

Épreuves informatisées des indicateurs des savoirs essentiels en lecture pour élèves francophones au début du primaire (INDISSE)

Computerized Measures of Reading for French-speaking Students in Kindergarten Through Grade Three Using Indicateurs dynamiques des savoirs essentiels en lecture (INDISSE)

Lise Saint-Laurent, *Faculté des Sciences de l'éducation, Université Laval, Québec*

Jocelyne Giasson, *Faculté des Sciences de l'éducation, Université Laval, Québec*

Abstract

This article presents computerized measures of reading for French-speaking students in kindergarten through grade three using the *Indicateurs dynamiques des savoirs essentiels en lecture (INDISSE)*. Their aims are identifying students at risk of difficulties with learning to read and following the progress of these students. The psychometric properties of the measures are examined with a sample of 1130 students evaluated three times during the school year (beginning, middle and end). Students at the risk, i.e. those whose result on one of the measures was below the 40th percentile, completed five assessments. The multitrait-multimethod (MTMM) matrices support the construct validity of the measures. Moreover, they are correlated with the success in reading at the end of the year and with the teacher judgement. The predictive validity of the measures on reading scores at the end of the year is higher in kindergarten (.56) and grade 1 (.34) than in grade 2 and 3 (.26). It is the same for the predictive validity of the measures on teacher judgement at the end of the year (K, .79; grade 1, .65; grade 2, .44; grade 3, .36). With all the school degrees, they are better predictors of the success than the teacher judgement at the beginning of the year. The INDISSE measures can supplement teachers' information based on their own observations in class and bring them useful data and essential knowledge of the students for their success in reading.

Résumé

Cet article présente des épreuves informatisées en lecture pour les élèves francophones de la maternelle et le début du primaire, les *Indicateurs dynamiques des savoirs essentiels en lecture (INDISSE)*. Elles visent à identifier les élèves à risque de difficulté en lecture et à suivre le progrès des élèves. Les qualités psychométriques ont été vérifiées auprès d'un échantillon composé de 1130 élèves évalués à trois reprises (début, milieu et fin d'année). Les élèves à risque, c'est-à-dire ceux dont le résultat à une des épreuves se situait sous le 40^e rang centile,

l'ont été à cinq reprises. Des analyses multitraits-multiméthodes soutiennent la validité de construit des épreuves. De plus, celles-ci sont corrélées avec la réussite en lecture à la fin de l'année et avec le jugement de l'enseignante. Leur validité prédictive des scores à l'examen de fin de l'année est plus élevée en maternelle (.56) et en 1^{re} année (.34) qu'en 2^e (.26) et 3^e année (.26). Il en est de même pour la validité prédictive du jugement de l'enseignante en fin d'année (maternelle, .79; 1^{re} année, .65; 2^e année, .44; 3^e année, .36). Par ailleurs, à tous les degrés scolaires, elles sont de meilleurs prédicteurs de la réussite que le jugement de l'enseignante en début d'année. Ces épreuves peuvent compléter les informations des enseignantes venant de leurs propres observations en classe et leur apporter des données utiles quant aux savoirs essentiels des élèves pour leur réussite en lecture.

Mots-clés

Épreuves de lecture informatisées, élèves à risque, suivi des élèves, DIBELS, validation d'épreuves, maternelle, primaire

Introduction

L'importance de la prévention de l'échec scolaire en début de scolarisation fait l'objet d'un large consensus dans le domaine de l'éducation. Pour la réaliser, il est nécessaire de bien identifier les élèves à risque de difficultés scolaires. Les problèmes en lecture sont les plus importants à prévoir parce que le savoir lire est une compétence transversale nécessaire à tous les autres apprentissages (Minskoff, 2005 ; Snow, Burns, & Griffin, 1998). Les recherches montrent que le moindre retard dans le développement de cette compétence au début du primaire a des conséquences néfastes à court et à long terme pour l'élève. Juel (1988) a trouvé que 88% des lecteurs faibles ou en difficulté en 1^{re} année le sont toujours en 4^e année. De plus, l'écart entre les lecteurs faibles et les autres élèves se creuse au fur et à mesure du cheminement scolaire (Cunningham & Stanovich, 1998 ; Phillips, Norris, Osmond, & Maynard, 2002).

La mise en place d'interventions pédagogiques préventives efficaces exige l'identification précoce des élèves à risque de difficultés en lecture (Coyne, Kame'enui, Simmons, & Harn, 2004 ; Menzie, Mahdavi & Lewis, 2008). Pour cela, il est nécessaire, en début de scolarisation, de bien répondre aux quatre questions suivantes: a) Quel élève est à risque de difficulté parce qu'il a des connaissances insuffisantes en lecture? b) Quel élève a besoin d'un enseignement particulier? c) Est-ce que l'enseignement est efficace? d) Quand l'élève n'est-il plus à risque de difficulté en lecture ? Ces questions doivent se poser d'une façon continue tout au long des premières années du primaire. Certains enfants peuvent ne pas être à risque à un moment et l'être plus tard parce qu'ils ne progressent pas suffisamment. Pour d'autres, grâce à des interventions, le risque de difficulté disparaît.

Au cours des deux dernières décennies, le courant de l'évaluation basée sur le programme scolaire (*curriculum-based assessment-CBA*) s'est développé afin de répondre au besoin de suivre de près les progrès des élèves et plus particulièrement de ceux à risque ou en difficulté. Ce modèle, dont Deno (1992 ; Deno, Fuchs, Marton, & Shin, 2001) est le chef de file, est basé sur le postulat que le programme d'études doit servir de point de référence pour identifier les besoins pédagogiques. Il ne s'agit plus de procéder à une évaluation unique à un moment donné dans le cheminement scolaire mais plutôt d'évaluer de façon continue (Simmons, Coyne, Kwok,

McDonagh, Harn, & Kame'enui, 2008 ; Stecker, Fuchs, & Fuchs, 2005). Ce faisant, le suivi des progrès donne à l'enseignant une rétroaction sur son enseignement, ce qui lui permet de mieux le différencier afin d'aider l'élève à risque (O'Connor, Fulmer, Harty, & Bell, 2005).

Dans ce modèle, on procède par l'administration d'épreuves standardisées et validées mesurant la réussite de l'élève dans les matières scolaires de base. Les consignes, la cotation, le matériel et l'interprétation des mesures sont standardisés. L'évaluation se veut facilement applicable en classe grâce à des épreuves courtes et faciles à administrer. De plus, comme celles-ci sont sensibles aux changements dans les apprentissages de l'élève, elles permettent à l'enseignant de connaître à court terme l'efficacité de son enseignement (Carnine, Silbert, Kame'enui, Tarver, & Jungjohann, 2006 ; Crawford & Torgensen, 2008). L'évaluation basée sur le programme scolaire porte sur les habiletés en lecture, en écriture et en mathématique.

Ce type d'évaluation ne vise pas à faire une évaluation globale des compétences des élèves mais plutôt à mesurer des indicateurs de réussite permettant d'identifier les élèves à risque plutôt que d'attendre que les problèmes s'amplifient (Coyne et al., 2004). Shinn (2000) compare ce type d'évaluation au thermomètre qui donne des indications de l'état de l'organisme mais qui ne donne pas un diagnostic comme tel. L'identification est basée sur la réussite à des épreuves scolaires et sur la trajectoire des progrès en comparaison avec les pairs de la classe (de l'école ou de la commission scolaire). Pour ce faire, on évalue chaque enfant à intervalles de temps fixes (par exemple, chaque mois). L'élève est comparé à des normes locales obtenues suite à l'administration des épreuves à toute la classe (ou l'école) deux ou trois fois durant l'année scolaire.

Développé par un groupe de chercheurs de l'Université de l'Orégon (Good & Kaminski, 2002 ; 1996 ; Good, Gruba, & Kaminski, 2002), *DIBELS (Dynamic indicators of basic early literacy skills)* s'inscrit dans l'approche d'évaluation basée sur le programme avec la particularité que les mesures portent sur les savoirs essentiels en lecture sous-tendant tout programme scolaire dans cette discipline au début du primaire (Good, Simmons, Kame'enui, & Kaminski, 2001 ; Kaminski & Good, 1996). S'adressant à des élèves de la maternelle à la 3^e année, les épreuves sont largement utilisées aux États-Unis [dans plus de 14,000 écoles en 2007-2008 (<http://dibels.uoregon.edu/data/index.php>)].

Les auteurs se sont basés sur les recherches qui ont identifié les variables prédisant la réussite en lecture (Good et al., 2001 ; National Reading Panel, 2000) afin de développer leurs épreuves visant à mesurer les indicateurs des connaissances en lecture, que sont la conscience phonologique, la connaissance des lettres, du principe alphabétique, le décodage, la fluidité en lecture orale et la compréhension de texte. Dans *DIBELS*, des évaluations sont effectuées auprès de tous les élèves de la classe au début, au milieu et à la fin de l'année. Elles servent à déterminer les élèves à risque pour lesquels l'enseignement devra être différencié et à suivre leur progrès tout au long de l'année (Coyne et al. 2004 ; Schilling, Carlisle, Scott, & Zeng, 2007).

