les échelles proposent une description des compétences que possèdent les élèves situés à différents niveaux. Il est possible, au départ de ces échelles, dans une perspective plus didactique, de définir le chemin que doit parcourir un élève pour accroître ses compétences de lecteur¹⁰.

Les évaluations menées sur le plan international sont la résultante de compromis multiples. Celles-ci doivent composer avec les contraintes propres à toute évaluation à large échelle, les exigences liées à la comparaison des données et aux différents contextes nationaux et les avancées des savoirs de référence (modèles cognitifs, modèles psychométriques, modèles linguistiques, psycholinguistiques et didactiques pour la lecture). Certes, la machinerie de ces évaluations est lourde, ce qui la met à l'abri des effets de mode passagers. L'évolution qu'elles ont connues en près de cinquante ans est perceptible et montre à suffisance que ces « grosses machines » ont réussi à s'adapter, dans un laps de temps raisonnable, et compte tenu des contraintes qui leur sont propres, à une conception de la lecture interactive qui intègre l'évaluation de compétences plus complexes que le simple repérage d'informations littérales via des questions à choix multiple. Plus récemment, le tournant pris pour intégrer la lecture en ligne constitue une avancée majeure et démontre le souci des organismes pilotant ces enquêtes d'être « à la page », voire des pionniers en la matière.

Les nouvelles possibilités ou défis qui se profilent à l'horizon pour la prochaine décennie sont d'une part l'adaptivité de l'évaluation, d'autre part son interactivité, toutes deux liées à la capacité qu'offre l'administration du test en ligne de corriger automatiquement les réponses d'un individu passant l'épreuve. On peut ainsi, sur la base des premières réponses que donne un individu à des tâches de difficulté moyenne, l'orienter vers des tâches plus simples s'il échoue, et plus complexes s'il réussit. Le diagnostic posé se trouve affiné et on évite de la sorte le découragement et la démotivation qui peuvent gagner l'individu confronté à des tâches de lecture qui sont toutes hors de sa portée, ou l'ennui qui guette celui pour qui les tâches sont trop faciles. Par ailleurs, en termes d'interactivité, il est désormais concevable, comme cela se fait déjà dans des études à plus petite échelle ou dans des tutoriels de lecture intelligents (Llorens & al. 2014), de proposer des aides à l'individu qui le souhaite, de voir si l'individu fait appel aux aides proposées (par exemple définition en ligne de certains mots), de tester si l'individu, après une indication que sa réponse est erronée ou incomplète, est capable de rectifier le tir... Une telle approche s'ancre sur le plan théorique dans la notion de zone de développement proximal de Vygotsky : il s'agit de tester si un individu est capable de se dépasser et de faire un pas de plus grâce au soutien ou à l'étayage d'un pair plus expert que lui (en l'occurrence l'ordinateur, avec toute l'intelligence du concepteur qui est derrière). La manière dont les performances de l'individu testé sont reportées sur les échelles IRT est parfaitement compatible avec cette approche : il suffit en effet d'attribuer un code moins élevé pour une tâche réussie avec soutien que pour une tâche réussie de manière autonome, en étendant le principe du crédit partiel attribué pour une réponse incomplète.

Cette perspective ouvre la possibilité de dépasser une des principales limites des évaluations internationales, largement partagées d'ailleurs avec la plupart des évaluations : celles-ci testent en effet le plus souvent les compétences d'un individu isolé, face à son test, empêché de collaborer, ou au risque d'être sanctionné s'il le fait. Entre ce que l'on ne sait faire seul et l'absence de compétences, il existe une zone large de compétences

¹⁰ Des capsules vidéo à l'intention des enseignants ont ainsi été réalisées au départ de l'évaluation PIRLS 2011. Voir <http://www.enseignement.be/index.php?page=25703&ne_id=2205>.

intermédiaires, où l'on peut démontrer certaines compétences, à condition de pouvoir s'appuyer sur d'autres ou de recevoir du soutien technique, cognitif voire psychologique (feedbacks formatifs et encouragements) qui permet d'aller plus loin, et d'apprendre. L'heure est sans doute venue de prendre en compte ces compétences dans l'interaction. C'est techniquement possible, et s'inscrit pleinement dans les bouleversements sociétaux liés à l'utilisation des technologies de l'information. Face à cette réalité incontournable du monde d'aujourd'hui – la constante connexion aux autres et aux machines intelligentes – l'heure est sans doute venue d'évaluer non plus seulement l'individu en solo, mais l'individu-plus (Perkins, 1995), avec ses ressources externes, qu'elles soient humaines ou matérielles.

