

Élaboration d'une épreuve de compréhension en lecture en 6^e année du primaire favorisant l'articulation enseignement-apprentissage-évaluation

Catherine Turcotte and Nadine Talbot

Volume 40, Number 3, 2017

URI: <https://id.erudit.org/iderudit/1048910ar>
DOI: <https://doi.org/10.7202/1048910ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)
2368-2000 (digital)

[Explore this journal](#)

Cite this article

Turcotte, C. & Talbot, N. (2017). Élaboration d'une épreuve de compréhension en lecture en 6^e année du primaire favorisant l'articulation enseignement-apprentissage-évaluation. *Mesure et évaluation en éducation*, 40(3), 37–67. <https://doi.org/10.7202/1048910ar>

Article abstract

Few reading comprehension tests can be used to inform 6th grade elementary teachers on students' knowledge and abilities, but also on their learning needs. The aim of this article is to present the conception and relevance of an assessment based on theoretical models of reading comprehension and developed according to DeVellis' (2003) steps. These steps are described in this article, in addition to the identification of the assessment's psychometric qualities and its potential to evaluate reading comprehension according to specific processes.

Élaboration d'une épreuve de compréhension en lecture en 6^e année du primaire favorisant l'articulation enseignement-apprentissage-évaluation

Catherine Turcotte

Nadine Talbot

Université du Québec à Montréal

Mots clés : évaluation en lecture, compréhension en lecture, élèves du primaire, modèle de Rasch, articulation enseignement-apprentissage-évaluation

Peu d'épreuves de compréhension en lecture peuvent informer les enseignants de la 6^e année du primaire à l'égard des connaissances et habiletés des élèves, mais aussi de leurs besoins d'apprentissage. L'objectif de cet article est donc de présenter l'élaboration et la pertinence d'une épreuve qui s'appuie sur des modèles théoriques de la compréhension en lecture et qui a suivi les étapes de conception d'un test suggérées par DeVellis (2003). Les étapes parcourues pour concevoir cette épreuve sont décrites dans cet article, en plus de la détermination de ses qualités psychométriques et de son potentiel pour évaluer la compréhension en lecture selon des processus précis.

Key words: reading assessment, reading comprehension, elementary school students, Rasch model, teaching-learning-assessment cycle

Few reading comprehension tests can be used to inform 6th grade elementary teachers on students' knowledge and abilities, but also on their learning needs. The aim of this article is to present the conception and relevance of an assessment based on theoretical models of reading comprehension and developed according to DeVellis' (2003) steps. These steps are described in this article, in addition to the identification of the assessment's psychometric qualities and its potential to evaluate reading comprehension according to specific processes.

Palavras-chave: avaliação da leitura, compreensão da leitura, alunos do ensino básico, modelo de Rasch, articulação ensino-aprendizagem-avaliação

Poucos testes de compreensão da leitura conseguem informar os professores do 6º do ensino básico sobre os conhecimentos e as habilidades dos alunos, mas também sobre as suas necessidades de aprendizagem. O objetivo deste artigo é, portanto, apresentar a elaboração e a pertinência de um teste que se baseia em modelos teóricos de compreensão da leitura e que seguiu as etapas da concepção de um teste sugerido por DeVellis (2003). As etapas percorridas para conceber este teste são descritas neste artigo, bem como a determinação das suas qualidades psicométricas e o seu potencial para avaliar a compreensão da leitura segundo processos específicos.

Note des auteurs: La correspondance liée à cet article peut être envoyée à [turcotte.catherine@uqam.ca] et [talbot.nadine@uqam.ca].

Les figures sont disponibles en version couleur auprès de madame Turcotte à [turcotte.catherine@uqam.ca].

Cette étude a été rendue possible grâce à un financement du ministère de l'Éducation et de l'Enseignement supérieur du Québec, dans le cadre du Programme de soutien à la formation continue du personnel scolaire.

Problématique

La compréhension d'un texte est difficilement observable de façon directe (Cain & Oakhill, 2006), car elle se fait généralement de façon silencieuse et ne laisse souvent pas de traces écrites par le lecteur. Des modèles illustrant la compréhension en lecture indiquent néanmoins qu'elle s'appuie sur des connaissances sur le monde et la langue ainsi que sur diverses habiletés et stratégies (Cormier, Desrochers & Sénéchal, 2016; Irwin, 2007; Kamhi, 2009; Kendeou & van den Broek, 2007; Perfetti, Landi & Oakhill, 2005; van Dijk & Kintsch, 1983). La compréhension en lecture est donc complexe et multidimensionnelle, en plus de se développer à long terme et à travers une diversité d'expériences relatives à l'écrit (Geva & Farnia, 2012; Paris, 2005; Perfetti et al., 2005). Cela n'est pas sans rappeler qu'au Québec, un élève sur quatre échoue en lecture aux épreuves ministérielles de la 6^e année et ce taux d'échec est plus élevé dans les milieux défavorisés (Desrosiers & Tétreault, 2012). En France, selon Cèbe et Goigoux (2014), les résultats de récentes évaluations nationales indiquent que 20 à 30 % d'élèves de la fin du primaire peinent à comprendre ce qu'ils lisent, en dépit d'un décodage efficace. Ces auteurs évoquent notamment le besoin de revoir la façon dont les enseignants abordent la compréhension en lecture au primaire, tellement cette dernière semble complexe à maîtriser.

Or, bien que cette complexité soit reconnue, les épreuves abordent souvent la compréhension comme étant un concept simple (Keenan, Betjemann & Olson, 2008), réduisant ainsi les possibilités d'analyser les connaissances et habiletés, mais aussi les besoins des élèves sur le plan de divers processus de lecture (Sweet, 2005). Aux prises avec des outils d'évaluation limités, les enseignants sont portés à émettre un jugement sur les habiletés des élèves à l'aide d'observations informelles ou d'instruments inappropriés (Falardeau, Guay & Valois, 2012). Ainsi, bien qu'ils puissent détecter les jeunes lecteurs éprouvant des difficultés, ils se sentent démunis lorsque vient le temps de planifier des tâches qui leur permettraient de progresser (Deum, Gabelica, Lafontaine & Nyssen, 2007).

Devant le besoin d'obtenir des données sur les performances des élèves, mais sans avoir en main des évaluations fiables et accessibles, les enseignants utilisent également des tests ministériels en lecture des années précédentes. Ces évaluations détiennent l'avantage de pouvoir être soumises à de grands groupes et d'évaluer les compétences attendues dans le curriculum. Cependant, elles ne sont pas créées pour mieux connaître les forces et besoins de chaque élève en vue de réajuster l'intervention. Selon le ministère de l'Éducation et de l'Enseignement supérieur (2016), cette épreuve obligatoire de «français langue d'enseignement» en 6^e année du primaire a davantage pour objectifs d'illustrer les exigences ministérielles au regard de la compétence à lire et à écrire à un moment clé de la formation des élèves, de reconnaître les acquis des élèves, et de fournir des données locales et nationales sur la lecture et l'écriture. Enfin, l'utilisation des épreuves ministérielles génère souvent des pratiques dites «d'enseignement du test» afin de s'assurer que les élèves maîtrisent la forme du test et donnent ce qui est attendu comme réponse dans chaque type de question, au détriment d'un enseignement exhaustif dans une grande variété de contextes (Jennings & Bearak, 2014; Popham, 2001; Volante, 2004).

Dans la même veine, les enseignants peuvent difficilement se fier aux données de tests standardisés disponibles pour mieux comprendre les différents processus de compréhension à travailler avec les élèves (Keenan & Meenan, 2014). Bien que certaines épreuves standardisées et validées auprès de populations francophones canadiennes existent, les données qu'elles fournissent aux praticiens et chercheurs dévoilent peu d'informations se rattachant à la complexité de la compréhension en lecture. À titre d'exemple, le test de rendement individuel de Wechsler (WIAT-II CDN-F; Wechsler, 2005) est couramment utilisé au Québec pour évaluer les apprentissages scolaires en lecture (conscience phonologique, décodage, compréhension), en écriture et en langue orale auprès des élèves, et ce, de façon individuelle. En ce qui a trait à la compréhension de l'écrit, des items gradués en difficulté présentent de courtes phrases ainsi que de courts textes suivis de questions de compréhension auxquelles le participant doit répondre oralement lors d'un entretien. Les concepteurs disent que ce test évalue la compréhension littérale et les inférences logiques. Selon l'idée que les modèles théoriques décrivent la compréhension en lecture comme étant multidimensionnelle, le WIAT-II s'attarde donc à un nombre très limité de processus de compréhension.