Plusieurs études confirment la validité de ces épreuves pour identifier les élèves à risque de difficulté en lecture. À la maternelle, Elliott, Lee et Tollefson (2001) ont trouvé que les épreuves administrées à la fin de l'année expliquent 73% des scores au Skills Cluster du Woodcock-Johnson-Revised (incluant des tâches d'identification et d'écriture de lettres) ainsi que 52% du jugement de l'enseignante sur le niveau d'éveil à la lecture (échelle de réponse allant de 1 à 5). Par ailleurs, les épreuves de maternelle ont également une valeur prédictive sur la réussite en

lecture à la fin de la 1^{re} année. Rouse et Fantuzzo (2006) ont trouvé que les scores aux trois épreuves de fin de la maternelle (Lettres, Segmentation de phonèmes et Pseudo-mots) expliquent 52% de la variance du niveau de lecture à la fin de la 1^{re} année évalué avec le Developmental Reading Assessment (Beaver, 1997). Les épreuves DIBELS de début 1^{re} année (Lettres et Pseudo-mots) prédiraient la compréhension en lecture à la fin de la 1^{re} année (GRA+DE, sous test Compréhension) avec un taux de précision de 71% et de 66% à la fin de la 2^e année (TerraNova, sous test Reading) ; celles de fin de 1^{re} année (Fluidité et Rappel) prédiraient la compréhension à 80% en 1^{re} année et à 74% en fin de 2^e année (Riedel, 2007). Schilling et al. (2007) ont trouvé que les épreuves administrées à l'automne et à l'hiver prédisent significativement le score total de lecture au Iowa Tests of Basic Skills en fin de 1^{re}, 2^e et 3^e année, la variance expliquée allant de 39% (pour les épreuves de début de 1^{re} année) à 57% (pour celles de l'hiver de la 2^e année).

En plus de ces travaux sur la validité prédictive des épreuves, plusieurs autres ont porté sur leur validité concurrente. Les résultats à chaque type d'épreuves ont été mis en lien avec les habiletés en lecture. La connaissance des lettres étant considérée comme le meilleur prédicteur des habiletés en lecture au début du primaire (voir recension des écrits de Scarborough, 1998), les scores à l'épreuve DIBELS des lettres en maternelle se sont avérés corrélés avec un test de préalable (Metropolitan Readiness Test) ($r=.70$) et avec le jugement de l'enseignante ($r=.75$) (Kaminski & Good, 1996). Elliott et al. (2001) obtiennent une corrélation de .75 entre l'épreuve des lettres et le score au Skills Cluster du Woodcock-Johnson ainsi qu'entre l'épreuve et le résultat d'élèves sur une échelle des préalables en lecture remplie par l'enseignant (Developing Skills Checklist, Pre-Reading Total Score). Un coefficient plus faible de .52 est rapporté par Rouse et Fantuzzo (2006) entre l'épreuve des lettres à la maternelle et le test TerraNova (comprenant trois sous-tests : Lecture, Vocabulaire, Langage) et de .62 entre la même épreuve et le niveau de développement en lecture (Developmental Reading Assessment, Beaver, 1997). Quant aux épreuves de lettres en 1^{re} année, selon les informations fournies sur le site internet DIBELS (Assessment Committee, 2002), leurs corrélations avec le score au Reading Cluster du Woodcock-Johnson (incluant les sous-tests d'identification de lettres et de compréhension en lecture) seraient en moyenne de .53. Schilling et al. (2007) rapportent un coefficient similaire ($r=.57$) entre l'épreuve en début d'année et le score de lecture à l'Iowa Test of Basic Skills (incluant les sous tests de Vocabulaire et de Compréhension). Par contre, Riedel (2007) obtient une corrélation de .15 entre l'épreuve de début d'année et les scores au sous-test de compréhension du GRA+DE.

En ce qui concerne les épreuves de conscience phonologique, celles-ci servent à mesurer une habileté qui s'avère le deuxième meilleur prédicteur de la réussite en lecture au début du primaire (voir recension des écrits de Scarborough, 1998). La conscience phonologique correspond à la capacité d'identification et de manipulation à l'oral d'unités linguistiques telles que les syllabes, l'attaque, la rime et les phonèmes. La conscience phonémique est le niveau le plus élevé de conscience phonologique. Pour apprendre à lire, l'enfant doit faire la correspondance entre les graphèmes et les phonèmes dans les mots, ce qui nécessite la conscience des phonèmes qui composent les mots. Dans les épreuves DIBELS, la conscience phonémique est mesurée par des tâches d'identification du phonème initial et de segmentation de phonèmes. À la maternelle, des coefficients de corrélation entre .28 et .51 entre les épreuves de maternelle et le score Skills Cluster du Woodcock-Johnson sont rapportés (Assessment Committee, 2002). Rouse et Fantuzzo (2006) obtiennent une corrélation de .43 avec le score du

sous-test de lecture du TerraNova. Par ailleurs, Elliott et al. (2001) trouvent que le résultat à cette épreuve d'identification du phonème initial corrèle négativement avec le jugement de l'enseignante. En 1^{re} année, Kaminski et Good (1996) ne trouvent pas de relation entre les scores de Segmentation des phonèmes et le Stanford Diagnostic Reading Test. Mais sur le site Internet DIBELS (Assessment Committee, 2002), un coefficient r moyen de .27 est rapporté entre les différentes épreuves en cours d'année et le score au Skills Cluster du Woodcock-Johnson. Un coefficient moyen de .38 entre les trois épreuves de 1^{re} année et le score de lecture de fin d'année est obtenu par Schilling et al. (2007). Selon Riedel (2007), la mesure de la conscience phonologique en 1^{re} année a une faible valeur prédictive de la compréhension en lecture parce que rendus au milieu de l'année, les lecteurs faibles maîtrisent suffisamment les habiletés phonologiques, ce qui fait que la mesure de la segmentation phonologique distingue mal les bons et les mauvais lecteurs à ce degré scolaire.

L'épreuve DIBELS des pseudo-mots en 1^{re} année évalue la connaissance du principe alphabétique c'est-à-dire l'habileté de l'élève à associer les lettres avec les sons ou à utiliser ces sons pour former des mots. L'élève doit lire le plus rapidement possible en une minute des non mots composés de voyelle-consonne (exemple : ov) et de consonne-voyelle-consonne (exemple : sig). On n'utilise pas de vrais mots pour éviter que l'élève ne reconnaisse des mots mémorisés et qu'il soit obligé de faire du décodage. Au début de la 1^{re} année, Hagan-Burke, Burke et Crowder (2006) rapportent des corrélations de .73 et de .75 entre l'épreuve et les scores de lecture globale (Sight Words) et de décodage (Phoneme Decoding Efficiency) d'un test standardisé (Test of Word Reading Efficiency, Torgesen, Wagner, & Rashotte, 1997). À la fin de la 1^{re} année, Riedel (2007) obtient une corrélation autour de .45 pour les trois épreuves de Pseudo-mots mises en lien avec une épreuve standardisée de compréhension (GRA+DE).

À partir du milieu de la 1^{re} année jusqu'à la fin de la 3^e année, les épreuves DIBELS évaluent l'habileté des élèves à lire rapidement et avec précision un texte. Selon les recherches, cette habileté serait le meilleur prédicteur de la réussite en compréhension (Fuchs, Fuchs, Hosp, & Jenkins, 2001). En effet, la lecture fluide libère l'énergie cognitive nécessaire pour comprendre un texte. Dans DIBELS, la fluidité en lecture (nombre de mots lus correctement en une minute) est calculée à partir de la lecture orale d'un texte qui fait par la suite l'objet d'une mesure de compréhension. Roberts, Good et Corcora (2005) obtiennent une corrélation de .76 entre les scores à l'épreuve de fluidité en 1^{re} année et les scores au Broad Reading Cluster du Woodcock Diagnostic Reading Battery (combinant l'identification de mots, la fluidité et la compréhension de texte). Par ailleurs, en 3^e année, Good et al. (2001) rapportent une corrélation de .67 entre l'épreuve de fluidité et une mesure standardisée de compréhension en lecture (Oregon Statewide Assesement-Reading/Literature). Toujours au même degré scolaire, Roehrig, Petscher, Nettles, Hudson et Torgesen (2008) rapportent des résultats similaires ($r=.66$ à l'automne et $r=.71$ à l'hiver) entre cette mesure et une épreuve nationale de fin d'année de la compréhension en lecture (Florida Comprehensive Assessment Test, FCAT-SSS). Ils obtiennent le même niveau de corrélation (de .68 à .72) entre les épreuves et un test standardisé de compréhension en lecture (Stanford Achievement Test, SAT-10). Donc, dans l'ensemble les résultats des études confirment que la mesure de fluidité dans DIBELS est en lien avec la compréhension en lecture en fin d'année scolaire. Cependant, Schilling et al. (2007) trouvent que le lien entre les deux diminue de la 2^e à la 3^e année. Les chercheurs concluent que plus l'élève devient habile en lecture, plus d'autres facteurs que la fluidité entrent en ligne de compte dans la compréhension en lecture dont le vocabulaire et les stratégies telles les inférences.

