REVUE FRANÇAISE DE PÉDAGOGIE

Revue française de pédagogie
Recherches en éducation

201 | 2017
Recherche, politique et pratiques en éducation : services rendus et questions posées d'un univers à l'autre /2

Évaluation des premiers effets d'un enseignement fondé sur l'outil didactique *Narramus* à l'école maternelle

The initial effects of Narramus, *an innovative teaching tool in nursery schools*

Isabelle Roux-Baron, Sylvie Cèbe et Roland Goigoux

OpenEdition Journals

Édition électronique
URL : https://journals.openedition.org/rfp/7284
DOI : 10.4000/rfp.7284
ISSN : 2105-2913

Éditeur
ENS Éditions

Édition imprimée
Date de publication : 31 décembre 2017
Pagination : 83-104
ISBN : 979-10-362-0148-6
ISSN : 0556-7807

Référence électronique
Isabelle Roux-Baron, Sylvie Cèbe et Roland Goigoux, « Évaluation des premiers effets d'un enseignement fondé sur l'outil didactique *Narramus* à l'école maternelle », *Revue française de pédagogie* [En ligne], 201 | 2017, mis en ligne le 01 janvier 2023, consulté le 03 janvier 2023. URL : http://journals.openedition.org/rfp/7284 ; DOI : https://doi.org/10.4000/rfp.7284

Tous droits réservés

Évaluation des premiers effets d'un enseignement fondé sur l'outil didactique *Narramus* à l'école maternelle

Isabelle Roux-Baron
Sylvie Cèbe
Roland Goigoux

Cette étude repose sur l'évaluation des effets d'un outil didactique novateur, *Narramus*, proposé à des enseignants de maternelle exerçant en réseau d'éducation prioritaire. Basé sur des lectures d'albums de la littérature de jeunesse, cet outil vise le développement de compétences langagières narratives en réception et en production. 78 classes de grande section et 87 de moyenne section sont réparties en trois groupes. Le premier (G1) met en œuvre *Narramus* et bénéficie de l'accompagnement de conseillers pédagogiques de circonscription. Le deuxième (G2) utilise *Narramus* sans accompagnement. Les enseignants qui composent le dernier (G3) visent les mêmes compétences et font étudier les mêmes albums mais sans l'aide de l'outil. La compréhension du langage écrit, les compétences narratives et les acquisitions lexicales de 990 élèves ont été mesurées à six reprises au cours de l'année. Les résultats montrent que ceux qui ont bénéficié de *Narramus* (G1 et G2) obtiennent, en moyenne, de meilleurs résultats que leurs camarades du groupe 3 sur l'ensemble des dimensions évaluées et que les différences se creusent au fil de l'année. En revanche les différences entre G1 et G2 ne sont pas significatives : l'effet de l'accompagnement n'est pas avéré.

Mots-clés (TESE) : école maternelle, compréhension, lecture, lexique, innovation pédagogique, évaluation

Introduction

Pour que l'école maternelle française puisse jouer le rôle compensatoire des inégalités sociales que le pays attend d'elle, elle doit permettre aux jeunes élèves d'exercer, en contexte scolaire, les habiletés que les plus favorisés d'entre eux construisent également en famille et qui constituent le meilleur viatique pour leur

scolarité future : comprendre les récits écrits, savoir les raconter et acquérir du lexique (Bonnéry, Mamede, Soulé *et al.*, 2013 ; Hindman, Skibbe & Foster, 2014).

C'est dans ce but que l'équipe dirigée par Sylvie Cèbe a entrepris de concevoir un outil didactique innovant baptisé *Narramus* (Cèbe & Goigoux, 2017). Basés sur des lectures d'albums, les scénarios didactiques proposés sont plus intensifs, explicites et systématiques que les dispositifs en usage dans les classes et visent un objectif rarement assigné aux jeunes élèves : apprendre à raconter. Tous les apprentissages sont organisés autour de ce projet. C'est pour mieux raconter que les enfants doivent mémoriser le vocabulaire et de nouvelles tournures syntaxiques, retenir les idées principales, s'interroger sur les pensées des personnages et comprendre l'implicite des récits. Le but est que chacun d'entre eux soit capable de relater seul les histoires étudiées en classe.

Narramus a été conçu, testé et modifié pendant trois ans en collaboration avec une trentaine d'enseignants de maternelle. Nous en avons ensuite proposé une version définitive à de nombreux autres professeurs pour savoir si l'outil ainsi élaboré produisait les effets escomptés sur le développement des compétences langagières des élèves même lorsqu'il n'était pas accompagné par un dispositif spécifique de formation[1].

Contexte théorique et problématique

Premier enjeu : évaluer l'efficacité d'un outil

Viser des compétences fortement prédictives mais peu enseignées

Avant leur entrée à l'école maternelle, les enfants de milieux favorisés ont pris de l'avance sur le versant du développement du vocabulaire, de la syntaxe (Le Normand, Parisse & Cohen, 2008) et de la compréhension des textes écrits (Dickinson & Tabors, 2001). Sénéchal et LeFevre (2002) montrent que les compétences langagières acquises en moyenne section prédisent fortement les performances cinq ans plus tard. D'autres recherches font apparaître que les différences restent stables au fil de la scolarité (Adlof, Perfetti & Catts, 2011 ; Elwer, Gustafson, Byrne *et al.*, 2015 ; Kendeou, Van Den Broek, White *et al.*, 2009 ; Nation, Cocksey, Taylor *et al.*, 2010), voire se creusent (Petscher, Justice & Hogan, 2017).

En France, la recherche « Lire-Écrire » a établi que le meilleur prédicteur en lecture-compréhension en fin de CE1 est la performance des élèves au début du CP à une épreuve de compréhension de textes narratifs entendus, c'est-à-dire lus à haute voix par l'évaluateur (Goigoux, Cèbe & Pironom, 2016). Le score obtenu explique, à lui seul, un tiers de la variance des performances en fin de CE1, soit plus que le niveau de décodage atteint par les élèves à la fin du CP et bien plus que leur niveau en conscience phonologique au début du CP. Si l'on ajoute que les performances en fin d'école maternelle sont fortement corrélées aux caractéristiques sociales et qu'au cycle 2 l'impact de l'effet-maître est minime dans le domaine de la compréhension (Jarlégan, Piquée, Fontanieu *et al.*, 2016), on mesure l'importance de ce qui se joue à l'école maternelle. Celle-ci peut cependant jouer un rôle compensatoire comme l'atteste la méta-analyse de Swanson, Vaughn, Wanzek *et alii* (2011) portant sur les effets des activités de « lecture partagée » menées en contexte scolaire avec de jeunes enfants de milieux populaires.

Pris ensemble, ces résultats plaident en faveur d'un enseignement précoce des compétences requises pour bien comprendre, ce qui n'est malheureusement pas le cas à l'école maternelle française, comme l'ont constaté les inspecteurs généraux de l'Éducation nationale (IGEN-IGAENR, 2011). En France, comme ailleurs (Catts, 2018 ; Spencer, Weddle, Peterson *et al.*, 2017), il semble y avoir une contradiction entre ce qui est le plus valorisé et évalué à l'école – la compréhension – et ce qui y est enseigné. On ne peut donc que suivre les IGEN quand ils recommandent que tous les enseignants soient « plus rigoureusement préparés à distinguer ce sur quoi doit porter l'enseignement de la compréhension et comment ils peuvent le conduire » (IGEN-IGAENR, 2011, p. 131). C'est dans ce but que Cèbe et Goigoux ont conçu *Narramus* après avoir identifié, dans la littérature scientifique, les caractéristiques des dispositifs pédagogiques ayant produit des effets avérés et durables (*evidence based practices*).

1 Cette recherche, réalisée dans le cadre d'un partenariat entre le laboratoire ACTé (université Clermont-Auvergne) et l'Institut français de l'Éducation (ENS Lyon), a bénéficié de trois sources de financement : l'Institut Carnot de l'Éducation (Auvergne-Rhône-Alpes), la direction générale de l'enseignement scolaire (DGESCO-MEN) et le rectorat de l'académie de Clermont-Ferrand. Nous remercions vivement l'ensemble des enseignants et des cadres de l'Éducation nationale qui ont collaboré à cette recherche.

Les caractéristiques des pratiques d'enseignement efficaces

– Apprendre à comprendre, ensemble, un texte, en profondeur

Plusieurs auteurs (Blewitt, Rump, Shealy *et al.*, 2009 ; Cabell, Justice, McGinty *et al.*, 2015 ; Wasik & Hindman, 2011) ont montré que les enseignants qui font discuter le sens des histoires aident leurs élèves à mieux les comprendre et à mieux les interpréter, mais à deux conditions : 1° qu'ils les sollicitent dans de multiples registres cognitifs : reformulation, description, explication, résumé, discussion, argumentation, mise en lien avec l'expérience personnelle… (Dickinson & Smith, 1994) ; 2° qu'ils leur fassent réaliser des activités présentant de véritables défis cognitifs (Massey, 2004 ; Mascareño, Deunk, Snow *et al.*, 2017). C'est pourquoi *Narramus* propose une approche intégrative de l'enseignement de la compréhension et de la narration en prenant appui sur des albums. Il se distingue en cela des approches modulaires qui travaillent des compétences isolées sur des supports morcelés (Bianco, Coda & Gourgue, 2006 ; Gourgue, Bianco & Coda, 2013).

– Apprendre à raconter l'histoire étudiée

L'activité de rappel, guidée par l'enseignant dans *Narramus*, offre un avantage avéré qui vient de sa distance par rapport aux événements alors que le langage des jeunes enfants est profondément ancré dans le contexte immédiat. Grâce à cette activité, les élèves apprennent à organiser les informations et à planifier leur discours, à soigner leur mise en mots (lexique et syntaxe) et à assurer la cohérence textuelle : désignation des personnages, rappel des principaux événements, reprises anaphoriques, etc. (Foorman, Beyler, Borradaile *et al.*, 2016).

– Un enseignement et une révision systématiques du vocabulaire du texte

Snell, Hindman et Wasik (2015) dressent la liste des pratiques qui ont fait la preuve de leur efficacité pour enseigner le vocabulaire à l'école maternelle. Elles concluent que si l'on veut que les élèves le mémorisent durablement, il faut expliquer le sens des mots et des expressions mais aussi le leur faire réviser, vérifier les acquis, lire les albums maintes fois, les engager dans des activités de rappel de récit et intégrer les expressions et les mots nouveaux dans d'autres activités scolaires. Ces principes sont opérationnalisés dans *Narramus*.

– Apprendre aux jeunes enfants à produire des inférences

Il ne suffit pas que les élèves connaissent le sens de tous les mots et de toutes les expressions d'un texte pour le comprendre en profondeur, il faut aussi qu'ils produisent des inférences. Cela suppose qu'ils apprennent à mettre en relation les informations présentes dans le texte de manière éparse et qu'ils les relient avec leur base de connaissances (Cook, Limber & O'Brien, 2001). Sur ce point, plusieurs études prouvent que la « théorie de l'esprit » – notamment la capacité à inférer les pensées d'autrui à partir de ses comportements / ou la capacité à inférer des relations entre les pensées d'un personnage et ses comportements – influe sur la qualité de la compréhension des textes narratifs (Atkinson, Slade, Powell *et al.*, 2017 ; Dore, Amendum, Golinkoff *et al.*, 2018 ; Graesser, Singer & Trabasso, 1994 ; Kim, 2016). *Narramus* met en pratique ces principes dont l'efficacité a été démontrée en proposant un ensemble de tâches scolaires centrées sur les états mentaux des différents personnages et l'explication des informations implicites.