Plusieurs autres questionnements à l'égard des tests standardisés méritent également d'être soulevés. Généralement, leur forme ne permet pas de déterminer quels sont précisément les processus impliqués dans le raisonnement de l'élève puisque l'examineur ne peut déroger du protocole de passation ni demander à l'élève, par exemple, de préciser sa réponse (Caffrey, Fuchs & Fuchs, 2008). De plus, dans toute discipline scolaire, il y a une énorme quantité de connaissances ou d'habiletés qu'un élève peut maîtriser sans toutefois être couverte par un test standardisé, car, dans un tel cas, cette évaluation serait interminable (Popham, 1999). Néanmoins, puisque ces tests individuels sont plutôt longs à faire passer (Keenan et al., 2008), leurs résultats sont parfois les seuls considérés pour poser un jugement sur les difficultés des élèves. Des chercheurs (Catts, Petscher, Schatschneider, Bridges & Mendoza, 2009) ont également révélé que ces tests se caractérisent parfois par un effet « plancher », qui détermine un trop grand nombre d'élèves comme étant à risque d'échec scolaire. L'identification de faux positifs diminue la valeur prédictive du test ainsi que sa pertinence pour réfléchir à l'intervention qui devrait être planifiée. Enfin, les travaux de Keenan et Meenan (2014) ont donné à voir que certains tests standardisés prétendant évaluer la compréhension en lecture montrent des corrélations très modestes entre eux. Leur étude comparant des résultats d'élèves ayant été soumis à trois tests standardisés montre qu'un participant ciblé comme étant en difficulté dans l'un des tests recevait ce même diagnostic lorsqu'évalué à l'aide d'un autre test selon une probabilité de seulement 43 %. Ces chercheuses mettent en garde les utilisateurs de ces tests, car, bien qu'ils prétendent évaluer les mêmes concepts, leur étude démontre que ce n'est pas le cas.

Préoccupés par l'ensemble de ces critiques, des chercheurs (O'Reilly, Sabatini, Bruce, Pillarisetti & McCormick, 2012) s'intéressent à la conception de nouveaux instruments plus appuyés sur les recherches et sur les modèles théoriques récents de l'apprentissage de la lecture afin d'aider les enseignants à prendre des décisions informées concernant leurs pratiques en classe. Les enseignants ont effectivement besoin d'épreuves pouvant être soumises à de grands groupes afin de cibler les difficultés et les forces des élèves, d'effectuer un premier dépistage, d'orienter leurs pratiques et de suivre les progrès des élèves (Sweet, 2005). Selon Compton, Fuchs, Fuchs et Bryant (2006), un tel dépistage universel peut aussi servir à déterminer les élèves qui ne présentent pas de risque d'éprouver des difficultés afin de se concentrer sur l'évaluation approfondie des élèves qui semblent

manifester davantage un retard. Ce dépistage s'appuie sur des tests pouvant détecter les processus qui posent problème à chaque élève à partir d'un modèle théorique connu des praticiens. Selon Loye et Barroso da Costa (2013), cela repose alors sur une modélisation psychométrique des réponses fournies par les élèves aux items du test. Cette modélisation permet de déterminer la probabilité de réussir un item selon le niveau de compétence de l'élève, en plus de cibler les besoins de chacun des élèves de façon spécifique et objective.

Ainsi, dans le but de mieux intervenir auprès des élèves qui éprouvent des difficultés de compréhension en lecture, des instruments sont nécessaires afin de cibler les difficultés des élèves, de concevoir des interventions adéquates et d'assurer un accompagnement soutenu. À l'instar des travaux de O'Reilly et ses collègues (2012), il importe à présent de concevoir des épreuves qui s'appuient sur des modèles psychométriques, mais aussi sur des modèles théoriques récents illustrant des processus précis sur le plan de la compréhension en lecture, ce qui a comme potentiel de mener à des pistes d'intervention en classe. Au lieu d'évaluer les items d'une épreuve existante et de tenter de découvrir les processus et les défis qu'elle suggère, cette nouvelle façon de faire exige de sélectionner un modèle théorique ainsi que les processus à évaluer dès les toutes premières phases de conception de l'épreuve. L'objectif de cet article est donc de présenter la conception d'une épreuve destinée aux élèves de la 6^e année pouvant être soumise à un grand groupe en faisant un premier pas dans cette direction. Cette année d'études a été sélectionnée puisque, à cet âge, les élèves sont amenés à lire des textes plus complexes sur des disciplines variées et également parce que ce moment est critique dans le cheminement scolaire, se situant juste avant l'entrée au secondaire.

Cadre de référence sur la compréhension en lecture : contenu de l'épreuve

L'élaboration de l'épreuve dont il est question dans cet article s'appuie sur une approche cognitive des processus de compréhension en lecture (Irwin, 2007). Ces derniers peuvent être locaux (p. ex., dans la phrase), tandis que d'autres sont plus globaux (p. ex., l'organisation des idées du texte). De plus, ce modèle a été sélectionné puisqu'il peut inspirer la conception de tâches de compréhension ayant des niveaux de complexité variés dans une même épreuve. En milieu scolaire, ce modèle est connu, car une première version de celui-ci (Irwin, 1986) a fait l'objet d'une traduc-

tion (Giasson, 1990) au Québec. De plus, de nombreux travaux s'inspirent de ce modèle (voir Beaulieu & Langevin, 2014; Clément, 2009; Falardeau, 2003; Martel & Lévesque, 2010; Thériault & Laurin, 2016). Les processus au cœur du modèle sont les suivants :

Processus 1 :

Le lecteur doit lire les mots inclus dans les phrases. Ces « microprocessus » réfèrent davantage à la maîtrise du système d'écriture alphabétique (p. ex., la relation graphème-phonème). Ces derniers sont nécessaires pour décoder les mots de tout texte, mais ils ne font pas partie des épreuves de compréhension.

Processus 2 :

Le lecteur doit faire des liens entre les phrases et entre les idées pour en construire une représentation cohérente, ce qui suscite la formulation de diverses inférences en s'appuyant sur des extraits du texte.

Processus 3 :

Il doit également dégager la structure du texte et déterminer les idées les plus importantes afin de comprendre la façon dont les idées sont organisées. À cet égard, le lecteur doit être en mesure d'extraire les idées les plus importantes en s'appuyant sur l'intention de l'auteur du texte, plutôt qu'en se centrant uniquement sur sa propre intention de lecture. Ces processus, aussi appelés « macroprocessus », exigent l'articulation de plusieurs connaissances et habiletés par rapport au texte, et non seulement au mot ou à la phrase. Dans une épreuve, les items visant ces processus sont donc souvent les plus difficiles, car ces processus sont plus complexes et requièrent un haut niveau de maîtrise (Block, Parris, Reed, Whiteley & Cleveland, 2009; Wharton-McDonald & Swiger, 2009).

Processus 4 :

Le lecteur combine les idées d'un texte avec ses connaissances. Il s'agit également d'une forme d'inférence, dite d'élaboration, amenant l'élève à sortir du texte et à faire des liens avec ses connaissances personnelles et sa compréhension de la situation générale décrite dans le texte.

Processus 5 :

Enfin, tous ces processus sont contrôlés par des stratégies métacognitives grâce auxquelles le lecteur relève ses bris de compréhension, y remédie, et réfléchit sur le contenu du texte et à sa propre compréhension.

Dans ce modèle d'Irwin (2007), l'importance d'effectuer des inférences entre les phrases et idées du texte est mise en lumière. Toutefois, les différents types d'inférences ne sont pas détaillés. Ces précisions se trouvent néanmoins dans le modèle de Bianco et Coda (2002), adapté par Dupin de Saint-André (2008), dans lequel les inférences causales, anaphoriques et lexicales sont présentées comme étant nécessaires à la compréhension.

En ce qui a trait aux inférences causales en lecture, les liens de cause à effet doivent s'effectuer adéquatement entre les phrases et entre les paragraphes afin de comprendre la chronologie d'un récit ou l'explication d'un phénomène (Boisclair, Makdissi, Sanchez Madrid, Fortier & Sirois, 2004). Parfois, certaines parties éloignées du texte doivent être mises en lien pour saisir une relation causale. À d'autres endroits, le lecteur doit comprendre le lien entre des informations rapprochées, mais dont la relation est demeurée implicite, par exemple sans la présence de marqueurs de relation.

Quant aux inférences anaphoriques, elles sont nécessaires pour utiliser adéquatement les mots de substitution, tels que les pronoms ou les synonymes, en s'appuyant sur la chaîne des référents afin de conserver la cohérence entre les propositions du texte (Irwin, 2007). Ces connaissances et habiletés, qui se rattachent notamment à la maîtrise de la grammaire, sont enseignées à partir du 2^e cycle du primaire au Québec. Or, les mots de substitution utilisés n'exigent pas toujours les mêmes habiletés ou le même degré de maîtrise. Par exemple, il est plus facile de faire un lien entre un pronom sujet (*Marie va au concert. Elle admire les chanteurs.*) et son référent, que de faire un lien entre un pronom personnel (*Marie les admire.*) et son référent.

