

# **Que nous apprend PIRLS sur la compréhension des élèves français de 10 ans ?**

**Martine Rémond, IUFM de Créteil et INRP, EducPol.**

---

Cette contribution porte sur l'analyse de corpus de réponses des élèves français à l'évaluation internationale PIRLS 2001. Ces élèves, scolarisés en CM1, présentent un score moyen de compréhension un peu supérieur à la moyenne internationale. Quelques explications sont avancées sur l'origine de leurs difficultés face à un matériau complexe, qu'ils ont dû gérer seuls et pour lequel, la fréquence des non-réponses aux questions nécessitant une réponse écrite les caractérise par rapport à d'autres pays. Ils ont une tendance marquée à ramener leurs réponses à un prélèvement d'information dans le texte alors que souvent, il faut mener des opérations plus complexes pour manifester sa compréhension. Ils limitent au maximum la production écrite de réponses. Confrontés à des écrits ou à des tâches ressemblant à du « connu », leurs performances s'accroissent notablement ; mais PIRLS propose une large palette de tâches à laquelle ils ne s'adaptent pas bien. Des questionnaires situent les pratiques et les contenus d'enseignement en CM1 et ouvrent des pistes de travail.

Pour acquérir la plasticité nécessaire pour gérer des tâches variées de compréhension, comme celles de PIRLS, les élèves doivent apprendre à comprendre et à interpréter des écrits de qualité, à devenir des lecteurs actifs et capables de gérer leur compréhension.

Le contexte des enquêtes internationales a déjà été présenté dans un article sur l'évaluation PISA diligentée par l'OCDE (Rémond, 2005), aussi, il n'est pas repris dans cet article consacré à la seconde grande évaluation internationale en cours, PIRLS menée sous la houlette de l'IEA (International Association for the Evaluation of Educational Achievement), organisme fondateur de ces grandes enquêtes. PIRLS (Progress in International Reading Literacy Study) porte sur la compréhension de l'écrit et sur les facteurs associés à son apprentissage ; il cible l'évaluation du niveau de compréhension d'élèves de 10 ans et de leurs expériences de l'écrit aussi bien à l'école qu'à la maison. Trente cinq pays ont participé en 2001 à cette enquête qui, comme PISA, a un caractère cyclique (première application en 2001, puis en 2006). Une méthodologie complexe conditionne la construction et l'organisation du matériel, élaboré en anglais pour former la « version source » servant de base commune à toutes les traductions.

## 1. L'organisation générale de PIRLS

Le programme PIRLS évalue les performances en lecture d'élèves de 4<sup>e</sup> année de scolarisation obligatoire. Il a été développé par une équipe de l'International Study Center de Boston College (États-Unis) et les outils de lecture ont été élaborés par NFER (Grande-Bretagne). Ce programme entretient de nombreuses similitudes avec PISA, tant par sa conception que dans ses analyses<sup>1</sup>. Ainsi, il repose sur un cadre conceptuel « Framework and specifications for PIRLS assessment 2001 » (Campbell *et al.*, 2001) dans lequel la « reading literacy<sup>2</sup> » est définie comme : « l'aptitude à comprendre et à utiliser les formes du langage écrit dont la société a besoin et / ou qui sont nécessaires à l'individu. Les jeunes lecteurs peuvent construire du sens à partir de textes très variés. Ils lisent pour apprendre, pour s'intégrer dans une société où la lecture joue un rôle essentiel, et pour leur plaisir. » (ibid., p. 3).

Cette définition assez large conditionne cette enquête. Le protocole d'évaluation de la compréhension est composé de huit unités (textes et items) bâties sur la base du croisement de deux dimensions : les Objectifs de lecture et les Processus de compréhension.

Deux **objectifs de la lecture** (« purposes for reading ») ont été retenus et fixent le choix des textes supports de l'épreuve :

- lire pour accéder aux textes littéraires,
- lire pour acquérir et utiliser des informations,

car « ils correspondent à la majorité des lectures faites par les jeunes élèves, à la fois, à l'école et en dehors, et donc aux raisons de lire à cet âge » (ibid., p. 15).

Quatre textes ont été sélectionnés pour chaque objectif (nous les appellerons désormais pour simplifier littéraire et informatif). pour l'objectif littéraire, il s'agit de quatre textes narratifs donnés dans une version complète de l'histoire. Ceux de l'objectif informatif se présentent sous des formes variées (texte documentaire portant sur des comportements animaux, sur la vie de Léonard de Vinci, prospectus, lettre, etc.).

Les items doivent évaluer les quatre **processus de compréhension** suivants<sup>3</sup> : prélever des informations explicites, faire des inférences directes, interpréter et assimiler des informations et des idées, examiner et évaluer le contenu, la langue et les éléments textuels (pour une description en français, voir annexe 1). La moitié des items<sup>4</sup> sont sous forme de QCM à 4 modalités, les autres nécessitent de construire des réponses selon une modalité écrite, même s'il est explicitement indiqué que l'écrit ne fait pas ici l'objet d'évaluation (Mullis *et al.*, 2003, p. 51).

---

1 Nous avons participé à ce dispositif en tant que membre du groupe d'experts internationaux.

2 Nous avons choisi de ne pas traduire cette expression.

3 Traduit par « compétences » par la DEP.

4 Nous emploierons sans les distinguer les termes items et questions.

Au travers du traitement de ces items, les élèves montreront pour chaque objectif comment ils maîtrisent les quatre processus de compréhension. Pour chaque texte, entre douze à treize items remplissent cette fonction ; cet ensemble correspond aux huit unités annoncées précédemment. Les textes comportent entre 500 et 800 mots, ils sont donc assez longs.

Les huit unités (textes et items associés, soit un total de 98 items) servent à constituer dix « cahiers tournants ». Chaque élève reçoit un cahier composé d'une unité de l'objectif littéraire et d'une unité de l'objectif informatif. Il dispose ensuite de 60 minutes fractionnées en deux temps pour le traiter.

Le codage des réponses aux questions ouvertes est fait en fonction des instructions figurant dans un guide afin d'effectuer des corrections aussi standardisées que possible, d'un pays à l'autre, d'une langue à l'autre. Les items les plus simples sont crédités d'un seul point tandis que d'autres, plus complexes, sont pondérés en fonction de la précision de la réponse. Un exemple d'item à codage « progressif » est donné en annexe, il s'agit d'un item se référant au texte *Les souris à l'envers* (Roald Dahl) qui figure également en annexe et qui servira de fil rouge à cet article.