– Mettre en scène les histoires étudiées

On a aussi montré que la mise en scène de l'histoire par les élèves (*via* des activités de théâtre ou l'usage d'une maquette et de marionnettes) exerce un effet sur la qualité de la compréhension, le développement des compétences narratives et l'acquisition du vocabulaire (Fischer & Zwaan, 2008 ; Glenberg, Gutierrez, Levin *et al.*, 2004 ; McLeod, Hardy & Kaiser, 2017 ; Newman, Dickinson, Hirsch-Pasek *et al.*, 2015 ; Pesco & Gagné, 2017). Tout ce qui anime dans l'espace ce qui n'est que mots sur une page constitue une aide puissante pour les jeunes élèves, à condition qu'ils ne soient pas livrés à eux-mêmes. Ces résultats s'expliquent en partie par la théorie de la cognition incarnée (Bara & Tricot, 2017). La mise en scène des histoires est travaillée de deux manières dans *Narramus* : en classe avec des figurines et une maquette, en salle de motricité avec des masques et du jeu corporel.

– Le recours aux supports multimédia numériques

Contrairement aux idées reçues, le recours aux vidéos, à la musique et aux sons, loin de distraire les élèves, permet un engagement plus durable que les livres traditionnels (Moody, Justice & Cabell, 2010). Ces supports offrent un ensemble d'outils qui, quand ils sont combinés avec le texte oral, améliorent la compréhension du texte, l'interprétation des événements et la production d'inférences (Myrtil, Justice, Pelfrey *et al.*,

2018 ; Zhou & Yadav, 2017). C'est pourquoi *Narramus* repose sur un usage important du numérique.

Description de l'outil *Narramus*

Pour chaque scénario, *Narramus* comprend trois éléments : un guide pédagogique qui décrit la manière de mener les activités et les justifie théoriquement (70 pages) ; un CD contenant tous les supports utiles pour la mise en œuvre ; l'album étudié. Chaque scénario comprend 8 modules répartis sur quatre semaines d'enseignement au cours desquelles les élèves vont découvrir ensemble l'histoire, épisode par épisode.

– *Avant la découverte de l'épisode*

À l'exception du 1er, chaque module commence par la révision du vocabulaire appris précédemment et l'enseignement des expressions et des mots nouveaux contenus dans le nouvel épisode. Puis un seul élève rappelle des épisodes précédents (les autres valident, corrigent et complètent à la fin).

– *Pendant la découverte du nouvel épisode*

L'enseignant commence par expliquer l'objectif du module. Il mène ensuite de nombreuses activités qui dépendent de la nature du texte étudié, de sa structure et de la place du module dans le scénario. C'est ainsi que les élèves vont avoir à s'interroger sur les états mentaux successifs des personnages et débattre de leur choix, expliciter l'implicite, s'interroger sur les suites possibles, discuter la morale de l'histoire…

– *Après la découverte du nouvel épisode*

Après qu'un seul élève a raconté l'épisode, l'enseignant donne un masque à quelques élèves chargés de jouer la scène devant les autres. Puis, un autre raconte l'histoire depuis le début (avec l'album ou la maquette), une aide pour lier les informations entre elles et assurer la cohérence textuelle. La classe part ensuite en salle de jeux pour mimer les mots de vocabulaire dictés par l'enseignante ou jouer l'histoire en équipes, chaque élève étant équipé du masque du personnage qu'il représente.

– *Au cours de la journée*

Les élèves jouent aux cartes de vocabulaire avec de multiples consignes (tris, catégorisation, dénomination rapide, loto, devinettes…), s'entraînent à raconter avec la maquette et les figurines ou l'album, écoutent l'histoire audio lue ou racontée.

Formulation de notre première hypothèse de recherche

L'objet de cet article est de vérifier que ces caractéristiques, lorsqu'elles sont mises en œuvre en respectant les contraintes contextuelles et la culture professionnelle des enseignants de l'école française, produisent des effets comparables à ceux observés dans les travaux que nous venons d'évoquer. Ces effets devraient toucher la qualité des rappels des histoires d'une part et la mémorisation du lexique enseigné d'autre part. La première hypothèse testée stipule donc que les élèves bénéficiant de *Narramus* (G1 et G2) progresseront plus que les autres (G3) sur ces deux dimensions évaluées.

Deuxième enjeu : évaluer le rôle de l'accompagnement dans l'innovation pédagogique

Un rôle controversé

Pour introduire des dispositifs innovants en classe et en garantir le succès, les chercheurs assurent généralement le suivi du travail des enseignants engagés dans l'expérimentation, jugeant qu'un tel soutien est nécessaire pour que la mise en œuvre des préconisations soit fidèle (Charlier, Biémar & Jorro, 2012 ; Rowan & Miller, 2007 ; Bélanger, Bowen, Cartier *et al*., 2012 ; Penneman, De Croix, Dellisse *et al*., 2016). La méta-analyse de Yoshikawa, Weiland, Brooks-Gunn *et alii* (2013) l'atteste : les innovations qui produisent les meilleurs résultats au préscolaire se caractérisent par la qualité des contenus et par le soutien continu offert aux enseignants. Pour Letor, Enthoven et Dupriez (2016), l'accompagnement est obligatoire « au risque de voir les conversations [entre enseignants] dévier vers des considérations périphériques ou de voir l'outil abandonné rapidement » (p. 52). Quand les expérimentations échouent, les chercheurs incriminent la formation et l'accompagnement insuffisant (Gather Thurler, 2000 ; Gentaz, Sprenger-Charolles, Colé *et al*., 2013 ; Zorman, Bressoux, Bianco *et al*., 2015) mais s'interrogent rarement sur la qualité des outils proposés.

Notre analyse est différente. Pour nous, l'accompagnement n'est indispensable que si les outils sont élaborés sans prendre en compte leurs premiers destinataires : les enseignants (Goigoux, 2017). Trop souvent, les dispositifs expérimentaux sont conçus en laboratoire à partir des connaissances scientifiques disponibles sur les savoirs à enseigner et les apprentissages des élèves mais sans tenir compte des contextes

scolaires. Rares sont les chercheurs qui étudient aussi les pratiques ordinaires des professeurs et identifient leur zone proximale de développement professionnel[2] pour anticiper une possible appropriation de l'outil. De plus, les innovations didactiques portent sur des points précis, mais négligent la cohérence d'ensemble de la pédagogie et les contraintes qui pèsent sur l'exercice du métier telles que l'organisation de la classe, la gestion d'un collectif hétérogène, le curriculum à respecter, les autres outils disponibles… (Fishman, Penuel, Allen *et al.*, 2013). Bref, faute de disposer d'un modèle de l'utilisateur (Rabardel & Pastré, 2005), l'ergonomie de l'outil est insatisfaisante. L'expérimentation ne réussit qu'au prix d'un encadrement externe de l'activité des enseignants, souvent réalisée par les chercheurs ou leurs équipes et elle est par conséquent très difficile, voire impossible, à généraliser (Gurgand, 2018 ; Klingner, Boardman & McMaster, 2013). Le « passage à l'échelle » (Odom, 2009) dans des conditions ordinaires d'enseignement échoue.

Notre analyse, partagée à l'étranger par de nombreux chercheurs en sciences de l'éducation (Snow, 2015) mais rarement concrétisée en France (voir les volumes 11-2 et 11-3 de la revue *Éducation et didactique*, 2017), n'est pas nouvelle. Dès 1995, Tyack et Cuban avaient attribué les échecs des innovations à l'insuffisante prise en compte des contextes d'enseignement et des habitus professionnels des enseignants (leur *Grammar of schooling*). Ils avaient montré que la réussite d'une innovation dépendait à la fois de sa compatibilité avec les pratiques d'enseignement habituelles (ce qui suppose de les connaître) et de l'efficience de l'intervention, c'est-à-dire du rapport entre son coût pour les professeurs (réorganisation cognitive, surcharge de travail, investissement émotionnel…) et ses bénéfices (satisfaction due aux progrès des élèves ou à l'accroissement du sentiment de compétence du professeur).

Ce sont ces conditions que Cèbe et Goigoux se sont efforcés de satisfaire. Ils ont fait tester pendant trois ans le prototype de leur outil par des enseignants « co-concepteurs » et ils l'ont modifié à plusieurs reprises avant l'expérimentation pour satisfaire aux principaux critères de l'ergonomie de conception (Béguin & Darses, 1998) : l'utilité, l'utilisabilité et l'acceptabilité de l'artéfact par ses futurs utilisateurs (Tricot, Plégat-Soutjis, Camps *et al.*, 2003). La démarche de conception utilisée pour élaborer *Narramus* (Cèbe & Goigoux, 2018) a donc été itérative et participative. Qualifiée de « conception continuée dans l'usage », elle s'inscrit dans la tradition de l'ergonomie de langue française (Béguin & Cerf, 2004) et diffère peu de celles préconisées dans d'autres pays au sein des courants *Design-based research* (Wang & Hannafin, 2005), *Design-Based Implementation Research* (Cobb, Confrey, Disessa *et al.*, 2003) ou, plus largement, *Educational Design* (Collins, 1992 ; McKenney & Reeves, 2014).

Formulation de notre seconde hypothèse de recherche

Puisque la conception de *Narramus* s'efforce de respecter les conditions énoncées, nous pensons qu'il n'aura pas besoin d'être accompagné pour produire les effets escomptés. Notre seconde hypothèse peut donc être formulée ainsi : conçu sur des bases scientifiques attestées et en partenariat avec les enseignants (au terme d'un processus de conception continuée dans l'usage), *Narramus* produira des effets similaires avec ou sans accompagnement.

Pour tester cette seconde hypothèse, nous avons construit un plan expérimental dans lequel certains enseignants étaient accompagnés et d'autres pas. Refusant que l'accompagnement soit assuré par les chercheurs pour ne pas condamner l'outil à n'être utilisable que dans des conditions exceptionnelles et pour anticiper sa diffusion à grande échelle, nous avons formé les conseillers pédagogiques et les inspecteurs de l'Éducation nationale dont le métier est précisément de conseiller les enseignants. Nous les avons initiés à la théorie et à la pratique de *Narramus*, à charge pour eux d'accompagner les équipes d'école dans le cadre de leurs missions statutaires. Avant de mesurer l'impact de cet accompagnement sur les apprentissages des élèves, nous avons évalué l'effet de ce suivi sur les pratiques des maîtres en considérant deux critères : le temps alloué à l'enseignement et la fidélité aux prescriptions de *Narramus*. Nous voulions savoir si l'accompagnement incitait les enseignants à être plus assidus et plus fidèles que les enseignants non accompagnés.

Méthodologie

Le dispositif expérimental

Les enseignants (N = 165)

En septembre 2016, à notre initiative, trente équipes de circonscriptions de onze académies (Aix-Marseille,

2 Définie comme l'intervalle entre ce que les enseignants réalisent ordinairement et ce qu'ils pourraient réaliser au cours d'une genèse instrumentale (Béguin, 2005).

Bordeaux, Clermont-Ferrand, Corse, Grenoble, Guadeloupe, Lille, Lyon, Mayotte[3], Strasbourg, Versailles) ont sollicité des écoles classées en éducation prioritaire (REP+ ou REP). Pour favoriser le travail collectif, trois enseignants au moins par école, un par niveau (PS, MS et GS), devaient accepter le contrat expérimental et le respecter jusqu'en juin 2017.

Dans cet article, nous étudions 165 classes de moyenne et grande sections de 79 écoles. 88,8 % des 165 professeurs sont des femmes. En moyenne, leur ancienneté professionnelle est de 15,5 ans (écart-type = 9,4) et leur expérience en école maternelle de 10,6 ans (écart-type = 8,0).

Nous avons affecté les écoles de manière aléatoire dans trois groupes expérimentaux puis avons pris soin de vérifier que ceux-ci étaient équivalents pour le critère de l'ancienneté en maternelle : 10,9 ans pour le groupe 1 (écart-type = 7,8), 10,6 ans pour le groupe 2 (écart-type = 8,1) et 10,1 ans pour le groupe 3 (écart-type = 8,4). Les différences ne sont pas significatives.

