

École Doctorale de Lettres, Sciences Humaines et Sociales (ED 370)
Mention : Sciences de l'éducation (70^e section)
Laboratoire Activité, Connaissance, Transmission, Éducation (EA 481)

Thèse de Doctorat en Sciences de l'Éducation
Présentée et soutenue publiquement le 06 décembre 2019 par

Isabelle Roux-Baron

Évaluer un outil didactique : effets sur les performances langagières des élèves et les pratiques des enseignants d'école maternelle en Réseau d'Éducation Prioritaire

Membres du jury :

Marie-France BISHOP, rapporteure
Professeure des Universités, Université de Cergy Pontoise

Stéphane BONNÉRY, rapporteur
Professeur des Universités, Université Paris 8

Sylvie CÈBE, co-encadrante
Maitresse de conférence, Université Clermont Auvergne

Benoit GALAND, examinateur
Professeur à l'Université catholique de Louvain, Belgique

Roland GOIGOUX, directeur
Professeur des Universités, Université Clermont Auvergne

Annette JARLÉGAN, examinatrice
Professeure des Universités, Université de Lorraine,

Ludovic MORGE, examinateur
Professeur des Universités, Université Clermont Auvergne

Résumé

Pour que l'école maternelle française puisse jouer un rôle compensatoire des inégalités sociales, 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. C'est dans ce but que l'outil didactique *Narramus* a été conçu.

Dans notre recherche doctorale, nous avons voulu savoir si cet outil produisait les effets escomptés sur le développement des compétences langagières des élèves et sur les pratiques des enseignants. Pour cela nous nous sommes posé trois questions :

- une première d'ordre socio-pédagogique : l'utilisation d'un outil didactique innovant peut-elle favoriser les apprentissages langagiers de tous les élèves et contribuer à réduire les inégalités de réussite à l'école ?
- une seconde d'ordre méthodologique : comment construire et mettre en œuvre un dispositif expérimental d'évaluation présentant des critères rigoureux pour tester l'efficacité d'un outil didactique ?
- une troisième relative au rôle de l'outillage et de l'accompagnement dans l'amélioration des pratiques d'enseignement : quels liens existe-t-il entre l'utilisation de l'outil, l'accompagnement assuré par des équipes de circonscription et le développement professionnel des enseignants ?

Pour répondre à ces questions, nous avons mené deux études quasi-expérimentales (impliquant 250 classes et 1 500 élèves de petite, moyenne et grande section) et élaboré deux protocoles expérimentaux qui respectent les critères méthodologiques exigeants définis par les promoteurs de l'*Evidence Based Education*.

La première étude visait à identifier un éventuel effet *Narramus* sur les apprentissages enfantins et à mesurer l'influence de l'accompagnement de cette innovation sur le développement professionnel des enseignants. Nos données montrent que les élèves qui ont bénéficié de *Narramus* obtiennent de meilleurs résultats que leurs camarades du groupe contrôle sur l'ensemble des dimensions évaluées. De plus, les différences entre les deux groupes outillés (avec ou sans accompagnement) ne sont pas significatives : l'accompagnement n'est pas indispensable pour qu'un outil, s'il est bien conçu, provoque les effets positifs attendus. Nous avons cependant montré que l'accompagnement, s'il n'affecte pas directement les performances des élèves, peut agir comme une aide à l'identification de gestes professionnels nouveaux, mobilisables dans d'autres domaines d'apprentissage et semble, à ce titre, être un vecteur de développement professionnel.

La deuxième étude a mis en évidence une réduction des inégalités sociales et un transfert des compétences construites avec *Narramus* vers d'autres textes que ceux étudiés en classe : les élèves qui en ont bénéficié comprennent et racontent mieux que leurs camarades de même condition sociale et leurs performances s'approchent de celles d'enfants favorisés. Leur vocabulaire est lui-aussi enrichi.

Ces conclusions valident la pertinence de l'outil et concourent à valoriser un enseignement explicite et intégré de la compréhension et du vocabulaire auprès de jeunes élèves.

Mots-clés : évaluation - inégalités sociales - innovation - école maternelle - langage – développement professionnel

Abstract

If French nursery schools want to play a compensatory role for social inequalities, they have to allow young learners to demonstrate skills (at school) that the most privileged among them can already develop within their family surrounding. These skills account for the best viaticum for their future school curriculums: understanding of written narratives, being able to recount them and acquiring a lexicon. The teaching tool *Narramus* was created for that purpose.

As part of our doctoral research, we wanted to know whether this tool has the expected impact on developing students' language skills and on teaching methods.

For this purpose, we were interested in three questions :

-The first one deals with a socio-pedagogical issue: does using an innovative teaching tool make the language learning process of students better and contribute to reduce unequal achievement in learning at school?

-The second one deals with a methodological issue: how can we build an experimental set-up in order to evaluate the efficiency of the teaching tool based on strict criteria ?

-The third one is related to the role played by the equipment and the support for the improvement of teaching methods: what is the connection between the use of the tool, the support provided by school district teams and the professional development of teachers?

In order to answer these questions, we conducted two studies almost experimental within 250 classes, including 1500 students from the first, second and third years of nursery schools. We drew up two experimental protocols abiding by strict methodological criteria defined by the developers of *Evidence Based Education*.

The first study aimed to identify some potential *Narramus* effect on children learning process and to examine the influence of the support for this innovative tool on the professional development of teachers. Our data showed that students who had access to *Narramus* had better results than their peers from the control group in all areas assessed. Besides, the differences between the two equipped groups (with or without support) are not significant: the support is not essential to the tool (if well-conceived) to produce the expected positive effects.

However, we showed that the support, if it does not directly affect the students' performances, can although help teachers to identify new professional methods that they can re-use in other learning settings, and thus seems to be a vector of professional development.

The second study brought to light the reduction of social inequalities and a transfer of skills built through *Narramus* to other texts than those studied in class: students who have implemented it understand and recount better than their peers from the same social background and their proficiency level is close to that of more privileged children. Their vocabulary is also enhanced.

These conclusions validate the relevance of the tool and contribute to promote an integrated and explicit teaching of comprehension and of vocabulary to young students.

Key words : evaluation – social inequalities – innovation – nursery school – language – professional development.

La recherche présentée ici n'aurait pas pu être réalisée sans l'impulsion du rectorat de l'académie de Clermont-Ferrand qui nous a mise à disposition de l'Université pendant trois ans pour coordonner l'étude et réaliser notre thèse.

Elle a aussi bénéficié du soutien financier de la direction générale de l'enseignement scolaire (DGESCO – MEN), de l'Institut Carnot de l'Éducation Auvergne-Rhône-Alpes et de 11 académies de l'Éducation nationale.

Les séminaires destinés aux formateurs et accompagnateurs ont été organisés grâce au concours technique et financier du centre Alain Savary et de l'Institut français de l'Éducation (IFé, ENS-Lyon).

La création d'une plateforme dédiée à la saisie des données et la vérification de nos traitements statistiques ont été assurées par des ingénieurs de l'IFé et du laboratoire ACTé dans le cadre d'un partenariat entre ces deux structures.

Nous remercions vivement toutes les personnes de ces différentes institutions qui nous ont aidée et, en particulier, madame le recteur Danièle Campion, Patrick Picard, Yvonnick Fesselier, Valérie Fontanieu et Julie Pironom.

Nous remercions aussi l'ensemble des enseignants et des cadres de l'Éducation nationale qui ont collaboré à cette recherche.

J'adresse mes sincères remerciements à mes Directeurs de thèse - Sylvie Cèbe et Roland Goigoux - pour la confiance qu'ils m'ont accordée, la possibilité de vivre cette belle aventure au sein du laboratoire Acté, ce saut de l'école à la recherche. Un duo de directeurs de thèse chic et choc pour m'accompagner sur ces trois années, un duo passionnant et passionné qui sait partager ses connaissances et donner l'envie d'en apprendre toujours plus...

Merci à toi Roland pour tes connaissances, tes conseils, ta rigueur, ton esprit de synthèse, ta patience, ton avis éclairé. Merci de t'être appliqué à comprendre mes difficultés rédactionnelles (et si on élaborait un deuxième, un troisième, un sixième plan) et de m'avoir particulièrement bien guidée dans cet exercice.

Merci à toi Sylvie pour ton écoute, ta grande disponibilité, tes conseils avisés, ton partage de connaissances, tous tes auteurs « idoles », ta maîtrise pointue du domaine associée à ta grande pédagogie. Tu m'as guidée, formée et accompagnée. Merci pour tout ce que tu partages avec modestie, passion et générosité.

Je remercie les membres du jury d'avoir accepté de lire et d'évaluer ce travail. Je remercie également les membres de mon comité de suivi de thèse, Benoit Galand et Pierre Sève qui ont contribué à l'évolution de ce travail.

Cette thèse n'aurait pas été possible sans la coopération des nombreux enseignants, élèves, conseillers pédagogiques, inspecteurs, etc... qui y ont participé. Nous n'aurions pas pu non plus aller très loin sans l'aide de l'IFé et du centre Alain Savary et nous tenions à remercier sincèrement ces participants. Des pensées particulières pour Valérie Fontanieu, toujours disponible et de bon conseil ; Yvonnick Fesselier, d'une réactivité extraordinaire ; Patrick Picard pour son brillant esprit de synthèse. Merci à vous trois pour le partage de vos compétences et votre extrême gentillesse. Une pensée sincère également pour Cécile, Francis, Anne, Pascale et Nadia.

Une pensée pour toutes les écoles et équipes (enseignants et ATSEM) avec lesquelles nous avons travaillé. Un grand merci aux écoles Daudet (Édith, Isabelle, Laurence, Florence, Adeline, Sylvie et les formidables ATSEM), Verne (Marie, Patricia, Aurore) et Jaurès (Joëlle, Isabelle, Ombeline) de Clermont-Ferrand et à l'école Marie-Noël de Montluçon. Il est difficile de citer tout le monde mais je pense sincèrement à chacun d'entre vous.

Julie, oh, Julie sans toi pas d'Anova, de T test, de corrélation... merci pour ton initiation aux statistiques, ta patience, ta pédagogie et le temps que tu as pris pour répondre à toutes mes questions (même celles que je t'avais déjà posées, même pendant tes congés) ! Lassée des sardines, tu as préféré la reine des neiges mais tu as eu la gentillesse de ne pas en abuser !

Merci pour les aides précieuses : du relevé de données, au codage, au relecteur ou traducteur en anglais. Je pense tout particulièrement à Élisabeth Roche, Annie Llombart, Pierre Sève, Marie Pellegry, Isabelle Lardon, Christophe Albert, Guillaume, Loïc, Kamilla et mon amie Claire. Merci à Fred Chassagniol toujours sur le pont pendant ces trois années.

Je pense à Margault. Quelle bouffée d'oxygène, quel plaisir d'avoir passé cette dernière année avec toi. Un peu de chocolat, un café, quelques crasses... et en avant, la motivation remonte.

Merci à mes relecteurs hors pairs. Un œil attentif et éclairé qui permet de réajuster, améliorer, finaliser : merci à mon père, Silvère, Margault et ma Clairette.

Merci à toute l'équipe du laboratoire Acté.

Merci à Géraldine Rix-Lièvre pour son aide, son écoute, ses conseils qui ont contribué à l'avancée de mon travail et à insuffler de l'énergie dès que le moral était en berne.

Merci à Pierre Sève pour ses encouragements, ses conseils, son amour des livres, son initiation à une autre grammaire et une autre littérature !

Merci à Mickaël Jury pour ses explications des analyses de contraste, son avis toujours prudent, sa bienveillance et sa bonne humeur.

Merci à Patricia Roche, un bonheur au quotidien, toujours un mot sympathique et un sourire.

Merci aux collègues de l'ESPE pour les échanges autour de ce travail et les moments de détente partagés le midi.

Merci à Jérôme Riou pour tes analyses, tu as tracé une voie que j'essaie de suivre.

Merci à ma famille qui m'a beaucoup encouragée.

Je pense particulièrement à ma mère Annick et à André, toujours disponibles pour donner un coup de main, s'occuper des enfants et nous rendre la vie plus facile.

Merci à mon père et à Dominique pour les mercredis partagés avec les enfants et les messages pour venir aux nouvelles.

Une pensée pour Caro qui m'a particulièrement apporté dans les moments compliqués, une oreille attentive et des conseils apaisants.

Merci aux collègues de l'école Aristide Briand pour leurs encouragements et leurs gestes d'affection. C'est une chance de travailler avec une telle équipe !

Je souhaite également remercier tous mes amis de longue date qui ont su m'apporter confiance et écoute à tous les moments. Vous avez compris les rendez-vous manqués, nous allons nous rattraper. Je pense évidemment à « la Sicile, les 40's à Amsterdam », mon Seb, ma Lolo, il y aura encore des dizaines à fêter. Je pense à Aude, les bulles arrivent et tu les partageras avec moi ! Une pensée pour Marc.

Mes plus sincères pensées pour Claire, Ophélie et Véro. Vous avez toujours veillé sur moi, pris le relais avec les enfants. Ma Phopho, une oreille attentive qui rebooste quand on est au plus bas. Véro et Claire, toujours présentes, d'une grande générosité, vous êtes au top !

Merci à Mary, toujours un message, un encouragement, un petit signe pour prendre des nouvelles et me donner du courage.

Merci à Éva pour tes messages, tes rappels à la réalité, ton discernement et ton amitié. Je te dois beaucoup, tu m'as appris à devenir l'enseignante que je suis et m'a donné l'envie d'aller plus loin.

Merci à Laure, trois ans et deux grandes aventures pour nous de chaque côté de l'océan. Tes « force, zen et courage », tes petites blagues m'ont accompagnée tout au long de la thèse. Quelle joie de pouvoir partager la fin de ce parcours de nouveau à tes côtés.

Merci à Steph d'avoir su m'écouter, m'encourager, m'épauler. Toujours là depuis 2011 (quelle bonne idée ce master !), depuis c'est un plaisir de travailler avec toi et d'être devenue ton amie.

Merci à Betty, un océan nous sépare mais tu as toujours suivi de près mon avancée. Vivement l'été 2020 !

Des pensées émues pour Julien et Laurent, le chemin se poursuit sans vous mais vous n'êtes jamais très loin...

Comment ne pas remercier Silvère, mon cher et tendre mari qui a su écouter, conseiller, encourager, dédramatiser voire, parfois, me bousculer mais surtout m'entourer de tout son amour pendant ces trois années. Toujours une parole reconfortante, un regard objectif et une confiance en mon travail. Sans toi le parcours aurait eu moins de saveur...

À mes deux grands bonheurs, mes enfants Salomé et Antoine. À ces petits déjeuners manqués, ces sorties au cinéma, au parc, « aux puces » le dimanche matin, à la piscine pour « laisser maman tranquille, elle travaille avec son ordinateur ». Je vous aime de tout mon cœur.

Table des matières

INTRODUCTION	13
CHAPITRE 1 ÉVALUER UNE INNOVATION PÉDAGOGIQUE	19
INTRODUCTION	19
1. DES INÉGALITÉS SOCIALES AUX DIFFICULTÉS SCOLAIRES	20
1.1. DES INÉGALITÉS PERSISTANTES	20
1.2. L'INFLUENCE DES PRATIQUES D'ENSEIGNEMENT SUR LA PRODUCTION DES INÉGALITÉS	21
1.2.1. Des résultats d'évaluation d'une stabilité préoccupante en fin de cycle 3	22
1.2.2. Des différences d'efficacité précoces	24
1.2.3. Les évaluations nationales CP.....	25
2. AMÉLIORER L'ÉCOLE, UNE NÉCESSITÉ	31
2.1. AMÉLIORER PAR LA FORMATION	32
2.1.1. Améliorer par l'utilisation d'outils	33
2.1.2. De l'outil didactique à l'outil professionnel	35
3. AMÉLIORER PAR L'INNOVATION	36
3.1. LE CONCEPT D'INNOVATION	37
3.2. INNOVER DANS L'ENSEIGNEMENT	38
3.3. AMÉLIORER, INNOVER ET ÉVALUER !	39
4. ÉVALUER UNE INNOVATION	40
4.1. ÉVALUER : UN DOUBLE ENJEU	41
4.2. ADMINISTRER LA PREUVE DE L'EFFICACITÉ D'UNE PRATIQUE EN ÉDUCATION	42
5. UNE MÉTHODOLOGIE RIGOREUSE	43
5.1. LES FONDEMENTS DE L'ÉDUCATION FONDÉE SUR LA PREUVE	43
5.2. DÉFINITION ET OBJECTIFS DE L'EVIDENCE-BASED EDUCATION	44
5.3. DES EXIGENCES MÉTHODOLOGIQUES FORTES	45
5.3.1. Les critères méthodologiques retenus par l'evidence-based education	46
5.3.1.1. Des standards méthodologiques	46
5.3.1.2. Les conditions d'une adhésion aux pratiques de recherche interventionnelle.....	48
5.3.1.3. La revue systématique ou « méta-analyse »	48
5.3.2. Comment procéder pour évaluer ?	50
5.3.3. Synthèse.....	52
6. LE DÉLICAT PASSAGE À L'ÉCHELLE	53
6.1. DÉFINITION DU PASSAGE À L'ÉCHELLE	53
6.2. LES OBSTACLES AU PASSAGE À L'ÉCHELLE	55
6.2.1. Une tension entre la prescription et l'autonomie des enseignants.....	55
6.2.2. L'accompagnement ou le suivi des équipes.....	58
6.2.2.1. Caractéristiques de l'accompagnement.....	58
6.2.2.2. L'accompagnement, une nécessité ?	59
6.2.3. La prise en compte des contextes ordinaires d'enseignement.....	61
6.2.4. Le respect du prescrit	63
6.2.5. L'environnement organisationnel.....	63
6.2.6. Un juste équilibre entre les gains et les pertes liés à l'innovation.....	64
6.2.7. Une réelle appropriation	65
7. UNE ALTERNATIVE : LA PROPOSITION DE L'ÉQUIPE CLERMONTOISE	66
8. CONCLUSION	68
CHAPITRE 2 DES PRATIQUES DE LECTURES PARTAGÉES À UN ENSEIGNEMENT EXPLICITE DE LA COMPRÉHENSION	70

INTRODUCTION	70
1. LE DÉVELOPPEMENT DU LANGAGE : UNE AFFAIRE DE FAMILLE	71
1.1. LES LECTURES FAMILIALES : LES MEMES ACTIVITES DANS TOUTES LES MAISONS ?	71
1.2. QUE NOUS APPRENNENT LES RECHERCHES SUR LES APPRENTISSAGES ENFANTINS ?	72
1.3. UNE RECHERCHE FRANÇAISE.....	73
1.3.1. <i>L'acquisition de connaissances spécifiques : le vocabulaire et les connaissances encyclopédiques</i>	75
1.3.2. <i>L'acquisition de connaissances spécifiques : les compétences narratives en réception ..</i>	77
1.4. QU'EN EST-IL A L'ÉCOLE MATERNELLE ?	80
1.4.1. <i>Les styles de lecture</i>	80
1.4.2. <i>Les styles d'interaction</i>	82
1.4.2.1. Styles d'interaction en cours de lecture.....	83
1.4.2.2. Conclusion	84
1.4.3. <i>Le rapport de l'Inspection générale (2011)</i>	84
1.4.4. <i>Lit-on les mêmes albums en REP et hors REP et propose-t-on les mêmes activités ?</i>	85
1.4.5. <i>Des pratiques éducatives familiales aux pratiques d'enseignement de l'école maternelle</i>	87
2. L'OUTIL NARRAMUS	88
2.1. UNE CONCEPTION PARTICULIERE	88
2.2. LES CONNAISSANCES ET COMPETENCES VISEES PAR NARRAMUS	90
2.2.1. <i>Des compétences narratives en réception</i>	90
2.2.2. <i>Des compétences narratives en production</i>	91
2.2.3. <i>Des compétences lexicales et syntaxiques</i>	91
2.2.4. <i>Des compétences inférentielles</i>	91
2.3. DESCRIPTION DE L'OUTIL	92
2.3.1. <i>Les objectifs</i>	92
2.3.2. <i>Le déroulé d'un scénario</i>	93
2.3.3. <i>Les caractéristiques de l'outil</i>	94
2.3.3.1. Un outil qui relève de l' <i>embedded instruction</i> (enseignement intégré)	94
2.3.3.2. Un enseignement explicite.....	94
2.3.3.3. Un enseignement ritualisé	97
2.3.3.4. Une attention conjointe favorisée	99
2.3.3.5. Un recours aux supports multimédia numériques.....	100
2.3.3.6. Un enseignement particulier du lexique	100
2.3.3.7. Une mise en scène des histoires	103
2.4. UN OUTIL QUI REPOND A LA PRESCRIPTION	103
2.4.1. <i>Le programme de l'école maternelle (BO mars 2015)</i>	104
2.4.2. <i>Les ressources de l'école maternelle : Éduscol</i>	105
2.4.3. <i>L'école maternelle, école du langage : note de service n°2019-084 du 28 mai 2019</i>	106
3. PROBLÉMATIQUE ET QUESTIONS DE RECHERCHE	107
3.1. PROBLEMATIQUE GENERALE	107
3.2. OBJECTIFS DE RECHERCHE	108
3.3. HYPOTHESES DE RECHERCHE	109
3.4. DESIGN DE RECHERCHE	110
3.5. SYNTHÈSE	111
CHAPITRE 3 MÉTHODE DE L'ÉTUDE 1	114
1. LE DISPOSITIF EXPÉRIMENTAL DE L'ÉTUDE 1	115
1.1. LES SEMINAIRES DE FORMATION	116
1.2. LES FORMATEURS	118
1.3. LES EVALUATEURS	120
1.4. LES ENSEIGNANTS.....	121
1.4.1. <i>Informers les enseignants avant le début de la recherche</i>	121
1.4.2. <i>Les engagements des enseignants des groupes 1 et 2</i>	122

1.4.3.	<i>Les engagements des enseignants du groupe 3</i>	123
1.4.4.	<i>Les caractéristiques des enseignants</i>	123
1.4.5.	<i>Le contrat établi avec les enseignants de chaque groupe</i>	124
1.5.	LES ELEVES	125
1.6.	LES ALBUMS	126
1.6.1.	<i>Les instruments de recueil des données</i>	127
1.6.2.	<i>Les épreuves d'évaluation</i>	127
1.6.2.1.	L'organisation temporelle des épreuves	127
1.6.2.2.	Épreuve de sélection des élèves de MS et GS : compréhension de phrases entendues	127
1.6.2.3.	Épreuves d'évaluation des apprentissages	129
1.6.2.3.1.	Rappel de l'histoire	130
1.6.2.3.2.	Lexique	133
1.6.2.3.2.1.	Les mots évalués	134
1.6.2.3.2.2.	Le déroulement de l'épreuve	135
1.6.2.3.2.3.	Le maintien du lexique	138
1.6.2.3.3.	Questionnaire	138
2.	LES PRATIQUES DES ENSEIGNANTS	139
2.1.	LES CARNETS DE BORD	139
2.1.1.	<i>Les carnets de bord des enseignants des groupes 1 et 2</i>	139
2.1.2.	<i>Les carnets de bord des enseignants du groupe 3</i>	140
2.1.2.1.	Les caractéristiques des carnets de bord des enseignants du groupe 3	140
2.1.2.2.	Synthèse des contenus des carnets de bord du G3	141
2.1.2.2.1.	Description des carnets de bord renseignés lors du travail sur l'album 1 par les enseignants du groupe 3	141
2.1.2.2.2.	Description des carnets de bord renseignés par le groupe 3 lors de l'enseignement sur l'album 2	142
2.1.2.2.3.	Description des carnets de bord renseignés par le groupe 3 lors de l'enseignement sur l'album 3	143
2.1.2.3.	Conclusion	145
2.2.	LES QUESTIONNAIRES	146
3.	LE RECUEIL DES DONNÉES SUR L'ACCOMPAGNEMENT	147
3.1.	DESCRIPTION DES ACCOMPAGNEMENTS	148
3.1.1.	<i>Descriptif général de la nature des accompagnements</i>	148
3.1.1.1.	Présentation synthétique des accompagnements	148
3.1.1.1.1.	Les objectifs ciblés par les accompagnateurs	149
3.1.1.1.2.	Les modalités et les contenus des accompagnements	150
3.1.2.	<i>Du travail collectif à un travail plus solitaire</i>	151
3.1.2.1.	Des accompagnements réalisés en « réseau » de formateurs	151
3.1.2.2.	Des accompagnements assurés par un seul et même formateur	154
3.2.	SYNTHESE	156
	CHAPITRE 4 RÉSULTATS DE L'ÉTUDE 1 ANNÉE 1 (2016-2017)	159
1.	RAPPEL DES HYPOTHÈSES ET DES QUESTIONS DE RECHERCHE	159
1.1.	VARIABLES DE CONTROLE DES PRATIQUES PÉDAGOGIQUES	159
1.1.1.	<i>Qualité de la mise en œuvre (implémentation) : la fidélité</i>	160
1.1.2.	<i>Contrôle du temps alloué à l'enseignement</i>	162
1.2.	CONSTITUTION DES TROIS GROUPES	164
1.3.	ÉQUIVALENCE INITIALE DES TROIS GROUPES	165
2.	LES RÉSULTATS DES ÉLÈVES DE MOYENNE ET GRANDE SECTION EN 2016-2017	167
2.1.	TEST DE L'HYPOTHÈSE 1 : G1 ET G2 VERSUS G3	167
2.1.1.	<i>Résultats en compréhension</i>	167
2.1.1.1.	Accord inter-juges	167
2.1.1.2.	Corrélations des épreuves de compréhension	168
2.1.1.3.	Rappel de récit et questionnaire : résultats des élèves de Grande Section	172
2.1.1.4.	Rappel de récit et questionnaire : résultats des élèves en Moyenne section	177
2.1.2.	<i>Résultats en lexique</i>	181

2.1.3.	<i>Le post-test immédiat : résultats des élèves de Grande Section</i>	181
2.1.3.1.	Le post-test immédiat : résultats des élèves de Moyenne section	183
2.1.3.2.	Le post-test différé : résultats des élèves de Moyenne et Grande sections	185
2.1.4.	<i>Contrôle de l'influence de la durée de l'enseignement en MS et GS</i>	186
2.2.	TEST DE L'HYPOTHESE 2 : G1 VERSUS G2	189
2.2.1.	<i>Les résultats en fonction du niveau initial de l'élève</i>	190
2.2.1.1.	<i>Narramus</i> : un effet différencié en fonction du niveau initial ?	191
2.2.1.1.1.	Un effet <i>Narramus</i> pour les élèves faibles de l'Éducation prioritaire ?	191
2.2.1.1.2.	<i>Narramus</i> : jusqu'où va la démocratisation ?	195
2.2.1.1.2.1.	Progrès des élèves de Grande Section	196
2.2.1.1.2.2.	Progrès des élèves de Moyenne Section	200
2.2.1.2.	Analyses inférentielles	204
2.2.1.2.1.	Rappel de récit en Grande Section et en Moyenne Section	204
2.2.1.2.2.	Lexique en Grande Section et en Moyenne Section	207
2.3.	LES RESULTATS DES ELEVES DE PETITE SECTION	210
2.3.1.	<i>Test de l'hypothèse 1</i>	211
2.3.1.1.	Rappel de récit et questionnaire : résultats des élèves de Petite Section	211
2.3.1.2.	Résultats en lexique	213
2.3.1.2.1.	Les post-tests immédiats : résultats des élèves de Petite Section	213
2.3.1.2.2.	Le post-test différé	214
2.3.1.3.	Synthèse des résultats des élèves de PS	215
2.3.1.4.	Contrôle de l'influence de la durée de l'enseignement en PS	215
2.3.2.	<i>Test de l'hypothèse 2</i>	217
CHAPITRE 5 RÉSULTATS DE L'ÉTUDE 1 : ANNÉES 2 ET 3 (2017-2019)		219
1.	RÉSULTATS ANNÉE 2 : UN EFFET DE CUMUL ?	220
1.1.	COMPARAISON DES PROGRES DES ELEVES : 1 AN DE <i>NARRAMUS</i> VS 2 ANS DE <i>NARRAMUS</i>	220
1.1.1.	<i>Comparaison des progrès des élèves de MS</i>	221
1.1.2.	<i>Comparaison des progrès des élèves de GS</i>	223
1.2.	SUIVI DES RESULTATS DES ELEVES SUR DEUX ANS.....	225
1.2.1.	<i>Suivi des élèves de PS vers MS</i>	225
1.2.2.	<i>Suivi des élèves en MS vers GS</i>	227
2.	RÉSULTATS ANNÉE 3	230
2.1.	UN MEME TEST SUR UN MEME ALBUM EN 2017 ET 2019	230
2.2.	LES RESULTATS	231
2.2.1.	<i>Variables temps et fidélité</i>	231
2.2.2.	<i>Comparaison des résultats</i>	232
2.2.3.	<i>Conclusion</i>	233
3.	DISCUSSION DES RÉSULTATS DE LA PREMIÈRE ÉTUDE	233
3.1.	L'EFFET <i>NARRAMUS</i>	234
3.1.1.	<i>Un enseignement spécifique du lexique</i>	234
3.1.2.	<i>Un enseignement innovant des compétences narratives</i>	238
3.1.3.	<i>Une mise en scène des histoires</i>	240
3.1.4.	<i>« Un temps d'attente » fréquemment proposé</i>	241
3.1.5.	<i>Conclusion</i>	243
3.2.	L'ABSENCE D'EFFET DE L'ACCOMPAGNEMENT	245
3.3.	L'ÉTUDE 2 : UN PROLONGEMENT DE CES PREMIERS RESULTATS.....	247
3.4.	LES LIMITES DE L'ÉTUDE	247
CHAPITRE 6 MÉTHODE, RÉSULTATS, DISCUSSION DE L'ÉTUDE 2		252
1.	LE DISPOSITIF EXPÉRIMENTAL DE L'ÉTUDE 2	253
1.1.	COMPOSITION DE L'ÉCHANTILLON	253
1.1.1.	<i>Les groupes</i>	253
1.1.2.	<i>Les élèves de Grande section</i>	254
1.2.	CONSTITUTION DES TROIS GROUPES	255
1.3.	LE CHOIX DE L'ALBUM SUPPORT DE L'ÉVALUATION	258

1.4.	LES EPREUVES D'ÉVALUATION	259
1.4.1.	<i>Déroulement de l'épreuve</i>	259
1.4.1.1.	Épreuve 1 : rappel de l'histoire	260
1.4.1.1.1.	Épreuve 2 : questionnaire	261
1.4.1.1.2.	Épreuve 3 : lexique	261
2.	RÉSULTATS	263
2.1.	HYPOTHESE 1 : VERS UN TRANSFERT DES COMPETENCES ?	263
2.1.1.	<i>Rappel de récit</i>	265
2.1.2.	<i>Questionnaire</i>	265
2.1.3.	<i>Lexique</i>	266
2.1.4.	<i>Conclusion</i>	266
2.2.	HYPOTHESE 2 : VERS UNE RÉDUCTION DES INÉGALITÉS ?	266
3.	DISCUSSION	273
4.	CONCLUSION	280
CHAPITRE 7 LA TRANSFORMATION DES PRATIQUES : INFLUENCE DE L'UTILISATION D'UN OUTIL ET RÔLE DE L'ACCOMPAGNEMENT		284
1.	QUELQUES CONSIDÉRATIONS THÉORIQUES	285
1.1.	LE DÉVELOPPEMENT PROFESSIONNEL	285
1.1.1.	<i>L'approche développementale</i>	285
1.1.2.	<i>L'approche professionnalisante</i>	286
1.2.	L'ACCOMPAGNEMENT	286
1.2.1.	<i>Considérations théoriques</i>	286
1.2.2.	<i>L'accompagnement dans le cadre de notre recherche</i>	287
2.	OUTILLER ET ACCOMPAGNER LES ENSEIGNANTS : DES VECTEURS DE DÉVELOPPEMENT PROFESSIONNEL ?	288
2.1.1.	<i>Question de recherche n° 1 : quels sont les liens qui unissent la prise en main d'un outil et le développement professionnel ?</i>	<i>288</i>
2.1.1.1.	Méthode	288
2.1.1.2.	Résultats	289
2.1.2.	<i>Question de recherche n° 2 : y a-t-il un lien entre accompagnement et développement professionnel, si oui lequel ?</i>	<i>298</i>
2.1.2.1.	Méthode	298
2.1.2.2.	Résultats : la modification des pratiques	298
2.1.3.	<i>L'accompagnement des enseignants : un accélérateur du développement professionnel ?</i>	<i>302</i>
2.1.3.1.	L'accompagnement : une aide à la prise en main de l'outil ?	302
2.1.3.2.	L'accompagnement : une aide à la compréhension des bonnes raisons que l'on a de faire ce que l'on nous demande de faire	302
2.1.3.3.	L'accompagnement favorise-t-il le transfert des pratiques ?	305
CONCLUSION GÉNÉRALE		307
RÉFÉRENCES BIBLIOGRAPHIQUES		312
TABLE DES FIGURES		337
TABLE DES TABLEAUX		339
TABLE DES ANNEXES		343

Introduction

Introduction

Notre travail de recherche se situe dans le prolongement de notre expérience d'enseignante en écoles primaires (2001-2016) et de formatrice à l'Institut de Formation des Maîtres puis à l'École Supérieure de Professorat et d'Éducation (2008-2019). Il est nourri par les échanges réguliers que nous avons eus avec nos collègues formateurs sur la question de l'amélioration de l'enseignement par l'introduction d'outils adaptés aux besoins d'enseignement des élèves : quel manuel proposer aux étudiants et aux enseignants et sur quelles bases le faire ? Comment les aider à porter un regard éclairé sur les outils disponibles ? Comment les conseiller sans être guidée par trop de subjectivité ?

À travers cette thèse, nous souhaitons contribuer à la réflexion portant sur la place des outils didactiques dans la formation. Nous jugeons nécessaire de disposer de résultats sur leur efficacité et de diffuser ceux-ci pour que les enseignants puissent choisir en toute connaissance de causes (et d'effets). Aujourd'hui, en effet, le choix de manuels fait par les enseignants, entre les multiples propositions éditoriales, repose sur de nombreux critères : leurs valeurs et leurs croyances, leurs habitudes professionnelles, la connaissance qu'ils ont des besoins de leurs élèves, leur contexte d'exercice, les prescriptions des formateurs ou des inspecteurs de l'Éducation nationale, l'aura des auteurs de manuels, etc. Pour éviter des choix trop hasardeux et fonder des recommandations de manière rationnelle, il est nécessaire de disposer de données probantes sur leur qualité.

Depuis quarante ans, les lois scolaires de la cinquième République ont toutes assigné à l'école un rôle compensatoire des inégalités sociales d'apprentissage. Les réformes éducatives se sont succédé, modifiant, au nom d'une école plus juste, les programmes et les organisations scolaires (création des REP, des cycles, CP à douze...). En dépit de ces nombreuses tentatives, l'échec scolaire touche encore massivement les élèves de milieux populaires et la France éprouve des difficultés pour enrayer ce phénomène. Le rapport « Grande pauvreté et réussite scolaire » (Delahaye, 2015) en témoigne : ce sont les enfants de milieux populaires qui sont le plus touchés par les inégalités d'accès aux savoirs scolaires. Si certains élèves sont, dès leur plus jeune âge, en connivence avec les demandes de l'école, d'autres s'y trouvent dans des situations inédites et n'entrent pas à l'école maternelle dotés des mêmes connaissances et des mêmes compétences. Ceci est particulièrement vrai dans le domaine du développement langagier (lexique, syntaxe, connaissances sur le monde, conceptualisation et compréhension

de l'écrit), domaine dans lequel les écarts d'efficience sont très marqués (Grobon & Panico, 2019; Le Normand, Parisse, & Cohen, 2008).