La mise en œuvre de méthodologies complexes a abouti à la publication d'un rapport international qui met largement l'accent sur la comparaison des performances entre les pays. Force est de constater que les pays sont rangés dans les innombrables tableaux selon leur score à l'évaluation PIRLS, ce qui revient à les ordonner (Mullis *et al.*, 2003). Les pays sont classés sur une dimension commune, la moyenne internationale a été fixée à 500 et l'écart-type à 100.

## **2. Les performances comparées en « reading literacy » des élèves de CM1**

Dans le tableau qui suit, nous rendons compte des résultats de quelques pays dont les caractéristiques présentent un intérêt particulier pour nos analyses. Dans la première colonne, le score moyen de compréhension est indiqué. Ce score a été ventilé en quatre **niveaux** qui rendent compte de la progression de la difficulté de la tâche ; ces niveaux ont été ainsi définis en fonction des quartiles (niveaux 1 à 3, correspondant aux scores de compréhension atteints par 25, 50 ou 75 % de la population générale) et d'un repérage spécifique des 10 % ayant les scores les plus élevés (niveau 4). Pour faciliter les comparaisons, dans la première ligne du tableau, figurent les normes de référence.

Tableau 1- Répartition des scores moyens de compréhension selon les pays et des pourcentages d'élèves par niveau de compréhension atteint à PIRLS.

	Score moyen	Niveau 4 615 et +	Niveau 3 570 et +	Niveau 2 510 et +	Niveau 1 435 et +
<b>Résultats théoriques</b>	<b>500</b>	<b>10</b>	<b>25</b>	<b>50</b>	<b>75</b>
Suède	561	20	47	80	96
Pays - Bas	554	14	40	79	98
Angleterre	553	24	45	72	90
États-Unis	542	19	41	68	89
Italie	541	14	36	69	92
Canada Ontario	548	19	40	70	92
Canada Québec	537	11	31	67	94
<b>France</b>	<b>525</b>	<b>9</b>	<b>26</b>	<b>60</b>	<b>90</b>

La France a obtenu un score de compréhension (525) significativement supérieur à la moyenne des pays participants (500)<sup>5</sup>, quinze pays ont des performances significativement inférieures aux siennes. Comme dans tous les pays, les filles obtiennent des résultats supérieurs à ceux des garçons et cette différence se vérifie pour l'objectif littéraire comme pour l'objectif informatif. La répartition des pourcentages d'élèves en fonction des Niveaux atteints est beaucoup plus informative.

À chaque niveau, correspondent des tâches de lecture que les élèves sont capables d'accomplir avec une certaine réussite, montrant ainsi quelles compétences ils sont capables de mobiliser. Nous avons condensé les caractéristiques propres à chaque niveau pour essayer de comprendre les profils de résultats, et tout particulièrement celui des élèves français scolarisés en CM1. Ces caractéristiques reflètent le contenu de la variable processus de compréhension qui n'a pas été exploitée de manière directe. Des exemples d'items seront donnés par la suite pour chacun des niveaux.

### *2.1. Niveaux de compréhension atteints*

Au niveau 1, les élèves savent prélever des informations explicites et établir des inférences simples. Au niveau 2, ils maîtrisent le niveau précédent et de plus, ils sont capables d'effectuer des inférences et des interprétations simples à partir d'informations puisées dans différentes parties du texte. Ils comprennent

<sup>5</sup> La DEP a publié une note d'information qui synthétise les résultats français (Colmant et Mulliez, 2003).

la structure générale du texte. Au niveau 3, ils sont également capables d'effectuer des inférences qui s'appuient sur différentes caractéristiques des personnages et des événements, et ils savent les justifier. Ils interprètent en faisant appel à leurs connaissances et expériences personnelles. Ils comprennent des procédés tels que la métaphore simple. Enfin au niveau 4, ils peuvent aussi interpréter les intentions, les sentiments, les comportements des personnages en se basant sur le texte, ils sont capables d'intégrer des idées pour dégager le thème.

Si l'on examine les profils des pays, les pourcentages d'élèves atteignant le niveau 1 sont très élevés et ils dépassent toujours, et parfois de loin, le pourcentage définissant le niveau 2. Autrement dit, pour les pays concernés entre 60 % et 80 % des élèves sont capables de prélever des informations explicites, de réaliser des inférences et interprétations simples. Mais au niveau 3 et encore plus au niveau 4, les résultats français deviennent proches des pourcentages théoriques et ceux du Québec, seule autre contrée francophone de l'enquête, s'infléchissent aussi.

Le niveau 2 est dépassé par 45 % des élèves en Angleterre, 41 % aux États-Unis, en Ontario ou en Suède, mais par 26 % des élèves Français. Autrement dit, seulement un quart de nos élèves de CM1 dispose des capacités inférentielles et interprétatives définissant le niveau 3. Précisons que les inférences en jeu appartiennent à la catégorie des inférences élaboratives, les inférences dites nécessaires effectuées lors du traitement des marques anaphoriques n'étant pas évaluées dans le contexte international car toutes les langues ne comportent pas de telles marques (pour des définitions, Rémond, 2005).

## **2.2. Poids différentiel des objectifs et des processus de compréhension**

La dimension « processus de compréhension » présidant à la construction du protocole d'évaluation n'a pas vraiment été exploitée, contrairement à ce qui pouvait être attendu, notamment par les pays participants désireux d'établir des comparaisons avec PISA ; leur requête a été prise en considération (Mullis *et al.*, 2004). Mais les apports restent très limités car pour des raisons méthodologiques, les processus ont été regroupés par deux pour aboutir à deux échelles de compréhension : une échelle « prélever / inférer » et une échelle « interpréter / apprécier<sup>6</sup> ».

Pour chaque pays, on dispose des scores à chaque échelle et des tests statistiques entre ces scores. La Suède et la France obtiennent de meilleures performances à l'échelle « prélever / inférer » qu'à l'échelle « interpréter / apprécier », mais pour la France, cette différence n'est pas significative. La plupart des pays anglophones (Angleterre, USA, Ontario, Nouvelle-Zélande) se rejoignent en présentant les plus fortes différences significatives entre les deux échelles, en faveur de l'échelle « interpréter / apprécier ».

---

6 Ces dénominations simplifiées correspondent aux quatre processus détaillés en annexe.

La France, et dans une moindre mesure le Québec, a de bien meilleurs scores aux items évaluant l'objectif informatif qu'à ceux qui relèvent de l'objectif littéraire. Dans tous les pays anglophones et en Italie, l'objectif littéraire donne lieu à des performances significativement meilleures que l'objectif informatif. Enfin, les Pays-Bas et la Suède ont des performances similaires pour les deux objectifs de lecture.