Enfin, les mots difficiles doivent être compris ou déduits en utilisant un ensemble d'indices dans le texte, dans la phrase ou à l'intérieur du mot lui-même. Certains mots créent peu d'ambiguïté, par exemple lorsqu'il y a un synonyme à même le texte. D'autres mots peuvent néanmoins nécessiter un autre type d'habileté ou de stratégie lorsqu'ils sont rares ou polysémiques. Le lecteur doit alors s'appuyer sur sa compréhension de la phrase ou des mots de la même famille pour en déduire le sens, parfois seulement de façon approximative. À cet égard, Schmitt, Jiang et Grabe (2011) estiment que 98% des mots d'un texte doivent être compris afin d'en construire une représentation mentale adéquate. Un lecteur détenant de bonnes habiletés et connaissances pour effectuer des inférences lexicales n'a donc pas à s'interrompre devant des mots nouveaux, et peut ainsi poursuivre sa lecture sans briser le fil de sa compréhension.

En somme, toutes ces opérations de mise en relation sont nécessaires pour traiter ce qui est implicite dans un texte puisque tout n'est pas explicitement décrit ou dévoilé par l'auteur (Cain, Oakhill, Barnes & Bryant, 2001 ; McNamara, Ozuru & Floyd, 2011). La combinaison des modèles d'Irwin (2007) et de celui de Bianco et Coda (2002), adapté par Dupin de Saint-André (2008), spécifie donc les stratégies suivantes comme étant au cœur de la compréhension en lecture :

- 1) identifier les idées principales (processus 3) ;
- 2) comprendre la structure du texte (processus 3) ;
- 3) formuler des inférences causales (processus 2 et 4) ;
- 4) formuler des inférences anaphoriques (processus 2) ;
- 5) formuler des inférences lexicales (processus 2) ;
- 6) formuler des inférences d'élaboration (processus 4).

En ce qui concerne la métacognition, elle est difficilement évaluable dans une épreuve soumise en groupe, car elle demande à l'élève de parler de ses méthodes. En revanche, il est important d'ajouter des items liés à la compréhension littérale du texte puisque cela permet de cibler les élèves qui ne peuvent pas, même en 6^e année, repérer l'information d'un texte apparaissant pourtant de façon textuelle ou explicite (Basaraba, Yovanoff, Alonzo & Tindal, 2013). Enfin, l'intérêt de centrer l'évaluation sur ces processus repose sur le fait qu'il est possible de cibler, à partir de questions sollicitant divers niveaux de compréhension, les processus à développer davantage auprès d'un élève. Des études démontrent en ce sens que des items d'un même test de compréhension peuvent évaluer des processus de compréhension distincts, selon les niveaux de compréhension sollicités (Basaraba et al., 2013).

Forme de l'épreuve

Selon De Ketele (2010), la fonction de l'évaluation se réfère au « pourquoi » (préparer, améliorer ou certifier), tandis que la démarche se réfère au « comment ». Dans ce projet, l'épreuve a une fonction d'amélioration, car elle est centrée sur la détermination de ce qui est déjà acquis ou de ce qui demeure à acquérir à un moment fixe afin d'orienter les pratiques d'enseignement visant l'apprentissage. L'analyse des réponses des élèves et l'émission d'hypothèses explicatives sur leurs sources deviennent des actions nécessaires, autant que la planification d'interventions s'adressant aux élèves qui montrent des besoins d'apprentissage.

En ce qui concerne la démarche ou le «comment», le fait de vouloir faire passer l'épreuve à plusieurs élèves de façon simultanée aurait pu mener à une forme à choix multiples, qui se veut relativement rapide à corriger et qui permet de détecter rapidement les élèves les plus faibles (Basaraba et al., 2013). Or, les répondants voient parfois cette activité non pas comme une tâche de compréhension en lecture, mais telle une résolution de problème dans laquelle la discrimination des leurres est centrale (Rupp, Ferne & Choi, 2006). De plus, les questions consistent souvent à retrouver des réponses inscrites littéralement dans le texte et cette habileté est reconnue pour être plus liée aux capacités à décoder qu'à comprendre (Cain & Oakhill, 2006). Avec les choix multiples, il est également difficile de savoir si l'élève aurait pu activer par lui-même des processus attendus. Par exemple, valider une inférence est moins complexe que d'en formuler une (Oakhill, Cain & Bryant, 2003). De plus, certaines questions factuelles peuvent être répondues sans même avoir lu le texte.

Plusieurs tests, par exemple le *Gates-MacGinitie Reading Test* (MacGinitie, MacGinitie & Dreyer, 2000) ou le *WIAT-II* (Wechsler, 2005), utilisent de courts passages pour évaluer la compréhension en lecture. Or, il est reconnu que ces courts passages mesurent davantage les habiletés de lecture de mots et de compréhension de phrases, car le lecteur ne peut alors pas s'appuyer sur un contexte ni une structure de texte complète (Keenan et al., 2008). À cet égard, selon Hammill (2004), les mesures les plus ressemblantes aux tâches de lecture habituellement confiées aux élèves sont celles étant les plus prédictives de leur compétence en lecture. Ainsi, c'est dans cette visée qu'une épreuve de compréhension en lecture s'appuyant sur des modèles théoriques de la compréhension en lecture s'est élaborée. Elle comprend un questionnaire dont les items exigent une réponse ouverte, et ce, à partir de la lecture d'un texte complet semblable aux textes soumis aux élèves de la 6^e année dans diverses disciplines scolaires, et non à partir de phrases ou paragraphes isolés. Les étapes parcourues pour concevoir cette épreuve sont décrites dans cet article, en plus de la détermination de ses qualités psychométriques.

Méthodologie

Cette épreuve de compréhension en lecture s'adressant aux élèves de la 6^e année a été élaborée à partir d'un modèle théorique (Irwin, 2007) bonifié (Bianco & Coda, 2002) et selon les étapes de conception d'un test formulées par DeVellis (2003):

- 1) déterminer clairement ce qui est à mesurer en se référant à la théorie;
- 2) générer des items;
- 3) déterminer le type d'items;
- 4) vérifier la clarté des items auprès d'experts du domaine;
- 5) envisager la possibilité d'ajouter des items;
- 6) soumettre les items à un groupe représentant la population cible;
- 7) analyser les items et la validité de construit;
- 8) améliorer la fidélité du questionnaire.

Le modèle de DeVellis a été retenu puisqu'il met de l'avant l'importance de s'appuyer sur des théories de la compréhension en lecture dès la première étape de conception de l'épreuve. Il exige également le recours à des experts du domaine, ce qui réfère dans ce projet à deux chercheurs et une praticienne, pour garder en tête des aspects pratiques et utiles en classe. La relation entre la théorie et la pratique est donc cruciale dans ce modèle.

Participants

L'échantillon est constitué de 223 élèves francophones de la 6^e année du primaire. De cet échantillon, 139 élèves ont répondu à cette épreuve en 2014, puis 84 élèves en 2016. Les élèves, tous âgés de 11 ou 12 ans, fréquentaient des écoles de niveau socioéconomique faible à moyen élevé. Les élèves de l'échantillon sont répartis en 12 groupes-classes appartenant à 8 écoles primaires francophones de la grande région de Montréal.

Instrument de mesure

Cet instrument, en plus de reposer sur des éléments théoriques, doit aussi tenir compte du nombre d'élèves, du temps alloué à cette évaluation, des caractéristiques des élèves ainsi que des utilisateurs et examinateurs de l'épreuve afin qu'il soit pratique en contexte scolaire. Cette épreuve a

donc été conçue en relation étroite entre chercheurs et praticiens, puis sur une période de trois ans entre 2011 et 2014, au cours de laquelle trois rondes de passation ont été nécessaires afin de la peaufiner.

Le texte, le questionnaire et la grille de correction ont été conçus par trois experts, dont deux chercheurs en littérature et une orthopédagogue diplômée au 2^e cycle, et ce, à partir des processus et des modèles théoriques choisis, puis en fonction des élèves ciblés. Un texte informatif a donc été rédigé en s'appuyant sur des critères de lisibilité correspondant à la 6^e année. Ce texte, qui porte sur le trafic des animaux exotiques, comprend 701 mots et 40 phrases. Le thème du texte a été déterminé en fonction de son accessibilité puisque les animaux dont il est question (tortues, singes, araignées, etc.) sont bien connus des élèves. Or, le texte porte sur le braconnage ainsi que sur les moyens entrepris pour faire le trafic de ces animaux, ce qui représente de l'information nouvelle.

L'organisation des idées du texte respecte une structure de type cause/effet (Meyer, 1985), c'est-à-dire qu'elle exprime les causes et les effets du trafic des animaux. Cette structure cause/effet a été sélectionnée, car elle permet tout particulièrement de générer des questions sur les inférences causales. L'indice Lisi (voir Mesnager, 2002, 2011) du texte est de 6,5, ce qui est considéré comme un niveau moyen. Cet indice est obtenu à l'aide d'un logiciel téléchargeable à partir d'un site du Réseau des observatoires locaux de la lecture¹. Il évalue le niveau de lisibilité textuelle à partir de la longueur des phrases et de la complexité du vocabulaire sur une échelle de 2 à 10, construite pour des élèves de 10 à 12 ans. Bien que, selon son auteur lui-même, il n'y aura jamais une mesure générale et exhaustive du « lisible », cet outil permet à l'enseignant d'avoir une certaine idée de la complexité d'un texte pour des élèves de cet âge.