Si l'on ne peut nier l'influence du milieu familial sur les performances scolaires, celui-ci ne détermine pas, à lui seul, le devenir des enfants. On a régulièrement montré que les enseignements proposés à l'école maternelle peuvent réduire les écarts à condition de prendre pour cibles les connaissances et compétences qui permettent d'envisager la suite de la scolarité avec confiance. C'est pourquoi politiques, chercheurs, formateurs et enseignants s'accordent pour allouer à l'école maternelle et aux pratiques d'enseignement qui ont fait la preuve de leur efficacité un rôle crucial dans la lutte contre les inégalités sociales d'apprentissage. Notre thèse de doctorat vise à y contribuer.

Nous nous demandons, en effet, comment aider les enseignants à intégrer dans leur pratique des activités cohérentes avec les résultats de la recherche. Les réponses de nombreux chercheurs (Berman, 1980; Cartier, Contant, & Janosz, 2012; Cèbe & Goigoux, 2018; Chartier, 2000; Dupriez, 2015; Rowan & Miller, 2007) vont dans le même sens : il faut doter les professeurs d'outils didactiques « efficaces » pour changer leurs pratiques et favoriser leur développement professionnel. Cette conception est largement partagée par les enseignants qui, quand on les interroge sur les caractéristiques d'une bonne formation continue, disent vouloir : « appréhender de nouvelles démarches pédagogiques ; disposer d'éléments concrets pour faire la classe ; découvrir de nouveaux outils ; améliorer et mutualiser leurs pratiques ; se tenir informés des avancées de la recherche ; s'adapter aux nouveaux programmes » (IGEN & IGAENR, 2017, p. 28).

Cependant, si tous les outils ne se valent pas, on sait peu de choses, aujourd'hui, sur leurs éventuels effets à la fois du côté des apprentissages des élèves et de celui des pratiques. C'est pour pallier ce manque que nous avons entrepris l'évaluation des effets d'un outil didactique nouveau et innovant, *Narramus*, qui organise l'enseignement de la compréhension de textes narratifs à l'école maternelle. Nous nous sommes assignée un objectif ambitieux, celui de respecter les critères méthodologiques les plus exigeants, tels qu'ils sont définis par les promoteurs de l'*Evidence Based Education*.

Mais on sait qu'infléchir les pratiques d'enseignement d'enseignants chevronnés n'est pas chose facile surtout quand les outils ont été élaborés, par des chercheurs, en laboratoire. Aussi beaucoup d'entre eux décident-ils d'accompagner la prise en main et l'usage à long terme. Nous l'avons fait aussi mais dans un objectif totalement différent puisque nous avons

fait l'hypothèse que le mode de conception de *Narramus*, très singulier, devrait rendre, *de facto*, l'accompagnement inutile.

Notre recherche vise par conséquent à apporter des éléments de réponse à trois interrogations :

- une première d'ordre socio-pédagogique : l'utilisation d'un outil didactique innovant peut-elle favoriser les apprentissages langagiers de tous les élèves et contribuer à réduire les inégalités de réussite à l'école ?

- une seconde d'ordre méthodologique : comment construire et mettre en œuvre un dispositif expérimental d'évaluation présentant des critères méthodologiques rigoureux pour tester l'efficacité d'un outil didactique ?

- une troisième relative au rôle de l'outillage et de la formation dans l'amélioration des pratiques d'enseignement : quels liens existe-t-il entre l'utilisation de l'outil, l'accompagnement assuré par des équipes de circonscription et le développement professionnel des enseignants ?

Nous rendons compte de notre réflexion en sept chapitres.

Nous précisons d'abord ce que l'on entend, habituellement, par les termes « évaluation d'une innovation » puis les principes sur lesquels repose une approche fondée sur les données probantes (*evidence based education*) qui fondent notre protocole expérimental. Nous présentons ensuite les résultats d'études portant sur l'accompagnement des innovations pédagogiques et de leurs effets à court terme (la mise en œuvre) et à long terme (la diffusion).

Notre deuxième chapitre traite des activités de « lecture partagée » et du rôle crucial qu'elles jouent dans le développement du langage oral et écrit des jeunes enfants. Nous verrons aussi quels sont les enjeux fondamentaux portés par l'école maternelle, particulièrement pour les élèves de milieux populaires. Cette partie fonde une partie de notre entreprise : l'évaluation des effets de *Narramus*, outil que nous décrivons ensuite en nous centrant sur ses différentes caractéristiques (la manière dont il a été conçu, les objectifs visés, les principes didactiques et pédagogiques sur lesquels il repose, les activités emblématiques et leurs justifications scientifiques ainsi que la mise en œuvre préconisée par les auteurs). Ce sont ces caractéristiques qui sont à l'origine de nos questions de recherche et d'une partie de notre méthodologie.

Les quatre chapitres suivants (3 à 6) relatent la méthode et les résultats de notre travail empirique, résultats que nous analysons et discutons à la fin de chaque étude. Nous y présentons deux études quasi-expérimentales de grande ampleur réalisées à l'École maternelle. Dans la

première qui ne concerne que des élèves scolarisés en Réseau d'Éducation Prioritaire, nous évaluons l'effet de l'outil *Narramus* sur les apprentissages langagiers, à différentes échéances temporelles (étude 1). Dans la seconde, nous cherchons à savoir si l'outil est à l'origine de progrès cognitifs solides, généralisables et fonctionnellement intégrés, bref s'il permet bien de réduire les inégalités entre élèves de milieux sociaux contrastés (étude 2).

Dans le dernier chapitre, nous analysons les effets de l'accompagnement sur le développement professionnel des enseignants en lien avec l'utilisation de l'outil.

Nous concluons en dégagant quelques principes de travail (et de recherche) qui pourraient être utilisés avec profit par les enseignants de maternelle qui veulent, comme nous, améliorer les conditions d'apprentissage faites aux élèves défavorisés et par les chercheurs qui veulent évaluer les effets d'outils innovants (ou pas).

Partie 1

Problématique

Chapitre 1

Évaluer une innovation pédagogique

Chapitre 2

Des pratiques de lectures partagées à un enseignement explicite de la compréhension

Chapitre 1

Évaluer une innovation
pédagogique

Chapitre 1 Évaluer une innovation pédagogique

« Si Avon et Toyota peuvent dépenser des millions de dollars en recherche pour créer de meilleurs produits, comment les écoles peuvent-elles continuer à utiliser de prétendues « pratiques exemplaires » sans recueillir de preuves sur ce qui fonctionne vraiment » (Fischer, Goswami, & Geake, 2010, p. 69).

Introduction

Faire réussir les élèves est une priorité pour notre système éducatif qui essaie de tendre vers une école plus juste. Cette ambition, partagée par les professionnels de l'enseignement, nécessite d'être accompagnée par des propositions d'actions concrètes. Ainsi, les projets foisonnent autour de nouvelles pratiques pédagogiques et les expérimentations d'outils didactiques novateurs capables de favoriser cette réussite (Hadji, 1991). C'est dans ce contexte très large du champ de l'innovation que se situe notre étude.

Si celle-ci répond à une demande sociale – offrir le meilleur aux enseignants et aux élèves pour favoriser la démocratisation scolaire – elle prend aussi place dans un débat initié par le politique. En décembre 2017, lors de la présentation des résultats de l'évaluation PIRLS 2016 (*Progress in International Reading Literacy Study*), Jean-Michel Blanquer, ministre de l'Éducation nationale, a annoncé que « tous les manuels ne se valent pas » et que « la liberté éditoriale ne doit pas être une anarchie éditoriale ». Un des axes de travail du conseil scientifique de l'éducation nationale (CSEN) consiste d'ailleurs à tenter de définir des critères « scientifiquement fondés » pour choisir les manuels. Cette question n'est pas récente (Séré & Bassy, 2010) et correspond à des besoins identifiés sur le terrain : « *Les enseignants manquent d'informations permettant d'évaluer la pertinence scientifique et pédagogique des manuels et leur adéquation aux programmes* » (Elalouf & Bois-Masson, 2014, p. 119). La volonté politique de recommander des ouvrages pour améliorer l'efficacité de l'École nous paraît être une intention louable. Toutefois, il ne faut pas oublier, qu'à quelques exceptions près, les éditeurs et les auteurs des manuels ne sont pas des chercheurs. En outre, les ouvrages que l'on trouve aujourd'hui sur le marché sont presque exclusivement des outils dont l'efficacité n'a pas été prouvée. Tout en respectant la liberté pédagogique des enseignants, nous croyons essentiel d'étayer scientifiquement leur choix en mettant nos connaissances théoriques et méthodologiques au service de l'évaluation d'outils existants. Notre ambition principale est d'apporter des connaissances scientifiques utiles aux décideurs, aux enseignants et aux formateurs.

1. Des inégalités sociales aux difficultés scolaires

Le troisième rapport de l'Observatoire des inégalités (2019) le rappelle : « *La France fait démarrer un système de compétition entre les enfants et introduit très vite des apprentissages formels* » (p. 65). Or, l'école accueille des enfants qui présentent des différences inter-individuelles importantes et tous ne démarrent pas leur scolarité avec les mêmes connaissances et compétences.

1.1. Des inégalités persistantes

Dans un contexte français marqué par les inégalités sociales de réussite (Andreu, Le Cam & Rocher, 2014; Benveniste, 2018 ; Bernardin, 2013 ; Felouzis, 2014 ; Felouzis, Fouquet-Chauprade & Charmillot, 2015), l'École affiche, depuis longtemps, un objectif de réussite pour tous (Charbonnier, 2018; MEN, circulaire de rentrée, 2019). Celui-ci est réaffirmé dans le texte de Loi de la refondation (2013) dans lequel on peut lire que « depuis près de vingt ans, notre école ne progresse plus. Le niveau global des compétences des élèves formés en France doit être amélioré pour parvenir à davantage de justice dans la réussite scolaire et pour pouvoir inscrire le pays sur une trajectoire de croissance structurelle forte dans une économie de la connaissance internationale ». Cette ambition légitime est fragilisée par deux constats : 1° la réussite reste fortement liée au milieu social d'origine et 2° en France, le statut économique, social et culturel des parents explique une plus grande part de la variation des scores des élèves qu'en moyenne dans l'ensemble des pays de l'OCDE (Daussin, Keskpaik, & Rocher, 2011; Rochex, 2013). Même si l'école primaire joue un rôle moins déterminant que l'enseignement secondaire dans l'accroissement des inégalités sociales (Duru-Bellat, 2006), les écarts ne s'y réduisent pas (Chapelle & Meuret, 2006 ; Daussin, Keskpaik & Rocher, 2011 ; Duru-Bellat, 2002). Ainsi, au CE2, le quart des élèves le moins favorisés¹ obtient un score de 58 sur 100 en français, contre 87 pour le quart issu du milieu social le plus favorisé (Observatoire des inégalités, 2019).

Depuis vingt ans, la quasi-totalité des enfants âgés de trois ans et plus est scolarisée (IGEN, 2017), mais l'allongement de la durée de scolarité n'a pas suffi pour réduire les inégalités (DEPP, 2013). Si les pratiques enseignantes jouent un rôle déterminant dans le devenir scolaire des enfants (Fayol, 2017; Felouzis, 2014), elles s'avèrent, souvent,

¹ Selon un indice de position sociale élaboré par le ministère de l'Éducation nationale (données 2017).

différenciatrices (Bonnéry, Crinon, & Simons, 2015; Joigneaux, 2009, 2013). Plusieurs auteurs ont cherché à expliquer ce phénomène. Pour certains (Bautier, 2018), les attentes de l'école restent parfois très éloignées des codes et des normes culturelles véhiculées dans les familles : « certains élèves, surtout lorsqu'ils sont issus de milieux populaires, sont mis en difficulté par les attendus scolaires ; ils ne participent pas des apprentissages prévus par l'enseignant et que d'autres élèves effectuent » (p. 23). Pour d'autres (Bautier & Goigoux, 2004 ; Bonnéry, 2009), les pratiques d'enseignement sont trop implicites. Partant, les élèves ne s'assignent pas tous les mêmes objectifs, ne se donnent pas tous les mêmes buts dans les tâches à accomplir. « Ainsi se creusent tôt les écarts d'apprentissage à l'insu de l'élève lui-même. Les moins performants des élèves réduisent souvent la visée de ces tâches à leur seule effectuation ; ils cherchent à s'en acquitter sans savoir qu'il s'agit d'en saisir la signification et les enjeux cognitifs, ou sans pouvoir le faire, sans même penser qu'il y en a » (Bautier, 2015, p. 23).

Ceux qui sont le moins outillés du point de vue socioculturel peuvent passer à côté des notions à acquérir tout en pensant accomplir correctement leur métier d'élève (Felouzis, 2014). Les supports utilisés par les enseignants peuvent également « renforcer ou atténuer les inégalités liées à une socialisation langagière et à un rapport aux savoirs plus ou moins proche des réquisits de l'école. C'est ainsi que les pédagogies qui, entre autres, explicitent peu les enjeux des tâches ou n'enseignent pas systématiquement ce que les élèves sont censés apprendre, se révèlent plus différenciatrices que des pédagogies plus explicites et plus systématiques, et redoublent ou prolongent les implicites ou les sollicitations différenciatrices des supports » (Bonnéry, Crinon, & Simons, 2015, p. 3). On observe que ces différences s'installent dès le début de la scolarité et perdurent.

1.2. L'influence des pratiques d'enseignement sur la production des inégalités

Depuis plus de vingt ans, les résultats des enquêtes nationales et internationales se rejoignent pour mettre en évidence un écart entre les performances des élèves de milieux sociaux contrastés. En outre, le pourcentage d'élèves en difficulté face à l'écrit a augmenté de manière significative : près d'un sur cinq est concerné en début de sixième et 20 % des élèves de 15 ans connaissent de grandes difficultés de maîtrise de la langue écrite (MEN, 2013). Ces évaluations soulignent l'incapacité du système éducatif de notre pays, malgré une volonté affichée et

répétée « d’offrir à tous les enfants des chances égales de réussite² » (MEN, 2008), de pallier cet écart ou au moins de le réduire.

1.2.1. Des résultats d’évaluation d’une stabilité préoccupante en fin de cycle 3

Depuis 2001, la performance française diminue légèrement à chaque évaluation PIRLS (*Program for International Student Assessment*, DEPP, 2017) qui mesure le niveau en lecture des écoliers de CM1 de cinquante pays de l’OCDE : 525 points en 2001, puis 520 en 2011, et 511 en 2016. À cette date, la France se situe toujours au-delà de la moyenne internationale (500 points) mais en deçà de la moyenne européenne (540 points) et de celle de l’OCDE (541 points). On observe de surcroît une surreprésentation des élèves français dans le groupe le plus faible. Ils constituent 39 % du quartile le plus faible (au lieu des 25 % attendus) selon le classement des scores des pays européens. En outre, 6 % des élèves français n’atteignent pas le niveau le plus élémentaire³ de PIRLS contre 4 % en Europe. Face à ces constats, nous nous interrogeons : pourquoi la France ne réussit-elle pas aussi bien que les autres pays, en particulier auprès des élèves les plus faibles ?

Différents facteurs expliquent ces résultats :

1° les pratiques d’enseignement. En lecture, ce sont les processus de compréhension les plus complexes « interpréter » et « apprécier »⁴ qui sont le moins bien réussis. Autrement dit, la compréhension de l’implicite pose plus de problème que le simple prélèvement d’informations : les élèves français déchiffrent bien mais comprennent mal. L’enquête PIRLS montre que l’enseignement explicite des stratégies est peu présent dans les classes françaises : les enseignants sont moins nombreux que leurs collègues européens à déclarer proposer des activités susceptibles de faire apprendre les stratégies et les compétences en compréhension de l’écrit. Quand 41 % des élèves français sont sollicités au moins une fois par semaine pour comparer ce qu’ils ont lu « à des faits qu’ils ont vécus », ce taux passe à 82 % pour les élèves européens.

2° la faiblesse de la formation. L’enquête PIRLS conclut que les résultats français ont notamment pour origine le peu de temps alloué à la formation continue des enseignants en

² Cette ambition de réussite présente dans les programmes de 2002 est réaffirmée dans ceux de 2008 puis de 2015.

³ Les élèves doivent atteindre un score d’au moins 400 pour le niveau bas.

⁴ Deux autres processus plus simples sont évalués : « prélever » et « inférer ».

lecture-compréhension. Ces derniers sont presque deux fois moins nombreux que leurs collègues européens à avoir suivi une formation récente (22 % contre 38 %). Ce constat a été récemment confirmé par l'enquête internationale TALIS (*Teachers and School Leaders as Lifelong Learners*, DEPP, 2019) : sur les 48 pays étudiés, la France est celui qui offre le moins de formation continue et celle-ci est, en outre, moins diversifiée qu'ailleurs.

Les résultats de PIRLS sont confirmés par ceux des évaluations nationales CEDRE (cycle d'évaluations disciplinaires réalisées sur échantillon) qui témoignent d'une stabilité des performances entre 2003 et 2015. À ces deux dates, seulement 60 % des élèves possèdent une maîtrise suffisante des compétences attendues en fin de scolarité primaire et l'écart de performances entre l'éducation prioritaire et l'enseignement public hors éducation prioritaire reste significatif en défaveur des premiers (DEPP, 2016). À l'instar de l'enquête PIRLS, l'évaluation CEDRE signale un déficit de connaissance des stratégies complexes utilisables par les élèves. Ces derniers réussissent davantage dans le prélèvement d'informations (compétence considérée comme étant de bas niveau) que dans l'analyse d'un document (compétence de haut niveau). Comment peut-on expliquer de tels résultats ? Premièrement, l'affirmation de la nécessité d'un enseignement explicite de la compréhension et de stratégies identifiées apparaît tardivement dans la prescription : pour le cycle 3, il est mis en avant par les programmes de 2015. Deuxièmement, les études permettant d'identifier les pratiques d'enseignement efficaces en compréhension se sont principalement développées dans les années 1990 (Duke & Pearson, 2002; Giasson, 1990; National Reading Panel, 2000). Ces travaux relativement récents méritent d'être diffusés plus largement pour aider les enseignants à introduire un enseignement explicitement centré sur les stratégies nécessaires à améliorer la qualité de leur compréhension des textes. Les recommandations rédigées par le jury de la conférence de consensus « Lire, écrire, comprendre » (CNETCO, 2016) étaient, à ce propos très explicites : « La compréhension doit faire l'objet d'un enseignement explicite. L'enseignant explicite les apprentissages visés (pourquoi), les tâches, les procédures et les stratégies (comment) et les apprentissages réalisés selon une scénarisation didactique et pédagogique anticipée, ajustable au fil du déroulement des activités et des réactions des élèves. Cet enseignement de la compréhension constitue une lutte contre les inégalités et leur reproduction et se concrétise par les actions suivantes : s'accorder sur les buts de la lecture, organiser et diriger des discussions auxquelles chaque élève est incité à participer, expliquer le vocabulaire et en vérifier l'acquisition, expliciter le type de texte, sa construction, sa fonction, sélectionner les informations importantes, identifier les liens référentiels et logiques, inciter à faire des inférences, aider au lien entre l'information nouvelle

et les connaissances préalables des élèves, revenir sur le texte (relecture) pour conforter la compréhension. Le rôle de l'enseignant est d'engager les élèves à réaliser ces tâches d'abord avec son aide puis de façon de plus en plus autonome. » (*idem*, 2016, p. 6).

Les données des enquêtes PIRLS et CEDRE, qui concernent des élèves respectivement âgés de 10 et 11 ans, montrent que les écarts d'efficacité sont importants à la fin de l'école élémentaire. Mais qu'en était-il au début ? En d'autres termes, l'école élémentaire a-t-elle créé la différence ? La recherche Lire-Écrire apporte une réponse à cette question grâce à l'évaluation des compétences en lecture de 2507 élèves entrant au CP en 2013 (Goigoux, 2016) (Goigoux *et al.*, 2016).

1.2.2. Des différences d'efficacité précoces

La recherche Lire-Écrire fait apparaître que le meilleur prédicteur des performances des élèves en lecture-compréhension à la fin du CE1 est le score qu'ils ont obtenu, au début du CP, à l'épreuve compréhension de textes narratifs entendus, c'est-à-dire lus à haute voix par un adulte (Goigoux, Cèbe, & Pironom, 2016). Elle a aussi permis d'établir que la qualité de la compréhension est fortement tributaire de l'appartenance sociale des enfants, bien plus que l'écriture ou la maîtrise du code alphabétique. En effet, à l'entrée au CP, la part de variance expliquée par l'origine sociale est de 8 % pour le code, 5 % pour l'écriture, quand elle s'élève à 25 % pour la compréhension de textes entendus.

L'épreuve de compréhension comportait deux sous-tests : compréhension de vingt-cinq phrases et compréhension de trois textes. On trouvera, dans le Tableau 1, les résultats des élèves en fonction de leur milieu social.

Tableau 1. Performances des élèves en compréhension de phrases et de textes en fonction du milieu social à l'entrée au CP

Milieu social	Compréhension de phrases : score sur 25 (et % de réussite)	Compréhension de textes entendus : score sur 15 (et % de réussite)
Favorisé	22,3 (90,0 %)	9,4 (62,6 %)
Intermédiaire	21,1 (84,4 %)	7,9 (52,6 %)
Défavorisé	19,6 (80,0 %)	6,4 (42,6 %)
Très défavorisé	17,9 (71,6 %)	5,5 (36,6 %)

Goigoux *et al.* (2016)

Les contrastes entre les élèves favorisés et très défavorisés sont plus forts en compréhension de textes qu'en compréhension de phrases. Pourquoi ? Ces deux épreuves ne font pas appel aux mêmes compétences qui sont inégalement sensibles aux pratiques sociales :

le test de compréhension de textes mobilise beaucoup de connaissances lexicales, textuelles et encyclopédiques alors que celui de compréhension de phrases très simples sur le plan lexical (adapté du test standardisé E.Co.S.Se) ne mesure que la maîtrise syntaxique des élèves.

Nous reviendrons sur ce point dans le chapitre 2 pour tenter d'expliquer l'origine de ces différences.

1.2.3. Les évaluations nationales CP

En septembre 2018, le ministère de l'Éducation nationale a demandé aux enseignants d'évaluer l'ensemble des élèves scolarisés dans les deux premières années du cycle des apprentissages fondamentaux (cycle 2). Le dispositif a permis de recueillir les réponses de plus de 790 000 élèves de cours préparatoire et de 780 000 élèves de cours élémentaire 1^{ère} année, répartis dans près de 31 000 écoles publiques et privées sous contrat en France métropolitaine, dans les DOM, la Polynésie Française et Saint-Pierre-et-Miquelon.

Nous allons en analyser les principaux résultats dans le domaine qui nous intéresse, soit la compréhension de textes écrits. En début de CP et de CE1, les épreuves proposées portaient sur la compréhension d'unités écrites lues à haute voix par le professeur (« compréhension entendue ») : mots, phrases et textes. Au milieu du CP, elles ne portaient plus que sur compréhension de phrases.

On trouve, dans le Tableau 2, les résultats des élèves en fonction 1° des seuils de réussite définis par le ministère et 2° du secteur scolaire (hors REP, REP ou REP+). Deux seuils de réussite, propres à chaque série d'exercices, ont été déterminés : un premier seuil en deçà duquel on peut considérer que les élèves présentent de graves difficultés et ont donc besoin d'un accompagnement spécifique (seuil 1) et un second pour ceux dont les acquis sont dits « fragiles » (seuil 2).

Tableau 2. Répartition des élèves dans les groupes en compréhension d'écrits oralisés au début du CP et à mi-CP selon le secteur de scolarisation (en %)

		Début CP (septembre 2018)			Mi-CP (janvier 2019)		
		Public Hors REP	REP	REP+	Public Hors REP	REP	REP+
Comprendre des mots entendus /15	Groupe sous le seuil 1 (élèves à besoins)	4,4	11,6	16,7			
	Groupe entre les seuils 1 et 2 (élèves fragiles)	18,7	33,7	38,8			
	Groupe au-dessus du seuil 2	76,9	54,7	44,4			
Comprendre des phrases entendues /14	Groupe sous le seuil 1 (élèves à besoins)	2,8	6,5	9,8	2,3	5,5	8,8
	Groupe entre les seuils 1 et 2 (élèves fragiles)	12,9	22,4	26,9	11,7	21,9	27,4
	Groupe au-dessus du seuil 2	84,3	71,1	63,4	86,0	72,6	63,9
Comprendre des textes entendus /18	Groupe sous le seuil 1 (élèves à besoins)	2,5	6,1	9,2			
	Groupe entre les seuils 1 et 2 (élèves fragiles)	5,9	11,5	14			
	Groupe au-dessus du seuil 2	91,6	82,3	76,9			

Source MENJ-DEPP, Note d'information, n° 19.13 et n° 19.15 © DEPP

Pour les trois épreuves dites de « compréhension entendue », on observe, qu'à l'entrée au CP, plus des trois quarts des élèves hors REP n'ont pas de difficulté : 76,9 % des élèves du secteur public présentent des acquis au-dessus du seuil 2 pour les mots, 84,3 % pour les phrases et 91,6 % pour les textes. Mais il n'en va pas de même en Éducation prioritaire.

Quand on s'intéresse aux scores selon les seuils de réussite et le secteur de scolarisation, on constate que 4,4 % des élèves du secteur public scolarisés hors REP sont sous le seuil 1 en compréhension de mots entendus, alors qu'ils sont 11,6 % en REP et 16,7 % en REP+. En compréhension de phrases entendues, les élèves identifiés comme étant « à besoins » ou « fragiles » (entre le seuil 1 et 2) sont trois fois plus nombreux en REP+ (9,8 %) qu'hors REP (2,8 %). Enfin, en compréhension de textes entendus, les élèves de REP+ sont quatre fois plus

nombreux qu'hors REP dans le groupe sous le seuil 1 (9,2 %) et deux fois plus nombreux entre les seuils 1 et 2 (6,1 %).

Ces résultats confortent des constats déjà bien établis selon lesquels les élèves de milieux populaires n'entrent pas à l'école élémentaire avec les mêmes habiletés que les autres. Ils nous renseignent aussi sur leurs besoins d'enseignement.

Pour la population globale (public et privé), les résultats obtenus en compréhension de textes entendus nous questionnent. En effet, la proportion d'élèves jugés sans difficulté semble anormalement élevée (90,1 %) en comparaison des résultats observés dans la recherche Lire-Écrire ou celle d'Écalle, Gomes, Auphan, Cros, et Magnan (2019) et au regard des difficultés inhérentes à l'épreuve « compréhension de textes entendus⁵ ». Rappelons que, dans la recherche Lire-Écrire, les élèves de milieux très défavorisés ne réussissaient que 36,6 % des items dans une épreuve sensiblement identique et 67,3 % dans celle d'Écalle *et al.*, plus simple. Comment expliquer que le ministère de l'Éducation nationale n'identifie que 10 % d'écoliers « en difficulté » en compréhension de textes entendus, soit trois fois moins qu'en lexique ?

Tout semble dépendre de la définition des seuils de réussite retenue par la DEPP comme le révèle le Tableau 3 qui présente ceux qui ont été arrêtés pour chaque épreuve et le pourcentage des élèves les ayant atteints.

Tableau 3. Définition des seuils pour chaque épreuve et pourcentage d'élèves les ayant atteints au début du CP

Compétences	Tests	Score sur	Seuil 1 « Besoin »	% d'élèves	Seuil 2 « Fragile »	% d'élèves
Compréhension Orale	Comprendre des mots lus	15	≤ 6	5,7	≤ 10	20,9
	Comprendre des phrases lues	14	≤ 4	3,5	≤ 10	14,4
	Comprendre des textes lus	18	≤ 4	3,2	≤ 7	6,7

Source MENJ-DEPP

Les taux de réussite retenus sont très différents d'une épreuve à l'autre : pour les mots, le seuil 2 « fragile » est fixé à 10/15 (soit les 2/3) et pour les phrases à 10/14 (un peu plus des 2/3) contre 7/18 (à peine plus d'1/3) pour les textes lus. Si on retenait aussi le seuil de deux tiers

⁵ Le lecteur intéressé trouvera les épreuves sur le site éducol : https://cache.media.eduscol.education.fr/file/CP/62/8/GUIDE_CP_1015628.pdf.

pour l'épreuve de compréhension de textes entendus, on modifierait considérablement la proportion d'élèves « en difficulté ».

Goigoux (2019) suggère que ces seuils ont été délibérément fixés très bas par le ministère pour ne pas orienter les efforts de l'école maternelle vers l'enseignement de la compréhension de textes entendus, mais vers celui du lexique et de la phonologie. La circulaire de rentrée (MEN, 2019) confirme cette analyse en donnant priorité à « un enseignement structuré du vocabulaire oral » et à celui de la phonologie : « La connaissance et la manipulation des unités sonores de la langue française font l'objet d'un enseignement progressif. Dès la petite section, la construction d'une conscience phonologique est régulièrement travaillée. Elle se structure jusqu'à la grande section par des activités appropriées. La connaissance du nom des lettres et du son qu'elles produisent est progressivement enseignée ». En outre, à la rentrée 2019, le ministère a diffusé deux nouveaux guides portant sur l'enseignement du vocabulaire et de la phonologie à l'école maternelle.

À la mi-CP, dans le domaine de la compréhension, seule l'épreuve évaluant les phrases entendues a été maintenue. On trouvera, dans le Tableau 4, une information intéressante pour notre propos.

Tableau 4. Écarts de performances entre élèves scolarisés dans le secteur public hors EP et élèves scolarisés en EP (différences de proportions et odds ratio⁶)

Domaine	Début CP				Mi-CP			
	Proportion d'élèves présentant une maîtrise satisfaisante (au-dessus du seuil 2)		Écart	Odds ratio	Proportion d'élèves présentant une maîtrise satisfaisante (au-dessus du seuil 2)		Écart	Odds ratio
	Public hors EP	EP (REP, REP+)			Public hors EP	EP (REP, REP+)		
Comparer des nombres	77,6	64,9	12,8	1,9	87,7	80,1	7,6	1,8
Comprendre des phrases à l'oral	84,3	68,2	16,1	2,5 ⁷	86,0	69,3	16,7	2,7
Connaître le nom des lettres et le son qu'elles produisent	77,8	68,8	9,0	1,6	83,1	79,8	3,4	1,2
Écrire des nombres sous la dictée	83,6	76,8	6,7	1,5	76,4	72,3	4,1	1,2
Manipuler des phonèmes	79,5	67,4	12,1	1,9	87,4	82,2	5,2	1,5
Placer un nombre sur une ligne numérique	45,8	31,8	14,0	1,8	66,4	57,0	9,4	1,5
Résoudre des problèmes	62,0	41,8	20,1	2,3	55,2	42,4	12,8	1,7

Source MENJ-DEPP, Note d'information, n° 19.15 © DEPP

La lecture du tableau fait apparaître que les écarts de performances observés en début d'année entre les élèves REP et hors REP diminuent dans toutes les épreuves sauf dans celle qui mesure la compréhension de phrases. On voit ici que les pratiques d'enseignement centrées sur la maîtrise du code alphabétique ont bien permis de réduire les différences excepté en compréhension de phrases, épreuve dans laquelle la différence augmente faute sans doute d'un enseignement adapté.

Ce déficit d'enseignement de la compréhension a-t-il des effets un an plus tard ? Autrement dit, les pratiques d'enseignement mises en œuvre dans la deuxième partie du CP ont-elles permis de réduire les écarts en compréhension ? On trouvera la réponse, négative, dans le Tableau 5.

⁶ Le « rapport de chances relatives » ou *odds ratio* apparaît comme un indicateur pertinent pour obtenir des résultats plus fins (Rochex, 2013).

⁷ En début de CP, faire partie du groupe d'élèves présentant une maîtrise satisfaisante de la compréhension de phrases est un événement qui a 2,5 fois plus de chance de se produire que la situation contraire pour les élèves scolarisés dans le secteur public hors éducation prioritaire que pour ceux scolarisés en éducation prioritaire. Cette valeur passe à 2,7 à mi-CP.

Tableau 5. Répartition des élèves dans les groupes en compréhension d'écrits oralisés au début du CE1 selon le secteur de scolarisation (en %)

		Début CE1 (septembre 2018)		
		Public Hors REP	REP	REP+
Comprendre des mots entendus	Groupe sous le seuil 1 (élèves à besoins)	3,7	12,5	18,8
	Groupe entre les seuils 1 et 2 (élèves fragiles)	14,9	27,1	30,6
	Groupe au-dessus du seuil 2	81,3	60,4	50,6
Comprendre des phrases entendues	Groupe sous le seuil 1 (élèves à besoins)	2,5	7,8	11,7
	Groupe entre les seuils 1 et 2 (élèves fragiles)	17,4	28,7	32,7
	Groupe au-dessus du seuil 2	80,1	63,6	55,6
Répondre à des questions lues par l'enseignant	Groupe sous le seuil 1 (élèves à besoins)	4,8	10,1	12,9
	Groupe entre les seuils 1 et 2 (élèves fragiles)	12,1	19,6	22,3
	Groupe au-dessus du seuil 2	83,1	70,3	64,8

Source MENJ-DEPP, Note d'information, n°19.14 © DEPP

Les différences observées au début du CP se retrouvent inchangées (voire amplifiées) à l'entrée du CE1 dans les trois épreuves de compréhension : si 81,3 % des élèves du secteur public hors REP sont en réussite (au-dessus du seuil 2) pour la compréhension de mots, ils ne sont que 60,4 % en REP et 50,6 % en REP+. Il y a cinq fois plus d'élèves sous le seuil 1 en REP+ (18,84 %) qu'hors éducation prioritaire (3,74 %).

Ajoutons que, parmi les dix-neuf compétences évaluées⁸ au CE1, les écarts les plus importants se situent dans le domaine de la compréhension du langage : comprendre des mots ou des phrases entendus et répondre à des questions lues par l'enseignant. Nous n'ignorons pas que les données ne portent pas sur les mêmes élèves et que la manière de fixer les seuils relève plus d'une option politique que scientifique. Mais, quoi qu'il en soit, elles prouvent, s'il en était encore besoin, que l'enseignement dispensé au CP n'a pas permis de réduire les écarts.

En résumé, quels que soient l'épreuve et le niveau d'enseignement, les élèves de milieux favorisés sont toujours plus performants que leurs congénères. Force est de constater que les inégalités sociales, à l'école, se portent toujours bien et que les pratiques mises en œuvre par les professeurs des écoles en compréhension en lecture ne suffisent pas pour compenser les

⁸ Dix compétences sont évaluées en mathématiques et neuf en français.