**En résumé**, les pays anglophones peuvent être ainsi caractérisés : ils obtiennent leurs meilleurs résultats quand les supports sont de type « littéraire », quand les questions mettent en jeu les capacités d'intégration et d'interprétation et lorsque la réponse doit être exprimée par écrit. Quand leurs élèves doivent évaluer ce qu'ils ont lu, leurs performances sont bonnes et même excellentes pour les Anglais. Mais ces pays réussissent moins bien les items où il faut retrouver de l'information explicitement présente dans le texte, ce phénomène est plus marqué pour les QCM. Sainsbury *et al.* (2004, p. 20, N.T.) en concluent : « les élèves de ces pays semblent avoir développé des compétences complexes en matière d'interprétation, en même temps qu'ils ont perdu leurs capacités de repérage d'éléments explicitement présents dans les textes ».

La France affiche une toute autre image : ses élèves lisent mieux les textes « informatifs » que « littéraires », se montrent beaucoup plus à l'aise pour retrouver des informations ou faire des inférences simples que pour traiter les items sollicitant l'interprétation ou la « réflexion critique »<sup>7</sup>. Cette image entretient une certaine cohérence avec une évaluation-bilan française mettant en évidence qu'une proportion importante d'élèves en fin CM2 a des capacités limitées d'analyse et de synthèse. À l'évidence, ces capacités sont sollicitées dans la mise en œuvre de certains processus de compréhension de l'écrit (voir Rémond, sous presse).

### **2.3. Nos élèves et leurs stratégies de réponse à des items de l'unité souris**

Afin de les dépasser, nous allons illustrer les termes techniques tels que processus et niveaux par une contribution originale issue d'une réponse à un appel d'offres de la DEP portant sur l'évaluation PIRLS 2001. Nous avons analysé le contenu de réponses d'élèves aussi bien en fonction de variables propres au matériel que de la complexité des opérations à engager pour traiter les questions. Les exemples seront pris dans l'unité *Les souris*, libre d'accès, contrairement à d'autres mises en réserve pour d'autres phases du cycle PIRLS. Les pourcentages moyens de réussite à ces items ont été reportés pour la France et la moyenne des pays (moyenne internationale) dans le tableau 2 figurant en annexe. Celui-ci met en relief la contribution de processus différents à un même degré de difficulté ; ainsi, au niveau 2, l'un des items a été conçu pour évaluer « interpréter », l'autre pour « apprécier ». Ceci montre qu'un même processus a été testé par des tâches de différents degrés de difficulté, le type d'information à rechercher pour répondre variant sur un continuum allant du très concret au très abstrait. Les éléments d'analyse que nous présentons

---

7 Concerne le processus « Examiner et évaluer le contenu, la langue et les éléments textuels ».

portent sur des items isolés, et à ce titre ne peuvent être généralisés. De plus, nous ne nous appuyons que sur une seule épreuve *Les souris* qui contribue au score général, sans être l'épreuve la plus facile pour la France, ni la plus difficile.

Le niveau 1 de difficulté peut être illustré par l'item M 10 et l'item M7 classés respectivement « prélever des informations » et « inférer », : « M10 - Où M. Labon met-il les souris après les avoir ramassées ? M7 - Retrouve et recopie **une** des phrases qui expriment la panique ressentie par les souris la deuxième nuit. » Ces questions ouvertes, très bien réussies par nos élèves, font appel à une production d'écrit très limitée puisqu'il suffit de recopier le segment judicieusement sélectionné dans le texte.

Pour le niveau 2 de difficulté, nous avons choisi l'item M6, classé « interpréter », question ouverte à codage progressif et le QCM M13 classé « apprécier ».

M6 - La deuxième nuit, où les souris pensent-elles se trouver ? Que décident-elles de faire alors ?

La première question peut constituer une aide pour traiter la suite. Si les non-réponses de nos élèves sont peu nombreuses, il y a une production de 14 % de réponses inexactes, de 50 % de réponses complètes (2 points) traduisant l'interprétation attendue, résultat identique à la moyenne et de 31 % de réponses incomplètes.

M13- Il s'agit de choisir, parmi quatre propositions « Quelle phrase décrit le mieux cette histoire ? » et donc d'avoir identifié l'idée générale du texte (voir énoncé complet en annexe 2). L'examen du pattern de réponses est intéressant puisque les deux contrées francophones connaissent un pourcentage de réponse similaire et inférieur à la moyenne des pays, et bien en deçà de celui de l'Ontario (écart de 13 points).

L'item M11, classé « interpréter », est au niveau 3 de difficulté, sa structuration délimite bien la tâche : « Penses-tu que les souris ont été faciles à tromper ? Donne une raison. » Quelle que soit la réponse à la première question, une interprétation plausible doit ensuite s'y rattacher en la basant sur des éléments textuels appropriés. Nos élèves le réussissent assez bien, mais ils restent à 10 points de moins que les pays les plus performants.

Pour illustrer le niveau 4, trois exemples d'items ont été insérés. L'item M4 est réussi par la moitié des élèves des pays les plus performants (et des anglophones), le Québec fait moins bien que l'Ontario, mais encore beaucoup mieux que la France. Le voici : « Pourquoi M. Labon sourit-il en voyant qu'il n'y a pas de souris dans les pièges ? » Un point est accordé si une intégration d'idées présentées dans le texte est identifiable dans la réponse. Par exemple, le fait que M. Labon n'est pas surpris, ou une indication de ses intentions pour mener un plan plus élaboré pour attraper les souris.

L'item M8 fait partie de ceux qui nécessitent « d'examiner et d'évaluer le contenu, la langue et les éléments textuels » ; il porte plus précisément sur la reconnaissance d'éléments de techniques narratives.

M8 - Comment l'histoire t'apprend-elle ce que les souris pensent de ce qui arrive ?

- A- En disant ce que M. Labon pense des souris.
- B- En décrivant où vivent les souris.
- C- En disant ce que les souris se racontent.
- D- En décrivant les souris.

Son énonciation complexe constitue une première barrière à la compréhension de la question. Nos élèves identifient la réponse correcte à ce QCM moins souvent que la moyenne des pays, mais comme les Québécois, ils ont une forte attirance pour la première modalité énoncée, la réponse A (environ 27 % de réponses). Le verbe « pense » a probablement favorisé un comportement d'appariement direct entre cette réponse A et des éléments de l'item. Dans la version anglaise, le terme "thought" est présent à la fois dans la question et dans la modalité de réponse A, choisie seulement par 14 % des petits Anglais.

