

Juin 2022

Effets combinés d'un programme d'apprentissage en classe ordinaire et d'un programme orthopédagogique sur le rendement en lecture d'élèves en première année

CHRISTIAN BOYER

Consultant en pédagogie et en orthopédagogie, SESSIONS

STEVE BISSONNETTE

Professeur titulaire au département d'éducation de la TÉLUQ

FRÉDÉRIC MORNEAU-GUÉRIN

Professeur adjoint au département d'éducation de la TÉLUQ

Collection
Des points sur les i et des barres sur les t



Éditions de l'Apprentissage
www.editionsdelapprentissage.com



Pour toi mon amour

Je suis allé au marché aux oiseaux
Et j'ai acheté des oiseaux
Pour toi
Mon amour

Je suis allé au marché aux fleurs
Et j'ai acheté des fleurs
Pour toi
Mon amour

Je suis allé au marché à la ferraille
Et j'ai acheté des chaînes
De lourdes chaînes
Pour toi
Mon amour

Et puis, je suis allé au marché aux esclaves
Et je t'ai cherchée
Mais je ne t'ai pas trouvée
Mon amour

Jacques Prévert, poème publié en 1946 dans le recueil *Paroles*

Ce document peut être reproduit partiellement ou en totalité librement si la source est explicitement mentionnée de la manière suivante :

Boyer, C., Bissonnette, S., Morneau-Guérin, F. (2022). Effets combinés d'un programme d'apprentissage en classe ordinaire et d'un programme orthopédagogique sur le rendement en lecture d'élèves en première année. Montréal : Éditions de l'Apprentissage.



© Éditions de l'Apprentissage
www.editionsdelapprentissage.com
Dépôt légal : 2^e trimestre 2022
Bibliothèque et Archives nationales du Québec
ISBN : 978-2-923805-65-8
Version 01/06/22

Introduction

Deux programmes scolaires pour l'enseignement de la lecture en anglais voient le jour au Québec au printemps de l'année 2003, le programme *Accelerated Development of Reading* (ADOR) en première année et le programme orthopédagogique *Intensive Intervention in Reading* (IIR) qui vise la clientèle des élèves faibles en lecture d'ADOR (ainsi que les élèves faibles des degrés supérieurs). Ces deux programmes sont des traductions et des adaptations de programmes québécois existants en français : le programme *Développement accéléré de la lecture en première année* (DAL) et le programme orthopédagogique *Développement intensif du raisonnement* (DIR) en lecture (connu à ses débuts sous l'appellation de *l'Intervention intensive en lecture*). Fait à noter, la traduction et l'adaptation vers l'anglais de matériel scolaire rédigé et conçu en français sont un phénomène plutôt inhabituel au Canada et au Québec. Les versions françaises de ces programmes ont pris forme en Estrie à l'automne 1990.

La conception de ces programmes s'appuie en grande partie sur les données probantes de l'époque et d'aujourd'hui. *L'Enseignement explicite et direct* (Boyer, 1993) représente le cadre de référence de ces programmes, tout en y intégrant d'autres composantes décrites un peu plus loin. De plus, l'élaboration ainsi que la diffusion de ces programmes s'inscrivent parfaitement dans l'esprit et la lettre de la *Gestion rationnelle axée sur les résultats* (GRAR) telle que décrite

dans Boyer et Bissonnette (2021).

En 2010, Boyer publie un livre qui présente des résultats du programme DIR en lecture (Boyer, 2010). Dion (2012) et Forget (2012) font référence à cet ouvrage en indiquant que, quoique le programme DIR soit basé sur des données probantes, il n'y a pas eu à leur connaissance de validation expérimentale dudit programme. Il importe donc de mentionner que deux études quasi expérimentales furent réalisées quelques années auparavant afin de mesurer les effets des programmes ADOR et IIR (Savage, 2005 ; Savage et Deault, 2007). Ces deux études ont été effectuées par une équipe de chercheurs externes, et ce, à la demande d'une commission scolaire anglophone qui expérimentait ces programmes. Malheureusement, à notre connaissance, ces résultats n'ont jamais été diffusés et n'ont donné lieu à aucune publication officielle, en dehors des rapports déposés par l'équipe de chercheurs à la commission scolaire. Ces deux études ont mesuré les effets combinés des programmes ADOR et IIR auprès de deux cohortes distinctes d'élèves en 1^{re} année. Une première cohorte en 2004-2005 et une seconde en 2006-2007 ont appliqué ces programmes sur une période d'une année scolaire complète. Nous présentons uniquement les résultats provenant du rapport de 2007 (Savage et Deault, 2007) puisqu'il intègre les résultats du rapport de 2005¹.

Méthodologie

Tel que mentionné, des chercheurs externes et indépendants du concepteur² des programmes ont mené les deux études dont nous présentons les résultats. Les habiletés en lecture de langue anglaise des élèves ont été évaluées au début de l'année scolaire (pré-test) et à la fin de celle-ci (post-test) à l'aide d'une mesure standardisée : le *Group Reading Assessment and Diagnostic Evaluation* (GRADE)³. Une mesure arithmétique est également incluse, le *Wide Range Achievement Test* (WRAT), afin de contrôler l'effet Hawthorne⁴.