Les enseignants du groupe 1 ont reçu les trois albums et les scénarios de *Narramus* (un chaque trimestre), ils les ont mis en œuvre et ont bénéficié d'un accompagnement assuré par leur équipe de circonscription. Ceux du groupe 2 ont reçu le même outillage didactique et l'ont mis en œuvre sans accompagnement ni formation d'aucune sorte. Ceux du groupe 3 ont étudié les trois mêmes albums mais en agissant à leur guise. Ils disposaient eux aussi de quatre semaines de travail entre le pré-test et le post-test et, comme les autres, étaient informés de la nature des évaluations auxquelles nous soumettions leurs élèves. Tous savaient, en effet, que leurs élèves seraient interrogés sur le lexique du récit et devraient le raconter seuls et sans support. Ils ont accepté d'être privés de l'outil *Narramus* (dont ils ignoraient tout) en échange de la promesse d'en bénéficier l'année suivante.

Le groupe 3 sert de groupe contrôle dans l'expérimentation, mais ne peut pas être considéré comme un groupe témoin représentatif des pratiques habituelles des enseignants de maternelle. En effet, le fait d'être enrôlé dans la recherche, de savoir que les élèves vont être évalués, de connaître la nature des épreuves, et de disposer de quatre semaines par album a sans doute influencé les enseignants du groupe 3. Dans les carnets de bord qu'ils ont renseignés individuellement, ils ont majoritairement déclaré[4] avoir accordé beaucoup plus de temps à l'étude des albums que d'ordinaire. Les utilisateurs de *Narramus* (G1 et G2) sont donc comparés à des enseignants plus mobilisés que des enseignants lambda, ce qui rend la comparaison plus exigeante et l'hypothèse 1 plus difficile à valider.

L'accompagnement du groupe 1 a duré 15 heures en moyenne et il a été réalisé par les conseillers pédagogiques et des inspecteurs dans le cadre des horaires dévolus à la formation continue en éducation prioritaire. Ces accompagnateurs ont bénéficié de quatre journées de formation à l'Institut français de l'Éducation (ENS Lyon) animées par les auteurs de *Narramus*[5]. Nous leur avons demandé de renseigner un journal de bord (au moins trois fois dans l'année) décrivant les objectifs, la durée, les modalités et les contenus de l'accompagnement. Le dépouillement de ces documents révèle que, si les accompagnements ont revêtu des formes différentes, ils ont toujours été organisés autour de trois axes : une aide organisationnelle et logistique à la mise en œuvre dans les classes, des apports théoriques sur les principes qui sous-tendent l'outil et des échanges de pratiques avec un retour des formateurs-évaluateurs sur les résultats des élèves.

Les élèves (N = 990)

Étant donné le nombre de classes engagées dans la recherche et la nature des évaluations, nous ne pouvions tester individuellement tous les élèves. Nous devions donc réduire la taille de notre échantillon tout en conservant la diversité des contextes pédagogiques. Dans ce but, au début de l'année, nous avons fait passer une épreuve de compréhension syntaxico-sémantique (É.Co.S.Se. ; Lecocq, 1996) aux 3 300 élèves nés entre mars et septembre et nous les avons rangés en trois terciles. Nous avons ensuite tiré au sort deux élèves de chaque tercile et avons demandé à leur enseignant de vérifier que le résultat au test était cohérent avec ce qu'il savait du niveau de l'enfant (seuls 3 élèves ont été exclus au vu de la discordance entre le jugement de l'enseignant et le résultat au test). En résumé, nous avons choisi, comme il est d'usage dans ce type d'étude (Neuman & Kaeffer, 2018), de ne

3 D'importants mouvements sociaux ayant perturbé le déroulement du travail scolaire à Mayotte en 2016-2017, nous avons exclu de notre analyse les élèves et les classes de cette île.

4 Dans un autre volet de la recherche, nous avons recueilli les informations touchant les pratiques effectives et déclarées des enseignants des trois groupes. Elles feront l'objet d'une publication ultérieure.

5 Nous remercions Patrick Picard et le centre Alain Savary pour leur aide précieuse dans l'organisation et l'animation de ces journées de formation.

Tableau 1. Nombre de classes et d'élèves dans chaque groupe en GS et MS

	Nombre de classes			Nombre d'élèves testés		
Groupe	GS	MS	Total	GS	MS	Total
G1	35	40	75	210	240	450
G2	25	28	53	150	168	318
G3	18	19	37	108	114	222
Total	78	87	165	468	522	990

Tableau 2. Répartition des évaluations au cours de l'année scolaire

Période	Octobre 2016	Trimestre 1 Nov-déc. 2016	Trimestre 2 Fév-mars 2017	Trimestre 3 Avril-mai 2017
Tests	É.Co.S.Se.	Pré-test et post-test album 1 : rappel, lexique	Pré-test et post-test album 2 : rappel, lexique	Pré-test et post-test album 3 : rappel, lexique

retenir que six élèves par classe en sélectionnant des enfants du même âge (plus ou moins trois mois), à parité garçons-filles, deux pour chaque niveau : faible, intermédiaire et fort.

On trouvera, dans le tableau 1, la répartition des élèves selon le niveau scolaire et le groupe d'appartenance.

L'échantillon final est composé de 990 élèves (49,6 % de filles et 50,4 % de garçons) ; 522 sont issus de 87 classes de MS et 468 proviennent de 78 classes de GS.

Les épreuves d'évaluation

Épreuve de sélection des élèves : compréhension de phrases entendues

L'épreuve É.Co.S.Se.[6] vise à évaluer la compréhension de phrases courtes, simples sur le plan lexical, mais dont la complexité syntaxique va croissante. L'évaluateur lit la phrase et invite l'enfant à montrer, parmi les quatre images qu'il a sous les yeux, celle qui correspond le mieux à l'énoncé entendu.

Épreuves d'évaluation des apprentissages

Nous avons conçu deux épreuves originales – rappel de récit et lexique – à partir des albums étudiés, qui ont été proposées, à l'identique, avant enseignement (pré-test) et après (post-test). Les élèves ont donc été testés six fois à propos de trois albums différents comme l'indique le tableau 2. Les passations étaient individuelles et prenaient environ 15 minutes. Les rappels de récit ont tous été enregistrés sur support numérique.

– *Rappel de l'histoire*

Pour le pré-test, l'enseignant raconte l'histoire sans montrer les illustrations (jour 1) puis la lit en montrant et en commentant les images (jour 2). Les évaluateurs sollicitent un rappel juste après. En post-test, ce dernier est réalisé après que les élèves ont bénéficié de quatre semaines d'enseignement.

Chaque rappel est codé deux fois comme il est de rigueur en psycholinguistique (Hughes, McGillivray & Schmidek, 1997 ; Justice, Bowles, Kaderavek *et al.*, 2006). Le premier codage porte sur la microstructure et permet de savoir combien de propositions sémantiques du récit sont relatées par l'élève (variable « mi »). Le second porte sur la macrostructure, soit le nombre d'idées essentielles rappelées (variable « M »).

– *Lexique*

Cette épreuve mesure les connaissances initiales dont disposent les élèves sur le lexique de l'album étudié (pré-test) et les effets de l'enseignement sur son apprentissage (post-test). L'évaluation du lexique en réception (choisir la bonne image parmi six) et en production (dénomination d'images, explication

6 Version 10 planches pour le *Test d'évaluation Compréhension MS* et celui de *GS/CP* réédités et diffusés en ligne par La Cigale (<https://www.editions-cigale.com/424224cc>).

Tableau 3. Données descriptives pour l'accord inter-juges en rappel de récit

M_T5 ρ = (IdC 95 %)	mi_T5 ρ = (IdC 95 %)	M_T6 ρ = (IdC 95 %)	mi_T6 ρ = (IdC 95 %)
0,93 (0,91-0,95)	*0,93 (0,89-0,95)*	*0,88 (0,84-0,91)*	*0,98 (0,97-0,98)*

Légende : IdC désigne l'intervalle de confiance du coefficient de corrélation intra classe ρ.

Tableau 4. Données descriptives pour la variable *Fidélité*

	Fidélité (en %)					
	GS			**MS**		
	G1	**G2**	**Test**	**G1**	**G2**	**Test**
	Effectif Moyenne (Écart-type)		Test de Student p	Effectif Moyenne (Écart-type)		Test de Student p
Album 1	N = 34 86 (15)	N = 22 86 (15)	0,884 NS	N = 37 85 (9)	N = 25 82 (11)	0,286 NS
Album 2	N = 31 88 (8)	N = 22 86 (14)	0,451 NS	N = 34 83 (8)	N = 23 81 (14)	0,568 NS
Album 3	N = 28 89 (8)	N = 23 88 (8)	0,697 NS	N = 33 87 (7)	N = 23 86 (9)	0,704 NS

Légende : les effectifs (N) sont des nombres de classes ; NS exprime la non-significativité du test de comparaison de moyennes.

verbale et mime) permet de construire un score unique (variable « el ») basé sur 25 items.

Accord inter-juges

La passation et le codage des épreuves ont été réalisés par les formateurs-évaluateurs préalablement formés par les chercheurs. Les enregistrements de récits enfantins et leurs cotations ont été déposés sur une plate-forme de l'ENS Lyon[7]. Afin de vérifier la fiabilité des codages, nous avons fait recoder de manière aléatoire 278 enregistrements de récits soit 5,3 % de l'ensemble. L'accord inter-juges est compris entre 0,88 et 0,98 (voir tableau 3), ce qui est excellent.

Variables de contrôle des pratiques pédagogiques

Qualité de la mise en œuvre (implémentation) : la fidélité

Pour évaluer l'efficacité d'un outil, il faut s'assurer qu'il a bien été utilisé comme prévu, autrement dit vérifier que sa mise en œuvre est satisfaisante. Le critère généralement retenu pour en juger est la proportion d'activités préconisées réellement effectuées. Les chercheurs s'accordent sur un taux minimal de 60 % et montrent que l'effet mesuré est d'autant plus fiable que l'implémentation est fidèle (Durlak & DuPre, 2008). Nous avons donc demandé aux enseignants de G1 et G2 de renseigner quotidiennement un carnet de bord décrivant leur mise en œuvre et d'estimer eux-mêmes, en pourcentage, leur degré de fidélité aux préconisations de *Narramus* (réalisation des tâches prescrites). Nous avons ensuite calculé une moyenne pour l'ensemble des huit modules de chaque scénario ; cette

7 Nous remercions vivement Yvonnick Fesselier (ENS Lyon) qui a développé et fait vivre cette plate-forme ainsi que Valérie Fontanieu (ENS Lyon) et Julie Pironom (laboratoire ACTé) qui nous ont apporté une aide précieuse pour valider et interpréter nos traitements statistiques.

Tableau 5. Données descriptives pour la variable *Temps*

	Temps											
	GS						**MS**					
	G1	**G2**	**G3**	**Tests post Anova**			**G1**	**G2**	**G3**	**Tests post Anova**		
	Effectif Moyenne (en minutes) (Écart-type)			G1 vs G3	G2 vs G3	G1 vs G2	Effectif Moyenne (en minutes) (Écart-type)			G1 vs G3	G2 vs G3	G1 vs G2
Album 1	N = 34 391 (163)	N = 22 347 (174)	N = 16 206 (121)	0,001 **	0,026 *	0,922 NS	N = 37 361 (134)	N = 24 300 (77)	N = 17 218 (154)	0,001 **	0,122 NS	0,195 NS
Album 2	N = 30 522 (199)	N = 22 504 (144)	N = 16 284 (151)	0,000 ***	0,001 **	1,000 NS	N = 34 469 (144)	N = 24 396 113)	N = 17 222 (118)	0,000 ***	0,000 ***	0,113 NS
Album 3	N = 25 441 (179)	N = 22 421 (176)	N = 16 171 (84)	0,000 ***	0,000 ***	1,000 NS	N = 31 407 (136)	N = 21 347 (103)	N = 17 201 (85)	0,000 ***	0,001 **	0.219 NS
Moyenne par album	7 h 31	7 h 04	3 h 40				6 h 52	5 h 48	3 h 33			

Légende : les effectifs (N) sont des nombres de classes ; NS exprime la non-significativité du test de comparaison de moyennes ; * p < 0,05 ; ** p < 0,01 ; *** p < 0,001.

moyenne constitue notre variable *Fidélité* (voir tableau 4).