Dominique Lafontaine / Patricia Schillings
 Université de Liège
 Département Education et formation
 Service d'Analyse des Systèmes et des Pratiques d'Enseignement (aSPe)
 Sart Tilman, B32, B-4000 Liège, Belgique
 Tél : ++32 4 366 20 97
 Fax : ++32 4 366 28 55
 Site Internet : <<http://www.aspe.ulg.ac.be>>
 <Dominique Lafontaine <dlafontaine@ulg.ac.be>
 <Patricia.schillings@ulg.ac.be>

Références

- Afflerbach, P. & Cho, B.Y. (2009). Determining and Describing Reading Strategies. Internet and Traditional Forms of Reading. Metacognition, strategy and instruction. In Waters, H.S. & Schneider, W. (Eds), *Metacognition, use and instruction*, New York, Guilford Press.
- Coiro, J. & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth- grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42-2, 214-257.
- Grisay, A., de Jong, J.H., Gebhardt, E., Berezner, A. & Halleux-Monseur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement*, 8-3, 249-266.
- Lafontaine, D. (2001). Quoi de neuf en littératie ? Regard sur trente ans d'évaluation de la lecture. In Collès, L., Dufays, J.L., Fabry, G. & Maeder, C. (Eds), *Didactique des langues romanes. Le développement de compétences chez l'apprenant*, Bruxelles, De Boeck, 67-82.
- Lafontaine, A. (2006). *PIRLS 2006. Progress in Reading Literacy Study*. Article de synthèse. <http://www.enseignement.be/index.php?page=23827&do_id=9096&do_check=>.
- Lafontaine, D. & Monsieur, C. (2003). Etude de l'influence du format de question et de la définition de la population sur les disparités garçons-filles en lecture. In Actes du 16^e colloque international de l'Admee-Europe, Evaluation entre efficacité et équité (Liège, 4-6 septembre 2003), Liège, Service de Pédagogie expérimentale, *Cahiers de Pédagogie expérimentale*, 15-16.
- Laveault, D. & Grégoire, J. (1997). *Introduction aux théories des tests en sciences humaines*. Bruxelles, De Boeck, coll. Méthodes en sciences humaines.
- Llorens, A.C., Vidal-Abarca E., Cerdán, R. & Ávila, V. (2014). Adaptive formative feedback to improve strategic search decisions in task-oriented reading. *Journal of Computer Assisted Learning*, 30-3, 233-251, doi: 10.1111/jcal.12050.
- McNamara, D.S. & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, 51, 297-384.
- Martin, M.O., Mullis, I.V.S. & Kennedy, A.M. (2007). *Pirls 2006, Technical report*. Chesnut Hill, MA, Boston College.
- Mullis, I.V.S., Martin, M.O., Kennedy, A.M., Trong, K.L. & Sainsbury, M. (2009). *PIRLS 2011 Assessment Framework*. Chesnut Hill, MA, Boston College.

- OCDE (1999). *Mesurer les connaissances et compétences des élèves. Un nouveau cadre d'évaluation*. Paris, OCDE.
- OCDE (2011). *Résultats du PISA 2009. Elèves en ligne. Technologies numériques et performance* (volume VI). Paris, OCDE.
- OCDE (2012). *Le cadre d'évaluation de PISA 2009. Les compétences clés en compréhension de l'écrit, en mathématiques et en sciences*. Paris, OCDE.
- OECD (2012). *PISA 2009 Technical report*. Paris, OECD. <<http://dx.doi.org/10.1787/9789264167872-en>>.
- OECD (2015). *Students, Computers and Learning: Making the connection*. Paris, OECD.
- Perkins, D.N. (1995). L'individu-plus. Une vision distribuée de la pensée et de l'apprentissage. *Revue française de pédagogie*, 111, 57-71.
- Purves, A.C. (1973). *Literature education in ten countries. International studies in evaluation II*. Stockholm, Almqvist & Wiksell.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, University of Chicago Press.
- Rouet, J.F. & Britt, M.A. (2011). Relevance processes in multiple document comprehension. In McCrudden, M.T., Magliano, J.P. & Schraw, G. (Eds), *Text relevance and learning from text*, Greenwich, CT, Information Age Publishing, 19-52.
- Rouet, J.F. & Britt, M.A. (2014). Learning from multiple documents. In Mayer, R.E. (Ed.) *Cambridge handbook of multimedia learning*, (2nd ed.), Cambridge, MA, Cambridge University Press.
- Rumelhart, D.E. (1985), Toward an interactive model of reading. In Singer, H. & Ruddell, R.B. (Eds), *Theoretical models and the processes of reading*, (3rd ed.), Newark, DE, International.
- Schillings, P., Hindryckx, G., Dupont, V., Matoul, A. & Lafontaine, D. (2012). *Pirls 2011. Enquête internationale sur le développement des compétences en lecture des élèves de 4^e année primaire. Note de synthèse*. <http://www.enseignement.be/download.php?do_id=9687&do_check=>>.
- Schillings, P. & Lafontaine, D. (2013). Les résultats de l'enquête PIRLS sur la lecture en quatrième année primaire : des compétences à mettre à l'étude. *Caractères*, 44, 5-15.
- Thorndike, R.L. (1973). *Reading comprehension education in fifteen countries. International studies in evaluation III*. Stockholm, Almqvist & Wiksell.