L'item M12 classé « interpréter » se caractérise par un énoncé long incluant des consignes enchaînées et concises. Les opérations à mener pour traiter cet item sont complexes, ce qui lui vaut de recevoir un codage progressif pouvant atteindre jusqu'à 3 points (voir annexe) : « Les actions de M. Labon qui sont racontées dans le texte nous informent sur son caractère. Décris ce caractère et donne **deux** exemples de ses actions. »

Les élèves français ont particulièrement échoué à cet item. Plus de 20 % d'entre eux n'a pas donné de réponse ; un tiers a donné une réponse erronée qui se focalise sur des actions, manifeste une compréhension inexacte du caractère de Labon ou ne le décrit pas. On ne peut exclure que le terme « caractère » ait été trop abstrait pour certains élèves qui ne le relient à aucun contenu du texte. Quand seule une réponse à la première partie de la question est pertinente, les élèves sont alors déclarés faire preuve d'une « compréhension limitée » et sont crédités d'un seul point. La plupart du temps, contrairement aux élèves qui obtiennent les scores de 2 ou 3 points, ils n'ont pas dégagé un trait de caractère pour lui associer une qualification et se contentent de prélever dans le texte des expressions comme « calme et paisible », « pauvre et heureux » (lignes 1 et 2). Dans le premier exemple, les deux adjectifs se rapportent effectivement au caractère de Labon. Dans le second, les élèves donnent l'impression d'avoir « ramené dans leur filet » des éléments qu'ils ne trient pas en fonction du critère de réponse.

Nous allons dépasser ces exemples de l'unité *Les souris* pour tirer quelques généralités sur le comportement de nos élèves à partir de notre expertise des réponses.



### **3. Comment se caractérisent nos élèves ?**

Nous avons déjà indiqué les tendances générales de résultats de nos élèves, nous allons maintenant montrer leur investissement dans les réponses aux questions, leur grande dépendance à l'habillage des items et leur manque d'autonomie dans les tâches.

#### **3.1. Stratégies de réponses et contournement de la production de réponses écrites**

De manière générale, les items demandant de rédiger une réponse par écrit sont moins bien réussis que les QCM. Comme dans PISA et plus que ceux d'autres pays, nos élèves ne répondent pas à un certain nombre d'items pour lesquels ils devaient rédiger une réponse par écrit, ce qui interroge (pour une discussion, voir Rémond, 2005). Selon le degré d'autonomie à engager dans la réponse, les résultats fluctuent.

#### **Une tâche limitée de rédaction de la réponse.**

Quand la tâche de rédaction est limitée et s'apparente à un prélèvement d'information (cf. recopier), même si des opérations cognitives complexes sont en jeu, leurs performances deviennent supérieures à celles de la moyenne des pays. Ils deviennent, par exemple, capables d'établir des liens entre cause et conséquence, ce qui est considéré comme une opération mentale de haut niveau.

**La présence d'une trame de réponse**, comportant éventuellement des organisateurs textuels tels que « parce que », permet aux élèves de réaliser de bien meilleures performances que lorsqu'ils doivent totalement gérer la tâche ; ils parviennent même à dégager des explications causales. Lorsque la tâche est fractionnée en étant présentée sous forme de questions successives, ils présentent la même aisance. Souvent la première question s'apparente à un prélèvement d'information et conduit à « recopier » un exemple, une phrase (voir item M6). Ces formes de « questions-ébauches de réponses » limitent leur planification globale de la tâche globale en leur fournissant un guidage qui jalonne les étapes de recherche d'information, et la construction écrite de la réponse est peu coûteuse. Cette forme d'étayage se répercute souvent vers les réponses aux questions suivantes où les élèves adoptent une trame parallèle de réponses qui peut se révéler efficace.

Dans les cas qui viennent d'être évoqués, leur charge cognitive est moins élevée, leur travail est plus sécurisant et probablement plus proche de celui de la classe habituelle (voir supra). Cependant, cette facilitation n'est pas valable pour toutes les questions, et certaines questions ouvertes reposent, à la fois, sur des inférences de haut niveau et sur des textes plus complexes et plus longs. On observe que parfois alors qu'il fallait glaner des informations au fil du texte pour construire la réponse, les élèves opèrent un traitement local du texte. Ceci explique notamment leurs faiblesses face à des tâches où l'interprétation est requise.

PIRLS n'évalue ni la production d'écrit, ni les qualités orthographique et syntaxiques et les meilleures réponses construites, recevant un score de 2 ou 3 points, ne sont pas exemptes de problèmes en relevant. Il faut remarquer ces réponses à des questions se situant aux niveaux 3 et 4 : elles sont structurées, comportent souvent des connecteurs, prennent appui sur des éléments de la consigne, sorte de mots-clés, servant d'indicateurs de traitement aux élèves qui les réinvestissent dans la rédaction de leur réponse, par exemple « au début », « à la fin », « la raison du changement ». Ces réponses reflètent de grandes qualités de traitement de l'écrit conduisant à une excellente compréhension. La distance entre ces élèves et d'autres plus faibles témoigne de la très grande hétérogénéité de la population scolarisée en CM1.

La fluctuation des performances en fonction de la structuration des items semble indiquer une très grande dépendance à l'égard des habitudes scolaires et un manque certain d'autonomie dans le traitement des tâches sous-jacentes, qui s'explique, entre autres, par des défauts métacognitifs.

### ***3.2. Défaillances métacognitives***

De nombreux élèves ne gèrent pas suffisamment la tâche et en tout premier lieu n'attachent pas assez d'importance à la question qui fixe leur tâche. Ils traitent partiellement la consigne, s'emparent d'un de ses éléments qu'ils s'efforcent d'apparier, de préférence, à un élément semblable ou synonyme du texte. Cette mise en correspondance assez systématiquement observée chez certains amène à s'interroger sur leur représentation rigide des tâches de compréhension et sur les pratiques de classe auxquelles ils sont soumis. Parfois, ils isolent le bon maillon du texte porteur de l'information requise, mais ne savent pas en tirer parti car ils ne tiennent pas compte de la précision de la consigne. Il faut s'interroger sur les aspects métacognitifs de leur compréhension : ils ne confrontent pas leur réponse à la consigne pour en éprouver la pertinence.