écarts d'efficience. Ces différences plaident en faveur d'un enseignement de la compréhension à l'école maternelle, particulièrement nécessaire pour les élèves de milieux populaires. Nous interrogeons alors les conditions d'enseignement qui peuvent concourir à la réduction des inégalités sociales. En d'autres termes, l'École peut-elle s'améliorer et si oui, comment ?

2. Améliorer l'école, une nécessité

Pour améliorer l'école et faire face aux inégalités, deux voies s'offrent à l'Institution scolaire. La première, pilotée à l'échelle nationale, dépend des propositions, des orientations, des décisions prises par les responsables des politiques publiques. Les transformations sont, par exemple, visibles en termes de structures, d'équipements, de budget, de postes. Récemment, le dédoublement des classes de CP et CE1 concrétise, par exemple, une volonté de lutte contre l'échec scolaire et la réduction des inégalités. La seconde est conduite à une échelle plus locale : les changements envisagés portent sur les pratiques professionnelles. Notre propos se situe à ce deuxième niveau : nous pensons qu'il est possible d'influencer durablement les pratiques des enseignants au profit des apprentissages des élèves qui ont le plus besoin de l'école pour apprendre. Se pose alors la question des modalités de changements : « quels sont les leviers pour améliorer les apprentissages des élèves et comment faire pour que les professeurs intègrent les avancées dans leur pratique ? » (Galand & Janosz, 2019). Ces auteurs identifient trois courants de recherche qui abordent cette question : le premier porte sur la formation et le développement professionnel des enseignants, le deuxième sur les effets de la diffusion des pratiques validées par des recherches expérimentales (*evidence-based*), le troisième sur l'impact des réformes politiques en éducation. Les trois partagent peu ou prou les mêmes constats d'impuissance : 1° les actions de formation visant le développement professionnel des enseignants n'ont qu'un effet marginal sur les pratiques de classe et les acquis des élèves, 2° les interventions réputées efficaces peinent à s'inscrire dans la durée et à se diffuser à large échelle, 3° les réformes politiques échouent le plus souvent à faire évoluer les pratiques d'enseignement et la réussite scolaire des élèves.

Refusant de nous résigner à l'impuissance, nous explorons une quatrième voie et formulons l'hypothèse que la conception et la diffusion d'outils didactiques novateurs offrent des possibilités d'amélioration.

2.1. Améliorer par la formation

La formation continue est en partie dispensée par des enseignants-chercheurs qui apportent des connaissances et présentent des résultats de recherches issues de différents champs : sciences de l'éducation, didactique, psychologie, sciences cognitives. Bon nombre d'entre eux espèrent que leur présentation et les recommandations qui en découlent conduiront les professeurs à modifier leurs conceptions et à transformer en conséquence, par souci de rationalité et sans aide particulière, leurs manières de faire. Ce faisant, ils leur laissent le soin de traduire les connaissances issues de la recherche en actions concrètes. C'est la thèse que défend Tricot (2017) : les chercheurs produisent des connaissances et il revient aux enseignants de concevoir les solutions. Cette tâche d'« ingénierie » (Tricot, 2018) incomberait aux professeurs qui devraient combiner leurs savoirs d'expérience aux savoirs scientifiques acquis lors des formations.

Malheureusement, il ne suffit pas que les enseignants soient informés ni même convaincus de la pertinence des résultats produits par la recherche pour qu'ils transforment leurs pratiques (Hamre et al., 2012; Prost, 2001). La quête de solutions qui intègrent les résultats scientifiques reste complexe et les obstacles sont multiples. Entre la présentation des résultats (ou des principes d'action jugés pertinents) et leur mise en œuvre concrète dans la classe, il y a souvent un gouffre que peu de chercheurs aident à franchir, laissant aux enseignants (ou aux formateurs) le soin de réaliser eux-mêmes le travail de transposition, d'opérationnalisation et d'ajustement à leur contexte (Coburn, Honig, & Stein, 2009). C'est sans doute pourquoi les actions de formation visant le développement professionnel des enseignants n'ont qu'un effet marginal sur les pratiques de classe et les acquis des élèves (Kennedy, 2016).

En France, la formation continue est en outre limitée par la faiblesse des moyens de remplacement qui lui sont consacrés. Majoritairement assurées par des conseillers pédagogiques, les actions de formation sont, pour l'essentiel, proposées dans le cadre des dix-huit heures d'animation pédagogique obligatoires organisées par chaque circonscription. Les conclusions du rapport des Inspecteurs Généraux (IGEN & IGAENR, 2017) remettent en cause cette modalité qui paraît trop éloignée des besoins des enseignants du premier degré : « le constat est largement partagé (par les enseignants et les responsables de formation) [...] des formations inadaptées aux attentes des professeurs des écoles et peu efficaces au regard des attentes institutionnelles ». Les rédacteurs ajoutent que « l'adaptation de la formation aux besoins des bénéficiaires est le parent pauvre du dispositif. Majoritairement utilisée comme

outils de pilotage des réformes ou comme réponse à des priorités nationales, la formation n'est pas perçue par le professeur comme construite pour lui » (p. 2). La formation continue ne semble pas assez porter sur les difficultés d'apprentissage des élèves et les problèmes ordinaires d'enseignement des professionnels. Un défi majeur s'impose : amener le travail réel en formation (Altet, Desjardins, Etienne, Paquay, & Perrenoud, 2014).

Les auteurs du rapport ajoutent que « l'apport de la recherche est, quant à lui, invisible en raison notamment de la faible propension des chercheurs à articuler les apports scientifiques aux enjeux professionnels des enseignants du premier degré. [...] Les écarts avec les attentes des enseignants se situent davantage au niveau des contenus, considérés comme insuffisamment articulés aux réalités des pratiques professionnelles, et sous-estimant les difficultés de mise en œuvre. Les professeurs sont en attente de dispositifs de formation variés, dès lors qu'ils répondent à leurs besoins en termes de compétences pour faire la classe ou d'évolution professionnelle. » (p. 7).

Loin de remettre en cause le caractère indispensable de la formation, ces constats interrogent les formes qu'elle prend.

2.1.1. Améliorer par l'utilisation d'outils

Dans le domaine des recherches portant sur l'amélioration de l'enseignement, il existe consensus international : l'utilisation de nouveaux outils ou le recours à des dispositifs innovants apparaît comme un vecteur puissant de développement professionnel (Fishman, Penuel, Allen, Cheng, & Sabelli, 2013). Mais, pour qu'ils produisent les effets recherchés, il faut mener un véritable travail de transposition des savoirs issus de la recherche en savoirs pour l'action, activité que les chercheurs ne peuvent pas sous-traiter aux enseignants, mais qu'ils doivent absolument réaliser avec eux (Goigoux & Cèbe, 2009 ; McKenney & Reeves, 2014 ; LeMahieu, Nordstrum, & Potvin, 2017 ; Snow, 2015 ; cf. Class & Schneider, 2013, pour une présentation en français du *Design-Based Implementation Research*).

Cette thèse, qui est la nôtre, devient aussi progressivement celle des tenants de l'approche EBP (*evidence-based practices*) qui recommandent désormais aux chercheurs de travailler en étroite collaboration avec les enseignants dès le début du processus de conception (Klingner, Boardman, & McMaster, 2013) pour répondre à leurs besoins et leurs préoccupations professionnelles tout en s'ajustant à leur contexte d'exercice (Bressoux, 2017; Bryk, 2015; Cobb & Smith, 2008; Durlak & DuPre, 2008). Elle se rapproche également de celle que les

américains appellent PEER (*Practice Embedded Educational Research*) précisément parce qu'elle est « intégrée à la pratique » (Goigoux, 2017). Celle-ci partage les quatre principes fondateurs du SERP, *Strategic Education Research Partnership* (Donovan, Snow, & Daro, 2013; National Research Council, 2003) : 1° reconnaître l'organisation systémique de toute réforme scolaire, 2° construire un partenariat structuré et soutenu entre les praticiens et les chercheurs, 3° partir des préoccupations des praticiens pour déterminer les problématiques de recherche, 4° étudier attentivement la façon dont les innovations sont conduites et traiter les variations de leur mise en œuvre comme une source majeure d'information.

Cette approche rejoint les constats du rapport déjà cité rédigé par Cristofari, Le Pivert et Lussiana intitulé « Évaluation de la politique publique sur la formation continue des professeurs du premier degré » (IGEN & IGEANR, 2017). On y apprend que les enseignants jugent les thèmes de formation intéressants mais les contenus insatisfaisants. Les dispositifs ne répondent pas aux problèmes ordinaires auxquels ils ont à faire face et les formateurs sont, selon eux, trop éloignés des pratiques professionnelles existantes.

Nous réalisons notre thèse dans un laboratoire⁹ qui intervient dans le champ de la didactique définie comme une science de l'intervention et de la conception (d'outils, de dispositifs d'enseignement et de formation, ou de curriculum) visant l'amélioration de la qualité du travail enseignant et des apprentissages scolaires. Les chercheurs de ce laboratoire postulent qu'agir directement sur les pratiques des enseignants est plus efficace pour modifier leurs conceptions et leurs connaissances que l'inverse (Goigoux & Cèbe, 2009).

Ils soulignent le rôle que peut jouer la prise en main d'outils didactiques novateurs dans le développement professionnel des enseignants notamment parce qu'ils proposent des scripts d'action précis et détaillés, cohérents avec les données scientifiques disponibles (Cèbe & Goigoux, 2018), critères généralement reconnus comme décisifs (März, 2014; Rowan & Miller, 2007). Il s'agit « de contribuer à accroître le pouvoir d'agir des professeurs en leur proposant de nouvelles ressources qui associent des techniques pédagogiques innovantes minutieusement décrites et l'élucidation de leurs fondements théoriques » (Cèbe & Goigoux, 2018, p. 77). Comme Bryk, Gomez, Grunow & LeMahieu (2015), ils pensent que le pouvoir d'agir des professeurs est dépendant des outils forgés dans leur sphère professionnelle. Selon eux, l'usage d'outils conçus pour améliorer les apprentissages scolaires peut favoriser leur développement

⁹ Laboratoire ACTé, *Activité Connaissance Transmission éducation*, de l'université Clermont-Auvergne

professionnel. Non seulement les outils sont des prolongements et des amplificateurs de l'activité des enseignants, de véritables « extensions » d'eux-mêmes au sens de McLuhan (1964), mais leur maniement produit en retour un effet sur l'utilisateur. C'est pourquoi, dans la perspective ouverte par le concept d'acte instrumental (Vygotski, 1925/1994), ils étudient comment l'évolution des outils influence l'activité des professeurs. L'effet formatif est perceptible en examinant, par exemple, comment apparaissent, disparaissent ou se recomposent les tâches didactiques choisies par les enseignants, comment elles sont ordonnées dans le temps (planification) et mises en œuvre en classe (régulation). Notre projet de thèse vise précisément à mettre à l'épreuve ces hypothèses et ces postulats en évaluant les effets d'une instrumentation.

Cette approche de la formation par l'outil vise à aider les enseignants à mieux comprendre et surmonter les difficultés des élèves en les incitant à recourir à de nouvelles manières de faire. Mais qu'entend-on précisément par outil ? Il nous semble nécessaire de définir ce terme.

2.1.2. De l'outil didactique à l'outil professionnel

En 2000, la revue *Repères* (n° 22) a consacré un numéro aux « outils d'enseignement du français » et, en 2014, la revue *Recherches* (n° 60) en a proposé un sur le même sujet. Dans ces publications, le terme outil désigne alternativement des modalités de travail, des objets d'apprentissages, des médias utilisées dans une démarche didactique... mais aucune définition unifiée n'apparaît. L'innovation que nous étudions et qui vise l'amélioration des compétences langagières des jeunes élèves est basée sur l'introduction d'un scénario didactique innovant dont la mise en œuvre est décrite avec force détails. Ce document écrit, accompagné de ressources multimédia, est un outil de travail pour l'enseignant destiné à faciliter son action au service des apprentissages des élèves (Plane & Schneuwly, 2000). Nous utilisons le mot outil, terme couramment employé par les enseignants, pour désigner ce « dispositif matériel » ou « artefact » servant les situations d'enseignement et d'apprentissage (Reuter, Cohen-Azria, Daunay, Delcambre, & Lahanier-Reuter, 2013, p. 155).

L'outil didactique est donc un artefact matériel « inscrit dans un usage, dans un rapport instrumental à l'action du sujet, en tant que moyen de celle-ci » (Rabardel, 1995, p. 59). Rabardel parle d'instrument pour décrire l'entité mixte composée de l'artefact (que nous appelons ici outil didactique) et des schèmes d'utilisation associés. Dans son cadre théorique, l'artefact est donc la partie neutre et universelle de l'instrument, indépendante de l'usage qu'en fait l'utilisateur : « chaque artefact a été conçu pour produire une classe d'effets, et sa mise en

œuvre, dans des conditions prévues par les concepteurs, permet d'actualiser ces efforts. Autrement dit, à chaque artefact correspondent des possibilités de transformations des objets de l'activité qui ont été anticipées, délibérément recherchées et qui sont susceptibles de s'actualiser dans l'usage. En ce sens, l'artefact (qu'il soit matériel ou non) concrétise une solution à un problème ou à une classe de problèmes socialement posés » (idem, p. 59). Dans notre travail, nous mobiliserons la théorie instrumentale de Rabardel pour examiner les schèmes d'utilisation des outils et pour analyser les effets des artefacts sur les processus d'enseignement et d'apprentissage.

L'outil *Narramus* est un artefact conçu par des chercheurs et des enseignants que d'autres enseignants – non concepteurs – peuvent transformer dans l'usage en un instrument d'enseignement de la compréhension. En cela, *Narramus* peut être envisagé comme un outil professionnel, c'est-à-dire utile dans la pratique des enseignants. Notre objectif sera de comprendre en quoi et comment l'introduction de *Narramus* est susceptible d'outiller l'enseignement et, à travers lui, d'influencer l'apprentissage de la compréhension de textes narratifs à l'école maternelle. Pour Goigoux (2007), l'activité enseignante est médiatisée par l'utilisation d'instruments didactiques et le développement de cette activité se réalise par l'appropriation et l'usage d'artefacts. Moyen d'action en situation, l'artefact devient donc instrument (Rabardel, 2005). Lorsqu'un enseignant utilise un nouvel outil didactique, cela transforme son action mais, en retour, l'action du professeur fait subir des transformations à l'outil. C'est ce double mouvement que Rabardel (1995) appelle « genèse instrumentale » qui est composée d'une « instrumentation », c'est-à-dire la modification de l'action de l'enseignant sous l'influence de l'outil et d'une « instrumentalisation », c'est-à-dire la transformation de l'outil lorsqu'il est utilisé par l'enseignant.

Les didacticiens partagent cette orientation et montrent que les interactions entre l'enseignant et l'outil génèrent des transformations (Kervyn & Rebière, 2014) aussi bien du côté du sujet et de son activité que du côté de l'outil. Utiliser un outil peut amener les enseignants à procéder autrement (Gaglio, 2011) et donc à modifier leurs pratiques habituelles.

3. Améliorer par l'innovation

Impulsées localement ou à plus grande échelle, prescrites ou simplement incitées par l'Institution, les expérimentations et les innovations prolifèrent dans le système éducatif français et sont mises en œuvre par des enseignants qui testent de nouveaux outils, de nouvelles

méthodes, des dispositifs d'enseignement originaux ou des contenus disciplinaires revisités (Alter, 2010; Cros, 2004). L'innovation transforme (Bronckart & Thurler, 2004 ; Cros, 1999) – voire bouscule – les conceptions, les valeurs, les pratiques et les relations entre les acteurs (Choplin *et al.*, 2007). Par l'usage d'un nouveau dispositif, la mise en place d'une nouvelle manière de faire, l'intégration d'un matériel innovant dans sa classe, l'enseignant est amené à transformer sa pratique. L'objectif premier et immédiat de la mise en œuvre d'une innovation est d'améliorer les apprentissages des élèves mais une visée à plus long terme participe à la « promotion d'idées novatrices destinées à changer l'École et peut-être aussi la société » (Reuter, Condette, & Boulanger, 2013, p. 3).

3.1. Le concept d'innovation

Nous allons définir ce que nous entendons par innovation en nous intéressant aux origines du concept. Le terme vient du latin « *innovare* » qui signifie « revenir à, renouveler ». Le mot *innovare* est lui-même composé du verbe *novare* (racine : *novus*) qui veut dire transformer, changer et du préfixe *in-*, qui signale un mouvement vers l'intérieur. Au Moyen-Âge, ce concept est d'abord utilisé dans un contexte strictement juridique : « introduire quelque chose de nouveau dans une chose établie¹⁰ ». Jusqu'au XII^e siècle, le mot désignait ce qui était jeune. Au milieu du XVI^e siècle, la définition évolue pour désigner ce qui est singulier, inattendu. C'est alors que le mot *innover* signifie « faire preuve d'inventivité, créer des choses nouvelles » (Rey, 1998), sens qu'il conserve en partie aujourd'hui.

L'innovation se démarque toutefois du concept d'invention, venant du verbe *inventer*, lui-même issu du latin « *invenire* » : « action de trouver, découvrir ». L'invention désigne une potentialité, un élément mis à disposition alors que l'innovation consiste en l'implantation effective et durable d'inventions dans un milieu social (Gaglio, 2011, p. 11). L'innovation se distingue ainsi de l'invention par son caractère opérationnel et sa mise en œuvre concrète. Contrairement à l'invention qui consiste à trouver une idée et sa possible concrétisation, l'innovation inclut le processus de diffusion auprès d'un public. Pour ce travail doctoral, nous retiendrons que si l'invention désigne une action envisagée comme individuelle, l'innovation, elle, comprend une dimension sociale.

¹⁰ Dictionnaire Le Petit Robert de la langue française, Édition 2015.

Dès son apparition, l'innovation revêt une dimension de changement social : elle consiste en l'introduction d'un changement par les humains dans un ordre établi (institution, système politique, etc.). Sous l'Ancien régime, elle était d'ailleurs vue comme un danger pour la société, un élément néfaste susceptible de déséquilibrer les institutions (Cros, 2004) et l'équilibre social. Autrement dit, elle s'inscrit dans un milieu qu'il est nécessaire de connaître pour favoriser sa diffusion. C'est avec l'essor du monde industriel et du capitalisme au XXI^e siècle que l'innovation reprend une connotation positive (Cros, 1997). Aujourd'hui, très valorisées voire plébiscitées, les innovations sont nombreuses dans le domaine de l'éducation.

3.2. Innover dans l'enseignement

Les travaux sur l'innovation dans l'enseignement soulignent l'extrême complexité du domaine (Alter, 2002, 2010; Cros, 2004; Monetti, 2002). En raison des nombreux champs de référence possibles (de l'innovation technique à l'économie en passant par la pédagogie) et des points de vue envisagés (enjeu, produit ou diffusion de l'innovation), les définitions divergent (Cros, 1997; Gaglio, 2011; West & Altink, 1996). Malgré cette variété, West et Altink ont identifié quatre thèmes communs récurrents : « la nouveauté (absolue ou nouvelle dans le contexte d'adoption), une composante d'application (c'est-à-dire pas seulement les idées mais leur application), la volonté d'améliorer (qui distingue les innovations du changement fortuit ou du sabotage volontaire) et une référence au processus de l'innovation¹¹ » (pp. 4-5). Nous présentons ici quelques définitions utiles à notre propos, elles comportent toutes une ou plusieurs de ces caractéristiques.

Selon la banque de données bibliographiques des documents écrits sur l'innovation en éducation et en formation Nova, l'innovation est envisagée comme « un processus qui a pour intention une action de changement et pour moyen l'introduction d'un élément ou d'un système dans un contexte déjà structuré » (Cros, 2003). Elle participe ainsi à la transformation et l'évolution des pratiques d'enseignement. Les définitions de Tricot (2017) et de Blignières-Legeraud (1998) lui ajoutent une dimension d'amélioration. Pour Tricot, l'innovation pédagogique correspond à l'introduction d'une nouvelle manière d'enseigner ; elle concerne les tâches réalisées par les enseignants et celles proposées aux élèves pour que ces derniers apprennent des connaissances au sein d'une organisation du temps, de l'espace et des relations. Cet auteur attribue trois finalités à l'innovation : l'apprentissage de connaissances, la mise en

¹¹ Traduction personnelle

œuvre d'un changement et la recherche d'une efficacité et d'une efficience toujours plus grandes. Blignières-Legeraud propose une définition plus complexe qui exprime aussi la visée de progrès et intègre l'idée d'un transfert des améliorations dans d'autres domaines ou environnements d'apprentissage. Pour cette auteure, l'innovation est une « production, assimilation ou exploitation de produits ou de pratiques de formation, dans un processus pluridimensionnel entre partenaires qui constituent des améliorations ou des réponses/solutions réussies – tout en dépassant la pratique usuelle – aux problèmes/besoins spécifiques dans leur contexte (à la fois spatial et temporel) et qui sont susceptibles d'être transférées dans d'autres contextes » (p. 13). Pour ce travail, nous retiendrons une définition de l'innovation qui met au premier plan la visée d'amélioration des apprentissages des élèves.

Les innovations peuvent donc être porteuses de changements positifs et produire des progrès, elles participent ainsi à relever des défis majeurs auxquels l'École est confrontée. C'est notamment pour leur contribution à l'amélioration des apprentissages que les enseignants introduisent des innovations (Huberman, 1973). Pour agir différemment, les maîtres se fixent des objectifs de transformation à atteindre, acceptent de modifier leurs pratiques et intègrent de nouvelles manières (Cros, 2004). L'innovation répond ainsi à des besoins de changement (Cros, 1997; Fullan, 1996; West & Altink, 1996) et est envisagée – par la Société, l'Institution et les enseignants – comme une opportunité pour faire progresser l'École et améliorer les performances des élèves. Notre propos est, jusqu'à présent, resté très général. Il est temps de passer au particulier : dans quel cadre spécifique se situe l'innovation qui nous intéresse ?

3.3. Améliorer, innover et évaluer !

Notre recherche s'inscrit dans un domaine spécifique qui est la didactique du français. Nous nous interrogeons sur l'introduction d'une innovation qui vise à aider les maîtres dans l'enseignement de la compréhension de textes narratifs à l'école maternelle.

De manière générale, trois grandes étapes se dégagent dans la mise en œuvre des innovations : la conception, la validation et la diffusion. Si notre travail porte sur la question de la validation, les deux autres dimensions seront incluses dans l'élaboration de notre questionnaire, nous y reviendrons plus tard. L'utilisation de *Narramus* devrait amener les enseignants à modifier leurs pratiques antérieures. Nous supposons qu'en utilisant l'outil ces dernières seront infléchies mais nous ne sommes pas assurée que ces transformations s'accompagneront d'une amélioration c'est-à-dire d'un accroissement des compétences

langagières des élèves. En d'autres termes, le caractère innovant de l'outil ne garantit ni une amélioration des performances des élèves, ni sa supériorité par rapport aux dispositifs d'enseignement habituels (Snow, 2015). C'est pourquoi il est nécessaire d'en évaluer les effets.

En contexte français, le besoin d'évaluation est légitimé par le fait que le système éducatif est fortement clivant socialement (Bablet, 2015). Si on souhaite réduire les inégalités scolaires, il faut recueillir des informations sur la validité et l'efficacité des innovations mises en œuvre. Or, en contexte francophone, on compte très peu de recherches de ce type. Fait remarquable, il n'y a jamais eu d'évaluation menée sur un dispositif d'enseignement de la compréhension chez les jeunes élèves de la petite section à la grande section sur un large échantillon. Ce manque de recherche évaluative ne signifie pas, bien sûr, que les dispositifs existants sont inefficaces, mais il indique qu'il y a peu de données probantes qui aideraient à opter pour un programme donné. Si l'objectif est d'améliorer les performances des élèves, on ne peut se soustraire à l'évaluation des dispositifs. C'est pourquoi nous chercherons à savoir si l'outil *Narramus* est efficace, c'est-à-dire s'il permet d'accroître les compétences de compréhension chez les jeunes élèves mais aussi à le valider dans une perspective comparatiste. Est-ce que le fait de travailler avec l'outil *Narramus* permet d'apprendre mieux qu'avec un autre dispositif d'enseignement ? Nous nous attacherons au processus selon lequel l'innovation est intégrée dans la pratique (Huberman, 1973) et mesurerons les effets de la mise en œuvre sur les performances des élèves et le développement professionnel des enseignants. Notre ambition sera double : elle portera sur l'évaluation de l'innovation didactique et pédagogique, mais également sur sa possible diffusion, intégrant pleinement la dimension sociale (les enseignants, les élèves et les formateurs) véhiculée par le concept même d'innovation.

4. Évaluer une innovation

Les innovations en didactique du français se succèdent et font l'objet de nombreuses propositions basées sur des expériences ou des croyances légitimes mais qui ne sont pas validées par des études reposant sur des protocoles expérimentaux fiables. Cette faiblesse nous questionne au regard des ambitions de réussite et d'amélioration de l'École. Pour nous, innover ne suffit pas. Il est nécessaire d'examiner si les innovations produisent les effets escomptés sur les apprentissages. Malgré une connotation généralement positive, le caractère bénéfique de l'innovation reste incertain (Bédard & Béchar, 2009). Autrement dit, elle n'est pas synonyme d'amélioration systématique (Dupriez, 2015 ; Tricot, 2017), ce qui remet en question la

croissance selon laquelle l'innovation serait nécessairement efficace. Il convient alors de penser l'évaluation des innovations si l'on souhaite proposer une recherche en éducation capable de fournir aux enseignants, aux formateurs et aux décideurs de politique publique des éclairages sur les questions cruciales d'éducation (Rey, 2014).

4.1. Évaluer : un double enjeu

Contrairement à d'autres pays, les recherches évaluant l'efficacité des pratiques pédagogiques restent exceptionnelles en France (Labat *et al.*, 2013), notamment en didactique du Français. Dans ce champ, depuis le début des années 70, les chercheurs (Romian, 1979) ont conçu de très nombreux outils ou dispositifs d'enseignement innovants mais ils n'ont pas utilisé de démarches expérimentales ou quasi-expérimentales pour valider leurs propositions. Leurs études ont permis de mieux comprendre les processus d'apprentissage et d'enseignement mais elles sont restées exploratoires. Réalisées sur des petits effectifs, elles n'ont pas quantifié les progrès accomplis par les élèves, ni administré la preuve de la supériorité d'une technique sur une autre. On compte de nombreuses recherches de ce type en contexte français (Boiron, 2014; Brigaudiot, 1993, 2000; Brigaudiot & Ewald, 1990; Canut, Masson ..., & Leroy-Collombel, 2018; Vinel, 2014).

En janvier 2019, lors d'une conférence donnée à l'Institut des hautes études de l'éducation et de la formation, Maryse Bianco, membre du conseil scientifique de l'éducation nationale, a déploré l'absence de recherche en contexte scolaire portant sur l'efficacité des pratiques de lecture inspirées des pratiques familiales efficaces. Notre projet devrait combler ce manque : l'innovation *Narramus* est basée sur des lectures d'albums de la littérature de jeunesse et a été introduite dans des classes en école maternelle. Des connaissances sur le langage oral et la compréhension des jeunes élèves de milieux populaires constitueront des apports importants : les données obtenues sur l'acquisition et la mémorisation du lexique en contexte, la narration et la compréhension permettront de compléter ce que l'on connaît actuellement des pratiques basées sur la littérature de jeunesse.

Dans l'optique d'une utilisation sociale de la recherche en éducation, administrer la preuve s'avère primordial (Duru-Bellat & Mingat, 1998), au niveau pédagogique et politique (Suchaut, 2003). Dans son ouvrage *L'École de demain* (2016), Jean-Michel Blanquer soutient qu'il est nécessaire de « diffuser des techniques pédagogiques qui ont fait leurs preuves, inspirées directement par la recherche et fondées sur les meilleures expériences internationales

(p. 30) ». Avant de diffuser de nouveaux outils didactiques, il conviendrait donc d'en évaluer la pertinence. *Narramus* a été conçu par des chercheurs qui, selon nous, ne sont pas exemptés de faire la preuve. Évaluer leur dispositif nous permettra d'éclairer les professionnels de l'éducation – enseignants et formateurs – et les politiques sur sa validité.

4.2. Administrer la preuve de l'efficacité d'une pratique en éducation

En 2014, France Stratégie (organisme public d'études, de prospective, d'évaluation des politiques publiques et de propositions, placé auprès du Premier ministre) a réalisé une synthèse des principaux travaux français cherchant à évaluer l'efficacité des pratiques pédagogiques, en présentant leur méthodologie et leurs résultats saillants. Dans le domaine de la compréhension, le document recense deux travaux jugés robustes d'un point de vue méthodologique. La première étude (Bianco *et al.*, 2010, 2012) correspond à notre propos, nous allons la présenter.

Cette recherche porte sur l'efficacité d'un enseignement précoce des stratégies de compréhension. Cette étude non randomisée portait sur un échantillon de 1 273 élèves âgés de 4 ans au début de l'enquête et scolarisés pour moitié dans des écoles de zone d'éducation prioritaire. Les élèves étaient suivis sur trois ans (de la moyenne section au cours préparatoire) et recevaient un programme d'enseignement parmi trois : le premier portait sur le développement de la compréhension, de manière implicite à travers la lecture et l'analyse d'histoires (SA), le deuxième proposait un enseignement explicite centré sur le développement de la conscience phonologique (PHO), le dernier un enseignement explicite de la compréhension avec des modules spécifiques sur la détection d'incohérences, les inférences logiques et la structure de l'histoire (CS). En fonction du programme suivi, trois groupes d'élèves ont été formés avec des sous-groupes constitués en fonction de la durée d'exposition au programme (un ou deux semestres). Les élèves ont été testés à plusieurs reprises : en début et en fin de moyenne section, en fin de grande section et à la fin du cours préparatoire. Dans le domaine de la compréhension, seul le programme centré sur l'enseignement explicite de la compréhension (CS), lorsqu'il était mis en œuvre pendant deux semestres, est efficace. À notre connaissance, cette recherche est la seule ayant évalué un outil visant l'amélioration de la compréhension chez les jeunes élèves, en contexte français. D'où notre intérêt pour évaluer un autre outil dans le même domaine d'apprentissage langagier mais qui se distingue sur deux points cruciaux : le choix des compétences visées et évaluées d'une part, la nature et la planification des activités proposées aux élèves d'autre part. Nous détaillerons précisément ces éléments dans le chapitre 2.

Pour administrer la preuve et savoir si les progrès des élèves peuvent être attribués à l'intervention, il est nécessaire d'utiliser une démarche expérimentale qui consiste à comparer les progrès des élèves d'un groupe expérimental avec celui d'un groupe témoin. C'est cette méthodologie – « robuste » – qui est utilisée dans l'étude précitée. Sans cette comparaison, les études ne permettent pas de savoir si le dispositif est plus efficace qu'un autre. D'où notre volonté de conduire une recherche empirique dans une perspective comparatiste (Duru-Bellat & Mingat, 1998) visant à mesurer si les élèves dont les enseignants utilisent l'outil *Narramus* progressent davantage que les autres dans le domaine de la compréhension et de la narration.

5. Une méthodologie rigoureuse

Pour construire notre protocole expérimental, nous nous appuyons sur des critères méthodologiques identifiés dans l'approche fondée sur les données probantes. La recherche de dispositifs et de solutions validés par la science, au nom de *l'evidence-based education*¹², alimente des échanges foisonnants au niveau international et prend une place de plus en plus importante dans les débats nationaux (Meuret, 2007). Dans le cadre de ce travail, l'approche fondée sur la preuve nous servira de cadre méthodologique pour évaluer l'innovation *Narramus*.

5.1. Les fondements de l'éducation fondée sur la preuve

Que ce soit aux États-Unis, au Royaume-Uni ou en France, l'approche fondée sur les preuves a émergé dans un contexte de critiques du fonctionnement et des résultats de la recherche en éducation. Celle-ci a notamment été remise en cause pour son manque d'utilité sociale (Davies, 1999; Hargreaves, 1997, 1999; Slavin, 2002) et de rigueur scientifique (Hargreaves, 1997, 1999 ; Lyon & Chhabra, 2004 ; Normand, 2006 ; Prost, 2001 ; Whitty, 2006). Au début des années 2000, des débats virulents ont eu lieu autour de la qualité de la recherche en éducation et des méthodologies utilisées (Gorard, Rushforth, & Taylor, 2004; Hadji & Baillé, 1998; Ravitch, 1985; Slavin, 2002; Snow, 2015).

¹² *L'evidence-based education* s'inscrit dans un contexte plus large d'*evidence movement*. Cette expression peut être traduite de plusieurs manières : « éducation fondée sur les données probantes », « pratique éducative basée sur les preuves », « politique d'éducation fondée sur la recherche », etc. (Rey, 2014). Dans notre travail, nous utiliserons principalement les termes *evidence-based education*, éducation fondée sur la preuve ou éducation fondée sur les données probantes.

À la suite de Saussez et Lessard (2009), Rey (2014) dresse une synthèse de ses principales insuffisances relevées dans la recherche en éducation. Les études sont, avant tout, critiquées pour :

1° le manque de cumul des connaissances : de nombreux travaux sur des sujets identiques ne prennent pas assez en compte les résultats antérieurs.

2° Leurs possibles dérives idéologiques, lorsqu'elles se rapprochent de professions de foi politiques. Fervents défenseurs de l'*evidence-based education* aux États-Unis, Davies (1999) et Slavin (2002, 2008) dénoncent ainsi les intérêts idéologiques ou politiques trop souvent présents dans la conduite des réformes éducatives.

3° Le manque de clarté et de transparence des dispositifs de recherche. Les démarches apparaissent confuses et peu explicites, que ce soit sur les questions posées ou sur les protocoles de recherche mis en œuvre. Les aspects qualitatifs et théoriques sont privilégiés au détriment de bases empiriques rigoureuses.

4° La quantité d'informations disponibles difficilement traitables par les enseignants qui ont du mal à trier et à apprécier les résultats produits par la recherche. Les études sont peu ou mal diffusées et restent donc peu connues des acteurs de terrains, enseignants et formateurs.

5.2. Définition et objectifs de l'*evidence-based education*

L'*evidence-based education* est née du mouvement de l'*evidence-based medicine* (la médecine fondée sur les preuves) qui désigne l'utilisation optimale de l'information médicale pour la pratique clinique (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). La médecine fondée sur les données probantes est fréquemment utilisée comme référence par les défenseurs de l'éducation basée sur la preuve (Hargreaves, 1997 ; Oakley, 2002 ; Slavin, 2002). Cette dernière prend appui sur une recherche exhaustive et comparée des recherches empiriques sur un sujet donné afin de dégager les meilleures pratiques. Il s'agit d'utiliser les données existantes issues de la recherche et de la littérature internationale en utilisant des critères de qualité et en définissant des niveaux de preuves exigeants.

Pour Davies (1999, pp. 109-110), « l'objectif de l'éducation fondée sur des données probantes est de faire en sorte que les recherches futures répondent aux critères de validité scientifique, de qualité et de pertinence pratique qui font parfois défaut dans les données existantes sur les activités, les processus et les résultats éducatifs ». Slavin (2002) avance l'idée

qu'appuyer l'adoption d'un programme sur les données probantes permet de réduire les différences de performances au sein de la communauté éducative. Cette approche répond aux critiques présentées ci-dessus et permet, en outre, d'éloigner les considérations trop idéologiques (Goigoux et al., 2015; Slavin, 2008) et d'éclairer les professionnels dans leur prise de décision.