Quant à la mesure de compréhension en lecture dans DIBELS, elle est administrée du milieu de la 1^{re} année à la fin de la 3^e année et consiste dans une tâche de rappel où l'élève doit redire dans ses mots l'information contenue dans un texte qu'il vient de lire. Les éléments rappelés sont ensuite cotés en fonction d'une grille. Deux études (Riedel, 2007 ; Roberts et al., 2005) rapportent des coefficients de corrélation modérés allant de .41 à .51 entre les scores à ces épreuves et ceux obtenus à un test standardisé mesurant la compréhension en lecture à la fin de la 1^{re} année soit le sous-test de compréhension du GRA+DE et le Broad Reading Cluster du Woodcock Diagnostic Reading Battery. Par ailleurs, Pressley, Hilden et Shankland (2005) ne trouvent pas de relation entre le rappel en 3^e année et la compréhension mesurée par un test standardisé (sous-test Lecture du TerraNova). Suite à ces résultats, ces chercheurs soulèvent de possibles problèmes de fiabilité dans l'administration et la cotation de ce type de mesure. De plus, ils se questionnent sur la validité des épreuves de Rappel dans DIBELS. Comme le souligne Riedel (2007), les études sur les qualités psychométriques des épreuves de compréhension dans DIBELS sont peu nombreuses. Des travaux sont nécessaires pour mieux en établir la validité et la fiabilité.

Les épreuves DIBELS sont destinées aux élèves anglophones. Une version francophone des épreuves d'un degré scolaire, soit de la 2^e année, ont été développées par Dufour-Martel (2003) mais elles furent administrées à des élèves anglophones de deux classes d'immersion en français. Les épreuves se sont avérées corrélées avec un test d'immersion en français (French Immersion Achievement Test). Ainsi, à l'heure actuelle, les épreuves DIBELS ne sont pas disponibles pour évaluer les savoirs essentiels en lecture des élèves francophones de la maternelle à la 3^e année

Buts de l'étude

Largement répandues dans les milieux scolaires américains, les épreuves DIBELS sont de plus en plus utilisées en recherche sur la prévention de l'échec en lecture. Plusieurs études en confirment la valeur psychométrique bien que des travaux restent à faire entre autres avec les épreuves de compréhension. Mais globalement, les épreuves DIBELS se sont avérées très utiles pour identifier les élèves à risque et faire le suivi des progrès dans l'apprentissage de la lecture en langue anglaise. Mais comme un tel type d'instrument n'est pas disponible en français, on n'en connaît pas l'utilité dans l'apprentissage de cette langue et dans la prévention des difficultés en lecture d'élèves francophones.

La présente recherche vise à développer et à tester la validité des *Indicateurs dynamiques des savoirs essentiels en lecture (INDISSE)* pour des élèves francophones de maternelle et du début du primaire (1^{re} à 3^e année). Les épreuves sont informatisées afin d'en favoriser la fiabilité en contrôlant les conditions d'administration et de cotation des épreuves et aussi afin d'en faciliter l'accès aux enseignantes dans les écoles.

Méthodologie

L'échantillon comprend 1130 élèves (603 garçons, 527 filles) de 62 classes provenant de neuf écoles de la ville de Québec situées en milieu socioéconomique moyen. Le nombre d'élèves par degré scolaire est le suivant : en maternelle, 274 élèves (18 classes) ; en 1^{re} année, 256 élèves (15 classes) ; en 2^e année, 323 élèves (16 classes), en 3^e année, 277 élèves (13 classes).

Tous les élèves ont été évalués en début d'année, au milieu et à la fin de l'année. Les élèves à risque, c'est-à-dire ceux dont le score se situe sous le seuil critique du 40^e rang centile à une des épreuves, ont été évalués quatre ou cinq fois durant l'année.

En maternelle, deux épreuves ont été développées en cinq versions de niveau de difficulté équivalent : Connaissance des lettres et Reconnaissance du phonème initial. En 1^{re} année, les épreuves de Catégorisation de phonèmes et d'Identification de pseudo-mots ont été développées en cinq versions et les épreuves de Connaissances des lettres et de Compréhension de phrases en trois versions. En 2^e et 3^e années, les mesures incluent cinq versions des épreuves de Décision lexicale et de Compréhension de texte.

Description des épreuves INDISSE

Bien qu'au départ, les épreuves aient été inspirées de DIBELS, les deux présentent des différences importantes quant aux tâches proposées aux élèves. Ces différences sont dues principalement à la langue qui n'est pas la même et au support matériel distinct (informatique vs papier/crayon). De plus, INDISSE ne comprend pas d'épreuves de fluidité en lecture parce que l'informatisation d'une telle épreuve aurait nécessité des ressources importantes au plan de la programme informatique.

Connaissance des lettres

L'épreuve a pour objectif d'évaluer l'habileté de l'élève à identifier les lettres minuscules. Quatre lettres sont présentées à l'écran de l'ordinateur accompagnées d'un enregistrement audio du type : « *Clique sur la lettre z* ». À l'aide de la souris, l'élève doit cliquer sur la lettre nommée qui est présentée avec trois autres lettres. Le score obtenu par l'élève correspond au nombre de lettres correctement identifiées en une minute. Comme les lettres reviennent plusieurs fois dans l'épreuve, le score des élèves peut être supérieur à 26.

Reconnaissance du phonème initial

L'épreuve a pour objectif d'évaluer une habileté de conscience phonologique chez les élèves de maternelle qui consiste à reconnaître un phonème initial dans des mots. Dans de cette épreuve, quatre images sont présentées à l'écran de l'ordinateur et sont nommées tour à tour par un enregistrement audio. Une directive demande ensuite à l'enfant de cliquer sur l'image qui débute par le son prononcé, par exemple: « *Sapin, four, chou, oreille. Clique sur le mot qui commence par le son /ch/.* » Le score obtenu par l'élève correspond au nombre d'items réussis en une minute.

Catégorisation du phonème initial

L'épreuve a pour objectif d'évaluer l'habileté des élèves de 1^{re} année à catégoriser les mots selon leur phonème initial. Dans cette épreuve, quatre images sont présentées sur l'écran de l'ordinateur et nommées tour à tour. Une directive enregistrée demande ensuite à l'enfant de cliquer sur le mot qui ne commence pas comme les autres. Par exemple: « *Regarde les quatre images et écoute bien le début de chaque mot que je vais dire. Voici boîte, menton, bonhomme, bain. Clique sur le mot qui ne commence pas comme les autres.*» Le score obtenu par l'élève correspond au nombre d'items réussis en deux minutes.

Identification de pseudo-mots

L'épreuve a pour objectif de vérifier si l'élève de 1^{re} année a compris et applique le principe alphabétique. La tâche consiste pour l'élève à choisir le bon pseudo-mot parmi un choix de trois (exemple : Lilo, Libo, Lico). La directive enregistrée suivante est donnée : « Sur l'écran, tu verras l'image d'un animal et trois noms écrits en dessous. Lorsque tu entendras le nom de cet animal, tu devras cliquer sur son nom ». Le score obtenu par l'élève correspond au nombre de mots bien identifiés en une minute.

Décision lexicale

L'épreuve a pour objectif de mesurer l'exactitude et la rapidité dans l'identification de mots chez des élèves de deuxième année et troisième année. Dans cette épreuve, on présente à l'élève un mot au centre de l'écran de l'ordinateur en lui demandant d'indiquer s'il s'agit d'un vrai mot ou d'un pseudo-mot. La consigne enregistrée est la suivante : « Sur l'écran de l'ordinateur tu vas voir apparaître des suites de lettres. Tu devras indiquer le plus rapidement possible si ce que tu vois est un vrai mot ou un pseudo-mot, c'est-à-dire un mot qui n'existe pas. Clique sur le bonhomme sourire si ce que tu vois est un vrai mot et sur le bonhomme grimace si ce que tu vois est un pseudo-mot ». Le score obtenu par l'élève correspond au nombre de mots correctement identifiés en une minute. Le degré de difficulté des mots (plus complexes, plus longs, plus rares) est plus élevé en 3^e qu'en 2^e année.

Compréhension de phrases

L'épreuve a pour objectif d'évaluer la compréhension en lecture de l'élève de 1^{re} année. Lors de cette épreuve, trois phrases sont présentées à l'écran ainsi qu'une image. L'enfant doit lire silencieusement les trois phrases et cliquer sur celle qui représente le mieux l'image affichée. Le score obtenu par l'élève correspond au nombre de phrases réussies en trois minutes.