L'expérimentation se déroule sur deux années scolaires complètes (2004-2005 et 2006-2007). Au cours de l'année 2004-2005, les programmes d'enseignement explicite ADOR-IIR, décrits à la page suivante,

ont été expérimentés auprès de 117 enfants provenant de six classes de deux écoles. En 2006-2007, les programmes ADOR-IIR ont été expérimentés à nouveau auprès de 87 élèves issus de quatre (4) classes dans les deux mêmes écoles de l'année précédente. Dans deux cas, le même enseignant a participé aux deux expérimentations. **Au total, 204 élèves ont vécu les programmes ADOR-IIR et constituent le groupe expérimental. Le groupe témoin est composé de 276 enfants provenant de 13 classes liées à sept (7) écoles ayant reçu un enseignement usuel de la lecture décrit ci-dessous.** L'enseignement de la lecture, tant dans le groupe expérimental que témoin, a été réalisé par des enseignantes dont l'expérience varie entre un (1) et 37 ans. Aucune différence significative concernant

le nombre d'années d'expérience des enseignantes a été notée entre le groupe expérimental et témoin. La commission scolaire a sélectionné les écoles participantes et leurs enseignantes ont accepté de participer à l'étude à la suite d'une rencontre d'informations.

Le groupe expérimental ADOR-IIR : application du programme ADOR dans les classes ordinaires avec l'ensemble des élèves en 1^{re} année

Le programme ADOR s'adresse à l'ensemble des élèves de la 1^{re} année du primaire et vise une éclosion de la lecture dans les 80 à 100 premières journées de l'année scolaire. Le programme ADOR (version anglaise) comme le programme DAL (version française) s'inspire, entre autres, des données probantes sur l'enseignement efficace de l'apprentissage de la lecture (National Reading Panel, 2000), de l'Enseignement explicite de la compréhension en lecture (Boyer, 1993), de l'Enseignement direct (Carnine, Silbert et Kameenui, 1997), des recherches portant sur la résilience et l'optimisme (Seligman, 1995), des travaux sur la charge cognitive de Kirschner (2002) ainsi que des recherches sur les interventions comportementales et les systèmes d'économie de jetons pour encourager l'effort et le rendement scolaire (Little et Akin-Little, 2003). Le programme ADOR, comme son acolyte en français, est clairement un programme d'enseignement explicite et direct dans lequel le développement du raisonnement en lisant, des habiletés de recherche d'information dans un texte lu, de l'acquisition systématique du décodage et de l'augmentation de la fluidité incluant la prosodie occupent une place prépondérante.

ADOR est également en conformité avec les données probantes actuelles qui recommandent le développement du vocabulaire et des connaissances générales sur le monde, éléments reconnus comme constituant la clef de voûte de la compréhension en lecture (Oakhill, Cain et Elbro, 2015 ; Stuart et Stainthorp, 2016 ; Willingham, 2017).

Au début de l'apprentissage, le contenu de ADOR comprend des activités quotidiennes d'habiletés phonologiques, d'apprentissage systématique des phonèmes des lettres, du décodage de mots simples,

d'enrichissement du vocabulaire, de développement de la compréhension ainsi que du raisonnement portant sur des phrases et de courts textes lus aux élèves ou avec eux. Après une quinzaine de jours d'enseignement, certaines combinaisons de lettres parmi les plus fréquentes en anglais sont systématiquement abordées (ex. : ay, ea, oo, sh, etc.), les activités des lettres sont grandement diminuées tandis que les activités phonologiques sont carrément abandonnées ainsi que certaines activités de décodage. Cependant, les activités portant sur le raisonnement, la compréhension et le vocabulaire sont maintenues toute l'année. Une activité visant la fluidité en lecture orale est intégrée au contenu des activités quotidiennes lorsque les élèves peuvent lire à voix haute entre 12 et 19 mots-minute en débit-exact⁵.

Des évaluations formatives brèves et fréquentes sont prévues tout au long du programme afin, entre autres, d'identifier les élèves les plus faibles et d'ajuster l'enseignement ainsi que les activités en fonction du rendement moyen du groupe. Les élèves plus faibles (deux à cinq élèves) de la classe reçoivent alors une aide *ad hoc* supplémentaire du titulaire, que l'on nomme récupération, qui se réalise en individuel ou en petits groupes à raison de cinq à 10 minutes par jour pendant une dizaine de jours, avant que les élèves suivis soient évalués à nouveau. Les résultats de cette nouvelle évaluation déterminent si l'intervention fut efficace et si l'élève continuera à être suivi en récupération ou non par la titulaire.

Les titulaires des classes ADOR ont participé à quatre jours de formation au cours de l'année scolaire animés par les entraîneurs formés de la commission scolaire. Les entraîneurs assurent un suivi en classe (en présence des élèves) pour faire des démonstrations des activités pédagogiques et de certaines techniques de gestion de classe, tout en effectuant aussi des observations du degré d'application de ces contenus par les titulaires. Des grilles d'observation sont utilisées et une rétroaction est donnée au titulaire. Chaque titulaire est observée au moins à vingt reprises au cours de l'année (la durée d'une observation/démonstration incluant la rétroaction varie de 40 à 75 minutes).

Le groupe expérimental ADOR-IIR : application du programme IIR pour les élèves les plus faibles en lecture et qui bénéficient du programme ADOR en 1^{re} année

Le programme ADOR en 1^{re} année est accompagné du programme orthopédagogique IIR pour tenter de récupérer les élèves les plus faibles. Comme mentionné précédemment, le programme est une traduction et une adaptation d'un programme québécois existant en français, le programme orthopédagogique DIR en lecture (voir Boyer, 2010 pour une description plus détaillée). Le programme IIR en 1^{re} année se déroule au cours de l'hiver ou du printemps et vise les élèves les plus faibles du degré scolaire qui bénéficient aussi du programme ADOR en classe ordinaire. Ces élèves sont identifiés à l'aide d'une mesure du débit oral à partir d'un texte qui n'a jamais été lu. Une autre mesure est prise portant sur la compréhension et le raisonnement en lecture silencieuse à partir d'un texte ou de plusieurs petits textes qui n'ont jamais été lus. Les huit (8) à douze (12) élèves les plus faibles de la cohorte de 1^{re} année de l'école sont alors sélectionnés, qu'ils soient hyperactifs ou non, dyslexiques ou non, qu'ils aient des problèmes de comportement ou non, que la langue en usage à la maison soit ou non la langue de l'enseignement et que les parents soient collaborateurs ou non.