L'établissement trop systématique d'une correspondance directe entre la question et le texte a conduit dans des impasses, car il fallait parfois mener des opérations cognitives plus complexes pour aller justement au delà du texte. Cette mise en correspondance, assez directe et systématique, entre des éléments de la question et du texte amène à s'interroger sur les modèles de tâches qu'ils rencontrent dans leur vie de classe et de manière associée à leur représentation de la lecture. Sans en avoir une évaluation directe, des représentations erronées de la lecture sont sous-jacentes comme par exemple : répondre à une question, c'est trouver directement l'information dans le texte. A l'évidence, certains ne savent pas ce qu'ils cherchent, en témoignent les réponses invalides et leur contenu parfois surprenant au regard de la tâche. Une représentation restrictive et erronée des tâches de lecture a été construite par certains qui ne s'en éloignent pas, contrairement à d'autres élèves. Cette représentation témoignant de connaissances métacognitives insuffisantes a été mise en évidence dans diverses recherches, comme celle d'un contrôle métacognitif insuffisant de l'activité de lecture (Rémond, 2003). D'autre part, les élèves ont travaillé en autonomie totale pendant deux fois une demi heure,

l'adulte n'a assuré aucune régulation de l'activité et de surcroît, les demandes ont parfois dû surprendre.

### **3.3. Une demande inhabituelle : exercer une « réflexion critique »**

Si les résultats à PIRLS dépeignent les capacités de compréhension de nos élèves, ils mettent aussi l'accent sur des demandes inhabituelles pour eux comme « examiner et évaluer le contenu, la langue et les éléments textuels » (résumé par « réflexion critique »). À cet égard, comme pour PISA, les items évaluant la « réflexion critique » dressent un écueil pour eux. Nous nous sommes demandé si certains pays favorisent « la réflexion critique » dans leurs Programmes scolaires. Pour répondre à cette question, nous avons exploré les données disponibles sur les Programmes scolaires des pays participant à PIRLS (Mullis *et al.*, 2002).

## **4. Quelques pistes d'explication des performances**

### **4.1. Contenu des Programmes d'enseignement**

La consultation de cet ouvrage révèle que la « réflexion critique » est inscrite dans les programmes d'enseignement de la lecture (*Reading curriculum and standards*) de tous les pays de langue anglaise, ainsi que de la Suède. Au delà des programmes, « la réflexion critique » fait l'objet d'évaluations dans des dispositifs centralisés, comme les évaluations britanniques (NCRT, *National Curriculum Reading Test in England*), où des compétences liées aux activités « réflexives » sont évaluées tous les ans, dès l'âge de 7 ans (fin KS 1, Key-Stage, étapes clés<sup>8</sup>). Aux États-Unis, la NAEP (*National Assessment of Educational Progress*) les prend en compte à l'évaluation effectuée en grade 4 (10 ans).

Quand la dimension de « réflexion critique » est intégrée dans les pratiques de l'enseignement scolaire, elle se reflète visiblement dans leurs performances aux évaluations internationales, un pôle linguistique lié à la langue maternelle anglaise étant facilement identifiable.

Des données provenant de questionnaires sont également disponibles sur les pratiques pédagogiques des maîtres des élèves passant PIRLS (pour des détails, Rémond, sous presse).

### **4.2. Les pratiques pédagogiques rencontrées par les élèves de CM1**

Nous rendons brièvement compte des réponses aux questions concernant la maîtrise de la langue. Celles-ci mettent en évidence un certain nombre d'éléments propres à la France en termes d'exposition à l'écrit au niveau CM1 exprimés dans la suite par rapport à la moyenne des autres pays. Le temps consacré à l'enseignement de la lecture est moins important en France et nos élèves seraient moins exposés à des activités de nature à développer leur compréhension de l'écrit : le « déficit » est particulièrement visible pour le

---

8 Pour des précisions, on peut consulter Céard *et al.*, 2003, p. 110-111, et Sainsbury *et al.*, 2006.

travail de nature inférentielle et pour les activités de mise en perspective avec des lectures antérieures ou avec l'expérience personnelle. Le matériel est très rarement adapté en tenant compte du profil des élèves. Les activités se révèlent peu variées tout comme les supports sur lesquelles elles portent. Les éléments apportés sur les modalités de questionnement après la lecture sont éclairants : le questionnement oral représente la pratique majoritaire, et parfois exclusive en France, les questions nécessitant une réponse exprimée par écrit sont peu fréquentes et les échanges verbaux portant sur la compréhension des textes sont plutôt rares.

Ce panorama basé certes sur des connaissances déclaratives ne peut laisser indifférent. Selon les résultats à PIRLS, les capacités inférentielles semblent faire plus souvent défaut à nos élèves qu'à d'autres, mais sont-elles suffisamment exercées dans nos classes, alors qu'elles jouent un rôle central dans l'élaboration de la signification des textes ? Les maîtres déclarent travailler de manière prépondérante sur des textes de fiction, mais le traitement des aspects littéraires des textes paraît moins efficace en France que pour la moyenne des pays. Certains des items qui contribuent essentiellement aux niveaux de compétences les plus élevés font appel au point de vue, à la reconnaissance de techniques narratives, ou encore à l'évaluation de la plausibilité d'une histoire, comment sont-ils abordés en France ?

N'a-t-on pas parfois évalué la familiarité des élèves à certains types de connaissances et de tâches plus que leur compréhension de l'écrit ? De là, il est facile de passer aux aspects inter-langues.

#### **4.3. Les effets inter-langues**

La supériorité des anglophones à divers items nous a fait avancer l'hypothèse d'un biais de traduction en faveur des pays où l'anglais est la langue d'enseignement, avis d'ailleurs partagé par Sainsbury *et al.* (2004) lors d'un travail comparatif sur les profils de résultats dans plusieurs langues. Notons que le matériel a été élaboré en anglais dans son équipe en Angleterre (NFER) et qu'il présente une forte parenté avec les évaluations annuelles mises en place dans l'ensemble du Royaume-Uni : des textes intégraux, assez longs, assortis d'un nombre assez important de questions. Les pays anglophones sont les seuls à avoir travaillé sur des versions du matériel qui n'ont pas été traduites, le matériel ayant été élaboré en anglais. Celui-ci a été traduit pour les autres pays dans leur langue nationale, et malgré tous les soins entourant les traductions, il n'a pas toujours été possible de restituer les nuances de la version originale dans toutes les langues pour des raisons sémantiques, grammaticales, etc. À noter que la France et le Québec ont travaillé chacun sur la traduction réalisée dans leur pays, en interaction avec l'organisme administrant localement PIRLS afin de vérifier que le matériel restait adapté aux élèves (vocabulaire, syntaxe, règles de présentation des questions...).

On s'attendait à ce que des profils similaires de réponses aux quatre modalités d'un même QCM se dessinent, sans effet de langue ; mais de temps en

temps, on observe des patterns de réponses différents selon les langues aux quatre modalités de réponse d'un même QCM. Ce phénomène est particulièrement facile à observer dans le cas du Canada où l'on constate pour nombre d'items un écart significatif et parfois important entre les résultats des deux provinces canadiennes, les résultats de l'Ontario étant proches de ceux des pays anglophones, ceux du Québec entretenant une forte similitude avec les résultats français. La plausibilité des distracteurs peut évoluer lors du passage à une autre langue et dans les deux Programmes, on a pu observer que certains distracteurs, qui ne l'étaient pas en anglais, sont devenus parfaitement plausibles en français.