Compréhension de textes

L'épreuve a pour objectif d'évaluer la compréhension écrite des élèves de 2^e et 3^e année. La tâche consiste à lire un texte présenté à l'écran et à répondre à 12 questions de compréhension. Chaque question est accompagnée de quatre options de réponses (dont l'option Je ne sais pas). Les questions et ses options de réponses sont lues à l'élève par enregistrement audio et celui-ci doit cliquer sur la réponse de son choix. La consigne enregistrée est la suivante: « Voici les différentes questions sur le texte que tu viens de lire. Pour chaque question, il y a quatre choix de réponses. Après chaque question, il faut que tu indiques ta réponse en cliquant sur ton choix. Si tu ne sais vraiment pas la réponse, tu peux cliquer sur Je ne sais pas. » Le score obtenu par l'élève correspond au nombre de bonnes réponses. Encore ici, les textes sont plus difficiles en 3^e année qu'en 2^e, les phrases sont plus longues et le pourcentage de mots longs (neuf caractères et plus) est plus élevé.

Mesures de la réussite des élèves

La validité des épreuves a été testée en reliant les résultats des élèves avec les scores en lecture à la fin de l'année et avec le jugement de l'enseignante en début et en fin d'année. À la maternelle, afin de faire un bilan des habiletés des élèves quant à l'éveil à la lecture, une traduction française de la sous-échelle Lecture du *Teacher Rating of Oral Language and Literacy (TROLL)*, Dickinson,

McCabe, & Sprague, 2001) a été utilisée. La sous-échelle comprend 11 items portant sur des aspects tels que l'intérêt de l'élève envers les livres, sa reconnaissance de mots, sa connaissance du lien entre des sons et des lettres. Les qualités psychométriques du *TROLL*, testées pour l'ensemble des 25 items répartis en trois sous-échelles (Langage, Lecture et Écriture), révèlent une bonne validité pour établir le niveau de développement des enfants du préscolaire en langage oral et écrit (Dickinson et al., 2001).

À la fin de l'année scolaire des élèves de 1^{re}, 2^e et 3^e années, un examen collectif de lecture leur fut administré. Les examens utilisés sont typiques de ceux utilisés dans les écoles québécoises pour les examens de fin d'année en lecture. À la fin de la 1^{re} année, l'examen de lecture mesure l'habileté des élèves à lire un texte à caractère informatif. Les élèves doivent identifier dans le texte les informations recherchées en effectuant une bonne sélection de l'information. D'une durée de 25 à 35 minutes, l'examen est corrigé sur 13 points. Une procédure d'accord inter-juges a été effectuée sur le score total avec 20% des protocoles en utilisant la formule Accords /Accords + Désaccords X 100. Un accord de 93% fut ainsi obtenu.

L'examen de lecture administrée à la fin de la 2^e année évalue l'habileté des élèves à lire un texte narratif. Ceux-ci doivent répondre à neuf questions de compréhension qui portent sur les parties essentielles du récit (personnage, problèmes, résolution). L'épreuve est d'une durée de 90 minutes. Le nombre maximum de points accordés est de 25. Le calcul de l'accord inter-juges selon la procédure décrite au paragraphe précédent donne un accord de 90.7%. L'examen de fin de 3^e année mesure aussi l'habileté des élèves à lire un texte narratif. Ceux-ci doivent répondre à des questions qui les amènent à reconstruire le déroulement du récit, à dégager les sentiments des personnages, à réagir au texte, à établir des liens entre les phrases et à reconstruire le sens des phrases. D'une durée de 120 minutes, l'épreuve est sur 33 points. L'accord inter-juges calculé sur 20% des protocoles est de 96 %.

Le jugement des enseignantes sur le niveau de réussite des élèves a été demandé à deux moments durant l'année scolaire : en début (septembre/octobre) et à la fin (mai/juin). En 1^{re}, 2^e et 3^e année, elles devaient indiquer sur une échelle de 1 (faible) à 5 (fort), le niveau de chacun de ses élèves en lecture. En maternelle, il s'agissait d'indiquer le niveau d'éveil à la lecture avec la même échelle en 5 points.

Résultats

Consistance interne des épreuves

La fiabilité des épreuves a été vérifiée au moyen de coefficients Kuder-Richardson 20 calculés sur chacune des versions des épreuves, c'est-à-dire des trois versions en 1^{re} année des épreuves de Connaissance des lettres et de Compréhension de phrases et des cinq versions de toutes les autres épreuves. Le tableau 1 présente les résultats obtenus. La fiabilité des mesures est adéquate en dehors de l'épreuve des pseudo-mots.

Validité convergente et divergente des épreuves

Afin d'analyser la validité des épreuves, une matrice multitrait-multiméthode (MTMM) a été effectuée pour chacun des degrés scolaires, les traits multiples consistant dans les types d'épreuves et les méthodes multiples, les temps de mesure. Les mesures répétées peuvent être considérées comme des méthodes multiples dans ce type d'analyse (Campbell & O'Connell,

1967 ; 1982 ; Marsh & Butler, 1984 ; Marsh & Hocevar, 1988). Pour ces analyses, les temps de mesure 1, 3 et 5 seulement ont été retenus parce qu'aux temps 2 et 4 où les élèves à risque uniquement ont été évalués, le nombre de sujets est insuffisant. De plus, dans ces matrices MTMM, ce ne sont pas toutes les épreuves INDISSE qui ont été incluses, seules des épreuves administrées à chacun des trois temps de mesure 1, 3 et 5 sont retenues. Par exemple, l'épreuve de Compréhension de phrases en 1^{re} année ne fut pas incluse parce qu'elle ne fut pas administrée au temps 1. Le tableau 2 présente la matrice MTMM calculée pour chacun des quatre degrés scolaires. Nous analysons ces matrices en fonction des conditions suggérées par Campbell et Fiske (1959).

Tableau 1: Indices de fiabilité pour les différentes versions des épreuves

Degré scolaire	Épreuve	Temps de mesure		
		Septembre/ Octobre	Janvier/ Février	Mai/ Juin
Maternelle	Lettres	0,8854	0,8939	0,8746
	Reconnaissance des phonèmes	0,6203	0,6280	0,6559
1 ^e année	Lettres	0,8592	0,8736	-
	Catégorisation des phonèmes	0,7531	0,7263	0,7797
	Pseudo-mots	0,5312	0,6214	0,5427
	Compréhension de phrases	-	0,6278	0,6467
2 ^e année	Décision lexicale	0,8744	0,8897	0,8870
	Fluidité en lecture silencieuse	0,8867	0,8948	0,9400
	Compréhension de texte	0,8309	0,7962	0,6450
3 ^e année	Décision lexicale	0,8758	0,9161	0,9083
	Fluidité en lecture silencieuse	0,8598	0,9292	0,8198
	Compréhension de texte	0,8420	0,7851	0,8456

D'abord, dans chacune des quatre matrices, la diagonale monotrait multiméthode (en gras et souligné), qui représente la validité convergente, est conforme aux attentes pour les quatre degrés scolaires, et traduit une bonne validité convergente sauf en 1^{re} année pour l'épreuve des Pseudo-mots. Le r convergent moyen est de .60 en maternelle (allant de .42 à .71), de .43 (allant de .24 à .63) en 1^{re} année, de .55 (allant de .39 à .72) en 2^e année et de .61 (allant de .44 à .77) en 3^e année.

Par ailleurs, la moyenne des coefficients des blocs hétérotraits-monométhodes ainsi que des blocs hétérotraits-multiméthodes est plus faible que la moyenne des coefficients des diagonales principale et secondaire. Pour la maternelle, le r convergent moyen est de .60 et le r divergent moyen est de .42 ; pour la 1^{re} année, le r convergent moyen est de .43 et le r divergent moyen est de .38 ; pour la 2^e année, le r convergent moyen est de .55 et le r divergent moyen est de .35 et enfin pour la 3^e année, le r convergent moyen est de .61 et le r divergent moyen est de .41. Cela signifie qu'à chaque degré scolaire, la validité discriminante des épreuves est bonne.