Pour chaque album, les enseignants des deux groupes expérimentaux ont été fidèles au protocole expérimental : l'indice calculé est supérieur à 86 % en GS et à 81 % en MS, bien au-dessus du seuil minimal exigé dans la littérature scientifique. Dans la suite de cet article, nous considérerons donc que la fidélité déclarée est suffisante pour juger que le protocole expérimental a été respecté par les deux groupes qui ne présentent aucune différence significative.

Contrôle du temps alloué à l'enseignement

Nous nous sommes donné une seconde exigence méthodologique, rarement retenue dans les recherches francophones : contrôler les durées effectivement consacrées, pour chaque album, à l'enseignement. Trop d'études concluent à la supériorité d'un dispositif A sur un dispositif B sans contrôler le facteur temps (Tricot, 2016) : on attribue alors à un outil des vertus qui ne sont que le fruit du temps consacré à l'enseignement de compétences habituellement peu travaillées. De nombreuses recherches expérimentales sont réalisées dans des conditions si extraordinaires (non-respect des programmes officiels, hypertrophie du temps dédié à l'entraînement d'un petit nombre d'habiletés, sur-encadrement pédagogique…) que leurs résultats n'ont guère de validité écologique et que les tentatives de généralisation échouent.

Pour éviter cet écueil, nous avons demandé aux enseignants des trois groupes d'indiquer, dans leur carnet de bord, le temps consacré à chaque module. L'addition de ces données nous renseigne sur le temps d'enseignement dédié à chaque album lors des quatre semaines qui séparent pré- et post-tests. Cette somme, exprimée en minutes, constitue la variable *Temps* (voir tableau 5).

Les différences de durée entre G1 et G2 ne sont pas significatives, ni en MS, ni en GS. On peut donc conclure que l'accompagnement n'a pas d'effet sur le temps alloué à l'enseignement.

En revanche, les temps d'enseignement déclarés par les enseignants sont significativement différents (presque du simple au double) entre le groupe 3 d'une part et les groupes 1 et 2 d'autre part. Comme on s'y attendait, l'outillage provoque un accroissement du temps consacré à l'enseignement de la compréhension et de la narration. Il sera donc nécessaire d'introduire la variable *Temps* dans les analyses ultérieures pour tester l'hypothèse 1 et pondérer ainsi les effets constatés.

Tableau 6. Résultats initiaux des élèves de GS et de MS (pré-test T1)

Épreuve	Tous les élèves		G1	G2	G3	Test
	Min-max	Effectif Moyenne (Écart-type)	Effectif Moyenne (Écart-type)			p (Anova)
Lexique GS	1-23	N = 449 14,0 (4,6)	N = 207 13,9 (4,9)	N = 139 14,2 (4,4)	N = 103 13,8 (4,4)	0,771 NS
Macro GS	0-19	N = 446 8,5 (5,4)	N = 202 8,4 (5,4)	N = 139 8,9 (5,6)	N = 105 8,0 (5,1)	0,408 NS
Micro GS	0-49	N = 446 13,2 (9,3)	N = 202 14,02 (10,2)	N = 139 13,0 (8,5)	N = 105 11,8 (8,1)	0,131 NS
Lexique MS	0-22	N = 490 10,7 (4,6)	N = 229 10,5 (4,5)	N = 156 10,6 (4,5)	N = 105 11,4 (5,0)	0,211 NS
Macro MS	0-18	N = 472 4,5 (4,4)	N = 222 4,2 (4,4)	N = 146 4,7 (4,4)	N = 104 4,8 (4,4)	0,416 NS
Micro MS	0-33	N = 472 6,7 (6,5)	N = 222 6,5 (6,7)	N = 146 6,9 (6,5)	N = 104 6,8 (5,8)	0,840 NS

Légende : les effectifs (N) sont des nombres d'élèves ; NS exprime la non-significativité.

Équivalence initiale des trois groupes

Nous avons d'abord vérifié que le niveau initial des élèves des trois groupes était équivalent. Pour cela, nous avons comparé les résultats au premier pré-test en lexique et en rappel de récit (macro et micro-propositions ; voir tableau 6).

En GS comme en MS, les groupes ne se distinguent ni en lexique, ni en rappel (macrostructure et microstructure). Nous pouvons donc considérer qu'ils sont équivalents sur les deux dimensions testées, ce que confirme le second calcul, réalisé à partir du test É.Co.S.Se. (voir tableau 7).

Dans la mesure où les données ne font apparaître aucune différence significative entre les groupes, en GS comme en MS, et ce pour les trois épreuves, nous pouvons tester notre première hypothèse.

Résultats

La partie suivante présente les résultats que nous avons recueillis tout au long de l'année pour tester nos deux hypothèses.

Test de l'hypothèse 1 : G1 et G2 *versus* G3

Notre première hypothèse porte sur les effets différenciateurs de l'usage de *Narramus*. Nous supposons que les groupes qui en ont bénéficié (G1 et G2) ont plus progressé que le troisième dont les enseignants ont travaillé sans l'outil (G3).

En grande section

En grande section, l'hypothèse est validée au vu des résultats exposés dans le tableau 8.

Quel que soit le groupe, tous les élèves progressent entre les pré- et les post-tests, mais, à chaque post-test (T2, T4 et T6), les résultats du G1 et G2 sont significativement supérieurs à ceux du G3 dans les deux épreuves et pour tous les scores (el, M et mi). Les différences G1 *vs* G3 et G2 *vs* G3 sont toutes significatives ($p < 0,05$), à l'exception du post-test lexique du premier album pour G2 *vs* G3. L'hypothèse 1 selon laquelle les élèves travaillant avec l'outil *Narramus* progressent plus que les autres est donc vérifiée en GS.

Tableau 7. Résultats des élèves de GS et MS au test É.Co.S.Se.

	Tous les élèves		G1	G2	G3	Test
	Min- max	Effectif Moyenne (Écart-type)	Effectif Moyenne (Écart-type)			p (Anova)
GS	0-10	N = 468 5,8 (2,1)	N = 210 6,0 (2,3)	N = 150 5,6 (2,0)	N = 108 5,7 (2,0)	0,201 NS
MS	0-10	N = 522 5,1 (2,3)	N = 240 5,1 (2,3)	N = 168 5,2 (2,3)	N = 114 5,2 (2,3)	0,825 NS

Légende : les effectifs (N) sont des nombres d'élèves ; NS exprime la non-significativité.

Pour évaluer l'ampleur de l'effet, nous avons calculé le coefficient *d* de Cohen[8] : une intervention est habituellement jugée inefficace si d < 0,2 ; faible si 0,2 < d < 0,5 ; moyenne si 0,5 < d < 0,8 et forte si d > 0,8 (Cohen, 1988). Nous avons comparé les performances de G1 puis de G2 à G3 (voir tableau 7). Au premier trimestre, les effets sont faibles à l'exception de la micro-structure (mi_T2) pour laquelle l'effet est moyen pour le groupe 1 (d = 0,73). Dès le deuxième trimestre, l'écart entre G1 et G2 d'une part et G3 de l'autre s'accroît et les effets sont particulièrement forts pour l'épreuve de rappel de récit. En macrostructure, le coefficient *d* est égal à 1,04 pour G1 *vs* G3 et à 0,87 pour G2 *vs* G3. En microstructure, il s'élève à 1,11 pour G1 *vs* G3 et à 0,93 pour G2 *vs* G3. Au troisième trimestre, les effets sont forts et s'amplifient encore.

Notons enfin qu'au deuxième pré-test (T3) et, plus encore, au troisième (T5), c'est-à-dire avant l'étude du deuxième et du troisième album, les scores du G1 et du G2 sont supérieurs à ceux du G3 alors qu'ils étaient équivalents au premier (T1). Les élèves du G1 et G2 sont plus compétents pour comprendre et raconter les histoires qu'un adulte vient seulement de leur lire et de leur raconter. Au troisième album, la taille de l'effet est égale à 0,48 pour le G2 et à 0,65 pour le G1. Autrement dit, les apprentissages réalisés au premier puis au deuxième trimestre se transfèrent : les élèves qui ont bénéficié de *Narramus* deviennent plus performants pour raconter des textes qu'ils n'ont pas encore étudiés.

En résumé, l'efficacité de l'intervention est avérée et va croissant au fil de l'année. Elle est jugée forte dès le second scénario, c'est-à-dire après huit semaines d'enseignement.

En moyenne section

Nous avons procédé exactement de la même manière et les résultats obtenus (présentés dans le tableau A1 placé en annexe) valident aussi l'hypothèse 1. À chaque post-test (T2, T4 et T6), les résultats des deux groupes *Narramus* (G1 et G2) sont significativement supérieurs à ceux du groupe 3 à une exception près : le rappel en T2. Comme en GS, les effets sont modestes au premier trimestre : ils ne deviennent importants qu'après deux modules de quatre semaines. Ils sont un peu moins forts qu'en GS et sont plus amples pour le rappel de récit que pour le lexique : les coefficients *d* atteignent 0,51 pour G1 et 0,55 pour G2 en macrostructure au troisième album (0,57 et 0, 63 en microstructure).

Contrôle de l'influence de la durée de l'enseignement en MS et GS

Nous l'avons indiqué plus haut (voir tableau 5), les maîtres du G3 allouent moins de temps à l'étude des albums que ceux du G1 et du G2. Dès lors, pour que l'effet *Narramus* soit plus clairement établi, il faut distinguer la part des progrès imputable à ses seules caractéristiques didactiques (choix des tâches et planification) de celle qui dépend de l'accroissement du temps d'enseignement qu'il génère.

Pour cela, nous avons procédé à une analyse de régression multiple prenant en compte les compétences initiales de chaque élève de GS et de MS après avoir centré et réduit les variables. Nous avons ainsi évalué l'influence de trois facteurs sur les performances des élèves en post-test : leur niveau initial au

8 La taille d'un effet correspond à la différence entre deux moyennes, divisée par l'écart-type de l'ensemble de la distribution : ((m1-m2)/σ).