L'éducation fondée sur les données probantes vise trois objectifs :

- 1° asseoir les réformes et les pratiques éducatives sur des résultats scientifiques avérés ;
- 2° améliorer la qualité de la recherche en éducation, notamment sa capacité à fournir des résultats probants de nature causale ;
- 3° favoriser les méthodologies qui répondent à cet objectif, en particulier les démarches expérimentales (ou quasi-expérimentales) ainsi que les « méta-analyses » ou « revues systématiques de recherche » (Rey, 2014).

Dans le cadre de cette recherche doctorale, nous poursuivons deux objectifs spécifiques :

- 1° mesurer les effets d'un outil ;
- 2° utiliser une méthode basée sur des critères rigoureux en élaborant un protocole expérimental inédit dans la recherche française en éducation.

Notre démarche est guidée par la volonté d'établir la preuve de l'efficacité d'un dispositif didactique en nous appuyant sur des principes issus de l'*evidence-based education*.

5.3. Des exigences méthodologiques fortes

De manière générale, une recherche qui s'inscrit dans le paradigme de l'éducation basée sur la preuve passe par des étapes clairement identifiées : 1° l'élaboration d'une revue systématique des recherches existantes sur le sujet traité ; 2° le développement de programmes d'intervention, 3° leur mise à l'essai dans le système éducatif en anticipant les adaptations nécessaires et définissant les critères d'évaluation de l'intervention ; 4° la diffusion des programmes d'intervention à plus grande échelle et enfin, 5° l'extension des domaines d'intervention.

Notre étude consiste à tester un outil lui-même élaboré à partir de résultats solidement établis (*cf.* chapitre 2). Elle correspond à l'étape 3 mais anticipe aussi largement l'étape 4 – la diffusion à grande échelle – par l'ampleur numérique de l'échantillon constitué.

5.3.1. Les critères méthodologiques retenus par l'*evidence-based education*

Les recherches « interventionnelles » consistent à évaluer les éventuels bénéfices d'interventions pédagogiques visant des compétences spécifiques. Quels sont les principes retenus pour réaliser ces évaluations ?

5.3.1.1. Des standards méthodologiques

L'éducation fondée sur la preuve définit des critères exigeants portant sur la conception de la recherche, la taille de l'échantillon, la qualité méthodologique et la taille de l'effet. Le rapport américain intitulé *Coalition for Evidence-Based Policy* (2003) détaille les caractéristiques méthodologiques des recherches évaluatives inscrites dans ce paradigme. Il vise à fournir aux praticiens de l'éducation des outils pour distinguer les données probantes rigoureuses de celles qui ne le sont pas. Dans la mesure où ces principes ont, en grande partie, guidé l'élaboration de notre protocole expérimental, nous avons choisi de les présenter ci-dessous.

Les essais contrôlés aléatoires (ou « randomisés ») dans lesquels les élèves, les classes ou les écoles sont assignés au hasard dans les différentes conditions expérimentales sont privilégiés pour évaluer l'efficacité d'une intervention (Hargreaves, 1997, 1999; Slavin, 2002, 2008). Ils sont considérés comme une norme d'excellence, un « étalon-or » (*gold standard*) en matière d'évaluation de l'efficacité d'un dispositif, d'un outil ou d'une politique publique (Saussez & Lessard, 2009). Pour Cheung et Slavin (2016), les recherches les plus fiables sont celles dans lesquelles on a réalisé des évaluations sur de grands échantillons et pour lesquelles les écoles ou les enseignants ont été assignés de manière aléatoire dans le groupe expérimental ou le groupe contrôle.

Si les études randomisées représentent la norme de l'*evidence-based education*, les protocoles expérimentaux ne peuvent pas toujours remplir cet exigeant cahier des charges en raison de contraintes temporelles, budgétaires ou de difficultés éthiques soulevées par l'affectation aléatoire des élèves dans le groupe expérimental ou dans le groupe témoin (Labrousse & Zamora, 2013). Ainsi, le rapport *Coalition for Evidence-Based Policy* note l'intérêt d'autres travaux dans lesquels le groupe témoin est correctement apparié au groupe

d'intervention, de telles études pouvant contribuer à établir des preuves de l'efficacité d'une intervention. Dans celles-ci, les groupes expérimentaux et les groupes témoins sont appariés sur des variables jugées clés – notamment des niveaux de réussite scolaire au pré-test et des caractéristiques démographiques – avant que les scores des post-tests soient connus. Dans ce cas-là, la recherche doit fournir des données montrant, qu'avant l'intervention, le groupe expérimental et le groupe témoin ne présentaient pas de différence dans la majorité des caractéristiques mesurées. Pour les enseignants, le groupe témoin doit être composé de maîtres qui n'ont pas eu la possibilité de bénéficier de l'intervention et non d'enseignants qui ont eu le choix mais qui ont refusé. On sait, par exemple, que la différence de motivation peut conduire à des effets différents pour les deux groupes et ainsi modifier les résultats. C'est l'ensemble de ces exigences que nous avons cherché à satisfaire dans notre recherche doctorale.

Le rapport dresse l'inventaire d'autres conditions à inclure pour construire des preuves, conditions exigeantes rarement réunies dans les évaluations des recherches interventionnelles, mais que nous nous sommes aussi efforcée de respecter.

- Décrire précisément l'intervention (qui l'a administrée, qui l'a reçue, qu'a-t-elle coûté, en quoi diffère-t-elle de ce que le groupe témoin a reçu et quels sont les résultats qui devraient en être affectés ?)
- Choisir des épreuves évaluatives pour mesurer les effets de l'intervention. En d'autres termes, évaluer les compétences enseignées.
- Limiter le taux d'attrition, c'est-à-dire vérifier qu'il y a peu de pertes de données au moment de la collecte des résultats et que les pertes sont comparables pour le groupe expérimental et le groupe témoin. Il est possible de compenser les pertes par des procédés statistiques mais la méthode la plus efficace est d'essayer de limiter autant que possible l'attrition (Duflo, 2004; Slavin, 2019).
- Collecter des données sur le long terme pour vérifier si les effets de l'intervention se maintiennent dans le temps.
- Indiquer l'ampleur des effets et montrer qu'ils sont statistiquement significatifs si l'intervention produit des améliorations sur les performances des élèves.
- Constituer des échantillons relativement grands (au moins trois cents élèves, cent-cinquante dans le groupe d'intervention et cent-cinquante dans le groupe témoin, pour obtenir un résultat statistiquement significatif).
- Communiquer tous les résultats de l'intervention, qu'ils soient positifs ou négatifs.

5.3.1.2. Les conditions d'une adhésion aux pratiques de recherche interventionnelle

En contexte français, Gentaz (2017, 2018) partage les standards méthodologiques internationaux et revendique lui aussi de ne pas cacher les échecs ou les résultats décevants. Il va plus loin cependant en inventoriant les conditions d'une adhésion aux pratiques de recherches interventionnelles :

- la confiance accordée par les tutelles, c'est-à-dire la confiance que place l'Éducation nationale en ses enseignants ;
- la participation volontaire des enseignants et des chercheurs ;
- l'implication des formateurs dans ces dispositifs, y compris dans les circonscriptions éloignées des centres universitaires (il souhaite à ce sujet une implication des Écoles Supérieures de Professorat et d'Éducation aux programmes de recherche interventionnelle) ;
- une autonomie financière, scientifique et pédagogique pour ne pas dépendre de financements privés ou publics qui pèsent sur la mise en œuvre, le maintien et la durée des dispositifs.

Dans une perspective de développement de recherches interventionnelles réalistes pour les chercheurs et pour les enseignants, ces conditions nous paraissent indispensables. Nous avons élaboré notre protocole expérimental en les prenant en compte mais en y ajoutant d'autres critères spécifiques utilisés par les chercheurs pour réaliser les méta-analyses.

5.3.1.3. La revue systématique ou « méta-analyse »

Outre l'élaboration d'une méthodologie évaluative des recherches interventionnelles très rigoureuse, les partisans de l'*evidence-based education* préconisent un recours, une synthèse et une comparaison systématiques des résultats existants. Cette revue systématique (*systematic review*) ou « méta-analyse » vise à identifier, analyser et évaluer la qualité des résultats produits par la recherche sur un sujet précis (Davies, 1999; Gough, 2007; Rey, 2014). La démarche consiste à : poser une question scientifiquement pertinente sur l'éducation ; utiliser des critères de recherche rigoureux pour collecter les données de manière systématique ; évaluer et comparer les données et les études de manière critique, en s'appuyant sur des références scientifiques et professionnelles ; définir la pertinence des recherches quant à la demande sociale en éducation.

Les chercheurs recensent toutes les études méthodologiquement acceptables et synthétisent leurs résultats, en utilisant des procédés statistiques pour juger de leur efficacité. La méta-analyse ne présente pas les perspectives théoriques, elle porte exclusivement sur les effets des programmes, des interventions ou des politiques. Les procédures utilisées sont très contrôlées et systématisées, les recherches retenues sont construites selon des approches expérimentales ou quasi-expérimentales et comparent donc les progrès d'un groupe expérimental à celui d'un groupe contrôle.

« En complément aux recherches évaluatives, la méta-analyse est un instrument conçu pour produire des synthèses de recherche. Au sens strict, la méta-analyse est une technique statistique qui consiste à combiner les résultats de plusieurs études empiriques afin de déterminer la force de certaines relations causales » (Lessard, 2007, p. 72). Cette définition souligne le double intérêt des méta-analyses : Elles balisent les standards méthodologiques et jouent donc le rôle de référence de ces standards en fournissant des critères de base pour qualifier les expérimentations valides. D'autre part, par le calcul de l'amplitude de l'effet, elles permettent d'établir ce que sont les bonnes pratiques. Elles visent non seulement à identifier les bonnes pratiques mais, avec les comparaisons et l'inclusion de la taille de l'effet, elles déterminent les meilleures pratiques.

Pour Saussez et Lessard (2009, p. 4), « il s'agit d'établir la meilleure preuve disponible sur le marché de la recherche à propos d'une question relative à l'efficacité d'une méthode, d'un programme, d'une politique ».

S'il existe différentes échelles pour évaluer la qualité de la recherche expérimentale, celles retenues pour réaliser les méta-analyses doivent être explicites et transparentes (Oakley, 2002). Malgré cette rigueur affichée, des chercheurs (Slavin, 2008) regrettent l'importante variation dans les méthodes utilisées pour établir ces synthèses, ce qui peut conduire à des conclusions incohérentes sur les programmes et les pratiques dont l'efficacité est démontrée. Certaines, parce qu'elles négligent le *design* des études en retiennent qui n'ont pas de groupe contrôle, évaluent un contenu dispensé aux groupes expérimentaux mais pas au groupe témoin, proposent une intervention de courte durée ou travaillent avec un échantillon trop limité (Slavin, 1995).

Ce chercheur propose donc une méthode d'investigation nouvelle intitulée « *best-evidence synthesis* », plus exigeante qu'une méta-analyse classique. S'il reprend les principes de celle-ci et si les critères retenus sont proches de ceux utilisés dans l'évaluation des recherches

primaires, il applique des critères d'inclusion plus stricts pour sélectionner les études. Trois d'entre eux ont retenu notre attention.

Pour Slavin, un nombre trop faible d'élèves pose des problèmes de fiabilité statistique : un nombre insuffisant de classes et d'écoles peut amener à une confusion entre les effets de l'intervention et ceux de l'effet-maitre et de l'effet-classe. Ce risque diminue avec l'augmentation du nombre d'unités indépendantes, les évaluations de plus de 250 élèves dans 10 classes ou écoles sont privilégiées. Il insiste sur la durée minimale des interventions testées. Selon lui, certaines études très courtes (parfois quelques heures) sont trop artificielles et présentent peu d'intérêt pour les praticiens. Dérivées d'expériences en laboratoire, elles ne permettent pas d'informer les professionnels de l'impact probable des programmes. Il propose d'exiger 12 semaines minimum d'intervention. Il prend pour exemple les études brèves (5 heures ou moins) et limitées en nombre (69 élèves ou moins) évaluant *Daisy Quest*, un programme informatique utilisé aux États-Unis pour enseigner la conscience phonémique en première année de maternelle alors que les groupes contrôles ne bénéficient d'aucun enseignement visant des compétences phonémiques. Enfin, Slavin souhaite que les auteurs de méta-analyses préfèrent les études prospectives aux études rétrospectives et s'assurent que les résultats mesurés soient cohérents avec l'hypothèse testée.

Nous reprenons à notre compte ces trois exigences supplémentaires relatives à la taille de l'échantillon (le nôtre sera vaste), à la durée des interventions (elle durera douze semaines chaque année) et aux contenus d'enseignement (les enseignants du groupe témoin viseront les mêmes apprentissages que ceux des groupes expérimentaux).

5.3.2. Comment procéder pour évaluer ?

Nous avons tenté d'inclure l'ensemble des critères méthodologiques inventoriés ci-dessus tout en les adaptant pour qu'ils s'appliquent à des conditions ordinaires d'enseignement. En d'autres termes, nous avons fait le choix d'élaborer un « protocole quasi expérimental dans un contexte écologique » (Daunay, 2009).

En premier lieu, nous souhaitons mesurer les progrès des élèves. Pour évaluer la capacité d'un outil pédagogique à faire progresser les élèves, nous mesurerons leurs performances à deux moments distincts, avant et après enseignement. Pendant l'expérimentation, ils seront testés à plusieurs reprises, sur plusieurs mois, voire sur plusieurs années pour les élèves de petite section et moyenne section (2016 à 2019).

Ensuite, deux groupes seront constitués : les élèves du groupe expérimental bénéficieront de l'intervention (ici, l'enseignement avec l'outil didactique dont on cherche à évaluer l'efficacité) et les élèves du groupe contrôle travailleront sur les mêmes objectifs et les mêmes textes avec les méthodes d'enseignement habituelles. Partant, nous mesurerons les effets moyens du dispositif sur un groupe d'élèves en les comparant avec un groupe « similaire » qui n'en a pas bénéficié.

En résumé :

- chaque groupe sera constitué d'un échantillon d'élèves et de classes important afin de pouvoir disposer de données traitables statistiquement et d'améliorer la puissance et la robustesse des analyses et des résultats.
- Nous choisirons les tests évaluatifs cohérents avec les apprentissages visés en retenant des épreuves classiques de psycholinguistique.
- Une analyse de la taille de l'effet de l'enseignement sera effectuée quand cela sera possible.
- L'enseignement se déroulera durant plusieurs mois, chaque année : trois fois quatre semaines pour chaque niveau.
- L'intervention durera une année scolaire pour les élèves scolarisés en grande section à la rentrée 2016, deux années pour les élèves de moyenne section et trois années pour les élèves de petite section.

Pour mesurer l'efficacité d'un outil, il faut s'assurer qu'il a bien été utilisé comme prévu par ses auteurs, autrement dit vérifier que son implémentation est satisfaisante. Un contrôle de la fidélité de la mise en œuvre de l'outil s'imposera donc. Si certains auteurs distinguent deux composantes dans l'implémentation d'un dispositif, la quantité et la qualité (Dusenbury, 2003; O'Donnell, 2008; Resnicow et al., 1998), nous engloberons ces deux dimensions et évaluerons la fidélité par la proportion d'activités préconisées réalisées par les enseignants (Dusenbury, 2003; Payne, Gottfredson, & Gottfredson, 2006). Il n'est pas réaliste de s'attendre à ce que tous mettent en œuvre l'ensemble du contenu d'un dispositif innovant. Durlak et DuPre (2008) considèrent qu'un taux de mise en œuvre de 80 % est un maximum pour de nombreuses interventions avec des attentes plus réalistes de l'ordre de 60 à 80 %. Ils montrent que l'effet mesuré est d'autant plus fiable que l'implémentation est fidèle. Le seuil de 60 % est généralement choisi comme seuil plancher en dessous duquel l'effet mesuré ne peut plus être imputé à l'outil testé. Nous vérifierons donc que ce seuil est dépassé. En outre, on a montré que la qualité de l'implémentation influe sur les performances des élèves (Resnicow et al., 1998).

Autrement dit, plus les enseignants sont fidèles dans la mise en œuvre des dispositifs proposés, meilleurs sont les résultats des élèves.

Notre expérimentation sera réalisée dans des conditions proches des conditions d'enseignement ordinaire. En effet, si les classes expérimentales bénéficient d'aides spécifiques, de soutiens financiers, de personnel supplémentaire, de formations assurées par les chercheurs ou de matériel rare, la mise en œuvre s'éloigne des pratiques ordinaires, ce qui réduit largement la possibilité d'une généralisation du dispositif après disparition des moyens extraordinaires alloués à l'expérimentation. En respectant ces conditions, la future diffusion de l'outil sera facilitée.

Nos résultats auront alors plus de force car la preuve est plus difficile à établir lorsque les expérimentations se déroulent dans le monde réel et non en laboratoire (Gomel & Serverin, 2013).

Les conclusions de notre travail seront présentées avec prudence et ne devront pas être généralisées outrancièrement : elles ne seront valides que dans le contexte de l'éducation prioritaire, avec de jeunes élèves et dans une temporalité précise.

5.3.3. Synthèse

L'introduction d'outils didactiques peut participer à l'amélioration des pratiques pédagogiques. Que ce soit en contexte francophone (Bélangier et al., 2012; Bonnéry, 2015; Cartier et al., 2012; Goigoux & Cèbe, 2009) ou en contexte anglo-saxon (Berman, 1980; Rowan & Miller, 2007), de nombreux chercheurs valorisent le recours à des outils qui peuvent prendre différentes formes (manuels, ressources numériques, plateforme d'échanges entre enseignants...). L'éducation fondée sur les données probantes joue ici une place centrale puisqu'elle met en avant les pratiques efficaces, basées sur des résultats de recherches (Davies, 1999; Dupriez, 2015; Slavin, 2002). Elle consiste en l'établissement de preuves solides : programmer, mettre en œuvre et publier des études répondant aux exigences les plus élevées de la recherche et de l'évaluation.

Dans un contexte où les critères permettant de juger l'efficacité font eux-mêmes l'objet de controverses, apporter la preuve qu'un dispositif est plus efficace qu'un autre reste difficile. Difficile mais pas impossible ! La qualité de la démarche d'évaluation dépend avant tout des principes retenus. Pour Pasquinelli (2011), des investigations empiriques qui se déroulent dans des contextes réels d'apprentissage et s'appuient sur des évaluations méthodologiquement

solides sont nécessaires pour établir une éducation fondée sur des preuves. Les conditions extrêmement exigeantes identifiées dans l'approche *evidence-based* sont rarement réunies dans les évaluations des recherches interventionnelles mais nous tenterons de nous en approcher. Lors de l'implémentation d'une innovation, l'enjeu premier est celui de l'adoption de l'innovation : les enseignants vont-ils s'emparer de l'outil ? Les principes didactiques de l'outil vont-ils être respectés ? L'outil va-t-il générer les effets escomptés sur les apprentissages des élèves ? Pour répondre à ces questions, nous mettrons en œuvre un dispositif expérimental. Il s'agira d'abord d'évaluer pour mesurer si les effets attendus sont présents. Ensuite, le questionnement se déplacera vers la diffusion des nouvelles pratiques. Dans une perspective d'amélioration de l'École, cet enjeu nous paraît aussi fondamental que le premier. En effet, la volonté affirmée de faire progresser les élèves ne sera effective que si elle s'accompagne d'une mise en œuvre à grande échelle, avec fidélité et au fil du temps. Trop souvent, des innovations qui ont fonctionné en contexte expérimental ne se révèlent pas fructueuses lorsqu'elles sont diffusées et généralisées. Le « passage à l'échelle » (Odom, 2009) échoue.

Nous allons à présent nous interroger sur les raisons de cet échec. Pourquoi les outils qui ont fait la preuve de leur efficacité dans des conditions privilégiées ne parviennent-ils pas à être transposés dans les contextes scolaires ordinaires ? Quels sont les principaux obstacles et quelles sont les conditions à réunir pour qu'une innovation bénéficie aux élèves auxquels elle est destinée ?

6. Le délicat passage à l'échelle

Si l'innovation est positivement évaluée, c'est sa diffusion dans le système éducatif qui lui donnera tout son sens (Cros, 2004). Le « passage à l'échelle » est le moment critique où des expériences réussies dans des conditions expérimentales et par un petit nombre d'enseignants sont étendues à une population plus importante dans des conditions ordinaires (Fixsen, Blase, Metz, & Van Dyke, 2013; Fixsen, Blase, Horner, & Sugai, 2009; Odom, 2009).

6.1. Définition du passage à l'échelle

Stringfield, Datnow, Ross et Snively (1998) définissent le passage à l'échelle comme « l'expansion délibérée dans de nombreux milieux d'un plan de restructuration scolaire élaboré à l'externe qui a été utilisé avec succès dans un ou un petit nombre de milieux scolaires (p. 271) ». Pour Coburn (2003), cette définition mérite d'être complétée car elle n'informe ni

sur la nature du changement, ni sur son maintien. En se focalisant uniquement sur des données chiffrées, cette approche du passage à l'échelle néglige des facteurs qualitatifs fondamentaux. Pour cette auteure, elle nécessite non seulement une extension à d'autres sites mais aussi des changements conséquents dans les salles de classe, un maintien des nouvelles pratiques au fil du temps et un changement des conceptions des enseignants. Il ne s'agit pas seulement de toucher un nombre plus important d'enseignants, d'écoles ou de circonscriptions mais la généralisation doit produire des changements profonds et durables.

Dunlap, Sugai, Lewis, Goodman et Horner (2009) complètent la définition de Coburn (2003) et identifient quatre phases dans la mise à l'échelle :

1° l'émergence est la phase pendant laquelle les chercheurs définissent l'innovation avec précision, anticipent le processus de mise en œuvre et de diffusion et décident du passage de mise à l'échelle ;

2° la démonstration des capacités est la phase où les chercheurs déterminent si la pratique est réalisable et si elle a un effet significatif sur les résultats visés ;

3° l'élaboration permet de passer de la démonstration à la mise en œuvre à grande échelle avec un recours à des formateurs locaux ;

4° l'adoption et la durabilité du système nécessitent l'intégration de l'innovation dans les pratiques pour assurer un usage dans le temps. La durabilité est la caractéristique la plus complexe du processus, le maintien des innovations – même les plus efficaces – étant souvent faible (Cobb & Smith, 2008; Fishman et al., 2013).

Odom (2009) définit le passage à l'échelle en distinguant la dimension d'efficacité de celle d'efficience. L'efficacité correspond aux effets avérés d'une intervention avec un contrôle de sa mise en œuvre dans un dispositif de recherche alors que l'efficience se caractérise par les effets des pratiques lorsqu'elles sont implémentées dans des conditions réelles avec moins de supervision. C'est ce passage qui est appelé « mise à l'échelle » dans le sens où les interventions sont mises en œuvre à grande échelle dans des conditions réelles.

Pour notre recherche, nous retiendrons la définition suivante proche de celle de Coburn et de Klingner *et al.* (2013) : la mise à l'échelle caractérise le processus par lequel les chercheurs mettent en œuvre des interventions à petite échelle, les valident, puis les appliquent plus largement dans des conditions réelles. L'évaluation menée ici est un peu particulière : en mesurant les effets de l'innovation sur un échantillon numériquement important et selon des

conditions de mises en œuvre exigeantes, nous considérons qu'il s'agit déjà d'une première étape dans la mise à l'échelle.

6.2. Les obstacles au passage à l'échelle

Pour essaimer, les changements produits par les innovations doivent perdurer après le départ des chercheurs ou des accompagnateurs de l'expérience (Vaughn, Klingner et Hughes, 2000). Or, généralement, la mise en œuvre diminue lorsque l'étude se termine. Supposer, en outre, que les pratiques fondées sur les données probantes vont se généraliser parce qu'elles ont été identifiées reste infructueux (Tseng, 2012) et la diffusion d'outils ou de dispositifs qui ont fait la preuve de leur efficacité en situation expérimentale est rarement concluante (Gentaz et al., 2013; James-Burdumy et al., 2012). Lorsqu'elles sont généralisées et évaluées en contexte ordinaire, les méthodes innovantes ne produisent pas toujours les effets positifs escomptés. Dans sa méta-analyse, Gurgand (2018) relève par exemple, que sur soixante-dix-sept expérimentations « randomisées » menées entre 2002 et 2011, sept seulement ont produit des effets positifs durables sur un temps relativement long.

Notre interrogation ici est la même que celle de Charlier et Peraya (2003): « À quelles conditions un dispositif innovant – dont le caractère expérimental et donc extra-ordinaire reste encore fort marqué – peut devenir à terme un mode de fonctionnement pérenne et donc ordinaire tout en respectant la diversité des contextes ? » (p. 15). Nous allons à présent examiner les facteurs qui influencent l'appropriation des outils et qui facilitent ou freinent la mise en œuvre et la diffusion des nouvelles pratiques.

6.2.1. Une tension entre la prescription et l'autonomie des enseignants

Bien documentés dans les travaux nord-américains, les changements pédagogiques à grande échelle empruntent généralement deux voies (Enthoven, Letor, & Dupriez, 2015). La première s'appuie sur le développement de pratiques standardisées jugées comme scientifiquement efficaces (*evidence-based education*). La seconde prône la nécessité de développer professionnellement les enseignants en favorisant notamment leur autonomie et leur apprentissage et en ayant confiance dans leur jugement professionnel (Biesta, 2007).

Citant les travaux de Bidwell (1965), Enthoven *et al.* (2015) évoquent la tension entre deux logiques organisationnelles dans le système éducatif actuel. La première « bureaucratique » se caractérise par la division du travail entre enseignants (répartition des

élèves par âge, par niveau...). Partant, il existe une segmentation entre les enseignants, les élèves et les savoirs à enseigner. La seconde dite « professionnelle » correspond à l'autonomie laissée aux enseignants (choix de la méthode ou des dispositifs d'évaluation, modalités d'enseignement, gestion des élèves...). Ces deux logiques mettent en tension la liberté d'action des enseignants avec d'un côté des aspects très prescriptifs et de l'autre une volonté d'encourager l'autonomie. Ainsi, cette opposition constitue « une première piste explicative de la perméabilité des organisations scolaires à l'innovation et de la difficulté d'implémentation d'une réforme » (Enthoven et al., 2015, p. 97). Pour Tardif et Lessard (1999), notre système éducatif correspond à une structure dite « cellulaire », c'est-à-dire « une organisation dans laquelle les travailleurs sont séparés les uns des autres et accomplissent une tâche à la fois complète et autonome dans un lieu soustrait au regard des autres » (p. 57). Cette approche illustre bien la tension entre les deux logiques « bureaucratique » et « professionnelle ». Selon eux, dans la mesure où les innovations sont faiblement prescriptives et laissent une grande autonomie aux enseignants, elles renforcent la dimension professionnelle du métier d'enseignant. Notre analyse diffère légèrement en ce sens qu'elle n'oppose pas ces deux approches de manière aussi catégorique : les outils peuvent assurer une fonction de prescription, certes secondaire par rapport aux programmes ou directives ministérielles, mais bien réelle.

Rowan et Miller (2007) ont déjà rapproché ces deux logiques lorsqu'ils ont analysé les *Comprehensive School Reforms* (CSR¹³). Ils se sont appuyés sur les logiques des changements de pratiques identifiées auparavant par Berman (1980). La première, qualifiée d'approche programmée, repose sur la conformité des pratiques d'enseignement aux modèles proposés. La seconde, nommée approche adaptative, encourage à adapter les innovations pédagogiques au contexte local dans lequel elles doivent être développées. Cette deuxième approche paraît plus motivante pour les enseignants (Rowan, 1990, cité par Dupriez, 2015). Suite aux travaux de Berman, Rowan et Miller (2007) ont caractérisé trois approches du changement permettant de distinguer les programmes adossés aux *Comprehensive School Reform*. Ils ont réalisé une enquête longitudinale dans cent-quatorze écoles primaires américaines et se sont intéressés à la perception des caractéristiques organisationnelles des différents modèles par les chefs d'établissement et les enseignants, ainsi qu'à la relation entre ces modèles et les changements de pratiques. Comme l'indique le Tableau 6, les principaux éléments permettant de caractériser

¹³ Aux États-Unis, le programme des *Comprehensive School Reform* a fourni un financement important pour permettre aux écoles d'adopter des « modèles » transférables d'une école à l'autre, à partir des données probantes.

ces approches sont la guidance pédagogique, l'autonomie des enseignants, le rapport aux valeurs et aux normes, le *leadership* pédagogique et le travail collaboratif entre les enseignants.

Tableau 6. Les modalités d'accompagnement du changement Rowan & Miller (2007)

	Approche procédurale	Approche professionnelle	Approche culturelle
Guidance et standardisation	+	+	-
Autonomie des enseignants	-	+/-	+
Valeur et normes	+/-	+/-	+
Leadership pédagogique local	-	+	+/-
Travail collaboratif	-	+	+

L'approche procédurale se caractérise par la mise en œuvre d'outils accompagnés d'instructions détaillées (objectifs, procédures et modalités de travail...). Elle est complétée par des outils d'évaluation permettant de vérifier si les pratiques prescrites sont bien mises en oeuvre. Dans cette approche, l'autonomie des enseignants n'est pas favorisée et le *leadership* pédagogique vise avant tout le contrôle.

L'approche culturelle correspond à une volonté de développer l'autonomie de l'enseignant et la réflexion collective. Elle s'appuie sur le partage de valeurs et de normes communes et laisse une grande liberté aux enseignants. Il s'agit avant tout de favoriser l'adoption des valeurs portées par le programme.

L'approche professionnelle du changement repose sur une orientation forte des pratiques enseignantes par la mise en place de standards professionnels élaborés par des experts. Si elle autorise une appropriation et une adaptation au contexte et à l'environnement des outils proposés, elle reste très cadrante avec des orientations pédagogiques et didactiques fortes. Elle repose sur un pilotage important et un travail collaboratif favorisant une dynamique d'échanges entre les professionnels et une appropriation des propositions didactiques et pédagogiques.

Les résultats recueillis montrent que ces différentes approches ne génèrent pas les mêmes effets sur les changements pédagogiques : ce sont les approches professionnelle et procédurale qui produisent davantage de changements de pratiques pédagogiques. La variable guidance – qui correspond au degré de standardisation des pratiques et d'accompagnement des enseignants – semble donc jouer un rôle important dans les transformations identifiées.

6.2.2. L'accompagnement ou le suivi des équipes

Pour certains chercheurs, « l'utilisation d'outils aussi pertinents fussent-ils ne suffit pas à transformer les pratiques » (Collin, 2012). L'accompagnement apparaît alors comme une nécessité dans l'adoption des innovations.

6.2.2.1. Caractéristiques de l'accompagnement

Qualifié de « nébuleuse sémantique » (Paul, 2004), l'accompagnement est un terme polysémique. Tantôt tutorat, soutien, suivi, conseil, mentorat, parrainage, compagnonnage ou guidance, il n'existe pas de définition unifiée et stable dans le champ éducatif : « on ne saurait réduire l'accompagnement à un mode uniforme, il désigne tant une fonction qu'une posture, renvoyant à une relation et à une démarche qui, pour être spécifiques, n'en sont pas moins vouées à devoir s'adapter à chaque contexte et chaque matrice relationnelle » (*idem*, p. 54). La notion d'accompagnement implique toutefois obligatoirement une relation entre un accompagnateur et un accompagné. Pour Paul (2009), la posture d'accompagnement se caractérise par un cheminement en commun dans une relation où accompagner, c'est se joindre à quelqu'un pour aller où il va en même temps que lui. Il s'agit d'être avec (ce qui suppose un dialogue) et d'aller vers (ce qui suppose un cheminement). Pour certains auteurs (Pineau, Robin, Boutinet, & Denoyel, 2007), la relation entre l'accompagnateur et l'accompagné doit s'équilibrer : le premier ne doit pas prendre une position de domination sur le second. La relation d'accompagnement diffère ainsi de la relation hiérarchique. Paul (2004) rejoint ces auteurs « s'il y a ascendance de l'accompagnateur sur l'accompagné, il n'y a plus d'accompagnement » (p. 140). Dans le cadre de ce travail, nous garderons l'idée d'un accompagnement correspondant à « la participation à un processus formatif permanent mettant un lien deux personnes : l'accompagnant et l'accompagné autour de points de bifurcation individuels » (Roquet, 2009, p. 6).

Pineau *et al.* (2007) identifie deux types d'accompagnement :

1° « l'accompagnement-visée » caractérise un accompagnement « vers l'atteinte de tel ou tel but : une performance, un emploi, une production, une réalisation, la réussite à telle ou telle épreuve »,

2° « l'accompagnement-maintien » correspond à un « accompagnement dans la situation existentielle momentanément donnée, un accompagnement qui assure une présence

pour encourager la persévérance au sein de tel ou tel type d'activité, de tel ou tel statut, dans la mesure où ce statut ou cette activité sont menacés de précarité, voire de disparition. En un mot, l'accompagnement-maintien est vecteur d'autonomie pour l'adulte accompagné » (p. 31). Il caractérise les situations d'analyse de pratiques dans lesquelles « l'accompagnateur aide le groupe à se repositionner, voire se réorienter en examinant ses pratiques individuelles et collectives, ses modes de fonctionnement, sa propre histoire, comme celle de son projet. Il s'agit alors d'un accompagnement réflexif » (p. 43).

Loin de fournir une liste exhaustive des fonctions de l'accompagnement, ce préambule nous permettra d'analyser et de caractériser les modalités retenues par les accompagnateurs dans le cadre de notre recherche expérimentale (chapitre 3).

6.2.2.2. L'accompagnement, une nécessité ?

Depuis longtemps, l'accompagnement et le soutien des innovations apparaissent comme indispensables (Ramel & Bovey, 2017). 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 (Bélanger et al., 2012; Charlier, Biémar, & Jorro, 2012; Kretlow & Bartholomew, 2010; Penneman et al., 2016; Rowan & Miller, 2007).

En outre, on a montré que les programmes accompagnés donnent de meilleurs résultats car les suivis réguliers aident les enseignants à respecter les procédures proposées (Rowan & Miller, 2007). La méta-analyse de Yoshikawa *et al.* (2013) l'atteste : les innovations les plus efficaces au préscolaire se caractérisent par la qualité des contenus et par le soutien continu offert aux enseignants pour assurer que les expériences immédiates des enfants, celles fournies par les activités et les interactions, sont riches en contenus et en stimulations, tout en étant émotionnellement favorables. Pour favoriser l'adoption et le maintien des innovations, le soutien doit aussi s'inscrire dans la durée (Little, 2006), les formations ponctuelles n'entraînant généralement pas de changement durable dans la pratique ou l'adoption d'une innovation (Garet, Porter, Desimone, Birman, & Yoon, 2001; Odom, 2009).