Tableau 2. Matrices multitraits-multiméthodes pour les épreuves INDISSE de la maternelle à la 3^e année

		M1 : Sept./Octobre		M2 : Janv./Février		M3 : Mai/Juin	
Maternelle		Lettres	Phonèmes	Lettres	Phonèmes	Lettres	Phonèmes
M1(n=271)	Lettres						
	Phonèmes	0,468					
M2(n=260)	Lettres	0,696	<i>0,318</i>				
	Phonèmes	0,452	0,442	0,515			
M3(n=267)	Lettres	0,594	<i>0,266</i>	0,71	<i>0,413</i>		
	Phonèmes	<i>0,422</i>	0,437	<i>0,474</i>	0,56	0,5	
1 ^{re} année		Phonèmes	Pseudo-mots	Phonèmes	Pseudo-mots	Phonèmes	Pseudo-mots
M1(n=253)	Phonèmes						
	Pseudo-mots	0,404					
M2(n=253)	Phonèmes	0,515	<i>0,252</i>				
	Pseudo-mots	0,516	0,292	0,455			
M3(n=254)	Phonèmes	0,46	<i>0,231</i>	0,638	<i>0,404</i>		
	Pseudo-mots	<i>0,386</i>	0,241	<i>0,405</i>	0,467	0,457	
2 ^e année		D.lexicale	Compréh.	D.lexicale	Compréh.	D.lexicale	Compréh.
M1(n=322)	D. lexicale						
	Compréh.	0,499					
M2(n=312)	D. lexicale	0,681	<i>0,46</i>				
	Compréh.	0,329	0,469	0,355			
M3(n=314)	D. lexicale	0,619	<i>0,469</i>	0,726	<i>0,405</i>		
	Compréh.	<i>0,22</i>	0,396	<i>0,228</i>	0,484	0,25	
3 ^e année		D.lexicale	Compréh.	D.lexicale	Compréh.	D.lexicale	Compréh.
M1(n=277)	D. lexicale						
	Compréh.	0,5					
M2(n=272)	D. lexicale	0,708	<i>0,522</i>				
	Compréh.	0,415	0,536	0,434			
M3(n=273)	D. lexicale	0,709	<i>0,484</i>	0,77	<i>0,478</i>		
	Compréh.	<i>0,28</i>	0,448	<i>0,313</i>	0,569	0,337	

Note. Les corrélations en gras correspondent aux liens entre les différentes épreuves dans le même temps de mesure (multitrait-monométhodes) ; les corrélations en italique correspondent aux liens entre les différentes épreuves administrées à différents temps de mesure (multitrait-multiméthode) ; les corrélations en gras souligné correspondent à la validité convergente (monotrait-multiméthodes). Les corrélations sont toutes significatives à $p < .01$.

Enfin, un dernière aspect analysé dans les matrices concerne les corrélations entre les différents traits qui doivent être semblables dans les trois méthodes (ou temps). En maternelle, les corrélations vont de .46 à .51, en 1^{re} année de .40 à .45, en 2^e année de .25 à .49 et en 3^e année de .33 à .50. Ainsi, la condition serait respectée avec les épreuves de maternelle et de 1^{re} année alors que les corrélations sont plus faibles en 2^e année, aux temps 3 et 5 et en 3^e année, au temps 3.

En somme, bien que toutes les conditions ne soient pas parfaitement respectées, les résultats des matrices multitraits-multiméthodes procurent un soutien pour la validité du construit des épreuves INDISSE.

Validité concomitante et prédictive

La validité des épreuves fut également vérifiée en analysant le lien entre les scores aux épreuves et la réussite à un examen de lecture en fin d'année ainsi que le lien entre les épreuves et le jugement de l'enseignante en début et fin d'année. Le tableau 3 présente les résultats des analyses corrélationnelles effectuées. Toutes les épreuves sont corrélées avec la réussite à l'examen de lecture de fin d'année et avec le jugement de l'enseignante sur la compétence des élèves en lecture en début et en fin d'année.

En maternelle, les deux épreuves sont en corrélation avec le bilan des acquis en lecture (Troll) à la fin de l'année ainsi qu'avec le jugement de l'enseignante sur le niveau d'éveil à la lecture des élèves.

En 1^{re} année, la Connaissance des lettres en début n'est pas reliée au jugement de l'enseignante en début d'année mais elle corrèle significativement (bien que modestement) avec le jugement de l'enseignante ($r=.28$) à la fin de l'année ainsi qu'avec la réussite en lecture ($r=.21$). Les épreuves de Catégorisation de phonèmes corrèlent avec la réussite à la fin de l'année surtout (r moyen de .32) et plus faiblement avec le jugement de l'enseignante (r moyen de .25). L'épreuve de Pseudo-mots en début d'année ne corrèle pas avec la réussite à la fin de l'année en lecture mais à partir du milieu de la 1^{re} année, cette mesure devient plus en lien avec les notes des élèves et le jugement de l'enseignante à la fin de l'année. La Compréhension de phrases corrèle significativement avec toutes les mesures des compétences des élèves mais le lien est en général faible à l'exception de la dernière prise de mesure (mai/juin) qui corrèle modérément ($r=.44$) avec la note à l'examen collectif de lecture.

En deuxième et troisième année, l'épreuve de Décision lexicale corrèle significativement mais modérément avec le jugement de l'enseignante ($r=.51$ et .47) et avec la réussite en lecture en fin d'année ($r=.35$ et .33). L'épreuve de Compréhension de textes en 3^e année est en lien modéré avec le jugement de l'enseignante ($r=.50$) et avec la réussite en fin d'année ($r=.50$) mais les liens sont plus faibles en 2^e année ($r=.38$ et .24).

Afin d'explorer davantage la relation entre les épreuves DIBELS et la réussite en lecture, des analyses de régression ont été effectuées. Pour chaque degré scolaire, deux types d'analyses ont été effectuées. Dans la première analyse de régression de type «enter», tous les scores des épreuves INDISSE aux temps 1, 3 et 5 sont entrés en même temps dans l'équation. Dans la seconde analyse de type hiérarchique, le jugement de l'enseignante au temps 1 (en début d'année) est entré dans le premier bloc et les mesures INDISSE sont entrées dans le second bloc.

Tableau 3. Coefficients de corrélation entre les résultats aux épreuves aux temps 1, 3 et 5 (début, milieu et fin d'année) et le jugement de l'enseignante en début et fin d'année ainsi que les scores en lecture à la fin de l'année

Degré scolaire	Épreuves	Version (temps)	Jugement Enseignante Septembre	Jugement Enseignante Juin	Examen Juin	
Maternelle	<i>Connaissance des lettres</i>	1	.43**	.62**	.50**	
		3	.35**	.56**	.49**	
		5	.32**	.46**	.42**	
	<i>Reconnaissance des phonèmes</i>	1	.41**	.47**	.36**	
		3	.30**	.53**	.52**	
		5	.40**	.55**	.51**	
	1 ^{re} année	<i>Connaissance des lettres</i>	1	.10	.28**	.21**
			3	.01	.09	.13*
			5	.01	.09	.13*
<i>Catégorisation des phonèmes</i>		1	.20**	.36**	.21**	
		3	.15*	.21**	.38**	
		5	.18**	.18**	.38**	
<i>Pseudo-mots</i>		1	.09	.21**	.04	
		3	.11	.21**	.39**	
		5	.17**	.28**	.37**	
<i>Compréhension de phrases</i>	3	.20**	.22**	.29**		
	5	.22**	.21**	.44**		
	5	.22**	.21**	.44**		
2 ^e année	<i>Décision lexicale</i>	1	.16**	.38**	.25**	
		3	.16**	.45**	.21**	
		5	.16**	.51**	.35**	
	<i>Compréhension de textes</i>	1	.23**	.48**	.36**	
		3	.27**	.45**	.34**	
		5	.20**	.38**	.24**	
	3 ^e année	<i>Décision lexicale</i>	1	.32**	.40**	.28**
			3	.26**	.45**	.27**
			5	.29**	.47**	.33**
<i>Compréhension de textes</i>		1	.38**	.43**	.28**	
		3	.35**	.33**	.27**	
		5	.33**	.50**	.50**	

** p<.01 ; * p<.05 ; Temps 1=Septembre/octobre ; 3=Janvier/février ; 5=Mai/juin.

Cette analyse permet d'analyser la contribution unique des épreuves à la réussite en lecture à la fin de l'année, après avoir tenu compte de la variance expliquée par le jugement de l'enseignant en septembre. Deux variables prédites sont analysées : le score de l'élève à l'examen de lecture en fin d'année et le jugement de l'enseignante des habiletés de l'élève en lecture en fin d'année. Les valeurs R^2 ajustées obtenues dans ces huit analyses de régression sont présentées au tableau 4.