Au hasard de notre travail, des éléments nous ont intriguée. Ainsi, le QCM M3 classé au niveau 3 élève la France à la première place avec 75 % de bonnes réponses en France, contre 53 % pour la moyenne internationale. L'Angleterre (59 % de réussite) travaille sur le matériel initial, sans traduction bien sûr. Dans le texte de la France, on lit : « elles croient à une bonne blague » et dans le QCM, modalité B « Elles trouvent que M. Labon a fait une bonne blague » (voir Annexe). Dans la version anglaise du texte, on lit : « *they thought it was a tremendous joke* » et pour la modalité B : « *they thought Labon had done something silly.* »

Il nous semble que le niveau d'inférences ne paraît pas être le même en anglais et en français. Dans la version française, l'appariement entre le texte et la modalité s'établit sans difficulté par une correspondance directe entre la question et le texte ; cela s'apparente davantage à un prélèvement d'information qu'à une inférence, alors que dans la version anglaise, le traitement requis est plus complexe, correspondant sans doute à l'inférence directe déclarée être évaluée par cet item. En outre, le distracteur<sup>9</sup> A attire beaucoup les élèves de certains pays ; ce distracteur, situé en première position dans le QCM, peut paraître assez plausible à des élèves qui se réfèrent au deuxième épisode de l'histoire au lieu du premier.

Autre item moins favorable à la France, cette fois, l'item M12 déjà présenté. Il a été très nettement mieux réussi par la plupart des pays anglophones, ce qui fait s'interroger sur sa traduction en français et en particulier celle de la phrase « *give two examples of what he does that show this.* » L'information apportée est beaucoup plus précise et explicite que dans la version soumise aux jeunes Français et elle souligne davantage le lien qu'il faut établir entre le caractère du personnage Labon et les actions qui permettent de dégager le trait de caractère.

Examinons maintenant un exemple particulièrement riche tiré de l'épreuve *Le lion et le lièvre* : « Vers la fin de l'histoire, le lion essaie de rassurer le lièvre. Donne **deux** exemples de la façon dont il s'y prend. » Parfois, des élèves prennent un exemple pour chacun des deux personnages mentionnés dans

---

9 La plausibilité des distracteurs est une variable spécifique aux QCM. Ce format d'item se caractérise par plusieurs modalités de réponse : la bonne réponse et les autres qui s'appellent distracteurs.

la consigne. La présence du pronom « il » nous a amenée à nous interroger sur son traitement. Il n'est pas impossible que par une stratégie de distance minimale, il ait été interprété comme le substitut du personnage « le lièvre ». Ce traitement défectueux d'une marque anaphorique met en lumière un problème introduit par la traduction, alors que les évaluations internationales ne portent pas sur le traitement des marques anaphoriques. Voici la question en anglais : « Write *two ways in which the lion tried to make the hare feel better at the end of the story.* » Aucune ambiguïté ne plane, ni sur la tâche, ni sur l'acteur concerné : le lion. Dans cette question, la « règle de l'ordre d'importance » est respectée : les informations présentées dans la question sont hiérarchisées et la tâche à réaliser est indiquée en premier lieu (voir Rémond, 2001). La version anglaise centre sans ambiguïté la question sur le lion alors que la version française a perdu cette caractéristique et introduit un biais. Cet exemple démontre remarquablement la complexité des phénomènes que fait émerger la non équivalence de la consigne dans ces deux langues.

Enfin, il faut noter des performances très contrastées à deux épreuves : la France se place en 10<sup>e</sup> position pour l'unité *Le lion et le lièvre* et à l'inverse, pour une autre, elle n'occupe que le 26<sup>e</sup> rang. Pour ces deux épreuves, le Québec obtient des performances répliquant celles de la France. Autrement dit, et malgré les précautions prises dans les procédures de pré-test, il existe une certaine variabilité des performances selon le support, et les effets liés à la langue et à la culture ne sont pas gommés.

L'éternel problème des comparaisons internationale reste posé : évaluons-nous des opérations de même nature au travers des diverses épreuves ?

## Conclusion

L'évaluation PIRLS a mis en exergue les performances moyennes des élèves français en matière de compréhension de l'écrit en CM1 et des facteurs qui la font varier. Quand les écrits qu'ils doivent traiter ressemblent à ce qu'ils connaissent, leurs performances sont remarquables. On citera l'exemple d'une épreuve de l'objectif informatif qui se fonde sur un dépliant publicitaire pour laquelle la France se classe 3<sup>e</sup>. On regrettera le manque de plasticité de leurs compétences au moment de les allouer à des tâches moins habituelles qu'ils ont alors beaucoup de mal à gérer, leur production de non-réponses et dans l'ensemble, un contrôle insuffisant de l'exécution des tâches. L'effet des comparaisons internationales est d'éclairer des comparaisons. Celles-ci ont montré que nos élèves sont beaucoup plus à l'aise avec des écrits « informatifs » et qu'ils doivent retrouver des informations ou faire des inférences simples. En revanche, et cela explique en partie les plus faibles pourcentages d'élèves français accédant aux niveaux 3 et 4, ils sont moins performants que ceux d'autres pays pour traiter des items sollicitant l'interprétation ou la « réflexion critique ». Nous avons pu relier cela aux pratiques déclarées par les maîtres : au

delà de la situation d'une population par rapport à d'autres, PIRLS met l'accent sur la manière de travailler en classe.

Les résultats à PIRLS laissent penser que l'empan des tâches proposées par l'école à nos élèves demeure trop étroit : ceux-ci sont déconcertés par des questions complexes, successives qu'il faut traiter sans guidage du maître. Les élèves se sont trouvés seuls en présence de textes longs et suivis de questions très variées. Certains peinent à dépasser le modèle habituel des tâches qu'ils vivent en classe et ont même une tendance affirmée à vouloir ramener les problèmes posés à cet éternel modèle.

Au risque de répéter ce que nous avons déjà dit ailleurs, il faut vraiment s'interroger sur les effets des questionnaires de lecture que l'on rencontre dans les classes, provenant de manuels ou directement fabriqués par les enseignants (Beltrami *et al.*, 2004 ; Tauveron, 2002). Trop souvent, même dans les plus récents, les questions visent une compréhension de surface et la réponse attendue reste limitée : celle-ci s'énonce sous forme d'un mot ou d'une expression. La maigreur de ce questionnement ne construit pas un lecteur actif, elle alimente au contraire des représentations erronées de la lecture (Guernier, 1999 ; Rémond, 2003).