L'IIR, comme son équivalent en français (DIR en lecture), est un programme d'enseignement explicite et direct de la lecture. Ce programme orthopédagogique s'échelonne sur une période de huit (8) à dix (10) semaines, à raison de deux heures consécutives d'intervention par jour. Au total, le programme IIR doit compter au minimum 76 heures d'intervention. Les assises du programme IIR incluent celles de l'ADOR (ou du DAL en français), avec quelques particularités spécifiques à la clientèle des élèves à risque ou en difficultés d'apprentissage. De plus, certaines données empiriques spécifiques à l'intervention orthopédagogique sont aussi intégrées au programme DIR (Boyer, 2010). Le contenu de l'intervention est similaire aux activités de la classe, mais avec un tempo plus rapide, un matériel pédagogique différent et certaines adaptations au déroulement des activités ainsi qu'aux techniques de gestion de classe. L'objectif du IIR est de réduire les écarts entre les élèves qui en bénéficient et

les autres élèves de la classe ordinaire.

Les orthopédagogues du programme IIR ont participé à huit (8) jours de formation au cours de l'année scolaire animés par les entraîneurs préalablement formés de la commission scolaire. Les entraîneurs assurent un suivi en classe (en présence des élèves) pour faire des démonstrations des activités pédagogiques et de certaines techniques de gestion de classe, tout en effectuant aussi des observations du niveau d'application du personnel enseignant impliqué. Des grilles d'observation sont utilisées et une rétroaction est donnée aux orthopédagogues. Chaque orthopédagogue est observé au moins à douze reprises au cours de l'année (la durée d'une observation/démonstration incluant la rétroaction : une journée).

Le groupe témoin : application du programme du ministère de l'Éducation dans les classes ordinaires avec l'ensemble des élèves en 1^{re} année

Les classes du groupe témoin appliquent le programme du ministère de l'Éducation (Ministère de l'Éducation du Québec, 2001) dont l'application est supervisée par les conseillères pédagogiques de la Commission scolaire. Le courant du *Langage intégré* (*Whole Language*) et de l'approche constructiviste teintent fortement le discours et les recommandations des hauts responsables pédagogiques de la commission scolaire et des conseillères pédagogiques dans la manière d'appliquer le programme du ministère de l'Éducation. En bref, l'approche de l'enseignement de la lecture recommandée consiste à faire quotidiennement la lecture collective de grands textes (de grands livres ou d'albums de littérature jeunesse), à animer la lecture par des questions, à faire globaliser certains mots, à mettre l'accent sur la compréhension et la recherche de sens en lecture, à oraliser les mots en s'aidant d'une anticipation souvent faite à partir de l'illustration, du titre, du contexte et des autres mots de la phrase ou du texte. La réalisation de ce qui précède se fait idéalement dans le cadre d'un projet pédagogique « signifiant et authentique » appuyé sur les intérêts des élèves et débouchant sur une production quelconque (écrite, dessinée, etc.). L'écriture et la lecture sont intimement intégrées dans ces projets. Des évaluations formatives informelles (entrevues avec l'élève, échantil-

lons des productions de l'élève, etc.) sont privilégiées. La constitution d'un portfolio (échantillons des productions de l'élève) devant refléter le développement de l'élève en lecture et en écriture est encouragée. Le décodage (*phonics*) n'est pas une activité bannie, mais il est conseillé d'en faire brièvement et seulement en fonction des cas et des difficultés rencontrées au cours de la lecture.

La supervision de l'application de cette pédagogie est assurée par diverses formations offertes par les services éducatifs de la Commission scolaire. Les conseillères pédagogiques répondent aux demandes du personnel enseignant et offrent habituellement des formations qui peuvent donner lieu à des suivis en classe. Par contre, nous ne connaissons pas le nombre de formations et de suivis qui a été offert ni le type de rétroaction en usage.

Le groupé témoin : application d'interventions orthopédagogiques variées pour les élèves les plus faibles du programme du ministère de l'Éducation en 1^{re} année

Toutes les classes du groupe témoin bénéficient d'un service d'orthopédagogie. La sélection des élèves faibles pouvant avoir accès au service orthopédagogique est généralement faite par le titulaire, l'opération étant supervisée par la direction de l'école. Le service orthopédagogique est variable. On retrouve du dénombrement flottant (élève sorti de la classe), à raison de 20 à 40 minutes par séance d'une à trois fois par semaine, selon les difficultés de l'enfant et les disponibilités de l'orthopédagogue. Les élèves peuvent être rencontrés en individuel ou en petits groupes. Le contenu des séances d'orthopédagogie comprend un déroulement et des activités similaires à celles de la classe ordinaire, parfois en usant du même matériel et d'autres fois en usant d'un matériel différent. Certaines orthopédagogues utilisent également des activités plus ludiques dans leurs interventions. D'autres tentent d'offrir des activités complémentaires à celles de la classe ordinaire. Certaines écoles demandent à leur orthopédagogue de travailler exclusivement ou en partie directement en classe ordinaire. Généralement, l'orthopédagogue qui œuvre en classe ordinaire épaula le titulaire en intervenant auprès des élèves les plus faibles. Parfois, la titulaire et l'orthopédagogue

travaillent clairement en équipe en se répartissant les tâches et l'ensemble des élèves. La supervision du service orthopédagogique est assurée par la Commission scolaire et ses conseillères pédagogiques. Ces dernières répondent habituellement aux demandes du personnel enseignant en offrant du matériel, des formations ou un soutien. À nouveau, nous ne connaissons pas le nombre de formations et de suivis qui a été offert ni le type de rétroaction en usage.