Tableau 4. Sommaire des résultats des analyses de régression simples et hiérarchiques avec l'examen de lecture et le jugement de l'enseignante en fin d'année comme variables dépendantes et toutes les épreuves INDISSE (en début, milieu et fin d'année) comme variables indépendantes

Modèle	Variable prédite: Score à l'examen de lecture en fin d'année			Variable prédite: Jugement de l'enseignante en fin d'année		
	R^2	dl	F	R^2	dl	F
<u>Maternelle</u>						
Modèle 1 Épreuves INDISSE	.42	1,193	23.33***	.54	1,252	49.70***
Modèle 2 Bloc 1: Jugement Enseignante Début d'année	.39	1,193	122.93***	.39	1,250	156.36***
Bloc 2: Épreuves INDISSE	.56	1,193	33.48***	.79	1,250	58.29***
<u>1^{re} année</u>						
Modèle 1 Épreuves INDISSE	.33	1,213	9.99***	.57	1,231	29.86***
Modèle 2 Bloc 1: Jugement Enseignante Début d'année	.14	1,195	33.58***	.49	1,210	192.23***
Bloc 2: Épreuves INDISSE	.34	1,195	8.89***	.65	1,210	33.79
<u>2^e année</u>						
Modèle 1 Épreuves INDISSE	.20	1,291	12.42***	.38	1,304	30.66***
Modèle 2 Bloc 1: Jugement Enseignante Début d'année	.12	,280	41.50***	.23	1,292	88.44***
Bloc 2: Épreuves INDISSE	.26	1,280	13.79***	.44	1,292	32.95***
<u>3^e année</u>						
Modèle 1 Épreuves INDISSE	.22	1,256	18.26***	.27	1,268	25.24**
Modèle 2 Bloc 1: Jugement Enseignante Début d'année	.18	1,250	54.96***	.27	1,262	100.33***
Bloc 2: Épreuves INDISSE	.26	1,250	17.20***	.36	1,262	29.67***

*** $p < .001$, ** $p < .01$

Chacune des régressions simples sont significatives. L'ensemble des mesures INDISSE expliquent une part significative de la variance des scores en lecture à la fin de l'année (de 20% à 42%, selon le degré scolaire) et de la variance du jugement de l'enseignante (de 20% et 42%, selon le degré scolaire). Dans les régressions hiérarchiques, l'ensemble des épreuves INDISSE continuent d'être significatives à chacun des degrés scolaires. L'examen des R^2 ajustés montre qu'après avoir contrôlé pour la variance dans le jugement de l'enseignante en début de l'année, l'ajout des mesures INDISSE explique un pourcentage significatif de la variance restante dans

les scores à l'examen de lecture, 56% à la maternelle, 34% en 1^{re} année, 26% en 2^e année et en 3^e année. Elles expliquent également une part importante de la variance restante dans le jugement de l'enseignante à la fin de l'année, 79% à la maternelle, 65% en 1^{re} année, 44% en 2^e année et 36% en 3^e année. Dans les deux types de régression, et ce pour les deux variables prédites à la fin de l'année, le pourcentage de la variance expliquée par les épreuves INDISSE est plus élevé en maternelle et en 1^{re} année qu'en 2^e et 3^e année.

Discussion

Dans cette étude, de nouvelles épreuves évaluant les savoirs essentiels en lecture chez des élèves francophones de la maternelle à la 3^e année sont présentées et leur validité est testée. Dans l'ensemble, les résultats obtenus révèlent de bonnes qualités psychométriques. Les épreuves ont une consistance interne adéquate. Également, les matrices multitraits-multiméthodes en appuient la validité du construit. Plus spécifiquement, les analyses indiquent des corrélations convergentes (entre les scores de chaque épreuve aux différents temps de mesure) élevées. Par ailleurs, les corrélations convergentes sont la plupart du temps plus élevées que les corrélations entre deux épreuves évaluées au même temps et entre deux épreuves différentes administrées à deux temps différents. De plus, les corrélations convergentes sont supérieures aux corrélations divergentes. Enfin, le niveau des corrélations entre les différentes épreuves est le même à travers les différents temps de mesure en maternelle et en 1^{re} année mais il est variable en 2^e et 3^e année.

Pour chacun de ces degrés scolaires, les résultats indiquent que les épreuves sont corrélées avec les notes à l'examen de lecture à la fin de l'année et avec le jugement des enseignantes quant au niveau de réussite en lecture des élèves. À chaque degré, les épreuves prédisent la réussite en lecture des élèves à la fin de l'année scolaire mais la validité prédictive est plus élevée pour les épreuves de maternelle et de 1^{re} année que pour celles de 2^e et 3^e année. Les corrélations avec la réussite en lecture sont parfois semblables ou plus élevées que celles rapportées pour DIBELS mais le plus souvent, elles sont plus faibles. Cela peut s'expliquer par les tâches qui sont différentes dans les deux ensembles d'épreuves et surtout par le fait que les mesures INDISSE sont informatisées, ce qui peut entraîner une moins grande sensibilité des épreuves à évaluer les savoirs visés par rapport aux tests papier/crayon administrées individuellement dans DIBELS. Certains enfants peuvent être malhabiles ou stressés devant l'ordinateur. De plus, comme le soulignent Vallerand, Guay et Blanchard (2000), lors de tests informatisés, l'absence de contact entre l'évaluateur et l'élève limite l'observation des comportements qui pourrait souvent ajouter de l'information à l'évaluation.

Épreuves INDISSE de maternelle

À la maternelle, la corrélation moyenne entre les épreuves des lettres dans INDISSE et le bilan des acquis en lecture à la fin de l'année est de .47 alors que pour le même type d'épreuve dans DIBELS, des niveaux de corrélations plus élevés (r de .70 à .75) sont rapportés (Elliott et al., 2001; Kaminski & Good, 1996). Par contre, le niveau modéré de corrélation que l'on obtient se rapproche des résultats fournis par Rouse et Fantuzzo (2006). La corrélation moyenne avec le jugement de l'enseignante en fin d'année que nous obtenons avec les épreuves des lettres dans INDISSE est de .50. Encore une fois, des corrélations plus fortes avec le jugement de l'enseignante (d'environ .75) ont été obtenues avec DIBELS (Kaminski & Good, 1996). Nos résultats plus modestes peuvent s'expliquer par la tâche différente (reconnaître vs nommer le nom ou le son de la lettre) et encore une fois, par le support différent (papier vs ordinateur).

En ce qui concerne les épreuves de conscience phonologique, les corrélations avec le bilan des acquis en lecture à la fin de la maternelle que l'on obtient vont de .36 à .51 (coefficient r moyen de .44). Ces résultats sont semblables à ce qui est rapporté pour le même type d'épreuve dans DIBELS (Assessment Committee, 2002 ; Rouse & Fantuzzo, 2006) avec une tâche similaire d'identification du phonème initial. L'informatisation des épreuves ne semble pas avoir ici joué sur la performance des élèves.

Par ailleurs, prises ensemble, les épreuves INDISSE à la maternelle prédisent les acquis des élèves en lecture à la fin de l'année. En contrôlant pour le jugement de l'enseignante au début de l'année, l'ajout des épreuves de lettres et de conscience phonologique expliquent 56% de la variance des scores obtenus par les élèves à la sous-échelle Lecture du TROLL et 79% du jugement de l'enseignante en fin d'année. Ces résultats confirment l'utilité des épreuves à la maternelle pour identifier les élèves à risque de difficulté en lecture. Les épreuves sont plus valides que le jugement de l'enseignante en début d'année pour prédire les acquis des élèves en lecture à la fin de l'année scolaire. Elliot et al. (2001) ont aussi trouvé que les épreuves DIBELS avaient une valeur prédictive sur un test à la fin de l'année mais le coefficient qu'ils obtiennent est plus élevé (73%). Il n'est pas surprenant qu'une telle corrélation soit obtenue puisqu'ils évaluent les mêmes habiletés dans les deux mesures qu'ils mettent en relation, soit l'identification des lettres.. Dans la présente étude, la prédiction a été effectuée avec un bilan des acquis fait par l'enseignante qui couvre un éventail beaucoup plus large de comportements émergents en lecture (dont la reconnaissance de mots et la lecture d'histoires). Par contre, la valeur prédictive de 79% des épreuves INDISSE sur le jugement des enseignantes en fin d'année est plus élevée que celle de 52% rapportée pour DIBELS par Elliott et al. (2001).

Épreuves INDISSE de 1^{re} année

En 1^{re} année, une corrélation modeste de .21 est obtenue entre l'épreuve des lettres en début d'année et la réussite en lecture à la fin de l'année et entre l'épreuve et le jugement de l'enseignante ($r=.28$). Dès le milieu de la 1^{re} année, la majorité des élèves connaissaient les lettres et un effet de plafond est obtenu. Riedel (2007) rapporte un niveau de corrélation faible ($r=.15$) entre l'épreuve des lettres DIBELS en début d'année et les scores de compréhension à un test standardisé en lecture. Cela suggère que cette mesure est surtout utile en maternelle pour identifier les élèves à risque et suivre leur progrès.