Tableau 8. Comparaison des progrès des élèves de grande section

	Tous les élèves			G1	G2	G3	Tests				
	Épreuves	Min-Max	(Effectif) Moyenne (Écart-type)	(Effectif) Moyenne (Écart-type)			P (Tests post Anova)			d de Cohen G1 vs G3	d de Cohen G2 vs G3
Album 1	Lexique T1 (el_T1) /25	1-23	(N = 449) 14,0 (4,6)	(N = 207) 13,9 (4,9)	(N = 139) 14,2 (4,4)	(N = 103) 13,8 (4,4)	0,771				
	Macro T1 (M_T1) /20	0-19	(N = 446) 8,5 (5,4)	(N = 202) 8,4 (5,4)	(N = 139) 8,9 (5,6)	(N = 105) 8,0 (5,1)	0,408				
	Micro T1 (mi_T2) /72	0-49	(N = 446) 13,2 (9,3)	(N = 202) 14,0 (10,2)	(N = 139) 13,0 (8,5)	(N = 105) 11,8 (8,1)	0,131				
	Lexique T2 (el_T2) /25	3-25	(N = 435) 18,8 (4,4)	(N = 205) 19,2 (4,5)	(N = 135) 19,0 (4,4)	(N = 95) 17,8 (4,3)	0,034*			**0,32**	
							G1vsG3 *	G2vsG3 NS	G1vsG2 NS		
	Macro T2 (M_T2) /20	0-20	(N = 429) 13,1 (5,3)	(N = 203) 13,8 (5,2)	(N = 132) 13,2 (5,2)	(N = 94) 11,5 (4,9)	0,002***			**0,43**	**0,32**
							G1vsG3 ***	G2vsG3 *	G1vsG2 NS		
	Micro T2 (mi_T2) /72	0-70	(N = 429) 29,2 (16,3)	(N = 203) 32,9 (17,0)	(N = 132) 29,0 (17,1)	(N = 94) 21,1 (13,0)	< 0,001***			**0,72**	**0,24**
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS		
Album 2	Lexique T3 (el_31) /25	2-25	(N = 447) 13,8 (4,9)	(N = 198) 13,9 (4,9)	(N = 145) 13,8 (5,0)	(N = 104) 13,5 (4,7)	0,844				
	Macro T3 (M_T3) /30	0-29	(N = 413) 10,4 (8,3)	(N = 182) 11,5 (8,7)	(N = 136) 10,7 (8,3)	(N = 95) 8,0 (7,0)	0,004***			**0,42**	**0,1**
							G1vsG3 **	G2vsG3 *	G1vsG2 NS		
	Micro T3 (mi_T3) /82	0-65	(N = 413) 17,2 (14,7)	(N = 182) 19,5 (16,3)	(N = 136) 16,8 (13,2)	(N = 95) 13,4 (12,4)	0,004***			**0,41**	
							G1vsG3 **	G2vsG3 NS	G1vsG2 NS		
	Lexique T4 (el_T4) /25	2-25	(N = 431) 20,0 (4,9)	(N = 197) 21,5 (3,5)	(N = 136) 20,1 (4,5)	(N = 98) 16,8 (5,0)	< 0,001***			**0,96**	**0,67**
							G1vsG3 ***	G2vsG3 ***	G1vsG2 **		
	Macro T4 (M_T4) /30	0-30	(N = 397) 20,0 (9,2)	(N = 188) 22,8 (8,1)	(N = 115) 21,2 (8,8)	(N = 94) 13,2 (8,0)	< 0,001***			**1,04**	**0,87**
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS		
	Micro T4 (mi_T4) /82	0-82	(N = 397) 42,5 (24,1)	(N = 188) 50,1 (22,1)	(N = 115) 45,8 (24,1)	(N = 94) 23,3 (16,3)	< 0,001***				
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS	**1,11**	**0,93**

Album 3	Lexique T5 (el_T5) /25	2-23	(N = 422) 14,0 (3,8)	(N = 191) 14,2 (4,0)	(N = 133) 14,4 (3,7)	(N = 98) 13,4 (3,5)	0,09				
	Macro T5 (M_T5) /48	0-46	(N = 416) 21,1 (12,7)	(N = 187) 23,6 (12,4)	(N = 128) 21,8 (12,2)	(N = 101) 15,7 (12,2)	< 0,001***				
							G1vsG3 ***	G2vsG3 NS	G1vsG2 NS	**0,62**	**0,48**
	Micro T5 (mi_T5) /109	0-89	(N = 409) 28,5 (19,2)	(N = 186) 32,4 (19,8)	(N = 128) 29,3 (18,1)	(N = 95) 20,0 (16,6)	< 0,001***				
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS	**0,65**	**0,48**
	Lexique T6 (el_T6) /25	6-25	(N = 414) 19,5 (4,1)	(N = 194) 20,6 (3,6)	(N = 130) 19,7 (3,8)	(N = 90) 16,9 (4,2)	< 0,001***				
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS	**0,9**	**0,68**
	Macro T6 (M_T6) /48	0-48	(N = 394) 31,5 (12,2)	(N = 188) 34,8 (11,4)	(N = 124) 32,0 (10,8)	(N = 82) 23,4 (12,4)	< 0,001***				
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS	**0,93**	**0,7**
	Micro T6 (mi_T6) /109	0-105	(N = 390) 56,3 (26,3)	(N = 189) 64,6 (24,9)	(N = 124) 57,6 (23,1)	(N = 77) 34,0 (21,5)	< 0,001***				
							G1vsG3 ***	G2vsG3 ***	G1vsG2 *	**1,16**	**0,89**

Légende : les effectifs (N) sont des nombres d'élèves ; NS exprime la non-significativité ; * p < 0,05 ; ** p < 0,01 ; *** p < 0,001.

pré-test, la durée de l'enseignement dont ils ont bénéficié et leur groupe d'appartenance. Nous avons pris le G3 comme groupe de référence (G1 *vs* G3 et G2 *vs* G3) et avons testé, pour chaque album, un modèle explicatif que l'on peut résumer par la formule : Score final = Score initial x Temps x Groupe (Sf = Si x Tps x G).

Nous avons reporté les résultats de l'analyse de régression multiple en grande section dans le tableau 9. On peut constater que l'influence de la variable *Temps* (durée de l'enseignement dispensé pour chaque scénario) n'annihile pas celle de la variable *Groupe* et qu'elle est moins importante que cette dernière. Autrement dit, les progrès des élèves sont doublement tributaires de *Narramus* : son utilisation provoque un accroissement du temps d'enseignement et un renouvellement des activités langagières, tous deux bénéfiques aux apprentissages.

Les modèles de régression permettent d'expliquer entre 40 % et 60 % de la variance du score final pour le rappel de récit (*M* et *mi*) et environ 60 % pour le lexique.

Au troisième trimestre, par exemple (case grisée sur le tableau 9), le modèle permet d'expliquer 60 % de la variance des micro-propositions (R^2 = 0,604). Les coefficients qui expriment le poids de l'effet de chacune des variables sont égaux à 0,60 pour le niveau initial, à 0,16 pour le temps, et à 0,27 ou 0,21 pour l'appartenance aux groupes 1 ou 2. On peut donc déduire que le poids du temps, s'il n'est pas négligeable, ne suffit pas à expliquer les différences de performances finales des élèves à niveau initial contrôlé. En d'autres termes, le fait d'appartenir aux groupes *Narramus* (G1 et G2) a un effet significatif en sus de l'effet-temps.

Si l'on considère l'ensemble des données, on note que, comme dans toutes les études de ce type (Jarlégan, Piquée, Fontanieu *et al.*, 2016), le niveau initial est le facteur qui a le coefficient le plus élevé dans l'estimation du score final : entre 0,40 et 0,60 par exemple pour le rappel de récit (*M* et *mi*), entre 0,63 et 0,75 pour le lexique. La durée de l'enseignement et l'outillage ont également une influence bénéfique, à peu près équivalente. Pour le rappel de récit, la variable *Temps* a un coefficient compris entre 0,16 et 0,27 et *Groupe* entre 0,12 et 0,29 (une seule différence est non significative, au premier trimestre). Au troisième trimestre, quand l'effet *Narramus* est maximal, le poids du temps est plus faible que celui du groupe (sauf en *Macro* pour G2). Pour le lexique, l'outil explique une plus grande part de variance que le temps au premier et au deuxième trimestre mais le résultat s'inverse au troisième trimestre.

Tableau 9. Analyses de régression multiple en GS : modèle Sf = Si x Tps x G

Évaluations		Macro	Micro	Lexique
Variables Album 1	Effectif	429	429	432
	R^2	0,405	0,406	0,603
		b =	b =	b =
	Score initial	0,583 ***	0,538 ***	0,747 ***
	Temps	0,162 ***	0,198 ***	0,095 **
	G1	0,124 *	0,208 ***	0,138 **
	G2	0,077 NS	0,124 *	0,117 **
Variables Album 2	Effectif	397	397	408
	R^2	0,465	0,518	0,598
		b =	b =	b =
	Score initial	0,444 ***	0,473 ***	0,627 ***
	Temps	0,258 ***	0,274 ***	0,172 ***
	G1	0,285 ***	0,281 ***	0,378***
	G2	0,241 ***	0,264 ***	0,206 ***
Variables Album 3	Effectif	378	378	378
	R^2	0,548	0,604	0,581
		b =	b =	b =
	Score initial	0,603 ***	0,600 ***	0,627 ***
	Temps	0,174 ***	0,158 ***	0,343 ***
	G1	0,177 **	0,268 ***	0,137 ***
	G2	0,129 *	0,214 ***	0,035 NS

Légende : NS exprime la non-significativité ; * p < 0,05 ; ** p < 0,01 ; *** p < 0,001. Les coefficients b sont standardisés.

En moyenne section, nous avons procédé de la même manière et avons obtenu des résultats voisins (voir tableau A2 en annexe) même si le temps et l'appartenance groupale expliquent une moindre part de la variance qu'en GS. Le pouvoir explicatif du modèle est fort en lexique (entre 54 % et 67 %) mais plus faible en rappel de récit (entre 29 % et 46 %). Le niveau initial est encore le facteur qui explique la plus grande part de variance mais l'influence des facteurs *Temps* et *Groupe* fluctue beaucoup d'un trimestre à l'autre. Le poids du temps est plus important que celui du groupe en rappel de récit (variables *M* et *mi*) : les coefficients varient entre 0,15 et 0,22 alors que le facteur *Groupe*, qui oscille entre 0,15 et 0,19, n'est pas significatif au troisième trimestre pour des raisons qui restent à expliquer, peut-être en raison du choix de l'album 3. Sur ce point, la MS se distingue donc de la GS, la part de variance expliquée par le temps étant supérieure à celle de l'appartenance groupale alors que c'était l'inverse en fin de GS.

Test de l'hypothèse 2 : G1 *versus* G2

Pour rappel, nous avons fait l'hypothèse que, compte tenu de son mode de conception, *Narramus* devrait produire des effets similaires, qu'il soit mis en œuvre avec ou sans accompagnement (G1 *vs* G2). Nous avons vérifié (voir tableau 5) que l'existence de l'accompagnement n'influençait pas la durée allouée à l'utilisation de *Narramus* : il n'y a aucune différence significative sur ce point entre G1 et G2, en MS comme en GS. Nous avons ensuite comparé les progrès des élèves des deux groupes :

– en MS (voir tableau A1 en annexe), aucune différence n'est significative, les performances des élèves de G1 et de G2 ne sont pas différentes, en T2 comme en T4 et T6 ;

– en GS (voir tableau 8), il en va de même à l'exception de 2 scores sur 11 : G1 surpasse G2 en lexique à T4 et en microstructure à T6. Bref, pour l'essentiel (18 fois sur 20), les élèves dont les enseignants sont accompagnés ne progressent pas plus que ceux dont les maîtres travaillent seuls.

L'hypothèse 2 selon laquelle l'accompagnement n'a pas d'influence sur les progrès des élèves, en éducation prioritaire et avec *Narramus* tel qu'il est conçu, est donc validée.

Discussion

L'étude que nous avons menée visait à évaluer l'effet de la mise en œuvre d'un nouvel outil didactique sur les apprentissages langagiers de jeunes élèves. Nous supposions que les progrès de ceux qui en bénéficieraient seraient supérieurs à ceux qui en seraient privés (hypothèse 1 : l'effet *Narramus*). Nous présumions aussi que, compte tenu de son mode de conception, l'outil produirait des effets similaires que les enseignants aient été soutenus par leur équipe de circonscription ou pas (hypothèse 2 : l'absence d'effet accompagnement).

L'effet *Narramus*

Nos résultats montrent que, dès la première année de l'expérimentation, l'effet est positif sur les deux dimensions évaluées (la qualité du rappel et la mémorisation du lexique) en MS et en GS. Ils font aussi apparaître que les enseignants utilisant *Narramus* (G1 et G2) consacrent significativement plus de temps à l'enseignement que les autres (G3). Si cette augmentation influence les performances des élèves, elle ne suffit pas cependant à expliquer les différences de progrès. L'outillage produit des modifications quantitatives (accroissement du temps) et qualitatives (introduction de nouvelles tâches et de nouvelles modalités pédagogiques) qui se combinent pour favoriser les apprentissages. Cet effet est plus fort en GS qu'en MS. On observe enfin que les élèves des trois groupes qui avaient le même niveau de performance linguistique au début de l'année ne sont plus égaux au pré-test du troisième trimestre : ceux de G1 et G2 sont meilleurs que ceux de G3. Autrement dit, les élèves qui ont préalablement bénéficié de l'outil *Narramus* restituent davantage et racontent bien mieux que les autres avant d'avoir étudié un nouvel album : il y a eu transfert d'apprentissage.