Instruments de mesure standardisés utilisés dans les études de Savage (2005) ainsi que Savage et Deault (2007)

Le *Group Reading Assessment and Diagnostic Evaluation* (GRADE) est un test standardisé en lecture et compréhension orale possédant des normes canadiennes. Ce test a montré sa fiabilité et sa validité pour mesurer les habiletés en lecture et en compréhension orale. Il est possible de le passer à une classe entière de jeunes enfants.

Pré-tests

Les sous-tests *Lecture de mots*, *Signification de mots* et *Compréhension orale* du GRADE ont été passés lors du prétest ayant eu lieu entre le 15 septembre et le 15 octobre des années 2004 et 2006. Ces sous-tests ont été administrés à tous les enfants des classes, en grand groupe et non en individuel, et ce tant pour le groupe expérimental que le groupe témoin.

Le sous-test *Lecture de mots* demande aux enfants de marquer d'un X un des quatre mots de l'item. Le mot est d'abord lu à haute voix par l'examineur, puis utilisé dans une phrase et enfin répété. Le sous-test *Signification de mots* demande aux enfants de lire un mot et de faire un X sous l'une des quatre images qui représente le mieux la signification de ce mot. Le sous-test *Compréhension orale* demande aux enfants d'écouter une courte phrase ou un court passage et de faire un X sous l'une des quatre images qui représente le mieux le sens de l'énoncé qu'ils viennent d'entendre.

Post-tests

Les post-tests ont été administrés entre le 15 avril et la fin de juin 2005 et en mai 2007. Les sous-tests *Compréhension orale*, *Lecture de mots* et *Signification de mots* ont été réutilisés. À ceux-ci sont ajoutés deux

sous-tests supplémentaires de GRADE, la *Compréhension de phrases* et la *Compréhension en lecture de passages*. Pour le sous-test *Compréhension en lecture de phrases*, l'élève lit une phrase silencieusement et il choisit, parmi quatre mots, celui qui s'intègre le mieux à la phrase en le marquant d'un X. Pour le sous-test *Compréhension en lecture de passages*, l'élève lit silencieusement un court passage composé de trois à quatre phrases. Ensuite, ce dernier lit une question et choisit une réponse en la marquant d'un X parmi quatre choix présentés.

Effet Hawthorne

Le sous-test d'arithmétique du *Wide Range Achievement Test (WRAT)* a été passé pour contrôler l'effet Hawthorne. Les dix premiers items du sous-test d'arithmétique du WRAT ont été administrés aux élèves tant lors des pré-tests que des post-tests.

Mesures complémentaires

Des évaluations additionnelles de type qualitatif ont également été réalisées. Les chercheurs externes ont effectué des entretiens avec les enseignantes ainsi que des observations en salle de classe à l'aide du AIMS (*Atmosphere, Instruction/Content, Management, Student Engagement*) développé par Pressley et ses collaborateurs (2001) afin de poser un regard sur la qualité de l'enseignement. De plus, les élèves ont complété le questionnaire CES (*Classroom Environment Scale*; Fraser 1986) mesurant leur perception de la classe. Les résultats à ces mesures supplémentaires ont été corrélés avec les progrès des élèves en lecture.

Hypothèses⁶

Étant donné que les programmes ADOR et IIR reposent en partie sur le cadre de l'Enseignement explicite et direct tout en s'appuyant sur les données probantes concernant l'apprentissage de la lecture, trois hypothèses sont formulées :

1. Le groupe ADOR-IIR devrait manifester un rendement supérieur au groupe témoin sur les mesures directes de l'apprentissage de la lecture en 1^{re} année, principalement sur la *Lecture de mots*, la *Compréhension en lecture de phrases* et sur la *Compréhension en lecture de passages*.

Comme l'apprentissage de la lecture et la lecture en

soi peut être influencé par le développement du langage oral (Lervåg, Hulme et Melby Lervåg, 2018) :

2. Le groupe ADOR-IIR devrait manifester un rendement supérieur au groupe témoin sur les mesures de la *Signification de mots* et de la *Compréhension orale*.

Comme la lecture est très peu sollicitée en mathématique en 1^{re} année et encore moins en arithmétique :

3. Le groupe ADOR-IIR devrait montrer un rendement semblable au groupe témoin sur la mesure d'arithmétique.

Résultats

Dans cette étude, moins de 1 % des données sont manquantes et elles sont liées à l'absence d'élèves lors des évaluations. Avant de procéder aux analyses statistiques, l'équipe de recherche externe a vérifié si les données étaient normalement distribuées. Les statistiques, variant de 1 à 9 avec une moyenne de 5, ont été utilisées pour cette analyse. Les analyses montrent qu'il n'y avait pas de kurtosis ou d'asymétrie significatifs dans les données ($s < 1$, ns, et $k < 1$, ns). Toutefois, les résultats du pré-test pour l'évaluation du sous-test en arithmétique présentaient une asymétrie positive. Cette variable a été transformée en racine carrée pour la normaliser.