Au contraire, la compréhension des textes se travaille en engageant les élèves à traiter des questions qui mettent en jeu des activités de raisonnement, un contrôle métacognitif pour s'assurer de la qualité de sa lecture (voir plusieurs contributions dans Gaonac'h et Fayol, 2003). Il faut apprendre à gérer les indices présents dans le texte, sans toutefois se laisser « entraîner, abuser » par des indices de surface et parfois gérer simultanément plusieurs paramètres. Il faut dépasser l'idée que la réponse est « dans sa tête » ou dans le texte et mobiliser ses connaissances « passives » pour les mettre au service de la tâche à réaliser. Cela suppose d'avoir constitué une palette de stratégies cognitives et métacognitives. C'est bien l'image d'un lecteur actif dans la construction de la représentation mentale du contenu du texte qui prévaut.

### **Bibliographie**

- BELTRAMI, D. & QUET, F. (2002) : « Lecture : l'espace d'un problème », *Le Français Aujourd'hui* n° 137, p. 57-71.
- CAMPBELL, J. *et al.* (2000) : *Framework and Specifications for PIRLS Assessment 2001-Progress in International Reading Literacy Study*; IEA, (2<sup>e</sup> édition) Chestnut Hill, MA, Boston, College.
- CEARD, M.-T., REMOND, M., VARIER, M. (2003) : *L'appréciation des compétences des élèves et des jeunes en lecture et écriture et l'évolution de ces compétences dans le temps*. Les rapports du haut conseil de l'évaluation de l'école, n° 15. Disponible sur Internet : <http://cisad.adc.education.fr/hcee>
- COLMANT, M. & MULLIEZ, A. (2003) : *Les élèves de CM1. Premiers résultats d'une évaluation internationale en lecture (PIRLS)*. Note d'Information, 03.22.

- GAONACH'H, D. & FAYOL, M. (éd.) (2003) : *Aider les élèves à comprendre. Du texte au multimédia*. Paris, Hachette-Collection Profession Enseignant.
- GUERNIER, M.-C. (1999) : « Interactions verbales et construction de sens. Lectures d'un texte narratif », *Repères*, n° 19, p. 167-182.
- MULLIS, I. V. S., MARTIN, M. O., GONZALES, E. & KENNEDY, A. (2003) : *PIRLS 2001 International Report*. Boston, International Study Center.
- MULLIS, I. V. S., MARTIN, M. O. & GONZALES, E. (2004) : *International achievement in the processes of reading comprehension. Results from PIRLS 2001 in 35 countries*. Chestnut Hill, MA, Boston College.
- OCDE (1999) : *Measuring student knowledge and skills. A new framework for assessment*. Paris, OCDE.
- RÉMOND, M. (2001) : « Adapter n'est pas traduire : Adaptation dans différents contextes culturels d'épreuves d'évaluation de la littéracie », in C. Sabatier et P. Dasen (éd.), *Cultures, développement et éducation. Autres enfants, autres écoles*. Paris : L'Harmattan.
- RÉMOND, M. (2003) : « Enseigner à comprendre : les entraînements métacognitifs », in D. Gaonach'h & M. Fayol (éd.), *Aider les élèves à comprendre. Du texte au multimédia*, Paris, Hachette-Collection Profession Enseignant.
- RÉMOND, M. (2004) : « Pourquoi enseigner la compréhension au cycle 3 ? » in ONL (coll.) *Nouveaux regards sur la lecture*, CNDP-Savoir-Livre.
- RÉMOND, M. (2005) : « Regards croisés sur les évaluations institutionnelles », *Repères*, n° 32, p. 113-140.
- RÉMOND, M. : « Éclairage des évaluations internationales PIRLS et PISA sur les élèves français », *Revue française de pédagogie*, n° 157, p. 71-84, INRP.
- SAINSBURY, M., SCHAGEN, I. & HAMMOND, P. (2004) : *What did PIRLS tell us about reading skills?* Conférence de l'International Association for Educational Assessment, Philadelphia, Juin 2004.
- TAUVERON, C. (éd.), (2002) : *Lire la littérature à l'école. Pourquoi et comment conduire cet apprentissage spécifique ? de la GS au CM*, Paris, Hatier-Pédagogie.



## **Annexe 1 : contenu des quatre processus à évaluer**

Le contenu est brièvement décrit ci-dessous ; entre parenthèses, figure le nom sous lequel chacun sera désigné par la suite.

**Prélever des informations explicites** (Prélever) 20 % des items : repérer les informations directement liées à l'objectif de la lecture, chercher des informations précises, repérer le contexte de l'histoire, retrouver l'idée principale, etc.

**Faire des inférences directes** (Inférer) 40 % des items : déduire l'élément principal d'une série d'arguments, ou quel événement en a entraîné tel autre, décrire la relation entre deux personnages, etc..

**Interpréter et assimiler idées et informations** (Interpréter) 25 % des items : déduire le message global ou le thème d'un texte, envisager une alternative aux actions des personnages, saisir l'atmosphère ou le ton du récit, etc..

**Examiner et évaluer le contenu, la langue et les éléments textuels** (Apprécier) 15 % des items : évaluer la probabilité que les événements décrits se passent réellement, décrire la manière dont l'auteur a amené la chute, juger de la clarté des informations fournies dans le texte, décrire comment le choix des adjectifs modifie le sens.

## **Annexe 2 : épreuve *Les souris***

### **Les souris à l'envers**

Il était une fois un vieil homme de 87 ans qui s'appelait Labon. Toute sa vie, il avait été calme et paisible. Il était à la fois très pauvre et très heureux.

Quand un jour M. Labon découvre qu'il y a des souris dans sa maison, il ne s'en inquiète pas trop au début. Mais les souris se multiplient. Elles commencent à le tracasser. Elles continuent tellement à se multiplier que M. Labon ne peut plus les supporter.

« C'en est trop. », se dit-il. « Cela va vraiment un peu trop loin. » Il sort de chez lui et se rend en clopinant jusqu'au magasin pour acheter des pièges à souris, un morceau de fromage et de la colle.

De retour à la maison, il met de la colle sous les pièges et les fixe au plafond. Puis il dispose soigneusement quelques morceaux de fromage sur les pièges ouverts.

Cette nuit-là, lorsque les souris sortent de leurs trous et voient les pièges au plafond, elles croient à une bonne blague. Elles se promènent sur le plancher, se donnent des coups de coude et montrent le plafond avec leurs pattes avant en se tordant de rire. Après tout, c'est plutôt marrant, des pièges à souris au plafond.