Ces résultats positifs devraient intéresser le Conseil scientifique de l'Éducation nationale car ils ont été établis en respectant les principales exigences méthodologiques d'une « Éducation fondée sur la preuve » (Slavin, 2002 ; Coalition for Evidence-Based Policy, 2003) :

– nous avons constitué deux groupes expérimentaux (avec et sans accompagnement) pour dissocier l'effet « outil » de l'effet « accompagnement de l'innovation » ;
– nous avons constitué un groupe témoin actif qui bénéficiait, comme les groupes expérimentaux, d'un enseignement long (trois fois quatre semaines d'enseignement basé sur les mêmes albums, durant la même période, avec la même connaissance de l'objectif final par les professeurs) ;
– nos effectifs étaient importants : 165 classes et 990 élèves évalués individuellement ;
– nous avons pris soin de vérifier que les trois groupes composés d'élèves tirés au sort et ayant le même âge moyen étaient équivalents au départ (prétests É.Co.S.Se. et microstructure en T1) et nous n'avons procédé à aucune sélection *a posteriori* ;
– les trois groupes ont été comparés à six reprises, avant et après l'intervention, à l'aide de tests classiques en psycholinguistique qui préexistaient à la recherche et qui étaient équitables pour le groupe témoin. Les évaluations ont été administrées sur chaque site par des professionnels formés ;
– la diversité des enseignants et des élèves était de bon aloi en zones d'éducation prioritaire disséminées sur le territoire national, loin des terrains d'influence didactique des concepteurs ;
– l'étude a une forte validité écologique : les conditions d'enseignement étaient les conditions habituelles, les horaires légaux et les programmes d'enseignement ont été respectés, aucun financement supplémentaire n'a été alloué aux écoles, la hiérarchie n'a exercé aucune pression sur les enseignants ;
– l'accompagnement était assuré par des professionnels du conseil n'appartenant pas à l'équipe de recherche et dans le cadre des horaires ordinaires de formation continue en éducation prioritaire ;
– nous avons contrôlé la fidélité de la mise en œuvre des scénarios pour tester le critère de qualité de l'implémentation ;
– nous avons contrôlé le temps alloué à l'enseignement dans les trois groupes ;
– nous avons mesuré l'ampleur des effets obtenus et constaté qu'elle était forte et allait croissante en cours d'année (le *d* de Cohen se situe, selon les variables considérées, entre 0,48 et 1,16 au troisième trimestre de GS par exemple) ;
– le domaine de validité de nos conclusions se limite à l'éducation prioritaire.

Ainsi établis, nos résultats peuvent être mobilisés comme arguments dans le débat relatif à l'efficacité des dispositifs pédagogiques. En effet, si l'importance

d'un enseignement précoce de la compréhension en lecture dès l'école maternelle fait aujourd'hui l'objet d'un consensus scientifique international, les chercheurs s'opposent sur la question de ses modalités. Les uns défendent une approche « modulaire » basée sur des batteries d'exercices décontextualisés, construits autour de blocs de compétences (par exemple, un module sur les inférences, un module sur les connecteurs, etc.) travaillées de manière isolée et indépendamment de toute lecture suivie de récits complets (Gourgue, Bianco & Coda, 2013 ; Keene & Zimmermann, 1997). Ils postulent qu'une fois ces compétences construites isolément, les élèves devraient être spontanément capables de les mobiliser ensemble et de les coordonner pour comprendre, seuls, un texte long. En France, si l'étude menée par Zorman et ses collaborateurs (2015) conclut bien sur l'efficacité de ce type de pratique, sa réplication menée conjointement par le laboratoire EMC de l'université Lyon 2, l'IREDU et la DEPP (2014) se solde par des résultats décevants dans le domaine de la compréhension.

D'autres chercheurs, dont les auteurs de *Narramus*, plaident au contraire en faveur d'une approche « multi-dimensionnelle intégrée » dans laquelle le fil conducteur du scénario didactique n'est pas la succession de compétences isolées, mais le texte étudié. Ce sont les caractéristiques textuelles qui suscitent et contextualisent l'enseignement de telle ou telle compétence : au cours d'une même séance, la priorité va être donnée d'abord à l'une (l'étude approfondie du lexique, par exemple) puis à une autre (la production d'inférences pour comprendre les états mentaux des personnages) avant que les deux soient aussitôt combinées au service d'une activité langagière : raconter, expliquer, évoquer, justifier…

Les résultats de l'étude présentée ici, les premiers en contexte francophone, apportent donc un argument en faveur de cette seconde option qui domine sur le plan international comme l'atteste la méta-analyse réalisée par Swanson, Vaughn, Wanzek *et alii* (2011). En France comme ailleurs, une approche intégrative, basée sur l'étude longue d'un album de la littérature de jeunesse proposée chaque trimestre, exerce une influence significative sur le développement de compétences langagières que l'on sait déterminantes pour la suite de la scolarité (Catts, Adlof & Weismer, 2006).

L'absence d'effet de l'accompagnement

Notre seconde hypothèse prédisait que *Narramus* produirait des effets positifs qu'il soit mis en œuvre avec ou sans accompagnement. Elle a été validée : aucune différence significative ne distingue les progrès des élèves appartenant aux deux groupes expérimentaux[9].

Ce résultat peut être interprété en examinant les caractéristiques intrinsèques de l'outil. Conçu en collaboration avec des professeurs des écoles et sur la base de connaissances scientifiques avérées, il est compatible avec les schèmes professionnels des enseignants et ne bouscule pas leur organisation pédagogique même s'il les conduit à réviser progressivement leur jugement sur les capacités de mémoire et d'attention des élèves et sur la durée des tâches collectives. L'outil est simple d'emploi et les objectifs des principales innovations qu'il introduit sont faciles à comprendre. Il propose un enseignement explicite et structuré dont l'organisation est stable, ce qui facilite son appropriation. Comme il est accompagné d'un guide très détaillé, lui-même longuement pré-testé, il ne nécessite pas d'explications complémentaires apportées par un tuteur ou un conseiller. Nous en avons obtenu confirmation en étudiant les carnets de bord des maîtres : la fidélité de la mise en œuvre et la durée d'enseignement allouée aux scénarios ne présentaient aucune différence significative entre les enseignants accompagnés et ceux qui ne l'étaient pas. Ces résultats sont proches de ceux de Jenkins, Duncan, Auger *et alii* (2018) qui montrent que la formation et l'accompagnement de certains dispositifs ne génèrent pas non plus de différences de fidélité.

Un autre argument peut être avancé pour interpréter l'absence d'effet de l'accompagnement sur l'enrôlement et la persévérance des enseignants impliqués. L'outil testé est ajusté aux compétences des élèves de l'éducation prioritaire et il suscite un grand intérêt de leur part. Il génère aussi des progrès rapidement perceptibles en narration et en lexique. Ces deux phénomènes nourrissent la motivation des enseignants-utilisateurs qui ne sont pas tentés d'abandonner l'expérience. Un indice conforte cette thèse : tous ceux qui

9 L'absence d'effet porte sur les deux groupes, elle n'exclut pas que dans certaines écoles cet accompagnement ait pu être efficace. Notre méthodologie ne nous permet cependant pas d'identifier ces « déviants positifs » (Bryk, 2017) car trop peu d'élèves ont été évalués dans chaque classe.

avaient été enrôlés dans l'expérimentation la première année ont continué à utiliser l'outil l'année suivante, même hors protocole. La méthode utilisée pour élaborer l'outil (Cèbe et Goigoux, 2018) avait pour objectif d'anticiper, dès la conception, le processus de généralisation et d'essaimage d'un dispositif qui aurait fait la preuve de son efficacité lors d'une expérimentation réussie. Les résultats présentés ici permettent d'entrevoir ce « passage à l'échelle ».

Notons enfin que si l'accompagnement n'affecte pas directement les performances des élèves, il semble en revanche influencer le développement professionnel des enseignants en leur permettant de mieux comprendre pourquoi les élèves progressent et comment transposer les principes pédagogiques de *Narramus* dans d'autres domaines d'apprentissage. Cette hypothèse fera l'objet d'une analyse future à partir des questionnaires que nous avons recueillis auprès des enseignants impliqués dans l'étude.

Les limites de l'étude

Nous ne pouvons conclure sans mentionner quelques-unes des limites de cette étude pilotée par une doctorante qui travaille dans l'équipe des concepteurs de l'outil. Cette appartenance affaiblit un peu la reconnaissance internationale de l'expérimentation malgré la rigueur de la démarche, la diversité et la neutralité des enquêteurs-évaluateurs et la vérification des traitements statistiques par des spécialistes indépendants.

Les résultats présentés ici ne portent que sur la première année de l'expérimentation d'une étude longitudinale qui en compte trois. Les premières données montrent que, dans le test de rappel, l'effet de l'intervention s'amplifie au fil des scénarios et qu'il est très fort au troisième pré-test. Nous pouvons donc raisonnablement soutenir que les habiletés construites dépassent le contexte de l'album étudié puisqu'après deux écoutes (lecture et narration), les élèves racontent plus, mieux et autre chose que les élèves du groupe contrôle. Mais nous ignorons encore si, après une année scolaire et trois scénarios, nous avons atteint un effet plafond ou si le fait de bénéficier d'une deuxième année d'intervention sera à l'origine de progrès plus importants encore. C'est pourquoi, en 2017-2018, nous avons doté les enseignants de GS de trois nouveaux scénarios et avons poursuivi notre enquête en évaluant leurs élèves (anciens MS) dans les mêmes conditions que celles décrites ici. Les évaluations sont en cours d'analyse.

Isabelle Roux-Baron
Université Clermont-Auvergne, laboratoire ACTé
isabelle.roux-baron@ext.uca.fr

Sylvie Cèbe
Université Clermont-Auvergne, laboratoire ACTé
sylvie.cebe@uca.fr

Roland Goigoux
Université Clermont-Auvergne, laboratoire ACTé
roland.goigoux@uca.fr

Bibliographie

ADLOF S. M., PERFETTI C. A. & CATTS W. (2011). « Developmental changes in reading comprehension: implications for assessment and instruction ». In S. J. Samuels & A. E. Farstrup (dir.), *What research has to say about reading instruction*. Newark : International Reading Association, p. 186-214.

ATKINSON L., SLADE L., POWELL D. & LEVY J. P. (2017). « Theory of mind in emerging reading comprehension: A longitudinal study of early indirect and direct effects ». *Journal of experimental child psychology*, n° 164, p. 225-238.

BARA F. & TRICOT A. (2017). « Le rôle du corps dans les apprentissages symboliques : apports des théories de la cognition incarnée et de la charge cognitive ». *Recherches sur la philosophie et le langage*. En ligne : <https://hal.archives-ouvertes.fr/hal-01628840/document> (consulté le 5 octobre 2018).

BÉGUIN P. (2005). « Concevoir pour les genèses professionnelles ». In P. Rabardel & P. Pastré (dir.), *Modèles du sujet pour la conception. Dialectiques, activités, développement*. Toulouse : Octarès, p. 32-52.

BÉGUIN P. & CERF M. (2004). « Formes et enjeux de l'analyse de l'activité pour la conception des systèmes de travail ». *Activités*, n° 1(1).

BÉGUIN P. & DARSES F. (1998). « Les concepteurs au travail et la conception des systèmes de travail : points de vue et débats ». *Deuxièmes journées Recherche et Ergonomie*, Toulouse, p. 17-33.

BÉLANGER J., BOWEN F., CARTIER S., DESBIENS N., MONTESINOS-GELET I. & TURCOTTE L. (2012). « L'appropriation de nouvelles pratiques d'interventions pédagogiques et éducatives en milieu scolaire : réflexions sur un cadre théorique intégrateur ». *Éducation et francophonie*, n° 40(1), p. 56-75.