Le *Tableau 1* (page suivante) présente les moyennes et les écarts-types pour chacun des sous-tests de lecture lors du pré-test et du post-test pour le groupe expérimental ADOR-IIR et le groupe témoin. Une grandeur d'effet est également présentée pour chacune des mesures. Il n'y avait pas de différences significatives entre les groupes lors des pré-tests à l'exception du sous-test de compréhension orale ($F = 8,392$, $p < 0,05$). Cette différence a été prise en compte dans les analyses ultérieures réalisées (ANOVA). Les résultats obtenus lors des post-tests pour chacun des sous-tests en lecture montrent que les élèves ayant reçu le programme ADOR-IIR ont des scores plus élevés que ceux du groupe témoin. Le sous-test de la *Lecture de mots* présente une différence significative ($p < 0,05$) avantageant le groupe ADOR-IIR avec une grandeur d'effet de 0,20. La différence observée au sous-test de

Tableau 1 — Résultats combinés exprimés en stanines des pré-tests et des post-tests en lecture et compréhension orale en 2004-2005 et 2006-2007

Groupe	Pré-test	Post-test	R max	Grandeur d'effet (d)
Lecture de mots				
Groupe ADOR et IIR	3,72 (1,60)	6,00 (2,19)	9	0,20
Groupe témoin et orthopédagogie	3,46 (1,57)	5,44 (2,32)		
Signification de mots				
Groupe ADOR et IIR	3,71 (1,77)	5,78 (1,95)	9	0,46
Groupe témoin et orthopédagogie	3,36 (1,79)	4,90 (2,19)		
Compréhension orale				
Groupe ADOR et IIR	4,56 (2,04)	5,62 (1,96)	9	0,29
Groupe témoin et orthopédagogie	4,14 (1,79)	5,01 (1,90)		
Compréhension en lecture de phrases				
Groupe ADOR et IIR	Nil	5,13 (2,29)	9	0,59
Groupe témoin et orthopédagogie		4,12 (2,30)		
Compréhension en lecture de passages				
Groupe ADOR et IIR	Nil	5,05 (2,07)	9	0,84
Groupe témoin et orthopédagogie		3,57 (2,12)		

Note 1 Les écarts-types sont montrés entre parenthèses.

Note 2 Les grandeurs d'effets présentées sont calculées selon d de Cohen. Les η^2 ont été transformés selon la formule : $\text{Cohen's } d = 2 \cdot \sqrt{(\eta^2 / (1 - \eta^2))}$.

Note 3 R Max représente le stanine maximal pour les sous-tests en lecture.

la *Signification de mots* est également significative ($p < 0,001$) avantageant le groupe ADOR-IIR avec une grandeur d'effet de 0,46. Le sous-test de la *Compréhension orale* présente une différence significative ($p < 0,01$) favorisant le groupe ADOR-IIR, la grandeur d'effet étant de 0,29.

De plus, les différences les plus élevées entre les deux groupes sont observées aux sous-tests *Compréhension de phrases* ($d = 0,59$; $p < 0,01$) et *Compréhension de passages* ($d = 0,84$; $p < 0,001$). Ces deux épreuves mesurent un des objectifs ultimes de la lecture : la compréhension.⁷

En ce qui a trait au test arithmétique, les résultats ne sont pas significatifs.

Les hypothèses 1 et 2 sont confirmées puisque le groupe ADOR-IIR obtient un rendement supérieur au groupe témoin sur toutes les mesures concernées. Le groupe ADOR-IIR présente aux sous-tests de *Lecture de mots*, de la *Compréhension en lecture de phrases* et de la *Compréhension en lecture de passages* (hypothèse 1) un rendement supérieur au groupe témoin et ces résultats sont tous significatifs ($p < 0,05$ à $p < 0,001$). Les mesures des sous-tests de la *Compréhension orale* et de la *Signification de mots* indiquent tous une différence significative ($p < 0,01$ à $p < 0,001$) qui avantage le groupe ADOR-IIR sur le groupe témoin.

La 3^e hypothèse, qui permet de vérifier l'effet Hawthorne, est également confirmée puisque le résultat de cette comparaison est non significatif, ce qui signifie que les deux groupes présentent sensiblement le même rendement. L'effet Hawthorne ne se manifeste pas.

En ce qui concerne les mesures complémentaires utilisées, la seule ayant des corrélations avec le progrès des élèves en lecture est le questionnaire CES complété par ces derniers. Comme les résultats présentés par les chercheurs externes ne permettent pas de comparer la perception des élèves du groupe ADOR-IIR à celle du groupe témoin, nous considérons qu'il est non pertinent d'en présenter les résultats.

Discussion

La présente recherche démontre que les programmes ADOR et IIR combinés entraînent un rendement supérieur en lecture, tel que mesuré par les sous-tests de l'instrument standardisé GRADE. Toutes les différences observées sont statistiquement significatives. Afin de bien saisir l'ampleur des effets engendrés par les programmes ADOR et IIR sur l'apprentissage de la lecture des élèves en 1^{re} année, nous allons placer ces grandeurs d'effet (d de Cohen) dans une perspective de gains d'apprentissage.