Quand M. Labon descend le lendemain matin, il constate qu'aucune souris n'est prise au piège. Il sourit en silence...

Il saisit alors une chaise, verse de la colle sous les pieds et la fixe à l'envers au plafond, à côté des pièges. Il fait la même chose avec la table, le téléviseur et la lampe. Finalement, il prend tout ce qui est au sol et le colle au plafond. Il y ajoute même un petit tapis.

Cette nuit-là, les souris sortent de leurs trous en ricanant et en faisant des plaisanteries sur ce qu'elles ont vu la veille. Mais cette fois, quand elles regardent au plafond, elles arrêtent de rire brusquement.

« Hé ! Regardez ! Voilà que le sol est là-haut ! », s'écrie l'une d'elles.

« Incroyable ! Alors, nous devons être au plafond ! », s'exclame une autre.

« Je commence à me sentir un peu étourdie », dit une autre.

« Le sang me descend à la tête », se plaint une quatrième.

« C'est horrible ! », dit une très vieille souris aux longues moustaches. « C'est vraiment horrible ! Il faut faire quelque chose tout de suite ! »

« Je vais m'évanouir si je dois me tenir une seconde de plus sur la tête ! », crie une jeune souris.

« Moi aussi ! »

« Je n'en peux plus ! »

« Au secours ! Vite, que quelqu'un fasse quelque chose ! »

Elles devenaient hystériques. « Je sais ce que nous allons faire », dit la très vieille souris. « Nous allons toutes nous tenir sur la tête et alors nous serons dans le bon sens. »

Docilement, elles se placent toutes sur la tête et au bout d'un long moment, le sang coulant vers leur cerveau, elles s'évanouissent l'une après l'autre.

Quand M. Labon descend le lendemain matin, le sol est couvert de souris. Il les ramasse rapidement et les met dans un panier.

Voici ce qu'il faut retenir de cette histoire : chaque fois que le monde semble à l'envers, mieux vaut rester les pieds sur terre.

« M 3 - Pourquoi, la première nuit lorsque les souris sortent de leurs trous, se donnent-elles des coups de coude et montrent-elles le plafond avec leurs pattes avant ?

A- Elles voient une chaise au plafond.

B- Elles trouvent que M. Labon a fait une bonne blague.

C- Elles veulent le fromage qui se trouve dans les pièges.

D- Elles ont peur de ce qu'elles voient.

M 4 - Pourquoi M. Labon sourit-il en voyant qu'il n'y a pas de souris dans les pièges ?

M 6 - La deuxième nuit, où les souris pensent-elles se trouver ? Que décident-elles de faire alors ?

M 7 - Retrouve et recopie **une** des phrases qui expriment la panique ressentie par les souris la deuxième nuit.

M 8 - Comment l'histoire t'apprend-elle ce que les souris pensent de ce qui arrive ?

A- En disant ce que M. Labon pense des souris.

B- En décrivant où vivent les souris.

C- En disant ce que les souris se racontent.

D- En décrivant les souris.

M 10 - Où M. Labon met-il les souris après les avoir ramassées ?

M 11 - Penses-tu que les souris ont été faciles à tromper ? Donne une raison.

M 12 - Les actions de M. Labon qui sont racontées dans le texte nous informent sur son caractère. Décris ce caractère et donne **deux** exemples de ses actions.

M 13 - Quelle phrase décrit le mieux cette histoire ?

A- Elle est sérieuse et triste.

B- Elle est effrayante et excitante.

C- Elle est amusante et ingénieuse.

D- Elle est palpitante et mystérieuse. »

### **Annexe 3 : exemple de codage progressif de l'item M12**

Le libellé de l'item M12 est le suivant :

« Les actions de M. Labon qui sont racontées dans le texte nous informent sur son caractère. Décris ce caractère et donne **deux** exemples de ses actions. »

Cette question contribue à l'évaluation du processus « interpréter », la réponse vaut 3 points si elle manifeste une compréhension approfondie, en ayant intégré les idées contenues dans le texte pour pouvoir apporter une interprétation du caractère de M. Labon.

Le guide de codage définit ainsi l'attribution des points :

3 points, il faut avoir indiqué au moins un trait de caractère et le justifier par deux exemples d'actions issus du texte.

2 points, il faut avoir décrit un trait de caractère et le justifier par un exemple d'action.

1 point, il faut avoir décrit un trait de caractère, sans mention pertinente d'une action qui le soutient.

0 point, si la réponse indique une compréhension erronée du caractère de Labon.

Le score obtenu traduit le degré de compréhension atteint : la réponse à 3 points montre que l'élève « a intégré », qu'il « a interprété » (2 points), que sa compréhension est limitée (1 point). Ici, les réponses à 1 ou à 2 points correspondent à des scores intermédiaires.

Voici des exemples de réponses à cet item :

*Il est très ingénieu il est aussi très clame car quand il voient que son premier plan n'a pas marcher il reste très clame est ne s'énerve pas. Il très ingénieu de collé tout se qu'il a au sol au plafond (3 points)*

*il est malin parce qu'il colle des meuble au plafond pour faire croire au souris qu'elle sont au plafond et il mets des piège au plafond pour les attraper (2 points)*

*M. Labon et calme et paisible et nerveux. Il va dans un magasin. Colle la chaise au plafon (1 point)*

*Il a pensé de métre la chèse au plafon et son carètère il est gentil (1 point)*

*Il a mauvais caractère. Il fait des pièges aux souris et il se rend au magasin en clopinant (0 point)*

#### Annexe 4

**Tableau 2 – Pourcentages moyens de réussites à des items de l'épreuve *Souris*, en fonction du niveau et du processus auxquels appartient l'item, pour les Français et l'ensemble de la population (moyenne internationale).**

Niveaux	N° et type de Questions*	Processus	Pourcentages moyens de réussite	
			Français	Internationaux
Niveau 1	M10 / RC M7 / RC	Prélever Inférer	93, 3 81	84 69
Niveau 2	M6 / RC M13 /QCM	Interpréter Apprécier	50 (2pts) / 31 (1pt) 63	50 (2pts) / 27 (1pt) 68
Niveau 3	M3 / QCM M11 / RC	Inférer Interpréter	76 50	53 37
Niveau 4	M4 / RC M8 / QCM M12 / RC	Interpréter Apprécier Interpréter	35 43 8 (3pts), 14 (2pts), 22 (1pt)	31 46 10 (3pts), 20 (2pts), 31 (1pt)

\* RC : question à réponse construite.