Par ailleurs, en ce qui concerne les épreuves de Catégorisation de phonèmes en 1^{re} année, les corrélations que l'on obtient entre ces scores et la réussite en lecture à la fin de l'année sont de .21, .38 et .38. Plus l'année avance, plus ces connaissances deviennent importantes pour lire et comprendre un texte. Ainsi, les épreuves INDISSE de catégorisation phonémique prédisent la compréhension en lecture en fin d'année. Cela confirme qu'elles fournissent des indications sur un savoir qui est essentiel pour la réussite en lecture en 1^{re} année. Concernant les épreuves DIBELS de conscience phonologique en 1^{re} année, des résultats contradictoires sont rapportées. En effet, des coefficients r moyen de .27 (Assessment Committee, 2002) et de .38 (Schilling et al. 2007) sont rapportés pour leur lien avec des scores de lecture à des tests standardisés alors que dans deux études (Kaminski & Good, 1996 ; Riedel, 2007), on ne trouve pas de corrélation significative entre les épreuves et des tests de lecture. Selon Riedel (2007), en 1^{re} année, la conscience phonologique distingue mal les bons et les mauvais lecteurs. L'auteur base son jugement sur une tâche de segmentation de phonèmes alors que dans INDISSE, il s'agit d'une tâche de catégorisation. Nos résultats contredisent les propos de Riedel (2007) car les mesures

INDISSE de la conscience phonologique donnent des informations sur les progrès des élèves qui sont reliées à leur réussite en lecture à la fin de la 1^{re} année.

Contrairement à ce qui était attendu, étant donné l'importance du principe alphabétique, l'épreuve INDISSE des Pseudo-mots en début de 1^{re} année ne corrèle pas avec les notes à l'examen de lecture en fin de 1^{re} année. Ce n'est qu'au milieu et à la fin de l'année que cette mesure devient en lien significatif avec la réussite en lecture. L'absence de corrélation entre la mesure en septembre et les notes à la fin de l'année peut s'expliquer par le degré de difficulté de la tâche pour des élèves qui commencent la 1^{re} année. En effet, plusieurs ont obtenu un score très faible à cette épreuve en septembre. Au milieu et à la fin de l'année, les épreuves de pseudo-mots sont corrélées ($r = .39$ et $.37$) avec la réussite à la fin de l'année, indiquant que ce savoir est un indicateur non négligeable pour la réussite en lecture en 1^{re} année. En comparaison avec les épreuves DIBELS, celles de Pseudo-mots dans INDISSE sont plus difficiles. En effet, les épreuves sont très différentes dans les deux systèmes de mesure. Dans DIBELS, les épreuves se passent à l'oral et les élèves doivent donner le son de chaque lettre ou lire le mot. Ils obtiennent des points dans les deux cas. Dans INDISSE, la tâche dans les épreuves de Pseudo-mots consiste à lire trois mots et à distinguer celui qui ne correspond pas au mot prononcé. À notre avis, cette difficulté plus grande des épreuves explique les corrélations moins élevées obtenus avec INDISSE en comparaison avec DIBELS. Hagan-Burke et al. (2006) ont obtenu des corrélations élevées (d'environ $.70$) entre les épreuves DIBELS de pseudo-mots en début d'année et les scores de lecture globale et de décodage à un test standardisé en lecture à la fin de la 1^{re} année. Le fait que nous ayons évalué la compréhension et non pas l'identification ou le décodage peut aussi expliquer en partie l'écart avec les résultats des épreuves DIBELS. Comme nous, Riedel (2007) a relié les épreuves DIBELS avec la compréhension en lecture. Nos résultats pour le milieu et la fin de l'année se rapprochent de ceux qu'il rapporte. Il obtient une corrélation autour de $.45$ pour chacune des trois épreuves de Pseudo-mots (début, milieu et fin d'année) mise en lien avec un score de compréhension à une épreuve standardisée de lecture à la fin de la 1^{re} année (GRA+DE). Par ailleurs, Schilling et al. (2007) ont obtenu une corrélation moyenne de $.57$ entre les trois épreuves et le sous-test de compréhension d'un test standardisé à la fin de l'année.

L'épreuve de compréhension de phrases en 1^{re} année est aussi en corrélation modérée avec les résultats à la fin de l'année. Alors qu'au milieu de l'année, cette épreuve distingue peu les lecteurs faibles, à la fin de l'année, elle corrèle à $.44$ avec la réussite en lecture. Cette corrélation est assez proche de celles rapportées (r de $.43$ à $.51$) dans deux études pour l'épreuve DIBELS (Riedel, 2007 ; Robert et al., 2005). Concernant la compréhension en lecture, rappelons que les tâches sont différentes dans INDISSE et DIBELS : compréhension de phrases par des questions à choix multiples vs rappel de récit. Il semble que ces différences n'ont pas affecté beaucoup les résultats aux deux types d'épreuves.

Bien que les épreuves INDISSE de 1^{re} année ne soient pas toutes également utiles comme indicateurs de réussite en lecture, l'ensemble de celles-ci possède néanmoins un pouvoir prédictif appréciable. Elles expliquent 34% des notes des élèves en lecture à la fin de l'année et 65% du jugement de l'enseignante. Nos résultats sont proches de ceux de Schilling et al. (2007) qui obtiennent des niveaux de prédiction de 39% pour les épreuves du début d'année et de 51% pour celles du milieu de l'année. Mais dans leur étude, à l'hiver, c'est l'épreuve de fluidité qui rend compte de 47% de la variance, les autres épreuves n'ont pas de contributions significatives. Riedel (2007) rapporte une valeur prédictrice plus élevée encore pour DIBELS : elles prédiraient

à 80% un niveau de compréhension satisfaisant en lecture à la fin de l'année. Encore ici, l'épreuve qui a le plus de poids est la fluidité en lecture orale. Cette épreuve ne fait pas partie des mesures INDISSE qui sont informatisées parce que ce type d'épreuves est très exigeant au plan de la programmation.

Épreuves INDISSE de 2^e et 3^e année

En 2^e et 3^e année, deux habiletés en lecture sont mesurées dans INDISSE: la décision lexicale et la compréhension de textes. En ce qui concerne les épreuves de décision lexicale, qui mesurent l'identification de mots, ce n'est qu'à la fin de l'année que ces épreuves deviennent en lien avec la réussite en lecture des élèves. Au début et au milieu de l'année, ces mesures semblent des indicateurs faibles de la réussite en lecture des élèves. Elles corréleront significativement mais modestement avec les notes à l'examen de lecture à la fin de l'année et avec le jugement de l'enseignante. Il est difficile de comparer ces résultats avec les épreuves DIBELS qui ne comprennent pas cette tâche de décision lexicale mais plutôt celle de la lecture orale de mots dans un texte. Cependant dans les deux cas, on mesure de la rapidité et l'exactitude dans l'identification de mots.

Quant aux épreuves de compréhension en lecture, les corrélations avec la réussite en lecture à la fin de l'année vont de .24 à .36 en 2^e année et de .28 à .50 en 3^e année. Des corrélations plus fortes étaient attendues. Il est possible que ces épreuves informatisées évaluent à la baisse les habiletés des élèves à cause de la lecture par paragraphe limitée dans le temps et de l'impossibilité de retourner en arrière pour répondre aux questions. Certains élèves peuvent ne pas avoir compris qu'ils avaient des temps minimum et maximum pour lire chaque paragraphe. Des travaux additionnels sont nécessaires pour améliorer la validité prédictive de ces épreuves. Concernant les mesures de compréhension dans DIBELS, peu d'études rapportent des données à ces degrés scolaires. En fin de 3^e année, Pressley et ses collaborateurs (2005) obtiennent une corrélation ($r=.45$) similaire à celle que nous obtenons à ce degré scolaire. Les problèmes de fiabilité dans l'administration et la cotation du rappel de texte que ces chercheurs reprochent à DIBELS ne s'appliquent pas à INDISSE où ces aspects sont standardisés. Néanmoins, l'utilité des deux types de mesures semble similaire pour distinguer les bons et faibles compreneurs en lecture.

Mais prises ensemble, que ce soit en 2^e ou en 3^e année, les deux épreuves d'identification de mots et de compréhension prédisent une part non négligeable de la réussite en lecture à la fin de l'année (26% de la variance expliquée) et du jugement de l'enseignante sur le niveau de compétences des élèves en fin d'année (36% de la variance en 2^e année et 44% de la variance en 3^e année). De plus, les épreuves prédisent mieux la réussite à l'examen de fin d'année des élèves que le jugement de l'enseignante en début d'année, ce qui suggère qu'elles peuvent être utiles aux enseignantes pour valider leur jugement en cours d'année sur les besoins pédagogiques de leurs élèves. En comparaison avec la valeur prédictive des épreuves INDISSE, Schilling et al. (2007) rapportent des niveaux de 49% et de 57% en 2^e année et de .43% et de 45% en 3^e année mais ces résultats reposent sur des épreuves différentes de celles que nous avons utilisées. En effet, les chercheurs n'ont pas utilisé les épreuves de compréhension (Rappel) mais seulement celles de la fluidité en lecture orale et d'utilisation de mot dans une phrase (*Word Usage*). Cette dernière épreuve s'est avérée avoir un poids négligeable dans la prédiction. Par ailleurs, les scores prédits sont différents également, il s'agit d'examen scolaire dans la présente étude alors

que Schilling et al. (2007) ont voulu prédire les scores de lecture à un test standardisé (Iowa Tests of Basic Skills).