Les tests standardisés de rendement scolaire employés aux États-Unis montrent qu'une année d'apprentissage en lecture pour les élèves de 1^{re} année équivaut en moyenne à $d = 0,97$ (Hill, Bloom, Black, et Lipsey, 2008). Par conséquent, si cela s'avère être le cas pour le test GRADE, dans son ensemble et pour ses sous-tests, nous pouvons alors calculer théoriquement le nombre de mois de gain d'apprentissage que sous-tendent les grandeurs d'effet rapportées dans le *Tableau 1* précédent. Par exemple, l'effet de 0,20 pour la *Lecture de mots* correspondrait à un gain d'environ deux (2) mois d'apprentissage sur cet aspect ($0,20/0,97 = 0,21$ d'une année scolaire de 10 mois, donc deux mois⁸). Sur la mesure de la *Signification de mots*, ADOR-IIR progresserait de près de cinq (5) mois de plus que le groupe témoins ($0,46/0,97 = 0,47$) et de trois (3) mois de plus d'apprentissage sur la mesure de la *Compréhension orale* ($0,29/0,97 = 0,30$). Là où le groupe ADOR-IIR obtient un rendement annuel hautement supérieur au groupe témoin est sur les mesures de la compréhension en lecture. Deux sous-tests de GRADE mesurent ce volet : la *Compréhension en lecture de phrases*, qui indiquerait un gain d'apprentissage de six (6) mois sur le groupe témoin ($0,59/0,97 = 0,61$) et la *Compréhension en lecture de passages*, qui présenterait un gain d'apprentissage assez exceptionnel de près de neuf (9) mois ($0,84/0,97 = 0,87$), donc presque qu'une année d'apprentissage supplémentaire. En d'autres mots, les élèves du groupe ADOR-IIR auraient effectué sur cette mesure en une année scolaire, presque deux ans d'apprentissage. Bien que ces conversions de grandeurs d'effets en mois d'apprentissage ne soient qu'hypothétiques⁹, elles tendent à indiquer que les effets des programmes ADOR-IIR sont importants, particulièrement en ce qui a trait à l'objectif ultime de l'enseigne-

ment de la lecture : la compréhension.

Dans la prochaine section, nous allons poser un regard critique sur la méthodologie de l'étude de Savage et Deault (2007).

Certains des éléments en usage dans l'élaboration de la Meilleure preuve (Best Evidence)

Robert Slavin et son équipe proposent depuis plusieurs années, dans le cadre du concept de la *Meilleure preuve*, de considérer certains éléments méthodologiques afin de déterminer et de sélectionner les meilleures recherches expérimentales pour réaliser une méta-analyse valide (Cheung et Slavin, 2016). La justification est fort simple : la méthodologie des recherches expérimentales est corrélée avec les effets obtenus (Cheung et Slavin, 2016). Nous allons appliquer certaines des indications de Slavin et de ses collègues à la recherche de Savage et Deault (2007) afin de jauger dans quelle mesure elle se rapproche du concept de la *Meilleure preuve*¹⁰.

L'emploi d'une mesure standardisée ou d'une mesure indépendante¹¹ est nettement préférable à l'usage d'une mesure maison

L'emploi de mesures standardisées ou d'une mesure indépendante plutôt que des mesures maison peut grandement influencer la grandeur d'effet. Slavin et Madden (2011) rapportent dans leur analyse de sept recherches en mathématique que la grandeur d'effet avec les mesures maison est de 0,45 et de -0,03 pour les mesures standardisées. Cheung et Slavin (2016) pour leur part observent un effet de 0,40 à partir de mesures maison et de 0,20 pour les mesures standardisées dans une méta-analyse de 645 recherches concernant l'apprentissage en lecture, mathématique, et sciences du préscolaire à la 12^e année.

L'étude à la base du présent texte (Savage et Deault, 2007) utilise une mesure indépendante du concepteur des programmes et des chercheurs de l'étude. Des sous-tests de l'instrument GRADE, qui a été standardisé avec un échantillon anglophone canadien, sont utilisés pour mesurer le rendement des élèves.

La taille de l'échantillon d'une étude a un impact important sur la grandeur d'effet observé

Slavin et Smith (2009) constatent que la grandeur d'effet varie selon le nombre de sujets de l'étude. Les études comptant moins de 51 sujets obtiennent une grandeur d'effet de 0,44 tandis que les études comptant de 51 à 100 sujets s'en sortent avec un effet de 0,29. Si l'échantillon comprend plus de 2000 sujets, l'effet n'est plus que de 0,09. Pellegrini (2017 ; voir Slavin, 2018) obtient une grandeur d'effet de 0,37 avec des recherches en pédagogie couvrant plusieurs matières comptant 60 sujets ou moins et un effet de 0,13 lorsque l'échantillon est supérieur à 250. Slavin et Smith (2009) pour leur part constate, dans une méta-analyse en mathématique au primaire et secondaire, qu'un échantillon composé de 50 sujets et moins peut obtenir une grandeur d'effet jusqu'à 3,67 supérieurs à un échantillon de 401 à 1000 sujets (un effet moyen de 0,44 pour les études de 50 sujets et moins et un effet moyen de 0,12 pour les études de 401 à 1000 sujets).

L'étude de Savage et Deault (2007) dont nous rapportons les résultats compte un échantillon de 480 élèves (204 élèves pour le groupe expérimental ADOR-IIR et 276 élèves pour le groupe témoin), ce qui réduit la possibilité que les résultats observés soient indûment gonflés par un faible nombre de sujets.

Le taux de mortalité des sujets ne doit pas être supérieur à 15 %

Neitzel, Lake, Pellegrini et Slavin (soumis) rappellent qu'une recherche ayant un taux de mortalité supérieur à 15 % ne devrait pas être retenue dans la réalisation d'une méta-analyse parce que cet élément pourrait influencer les résultats observés. Le What Works Clearinghouse (2020), qui propose également certains standards pour effectuer une méta-analyse, fait la même proposition.