Parallèlement, un questionnaire sur ce texte a été conçu en conservant en tête les modèles théoriques de départ sur la compréhension en lecture (Bianco et Coda, 2002 ; Irwin, 2007). Il comprend, dans sa version finale, les 13 items suivants:

- 1) repérage d'une information en début de texte ;
- 2) repérage d'une information en fin de texte ;
- 3) identification d'une idée principale explicite dans un paragraphe ;
- 4) identification d'une idée principale implicite dans un paragraphe ;
- 5A-5B) formulation d'une inférence anaphorique (2 items) ;

- 6-7) formulation d'une inférence causale logique (2 items);
- 8A-8B) formulation d'une inférence lexicale (2 items);
- 9) formulation d'une inférence causale pragmatique;
- 10) formulation d'une inférence logique;
- 11) organisation d'informations référant à la structure cause/effet du texte.

Chaque réponse vaut 0 ou 1 point, donc les participants peuvent obtenir un score entre 0 et 13. Pour l'item 11, qui exige d'inscrire six éléments de réponse, les élèves ayant rapporté quatre éléments sur six obtiennent 1 point, tandis que ceux ayant rapporté trois éléments ou moins n'obtiennent aucun point. Notons que trois causes et trois effets sont attendus, donc l'élève qui inscrit quatre éléments adéquats démontre qu'il distingue ces deux concepts et qu'il a saisi au moins partiellement la structure du texte.

Il existe deux items évaluant la formulation d'une inférence anaphorique (5A et 5B), puis deux pour la formulation d'une inférence lexicale (8A et 8B) mettant en jeu des habiletés différentes. En ce qui concerne l'inférence anaphorique, l'item 5A demande à l'élève de trouver le référent d'un pronom sujet «elles» (les tortues), tandis que l'item 5B demande à l'élève de trouver le référent d'un pronom personnel dans l'extrait «des personnes sont prêtes à tout pour se les procurer» (les réfère aux animaux exotiques). Ces deux items sont donc d'un niveau de complexité très différent, font appel à des habiletés distinctes et permettent de détecter des difficultés précises. Les élèves de la 6^e année pourraient réussir facilement l'item 5A puisque la reprise d'informations dans un texte est abordée dans le 1^{er} cycle du primaire. Néanmoins, le pronom «les» génère une confusion, car il est plus souvent utilisé comme déterminant. Le fait de détenir deux items ayant une telle indépendance s'avère ainsi utile pour mieux cibler les défis d'élèves faibles lecteurs.

En ce qui a trait à l'inférence lexicale, l'item 8A demande à l'élève de définir le mot «braconnier», pour lequel un synonyme se trouve dans le texte. L'item 8B, quant à lui, nécessite de comprendre la phrase dans laquelle se trouve le mot «épingler» (se faire prendre) puisque ce mot hors contexte pourrait revêtir un autre sens plus littéral (attacher avec une épingle). Il s'agit une fois de plus de deux items sollicitant des habiletés différentes.

La révision des items et l'ajustement du texte à la suite de trois rondes de passation menées auprès de 115 à 122 élèves chaque fois (en 2011, 2012 et 2013) ont servi à ajuster le niveau de difficulté du texte et la clarté des items du questionnaire, puis à s'assurer que chacun évaluait bien ce qu'il prétendait évaluer. Cette étape s'est réalisée par les trois experts en tenant compte de la variété des réponses des élèves, des commentaires des enseignants participants et de leurs observations lors de la passation. Par exemple, le manque de clarté de l'item 5A conduisait les élèves à ne pas se référer au texte pour y répondre. Le voici : *Le mot de substitution suivant remplace un autre mot. Quel est-il?* Cette formulation a été changée pour la suivante : *Le mot encadré est un mot de substitution. Retrouve dans le texte le mot ou les mots qu'il remplace.* À cet effet, le choix des mots utilisés dans les questions peut avoir une influence sur les réponses des participants (Blais, Grondin, Loye & Raïche, 2007), mais il n'est pas toujours aisé de comprendre quels sont ces mots et quel est leur effet sans avoir mis à l'essai des formulations différentes. C'est notamment sur ce type de considération que l'équipe d'experts s'est concentrée pendant trois années. Par exemple, l'expression « mot de substitution » a été employée puisqu'elle est couramment utilisée en milieu scolaire ainsi que dans le document qui réfère à la progression des apprentissages (MELS, 2011).

La grille de correction de ce questionnaire a également été conçue dès le départ en consignnant les réponses attendues ainsi que leurs variantes acceptées, puis des exemples de réponses refusées avec une justification. Cette grille a été alimentée à chaque tour de passation à partir des réponses des élèves. Certaines ont été ajoutées dans la catégorie « réponses acceptées » et d'autres dans la catégorie « réponses refusées ».

Déroulement de la collecte des données

La collecte des données présentées dans cet article s'est effectuée en septembre 2014 et en septembre 2016, avec la même version finale de l'épreuve. Lors de la passation de l'épreuve, les élèves disposaient d'une heure pour lire le texte et pour répondre aux items du questionnaire. De plus, l'enseignant ne répondait pas aux questions des élèves pendant la passation de l'épreuve et ne lisait pas le texte. Le dictionnaire n'était pas autorisé.

Méthodes d'analyse des données

Puisque l'intérêt de cette épreuve est d'examiner les processus qui pourraient présenter des difficultés chez les élèves, il serait peu pertinent de considérer simplement le score final à cette épreuve ou encore le taux de réussite pour chaque item. La validation de cette épreuve s'appuie donc sur la possibilité de comprendre la complexité de chaque item, sur la gradation de difficultés entre les items, sur la possibilité de réussir chaque item pour des élèves ayant des habiletés différentes, puis sur le potentiel de chaque item d'informer l'examineur sur les pratiques à favoriser pour répondre aux besoins des élèves.

La détermination de la difficulté de chacun des items ainsi que des items les plus informatifs a été effectuée à l'aide du modèle de Rasch (1960/1980). Ce modèle se conjugue particulièrement bien à la conception de cette épreuve, car la compréhension en lecture est complexe, mais renferme des processus qui peuvent être ciblés et distingués les uns des autres. Bien que cette distinction existe sur le plan théorique (Irwin, 2007), le modèle de Rasch permet de déterminer si les processus peuvent être évalués de façon indépendante ou s'ils forment un amalgame pour lequel il est difficile de relever des difficultés précises. En effet, puisque cette épreuve est conçue pour les enseignants et autres intervenants scolaires, chaque item, relié à un processus, doit contribuer à distinguer les élèves entre eux, et ce, dans la mesure où le niveau de difficulté de chaque item est connu.

Certaines analyses préalables à l'utilisation du modèle de Rasch (1960/1980) ont dû être effectuées. À ce titre, une analyse descriptive des données s'est avérée nécessaire afin de regarder les données dans leur globalité. Par la suite, la fidélité du questionnaire a été déterminée. Une analyse des corrélations ainsi que leur adéquation pour mener une analyse de la dimensionnalité s'est imposée. L'analyse de la dimensionnalité, suivie de l'analyse selon le modèle de Rasch, termine la séquence d'analyses afin d'atteindre l'objectif de recherche. L'ensemble de ces analyses a été mené à l'aide du logiciel R et des bibliothèques psych, nFactors et ltm.

Résultats

Analyse descriptive

Les statistiques descriptives des données, présentées au tableau 1, montrent que les données sont présentes pour l'ensemble des items. En effet, seule une donnée est manquante à l'item 8B, pour lequel la réponse de l'élève est demeurée illisible sur le plan calligraphique. Les élèves ont obtenu la bonne réponse de façon variable selon les items, mais les items 4 et 11 ont été particulièrement échoués. D'ailleurs, il s'agit des deux seuls items dont l'asymétrie est légèrement positive. Bien que l'ensemble des items présente une asymétrie, ce constat ne remet pas en cause l'utilisation du modèle de Rasch (1960/1980) puisque celui-ci ne requiert pas une distribution normale des données.

Tableau 1
Statistiques descriptives des données

| Item | Fréquences | | | a ³ |
|------|--------------------------------------|---|---|----------------|
| | N ^b re de bonnes réponses | N ^b re de mauvaises réponses | N ^b re de données manquantes | |
| 1 | 198 | 25 | 0 | -2,476 |
| 2 | 180 | 43 | 0 | -1,568 |
| 3 | 120 | 103 | 0 | -0,154 |
| 4 | 72 | 151 | 0 | 0,763 |
| 5A | 194 | 29 | 0 | -2,215 |
| 5B | 191 | 32 | 0 | -2,048 |
| 6 | 136 | 87 | 0 | -0,454 |
| 7 | 136 | 87 | 0 | -0,454 |
| 8A | 146 | 77 | 0 | -0,655 |
| 8B | 158 | 64 | 1 | -0,941 |
| 9 | 153 | 70 | 0 | -0,807 |
| 10 | 181 | 42 | 0 | -1,605 |
| 11 | 97 | 126 | 0 | 0,264 |

Note. a³ = coefficient d'asymétrie.