Il existe une pluralité de modalités de soutien (Penneman, 2018) : ce dernier peut consister en un appui technique, logistique ou organisationnel, une collaboration entre des chercheurs et des acteurs de terrains (Levin & Fullan, 2008) ou une formation et/ou un accompagnement (Bélanger et al., 2012; Dupriez, 2015). La formation apparaît comme un

facteur décisif dans la capacité des enseignants à s'approprier un outil et à s'impliquer dans un dispositif (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001; Durlak & DuPre, 2008; Gentaz, Sprenger-Charolles, Colé, Theurel, & Gurgan, 2013). Ainsi, l'équipe de Gentaz *et al.* (2013) identifie la formation et le suivi des enseignants comme une des limites de leur dispositif. Les chercheurs regrettent, par exemple, de ne pas avoir proposé une « formation approfondie régulière et un suivi personnalisé pour chacun des enseignants ». On peut toutefois se questionner sur cette limite puisque, dans cette étude, chaque enseignant a reçu une trentaine d'heures de formation sur une année scolaire, ce qui est déjà bien au-delà des 18 heures d'animation pédagogique annuelles règlementaires en France. La généralisation du dispositif expérimental serait par conséquent impossible en contexte scolaire ordinaire, mais Gentaz (2018) ne semble pas en tirer de conséquences : il réaffirme le rôle indispensable de l'accompagnement et attribue les différences d'efficacité constatées entre les interventions à petite échelle et à grande échelle au manque de formation. Pour d'autres chercheurs, la formation ne suffit pas : elle ne permet ni de transformer les pratiques, ni d'en développer de nouvelles de manière durable, c'est pourquoi un accompagnement prolongé et régulier est préconisé. Lafortune (2008a, 2008b), par exemple, affirme que l'accompagnement privilégie la réflexion dans et sur l'action et assure une continuité et une cohérence avec les composantes du changement à mettre en œuvre. En résumé, tous ces chercheurs s'accordent pour soutenir que la mise en place de dispositifs de formation et d'accompagnement des enseignants permet de renforcer le changement pédagogique produit par l'introduction d'un outil didactique (Bélanger *et al.*, 2012 ; Dupriez, 2015, Enthoven, Letor *et Dupriez*, 2015).

L'accompagnement assure des échanges organisés autour des pratiques, des expériences positives et des difficultés rencontrées lors de l'utilisation de l'outil (Horn & Little, 2010). Pour Letor, Enthoven *et Dupriez* (2016), il apparaît comme une nécessité « au risque de voir les conversations dévier vers des considérations périphériques ou de voir l'outil abandonné rapidement ». En France, les auteurs du programme PARLER (2005) ont opté pour un suivi des équipes pédagogiques et une collaboration entre chercheurs, enseignants et équipe de circonscription. L'efficacité de l'outil, son appropriation par les enseignants et l'implication de ces derniers sont imputées au suivi des équipes. Aussi, lorsque par la suite, l'introduction de plusieurs interventions proches du programme PARLER (Ecalte *et al.*, 2015; FEJ, 2014; Gentaz *et al.*, 2013) n'a pas produit pas les mêmes effets, les chercheurs ont-ils incriminé le manque d'accompagnement et de suivi des équipes (Zorman *& al.*, 2015).

Tous ces arguments sont séduisants mais ils se heurtent à la réalité des systèmes scolaires : aucun n'a les moyens d'accompagner des centaines de milliers d'enseignants chargés d'appliquer une réforme pédagogique. Le passage à l'échelle est donc compromis par la nature même des dispositifs expérimentés qui ne peuvent se déployer de manière autonome.

Notre raisonnement semble partagé par d'autres chercheurs. Biesta (2007), par exemple, suggère que les conditions expérimentales réunies pour évaluer les effets d'un dispositif (« prouver qu'il marche ») ne doivent pas être trop différentes de celles de son essaimage en conditions ordinaires. Gurgand (2018) et Klingner *et al.* (2013) soutiennent que les expérimentations ne réussissant qu'au prix d'un encadrement externe de l'activité des enseignants, souvent réalisée par les chercheurs ou leurs équipes, sont très difficiles, voire impossibles, à généraliser. L'intensité du programme (le nombre d'heures allouées au suivi des enseignants, leur répartition dans la durée ou encore le volume d'informations transmises) n'y change rien : dans la méta-analyse réalisée par Kennedy (2016), les programmes les plus intenses ne se révèlent pas toujours les plus efficaces.

On considère que la diffusion est réussie quand les changements produits par les innovations perdurent après le départ des chercheurs ou des accompagnateurs de l'expérience (Vaughn, Klingner & Hughes, 2000). C'est pourquoi la qualité intrinsèque des outils influence la réussite ou l'échec du passage à l'échelle : bien conçus, ils doivent être immédiatement utiles, utilisables et acceptables (Goigoux, Renaud, & Roux-Baron, 2019). Autrement dit, s'ils n'avaient pas besoin d'être accompagnés pour être efficaces, les outils seraient plus faciles à diffuser. Par conséquent, si *Narramus* avait les qualités requises, il pourrait générer des effets positifs sur les apprentissages même sans accompagnement. C'est ce que nous chercherons à vérifier à travers notre protocole expérimental : certains enseignants seront accompagnés dans l'utilisation de l'outil, d'autres pas.

6.2.3. La prise en compte des contextes ordinaires d'enseignement

L'innovation peut être envisagée selon trois composantes : les antécédents, le processus lui-même et les résultats (Rogers, 1962). Si Rogers utilise la notion d'« antécédents », nous la reformulons et la rapprochons de l'analyse *a priori* théorisée par Goigoux et Cèbe lorsqu'ils décrivent le processus de conception de leurs outils didactiques (2018). L'innovation doit tenir compte des pratiques habituelles des enseignants et doit pouvoir s'insérer dans leur genre professionnel (Clot, 1999; Clot & Faïta, 2000; Goigoux, 2007). Cette analyse, partagée à

l'étranger par de nombreux chercheurs en sciences de l'éducation (pour une synthèse, voir Snow, 2015), mais rarement concrétisée en France (cf. 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, dans leur mise en œuvre, des contextes d'enseignement et des habitus professionnels des enseignants (leur *Grammar of schooling* que nous appelons genre professionnel). Ils avaient montré, par exemple, que les chances de réussite d'une innovation dépendaient de deux facteurs principaux : sa compatibilité avec les pratiques habituelles des professeurs (ce qui suppose de les connaître) et, de l'efficacité de l'intervention, c'est-à-dire du rapport entre son coût pour les enseignants (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). Bryk (2015) ajoute que les « bonnes idées » mises en œuvre trop rapidement et sans souci de leur intégration dans la diversité des contextes professionnels sont vouées à l'échec si l'on ne dispose pas de stratégies adossées à une théorie du changement des pratiques professionnelles. Selon lui, les décideurs négligent trop souvent un facteur pourtant décisif : l'engagement des acteurs de terrain, enseignants et cadres, dans les projets innovants. Autrement dit, ils diffusent de nouveaux dispositifs pédagogiques sans prendre suffisamment en compte leurs futurs utilisateurs.

Même si elles se sont avérées efficaces, certaines innovations pédagogiques ont du mal à perdurer après la fin du financement et le départ des chercheurs (Cobb & Smith, 2008). Cette difficulté peut être associée au manque de prise en compte des caractéristiques contextuelles locales ou au manque d'appropriation de la part des enseignants (Elmore, 1996). En outre, les transformations ne sont souvent pas explicitement liées à des changements fondamentaux dans la façon dont les connaissances sont construites, ni à la répartition des responsabilités entre l'enseignant et l'élève, ni à la manière dont ils interagissent.

Comme nous l'avons souligné plus haut, l'adoption à grande échelle de pratiques prometteuses s'avère peu réalisable dans des conditions ordinaires d'enseignement (Gurgand, 2018; Klingner et al., 2013; Odom, 2009). Les problèmes d'échelle sont profondément enracinés dans les incitations et les normes culturelles des institutions. La question de la généralisation des bonnes pratiques éducatives nécessite une réflexion et une prise en compte, dès le départ, des premiers destinataires : les enseignants. C'est pourquoi nous évaluons un outil conçu par des chercheurs qui ont associé des enseignants à son élaboration. L'intégration des

futurs utilisateurs dès le début de la conception devrait faciliter l'adoption de l'innovation par le plus grand nombre.

6.2.4. Le respect du prescrit

Fonctionnaires de la fonction publique d'État, les enseignants répondent à des prescriptions (les programmes, les ressources proposées par le ministère, les orientations données par l'Inspecteur de l'Éducation nationale...) et souhaitent les respecter. Faute de la prise en compte des éléments du prescrit – et des contextes d'enseignement évoqués précédemment –, les maîtres ajoutent, empilent mais ne changent pas profondément leur manière de faire (Huberman, 1973). Parfois, ils essaient de transformer leurs pratiques mais finissent inévitablement par rejeter l'innovation (Huberman, 1982; Imbert, 1986; Rogers, 2010) si elle se révèle contradictoire ou incompatible avec les exigences du terrain. « L'introduction d'un élément novateur dans un système pédagogique donné a pour effet soit une simple juxtaposition, par manque de cohérence explicite du système ; soit un rejet pur et simple par excès de fermeture du système ; soit une transformation du système. » (Imbert, 1986, p. 53)

Nous nous sommes donc assurée de la cohérence de l'outil avec les instructions officielles : *Narramus* respecte les éléments du prescrit. Cette caractéristique, prise en compte par les auteurs dans l'élaboration de l'outil, devrait aussi faciliter sa diffusion.

6.2.5. L'environnement organisationnel

La collaboration entre les professionnels de l'école assure une fonction importante dans le développement professionnel et l'amélioration de la qualité de l'enseignement. Les maîtres enrichissent leur enseignement en collaborant avec d'autres professionnels (Bryk & Schneider, 2003; Hallinger, 2003; Min, Modeste, Salisbury, & Goff, 2016). Ainsi, ceux qui participent à des activités d'apprentissage avec des collègues de la même école, du même département ou du même niveau scolaire sont plus susceptibles d'améliorer leurs pratiques d'enseignement (Dolle, Gomez, Russell, & Bryk, 2013). Pour certains chercheurs (Horn & Little, 2010 ; Dupriez, 2015), la collaboration favorise l'entraide, la communication au sein des équipes éducatives et la construction d'une vision commune de l'enseignement. Or, les innovations introduites dans une seule classe ou dans plusieurs éloignées géographiquement ne permettent pas aux enseignants de collaborer ni d'échanger sur leurs pratiques. Un travail trop individuel et une absence de mobilisation collective (Deaudelin, Dussault, & Brodeur, 2002) peuvent freiner l'adoption d'une innovation.

Notre plan expérimental anticipe cet éventuel obstacle en mobilisant un collectif : il stipule d'engager des équipes (3 enseignants *a minima* par école : PS, MS et GS) afin de ne pas laisser les professionnels découvrir l'innovation de manière isolée. Cette exigence permettra de surcroît de réaliser un suivi longitudinal d'élèves bénéficiant de deux, voire trois années consécutives d'expérimentation *Narramus*. Les élèves de GS évalués en juin 2019 ont ainsi profité de 4 semaines d'enseignement chaque trimestre durant la petite, la moyenne et la grande section de maternelle, soit un total de 36 semaines.

6.2.6. Un juste équilibre entre les gains et les pertes liés à l'innovation

Le contexte de l'enseignement crée souvent des « réalités » qui constituent des obstacles concrets à la mise en œuvre réussie de pratiques fondées sur la preuve. Le rapport gain - perte pèse dans les appropriations du changement (Tyack & Cuban, 1995). Si la mise en œuvre est trop coûteuse et que les avantages sont limités, les enseignants préfèrent ne pas modifier leurs habitudes (Marks & Gersten, 1998). Vaughn & al (2000) expliquent les faibles niveaux de fidélité et de durabilité des pratiques fondées sur la recherche par les éléments suivants : 1° les effets de la mise en œuvre ne sont pas immédiatement visibles pour la plupart des enseignants ; 2° ces derniers estiment que leurs pratiques habituelles sont déjà, en moyenne, efficaces ; 3° la majorité perçoit qu'il existe de nombreuses façons d'enseigner aux élèves et qu'il y a peu de consensus dans la recherche qui justifieraient un changement dans leur pratique pédagogique ; 4° le changement proposé est extrêmement difficile à mettre en œuvre.

Le temps constitue un autre obstacle dans la diffusion des innovations, et ce à double titre. D'abord, il représente une contrainte permanente pour les enseignants qui sont toujours vigilants à ne pas mettre en œuvre des pratiques trop chronophages, peu importe les bénéfices possibles (Stone, 1998; Vaughn, Hughes, Schumm, & Klingner, 1998). Le temps d'enseignement n'est pas extensible et le fait d'introduire une nouvelle activité s'accompagne de la diminution voire de la disparition d'autres tâches.

Ensuite, l'enseignement proposé est parfois beaucoup trop bref : les dispositifs limités à deux semaines d'intervention sont généralement peu efficaces et leurs effets de courte durée (Pullen, Tuckwiller, Konold, Maynard, & Coyne, 2010). « Les écoles et les districts sont trop souvent impatients d'obtenir des résultats à court terme et ne veulent pas accorder à des pratiques novatrices le temps dont elles ont besoin pour être maîtrisées, adaptées, évaluées et institutionnalisées » (Snow, 2015, p. 462). Le changement prend du temps et l'appropriation de

nouvelles pratiques nécessite jusqu'à cinq années de mise en œuvre d'un outil (Borman, Hewes, Overman, & Brown, 2003). Aux États-Unis, Borman *et al.* (2003) ont réalisé une méta-analyse sur le changement à partir de vingt-neuf programmes d'intervention. Si les programmes gagnent en efficacité sur le versant des apprentissages des élèves et sur celui du développement professionnel lorsqu'ils sont mis en œuvre sur une longue période, les effets observés sur les performances des élèves restent faibles ($d = 0,15$).

Considérant ces questions de temps, nous mesurerons les effets de l'intervention sur les apprentissages des élèves plusieurs fois par an et sur plusieurs années.

6.2.7. Une réelle appropriation

Pour Coburn (2003), la généralisation d'une innovation suppose un partage des croyances, des normes et des principes qui la sous-tendent. Diffuser les scripts didactiques et les matériaux associés ne suffit donc pas : pour « passer à l'échelle » avec succès, les caractéristiques des dispositifs doivent être intégrées aux schèmes professionnels de la communauté professionnelle à laquelle elle (ou il) s'adresse et ne doivent plus être portées par des membres externes à cette communauté tels que les chercheurs. Cette appropriation (au sens « d'intégration partielle ou totale à la culture propre de l'acteur », Theureau, 2011) doit dépasser la simple acceptation tacite et se transformer en une approbation plus profonde, plus large et plus substantielle (McLaughlin & Mitra, 2001). On peut alors poser la question suivante : quelles sont les conditions indispensables à réunir pour soutenir une mise en œuvre durable d'un dispositif innovant ?

Selon plusieurs chercheurs (Bryk, 2015; Cèbe & Goigoux, 2018; Gentaz, 2019; Goigoux & Cèbe, 2009; Klingner *et al.*, 2013), le meilleur moyen d'y parvenir est de l'amorcer dès le début du processus de conception, en collaboration avec les enseignants, pour anticiper sa possible intégration à leurs habitus professionnels. En d'autres termes, les concepteurs doivent être attentifs aux continuités avec les pratiques existantes tout autant qu'aux ruptures introduites par l'innovation. C'est la position adoptée par l'équipe dans laquelle nous effectuons notre recherche doctorale.

Dans le cadre de cette étude, nous faisons l'hypothèse que la qualité de la démarche de conception des outils détermine la réussite ou l'échec de leur diffusion. Le processus de conception continuée dans l'usage permet d'ajuster l'outil au plus près des utilisateurs et

anticipe ainsi le futur passage à l'échelle. Autrement dit, ce processus de conception intègre des dimensions qui devraient faciliter la diffusion de l'outil.

7. Une alternative : la proposition de l'équipe clermontoise

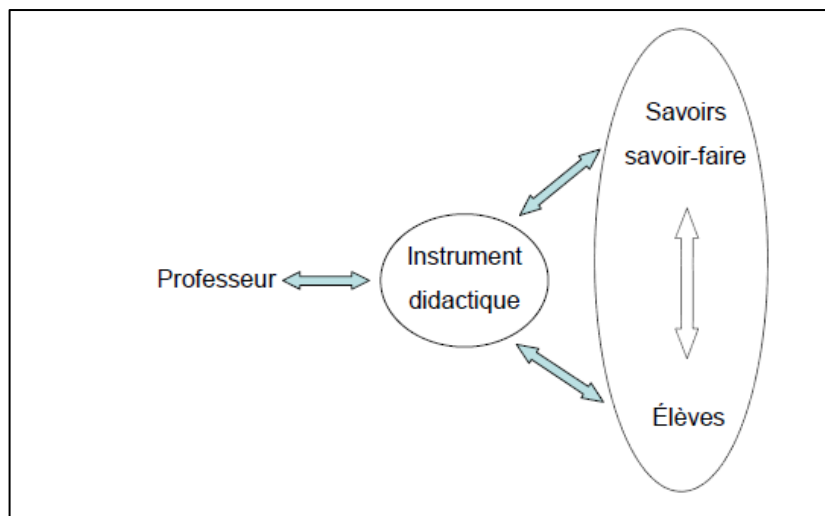
Dans les années 90, les partisans de l'éducation fondée sur la preuve ont cru que les dispositifs testés et évalués positivement en contexte expérimental pouvaient être diffusés aussitôt en vue de générer des améliorations substantielles. Or, il existe aujourd'hui un consensus pour contester cette vision trop linéaire. Malheureusement, lorsque les expérimentations échouent, seuls la formation et l'accompagnement insuffisants des enseignants sont incriminés (Gentaz et al., 2013; Thurler, 2000; Zorman et al., 2015), pas la pertinence des outils proposés.

Goigoux et Cèbe (2018; 2009) proposent une alternative à ce raisonnement : selon eux, c'est parce qu'ils sont mal conçus (en laboratoire et sans modèle de l'utilisateur) que les outils ne marchent pas ou se diffusent mal. Ils préconisent d'employer une méthode différente, la conception continuée dans l'usage inspiré de l'ergonomie de conception (Béguin & Cerf, 2004). Leur approche est voisine du mouvement *Design-Based Implementation Research* (DBIR) (Cobb, Confrey, Disessa, Lehrer, & Schauble, 2003; Fishman et al., 2013). Elle rejoint aussi les positions de Gentaz (2018) qui recommande de « mobiliser conjointement les acteurs du monde de la recherche et ceux du monde éducatif ». Si l'on veut favoriser la mise en place d'interventions efficaces à grande échelle, il faut que les outils soient co-construits par les chercheurs et les enseignants.

Goigoux et Cèbe interviennent dans le champ de la didactique définie comme une science de l'intervention et de la conception (d'outils, de dispositifs d'enseignement et de formation, ou de curriculum) visant l'amélioration de la qualité du travail enseignant et des apprentissages scolaires. Leur programme de recherche est à finalité technique (identifier les caractéristiques des outils qui sont vecteurs d'amélioration et laissent des marges de manœuvre aux enseignants) et à finalité méthodologique (conceptualiser et conduire le processus de conception). Il associe les enseignants à la démarche de conception de leurs instruments en organisant, comme les ergonomes (Barcellini, Van Belleghem, & Daniellou, 2013), un dialogue entre opérateurs et concepteurs au service du développement de leur activité. Il emprunte divers concepts et méthodes à la psychologie ergonomique (Leplat, 2000) et à l'ergonomie de conception (Daniellou, 2004; Pinsky & Theureau, 1992) ; il repose sur une analyse fine des

pratiques professionnelles pour construire et faire évoluer un modèle de l'utilisateur-enseignant (Rabardel & Pastré, 2005) qui complète celui de l'élève étudié par les sciences cognitives. Autrement dit, pour concevoir de nouveaux instruments, les chercheurs s'efforcent de tenir compte des connaissances scientifiques sur les apprentissages scolaires mais aussi sur l'activité d'enseignement et sur les savoirs des professeurs, leurs publics et leurs contextes de travail, comme l'illustre la Figure 1 (Goigoux et Cèbe, 2009).

Figure 1. Les connaissances utiles à la conception d'un outil didactique



Légende : l'instrument est au centre du triangle didactique. Sa conception exige de prendre simultanément en compte toutes les connaissances disponibles sur les savoirs en jeu, sur les élèves et leurs apprentissages ainsi que sur les professeurs appelés à l'utiliser (leurs conceptions et savoir-faire).

Les chercheurs se démarquent ainsi des démarches applicationnistes issues des sciences cognitives qui ne prennent en compte que le pôle apprentissage et des démarches d'inspiration ergonomique qui ne s'intéressent qu'à l'activité du professeur sans prendre en compte la spécificité des contenus et des processus des apprentissages scolaires (Boyer, 2019).

Partir des préoccupations des enseignants suppose de remettre en cause la logique actuelle des communautés de recherche qui décident, seules, des problèmes à résoudre en fonction du développement des connaissances scientifiques dans leur champ. Cela nécessite aussi de reconnaître qu'enseignants et chercheurs sont détenteurs et producteurs de connaissances d'égale valeur même si elles sont de nature différente. Dans le modèle de partenariat qui en découle, les chercheurs doivent faire l'effort de connaître les réalités de la pratique, les praticiens d'identifier la rigueur et les exigences de la recherche (Bressoux, 2017). Les problèmes rencontrés sont, en effet, à résoudre ensemble (Snow, 2015). En résumé, si l'on

veut que la formation par les outils ait la moindre chance d'atteindre le double objectif que les auteurs lui assignent – être acceptée par le plus grand nombre de maîtres et réduire les inégalités sociales d'apprentissage –, il faut que lesdits outils présentent plusieurs caractéristiques : ils doivent être, dès le départ, conçus pour répondre aux préoccupations des enseignants et aux besoins d'enseignement des élèves. Il faut aussi qu'ils puissent s'intégrer sans trop de bouleversements dans les conditions d'exercice ordinaire des maîtres débutants ou chevronnés. Il faut enfin qu'ils incluent des justifications théoriques et empiriques permettant aux utilisateurs de comprendre les principes qui les sous-tendent et la nature des activités proposées.

8. Conclusion

Ce premier chapitre nous a permis de déterminer le cadre méthodologique que nous souhaitons respecter : conjuguer la pertinence de l'expérimentation et la qualité de l'évaluation.

L'objectif principal de notre recherche est d'évaluer les effets de l'utilisation d'un outil sur les performances de compréhension d'élèves scolarisés en Réseau d'Éducation Prioritaire en comparaison avec un enseignement classique. Pour cela, nous appliquerons de la manière la plus rigoureuse possible une méthode quasi-expérimentale, à grande échelle et dans des conditions réalistes respectueuses des contextes ordinaires d'enseignement. Si la mise en œuvre de ce dispositif produit des effets bénéfiques sur les apprentissages des élèves, les recommandations de diffusion de l'outil seront légitimes.

Le second objectif concerne le soutien et le suivi habituellement proposés aux enseignants dans l'implémentation de l'innovation. Les auteurs de *Narramus* postulent que la qualité de la démarche de conception des outils détermine la réussite ou l'échec de leur utilisation et de leur diffusion.

Nous faisons de cette affirmation une première hypothèse que nous mettrons à l'épreuve en intégrant la question de l'accompagnement dans notre protocole expérimental : le soutien des enseignants qui utilisent *Narramus* est-il indispensable ? Est-ce qu'un accompagnement rend l'outil plus efficace ? Est-ce qu'il facilite son appropriation ? Est-ce que la nature singulière de *Narramus*, liée à son processus de conception, permet l'adoption de l'innovation sans autre forme de soutien ?

Chapitre 2

Des pratiques de lectures
partagées à un enseignement
explicite de la compréhension

Chapitre 2 Des pratiques de lectures partagées à un enseignement explicite de la compréhension

« Avant l'âge de la lecture, il y a l'âge où l'on nous raconte des histoires, l'âge où des plus grands nous choisissent des textes et les lisent pour nous » (Grossmann, 1996, p. 1)

Introduction

Dans ce chapitre, nous traiterons d'abord des activités de « lecture partagée¹⁴ » (*shared reading*) mises en œuvre dans les familles et à l'école puis recenserons les recherches montrant les effets différenciateurs qu'elles exercent sur le développement du langage oral et écrit des jeunes enfants. Cette analyse permettra d'expliquer pourquoi une intervention pédagogique visant l'enseignement de la compréhension et de la narration de récits peut jouer un rôle compensateur des inégalités sociales, rôle que notre thèse vise à évaluer.

Nous présenterons ensuite les caractéristiques de l'outil dont nous avons évalué les effets, *Narramus*, en nous centrant successivement sur :

- la méthode de conception particulière employée par les chercheurs,
- le choix du matériel utilisé,
- les principes pédagogiques et didactiques sur lesquels repose l'enseignement,
- les activités et les tâches proposées,
- le déroulé d'un scénario type,
- les spécificités de l'outil,
- la place de l'outil dans la prescription.

Nous terminerons par la formulation du second volet de notre problématique.

¹⁴ Nous qualifions de « lecture partagée » toute expérience dans laquelle les jeunes enfants d'âge préscolaire (avant 3 ans) ou scolaire (de 3 à 6 ans) écoutent un lecteur (un parent, une enseignante, un frère, une nourrice...) leur lire un livre (Hindman, Skibbe, & Foster, 2014).

1. Le développement du langage : une affaire de famille

L'acquisition du langage oral chez le petit d'homme correspond à un besoin inné (sauf cas pathologique). Personne n'a eu l'idée de proposer une méthode pour aider les parents à apprendre à parler à leur bébé (Florin, 2016) même si, dans ce domaine, ils sont les premiers et les plus importants « professeurs » (Bronfenbrenner, 1986). La plupart des familles proposent très tôt et régulièrement à leurs enfants des activités dites de « lecture partagée » mais celles-ci sont-elles partout de même nature, sont-elles également présentes et fréquentes dans tous les milieux sociaux ?

1.1. Les lectures familiales : les mêmes activités dans toutes les maisons ?

C'est à ces deux questions que répond l'étude de grande ampleur menée par Hindman et ses collaboratrices (2014) qui ont observé 700 dyades mère / enfant de toutes origines ethniques et socio-économiques¹⁵ vivant aux États-Unis. Les chercheuses ont d'abord interrogé les familles pour connaître la fréquence à laquelle elles proposaient ces activités de lecture. Leurs données font apparaître que, contrairement à une idée reçue et largement répandue dans les milieux éducatifs, les familles dans lesquelles on ne lit jamais constituent une exception¹⁶ (2 %) : un quart des mères déclarent lire une ou deux fois par semaine, 35 % de 3 à 6 fois par semaine et 38 % tous les jours. Selon les auteures, ce résultat s'explique par les efforts considérables faits aux États-Unis pour convaincre les parents que la lecture est l'un des moyens le plus efficace pour favoriser le développement langagier. Les conseils donnés sont toutefois assez généraux : on recommande aux parents de lire fréquemment et de mener une discussion sur l'histoire et les illustrations sans leur donner d'indications sur ce dont il faut discuter et pourquoi.

L'étude de Hindman *et. al* (2014) visait surtout à caractériser les activités menées et la nature des échanges. Aussi les dyades ont-elles été filmées à trois reprises (quand l'enfant avait neuf mois, deux ans puis quatre ans). Pour traiter leurs observables, les chercheuses ont constitué deux catégories : dans la première, sont rangées les interactions qui portent sur le code écrit (nom des lettres, correspondances lettres-sons, déchiffrage, etc.), dans la seconde celles centrées sur le sens de l'histoire et les illustrations.

¹⁵ 44 % des mères ne sont pas allées à l'école secondaire.

¹⁶ Ce constat est le même en France où les pratiques familiales de lecture partagée sont présentes quel que soit le milieu social d'origine (Frier, 2006).

Ce classement permet d'observer que les échanges sur le code alphabétique sont extrêmement rares (10 % dans lesquels la mère invite l'enfant à lire un mot ou un groupe de mots) et ceux sur les lettres ou les sons le sont plus encore (moins de 1 %). Ce résultat complète celui d'Evans et Saint Aubin (2005) qui avaient observé les mouvements des yeux de jeunes enfants auxquels on lisait un album. Cette expérience révélait que 93 % des regards étaient portés sur l'image et 7 % sur l'écrit quels que soient l'arrangement spatial du texte, la nature des illustrations et leur attractivité supposée (couleurs ou noir et blanc). On peut déduire de ce qui précède que, quel que soit le milieu social, pour les enfants comme pour les adultes, les activités de lecture partagée n'ont pas pour fonction de découvrir (ou de faire découvrir) le code, mais le sens de l'histoire.

L'étude d'Hindman fait aussi apparaître que certaines pratiques sont utilisées par la plupart des mères : 85 % lisent en dirigeant l'attention des enfants sur l'image et 64 % produisent une expansion de l'histoire. D'autres sont moins communes : 46 % font des liens avec la vie de l'enfant ou des livres qu'il connaît, 43 % mettent en scène l'histoire, 16 % discutent sur la manière de la mémoriser, la rappeler ou la résumer et seulement 14 % expliquent le vocabulaire.

Les auteures avaient un dernier objectif, celui de démêler le poids de trois variables : le niveau scolaire atteint par la mère, l'appartenance ethnique et la langue parlée à la maison. Leurs résultats mettent en évidence que si la première exerce un effet significatif sur la nature des échanges, les deux autres n'affectent pas de manière indépendante la gamme d'expériences proposées aux enfants. On apprend enfin que ce sont les mères le plus diplômé qui centrent le plus l'attention sur le sens de l'histoire, pratique qui est corrélée avec un accroissement du vocabulaire chez l'enfant.

Mais qu'en est-il en France où, jusqu'à très récemment, les campagnes médiatiques cultivaient plus la méfiance envers les écrans que la valorisation des lectures partagées ? Observe-t-on les mêmes types de pratiques et les mêmes différences ? L'étude dirigée par Bonnéry (Bonnéry, 2014; Bonnéry & Mamede, 2013) répond en partie à la question.

1.2. Que nous apprennent les recherches sur les apprentissages enfantins ?

Les résultats que nous rapportons ici portent sur les effets des activités de lectures partagées : littératie émergente, littératie précoce, compétences préparatoires à la lecture, compétences de

pré-lecture¹⁷ sont les termes employés dans la littérature scientifique pour désigner les apprentissages enfantins qui touchent les connaissances relatives au langage oral et au langage écrit (Sénéchal, Whissell, & Bildfell, 2017). Compte tenu de l'importance que revêtent les compétences précoces en lecture dans la réussite scolaire ultérieure et étant donné l'impact des pratiques parentales sur leur développement, nous supposons que, bien que la lecture conjointe ne vise pas à apprendre aux jeunes enfants à décoder, « elle joue toujours un rôle clé dans le processus d'alphabétisation » (Teale & Sulzby, 1999, p. 147). Reste à comprendre ce que ces pratiques permettent d'apprendre.

1.3. Une recherche française

Les chercheurs de l'équipe de Bonnéry ont eux-aussi filmé les modes de faire de trois types de familles (populaires, aisées ou cultivées) qui partageaient des temps de lecture avec leurs enfants : « les familles populaires avec peu de capitaux, les familles aisées où le capital économique est plus important que le capital culturel, à l'inverse des familles cultivées¹⁸ » (Bonnéry & Mamede, 2013, p. 85). Quatre profils de lecture se dégagent de leurs analyses.

Profil 1 : les lectures « populaires - oralisantes »

Les parents n'interrompent pas leur lecture et échangent peu avec leur enfant après qu'ils ont fini de lire. Ils apportent peu de connaissances extérieures au livre, expliquent rarement le vocabulaire, ne font pas de lien avec des connaissances culturelles ou encyclopédiques et reformulent peu les passages complexes : « leur lecture traite le texte comme autosuffisant » (p. 86). On observe aussi qu'ils relisent de la même manière, sans apport de nouvelles informations considérant, sans doute, que le sens des albums est directement accessible par la simple oralisation du texte.

Profil 2 : les lectures « populaires - restitutives »

Ici, les familles centrent essentiellement l'attention de l'enfant sur les éléments explicites du texte et sur un mode qui s'apparente à une interrogation scolaire. On invite ce dernier à reformuler le texte, pas à pas, avec des questions fermées qui appellent des réponses factuelles (et non interprétatives), questions qui sont suivies par une correction immédiate visant à élaborer une « interprétation univoque ». Après la lecture, l'adulte et l'enfant échangent un peu

¹⁷ *Emergent literacy, early literacy, reading readiness skills, pre-reading skills*

¹⁸ Les auteurs distinguent les familles « aisées » dans lesquelles le capital économique est plus important que le capital culturel des familles « cultivées » qui présentent un profil inverse.

plus longtemps que ceux du premier type, essentiellement sur leur ressenti. « L'adulte, par des questions fermées et des reformulations, guide ici l'enfant vers le repérage de l'explicite et l'interprétation univoque qu'il juge pertinente » (p. 89).

Profil 3 : les lectures « cultivées - élaboratives »

Les parents qui présentent ce profil invitent l'enfant à s'interroger sur le sens de l'histoire et des images, à formuler des hypothèses sur la suite de l'histoire, à chercher les indices qui les accréditent, à exprimer les émotions provoquées par le texte. Ils lui demandent aussi d'argumenter et de justifier ses positions. Ici, les relectures ne se ressemblent pas : elles apportent des niveaux successifs de compréhension, des déplacements du point de vue... Une attention particulière est portée sur ce que l'auteur veut faire penser au lecteur : sous-entendus, préparation du suspens, fausses-pistes...

Profil 4 : les lectures « aisées - motivantes »

Dans ce groupe, la lecture est régie par le plaisir du moment partagé et la découverte de l'histoire. Intermédiaires entre les profils 2 et 3, elles offrent un rapport à la lecture et à la culture qui est plus éloigné de la « culture cultivée » et des attentes de l'école. Les parents invitent les enfants à s'interroger sur différents niveaux de lecture mais, « il n'y a pas une systématisme dans la sollicitation de ces postures réflexives ni sur les références culturelles à transmettre » (p. 93).

Il ressort de ces deux études que, si on lit dans tous les milieux sociaux¹⁹, les manières de faire divergent²⁰. Les deux premiers profils s'observent majoritairement dans les familles de type 1, les familles populaires avec peu de capital culturel et économique. On peut donc supposer que ces différentes modalités de lecture ne produisent pas les mêmes effets sur le développement du langage oral et écrit²¹ et qu'elles contribuent à expliquer les différences de performance des élèves en compréhension de textes entendus à l'entrée au cours préparatoire.

¹⁹ Rappelons toutefois la donnée avancée par Delahaye (2015) : à l'entrée à l'école maternelle, certains enfants ont bénéficié de 1000 heures de lecture, d'autres d'aucune.

²⁰ Ce constat n'est pas nouveau, les travaux menés par Heath (1982) avaient déjà montré que les manières de lire étaient fortement associées au milieu social d'origine.

²¹ Nous sommes toutefois bien consciente que ces diverses modalités de lecture participent toutes au développement du langage et de la littératie (Haden, Reese, & Fivush, 1996).