À cet égard, l'étude de Savage et Deault (2007) indique que moins de 1 % des données sont manquantes, ce qui indique un taux de mortalité très faible.

L'expérimentation doit s'échelonner sur une durée de 12 semaines au minimum entre le pré-test et le post-test

Cheung et Slavin (2016) recommandent que les recherches aient une durée d'au moins douze (12) semaines afin qu'elles puissent refléter, entre autres, les effets de pratiques usuelles s'étendant sur une année scolaire. Les interventions les plus courtes (moins de dix heures d'enseignement ; voir par exemple Wanzek et ses collègues, 2017) semblent produire les effets les plus élevés dont on peut soupçonner l'artificialité.

La recherche de Savage et Deault (2007) mesure les effets d'une année scolaire complète, et ce, à deux reprises. Conséquemment, leur étude reflète l'effet d'une pratique usuelle annuelle.

L'expérimentation doit utiliser idéalement une assignation aléatoire aux groupes expérimental et témoin (méthode expérimentale) ou, à défaut, à des ajustements statistiques entre les groupes (méthode quasi expérimentale)

L'assignation aléatoire est une caractéristique importante de la méthode scientifique (Cheung et Slavin, 2016) qui est peu souvent appliquée en pédagogie. La recherche de Savage et Deault (2007) ne possède pas cette caractéristique, mais elle a appliqué les ajustements statistiques nécessaires.

En somme, sur plusieurs des éléments importants sous-jacents au concept de *Meilleure preuve* de Slavin, l'étude de Savage et Deault (2007) s'en tire très bien et serait sûrement retenue dans la réalisation d'une méta-analyse.

Limites de l'étude

Cette recherche possède certaines limites. Par exemple, il est impossible de déterminer l'effet respectif du programme ADOR et du programme IIR, puisqu'aucune mesure séparée des deux programmes n'a été prise. Cela n'invalide pas les résultats observés, mais limite la compréhension fine de la contribution de chacun des programmes aux résultats.

Même si l'étude de Savage et Deault (2007) n'en fait pas mention, la fidélité de l'application des programmes ADOR et IIR a été assurée par les entraîneurs formés à ces programmes¹², et ce, pour les deux années de l'expérimentation¹³. Par contre, la fidélité de l'application des programmes d'enseignement en classe ordinaire et en orthopédagogie du groupe témoin n'a pas bénéficié d'un suivi précis et aussi intense que les programmes du groupe expérimental. Bien que l'on puisse authentifier que les programmes ADOR et IIR ont été réellement appliqués, nous n'avons pas d'information sur le degré d'application des programmes dans le groupe témoin. Cette faiblesse est fréquente dans la recherche en pédagogie, et l'étude de Savage et Deault (2007) n'y échappe pas.

Conclusion

L'étude de Savage et Deault (2007) rencontre plusieurs des éléments constitutifs du concept de la *Meilleure preuve* de Slavin que ce chercheur et son équipe utilisent pour identifier les recherches les plus rigoureuses afin de réaliser des méta-analyses de haute qualité scientifique.

Deux programmes combinés en enseignement de la lecture en 1^{re} année, inspirés de l'*Enseignement explicite*, de l'*Enseignement direct* ainsi que d'autres éléments issus en grande partie de données probantes ont démontré une efficacité supérieure aux programmes du ministère de l'Éducation employés en classe ordinaire et en orthopédagogie dans une commission scolaire anglophone au Québec, et ce, sur toutes les mesures utilisées. Les effets les plus importants ont été obtenus aux sous-tests mesurant la compréhension en lecture. D'ailleurs, on peut observer un effet sur la compréhension de courts textes qui pourrait correspondre théoriquement à un gain équivalent à une année d'apprentissage supplémentaire. Comme d'autres chercheurs l'ont démontré (National Reading Panel, 2000 ; Bissonnette, Gauthier, Richard et Bouchard, 2010 ; Stockard, Wood, Coughlin et Rasplia Khoury, 2018), la présente recherche semble confirmer que les programmes axés sur l'*Enseignement explicite* et *direct* peuvent favoriser un meilleur apprentissage comparativement à d'autres programmes.