Fidélité du questionnaire

La cohérence interne a été établie à l'aide du coefficient alpha de Cronbach. Celui-ci est estimé à 0,74. Ces résultats démontrent une bonne fidélité du questionnaire, car, selon Nunally (1978), un coefficient de 0,70 et plus est jugé suffisant pour attester d'une bonne fidélité. Bien qu'un coefficient alpha de Cronbach de 0,74 n'indique pas l'unidimensionnalité du test, il n'en demeure pas moins qu'une bonne fidélité demeure une condition nécessaire à l'homogénéité des items. Ainsi, ce coefficient permet de soutenir l'interprétation des résultats obtenus à l'analyse de la dimensionnalité par l'analyse en composantes principales.

Analyse des corrélations et de leur adéquation pour une analyse de la dimensionnalité

Puisqu'il y a une seule donnée manquante pour un seul item, il n'a pas été jugé nécessaire d'utiliser des méthodes pour remplacer les données manquantes. Ainsi, l'analyse des corrélations a été effectuée selon la méthode *pairwise*, soit en effectuant le calcul des corrélations par paires de variables disponibles. La matrice des corrélations de Pearson présente des corrélations inter-items se situant entre 0,01 et 0,42. Les résultats au test de sphéricité de Bartlett ($\chi^2=41$; $n=223$; $p<0,001$) ainsi que l'excellence de l'indice KMO (KMO=0,80) démontrent que les 13 variables manifestes peuvent être regroupées en un nombre restreint de variables latentes. Une analyse de la dimensionnalité peut alors être effectuée afin de déterminer le nombre de dimensions évaluées par le questionnaire.

Analyse de la dimensionnalité

D'après le tableau 2 et la figure 1, une seule dimension est à retenir. En effet, la première valeur propre explique 26% de la variance, ce qui répond au minimum de 20% proposé par Reckase (1979). À cet égard, la première valeur propre se démarque clairement des valeurs propres subséquentes, qui n'expliquent que très peu de variance. De plus, selon l'analyse parallèle de Horn (1965), une seule dimension s'impose. En effet, la seconde dimension se situe en deçà du trait représentant les valeurs propres moyennes, générées de façon aléatoire. Par contre, il est tout de même à préciser que le test de l'éboullis de Cattell (1966) suggérerait deux dimensions, alors que la règle de Kaiser-Guttman (Guttman, 1954; Kaiser, 1960) en proposerait trois. L'indépendance locale de chaque item a été assumée

Tableau 2
*Valeurs propres d'une analyse en composantes principales
selon la matrice des corrélations de Pearson*

| Items | Valeurs propres | % | Cumul (%) |
|-------|-----------------|------|-----------|
| 1 | 3,33 | 0,26 | 0,26 |
| 2 | 1,18 | 0,09 | 0,35 |
| 3 | 1,14 | 0,09 | 0,43 |
| 4 | 0,98 | 0,08 | 0,51 |
| 5 | 0,97 | 0,07 | 0,58 |
| 6 | 0,87 | 0,07 | 0,65 |
| 7 | 0,82 | 0,06 | 0,71 |
| 8 | 0,76 | 0,06 | 0,77 |
| 9 | 0,72 | 0,06 | 0,83 |
| 10 | 0,67 | 0,05 | 0,88 |
| 11 | 0,58 | 0,04 | 0,92 |
| 12 | 0,54 | 0,04 | 0,97 |
| 13 | 0,44 | 0,03 | 1,00 |

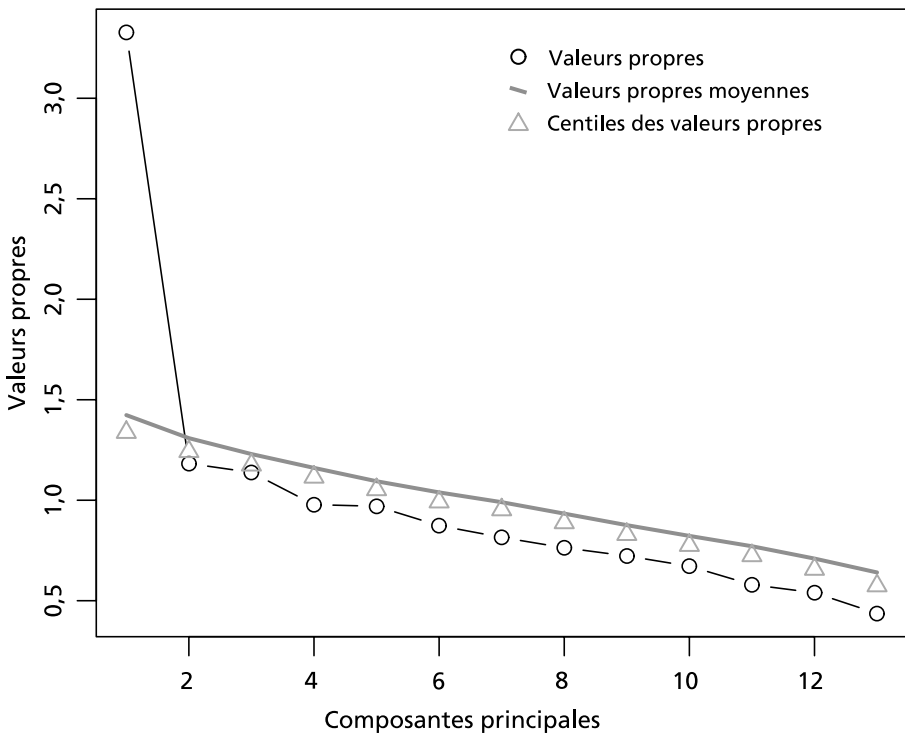


Figure 1. *Graphique des valeurs propres et analyse parallèle
selon une analyse en composantes principales*

puisque la réussite d'aucun item n'est préalable à la réussite d'un autre item. Bien que certains items évaluent des processus semblables, ils ne sollicitent pas les mêmes habiletés chez les élèves.

Analyse selon le modèle de Rasch

Paramètres de difficulté et de discrimination pour chacun des items

Alors que les paramètres de difficulté pour chacun des items ont été calculés, le paramètre de discrimination a été fixé à 1,19 pour chacun des items puisque le modèle de Rasch (1960/1980) considère une discrimination identique pour l'ensemble des items. Le test cible donc les élèves ayant des habiletés de compréhension en lecture plus élevées et ceux ayant des habiletés de compréhension en lecture plus faibles.

À la lecture du tableau 3 et de la figure 2, il est possible de constater que les items 1, 5A, 5B et 10 sont les items les plus faciles. Il s'agit d'items évaluant le repérage d'une information en début de texte, la formulation d'une inférence anaphorique ainsi que la formulation d'une inférence logique. Selon le curriculum, en 6^e année du primaire, le simple repérage d'une information textuelle doit être acquis, de même que la reprise et l'ajout d'informations dans un texte. Ainsi, les items 1, 5A et 5B se révèlent faciles, ce qui était attendu. Néanmoins, l'item 2, qui exige le repérage d'une information en fin de texte, est moins bien réussi que l'item 1. Cela laisse entendre que certains élèves ne relisent peut-être pas le texte au complet avant de répondre à certains items et se précipitent sur la première information qui pourrait sembler pertinente. Enfin, sur le plan de l'inférence logique, l'item 10 se révèle facile à réussir. Étant donné que plusieurs réponses apparaissent dans le texte et que les élèves peuvent extraire cette information à plusieurs endroits, il s'agit d'un item qui ne détermine donc pas les lecteurs les plus compétents.

En revanche, et comme déjà décrit, les items 4 et 11 sont définis comme étant les plus difficiles. Alors que l'item 4 permet d'évaluer l'identification d'une idée principale implicite dans un paragraphe, l'item 11 cible la compréhension des causes et effets du phénomène (structure cause/effet) selon l'organisation des idées dans le texte. Ces processus de compréhension de la macrostructure du texte sont d'ailleurs considérés comme étant les plus complexes, selon le modèle d'Irwin (2007). Néanmoins, au regard de la globalité du test, le test est jugé plutôt facile.