1.3.1. L'acquisition de connaissances spécifiques : le vocabulaire et les connaissances encyclopédiques

Au cours des dernières décennies, on a régulièrement démontré que les conversations quotidiennes permettent d'enrichir le vocabulaire de l'enfant (Sénéchal, LeFevre, Hudson, & Lawson, 1996), quelle que soit la situation socioéconomique des parents (Snow & Beals, 2006). Celles-ci lui offrent l'occasion d'entendre et d'utiliser de nouveaux mots et expressions en bénéficiant de *feed-back* à la fois sur leur compréhension et leur prononciation (Hindman, Wasik, & Bradley, 2019). Celles qui s'organisent autour de la lecture de textes sont particulièrement efficaces parce que le langage y est plus décontextualisé, le vocabulaire rencontré plus rare et élaboré (Sénéchal, 2000; Sénéchal, Thomas, & Monker, 1995), la syntaxe plus complexe (Wasik & Bond, 2001), et l'énoncé entendu plus stable. On a aussi régulièrement observé qu'entre 3 et 6 ans la lecture répétée d'une même histoire amplifie les effets d'apprentissage du lexique (Cunningham & Stanovich, 1998; Damhuis, Segers, & Verhoeven, 2015; McLeod & McDade, 2011; Montag, Jones, & Smith, 2015; Snow, 1991).

Mais on sait aussi que la multiplication des expériences de lectures à haute voix ne suffit pas toujours pour accroître le stock lexical, en réception et en production, de tous les enfants (Florin, 2010).

En nous basant sur les résultats de Bonnéry *et al.* (2013), on peut supposer qu'à temps de lecture équivalent ceux qui grandissent dans une famille du profil 1 devraient apprendre moins de vocabulaire que leurs congénères. L'étude longitudinale menée par Le Normand, Paris et Cohen (2008) confirme cette hypothèse : les enfants de milieux populaires connaissent moins de noms et de verbes que leurs camarades de milieux favorisés et produisent des phrases moins longues. Ces différences, visibles dès l'âge de 2 ans, se creusent les années suivantes.

On pourrait cependant soutenir que les tout jeunes enfants apprennent le sens de mots inconnus de manière incidente (en entendant des conversations ou des histoires, en regardant la télévision) c'est-à-dire sans que le langage leur soit directement adressé et sans que les objets soient nommés ou les mots expliqués (Akhtar, 2005 ; Robbins & Ehri, 1994).

La méta-analyse réalisée par Flack et ses collaborateurs (2018) permet, selon nous, de trancher entre ces deux hypothèses au profit de la première. Après avoir constaté que, sans aide et sans enseignement spécifique, « les enfants sont capables de comprendre un peu moins de la moitié des nouveaux mots du texte qu'on leur lit » (p. 17), ils observent que le « style de lecture » employé *via* l'utilisation de plusieurs techniques dialogiques différentes – décrire

l'image, pointer l'image du mot avec le doigt, faire répéter le mot, fournir des explications avant ou pendant la lecture – influent considérablement sur la quantité de mots appris. Whitehurst, Falco, Lonigan *et al.* (1988) observent que, quand les parents discutent avec l'enfant sur le sens des mots et les leur font répéter, les gains en vocabulaire expressif sont plus importants. La nature des interactions entre les enfants et les adultes exerce donc une influence dans le développement du vocabulaire (Canut & Vertalier, 2012; Grossmann, 2001). Les enfants qui sont activement engagés dans une lecture de livre avec un adulte apprennent plus de vocabulaire que ceux qui l'écoutent passivement (Sénéchal *et al.*, 1995).

De son côté, Sénéchal (1997) compare les effets de trois pratiques contrastées proposées à des enfants de 3 et 4 ans : 1° lecture unique, 2° lectures réitérées du même album, 3° lectures réitérées + questions adressées à l'enfant (qu'est-ce que c'est ? Montre-moi où est le...). Ses résultats montrent que les lectures réitérées (2 et 3) augmentent la probabilité d'accroître le vocabulaire en réception mais celles qui donnent un rôle actif à l'enfant (3) exercent un effet supplémentaire sur le vocabulaire en production (l'enfant ayant eu à récupérer en mémoire les représentations phonologiques des mots stockés en mémoire).

Multiplier les occasions de lire pour réduire les différences lexicales initiales permet donc bien d'augmenter la probabilité que les enfants rencontrent des mots qu'ils ne connaissent pas, mais rien ne garantit que ceux-ci profitent de l'occasion pour en apprendre le sens (Baker, Simmons, & Kame'enui, 1995; Compton, Elleman, Olinghouse, Gilbert, & Gilbert, 2009).

Sachant que le vocabulaire est le premier prédicteur de la compréhension chez les jeunes enfants (Kendeou, Bohn-Gettler, White, & Van Den Broek, 2008; Stanley, Petscher, & Catts, 2018), se pose la question de savoir comment l'enseigner de manière efficace et durable.

Les résultats des études les plus récentes plaident en faveur d'une approche « intégrative » dans laquelle les mots et les expressions sont présentés dans des récits complets et cohérents : on a, en effet, régulièrement montré que cette méthode est plus efficace à court et à long terme que les leçons de vocabulaire « décrochées » portant sur l'enseignement de mots isolés ou catégorisés (*cf.* la revue de question menée par Snell, Hindman, & Wasik, 2015, et la méta-analyse de Wright & Cervetti, 2017).

C'est cette approche intégrative que les auteurs de *Narramus* ont retenue et que les anglo-saxons désignent par les termes *embedded instruction* (Goldstein *et al.*, 2016). Aussi chercherons-nous, dans le cadre de notre thèse, à savoir si leurs propositions produisent bien

les effets escomptés. Nous évaluerons donc les acquisitions lexicales comme étant une manifestation de l'effet d'un scénario didactique basé sur des pratiques d'enseignement intégrées à l'étude de textes complets.

D'autres données prouvent que, pour accroître les effets de l'enseignement, il faut multiplier les modalités de présentation et d'explication (Wright & Cervetti, 2017) ainsi que les rencontres avec les mêmes mots (Loftus-Rattan, Mitchell, & Coyne, 2016; Wasik, Hindman, & Snell, 2016). Les auteurs de *Narramus* ont aussi inclus, dans leurs scénarios, plusieurs manières d'expliquer le vocabulaire et de multiples tâches de reprises, de révisions et de remobilisations. Dans notre évaluation, nous examinerons si ces choix produisent les effets attendus.

1.3.2. L'acquisition de connaissances spécifiques : les compétences narratives en réception

Il ne suffit pas, bien sûr, que l'enfant connaisse le sens de tous les mots d'un texte pour le comprendre. Les interventions de lectures partagées qui sont à l'origine des progrès les plus solides en compréhension incluent des discussions visant à favoriser un traitement en profondeur du texte étudié et la réalisation, par les enfants, de multiples activités basées sur des tâches cognitives de haut niveau (Blewitt, Rump, Shealy, & Cook, 2009; Elleman, Lindo, Morphy, & Compton, 2009; Reese, Cox, Harte, & McAnally, 2003; Sigel, 1982). Ces activités sont privilégiées par les parents de milieux favorisés qui ont tendance à employer un discours non immédiat (plus décontextualisé) incluant des discussions extra-textuelles par exemple quand ils demandent à l'enfant d'expliquer des idées, de produire des hypothèses et des inférences, de lier les contenus avec ses propres expériences, de résumer l'histoire ou d'en rappeler les idées importantes (Blewitt et al., 2009; Bonnéry & Joigneaux, 2015; De Temple & Snow, 1996; Ganea, Pickard, & DeLoache, 2008; Girolametto, Weitzman, van Lieshout, & Duff, 2000; Justice & Ezell, 2002).

En revanche, dans les familles populaires, les conversations autour de l'album sont plutôt de l'ordre de la paraphrase. Elles sont aussi plus descriptives, concrètes ou contextualisées, la majorité des échanges portant sur la description des illustrations, des personnages et de leurs actions (Aukrust & Rydland, 2011; Bonnéry & Joigneaux, 2015; Dickinson & Tabors, 2001; Korat, Klein, & Segal-Drori, 2007).

Quand on met ces pratiques éducatives en lien avec les compétences requises pour comprendre les textes, on n'est pas étonné d'observer que les enfants de milieux favorisés sont très tôt plus performants que leurs camarades moins bien lotis. Ces compétences sont de trois

ordres :

1. les connaissances stockées en mémoire : connaissances encyclopédiques, vocabulaire (étendue et qualité), morphologie, syntaxe, structures textuelles ;
2. l'efficacité cognitive générale : attention, mémoire de travail, raisonnement ;
3. les habiletés propres au traitement des discours continus : traitement de la phrase (cohérence locale), intégration et organisation sémantiques, production d'inférences fondées sur le texte et sur les connaissances antérieures, auto-évaluation de la compréhension et régulation.

Touchant la production d'inférences, Dunn, Brown, Slomkowski, Tesla et Youngblade (1991) ont établi que les enfants qui, à 3;3 ans, sont capables d'expliquer les sentiments et les actions de personnages fictifs (poupées) sont ceux qui ont très tôt été impliqués dans des conversations familiales portant sur les états mentaux et la causalité psychologique. Certains parents profitent des lectures partagées pour expliquer à leur enfant :

1. les états mentaux des personnages c'est-à-dire ce qu'ils ressentent (surprise, tristesse, peur, bonheur...), désirent, ce dont ils ont besoin ou ce dont ils manquent,
2. les états cognitifs des personnages : ce qu'ils savent, croient, pensent, se rappellent, supposent, imaginent, espèrent.

Ceux qui incitent régulièrement les jeunes enfants à s'interroger sur ce que pensent et ressentent les protagonistes leur font prendre l'habitude de se poser ce type de questions et de chercher à y répondre (Symons, Peterson, Slaughter, Roche, & Doyle, 2005). Garner, Jones, Gaddy et Rennie (1997) ont demandé à des mères de discuter d'un livre d'histoire sans texte avec leurs enfants et ont étudié les références qu'elles faisaient aux états mentaux concernant les émotions des personnages. Ils concluent que les enfants dont les mères expliquaient les causes et les conséquences des émotions réussissaient mieux dans les tâches standard de compréhension des émotions que ceux dont les mères n'y faisaient jamais référence. Le Sourn-Bissaoui et Deleau (2001) démontrent que les enfants qui, dans les situations de lecture partagée, ont été invités à discuter et à interpréter les états mentaux et cognitifs des personnages développent des connaissances sur des éléments de la psychologie humaine. Pour Deleau (2007), tous les enfants n'entrent pas à l'école avec les mêmes habiletés préparatoires à la théorie de l'esprit et sont donc différemment réceptifs au travail mené sur les états mentaux en classe. Partant, certains élèves ne sont pas en mesure de comprendre les apprentissages qui

nécessitent de les mobiliser. Aussi, Deleau (2007) invite-t-il les enseignants à mener des activités spécifiques centrées sur le développement des états mentaux des élèves.

Cette idée est soutenue par Nader-Grobois (2011) qui propose un éventail de cibles d'intervention susceptibles d'influencer le développement de la théorie de l'esprit en milieu scolaire. L'auteure précise qu'il est possible d'entraîner simultanément plusieurs de ces compétences. Ces propositions peuvent être rangées selon qu'elles concernent plus des connaissances, des compétences, des fonctions exécutives générales.

Ces pratiques éducatives familiales sont très efficaces : les enfants habitués à réfléchir aux états mentaux des personnages sont ceux qui comprennent le mieux les textes narratifs précisément parce que ceux-ci sont lacunaires : ils sollicitent la coopération du lecteur pour lever l'implicite qui relève surtout des intentions des divers personnages. C'est ce que montre par exemple la recherche longitudinale menée par Atkinson, Slade, Powell et Levy (2017) : une bonne théorie de l'esprit à quatre ans prédit la compréhension en lecture à six ans et demi.

C'est pourquoi on n'est pas étonné de constater que les enfants de milieux populaires accusent un retard par rapport à leurs camarades plus favorisés dans ce domaine (Korat et al., 2007; Weimer & Guajardo, 2005) : sur ce plan aussi, les pratiques parentales sont différenciatrices.

En résumé, les recherches sur les pratiques éducatives familiales montrent que les enfants dont les parents explicitent les états mentaux des protagonistes de récits sont mieux préparés à la compréhension de textes qu'ils vont ensuite apprendre à lire seuls. Sur ce point encore, les auteurs de *Narramus* sont en accord avec les conclusions des recherches internationales : ils visent les mêmes buts que les pratiques parentales efficaces mais par des moyens pédagogiques. Nous chercherons à savoir s'ils ont eu raison.

Conclusion

Les occasions offertes aux jeunes enfants pour développer le langage oral et la compréhension de l'écrit prennent donc des formes différentes selon les pratiques éducatives familiales et produisent nécessairement des effets contrastés. C'est pourquoi chercheurs et politiques pensent que des interventions scolaires précoces pourraient éviter que les différences observées à l'entrée à l'école maternelle se transforment en difficultés, s'enracinent et deviennent alors difficiles à surmonter (Coyne, Simmons, Kame'enui, & Stoolmiller, 2004).

Si cet objectif est consensuel, les moyens pour l'atteindre – principes pédagogiques et didactiques, activités proposées, temps alloué et modalités d'organisation sociale – ne sont pas les mêmes.

1.4. Qu'en est-il à l'école maternelle ?

Une donnée est aujourd'hui bien établie : tous les enseignants de maternelle proposent des activités de lecture partagée (plus communément désignée par les termes « lecture offerte »). Ceci n'est pas surprenant : le programme le leur enjoint et le MEN conseille une liste d'albums adaptés à chaque section. Mais comment s'y prennent-ils réellement ?

1.4.1. Les styles de lecture

Pour comprendre les expériences de lecture partagée, il faut tenir compte des comportements du lecteur adulte et des enfants participants. Les pratiques de lecture directement observées dans des classes à l'école maternelle sont peu documentées dans la littérature. L'étude de Bastide et Joigneaux (2014) apporte toutefois des éclairages pertinents pour notre propos. Les chercheurs se sont intéressés aux liens existants entre les modalités de lecture et la nature des albums dans des classes de petite, moyenne ou grande sections. À partir d'observations filmées (deux heures²² par classe en moyenne) et d'un corpus support de douze albums, les chercheurs distinguent quatre types de lectures conjointes mises en œuvre actuellement à l'école maternelle.

La lecture oralisation

Ce premier type regroupe les lectures « linéaires », « intégrales » et « continues ». Celles-ci vont « rarement au-delà d'une oralisation, assez linéaire, du texte des albums, dans son intégralité et sa littéralité ». L'enseignant respecte l'écriture et n'interrompt pas la découverte du texte par des commentaires ou des questions. Il procède à une mise en voix très fidèle à la ponctuation et à la syntaxe des phrases, sans mobilisation du corps et des éléments présents dans les illustrations. « Tout se passe comme si ces lecteurs considéraient qu'une telle lecture est suffisante pour que les élèves partagent leur interprétation du texte de l'album, ou construisent la leur de manière autonome » (p. 11). Les échanges entre le lecteur et les auditeurs sont très rares. Le premier ne répond quasiment jamais aux sollicitations ou remarques

²² Avant de procéder aux enregistrements filmés, les auteurs avaient observé un plus grand nombre de situations de lecture pour en appréhender la variété.

effectuées par les seconds, que celles-ci soient effectuées au cours de la lecture ou à la fin de l'histoire. Cette manière de lire est proche de celle des enseignantes « lectrices » (au sens de Grossmann, 1996) qui effectuent des lectures strictement littérales du texte sans ajout, suppression ou expansion.

La lecture mise en scène

Une des caractéristiques de ce deuxième type de lecture est la mobilisation du mime facial et corporel. L'interprétation du texte est portée visuellement par le mime des actions ou des émotions. Les illustrations jouent un rôle secondaire et ne sont montrées qu'après la phase de mime. Le sens du texte est principalement porté par le langage corporel. « Le support lui-même devient accessoire dans la transmission du texte, puisque l'essentiel passe par le corps » (p. 12). La lecture mise en scène constitue cependant bien une entrée dans l'écrit puisque, avant et après la lecture, le lecteur fait notamment référence au support en utilisant des termes associés à l'univers de l'écrit (auteur, couverture...). En outre, celui qui lit revient parfois au texte, « objet stabilisé par l'écriture servant de référence au mime. »

La lecture dialogique

Comme dans la lecture mise en scène, « le langage oral est ici utilisé pour traduire le sens de ce qui est lu. ». Dans ce troisième type de lecture, un accent particulier est porté sur la prosodie et le rythme du débit oral. Ces techniques permettent à l'enseignant 1° d'insister volontairement sur certains passages ou événements « clés » identifiés au préalable et 2° de favoriser l'attention conjointe. Ici, le geste est secondaire et ne prend que ponctuellement le relais de l'oral. Les illustrations, toujours visibles par les élèves (le livre n'est jamais orienté vers le lecteur), servent d'appui pour montrer ou échanger sur des passages du texte : « à l'approche de mots "stratégiques" pour la compréhension », celui qui lit ralentit son débit et l'interrompt pour interroger les élèves et élaborer le sens de ces mots en contexte.

La lecture enquête : quand le lecteur sème des indices que les élèves doivent collecter

Assez proche de la lecture dialogique, cette modalité mobilise les éléments picturaux mais également d'autres modes de communication. « Ainsi l'oralisation du texte est dramatisée non pas pour permettre aux auditeurs d'accéder relativement directement au sens de ce qui est lu, mais pour les inciter à le reconstruire à partir d'un faisceau d'indices plus ou moins convergents » (*idem*, p. 13). Les échanges entre élèves sont particulièrement sollicités prenant la forme de débats interprétatifs. Le lecteur cache sciemment certains indices (par exemple, en

ne répondant pas immédiatement aux sollicitations des élèves ou en différant la présentation des illustrations). Dans ce type de lecture partagée, « l'album est utilisé comme support d'entrée non seulement dans l'écrit mais aussi la littérature, au moyen d'un système de mise en relations internes (entre des éléments présents sur des pages différentes ou entre le texte et l'image) et externes (à d'autres histoires ou personnages [...]), mettant ainsi en avant l'importance du sens et de l'existence de liens entre les éléments appartenant à la sphère de l'écrit (support, composantes (texte, images...), registre narratif de ce type d'écrits, personnages, références à d'autres lectures...) » (p. 14).

Conclusion

L'analyse menée par Bastide et Joigneaux montre qu'il existe différents styles de lectures partagées à l'école maternelle²³. En outre, les diverses intentions portées par le lecteur dans la narration d'histoire influencent probablement les manières de lire. Grossmann (1996) distingue deux tendances principales : « soit il s'agit d'un temps faible, c'est-à-dire peu didactisé, pouvant viser le retour au calme (après une activité intense, physiquement ou intellectuellement), ou la récompense lorsqu'on a été sage) ». La lecture d'histoire est alors souvent un moment relativement autonome, peu relié à ce qui précède. « Soit il s'agit d'un temps peu autonome parce qu'il sert de point de départ à une série d'activités didactiques parfois très diverses : rappel d'histoire, lecture, écriture, dessin, mime... » (p. 75).

On peut donc supposer que, selon la nature et la diversité des modalités de lectures conjointes mises en œuvre par les enseignants, tous les élèves ne construisent pas les mêmes connaissances et pratiques de l'écrit et que « différentes techniques et styles de lecture peuvent avoir des effets différentiels sur le développement du langage des élèves » (Mascareño, Deunk, Snow, & Bosker, 2017, p. 137).

1.4.2. Les styles d'interaction

Au-delà des conduites de lecture, les interactions verbales diffèrent dans les moments de lectures partagées.

²³ Aux États-Unis, Martinez et Teale (1993) avait déjà établi le même constat : les conduites de lecture partagées mises en œuvre par les enseignants sont variées. Malgré cette diversité, les auteurs indiquent que chaque enseignant conserve un style relativement constant à travers les différentes lectures qu'il effectue.

1.4.2.1. Styles d'interaction en cours de lecture

Les travaux de Dickinson et Keebler (1989) et Grossmann (1996) distinguent deux principaux styles d'interaction dans les lectures conjointes : un style d'interaction « faible » et un style d'interaction « fort ».

Le style d'interaction faible

Le style d'interaction faible se caractérise par une lecture continue du texte avec très peu d'échanges entre les adultes et les enfants au cours de la lecture (Bastide & Joigneaux, 2014; De Temple & Tabors, 1994; Frier, 2006; Grossmann, 1996). Le lecteur n'instaure pas de dialogue, ne pose pas de question pendant la découverte du texte mais il ajoute des informations, effectue de nombreux commentaires ou développe certains passages de manière explicite au cours de la lecture. Dans ce style d'interaction, le texte est mis en position centrale et la compréhension est guidée par l'adulte.

On retrouve dans ce mode d'interaction les styles de lecture évoqués précédemment comme la lecture oralisation²⁴ (Bastide & Joigneaux, 2014) au cours de laquelle l'adulte lit le texte de manière fidèle et linéaire; la lecture mise en scène²⁵ (Dickinson & Keebler, 1989) où le texte est joué et le mime facial et gestuel est largement utilisé.

Le style d'interaction forte

Le style d'interaction forte se caractérise par des échanges nombreux et des questionnements au cours de la lecture. L'enfant est particulièrement actif puisqu'il est « poussé à réagir verbalement aux énoncés et, plus souvent aux images qu'il vient d'entendre ou de regarder, dans un rapport de dialogue continu avec l'adulte » (Cardarello, 1991, cité par Grossman, 1996, p. 26). Ce style d'interaction développe davantage les capacités d'expression et de réaction que d'écoute d'un texte lu. La communication est au centre (Dickinson & Keebler, 1989) et les discussions font l'objet de multiples reformulations. Partant, le texte est placé au second plan.

La lecture « dialogique » et la lecture « enquête » (Bastide & Joigneaux, 2014) se retrouvent dans le style d'interaction forte. Dans la première, les adultes jouent sur le rythme de lecture et l'interrompent régulièrement pour discuter de l'histoire, expliciter certains passages... Dans la seconde, les enfants sont sollicités pour s'impliquer de manière plus

²⁴ Ce style de lecture se rapproche également de la lecture simple évoquée par De Temple & Tabors (1994).

²⁵ La lecture « mise en scène » correspond aussi au style théâtral (Frier, 2006).

personnelle dans l'histoire et interpréter celle-ci. Le support est, en outre, questionné par ce type de lecture.

1.4.2.2. Conclusion

L'identification de ces deux styles d'interaction peut questionner : est-il souhaitable de privilégier un mode d'interactions au détriment d'un autre ? Pour Grossman (1996), le style d'interaction « forte » rend les élèves plus actifs dans la construction du sens, développe leur esprit critique et permet davantage de complicité entre enfants et adultes. Le style d'interaction « faible » joue davantage sur le côté « affectif », en développant l'imagination et la sensibilité des élèves. Dramatisé²⁶, ce mode d'interaction facilite l'appropriation syntaxique et prosodique lorsque les élèves écoutent plusieurs fois la même histoire. Toutefois, aucun style n'est universellement efficace (Teale, 2003).

1.4.3. Le rapport de l'Inspection générale (2011)

En France, on ne connaît pas d'étude de grande ampleur basée sur l'observation des pratiques effectives des enseignants de maternelle. C'est pourquoi, en sus des recherches précédemment citées, nous recourons au rapport des Inspecteurs généraux de l'Éducation nationale (2011) qui dresse un constat assez sombre de l'état de l'enseignement de la compréhension en lecture à l'école maternelle. Dans les 97 écoles qu'ils ont visitées, les IGEN n'ont observé aucune activité en lien avec ce domaine. Sur les 221 rapports d'inspection qu'ils ont analysés, la compréhension n'est citée que 18 fois (soit 8,1 %) quand la « reconnaissance de mots » l'est dans près d'un tiers des cas (26,7 %). Les inspecteurs signalent que si tous les emplois du temps intègrent bien des moments de lecture « offerte », rares sont ceux qui affichent des « séances de travail dévolues à la compréhension des textes entendus, en eux-mêmes et dans la relation qu'ils entretiennent avec les images qui le plus souvent les illustrent » (p. 130). Ils constatent que les maîtres attirent souvent l'attention des élèves sur les illustrations dans le souci de les aider, mais « l'absence d'explicitation de ce qui permet de comprendre dans le texte ou des relations entre le texte et les images reste dommageable » (p. 131). Ils notent également qu'un véritable travail oral à propos de la compréhension de textes écrits est peu fréquent : les élèves sont très peu sollicités pour une « expression en continu » (récits, explications, etc. dans lesquels les enfants enchaînent plusieurs énoncés) ; ils sont peu impliqués dans des échanges réglés et organisés,

²⁶ La compréhension des récit est fortement influencée par cette lecture « dramatisée » du texte qui oriente les modes de réception des élèves (Boiron, 2010).

dans lesquels ils sauraient qu'ils apprennent à mieux parler, à mieux raconter, à mieux comprendre une histoire (IGEN, 2011).

Toutefois, les IGEN ne différencient pas lesdites pratiques selon le contexte d'exercice des enseignants (REP vs hors REP). S'ils l'avaient fait, aurait-on observé des différences ? Autrement dit, les maîtres de REP adaptent-ils leurs pratiques aux caractéristiques de leurs élèves ? Lisent-ils les mêmes albums et proposent-ils les mêmes activités ? Là encore, l'étude menée par Bonnéry *et al.* (2014 ; 2013) nous éclaire.

1.4.4. Lit-on les mêmes albums en REP et hors REP et propose-t-on les mêmes activités ?

Pour répondre à la première question, les chercheurs classent les albums en trois catégories²⁷ :

1. les « explicites » qui délivrent au lecteur l'essentiel des informations nécessaires pour comprendre le sens global de l'histoire.
2. Les « patrimoniaux » qui, s'ils sont aussi explicites, se distinguent des précédents par le fait qu'ils racontent des histoires qui font référence pour les générations précédentes,
3. Les « complexes » obligent le(s) lecteur(s) de déduire le (ou les) sens, parfois en déjouant les sous-entendus, les fausses pistes ou les pièges volontairement tendus par l'auteur.

Ils observent qu'à l'école maternelle ce sont les albums « explicites » (40 %) qui sont le plus utilisés, les « complexes » représentant 31 % de l'offre et les « patrimoniaux » 29 %. Mais la distinction en fonction du contexte scolaire fait apparaître que les écoles classées en REP proposent moins d'albums patrimoniaux (21 % contre 33 % hors REP), et plus d'albums « explicites » (48 % contre 37 % hors REP), mais autant d'albums « complexes » (31 %). Autrement dit, si on lit bien dans toutes les classes de maternelle, on ne lit pas tout à fait la même chose.

L'étude fait aussi apparaître que, quand les ouvrages étudiés sont les mêmes en REP et hors REP, les usages qu'en font les enseignants ne le sont pas : ceux qui exercent en REP ont tendance à poser des questions fermées portant sur le sens littéral du texte, à aborder les albums par leur thème (p. ex. on lit *Roule galette* au moment de l'Épiphanie avec très peu d'attention portée aux indices narratifs), là où, hors REP, ils sont plutôt traités pour leur aspect littéraire

²⁷ Dans deux autres études, Bonnéry et Joigneaux (2015) et Bonnéry, Crinon et Marin (2015) distinguent deux types idéaux d'albums : les albums « explicites » et les albums plus « implicites » ou indiciaires.

(où on lit *Roule galette* pour attirer l'attention des élèves sur la ruse du renard, figure typique). Toutefois, les auteurs signalent avoir fréquemment observé des échanges structurés par une articulation entre questions ouvertes et fermées qui ont pour enjeu de relayer les sollicitations faites aux enfants par l'album en termes de production d'inférences, de mobilisation des postures réflexives, etc. Selon eux, ce mode de faire place les élèves qui n'ont pas développé hors l'école les connaissances, les compétences et les postures de lecteur réflexif sont d'emblée placés hors-jeu de la situation didactique.

L'étude longitudinale menée par Bianco, Bressoux, Doyen, Lambert, Lima, Pellenq et Zorman (2010) montre de surcroît que les « lectures offertes » à des élèves de milieux populaires se soldent par des échecs quand elles ne sont pas accompagnées d'activités explicitement construites pour enseigner les compétences requises pour comprendre. Une partie de leur étude portait en effet sur un « entraînement à l'analyse d'histoires » basé sur la lecture et l'analyse répétées d'un livre d'histoires avec de petits groupes d'enfants de moyenne et de grande section sur une période relativement longue (4 à 8 semaines). « Cette intervention, bien qu'axée sur la langue, était implicite à la fois dans son format d'enseignement et dans la manière dont elle attirait l'attention des enfants sur leurs capacités de compréhension » (p. 216). Les enseignants avaient seulement reçu comme consigne « d'appliquer les principes des lectures partagées, c'est-à-dire de créer des conditions de lecture et d'écoute des histoires proches de celles observées en milieu naturel, comme la narration d'histoires à la maison » (p. 220). Les auteurs précisent « qu'aucune instruction n'a été donnée aux enseignants sur les compétences de compréhension nécessaires à prendre en considération » et qu'ils leur ont simplement dit de « lire à plusieurs reprises à haute voix les histoires ou des parties d'entre elles et d'encourager les discussions sur le texte et l'histoire tout en donnant suite, dans la mesure du possible, aux commentaires des élèves » (p. 220). Bien que deux livres d'histoires aient été étudiés en moyenne section (huit séances pour chaque livre) et trois autres en grande section (quatre séances pour le premier et six pour les deux autres), aucun transfert d'apprentissage n'a pu être identifié en post-test. En d'autres termes, les lectures offertes ne produisaient aucun effet sur la compréhension orale.

Ces résultats ont de quoi alerter quand on sait que les écarts observés à la fin de l'école maternelle dans le domaine de la littératie précoce se révèlent particulièrement stables (Catts, Adlof, & Weismer, 2006; Catts, Hogan, & Adlof, 2005; Å. Elwér, Keenan, Olson, Byrne, & Samuelsson, 2013; Nation, Cocksey, Taylor, & Bishop, 2010). Les enfants issus de

communautés minoritaires et d'environnements pauvres pâtissent plus que les autres de l'écart entre la façon dont la langue est utilisée chez eux et celle dont on use à l'école et dans la littérature de jeunesse (Bonnéry, 2014; Reardon, Valentino, & Shores, 2012; Uccelli, Galloway, Barr, Meneses, & Dobbs, 2015). L'étude longitudinale menée par Cabell, Justice, Logan et Konold (2013) qui porte sur des enfants d'âge préscolaire issus de milieux populaires révèle que, parmi ceux qui font partie du groupe le plus faible à l'automne, seuls 21 % se situent dans la moyenne au printemps (et ce sont ceux qui, au départ, avaient le niveau le plus élevé).

Ces observations permettent d'expliquer en partie les résultats de la recherche *Lire-Écrire* (Goigoux *et al.*, 2016) dont nous avons parlé plus haut. Si, à l'issue de la scolarité maternelle, les performances en compréhension sont fortement corrélées avec les caractéristiques sociales des élèves, c'est que l'enseignement de la compréhension – qui repose sur un ensemble quasi infini de compétences langagières et de connaissances culturelles – y est délaissé au profit de celui de compétences dites « finies » : la conscience phonologique, la compréhension du principe alphabétique et la connaissance des lettres (Spencer, Weddle, Petersen, & Adams, 2017).

Pris ensemble, ces résultats soulignent la nécessité d'interventions scolaires précoces, fondées sur des données probantes, pour faire acquérir aux élèves de milieux populaires les connaissances et compétences qui permettraient d'envisager la suite de leur scolarité avec confiance. Mais se pose alors la question des cibles à viser, des activités et des tâches à proposer.

1.4.5. Des pratiques éducatives familiales aux pratiques d'enseignement de l'école maternelle

Nous l'avons vu plus haut, il ne suffit pas d'enjoindre aux parents de milieux populaires de lire chez eux²⁸ et d'inviter les enseignants à singer les pratiques des familles favorisées pour obtenir une amélioration. Les deux contextes sont bien trop différents en termes de temps, de régularité, d'attention possible à l'enfant, de quantité de langage... Mais on peut soutenir que si les

²⁸ C'est pourtant l'idée portée aujourd'hui par le ministère de l'Éducation nationale dans sa circulaire de rentrée (2019). On peut y lire, en effet, que « *L'intensité de l'exposition des enfants au langage parlé et la qualité de ce langage sont essentielles dans ce processus d'appropriation. C'est l'une des raisons qui conduit à **encourager les parents** à engager le plus souvent possible des dialogues avec leur enfant ainsi qu'à lui lire des histoires* ». On notera que, dans un cas comme dans l'autre, on considère que le développement des compétences requises pour comprendre des textes emprunte la même voie que celui de l'oral : naturellement, par un simple accroissement de la quantité d'oral et d'écrit adressée à l'enfant.

pratiques des uns s'avèrent particulièrement efficaces c'est parce qu'elles reposent sur des principes psychologiques valides (et validés par la recherche) dans le domaine du développement du langage oral et du langage écrit.

Il faut donc chercher à développer des moyens qui sont propres à l'école et qui « didactisent » les buts, les cibles, les activités, les tâches et les manières de faire en tirant profit des recherches qui ont déjà fait la preuve de leur efficacité à l'école maternelle. C'est ce qu'ont cherché à faire Cèbe et Goigoux (2017) avec la conception d'un outil destiné aux enseignants : *Narramus*.

2. L'outil *Narramus*

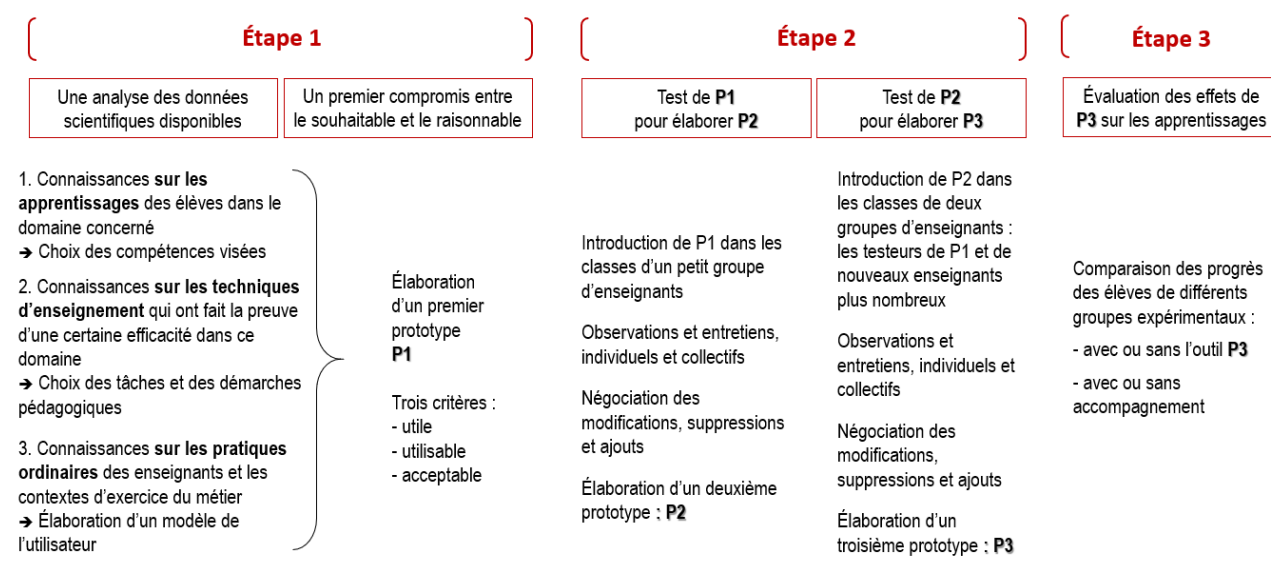
Nous l'avons écrit précédemment, la conception de l'outil *Narramus* s'inscrit dans une démarche de formation qui se démarque des dispositifs classiques de formation (Cèbe & Goigoux, 2013). Les auteurs postulent que la prise en main d'un nouvel outil qui associe des techniques pédagogiques innovantes minutieusement décrites et l'élucidation de leurs fondements théoriques contribue à accroître le pouvoir d'agir des enseignants (Goigoux & Cèbe, 2009).