BIANCO M., CODA M. & GOURGUE D. (2006). *Compréhension, Moyenne section*. Grenoble : Éditions de la Cigale.

BLEWITT P., RUMP K. M., SHEALY S. E. & COOK S. A. (2009). « Shared book reading: When and how questions affect young children's word learning ». *Journal of Educational Psychology*, n° 101(2), p. 294-318.

BONNÉRY S., MAMEDE M., SOULÉ V., GRESSIER N., BASTIDE I., CRINON J., JOIGNEAUX C., MARIN B. & PINTO S. (2013). « Les usages de la littérature de jeunesse (4-8 ans) dans les structures éducatives et culturelles de la ville. Risques d'inégalités et conditions d'appropriation ». *Rapport de recherche commandé par la ville de Gennevilliers*.

BRYK S. A. (2017). « Accélérer la manière dont nous apprenons à améliorer ». *Éducation & didactique*, n° 11(2), p. 11-29. En ligne : <https://www.cairn.info/revue-education-et-didactique-2017-2-page-11.htm> (consulté le 25 juin 2018).

CABELL S. Q., JUSTICE L. M., MCGINTY A. S., DECOSTER J. & FORSTON L. D. (2015). « Teacher-child conversations in preschool classrooms: Contributions to children's vocabulary development ». *Early Childhood Research Quarterly*, n° 30, p. 80-92.

CATTS H. W. (2018). « The simple view of reading: Advancements and false impressions ». *Remedial and Special Education*, n° 39(5), p. 317-323.

CATTS H. W., ADLOF S. M. & WEISMER S. E. (2006). « Language deficits in poor comprehenders: A case for the simple view of reading ». *Journal of Speech, Language, and Hearing Research*, n° 49, p. 278-293.

CÈBE S. & GOIGOUX R. (2017). *Narramus. Apprendre à comprendre et à raconter*. Paris : Retz.

CÈBE S. & GOIGOUX R. (2018). « Former autrement pour lutter contre les inégalités ». *Recherche & Formation*, n° 87, p. 77-96.

CHARLIER E., BIÉMAR S. & JORRO A. (2012). *Accompagner : un agir professionnel*. Bruxelles : De Boeck.

CLOT Y. & FAÏTA D. (2000). « Genres et styles en analyse du travail : concepts et méthodes ». *Travailler*, n° 4(7), p. 7-42.

COALITION FOR EVIDENCE-BASED POLICY (2003). *Identifying and implementing educational practices supported by rigorous evidence. A user friendly guide*. Washington : US Department of Education, Institute of Education Sciences, National Center For Education Evaluation and Regional Assistance.

COBB P., CONFREY J., DISESSA A., LEHRER R. & SCHAUBLE L. (2003). « Design experiments in educational research ». *Educational Researcher*, n° 32(1), p. 9-13.

COHEN J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale : Erlbaum Associates.

COLLINS A. (1992) « Toward a design science of education ». In E. Scanlon & T. O'Shea (dir.), *New Directions in Educational Technology*. Berlin : Heidelberg, p. 15-22.

COOK A. E., LIMBER J. E. & O'BRIEN E. J. (2001). « Situation-based context and the availability of predictive inferences ». *Journal of Memory and Language*, n° 44(2), p. 220-234.

DEPP, LABORATOIRE EMC & IREDU (2014). *Projet lecture, rapport d'évaluation*. En ligne : <http://www.agirpourlecole.org/wp-content/uploads/2016/04/Rapport-FEJ-GS-et-CP.pdf> (consulté le 12 juillet 2018).

DICKINSON D. K. & SMITH M. W. (1994). « Long-term effects of preschool teachers book readings on low-income children's vocabulary and story comprehension ». *Reading Research Quarterly*, n° 29(2), p. 105-122.

DICKINSON D. K. & TABORS P. O. (2001). *Beginning literacy with language: Young children learning at home and school*. Baltimore : Paul H. Brookes Publishing.

DORE R. A., AMENDUM S. J., GOLINKOFF R. M. & HIRSH-PASEK K. (2018). « Theory of mind: A hidden factor in reading comprehension? ». *Educational Psychology Review*, n° 30(3), p. 1067-1089.

DURLAK J. A. & DUPRE E. P. (2008). « Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation ». *American journal of community psychology*, n° 41(3), p. 327-350.

ELWER A., GUSTAFSON S., BYRNE B., OLSON R. K., KEENAN J. M. & SAMUELSSON S. (2015). « A retrospective longitudinal study of cognitive and language skills in poor reading comprehension ». *Scandinavian Journal of Psychology*, n° 56, p. 157-166.

FISCHER M. H. & ZWAAN R. A. (2008). « Embodied language: A review of the role of the motor system in language comprehension ». *The Quarterly Journal of Experimental Psychology*, n° 61(6), p. 825-850.

FISHMAN B. J., PENUEL W. R., ALLEN A. R., CHENG B. H. & SABELLI N. O. R. A. (2013). « Design-based implementation research: An emerging model for transforming the relationship of research and practice ». *National society for the study of education*, n° 112(2), p. 136-156.

FOORMAN B., BEYLER N., BORRADAILE K., COYNE M., DENTON C. A., DIMINO J., FURGESON J., HAYES L., HENKE J., JUSTICE L., KEATING B., LEWIS W., SATTAR S., STREKE A., WAGNER R. & WISSEL S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade (NCEE 2016-4008)*. Washington : National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences.

GATHER THURLER M. G. (2000). « L'innovation négociée : une porte étroite ». *Revue française de pédagogie*, n° 130, p. 29-42.

GENTAZ E., SPRENGER-CHAROLLES L., COLÉ P., THEUREL A. & GURGAND M. (2013). « Évaluation quantitative d'un entraînement à la lecture à grande échelle pour des enfants de CP scolarisés en réseaux d'éducation prioritaire : apports et limites ». *Approche neuropsychologique des apprentissages chez l'enfant (ANAE)*, n° 123, p. 172-181.

GLENBERG A. M., GUTIERREZ T., LEVIN J. R., JAPUNITICH S. & KASCHAK M. P. (2004). « Activity and imagined activity can enhance young children's reading comprehension ». *Journal of Educational Research*, n° 96(3), p. 424-436.

GOIGOUX R. (2007). « Un modèle d'analyse de l'activité des enseignants ». *Éducation et didactique*, n° 1(3), p. 47-69.

GOIGOUX R. (2017). « Associer chercheurs et praticiens à la conception d'outils didactiques ou de dispositifs innovants pour améliorer l'enseignement ». *Éducation et didactique*, n° 11(3), p. 135-142.

GOIGOUX R., CÈBE S. & PIRONOM J. (2016). « Les facteurs explicatifs des performances en lecture-compréhension à la fin du cours préparatoire ». *Revue française de pédagogie*, n° 196, p. 67-84.

GOURGUE D., BIANCO M. & CODA M. (2013). *Compréhension, Grande section*. Grenoble : Éditions de la Cigale.

GRAESSER A. C., SINGER M. & TRABASSO T. (1994). « Constructing inferences during narrative text comprehension ». *Psychological Review*, n° 101(3), p. 371-395.

GURGAND M. (2018). « Expérimentation scolaire : du laboratoire à la classe ». *Communication au colloque « Le rôle de l'expérimentation dans le domaine éducatif »*, organisé le 1er février 2018 au Collège de France. En ligne : <https://www.college-de-france.fr/site/stanislas-dehaene/symposium-2018-02-01-14h00.htm> (consulté le 15 mars 2018).

HINDMAN A. H., SKIBBE L. E. & FOSTER T. D. (2014). « Exploring the variety of parental talk during shared book reading and its contributions to preschool language and literacy: evidence from the Early Childhood Longitudinal Study-Birth Cohort ». *Reading and Writing*, n° 27(2), p. 287-313.

HUGHES D. L., MCGILLIVRAY L. & SCHMIDEK M. (1997). *Guide to narrative language: Procedures for assessment*. Eau Claire : Thinking Publications.

IGEN-IGAENR (2011). « L'école maternelle ». *Rapport n° 2011-108*.

JARLÉGAN A., PIQUÉE C., FONTANIEU V. & GOIGOUX R. (2016). « Efficacité et équité dans l'enseignement de la lecture-écriture au cours préparatoire ». *Revue française de pédagogie*, n° 196, p. 7-22.

JENKINS J. M., DUNCAN G. J., AUGER A., BITLER M., DOMINA T. & BURCHINAL M. (2018). « Boosting school readiness: Should preschool teachers target skills or the whole child? » *Economics of Education Review*, n° 65, p. 107-125.

JUSTICE L. M., BOWLES R. P., KADERAVEK J. N., UKRAINETZ T. A., EISENBERG S. L. & GILLAM R. B. (2006). « The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative performances ». *American Journal of Speech-Language Pathology*, n° 15(2), p. 177-191.

KEENE E. O. & ZIMMERMANN S. (1997). *Mosaic of Thought: Teaching Comprehension in a Reader's Workshop*. Portsmouth : Heinemann.

KENDEOU P., VAN DEN BROEK P., WHITE M. J. & LYNCH J. S. (2009). « Predicting reading comprehension in early elementary school: the independent contributions of oral language and decoding skills ». *Journal of Educational Psychology*, n° 101(4), p. 765-778.

KIM Y. S. G. (2016). « Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children ». *Journal of Experimental Child Psychology*, n° 141, p. 101-120.

KLINGNER J. K., BOARDMAN A. G. & MCMASTER K. L. (2013). « What does it take to scale up and sustain evidence-based practices? ». *Exceptional Children*, n° 79(2), p. 195-211.

LE NORMAND M.-T., PARISSE C. & COHEN H. (2008). « Lexical diversity and productivity in French preschoolers: developmental, gender and sociocultural factors ». *Clinical Linguistics & Phonetics*, n° 22(1), p. 47-58.

LECOCQ P. (1996). *L'É.Co.S.Se. une épreuve de compréhension syntaxico-sémantique*. Villeneuve-d'Ascq : Presses universitaires du Septentrion.

LETOR C., ENTHOVEN S. & DUPRIEZ V. (2016). « L'influence conjointe des outils pédagogiques et du travail collaboratif sur le changement de représentations et de pratiques des enseignants ». *Les Dossiers des sciences de l'éducation*, n° 35, p. 37-55.

MASCAREÑO M., DEUNK M. I., SNOW C. E. & BOSKER R. J. (2017). « Read-alouds in kindergarten classrooms: a moment-by-moment approach to analyzing teacher-child interactions ». *European Early Childhood Education Research Journal*, n° 25(1), p. 136-152.

MASSEY S. L. (2004). « Teacher-child conversation in the preschool classroom ». *Early childhood education journal*, n° 31(4), p. 227-231.

MCKENNEY S. & REEVES T. C. (2014). « Educational design research ». In *Handbook of research on educational communications and technology*. New York : Springer, p. 131-140.

MCLEOD R. H., HARDY J. K. & KAISER A. P. (2017). « The effects of play-based intervention on vocabulary acquisition by preschoolers at risk for reading and language delays ». *Journal of Early Intervention*, n° 39(2), p. 147-160.

MOODY A. K., JUSTICE L. M. & CABELL S. Q. (2010). « Electronic versus traditional storybooks: Relative influence on preschool children's engagement and communication ». *Journal of Early Childhood Literacy*, n° 10(3), p. 294-313.

MYRTIL M. J., JUSTICE L. M., PELFREY L., LOGAN J. A., XIE K. & BARNES L. (2018). « Preschool teachers' implementation fidelity when using a technology-mediated language and literacy intervention ». *Child & Youth Care Forum*, n° 47(6), p. 771-786.

NATION K., COCKSEY J., TAYLOR J. S. H. & BISHOP D. V. M. (2010). « A longitudinal investigation of early reading and language skills in children with poor reading comprehension ». *Journal of Child Psychology and Psychiatry*, n° 51(9), p. 1031-1039.