Tableau 3
Paramètres de difficulté et de discrimination pour chacun des items

| Item | Difficulté (b) | Discrimination (a) |
|------|----------------|--------------------|
| 1 | -2,15 | 1,19 |
| 2 | -1,51 | 1,19 |
| 3 | -0,16 | 1,19 |
| 4 | 0,80 | 1,19 |
| 5A | -1,98 | 1,19 |
| 5B | -1,87 | 1,19 |
| 6 | -0,47 | 1,19 |
| 7 | -0,47 | 1,19 |
| 8A | -0,68 | 1,19 |
| 8B | -0,96 | 1,19 |
| 9 | -0,83 | 1,19 |
| 10 | -1,54 | 1,19 |
| 11 | 0,29 | 1,19 |

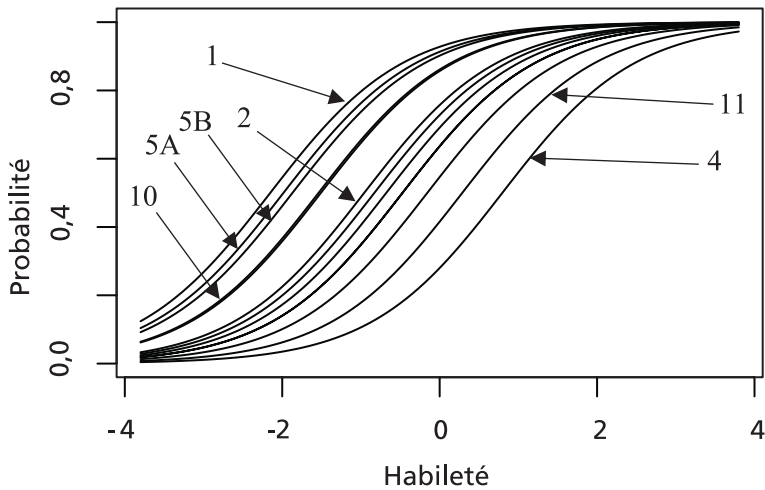


Figure 2. *Courbes caractéristiques pour chacun des items*

Qualité de l'ajustement entre les données et le modèle

La qualité de l'ajustement des données (*item fit*) au modèle de Rasch (1960/1980) a été déterminée à l'aide d'un test de khi carré. Le seuil de signification retenu pour l'interprétation du rejet ou non-rejet de l'hypothèse nulle, à savoir si la distribution des paramètres de chacun des items suit la distribution proposée par le modèle de Rasch, est de 0,05. Les résultats obtenus (voir Tableau 4) démontrent un assez bon ajustement au modèle. Or, les items 7 et 11 n'adhèrent pas au modèle puisque le seuil de signification est inférieur à 0,05.

Tableau 4
Valeurs du khi carré et du seuil de signification pour le test de la qualité de l'ajustement des données pour chacun des items

| Item | χ^2 | Seuil de signification |
|------|----------|------------------------|
| 1 | 5,3 | 0,6231 |
| 2 | 9,6 | 0,2115 |
| 3 | 12,0 | 0,0996 |
| 4 | 7,7 | 0,3641 |
| 5A | 5,5 | 0,6046 |
| 5B | 9,8 | 0,2018 |
| 6 | 7,7 | 0,3603 |
| 7 | 20,2 | 0,0052 |
| 8A | 11,0 | 0,1374 |
| 8B | 11,0 | 0,1405 |
| 9 | 10,7 | 0,1513 |
| 10 | 6,4 | 0,4930 |
| 11 | 16,6 | 0,0202 |

Courbes d'information

Les courbes d'information pour chacun des items ainsi que pour le test dans son ensemble sont présentées respectivement aux figures 3 et 4. Selon les courbes présentées, l'ensemble des items permet d'obtenir de l'information pour des élèves ayant essentiellement des habiletés se situant entre -4 et 4, mais plus particulièrement pour ceux ayant des niveaux d'habileté se situant entre -2,5 et 1. Or, il est possible d'obtenir de l'information pour l'ensemble des niveaux d'habileté des élèves en compréhension en lecture. Plus précisément, les items 3, 4 et 11 sont plus informatifs pour les élèves ayant des habiletés de compréhension en lecture plus élevées. Par contre, les items 1, 5A et 5B permettent d'obtenir plus d'informations pour les élèves plus faibles. Comme nous l'avons exprimé plus haut, bien que les items 5A

et 5B semblent évaluer le même type d'inférence, les habiletés qu'ils suscitent sont distinctes. En somme, l'ensemble du test se révèle plus informatif pour les élèves ayant une habileté de compréhension en lecture plutôt faible.

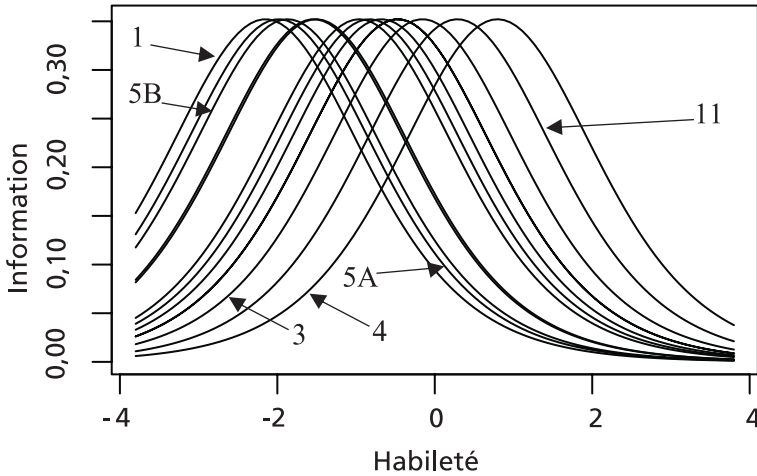


Figure 3. *Courbes d'information pour chacun des items*

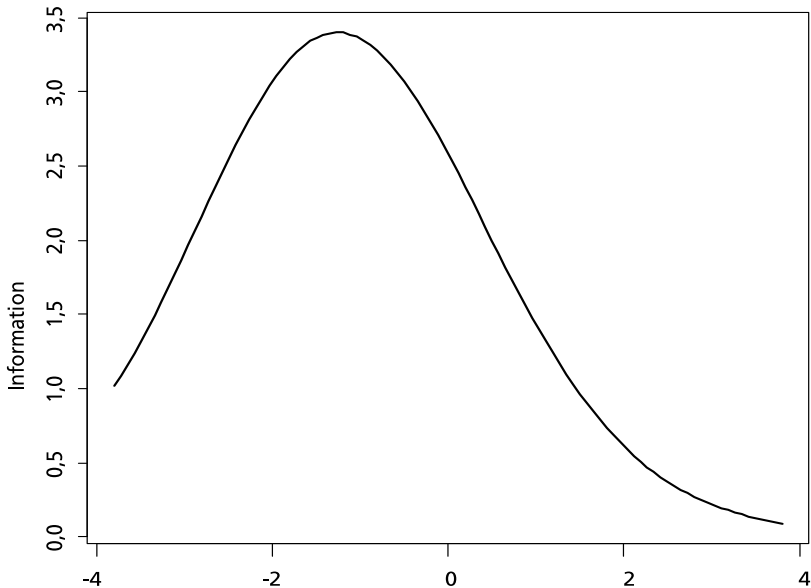


Figure 4. *Courbe d'information pour l'ensemble du test*

Discussion

La conception d'une nouvelle épreuve exige du temps et des ressources considérables consacrés à son élaboration, à sa passation, à son analyse et à ses multiples révisions. En contexte scolaire, plusieurs considérations pratiques doivent se trouver au cœur de cette élaboration lorsque les enseignants et les chercheurs veulent pouvoir l'utiliser régulièrement. L'épreuve conçue dans ce projet représente ainsi une contribution puisqu'elle s'appuie sur des modèles théoriques connus, a été élaborée par des experts chercheurs et praticiens, et peut être soumise rapidement à un grand groupe, c'est-à-dire en 60 minutes. Les démarches effectuées lors des premières années vouées à cette conception ont été décrites dans cet article. Le cadre théorique choisi, la conception et l'analyse des réponses aux items donnent à voir les qualités de l'épreuve ainsi que son potentiel pour mieux cibler les difficultés en compréhension des élèves de façon objective, puis pour orienter l'évaluation tout comme l'enseignement subséquent.

En ce qui a trait à la pertinence de cette épreuve, il est important de rappeler qu'elle a été élaborée en fonction de modèles théoriques de la compréhension en lecture, tout en ayant comme objectif de satisfaire à des standards psychométriques et pratiques acceptables pouvant mener à des pistes d'intervention en classe (O'Reilly et al., 2012). En effet, cet instrument se veut complémentaire à ceux utiles pour détecter les élèves qui éprouvent des difficultés, mais qui aident peu les enseignants à planifier des tâches précises permettant aux élèves de progresser (Deum et al., 2007). De plus, cette épreuve a pour objectif de favoriser l'articulation enseignement-apprentissage-évaluation, ce qui se distingue des tests soumis par le ministère de l'Éducation et de l'Enseignement supérieur (2016), plus centrés sur la reconnaissance des acquis des élèves ou sur la collecte de données locales et nationales sur la lecture et l'écriture.