2.1. Une conception particulière

La démarche de conception utilisée pour élaborer *Narramus* a 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* (Paul Cobb et al., 2003) ou, plus largement, *Educational Design* (Collins, 1992; McKenney & Reeves, 2014). Cette approche « intégrée à la pratique » (Goigoux, 2017) part des préoccupations des enseignants et est basée sur une relation réciproque entre les enseignants et les chercheurs (Snow, 2015).

Le processus de conception utilisé par les chercheurs peut être résumé en trois étapes (pour une présentation complète, voir Cèbe & Goigoux, 2018), comme le fait apparaître la Figure 2.

Figure 2. Les trois étapes d'une conception continuée dans l'usage (Schéma de Cèbe et Goigoux, 2018)



L'étape 1 consiste à effectuer une synthèse des savoirs disponibles. Pour concevoir *Narramus*, les chercheurs ont pris appui sur trois principales sources de connaissances :

- les contenus et les processus d'apprentissage des élèves dans le domaine de la compréhension chez les jeunes élèves ;
- les techniques qui ont fait la preuve de leur efficacité dans ce domaine ;
- les pratiques ordinaires des enseignants.

À l'issue de cette analyse, ils ont élaboré un premier prototype envisagé « comme un compromis entre ce qui pourrait paraître souhaitable du point de vue des apprentissages des élèves et ce qui semble raisonnable du point de vue de l'action pédagogique contextualisée » (*idem*, p. 83).

L'étape 2 consiste en une mise à l'essai du prototype. Les chercheurs ont fait tester pendant trois ans les prototypes de leur outil par des enseignants « co-concepteurs ». Ces derniers suggéraient à plusieurs reprises des modifications pour satisfaire aux principaux critères de l'ergonomie de conception : l'utilité, l'utilisabilité et l'acceptabilité de l'artéfact par ses futurs utilisateurs (Tricot *et al.*, 2003). À ce stade, les enseignants jouaient un rôle actif qui réduisait progressivement l'asymétrie initiale.

L'outil *Narramus* a donc été conçu selon un processus original qui nous amène à réinterroger les obstacles identifiés (chapitre 1) dans l'adoption et la diffusion des innovations. Notre questionnement est double :

1° cet outil co-conçu par des enseignants et des chercheurs et basé sur des fondements scientifiques attestés produit-il les effets escomptés (sur les pratiques des enseignants et les apprentissages des élèves) ?

2° sa mise en œuvre gagne-t-elle à être accompagnée par des formateurs issus des équipes de circonscription ?

C'est pour répondre à ces deux principales questions que nous avons mis en place une évaluation.

Cette dernière, qui constitue le socle de cette thèse, correspond à la troisième étape du processus de conception continuée dans l'usage. Il s'agit de tester l'outil en comparant les progrès des élèves de deux groupes expérimentaux (outillés, l'un avec accompagnement et l'autre sans) à ceux d'un groupe témoin (les pratiques d'enseignement ordinaires que nous venons de décrire).

Avant de présenter de manière détaillée la méthode utilisée pour élaborer notre protocole expérimental, nous allons présenter l'outil *Narramus* : les principes didactiques sur lesquels il repose, le matériel support, les objectifs visés par chaque scénario et les caractéristiques de l'enseignement.

2.2. Les connaissances et compétences visées par *Narramus*

Les principes didactiques et pédagogiques de *Narramus* ont été choisis pour soutenir le développement de quatre ensembles de compétences. L'enseignement est systématiquement basé sur des textes issus d'albums de littérature de jeunesse.

2.2.1. Des compétences narratives en réception

Les élèves découvrent l'histoire épisode par épisode afin de la comprendre en profondeur. Les enseignants les amènent à discuter le sens de l'histoire pour mieux l'interpréter (Blewitt et al., 2009). La présentation du texte précède toujours celle de l'image. La dissociation systématique du texte et de l'image oblige les élèves à traiter l'écrit et attire leur attention sur l'information qui est sémantiquement centrale (Brigaudiot, 2000). On a montré que si le texte et l'image sont

présentés simultanément, l'attention des jeunes élèves se porte davantage sur l'information visuellement attractive que sur l'information linguistique importante sur le plan sémantique (Justice, Pullen, & Pence, 2008). « Être privés de l'image dans un premier temps oblige les élèves à traiter l'écrit entendu et à fabriquer une représentation dynamique qui intègre les différents personnages, leurs intentions et leurs actions en contexte. Dans un second temps, les images que l'on a décrites et appris à observer servent de support à la mémoire pour faciliter le rappel de récit » (Cèbe & Goigoux, 2017, p. 10).

2.2.2. Des compétences narratives en production

L'activité de rappel, prise en charge alternativement par l'enseignant et les enfants, donne un but à la tâche d'écoute de la lecture oralisée. Pour raconter l'histoire, les élèves élaborent une représentation mentale cohérente de l'histoire (Duke & Pearson, 2002), mémorisent les idées du texte, planifient et organisent leur discours. De multiples activités leur permettent de jouer l'histoire et d'impliquer leur corps dans la mise en scène : théâtre, mimes, maquette, marottes... 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.

2.2.3. Des compétences lexicales et syntaxiques

Tous les scénarios proposent une étude systématique du vocabulaire. Les tâches d'enseignement, de mémorisation et de réutilisation régulière des mots et expressions enseignés soutiennent l'apprentissage. Elles prennent appui sur des contextes riches et variés et sollicitent des échanges avec les élèves autour des caractéristiques sémantiques des mots (Chi & Koeske, 1983). De nombreuses lectures ou narrations de l'histoire sont effectuées (avec l'album, avec la maquette, avec la projection des images, sans support...). Ces lectures répétées favorisent également la mémorisation du lexique (Snell et al., 2015).

2.2.4. Des compétences inférentielles

Les auteurs de *Narramus* proposent un ensemble de tâches scolaires centrées sur les états mentaux des différents personnages et l'explication des informations implicites. 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).

Un travail sur l'implicite et les inférences est proposé pour apprendre aux élèves à remplir les blancs laissés par l'auteur et à relier les informations délivrées par le texte avec leurs propres connaissances. La compréhension des histoires nécessite d'accéder aux pensées et aux motivations des personnages (Deleau, 2007; Kim & Phillips, 2014), c'est pourquoi des temps sont explicitement dédiés à l'étude des états mentaux de tous les personnages pour aider les élèves à mieux saisir et à interpréter leurs actions et leurs intentions.

2.3. Description de l'outil

Narramus propose différents scénarios basés sur des albums de littérature de jeunesse, adaptés à l'âge des élèves et comprenant 7 à 9 modules répartis sur quatre semaines d'enseignement. Chaque opus est composé de l'album étudié, d'un guide pédagogique qui précise les activités à mener en classe et d'un CD²⁹ avec les annexes numériques nécessaires à leur mise en œuvre. Toutes les activités sont finalisées par un seul et même objectif : être capable de raconter l'histoire tout seul aux autres élèves en classe, puis à la famille à la maison. « C'est pour mieux raconter que les enfants devront mémoriser le vocabulaire, acquérir de nouvelles tournures syntaxiques, retenir les idées principales, s'interroger sur les pensées des personnages et comprendre l'implicite du récit » (Cèbe & Goigoux, 2017, p. 5).

L'enseignement, très explicite et ritualisé, vise à permettre aux élèves de se familiariser avec la langue écrite et de mieux la comprendre.

2.3.1. Les objectifs

D'un scénario à l'autre, on retrouve les mêmes objectifs. Ils sont ainsi déclinés :

- comprendre un texte narratif, en profondeur ;
- mémoriser la syntaxe et le lexique (mots et expressions) et savoir les utiliser ;
- apprendre des connaissances sur le monde (encyclopédiques) ;
- apprendre à intégrer (à mettre en mémoire) les informations successives pour fabriquer une représentation mentale cohérente ;
- développer des compétences narratives pour pouvoir raconter l'histoire, tout seul, avec un support imagé (maquette, illustrations...) ;
- apprendre à produire des inférences, expliciter l'implicite ;

²⁹ Pendant l'expérimentation, le guide était au format PDF et le CD n'était pas encore disponible mais son contenu était téléchargeable par les enseignants engagés dans le protocole.

- s'interroger sur les états mentaux de tous les personnages (ce qu'ils savent, pensent, croient, ressentent, espèrent...);
- apprendre à interpréter³⁰.

2.3.2. Le déroulé d'un scénario

Chaque scénario est détaillé dans un guide du maître qui décrit avec précision le déroulement des modules. Un module correspond à un ensemble d'activités à mener mais il peut être découpé et réorganisé en plusieurs parties et peut se dérouler sur deux voire trois jours. Les enseignants disposant de l'outil ont été informés, dès le début de l'étude, qu'ils étaient libres d'organiser temporellement les modules comme ils le souhaitaient. Ils pouvaient proposer un module en un seul bloc mais avaient la liberté de le couper en différentes étapes. Le déroulé d'un scénario est stable et la découverte du texte se fait épisode par épisode.

- *Avant la découverte de l'épisode*

À l'exception du premier, 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 les 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.

³⁰ Cet objectif n'apparaît pas dans tous les scénarios : il dépend de l'album proposé et de l'âge des élèves.

- *Au cours de la journée*

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

2.3.3. Les caractéristiques de l'outil

Narramus apparaît comme un outil innovant dans le domaine de la compréhension à l'école maternelle. Nous allons nous intéresser aux caractéristiques didactiques et pédagogiques spécifiques de cet outil. Cette description nous permettra de mieux comprendre les effets sur les élèves et les pratiques des enseignants (chapitres 3 à 5).

2.3.3.1. Un outil qui relève de l'*embedded instruction* (enseignement intégré)

Jusqu'à présent, très peu d'outils visant un enseignement de la compréhension à l'école maternelle ont été évalués. À notre connaissance, le seul outil francophone testé s'intitule « Compréhension » (Gourgue, Bianco, & Coda, 2013). Il propose 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. Les auteurs 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.

D'autres chercheurs proposent une approche intégrative ou *embedded instruction* (Goldstein *et al.*, 2016), c'est le cas des auteurs de *Narramus* : l'enseignement de la compréhension est basé sur des albums de littérature de jeunesse et différentes compétences (lexique, inférences, états mentaux...) sont travaillées successivement voire simultanément dans un même module.

Nous verrons dans les chapitres suivants si cette approche permet de produire les effets attendus par les auteurs.

2.3.3.2. Un enseignement explicite

En 2011, les inspecteurs généraux de l'Éducation nationale écrivaient : « il faudrait que tous les maîtres soient plus rigoureusement préparés à distinguer ce sur quoi doit porter cette activité [la compréhension] et comment ils peuvent la conduire, sensibilisés au fait qu'il y a des niveaux

de compréhension de plus en plus fins qu'une unique rencontre des textes ne peut épuiser » (p. 131). C'est dans ce contexte et pour répondre aux préoccupations et besoins des enseignants que les auteurs de *Narramus* ont rédigé des guides du maître particulièrement détaillés.



Formulées pour les enseignants ou pour les élèves, les modalités de guidage proposées dans les scénarios pédagogiques sont très précises et clairement explicitées. Ce guidage structuré correspond aux indications formulées dans les programmes : « L'enseignant rend lisibles les exigences de la situation scolaire par des mises en situations et des explications qui permettent aux enfants – et à leurs parents – de les identifier et de se les approprier. Il aide à identifier les objets sur lesquels portent les apprentissages, fait acquérir des habitudes de travail qui vont évoluer au fil du temps et que les enfants pourront transférer. Pour ce faire, il s'attache à faire percevoir la continuité entre les situations d'apprentissage, les liens entre les différentes séances. Pour stabiliser les premiers repères, il utilise des procédés identiques dans ses manières de questionner le groupe, de faire expliciter par les enfants l'activité qui va être la leur, d'amener à reformuler ce qui a été dit, de produire eux-mêmes des explications pour d'autres à propos d'une tâche déjà vécue » (MEN, 2015, p. 5).

La précision du guide du maître vise à clarifier les objectifs, à faciliter la préparation et la mise en œuvre des différentes activités proposées. Différents documents sont fournis aux enseignants :

- une fiche de préparation (*cf.* annexe 1) contenant une description des objectifs, des différentes étapes à suivre, des tâches à réaliser, des consignes à donner avec même, parfois, des propositions d'énoncés.
- Des commentaires de natures très différentes : il peut s'agir d'un conseil pratique, de la justification d'une option didactique contre-intuitive ou inhabituelle, d'une explicitation de la théorie sous-jacente aux propositions des auteurs... Pour les rédiger, les auteurs ont recensé les questions posées par les enseignants-concepteurs qui ont testé les prototypes et imaginé celles que ne manqueraient pas de se poser ceux qui, contrairement aux précédents, ne pourraient dialoguer directement avec les auteurs.

Le Tableau 7 montre quelques exemples de la nature des commentaires présents dans le guide du maître.

Tableau 7. Exemples de commentaires rédigés dans les guides du maître *Narramus*³¹

<p>Conseil pratique (extrait du scénario <i>Le Machin</i>, PS^o)</p>	
	<p>Les enseignantes-conceptrices ont affiché sur un mur les images des mots étudiés au fur et à mesure de l'avancée dans le scénario. Ce support permet de désigner les mots pendant les narrations et d'organiser de petits jeux centrés sur le lexique (montrez-moi tous les animaux, tous les vêtements...).</p> <p>Une enseignante a observé que ses élèves allaient très souvent pointer et nommer les images. Parfois plusieurs d'entre eux échangeaient autour de ces images en totale autonomie pendant le temps de l'accueil ou de jeux libres.</p>
<p>Justification d'une option didactique contre-intuitive ou inhabituelle (extrait du scénario <i>Le Machin</i>, PS)</p>	
 <p>D8</p>	<p>Afficher D8 – Expliquer : je vais d'abord vous raconter le début de l'histoire sans vous montrer l'image. Vous, vous devez essayer de fabriquer l'image, de voir dans votre tête ce qui se passe.</p> <p>Les enseignants-concepteurs étaient sceptiques quant à la capacité des tout jeunes enfants à comprendre cette consigne. Nous l'avons donc fait précéder d'un jeu dans lequel nous avons demandé aux élèves de fabriquer, dans leur tête, l'image d'un paquet de bonbons qui était caché dans la boîte des doudous. Nous leur avons ensuite proposé d'aller chercher un bonbon. Que croyez-vous qu'il arriva ? Ils se sont précipités sur la boîte des doudous.</p>
<p>Explicitation de la théorie sous-jacente à l'une des activités (extrait du scénario <i>Les deniers de compère lapin</i>, GS)</p>	
<p>Rosenthal et Ehri (2011) ont montré qu'en demandant systématiquement et régulièrement aux élèves de prononcer les mots nouveaux qu'on vient d'expliquer, on améliore considérablement leur mémorisation à long terme et leur compréhension. En procédant de la sorte, on permet aux élèves de stocker en mémoire deux connaissances d'un même mot ou d'une même expression : l'une sémantique, l'autre phonologique. Ce sont ces deux modalités qui permettent ensuite aux élèves de reconnaître, de comprendre et de produire le vocabulaire nouvellement appris.</p> <p>Nous vous engageons donc vivement à ritualiser cette courte activité qui consiste à inviter tous les élèves à redire les mots nouveaux à haute voix, au cours de chaque module, même si nous ne l'écrivons pas systématiquement dans le déroulé pour ne pas alourdir inutilement notre guide.</p>	

3° Des supports numériques préfabriqués : ils comprennent un powerpoint avec les tâches proposées aux élèves (découverte du texte, description des illustrations, cartes de vocabulaire...) pendant les différents modules, les supports (texte écrit, images à projeter et à découper, activités spécifiques, gifs, vidéos...) permettant de mener les activités centrées sur la

³¹ Ces extraits sont tous issus des guides pédagogiques fournis aux enseignants en format pdf pendant l'expérimentation.

compréhension ou sur le lexique, l'album scanné, les gabarits des masques et des marottes pour raconter, les enregistrements audio pour écouter la version lue ou racontée de l'album et parfois des vidéos (dessins animés, théâtre animé...).

Des chercheurs se sont intéressés aux liens entre les caractéristiques des activités de langage oral et écrit proposées par les enseignants et les performances des élèves dans le domaine de la lecture, de la compréhension et du vocabulaire. (Connor, Morrison, Petrella, 2004 ; Connor, Morrison, & Slominski, 2006). Leurs études font apparaître des liens forts entre un enseignement explicite et les progrès des élèves. Est-ce que ce sera le cas avec l'outil *Narramus* ? Nous analyserons les résultats des élèves en les mettant en lien avec ces constats et les caractéristiques de l'outil.

D'autres ont montré une interaction très significative entre le niveau initial de l'élève et le style d'enseignement reçu. Les élèves considérés dans la norme des apprentissages et ceux présentant des besoins d'enseignement plus importants en début d'année progressent, par exemple, davantage dans les classes où l'enseignement de la compréhension est réalisé de manière explicite et dirigée par l'enseignant (Bianco & Bressoux, 2009). L'enseignement proposé dans *Narramus* produira-t-il des progrès différents selon le niveau des élèves ? Dans notre recherche doctorale, nous mesurerons les progrès des élèves en tenant compte de leur niveau initial.

2.3.3.3. Un enseignement ritualisé









Narramus propose un déroulement ritualisé. La ritualisation inter et intra scénarios vise à permettre aux élèves de refaire plusieurs fois la même chose, favorisant ainsi 1° le développement progressif d'un sentiment de compétence et 2° un contrôle de plus en plus grand sur leur activité, pour pouvoir, à terme, la réguler sans aide.

Les auteurs postulent que le déroulement de modules stabilisés assure un repérage facilité dans l'activité et un engagement accru. Partant, de multiples activités reviennent au cours d'un seul et même album et se retrouvent d'un scénario à l'autre. Cet apprentissage collectif et progressif s'inscrit dans la durée avec un scénario travaillé sur un mois. On peut faire l'hypothèse que, parce qu'ils vont faire les mêmes activités plusieurs fois, guidées par l'enseignant, les élèves qui travaillent avec *Narramus* adopteront progressivement les manières de faire qui ont été fortement guidées au départ puis de moins en moins. Notre protocole expérimental testera cette hypothèse.

Narramus propose un enseignement explicite de la compréhension, la démarche retenue est très ritualisée afin de favoriser la clarté cognitive. Les objectifs sont toujours annoncés aux élèves et l’affichage de pictogrammes permet à ces derniers de savoir ce qu’ils sont en train de faire et d’apprendre. Les auteurs affichent une réelle volonté d’associer les élèves de maternelle à leurs apprentissages en leur disant clairement et précisément ce qu’ils vont faire et ce que l’enseignant attend d’eux.

Huit pictogrammes correspondent à des consignes ritualisées (cf. Figure 3).

Figure 3. Pictogrammes présentés aux élèves de l’outil *Narramus*

Au fil des diapositives, huit logos logos, huit consignes ritualisées	
	L'enseignant·e lit le texte.
	L'enseignant·e raconte le texte (ou un extrait) sans support de texte.
	Les élèves doivent imaginer le dessin qui illustre le texte.
	Les élèves doivent imaginer le texte, prévoir la suite de l'histoire.
	Un·e élève raconte seul·e l'histoire.
	La classe complète le récit de l'élève.
	La boîte « mémoire des mots » permet de stocker les images des mots appris et de les réviser.
	Les élèves jouent la scène.

(extrait du scénario *Le jour où loup gris est devenu bleu*, GS)

Le logo « L’enseignant·e lit le texte. » assure deux fonctions : la première indique à l’adulte que c’est lui qui effectue la lecture³², la seconde oriente l’activité de l’élève (la maitresse lit et l’élève fabrique l’illustration dans sa tête, transforme les mots en images). C’est pour favoriser cet engagement cognitif que les auteurs proposent systématiquement ces pictogrammes : « Les élèves les retrouveront au fil des modules et des différents scénarios *Narramus*. Placé en haut à gauche de chaque diapositive, le pictogramme permet de gagner un temps précieux puisque, très vite, les élèves saisissent ce qu’ils vont devoir faire et/ou apprendre

³² Cette indication est clairement écrite dans le guide du maitre.

et comment ils doivent le faire. Ils prennent ainsi de plus en plus de contrôle sur leur activité, conséquence attendue d'un apprentissage explicite. » (Cèbe et Goigoux, 2017, p. 26).

Avec le pictogramme « L'enseignant·e raconte l'histoire (ou un extrait) sans support du texte », les auteurs invitent les enseignants à « Reformuler le texte dans une langue adaptée aux élèves, mimer, jouer sur l'intonation, ajouter des informations, expliciter l'implicite... Ceci leur permet de compléter l'image mentale qu'ils ont commencé à fabriquer » (*idem*, p. 30).

Au début de chaque module, lorsque les élèves découvrent et mémorisent le vocabulaire inconnu et révisent les mots et les expressions déjà apprises. Le pictogramme « La boîte mémoire des mots » est affiché : les élèves savent ainsi qu'ils vont apprendre de nouveaux mots et de nouvelles expressions dont ils ont besoin pour bien comprendre l'histoire et bien la raconter. Cette accumulation du savoir lexical et l'entraînement systématique proposé est très singulière.

Avec le logo point d'interrogation, les élèves doivent imaginer le dessin qui illustre le texte. Pour les auteurs de *Narramus*, la dissociation du texte et de l'image permet de travailler la compétence inscrite dans le programme « Comprendre un texte sans autre aide que le langage entendu ».

Quel que soit le niveau auquel s'adresse le scénario (PS, MS ou GS), les pictogrammes utilisés sont toujours les mêmes. C'est dans le même but que les auteurs proposent un format stabilisé et un déroulement similaire des modules. « Cette option pédagogique a le mérite de rendre le monde scolaire plus prévisible : pour chaque élève, se sentir suffisamment en sécurité, grâce à un environnement dont les règles de fonctionnement et les repères sont stables, est à nos yeux une condition nécessaire à la réflexion et à l'apprentissage » (*idem*, p. 26).

En outre, le rappel donne un but intégrateur à toutes les activités. Les élèves savent, dès le départ, que toutes les activités mises en œuvre visent à leur apprendre à raconter sans aide l'histoire travaillée en classe. Ils prennent ainsi conscience du caractère actif et intentionnel de la compréhension. Après la découverte ou la révision lexicale, chaque module débute par un rappel de récit des épisodes précédents pris en charge soit par un élève soit par l'enseignant.

2.3.3.4. Une attention conjointe favorisée

Dans une classe, mobiliser (puis maintenir) l'attention des élèves représente un objectif crucial pour les enseignants. Ces derniers doivent mettre en œuvre des conditions d'apprentissage et

utiliser des supports pertinents pour orienter l'attention des élèves. « Pour qu'il y ait attention conjointe, il faut une référence spatiale autocentrée nécessaire : une personne, sujet agissant et raisonnant. Elle entre en contact visuel avec une autre personne (intersubjectivité primaire) puis oriente son regard vers un élément d'intérêt commun (personne, objet, événement). [...] C'est un acte perceptif (voir ensemble) qui implique un acte cognitif (savoir ensemble) » (Aubineau, Vandromme, & Driant, 2015, p. 143).

Les auteurs de *Narramus* ont proposé un enseignement qui repose sur une utilisation régulière du vidéoprojecteur. Tous les enseignants engagés dans l'expérimentation en disposaient. Pour la plupart des activités collectives, les informations sont affichées au tableau à l'aide de ce support. Ce dispositif facilite le guidage de l'enseignant et le maintien de l'attention des élèves sur ce qui fait l'objet de la discussion collective.

Pour les auteurs, le numérique est un auxiliaire précieux quand on veut centrer l'attention des élèves sur un point précis, faire disparaître des informations, ajouter du texte, montrer toutes les illustrations d'un même épisode sur une même diapositive et soutenir ainsi le rappel, organiser une activité particulière. Pour eux, l'affichage permet de réguler plus aisément les interactions et diminue les contraintes qui pèsent sur l'activité, celle de l'enseignant et celle des élèves.

2.3.3.5. Un recours aux supports multimédia numériques

Le recours à des vidéos, de la musique et des sons, loin de distraire les élèves, permet un engagement plus durable dans l'activité que les livres traditionnels (Moody, Justice, & Cabell, 2010). Ces supports, quand ils sont combinés avec le texte oral, améliorent la compréhension du texte, l'interprétation des événements et la production d'inférences (Myrtil et al., 2018; Zhou & Yadav, 2017). C'est pourquoi les auteurs de *Narramus* font un usage important de ces ressources multimédia.

2.3.3.6. Un enseignement particulier du lexique

En moyenne, 5 à 7 mots sont enseignés en collectif par jour. Les enseignants introduisent le mot en donnant une explication en employant un vocabulaire adapté au niveau de compréhension des élèves. Les mots et expressions sont présentés à l'aide du vidéoprojecteur : une représentation visuelle et parfois sonore est proposée sous forme d'images fixes ou animées (gif ou vidéos) ou d'enregistrements (pour le verbe *aboyer* par exemple).

Après la phase de découverte du nouveau vocabulaire, l'enseignant présente les cartes des mots du jour. Ce matériel manipulable peut être affiché dans un endroit de la classe réservé à cet effet et/ou mis dans une boîte à mots régulièrement utilisée en séances dirigées ou autonomes pour des révisions. Tout au long de la journée, les élèves sont invités à réutiliser le vocabulaire nouvellement appris et l'enseignant veille à l'intégrer dans les conversations quotidiennes qui vont au-delà des échanges sur le livre. Des « dictées de mots » sont régulièrement menées en salle jeux (ou en classe) : les élèves sont alternativement invités à mimer le mot (ou l'expression) prononcé par l'adulte ou un camarade puis à mimer le mot qu'ils voient sur une image. Ces activités peuvent être conduites en collectif ou en atelier, en classe entière ou en groupe.

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 leur 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* avec un entraînement et une révision systématiques des mots et expressions enseignés.

Reste à savoir si cette mise en œuvre produit les effets attendus sur les apprentissages des élèves et si ces effets sont solides et durables. L'étude de Pullen, Tuckwiller, Konold, Maynard et Coyne (2010) nous a, en effet, alertée sur ce second point.

Pullen et ses collaborateurs ont cherché à opérationnaliser les résultats de recherche sur l'enseignement explicite du vocabulaire en proposant une intervention différenciée à des élèves présentant des troubles du langage et des difficultés en lecture. 224 élèves de première année (CP) de trois écoles élémentaires en ont bénéficié au cours de deux semaines.

Tous ont reçu un enseignement collectif (niveau 1) assuré par l'enseignant. Chaque semaine, un album était lu à haute voix (les jours 1 et 3) et l'étude du vocabulaire était proposée en collectif à partir de 8 mots et expressions présents de l'histoire. Seize mots ont donc été enseignés pendant la durée de l'intervention (trente minutes).

Les élèves ont été répartis en trois groupes. Le groupe 1 (G1) est composé de 98 élèves identifiés « à risque » : 49 ont reçu une intervention de niveau 1 (G1 niveau 1) et 49 une intervention de niveau 2 (G1 niveau 2). Les élèves de niveau 1 n'ont reçu que l'enseignement

collectif alors que les élèves de niveau 2 ont bénéficié d'interventions supplémentaires. Un membre de la recherche proposait un apprentissage plus intensif (les jours 2 et 4) en petits groupes de 4 ou 5 élèves pendant 20 minutes. Il faisait réviser les expressions et les mots appris la veille et proposait de multiples occasions d'utilisation du vocabulaire. Le groupe 2 (G2) comprenait 126 élèves qui ne présentaient pas de difficultés majeures pouvant faire obstacle à l'apprentissage de la lecture. L'évaluation mesurait l'apprentissage des 16 mots et expressions avec trois épreuves en post-test immédiat. Une épreuve évaluait le langage en production pour un score maximal de 8 points ; une autre mesurait le niveau de connaissance contextuelle des mots (8 points) et une dernière portait sur le langage en réception (8 points).

Au post-test immédiat après les deux semaines d'intervention, les élèves « à risque » qui avaient bénéficié de l'intervention additionnelle de niveau 2 ont obtenu des scores significativement plus élevés aux épreuves « contextuelle » et « réception » que les élèves à risque qui n'ont bénéficié que de l'intervention de niveau 1. On ne notait pas de différence significative pour le langage expressif.

Pour étudier la stabilité de l'effet de l'intervention, les chercheurs ont proposé un post-test différé quatre semaines après l'intervention en employant les mêmes épreuves. Les résultats font apparaître que les différences entre les élèves à risque (G1 niveau 1 et G1 niveau 2) ne sont plus statistiquement significatives. Ces deux groupes d'élèves ont perdu une partie de leurs acquis (sur les niveaux réceptifs et contextuels), alors que ceux qui n'étaient pas considérés comme « à risque » ont maintenu leurs connaissances du vocabulaire. Le principal résultat de Pullen *et al.* porte sur ce point : l'avantage pris par le groupe des élèves ayant bénéficié de plus d'occasions d'apprendre (niveau 1 et niveau 2) a été de courte durée puisqu'un mois après la fin de l'intervention, les effets ont disparu.

Pour expliquer cette perte, les chercheurs ont remis en cause la brièveté du dispositif et l'absence d'usage continué des mots par les enseignants après l'intervention. Précisons que, pour assurer la rigueur de l'étude et contrôler la variabilité d'une classe à une autre, les professeurs participant à la recherche avaient reçu pour consigne de ne pas faire réviser les mots enseignés. Il ressort donc de cette étude que les activités contextualisées ne produisent pas d'effets durables si on n'accorde pas suffisamment de temps à la mémorisation des mots et à leur utilisation répétée au fil des semaines. C'est pourquoi la question du maintien du lexique des albums étudiés a retenu notre attention : notre protocole expérimental comporte des

épreuves de post-test différé de lexique afin d'évaluer la stabilité des apprentissages lexicaux eux-mêmes stimulés par les activités régulières de révision prévues par *Narramus*.

2.3.3.7. Une mise en scène des histoires

Dans *Narramus*, une mise en scène des histoires étudiées est systématiquement proposée (via des activités de théâtre ou l'usage d'une maquette et de marionnettes). On a montré que celle-ci exerce un effet sur la qualité de la compréhension, le développement des compétences narratives et l'acquisition du vocabulaire (Fischer & Zwaan, 2008 ; Glenberg *et al.*, 2004 ; McLeod, Hardy & Kaiser, 2017 ; Newman *et al.*, 2015 ; Pesco et 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 (*embodied cognition*) (Glenberg, 2011 ; Bara & Tricot, 2017). Intégrer des activités corporelles aux activités habituelles favorise la compréhension en aidant à la mémorisation des informations importantes et de l'ordre des événements (Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004). 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.

Conclusion

Les auteurs postulent que, pour compenser les inégalités, il faut commencer tôt l'enseignement de la compréhension : habituer les élèves à réaliser des tâches de haut niveau, les inciter à raisonner... C'est pourquoi *Narramus* propose un ensemble de tâches complexes. Les activités sont ritualisées d'un module à un autre et d'un scénario à un autre. Nous verrons si les ambitions portées par les auteurs sont bénéfiques pour les élèves et les enseignants dans les chapitres suivants.

2.4. Un outil qui répond à la prescription

Comme nous l'avons écrit dans le chapitre 1, les enseignants répondent à des prescriptions. Il faut nous assurer que les activités proposées répondent au prescrit : c'est pourquoi nous mettons en lien l'outil que nous évaluons avec les programmes et les ressources pour l'école maternelle.

2.4.1. Le programme de l'école maternelle (BO mars 2015)

Depuis 2014, l'école maternelle forme un cycle unique. La place du langage y est affirmée comme condition essentielle de la réussite. La stimulation et la structuration du langage oral ainsi que l'entrée progressive dans la culture de l'écrit constituent des priorités et concernent l'ensemble des domaines d'apprentissage. « Tout au long de l'école maternelle, l'enseignant crée les conditions bienveillantes et sécurisantes pour que tous les enfants (même ceux qui ne s'expriment pas ou peu) prennent la parole, participent à des situations langagières plus complexes que celles de la vie ordinaire. [...] Ainsi, il contribue à construire l'équité entre enfants en réduisant les écarts langagiers » (p. 7).

Les attendus de fin de maternelle sur la compréhension apparaissent dans le domaine d'apprentissage intitulé « Mobiliser le langage dans toutes ses dimensions ».

Parmi les différents objectifs déclinés, on trouve :

1° « Écouter de l'écrit et comprendre³³ » : « L'école maternelle occupe une place privilégiée pour offrir aux enfants une fréquentation de la langue de l'écrit, très différente de l'oral de communication. L'enjeu est de les habituer à la réception de langage écrit afin d'en comprendre le contenu. L'enseignant prend en charge la lecture, oriente et anime les échanges qui suivent l'écoute. La progressivité réside essentiellement dans le choix de textes de plus en plus longs et éloignés de l'oral ; si la littérature de jeunesse y a une grande place, les textes documentaires ne sont pas négligés » (p. 9).

2° « Comprendre et apprendre » : « Ils [les enfants] sont incités à s'intéresser progressivement à ce qu'ils ignoraient grâce à l'apport de nouvelles notions, de nouveaux objets culturels et même de nouvelles manières d'apprendre. Les moments de réception où les enfants travaillent mentalement sans parler sont des activités langagières à part entière que l'enseignant doit rechercher et encourager, parce qu'elles permettent de construire des outils cognitifs : reconnaître, rapprocher, catégoriser, contraster, se construire des images mentales à partir d'histoires fictives, relier des événements entendus et/ou vus dans des narrations ou des

³³ Quand on sait qu'écouter une histoire ne suffit pas pour la comprendre, cette formulation peut, tout d'abord, surprendre mais elle a le mérite de distinguer la compréhension élaborée lors des échanges oraux de celle portant sur un texte écrit. La note de service n°2019-084 du 28 mai 2019 reformule d'ailleurs cet objectif puisqu'on peut y lire : « Un travail sur la compréhension est d'autant plus nécessaire dès l'école maternelle que cette activité langagière est invisible pour un enfant. Il ne suffit pas d'écouter pour comprendre. » (p. 2)

explications, dans des moments d'apprentissages structurés, traiter des mots renvoyant à l'espace, au temps, etc. » (p. 7).

3° « Échanger et réfléchir avec les autres » : « [...] L'école demande régulièrement aux élèves d'évoquer, c'est-à-dire de parler de ce qui n'est pas présent (récits d'expériences passées, projets de classe...). Ces situations d'évocation entraînent les élèves à mobiliser le langage pour se faire comprendre sans autre appui, elles leur offrent un moyen de s'entraîner à s'exprimer de manière de plus en plus explicite. Cette habileté langagière relève d'un développement continu qui commence tôt et qui ne sera constitué que vers huit ans. » (p. 8).