En somme, les épreuves INDISSE s'avèrent avoir une bonne valeur psychométrique. Leur validité prédictive est élevée en maternelle surtout et également en 1^{re} année mais en 2^e et 3^e année, elle est plus modeste bien que non négligeable. Ainsi, elles peuvent être utiles pour identifier les élèves francophones à risque de difficultés en lecture au début du primaire. Les épreuves INDISSE seront éventuellement accessibles à la communauté francophone par Internet

Certaines limites doivent être prises en considération dans l'interprétation des données. D'abord, les travaux doivent être effectués afin d'améliorer la consistance interne de certaines épreuves de pseudo-mots. De plus, les fiabilités test-retest révélées par les matrices MTMM ne sont pas toujours très élevées, surtout en 2^e et 3^e année. De plus, des travaux additionnels sont souhaitables afin d'améliorer la validité prédictive des épreuves de compréhension en 2^e et 3^e année.

Conclusion

Les épreuves INDISSE présentées dans cette étude comportent deux aspects novateurs importants. D'abord, elles sont les premières mesures informatisées des savoirs essentiels en lecture pour les élèves de maternelle et du début du primaire. De plus, elles constituent les premières mesures de ce type destinées à des élèves francophones de ces degrés scolaires. Bien que des recherches additionnelles sont nécessaires sur ces nouvelles épreuves INDISSE, la présente étude révèle que les qualités psychométriques sont adéquates. En comparaison avec les épreuves DIBELS, leur validité prédictive est moins grande en maternelle et en 1^{re} année mais cela s'explique par les tâches qui sont différentes et par l'informatisation des épreuves. En 2^e et 3^e année, le peu d'études avec les épreuves DIBELS à ces degrés scolaires ne nous permet pas de faire beaucoup de comparaisons. L'utilisation des épreuves INDISSE dans les classes peut apporter un éclairage utile aux enseignantes quant aux savoirs essentiels des élèves francophones pour leur réussite en lecture.

Cette recherche a été effectuée grâce à une subvention du Conseil de recherches en sciences humaines du Canada (Subvention numéro 410-2005-0434).

Références

- Assessment Committee (2002, Février). Analysis of reading assessment measures, coding form for Dynamic Indicators of Basic Early Literacy Skills (En ligne). Repéré à <http://DIBELS.uregon.edu>
- Beaver, J. (1997). *Developmental reading assessment. K-3*. Glenview, IL: Celebration Press.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 30, 81-105.
- Campbell, D.T. & O'Connell, E.J. (1967). Methods factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research*, 2, 409-426.
- Campbell, D.T. & O'Connell, E.J. (1982). Methods as diluting trait relationship rather than adding irrelevant systematic variance. *New Directions for Methodology of Social and Behavioral Science*, 12, 93-111.
- Carnine, D.W., Silbert, J., Kame'enui, E.J., Tarver, S.G., & Jungjohann, L. (2006). *Teaching struggling and at-risk readers*. Columbus, OH: Merrill/Prentice Hall.
- Coyne, M.D., Kame'enui, E., Simmons, D.C., & Harn, B.A. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten instruction. *Journal of Learning Disabilities*, 37, 90-103.
- Crawford, E. & Torgesen, J. (2007). [*Teaching all students to read: Practices from "Reading First" schools with strong intervention outcomes. Rapport*](#) non publié. Florida Center for Reading Research, Florida State University, Tallahassee, FL.
- Cunningham, A.E. & Stanovich, K.E. (1998). What reading does for the mind. *American Educator*, 22(1-2), 89-97.
- Deno, S. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure*, 36, 5-10.
- Deno, S. Fuchs, L., Marton, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*, 30, 507-524.
- Dickinson, D. K., McCabe, A., & Sprague, L. K. (2001) *What is the Teacher Rating of Oral Language and Literacy (TROLL)? Can it provide an accurate measure of students' literacy acquisition skills?* Center for the Improvement of Early Reading Achievement, University of Michigan School of Education. Internet: www.ciera.org
- Dufour-Martel, C. (2003). *Assessing French reading skills of elementary French immersion students: Utility of DIBELS in French*. Thèse de doctorat non publiée, University of Oregon, Eugene, OR.
- Elliott, J., Lee, S.W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills-modified. *School Psychology Review*, 30, 33-49.

- Fuchs, L.S., Fuchs, D., Hosp, M.K., & Jenkins, J.R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5, 239-256.
- Good, R. H., Gruba, J., & Kaminski, R. A. (2002). Best practices in using dynamic indicators of basic early literacy skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology IV* (pp. 699-720), Bethesda, MD: National Association of School Psychologists.
- Good, R.H. & Kaminski, R.A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly*, 11, 126-136.
- Good, R.H. & Kaminski, R.A. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6^e édition). Eugene: OR: Institute for the Department of Educational Achievement. Adresse Internet: <http://dibels.uoregon.edu>
- Good, R.H., Simmons, D.C., & Kame'enui, E.J.. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5, 257-288.
- Hagan-Burke, S., Burke, M. D., & Crowder, C. (2006). The convergent validity of the Dynamic Indicators of Basic Early Literacy Skills and the Test of Word Reading Efficiency for the Beginning of First Grade. *Assessment for Effective Intervention*; 31; 1-15. Revue en ligne: <http://aei.sagepub.com>
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80, 437-447.
- Kaminski, R.A. & Good, R.H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, 25, 215-227.
- Marsh, H.W. & Butler, S.(1984). Evaluating reading diagnostic tests: An application of confirmatory factor analysis to multitrait-multimethod data. *Applied Psychological Measurement*, 8, 307-320.
- Marsh, H.W. & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology*, 73, 107-117.
- Menzies, H.M., Mahdavi, J.N., & Lewis, J.L. (2008). Early intervention to reading: From research to practice. *Remedial and Special Education*, 29, 57-77.
- Minskoff, E. (2005). *Teaching reading to struggling learners*. Baltimore: Paul H. Brookes.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development/National Institutes of Health.

- O'Connor, R.E., Fulmer, D., Harty, K.R., & Bell, K.M. (2005). Layers of reading intervention in kindergarten through third Grade: changes in teaching and student outcomes. *Journal of Learning Disabilities, 38*, 440-455.
- Phillips, L.M., Norris, S.P., Osmond, W.C. & Maynard, A.M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology, 94*, 3-13.
- Pressley, M., Hilden, K., & Shankland, R. (2005). *An evaluation of end-grade-3 Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Speed reading without comprehension, predicting little* (Tech. Rep.). East Lansing, MI: Michigan State University, Literacy Achievement Research Center.
- Riedel, B.W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42*, 546-567.
- Roberts, G., Good, R., & Corcora, S. (2005). Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly, 20*, 304-317.
- Roehrig, A.D., Petscher, Y., Nettles, S.M., Hudson, R.F., & Torgesen, J.K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46*, 343-366.
- Rouse, H.L. & Fantuzzo, J.W. (2006). Validity of the Dynamic Indicators for Basic Early Literacy Skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review, 35*, 341-355.
- Scarborough, H.S. (1998). Identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B.K. Shapiro, A.J. Capute, & B. Shapiro. (Éds.), *Specific reading disability: A view of the spectrum* (pp. 77-121). Hillsdale, NJ: Erlbaum.
- Schilling, S.G., Carlisle, J.F., Scott, S.E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal, 107*, 431-448.
- Shinn, M.R. (2000). Best practices in using curriculum-based measurement in a problem solving model. In National Association of School Psychologists (Éd.). *Best practice in school psychology* (pp. 671-697). Bethesda, MA: National Association of School Psychologists.
- Simmons, D., Coyne, M., Kwok, O., McDonagh, S., Harn, B., & Kame'enui, E. (2008). Indexing response to intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41*, 158-173.
- Snow, C.E., Burns, M.S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington: DC: National Academy Press.
- Stecker, P., Fuchs, L.S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: review of research. *Psychology in the Schools, 42*, 795-819.

Torgensen, J.W., Wagner, R., & Rashotte, C. (1997). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

Vallerand, R.J., Guay.F., & Blanchard, C. (2000). Les méthodes de mesure verbales en psychologie. In R.J. Vallerand & U. Hess (Éds.). *Méthodes de recherche en psychologie* (pp. 241-284). Boucherville : Gaëtan Morin Éditeur.

Auteurs

Lise Saint-Laurent
Jocelyne Giasson

La correspondance concernant cet article doit être adressée à Lise Saint-Laurent, Département d'études sur l'enseignement et l'apprentissage, Faculté des Sciences de l'éducation, Pavillon des Sciences de l'éducation, bureau 918, 2320, rue des Bibliothèques, Québec, CA, G1V 0A6.
Courriel: lise.saint-laurent@fse.ulaval.ca.



Cette création est mise à disposition sous un contrat Creative Commons.