Au cours des diverses étapes de conception de l'épreuve, les experts se sont assurés de la cohérence entre les items du questionnaire et les modèles théoriques sélectionnés. Les trois rondes de passation entre 2011 et 2014 ont permis d'ajuster le texte et les items du questionnaire ainsi que la grille de correction en ce sens. Par la suite, les étapes de conception se sont centrées sur la valeur psychométrique de la version finale de l'épreuve soumise à 223 élèves en 2014 et en 2016 en dévoilant le niveau de complexité de chaque item. Il est d'ailleurs intéressant de constater que les items les plus difficiles, soit l'item 4 (qui exige de formuler l'idée principale d'un

paragraphe) et l'item 11 (qui demande aux élèves de sélectionner et d'organiser des informations du texte, ce qui démontre leur compréhension de la structure cause/effet), s'inscrivent dans un processus de compréhension considéré comme plus complexe (Block et al., 2009; Wharton-McDonald & Swiger, 2009). L'item 11, qui ne semble d'ailleurs pas adhérer correctement au modèle de Rasch (1960/1980), est probablement le plus ardu à répondre, car il requiert des connaissances et une réflexion de haut niveau sur le plan cognitif. Au sein d'une telle épreuve, même si son adhérence au modèle pourrait être remise en question sur le plan psychométrique, cet item détient néanmoins une utilité plutôt intéressante en permettant de déterminer les élèves les plus forts de la classe en compréhension en lecture. Quant à l'item 7, il exige une inférence causale difficile à formuler pour plusieurs élèves puisque la réponse n'apparaît pas explicitement dans le texte. Alors que cet item exige l'inférence d'une cause, des élèves vont plutôt évoquer un effet. Cela peut être une piste expliquant pourquoi cet item n'adhère pas correctement au modèle non plus, mais une analyse qualitative des réponses données par les élèves pourrait éclairer davantage cette hypothèse.

Les items se référant à la compréhension littérale (la réponse est directement dans le texte) et aux inférences anaphoriques (mots de substitution) se sont révélés quant à eux plus faciles pour des élèves de la 6^e année, ce qui concorde avec les habiletés attendues à cette année d'études. Rappelons d'ailleurs que, dès la 2^e année du primaire au Québec, les élèves apprennent qu'un groupe du nom peut être remplacé par un pronom sujet dans une phrase. Les élèves qui ne réussissent pas ces items pourraient donc être considérés comme étant les plus faibles et, à ce titre, les items 1, 2, 5A et 5B sont fort utiles. De plus, l'analyse du niveau de difficulté indique que les élèves faibles sont plus susceptibles d'être repérés dans cette épreuve. Cela évitera peut-être un effet «plancher», qui détermine un trop grand nombre d'élèves comme étant à risque (Catts et al., 2009).

Cette épreuve indique la réussite ou l'échec à des items se référant à des processus de compréhension variés, ce qui peut donner à voir différents portraits d'élèves ou une progression sur le plan d'un ou de plusieurs processus, si elle est soumise en début et en fin d'année. À cet égard, la réussite d'un item ne permet pas de prédire la réussite d'un autre item puisque ces derniers sont indépendants. Ainsi, en connaissant le niveau de difficulté de chaque item, mais surtout la possibilité de réussir un item selon

les compétences d'un élève, l'enseignant peut interpréter la performance de chaque élève de la classe de façon objective, puis réfléchir à des pistes d'enseignement. Un élève qui échoue un item réussi par la majorité de ses pairs est donc susceptible d'attirer l'attention de son enseignant.

En somme, la modélisation de Rasch (1960/1980) a permis de créer une épreuve offrant un portrait détaillé des habiletés de chaque élève à l'égard de la compréhension d'un texte afin d'orienter l'enseignement sur des objectifs précis. Cette modélisation a permis également de comprendre que l'épreuve était particulièrement pertinente pour les élèves éprouvant des difficultés. En ce sens, il est utile de préciser que le score total des élèves à ce type d'épreuve dévoile peu de détails sur ce que chaque élève sait ou ne sait pas faire. Ainsi, deux élèves ayant un score total de 8 pourraient avoir, par exemple, des besoins d'apprentissage considérablement différents, selon les items échoués et réussis par chacun. Sachant cela, puis connaissant les modèles théoriques derrière l'élaboration de cette épreuve, l'enseignant peut concevoir des activités d'apprentissage qui visent le développement des processus de compréhension dont les élèves ont le plus besoin. Si la majorité des élèves de la classe échoue les items qui traitent des idées principales (items 3 et 4), un enseignement des stratégies pour trouver et formuler cette idée dans les paragraphes s'avère probablement utile. Une telle façon de déterminer des objectifs d'enseignement pourrait également s'appliquer auprès de sous-groupes d'élèves ayant des besoins semblables. Les enseignants qui connaissent les modèles théoriques derrière la conception de cette épreuve peuvent également bonifier l'évaluation des élèves à l'égard des processus qui semblent les plus complexes à l'aide d'autres textes et questionnaires, concevoir des interventions adéquates et assurer du suivi soutenu.

Toujours dans le but d'approfondir l'évaluation de la compréhension en lecture chez les élèves, un enseignant pourrait vouloir soumettre un élève à cette épreuve sous forme orale uniquement, dans le cas où ce dernier démontrerait de très faibles habiletés de décodage ou encore présenterait un trouble ou un handicap sur le plan moteur. Contrairement à ce qui est possible avec les tests standardisés (Caffrey et al., 2008), l'enseignant pourrait également vouloir questionner individuellement l'élève sur ses procédures pour répondre à chaque item afin d'obtenir des informations sur son contrôle de la tâche et sur ses stratégies métacognitives. Puisque cette épreuve peut s'articuler avec l'enseignement et l'apprentissage

de la compréhension de l'écrit en dévoilant les processus réussis ou échoués pour chaque élève, il revient aux enseignants de l'utiliser en connaissant les modèles théoriques.

Limites et implications

Cette épreuve à elle seule ne peut suffire pour évaluer la compréhension en lecture. D'une part, peu d'items sont reliés à chaque processus ciblé. Il aurait été possible d'ajouter des items supplémentaires, mais cette épreuve exige déjà en moyenne 60 minutes à réaliser pour les élèves de la 6^e année. L'ajout d'items augmente le temps de passation et il devient moins réaliste de soumettre ainsi les élèves à cette condition, d'autant plus que le texte devrait en ce sens contenir des paragraphes supplémentaires afin d'augmenter la quantité d'informations pouvant par la suite faire l'objet de questions. D'autre part, cette épreuve utilise un texte informatif qui a une structure cause/effet. La compréhension de tout autre type de texte, surtout littéraire, n'est donc pas évaluée dans cette épreuve.

De plus, cette épreuve consiste en une première évaluation soumise à toute la classe qui permet de mettre en lumière des connaissances, des habiletés et des difficultés liées à des processus de compréhension ayant fait ou non l'objet d'un enseignement afin d'orienter l'intervention en classe, en sous-groupes ou de façon individuelle. Des évaluations qui approfondissent les processus les moins bien réussis devraient compléter les données de cette épreuve à des fins de confirmation ou d'infirmerie puisqu'il est bien connu qu'une mesure de compréhension exhaustive semble irréaliste (Cain & Oakhill, 2006), tant sa complexité est grande.

Sur le plan psychométrique, il serait intéressant de faire une analyse de fonctionnement différentiel d'item (FDI, ou DIF pour *differential item functioning*) ou encore une analyse d'invariance factorielle. Cela permettrait de déterminer la stabilité de l'épreuve lors de la passation auprès de différents groupes, notamment auprès des garçons et des filles, des élèves en milieu rural ou urbain, puis en fonction de la date de naissance des participants.

Réception : 14 février 2017

Version finale : 3 octobre 2017

Acceptation : 27 octobre 2017

NOTES

1. <http://lettres-histoire-geo.ac-amiens.fr/188-logiciel-lisi.html>

RÉFÉRENCES

- Basaraba, D., Yovanoff, P., Alonzo, J., & Tindal, G. (2013). Examining the structure of reading comprehension : Do literal, inferential, and evaluative comprehension truly exist? *Reading and Writing, 26*(3), 349-379. doi: 10.1007/s11145-012-9372-9
- Beaulieu, J. & Langevin, J. (2014). L'élève qui a des incapacités intellectuelles et la lecture. *Revue francophone de la déficience intellectuelle, 25*, 52-69. doi: 10.7202/11027327ar
- Bianco, M. & Coda, M. (2002). La compréhension en quelques points. Dans M. Bianco, M. Coda & D. Gourgue (dir.), *La compréhension* (pp. 93-97). Grenoble: Éd. de la Cigale.
- Blais, J.-G., Grondin, J., Loye, N., & Raïche, G. (2007, April). *A transverse study of items' wording impact with Rasch's rating scale model*. Paper presented at the Annual Conference of the American Educational Research Association (AERA), Chicago, IL.
- Block, C. C., Parris, S. R., Reed, K. L., Whiteley, C. S., & Cleveland, M. D. (2009). Instructional approaches that significantly increase reading comprehension. *Journal of Educational Psychology, 101*(2), 262-281.
- Boisclair, A., Makdissi, H., Sanchez Madrid, C. P., Fortier, C. & Sirois, P. (2004, août). *La structuration causale du récit chez le jeune enfant*. Actes du 9^e colloque de l'Association internationale pour la recherche en didactique du français (AIRDF), Québec. Repéré à www.colloqueairdf.fse.ulaval.ca/fichier/Communications/Boisclair-makdissi-sanchez-fortier.pdf
- Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education, 41*(4), 254-270. doi: 10.1177/0022466907310366
- Cain, K., & Oakhill, J. V. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76*(4), 697-708. doi: 10.1348/000709905x69807
- Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition, 29*(6), 850-859. doi: 10.3758/bf03196414
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276. doi: 10.1207/s15327906mbr0102_10

- Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*(2), 163-176. doi: 10.1177/0022219408326219
- Cèbe, S. & Goigoux, R. (2014). Avec Lectorino et Lectorinette. *Cahiers pédagogiques, 516*, 34-36.
- Clément, N. L. (2009). Une rencontre affective avec la lecture : un modèle innovateur de l'enseignement de la lecture. *Revue du Nouvel-Ontario, 34*, 175-217. Repéré à <http://id.erudit.org/iderudit/038724ar>
- Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*(2), 394-409. doi: 10.1037/0022-0663.98.2.394
- Cormier, P., Desrochers, A. & Sénéchal, M. (2006). Validation et consistance interne d'une batterie de tests pour l'évaluation multidimensionnelle de la lecture en français. *Revue des sciences de l'éducation, 32*(1), 205-225. doi: 10.7202/013483ar
- De Ketele, J.-M. (2010). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée, 1*(15), 25-37. Repéré à www.cairn.info/revue-francaise-de-linguistique-appliquee-2010-1-page-25.htm
- Desrosiers, H. & Tétreault, K. (2012). *Les facteurs liés à la réussite aux épreuves obligatoires de français en sixième année du primaire: un tour d'horizon*. Étude longitudinale du développement des enfants du Québec (ÉLDEQ 1998-2010): de la naissance à 12 ans. Institut de la statistique du Québec, 7(1), 40 p. Repéré à www.stat.gouv.qc.ca/statistiques/education/prescolaire-primaire/reussite-epreuve-francais.pdf
- Deum, M., Gabelica, C., Lafontaine, A. & Nyssen, M.-C. (2007). *Outil pour le diagnostic et la remédiation des difficultés d'acquisition de la lecture en 1^{re} et 2^e années primaires*. Liège, Belgique : Ministère de la Communauté française, Service général du pilotage du Système éducatif.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Dupin de Saint-André, M. (2008). Pistes pour travailler les inférences au primaire. *Québec français, 150*, 66-67. Repéré à <http://id.erudit.org/iderudit/44010ac>
- Falardeau, É. (2003). Compréhension et interprétation: deux composantes complémentaires de la lecture littéraire. *Revue des sciences de l'éducation, 29*(3), 673-694. doi: 10.7202/011409ar
- Falardeau, É., Guay, F. & Valois, P. (2012). Expliciter la compétence à lire et apprécier pour mieux l'évaluer. *Québec français, 164*, 71-72. Repéré à <http://id.erudit.org/iderudit/65900ac>
- Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Reading and Writing, 25*(8), 1819-1845. doi: 10.1007/s11145-011-9333-8
- Giasson, J. (1990). *La compréhension en lecture*. Boucherville : Gaëtan Morin.
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*(2), 149-161. doi: 10.1007/bf02289162

- Hammill, D. D. (2004). What we know about correlates of reading. *Exceptional Children*, 70(4), 453-469. doi: 10.1177/001440290407000405
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185. doi: 10.1007/bf02289447
- Irwin, J. W. (1986). *Teaching reading comprehension processes*. Englewood Cliffs, NJ: Prentice Hall.
- Irwin, J. W. (2007). *Teaching reading comprehension processes* (3rd ed.). Boston, MA: Pearson Allyn and Bacon.
- Jennings, J. L., & Bearak, J. M. (2014). "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher*, 43(8), 381-389. doi: 10.3102/0013189x14554449
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151. doi: 10.1177/001316446002000116
- Kamhi, A. G. (2009). The case for the narrow view of reading. *Language, Speech, and Hearing Services in Schools*, 40(2), 174-177. doi: 10.1044/0161-1461(2009/08-0068)
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300. doi: 10.1080/10888430802132279
- Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*, 47(2), 125-135. doi: 10.1177/0022219412439326
- Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, 35(7), 1567-1577. doi: 10.3758/bf03193491
- Loye, N. & Barroso da Costa, C. (2013). Hiérarchiser les besoins de diagnostic en mathématique en FP à l'aide d'un modèle de Rasch. *Mesure et évaluation en éducation*, 36(2), 59-85. doi: 10.7202/1024415ar
- MacGinitie, W. H., MacGinitie, R. K., & Dreyer, L. G. (2000). *Gates-MacGinitie reading tests* (4th ed.). Itasca, IL: Riverside Publishing Company.
- Martel, V. & Lévesque, J.-Y. (2010). La compréhension en lecture aux deuxième et troisième cycles du primaire: regard sur les pratiques déclarées d'enseignement. *Revue canadienne de linguistique appliquée*, 13(2), 27. Repéré à <https://journals.lib.unb.ca/index.php/CJAL/article/view/19883>
- McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education*, 4(1), 229-257. Retrieved from <http://files.eric.ed.gov/fulltext/EJ1070456.pdf>
- Mesnager, J. (2002). Pour une étude de la difficulté des textes: la lisibilité revisitée. *Le français aujourd'hui*, 137, 29-42. doi: 10.3917/lfa.137.0029
- Mesnager, J. (2011). *Le vocabulaire et son enseignement – Évaluation de la difficulté des textes: la part du lexique*. Saint-Étienne, France: Ministère de l'Éducation nationale, de la Jeunesse et de la Vie associative. Repéré à http://cache.media.eduscol.education.fr/file/Dossier_vocabulaire/57/4/Jean_Mesnager_111202_C_201574.pdf

- Meyer, B. J. F. (1985). Prose analysis: Purposes, procedures, and problems. In B. K. Britton & J. Black (Eds.), *Understanding expository text: A theoretical and practical handbook for analyzing explanatory text* (pp. 269-304). Hillsdale, NJ: Erlbaum.
- Ministère de l'Éducation, du Loisir et du Sport (2011). *Progression des apprentissages au primaire : français, langue d'enseignement*. Québec: Gouvernement du Québec. Repéré à www1.education.gouv.qc.ca/progressionPrimaire
- Ministère de l'Éducation et de l'Enseignement supérieur (2016). *Document d'information, Janvier 2017, juin 2017 – Épreuves obligatoires, Enseignement primaire, 3^e cycle – Français, langue d'enseignement, 6^e année du primaire*. Québec: Gouvernement du Québec. Repéré à www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/evaluation/DI_FLE_6e_primaire_Jan_Juin_2017.pdf
- Nunally, J. C. (1978). *Psychometric theory* (2nd ed.), New York, NY: McGraw-Hill.
- Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18*(4), 443-468. doi: 10.1080/01690960344000008
- O'Reilly, T., Sabatini, J., Bruce, K., Pillarisetti, S., & McCormick, C. (2012). Middle school reading assessment: Measuring what matters under a RTI framework. *Reading Psychology, 33*(1/2), 162-189. doi: 10.1080/02702711.2012.631865
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184-202. doi: 10.1598/rrq.40.2.3
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227-247). Malden, MA: Blackwell Publishing. Retrieved from [www.pitt.edu/~perfetti/PDF/Acquisition%20\(Oakhill%20chapter\).pdf](http://www.pitt.edu/~perfetti/PDF/Acquisition%20(Oakhill%20chapter).pdf)
- Popham, W. J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership, 56*(6), 8-15. Retrieved from www.ascd.org/publications/educational-leadership/mar99/vol56/num06/Why-Standardized-Tests-Don%27t-Measure-Educational-Quality.aspx
- Popham, W. J. (2001). Teaching to the test? *Educational Leadership, 58*(6), 16-20. Retrieved from www.ascd.org/publications/educational-leadership/mar01/vol58/num06/Teaching-to-the-Test%C2%A2.aspx
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: Mesa Press.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*(3), 207-230. doi: 10.3102/10769986004003207
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing, 23*(4), 441-474. doi: 10.1191/0265532206lt337oa
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26-43. doi: 10.1111/j.1540-4781.2011.01146.x
- Sweet, A. P. (2005). Assessment of reading comprehension: The RAND Reading Study Group vision. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 3-12). Mahwah, NJ: Lawrence Erlbaum Associates.

- Thériault, P. & Laurin, N. (2016). Enseigner la lecture au primaire : un aperçu des pratiques d'enseignants des 2^e et 3^e cycles du primaire. Dans I. Carignan, M.-C. Beaudry & F. Larose (dir.), *La recherche-action et la recherche-développement au service de la littératie* (pp. 15-35). Sherbrooke : Éd. de l'Université de Sherbrooke.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.
- Volante, L. (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, 35. Retrieved from <http://files.eric.ed.gov/fulltext/EJ848235.pdf>
- Wechsler, D. (2005). *Wechsler Individual Achievement Test, 2nd edition (WIAT-II)*. London: The Psychological Corp.
- Wharton-McDonald, R., & Swiger, S. (2009). Developing higher order comprehension in the middle grades. In S. Israel & G. Duffy (Eds.) *Handbook of research on reading comprehension* (pp. 510-530). New York, NY: Routledge. Retrieved from <https://lwpsfor.wikispaces.com/file/view/Developing+higher+order+comprehension+in+the+middle+grades.pdf>