Références

- Bissonnette, S., Richard, M., Gauthier, C., Bouchard, C. (2010). Quelles sont les stratégies d'enseignement efficaces favorisant les apprentissages fondamentaux auprès des élèves en difficulté de niveau élémentaire ? Résultats d'une méga-analyse. *Revue de recherche appliquée sur l'apprentissage*, 3(1), 1-35.
- Boyer, C. (1993). *L'enseignement explicite de la compréhension en lecture : modèles d'activités d'enseignement*. Boucherville : Graficor.
- Boyer, C. (2010). *Le programme orthopédagogique DIR en lecture — L'Intervention intensive en lecture*. Montréal : Éditions de l'Apprentissage.
- Boyer, C., Bissonnette, S. (2021). Comment exercer une gestion rationnelle axée sur les résultats ? Exemple de la mesure de l'effet d'un programme orthopédagogique sur le rendement scolaire des élèves. *Enfance en difficulté*, 8, 97-128.
- Carnine, D., Silbert, J., Kameenui, E. J., Tarver, S. G. (1997). *Direct instruction reading*. Columbus, OH: Merrill.
- Cheung, A. C., Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283-292.
- Dion, É. (2012). Compte rendu de [Boyer, C. (2010). *Le programme orthopédagogique DIR en lecture*. Montréal, Qc : Éditions de l'apprentissage]. *Revue de psychoéducation*, 41(2), 283–286. <https://doi.org/10.7202/1061807ar>
- Forget, J. (2012). Les contributions de la pédagogie comportementale à l'éducation spécialisée. *Enfance en difficulté*, 1, 7–45. <https://doi.org/10.7202/1012122ar>
- Fraser, B. J. (1986). *Classroom environment*. London: Croon Helm.
- Hill, C. J., Bloom, H. S., Black, A. R., Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives*, 2(3), 172-177.
- Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. https://dspace.library.uu.nl/bitstream/handle/1874/16884/kirschner_02_cognitive_load_theory_implications_cognitive.pdf?sequence=1
- Lervåg, A., Hulme, C., Melby Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, 89(5), 1821-1838.
- Little, S. G., Akin-Little, K. A. (2003). Classroom management. Dans O'Donohue, W., Fisher, J., et Hayes, S. (Éditeurs), *Empirically supported techniques of cognitive behavioral therapy : A step-by-step guide for clinicians* (65-70). New York: Wiley.
- Ministère de l'Éducation du Québec (2001). Programme de formation de l'école québécoise. Éducation préscolaire, Enseignement primaire. Québec : Gouvernement du Québec.
- National Reading Panel (2000). Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction : Reports of the subgroups. National Institute of Child Health and Human Development, National Institutes of Health.
- Neitzel, A., Lake, C., Pellegrini, M., Slavin, R. (soumis). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly*.
- Oakhill, J., Cain K., Elbro, C. (2015). *Understanding and Teaching Reading Comprehension — A handbook*. Oxon: Routledge.
- Pressley, M., Wharton-Mc Donald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., et al. (2001). A study of effective first-grade literacy instruction. *Scien-*

tific Studies of Reading, 5,35-58.

Savage, R. (2005). An independent evaluation of the ADOR early reading program and other forms of effective teaching of reading in Level 1 Cycle 1. Montréal : Université McGill.

Savage, R., Deault, L. (2007). An independent evaluation of the ADOR early reading program and other forms of effective teaching of reading in Level 1 Cycle 1 2004-2007. Montréal : Université McGill

Seligman, M. E. P. (1995). *The optimist child. A Proven Program to Safeguard Children Against Depression and Build*. New York : Houghton Mifflin Harcourt.

Slavin, R. (2018). Small studies, big problems. Repéré à <https://robertslavinsblog.wordpress.com/2018/09/13/small-studies-big-problems/>

Slavin, R., Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, 4(4), 370-380.

Slavin, R., Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational evaluation and policy analysis*, 31(4), 500-506.

Stockard, J., Wood, T. W., Coughlin, C., Rasplia Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half-century of research. *Review of Educational Research*, 88(4), 479-507.

Stuart, M., Stainthorp, R. (2016). *Reading Development and Teaching*. Sage Publications.

Wanzek, J., Petscher, Y., Al Otaiba, S. A., Rivas, B. K., Jones, F. G., Kent, S. C., Schatschneider, C., Mehta, P. (2017). Effects of a year-long supplemental reading intervention for students with reading difficulties in fourth grade. *Journal of Educational Psychology*, 109(8), 1103–1119.

What Works Clearinghouse (2020). *Procedures handbook (Version 4.1)*. Institute of Education Sciences, US Department of Education.

Willingham, D. T. (2017). *The Reading Mind : A Cognitive Approach to Understanding How the Mind Reads*. San Francisco : Jossey-Bass—A Wiley Brand.

Notes

1. La direction générale de la Commission scolaire concernée a donné son autorisation à l'usage des résultats de ces rapports.
2. Le concepteur, Christian Boyer, est le premier auteur de ce texte.
3. Le *Group Reading Assessment and Diagnostic Evaluation* (GRADE) est un test diagnostique de lecture qui permet de déterminer le niveau atteint d'habiletés en lecture et en compréhension orale par les élèves de la maternelle à la 12^e année.
4. L'effet Hawthorne désigne l'effet que produit le seul fait de participer à une expérience, ce qui a tendance à entraîner une motivation plus grande chez les participants du groupe expérimental par rapport aux participants du groupe témoin, pouvant influencer positivement les perceptions et les résultats obtenus des sujets du groupe expérimental.
5. Le débit-exact est une mesure de lecture orale d'un texte jamais lu auparavant et correspond au nombre de mots bien lus sans aide au cours d'une minute.
6. Ce sont les hypothèses formulées par les auteurs de ce texte (Boyer, Bissonnette et Morneau-Guérin, 2022) et non celles de Savage et Deault (2007).
7. Les grandeurs d'effet du rapport réalisé en 2004-2005 (Savage 2005) sur ces mêmes mesures sont respectivement de $d = 0,59$ ($p < 0,001$) et $d = 0,87$ ($p < 0,001$).
8. Autrement dit, les élèves du groupe ADOR-IIR auraient progressé de deux mois d'apprentissage de plus que le groupe témoin.
9. Cette analyse vise à illustrer la portée des résultats obtenus.
10. Le concept de la *Meilleure preuve* de Slavin consiste à considérer certains éléments méthodologiques dans la sélection des recherches composant une méta-analyse ainsi que dans l'analyse de ces éléments comme variables modératrices des grandeurs d'effet.
11. Une mesure indépendante des concepteurs ou des chercheurs de l'étude est préférable à une mesure maison élaborée par les concepteurs et les chercheurs.
12. Les entraîneurs ont été formés et supervisés par Christian Boyer.
13. Les données étant perdues, selon les souvenirs de Christian Boyer, la fidélité de l'application des programmes variait entre 50 % et 98 % pour ADOR avec une moyenne de 79 % et, de 40 % à 98 % avec une moyenne de 71 % pour IIR.).