NEUMAN S. B. & KAEFER T. (2018). « Developing low-income children's vocabulary and content knowledge through a shared book reading program ». *Contemporary Educational Psychology*, n° 52, p. 15-24.

NEWMAN K. M., DICKINSON D. K., HIRSH-PASEK K. & MICHNICK-GOLINKOFF R. (2015). « Using play to promote language comprehension in preschoolers ». In A. DeBruin-Parecki, A. van Kleeck & S. Gear, *Developing Early Comprehension: Laying the Foundation for Reading Success*. Baltimore : Paul H. Brookes Publishing, p. 35-52.

ODOM S. L. (2009). « The tie that binds: Evidence-based practice, implementation science, and outcomes for

children ». *Topics in Early Childhood Special Education*, n° 29(1), p. 53-61.

PENNEMAN J., DE CROIX S., DELLISSE S., DUFAYS J.-L., DUMAY X., DUPRIEZ V., GALAND B. & WYNS M. (2016). « Outils didactiques et changement pédagogique : analyse longitudinale de l'appropriation de l'outil *Lirécrire* par des enseignants du secondaire ». *Revue française de pédagogie*, n° 197, p. 79-98.

PESCO D. & GAGNÉ A. (2017). « Scaffolding narrative skills: A meta-analysis of instruction in early childhood settings ». *Early Education and Development*, n° 28(7), p. 773-793.

PETSCHER Y., JUSTICE L. M. & HOGAN T. (2017). « Modeling the early language trajectory of language development when the measures change and its relation to poor reading comprehension ». *Child development*, n° 89(6), p. 2136-2156.

RABARDEL P. & PASTRÉ P. (dir.) (2005). *Modèles du sujet pour la conception*. Toulouse : Octares Éditions.

ROWAN B. & MILLER R. J. (2007). « Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers ». *American Educational Research Journal*, n° 44(2), p. 252-297.

SÉNÉCHAL M. & LEFEVRE J. A. (2002). « Parental involvement in the development of children's reading skill: A five-year longitudinal study ». *Child development*, n° 73(2), p. 445-460.

SLAVIN R. E. (2002). « Evidence-based education policies: Transforming educational practice and research ». *Educational researcher*, n° 31(7), p. 15-21.

SNELL E. K., HINDMAN A. H. & WASIK B. A. (2015). « How can book reading close the word gap? Five key practices from research ». *The Reading Teacher*, n° 68(7), p. 560-571.

SNOW C. E. (2015). « 2014 Wallace Foundation Distinguished Lecture. Rigor and realism: Doing educational science in the real world ». *Educational Researcher*, n° 44(9), p. 460-466.

SPENCER T. D., WEDDLE S. A., PETERSEN D. B. & ADAMS J. L. (2017). « Multi-tiered narrative intervention for preschoolers: A head start implementation study ». *NHSA Dialog*, n° 20(1).

SWANSON E., VAUGHN S., WANZEK J., PETSCHER Y., HECKERT J., CAVANAUGH C. & TACKETT K. C. (2011). « A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties ». *Journal of Learning Disabilities*, n° 44, p. 258–275.

TRICOT A. (2016). « Dans quelle mesure les supports numériques peuvent-ils compliquer ou faciliter l'apprentissage et la pratique de la lecture ? ». Communication donnée dans le cadre de la conférence du consensus *Lire, comprendre, apprendre : Comment soutenir le développement de compétences en lecture ?* En ligne : <http://www.cnesco.fr/fr/lecture/paroles-dexperts/lecture-a-lheure-du-numerique/> (consulté le 15 juin 2016).

TRICOT A., PLÉGAT-SOUTJIS F., CAMPS J.-F., AMIEL A., LUTZ G. & MORCILLO A. (2003). « Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH ». In *Environnements Informatiques pour l'Apprentissage Humain*, ATIEF-INRP, p. 391-402.

TYACK D. B. & CUBAN L. (1995). *Tinkering toward utopia*. Cambridge : Harvard University Press.

WANG F. & HANNAFIN M. J. (2005). « Design-based research and technology-enhanced learning environments ». *Educational technology research and development*, n° 53(4), p. 5-23.

WASIK B. A. & HINDMAN A. H. (2011). « Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development ». *Journal of Educational Psychology*, n° 103(2), p. 455-469.

YOSHIKAWA H., WEILAND C., BROOKS-GUNN J., BURCHINAL M., ESPINOSA L., GORMLEY W. & ZASLOW M. J. (2013). *Investing in our future : The evidence base on preschool education*. New York : Foundation for Child Development.

ZHOU N. & YADAV A. (2017). « Effects of multimedia story reading and questioning on preschoolers' vocabulary learning, story comprehension and reading engagement ». *Educational Technology Research and Development*, n° 65(6), p. 1523-1545.

ZORMAN M., BRESSOUX P., BIANCO M., LEQUETTE C., POUGET G. & POURCHET M. (2015). « "PARLER" : un dispositif pour prévenir les difficultés scolaires ». *Revue française de pédagogie*, n° 193, p. 57-76.

Annexe. Résultats en moyenne section

Tableau A1. Comparaison des progrès des élèves de moyenne section

	Tous les élèves			G1	G2	G3	Test			Test	
	Épreuves	Min-Max	(Effectifs) Moyenne (Écart-type)	(Effectifs) Moyenne (Écart-type)			p Tests post Anova			d de Cohen G1 vs G3	d de Cohen G2 vs G3
Album 1	Lexique T1 (el_T1) /25	0-22	(N = 490) 10,7 (4,6)	(N = 229) 10,5 (4,5)	(N = 156) 10,6 (4,5)	(N = 105) 11,4 (5,0)	0,211				
	Macro T1 (M_T1) /20	0-18	(N = 472) 4,5 (4,4)	(N = 222) 4,2 (4,4)	(N = 146) 4,7 (4,3)	(N = 104) 4,8 (4,4)	0,416				
	Micro T1 (mi_T2) /72	0-33	(N = 472) 6,7 (6,5)	(N = 222) 6,5 (6,7)	(N = 146) 6,9 (6,5)	(N = 104) 6,8 (5,8)	0,840				
	Lexique T2 (el_T2) /25	0-25	(N = 465) 15,3 (5,5)	(N = 219) 15,6 (5,7)	(N = 145) 15,0 (5,5)	(N = 101) 15,2 (5,2)	0,551				
	Macro T2 (M_T2) /20	0-20	(N = 465) 8,5 (6,0)	(N = 220) 8,3 (6,1)	(N = 144) 9,8 (6,2)	(N = 101) 7,2 (5,4)	0,03*				**0,43**
							G1vsG3 NS	G2vsG3 **	G1vsG2 NS		
	Micro T2 (mi_T2) /72	0--67	(N = 465) 16,8 (14,6)	(N = 220) 17,0 (15,3)	(N = 144) 19,1 (15,2)	(N = 101) 12,9 (11,2)	0,004***				**0,42**
							G1vsG3 NS	G2vsG3 **	G1vsG2 NS		
Album 2	Lexique T3 (el_31) /25	0-19	(N = 482) 6,6 (3,8)	(N = 226) 6,3 (3,6)	(N = 154) 6,7 (4,0)	(N = 102) 7,1 (4,0)	0,257				
	Macro T3 (M_T3) /30	0-12	(N = 436) 1,4 (2,2)	(N = 191) 1,2 (2,0)	(N = 146) 1,4 (2,2)	(N = 99) 1,8 (2,5)	0,047*			**- 0,27**	
							G1vsG3 *	G2vsG3 NS	G1vsG2 NS		
	Micro T3 (mi_T3) /82	0-19	(N = 437) 2,5 (3,5)	(N = 192) 2,1 (3,1)	(N = 146) 2,8 (3,8)	(N = 99) 2,9 (3,5)	0,084				
	Lexique T4 (el_T4) /25	0-25	(N = 445) 13,1 (6,0)	(N = 214) 13,8 (6,1)	(N = 131) 13,9 (5,7)	(N = 100) 10,8 (5,6)	< 0,001***			**0,5**	**0,52**
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS		
	Macro T4 (M_T4) /30	0-19	(N = 431) 4,3 (4,3)	(N = 199) 4,7 (4,6)	(N = 130) 4,8 (4,1)	(N = 102) 2,8 (3,6)	< 0,001***			**0,44**	**0,46**
							G1vsG3 **	G2vsG3 **	G1vsG2 NS		
	Micro T4 (mi_T4) /82	0-45	(N = 429) 9,5 (9,2)	(N = 198) 10,3 (9,4)	(N = 129) 11,4 (9,8)	(N = 102) 5,5 (6,3)	< 0,001***				
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS	**0,52**	**0,64**

Album 3	Lexique T5 (el_T5) /25	0-24	(N = 472) 11,0 (4,5)	(N = 230) 10,8 (4,6)	(N = 141) 11,2 (4,5)	(N = 101) 11,2 (4,5)	0,668				
	Macro T5 (M_T5) /48	0-34	(N = 450) 7,8 (7,0)	(N = 206) 7,3 (7,6)	(N = 138) 9,0 (7,2)	(N = 106) 7,0 (5,4)	0,039*				
							G1vsG3 NS	G2vsG3 NS	G1vsG2 NS		
	Micro T5 (mi_T5) /109	0-59	(N = 450) 11,2 (10,9)	(N = 206) 11,1 (11,4)	(N = 128) 12,9 (11,8)	(N = 106) 9,1 (8,0)	0,024*				
							G1vsG3 NS	G2vsG3 *	G1vsG2 NS		**0,35**
	Lexique T6 (el_T6) /25	0-25	(N = 455) 15,8 (4,8)	(N = 220) 16,2 (4,9)	(N = 135) 16,1 (4,7)	(N = 100) 14,5 (4,5)	0,008**				
							G1vsG3 **	G2vsG3 *	G1vsG2 NS	**0,35**	**0,33**
	Macro T6 (M_T6) /48	0-42	(N = 434) 16,3 (11,0)	(N = 202) 17,4 (11,5)	(N = 134) 17,9 (10,9)	(N = 98) 11,8 (8,4)	< 0,001***				
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS	**0,51**	**0,55**
	Micro T6 (mi_T6) /109	0-85	(N = 434) 26,3 (20,1)	(N = 202) 28,5 (20,6)	(N = 134) 29,7 (21,0)	(N = 98) 17,1 (14,2)	< 0,001***				
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS	**0,57**	**0,63**

Légende : les effectifs (N) sont des nombres d'élèves ; NS exprime la non-significativité ; * p < 0,05 ; ** p < 0,01 ; *** p < 0,001.

Tableau A2. Analyses de régression multiple en moyenne section : modèle Sf = Si x Tps x G

Compétences		**Macro**	**micro**	**Lexique**
	Effectifs	465	465	465
	R2	0,398	0,300	0,674
Variables **Album 1**	Score initial Temps G1 G2	b = 0,606 *** 0,138 ** 0,059 NS 0,193 ***	b = 0,506 *** 0,178 *** 0,047 NS 0,167 **	b = 0,817 *** 0,094 ** 0,079* 0,023 NS
	Effectifs	431	429	445
	R2	0,293	0,339	0,539
Variables **Album 2**	Score initial Temps G1 G2	b = 0,449 *** 0,222 *** 0,152 * 0,135 *	b = 0,472 *** 0,209 *** 0,189 * 0,179 **	b = 0,694 *** 0,129 ** 0,234 *** 0,205 ***
	Effectifs	414	414	414
	R2	0,424	0,461	0,554
Variables **Album 3**	Score initial Temps G1 G2	b = 0,580 *** 0,165 ** 0,105 NS 0,074 NS	b = 0,605 *** 0,153 ** 0,092 NS 0,088 NS	b = 0,703 *** 0,163 *** 0,106 NS 0,071 NS

Légende : NS exprime la non-significativité ; * p < 0,05 ; ** p < 0,01 ; *** p < 0,001. Les coefficients b sont standardisés.