Les quatre compétences attendues en fin de grande section sont les suivantes :

- 1° s'exprimer dans un langage syntaxiquement correct et précis ; reformuler pour se faire mieux comprendre ;
- 2° comprendre des textes écrits sans autre aide que le langage entendu ;
- 3° pratiquer divers usages du langage oral : raconter, décrire, évoquer, expliquer, questionner, proposer des solutions, discuter un point de vue ;
- 4° manifester de la curiosité par rapport à l'écrit : pouvoir redire les mots d'une phrase écrite après sa lecture par l'adulte, les mots du titre connu d'un livre ou d'un texte (MEN, 2015, p. 11).

L'enseignement proposé dans *Narramus* vise bien le développement des compétences attendues en fin d'école maternelle. Comme nous allons le voir, il répond aussi aux « prescriptions » complémentaires des documents « ressources » publiés sur le site Éduscol.

2.4.2. Les ressources de l'école maternelle : Éduscol

Les ressources publiées par le ministère de l'Éducation nationale distinguent deux formes de langage oral à développer à l'école maternelle : le langage en situation et le langage scriptural. Le second suppose un langage décontextualisé, très différent du premier « Il est nécessairement précis et structuré, conditions obligatoires pour qu'il soit compris. [...] ». Il s'apparente au langage écrit en tant que forme produite hors du contexte immédiatement vécu, comme l'est l'écrit. Il est élaboré. Pour référer à ce type de forme élaborée, des chercheurs parlent d'« oral scriptural » [...]. Il recouvre des formes d'oral élaborées, structurées, requises par l'école scolairement efficaces. Ce sont elles qui sont visées par le programme quand il indique : « Pratiquer divers usages du langage oral : raconter, décrire, évoquer, expliquer, questionner, proposer des solutions, discuter un point de vue » (MEN, p. 7).

Ce texte pointe deux éléments : quand on sort de la situation d'énonciation, il faut tendre vers des formulations de plus en plus explicites et donc développer le lexique. Le langage doit

être plus structuré, le lexique plus soutenu. Ce rapport au langage et au lexique est particulièrement développé dans *Narramus*, permettant d'assurer une compatibilité entre la prescription et l'outil.

« [...] L'école demande un autre rapport au langage oral qui n'est plus fondé sur les expériences immédiates et agies mais qui les met à distance, les reconfigure dans des discours construits parfois élaborés collectivement. Il oblige à se décaler par rapport à l'évènement, à le considérer avec un autre point de vue, à le décontextualiser pour le reconstruire par le langage. Ce rapport second est privilégié par l'école car c'est l'instrument d'entrée dans les apprentissages qui relèvent aussi de l'écrit. En effet, il est de très nombreuses situations orales scolaires où il s'agit de décrire, de mener une explication, de développer un questionnement, de justifier, de relater des faits précis, d'organiser un raisonnement chronologique ou un rapport de cause à conséquence. Ces conduites langagières exigent des usages élaborés du langage oral, produits d'une manière consciente, avec une attention portée à leur construction autant qu'aux significations qu'ils véhiculent. Cet oral est fortement apparenté à l'écrit ; même s'il n'en a pas toutes les formes (les deux termes de la négation ne sont pas nécessairement présents, la reprise du sujet, fréquente dans un oral quotidien, peut fonctionner), il en a les caractéristiques essentielles en termes d'organisation et d'élaboration : c'est pourquoi on emploie l'expression oral scriptural. Il constitue l'objectif majeur de l'école maternelle. Et c'est sur sa capacité à l'utiliser que peut se jouer l'avenir scolaire d'un enfant. » (*idem*, p. 7).

Les objectifs du programme de l'Éducation nationale et les orientations ministérielles successives témoignent de l'importance de la maîtrise de la langue dans les apprentissages ultérieurs. Qu'en est-il de la récente note de service (MEN, 2019) adressée par le ministre de l'Éducation nationale à tous les enseignants du premier degré ?

2.4.3. L'école maternelle, école du langage : note de service n°2019-084 du 28 mai 2019

Cette note de service réaffirme certains éléments essentiels dans le développement du langage oral et de la compréhension.

« Le cycle des apprentissages premiers est mis à profit pour faire progresser les élèves depuis la petite section jusqu'à la grande section vers la compréhension et l'usage d'une langue française orale de plus en plus élaborée sur laquelle ils pourront s'appuyer lors de l'apprentissage de la lecture. »

« En situation scolaire, le langage correspond aux activités de compréhension (écouter, lire) et aux activités de production (parler, écrire). L'appropriation par les élèves d'un langage oral riche, organisé et compréhensible requiert la mise en œuvre d'un enseignement structuré et systématique. »

« L'école accompagne aussi les élèves, dès la petite section, dans le développement d'un langage oral de plus en plus construit, diversifié dans ses fonctions. L'enseignant concentre progressivement son action sur le développement des compétences communicationnelles : écoute, attention partagée, mémoire, expression. L'évolution attendue des compétences des élèves dans ce domaine est très importante. En grande section, les élèves doivent en effet pouvoir se faire comprendre par le seul usage du langage. L'enjeu est de les rendre capables de raconter, d'expliquer une réalité passée ou à venir, de créer une histoire portant sur des événements, lieux ou personnages inconnus d'au moins un des partenaires de l'échange. »

Les programmes et les ressources de l'école maternelle mettent en avant l'acquisition et la structuration du langage oral. L'outil *Narramus* semble satisfaire ces deux exigences mais est-il suffisant pour faire progresser les élèves ? Les chapitres 3 et 4 apporteront des réponses à ce questionnement.

3. Problématique et questions de recherche

Des données probantes attestent que l'introduction d'outils didactiques peut participer à l'amélioration des pratiques pédagogiques. L'efficacité de ces outils (manuels, dispositifs ou scénarios d'enseignement, ressources numériques...) est étudiée par des chercheurs en contexte francophone (Bélangier *et al.*, 2012 ; Dupriez, 2015 ; Cartier, Contant, & Janosz, 2012) et anglo-saxon (Berman, 1980 ; Rowan & Miller, 2007). Dans cette perspective, notre recherche doctorale vise à étudier si l'outil conçu par l'équipe de Cèbe et Goigoux est opérationnel quand il est utilisé par des enseignants n'ayant pas participé à son élaboration et si sa mise en œuvre produit les effets attendus sur leurs pratiques et les apprentissages langagiers de leurs élèves. Notre protocole expérimental s'appuie pour cela sur des critères méthodologiques exigeants, ceux de l'*evidence based-education* (Davies, 1999 ; Slavin, 2002).

3.1. Problématique générale

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, 2014-a; Hindman *et al.*, 2014).

C'est dans ce but que Cèbe et Goigoux ont conçu *Narramus* (2017) après avoir identifié, dans la littérature scientifique, les caractéristiques des dispositifs pédagogiques ayant produit des effets avérés et durables (*evidence based practices*). 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. *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 les pratiques des enseignants et le développement des compétences langagières des élèves même lorsqu'il n'était pas soutenu par un dispositif spécifique d'accompagnement. Pour atteindre cet objectif et étudier le rôle compensatoire que peut jouer l'école maternelle sur le développement des compétences langagières des jeunes élèves, nous nous sommes aussi intéressée aux modalités et contenus des accompagnements mis en œuvre par les formateurs et au rôle desdits accompagnements sur les pratiques des maîtres.

3.2. Objectifs de recherche

Pour rappel, notre recherche tente de répondre à trois types d'interrogations :

- une première d'ordre socio-pédagogique : l'utilisation d'un outil didactique innovant peut-elle favoriser les apprentissages langagiers de tous les élèves et contribuer à réduire les inégalités de réussite à l'école ?
- une seconde d'ordre méthodologique : comment construire et mettre en œuvre un dispositif expérimental d'évaluation présentant des critères méthodologiques rigoureux pour tester l'efficacité d'un outil didactique ?
- une troisième relative au rôle de l'outillage et de l'accompagnement dans l'amélioration des pratiques d'enseignement : quels liens existe-t-il entre l'utilisation de l'outil, l'accompagnement assuré par des équipes de circonscription et le développement professionnel des enseignants ?

3.3. Hypothèses de recherche

Notre recherche doctorale teste cinq principales hypothèses.

- **Première hypothèse de recherche (l'effet *Narramus*)**

La première hypothèse stipule que les élèves bénéficiant de *Narramus* progresseront plus que les autres sur toutes les dimensions évaluées.

- **Seconde hypothèse de recherche (l'absence d'effet de l'accompagnement)**

Puisque *Narramus* a été conçu dans le cadre d'un processus de conception continuée dans l'usage, 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, *Narramus* produira des effets similaires avec ou sans accompagnement.

- **Troisième hypothèse de recherche (le développement professionnel)**

Au regard de notre seconde hypothèse, il conviendra toutefois d'en tester une autre selon laquelle l'accompagnement n'affecterait pas directement les performances des élèves, mais toucherait le développement professionnel des enseignants.

- **Quatrième hypothèse de recherche (le transfert)**

En proposant des scénarios ritualisés, en multipliant les activités et en travaillant l'attention conjointe, *Narramus* vise à donner des habitudes aux élèves et à permettre le développement d'attitudes d'écoute, de mémorisation, de raisonnement tout en favorisant l'engagement.

Nous émettons une quatrième hypothèse selon laquelle les élèves ayant bénéficié d'un enseignement explicite de la compréhension avec l'outil *Narramus* construisent des compétences suffisamment solides pour mieux comprendre un texte nouveau que leurs camarades d'égale condition sociale. Autrement dit, nous vérifierons si les compétences construites se transfèrent vers d'autres écrits (livres, histoires, albums) que ceux travaillés en classe.

- **Cinquième hypothèse de recherche (la réduction des inégalités)**

La cinquième hypothèse stipule qu'un usage prolongé de l'outil *Narramus* (deux années scolaires, soit six modules de quatre semaines chacun) permet de réduire les écarts entre élèves de milieux sociaux contrastés.

3.4. Design de recherche

Pour mettre à l'épreuve ces hypothèses, nous avons mené deux études qui poursuivent le même objectif général : analyser les effets de l'outil *Narramus* sur les compétences langagières des élèves. Le protocole expérimental élaboré s'appuie sur des critères méthodologiques exigeants notamment utilisés dans *l'evidence based-education*.

La première étude³⁴ vise à identifier un éventuel effet *Narramus* sur les apprentissages des élèves (effet de l'enseignement) mais aussi à mesurer l'influence de l'accompagnement (effet de l'accompagnement) sur le développement professionnel des enseignants (hypothèses 1, 2 et 3). Pour observer les effets de l'outil sur les élèves, nous avons mesuré les compétences en compréhension des élèves (de petite, moyenne et grande section) avant (pré-test) et après (post-test) enseignement, et ce à plusieurs reprises. Nous avons ainsi comparé les résultats des élèves en fonction de leur groupe expérimental d'appartenance.

Pour évaluer et mieux comprendre les effets de l'outil et/ou de l'accompagnement sur les pratiques des maîtres, nous avons mené une analyse qualitative des questionnaires renseignés par les enseignants.

Des enregistrements filmés des temps d'accompagnement et des compte-rendu de ces séances nous ont permis de distinguer les indices de développement professionnel liés à l'accompagnement de ceux liés à l'utilisation de l'outil.

La deuxième étude cherche à identifier un éventuel effet « transfert » et un effet « réduction des inégalités » (hypothèses 4 et 5). Pour cela, nous avons évalué la compréhension du langage écrit à partir d'un album de la littérature de jeunesse non étudié en classe. Une épreuve de rappel de récit, une de lexique et un questionnaire de compréhension ont été proposés en fin de grande section à trois groupes d'élèves scolarisés dans des écoles classées en REP pour les uns (certains élèves avaient travaillé avec l'outil *Narramus*, d'autres non), d'écoles dans un quartier très favorisé pour les autres.

³⁴ 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.

3.5. Synthèse

À travers deux études quasi-expérimentales, nous mettons à l'épreuve cinq hypothèses desquelles découlent différentes questions de recherche.

Tableau 8. Hypothèses et questions de recherche

		Questions de recherche
Hypothèse n° 1	Les élèves ayant travaillé avec <i>Narramus</i> progresseront plus que les autres sur les dimensions évaluées (lexique, rappel de récit, questionnaire).	<ol style="list-style-type: none"> 1. L'intervention produit-elle des effets positifs sur les apprentissages des élèves ? 2. Sur quels aspects spécifiques les effets sont-ils les plus forts ? 3. Les effets sont-ils différents selon le niveau initial des élèves ? 4. Les effets sont-ils cumulatifs ? Quels sont les résultats après deux scénarios ? Après trois scénarios ? 5. Le vocabulaire enseigné au premier et au deuxième trimestre est-il encore connu à la fin de l'année ?
Hypothèse n° 2	Conçu sur des bases scientifiques attestées et en partenariat avec des enseignants, <i>Narramus</i> produira des effets similaires sur les apprentissages des élèves avec ou sans accompagnement.	<ol style="list-style-type: none"> 1. L'usage des outils didactiques proposés est-il différent selon que les enseignants bénéficient d'un accompagnement régulier ? 2. Quelles sont les modalités d'accompagnement mises en œuvre ? 3. Quels sont les effets de cet accompagnement sur les apprentissages des élèves ?
Hypothèse n° 3	L'accompagnement devra impacter le développement professionnel des enseignants.	<ol style="list-style-type: none"> 1. Dans quelle mesure l'utilisation de <i>Narramus</i> transforme-t-elle les pratiques professionnelles et les connaissances dont les maîtres disposent ? 2. Y a-t-il un lien entre accompagnement et développement professionnel, et si oui lequel ?
Hypothèse n° 4	Les élèves ayant travaillé avec <i>Narramus</i> développeront des attitudes et des habitudes de traitement qui les rendront plus efficaces que leurs camarades de même milieu, dans une situation inédite.	<ol style="list-style-type: none"> 1. Les élèves ayant travaillé avec <i>Narramus</i> développent-ils des compétences suffisamment solides pour mieux comprendre un album inconnu ? 2. Les résultats de ces élèves se distinguent-ils de ceux obtenus par d'autres élèves de REP n'ayant pas utilisé <i>Narramus</i> ?
Hypothèse n° 5	La mise en œuvre de <i>Narramus</i> au cours de deux années scolaires devrait permettre de réduire les écarts entre élèves de milieux sociaux différents.	<ol style="list-style-type: none"> 1. Quel degré de transfert peut-on attendre suite à ces interventions ? 2. Les élèves ayant travaillé avec <i>Narramus</i> développent-ils des compétences suffisamment solides pour se rapprocher des résultats obtenus par des élèves de milieu favorisé ?

La méthode, les résultats et l'analyse de l'étude 1 sont présentés dans les chapitres 3, 4 et 5 (hypothèses 1 et 2). Les éléments relatifs à l'étude 2 constituent le chapitre 6 (hypothèses 4 et 5). L'hypothèse n°3 sur le développement professionnel des enseignants et le rôle de l'accompagnement sera traitée dans le chapitre 7.

Partie 2

Étude 1 Un effet *Narramus* ?

Chapitre 3

Méthode de l'étude 1

Chapitre 4

Résultats de l'étude 1 : année 1 (2016-2017)

Chapitre 5

Résultats de l'étude 1 : années 2 et 3 (2017-2019)

Chapitre 3

Méthode de l'étude 1

Chapitre 3 Méthode de l'étude 1

L'objet de cette première étude est de vérifier que la mise en œuvre des caractéristiques des pratiques identifiées comme efficaces sur le plan international produit les effets attendus en France par les concepteurs de *Narramus* qui ont pris soin de respecter les contraintes contextuelles et la culture professionnelle des enseignants. En d'autres termes, nous chercherons à savoir si les élèves qui ont bénéficié de l'outil progressent davantage que ceux qui ont suivi un enseignement ordinaire de la compréhension et de la narration. Si tel est le cas, ces effets s'observent-ils à la fois sur la qualité des rappels des histoires, des réponses aux questions de compréhension et de la mémorisation du lexique enseigné ?

Pour comparer les effets, nous avons constitué deux groupes : un groupe expérimental et un groupe contrôle. Les enseignants qui constituent le premier utilisent *Narramus* quand ceux du second travaillent les mêmes albums comme ils le souhaitent. Puisque *Narramus* a été conçu dans le cadre d'un processus de conception continuée dans l'usage, nous avons émis l'hypothèse qu'il n'aurait pas besoin d'être accompagné pour produire les effets escomptés.

Nous nous sommes donc interrogée sur ce facteur et avons subdivisé le groupe expérimental en deux : un sous-groupe comprend les enseignants outillés et accompagnés par des formateurs de terrain, un deuxième sous-groupe utilise l'outil sans accompagnement.

Notre échantillon est donc composé de trois groupes : G1 (outil avec accompagnement), G2 (outil sans accompagnement) et G3 (pratiques habituelles).

- **Hypothèse 1**

La première hypothèse testée stipule que les élèves bénéficiant de *Narramus* (groupes G1 et G2) progresseront plus que les autres (G3) sur les trois dimensions évaluées (rappel, compréhension et lexique).

- **Hypothèse 2**

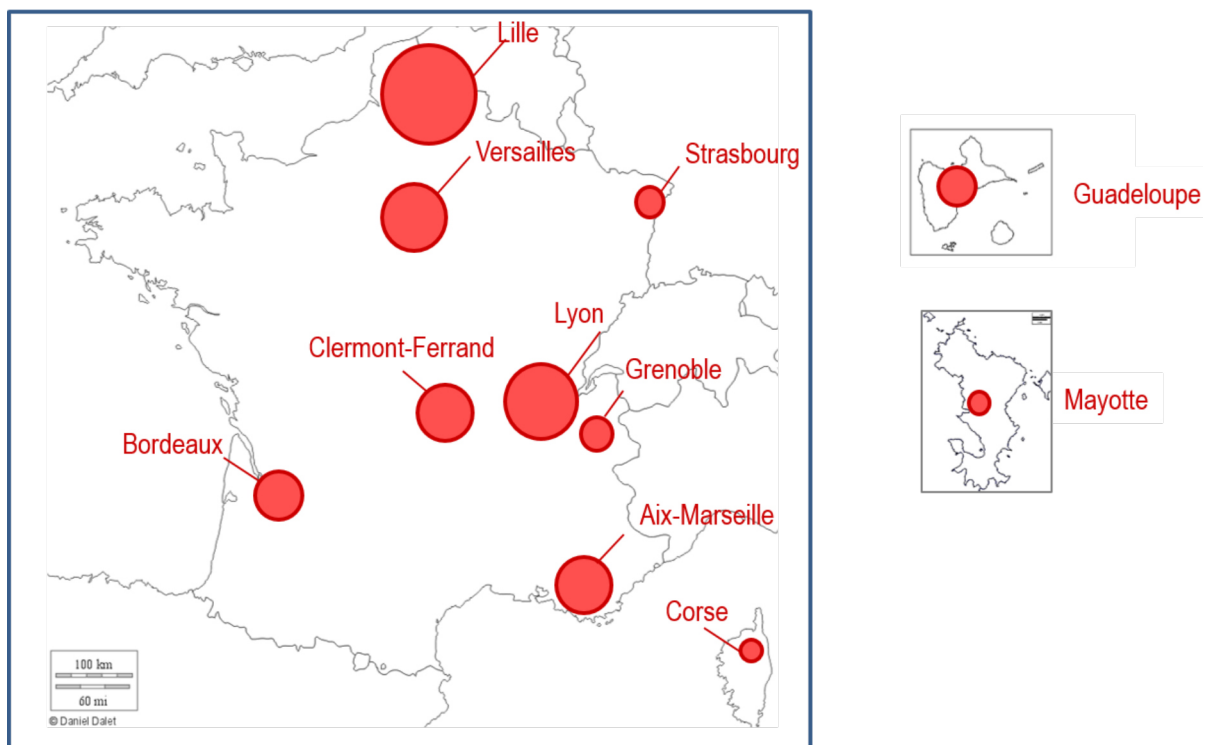
Puisque la conception de *Narramus* s'efforce de respecter les conditions énoncées dans le chapitre précédent, nous pensons que l'outil n'aura pas besoin d'être accompagné pour produire les effets escomptés. Notre seconde hypothèse est formulée ainsi : conçu sur des bases scientifiques attestées et en partenariat avec les enseignants, *Narramus* produira des effets similaires avec ou sans accompagnement.

1. Le dispositif expérimental de l'étude 1

En septembre 2016, à notre initiative et avec l'aide du centre Alain Savary (IFé – ENS Lyon), trente équipes de circonscriptions de onze académies (Aix-Marseille, Bordeaux, Clermont-Ferrand, Corse, Grenoble, Guadeloupe, Lille, Lyon, Mayotte³⁵, Strasbourg, Versailles) ont sollicité des équipes d'enseignants exerçant dans des écoles classées en Éducation Prioritaire (REP + ou REP). Pour favoriser le travail collectif et nous permettre de mener notre étude sur trois ans, trois enseignants au moins par école, un par niveau (PS, MS et GS), ont accepté le contrat expérimental et se sont engagés à le respecter jusqu'en juin 2019.

La Figure 4 présente la répartition géographique des secteurs de REP avec lesquels nous avons travaillé.

Figure 4. Répartition géographique des académies engagées dans la recherche



Cette étude longitudinale a permis d'évaluer des élèves entrant en PS pendant trois années scolaires (PS, MS et GS, de septembre 2016 à juin 2019) et des élèves de MS sur deux années (MS et GS, de 2016 à 2018). En juin 2017 et juin 2018, les formateurs de terrain se sont assurés que les enseignants souhaitent poursuivre la deuxième puis la troisième année de

³⁵ 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.

recherche avec *Narramus*. Aucun retrait n'a été identifié et tous ont poursuivi l'expérimentation jusqu'à la fin de celle-ci (juin 2019). On note toutefois quelques changements dans la composition de l'échantillon professeurs des écoles : quelques enseignants ont intégré le dispositif en année 2 ou en année 3 en raison de la mobilité professionnelle (changements d'affectation ou de niveau). Le dispositif était systématiquement présenté aux maîtres nouvellement nommés qui avaient la possibilité de l'intégrer s'ils le souhaitaient, ce qui a été le cas pour l'ensemble des nouveaux enseignants ; nous avons ainsi pu ne pas perdre trop d'élèves dans notre suivi longitudinal. En année 1 (2016-2017), 396 élèves de petite section (des groupes 1 et 2) ont débuté la recherche ; en année 3, ils étaient 315 à être présents en grande section, ce qui équivaut à un suivi de 79,5 % d'élèves entre le début et la fin de la recherche.

À l'issue de la première année, les enseignants de PS pouvaient poursuivre l'expérience même si leurs nouveaux élèves n'étaient pas évalués : nous leur proposons alors de nouveaux scénarios didactiques. Cette possibilité a également été offerte aux enseignants de MS la troisième année.

1.1. Les séminaires de formation³⁶

Deux séminaires de deux jours de formation ont été organisés en septembre 2016 et mars 2017 à l'Institut français de l'Éducation au cours desquels les accompagnateurs/formateurs³⁷ ont principalement été initiés aux théories qui fondent *Narramus* et à sa mise en œuvre concrète, charge à eux, ensuite, d'accompagner les équipes d'école dans le cadre de leurs missions statutaires.

Le premier séminaire (septembre 2016) visait à :

1° présenter les chercheurs, les objectifs de la recherche et les outils nécessaires au recueil de données. Deux documents importants ont été distribués aux formateurs afin de recueillir des données sur les élèves et les enseignants. Les informations demandées permettent de caractériser les élèves (âge, sexe, langue (s) parlée (s) à la maison, présence en Toute Petite Section, niveau de langage) et les enseignants (ancienneté générale dans le métier, ancienneté en maternelle, statut ou non de formateur).

2° Détailler le calendrier du recueil de données sur l'année : présentation des épreuves de pré-test (utilisées pour constituer l'échantillon « élèves »), la nature des épreuves et leurs

³⁶ Trois autres séminaires ont été assurés lors des deuxième et troisième années de l'étude : deux journées en septembre 2017, une en mars 2018 et une en octobre 2018.

³⁷ Nous emploierons indifféremment les termes formateur ou accompagnateur pour désigner celui qui forme et/ou accompagne, et ce, quel que soit son statut.

consignes, le calendrier de passation, les modalités de saisie des résultats, le cadrage des temps d'enseignement propres à chaque académie en fonction des congés scolaires.

3° Découvrir l'interface de la plateforme³⁸ dédiée à la recherche et indispensable pour saisir les résultats des élèves et déposer les documents (scénarios et épreuves pour le chercheur, enregistrements des élèves, carnets de bord des enseignants et journaux pour les accompagnateurs).

4° Former au codage des données. Toutes les épreuves sur le premier album (rappel, compréhension et lexique) ont d'abord été présentées de manière détaillée aux formateurs qui étaient donc également les évaluateurs des performances des élèves. Ensuite, un entraînement au codage du rappel de récit a été réalisé à partir d'un enregistrement d'élève.

5° Présenter les principes didactiques de *Narramus* : les quatre cibles de l'apprentissage ont été justifiées théoriquement et mises en lien avec les différentes activités et tâches proposées aux élèves dans tous les scénarios.

6° Découvrir le premier scénario d'enseignement *La sieste de Moussa*.

Le second séminaire, organisé en mars 2017, visait à réaliser un premier bilan de la mise en œuvre du scénario par les enseignants travaillant avec *Narramus*, à compléter les apports théoriques mais aussi à assurer des échanges entre formateurs sur les modalités d'accompagnement qu'ils avaient retenues.

Le contenu a été le suivant :

1° bilan et perspective des scénarios : les formateurs ont effectué un retour sur la mise en œuvre par niveau, par type de tâches et par modalités pédagogiques. Ils ont rapporté les propos des enseignants sur leur mise en œuvre des scénarios (réussites, difficultés).

2° Évolutions observées par les enseignants des groupe 1 et 2 : manières de faire et manières de voir.

3° Retour sur les fondements théoriques et empiriques des scénarios.

4° Focus sur l'enseignement du vocabulaire.

5° Les épreuves d'évaluation : justifications théoriques, intérêts et limites et présentation des résultats du premier trimestre.

6° Temps dédié à l'accompagnement des équipes : témoignages de formateurs sur les modalités et les tâches proposées au cours de cette première année d'étude.

7° Présentation de l'année 2 de la recherche expérimentale (2017-2018).

³⁸ Yvonnick Fesselier (ENS-Lyon) a développé et fait vivre cette plateforme, nous le remercions vivement.

Ces séminaires ont été plébiscités par les participants qui y ont trouvé un espace de formation (assurée par des chercheurs), leur permettant de réinvestir les contenus théoriques dans les accompagnements (réalisés pour le G1 uniquement) et de se sentir ainsi mieux armés pour former les enseignants.

Les échanges avec les autres formateurs ont également été appréciés : ils ont permis d'aborder les différentes modalités d'accompagnement (les contenus, les différentes organisations temporelles, les outils utilisés...).

Pour soutenir les activités de formateurs, à l'issue de chaque séminaire, tous les documents supports étaient données aux personnes présentes : power-points utilisés, références bibliographiques, articles de recherche.

Les trois autres séminaires (en 2017-2018 et 2018-2019) ont permis d'approfondir les mêmes questions. Même s'ils n'ont pas eu d'impact sur l'étude 1 réalisée en 2016-2017, ils ont concouru au perfectionnement didactique des formateurs de terrain.

1.2. Les formateurs

Refusant que l'accompagnement soit assuré par les chercheurs pour ne pas condamner l'outil à n'être utilisable que dans des conditions exceptionnelles (Charlier & Peraya, 2003) et pour anticiper sa diffusion à grande échelle, nous avons confié cette tâche aux conseillers pédagogiques et aux inspecteurs de l'Éducation nationale dont le métier est justement de conseiller les enseignants : « Dans l'institution éducative existent des personnels (Inspecteurs, conseillers pédagogiques) dont la fonction est précisément d'accompagner les enseignants et de les aider à perfectionner leur pratique. Si les chercheurs universitaires peuvent contribuer à cet accompagnement, ils ne sauraient se substituer à ces personnels. L'expertise apportée par le chercheur soucieux de méthodologie de la recherche n'est ni celle du conseiller pédagogique, ni celle de l'enseignant lui-même. » (Bru, 2002, p. 66).

Dans notre étude, les formateurs assuraient un rôle essentiel, celui de « passeur », entre le travail des enseignants et celui des chercheurs. Ils avaient pour mission 1° d'assurer la mise en œuvre et le suivi de la recherche sur le terrain et 2° de former les enseignants du groupe accompagné (G1).

Ainsi, ils prenaient en charge différentes tâches indispensables au bon déroulement de la recherche :

- la diffusion du matériel pédagogique,
- la formation des enseignants du G1 (groupe accompagné),
- la rédaction d'un « journal de l'accompagnateur »,

- la coordination avec les enquêteurs qui assurent l'évaluation des élèves,
- le recueil des carnets de bord des enseignants.

Tous les trois mois, les accompagnateurs diffusaient³⁹ aux enseignants (du G1 et du G2) les scénarios et leurs annexes (le guide du maître, les supports numériques power-point qui accompagnent toutes les séances, les enregistrements audio de l'histoire, le dessin animé le cas échéant).

Ils planifiaient et menaient les séances de formation comme ils le souhaitaient. Ils pouvaient, par exemple, présenter la recherche, les principes qui sous-tendent les propositions didactiques des scénarios et leurs justifications théoriques, décrire et justifier les tâches emblématiques de *Narramus*, présenter en détail un scénario. Ils pouvaient également faire écouter aux maîtres les enregistrements de récits pour que ces derniers se fassent une représentation de ce que l'usage dudit scénario dans différentes classes de REP + a donné comme résultat à Clermont-Ferrand - lors de la phase de conception - en PS, MS et GS.

Les modalités de cet accompagnement (présentées plus bas) ont été définies localement, en concertation avec les enseignants. Nous avons toutefois effectué quelques demandes qui ont certainement orienté l'organisation des formations :

1° les accompagnateurs s'engagent à offrir au G1 une action de formation spécifique – avant la mise en œuvre du premier scénario – dans laquelle ils présentent *a minima* le contrat qui lie les enseignants à la recherche, les scénarios didactiques et le carnet de bord que ces derniers doivent renseigner après chaque séance réalisée ;

2° ils peuvent, s'ils le jugent possible et souhaitable, accompagner les équipes pendant la phase de mise en œuvre des scénarios.

3° ils doivent, à la fin du scénario, organiser une réunion de débriefing et de bilan selon des modalités de leur choix. Cet échange peut porter sur

- la mise en œuvre du scénario lui-même (les réussites et les difficultés de mise en œuvre),

- sur les enseignants eux-mêmes : qu'ont-ils appris ? De quoi sont-ils fiers ? Sur quels aspects précis la mise en œuvre du scénario a-t-elle infléchi leurs pratiques d'enseignement ? En quoi leur regard sur les élèves a-t-il changé ? Leurs conceptions et/ou connaissances didactiques et pédagogiques ont-elles été affectées ? Ont-ils observé des changements dans

³⁹ L'accès à la plateforme qui héberge les documents pour les enseignants et les résultats des élèves est exclusivement réservé aux chercheurs et aux accompagnateurs. L'adresse n'est pas communiquée aux enseignants qui reçoivent une clé USB sur laquelle tous les documents sont rassemblés.

leurs relations avec les parents ? Ont-ils été sensibles à l'accompagnement dont ils ont bénéficié ?

Compte tenu de l'hypothèse majeure de notre étude touchant le rôle de l'accompagnement dans la diffusion des innovations, les équipes doivent renseigner un « journal de l'accompagnateur » (au moins trois fois dans l'année) qui garde la trace des actions de formation menées, du travail réalisé dans et hors les classes, des questions auxquelles ils ont eu à répondre, des discussions voire des controverses... Ces journaux décrivent *a minima* les objectifs, la durée, les modalités et les contenus de l'accompagnement.

1.3. Les évaluateurs

Dans la majorité des circonscriptions, les formateurs-accompagnateurs étaient aussi évaluateurs mais, compte tenu de la nature des épreuves et du nombre d'élèves testés, ils ont recruté d'autres professionnels pour les seconder. Tous étaient membres de l'Éducation nationale : conseillers pédagogiques, inspecteurs, formateurs en éducation prioritaire, enseignants spécialisés. Au total, cette recherche a donc mobilisé une centaine d'évaluateurs.

La plupart n'avaient la responsabilité que d'une seule classe (6 élèves), certains deux ou trois au maximum (soit 18 élèves). Notre situation personnelle était particulière. Au cours de la première année de l'étude, nous avons suivi 84 élèves sur Clermont-Ferrand, ville dans laquelle trois écoles étaient engagées, soit 13 classes. Le codage des épreuves et la saisie sur la plateforme étaient de notre ressort.

Sur les autres sites, les évaluateurs devaient administrer les épreuves, les coder, puis placer les enregistrements audio et leurs codages sur la plateforme numérique.

Tous ont bénéficié d'une formation assurée par les équipes de circonscription qui leur ont présenté les évaluations, les modalités de passation (consignes, relances...) et de codage. En cas de difficulté ou de doute, ils pouvaient s'adresser aux formateurs ou prendre directement contact avec nous pour obtenir des précisions.

Au cours des trois années de la recherche et particulièrement des deux premières, nous avons eu des échanges très réguliers avec les formateurs/accompagnateurs et les évaluateurs. Nous avons, en effet, assuré le pilotage et le suivi des évaluations : l'accès à la plateforme ; la saisie des résultats, l'absence des élèves, un doute sur la manière de coder une production enfantine... Quand, pour une raison ou une autre, les évaluateurs ne pouvaient réaliser les codages (manque de temps ou incertitude dans la manière de coder) nous nous sommes chargée des codages (à partir des enregistrements audio) et de la saisie des résultats (à partir des protocoles sur papier).

Durant l'étude, nous n'avons pas eu de contact direct avec les enseignants, hormis ceux de notre académie. Ce sont les évaluateurs et les formateurs qui assuraient le relais sur le terrain.

1.4. Les enseignants

Avant que les enseignants acceptent de participer à notre recherche, les formateurs ont apporté différents éléments définis et présentés préalablement par les chercheurs, lors du premier séminaire (septembre 2016).

1.4.1. Informer les enseignants avant le début de la recherche

Une « charte de l'accompagnateur » synthétisant les informations relatives à la recherche a été adressée à tous. Elle indiquait qu'avant le début de l'étude les accompagnateurs devaient présenter la recherche aux enseignants des trois groupes.

Les professeurs des écoles ont ainsi été informés :

1° du contexte scientifique dans lequel s'inscrit cette étude : validée par trois experts sollicités par l'Institut Carnot de l'Éducation, elle fédère des chercheurs du laboratoire ACTÉ, des équipes de circonscription et des enseignants d'une vingtaine de départements de France entière. Elle est également soutenue par le ministère de l'Éducation nationale, de l'enseignement supérieur et de la recherche ainsi que par l'Institut Français de l'éducation et le Centre Alain Savary.

2° Des principes éthiques retenus : cette étude respecte la charte internationale pour la recherche éthique impliquant les enfants et les préconisations touchant les meilleures pratiques dans le domaine de la recherche. Notre préoccupation éthique concerne également les enseignants qui participent à l'étude. Une fois les résultats des élèves entrés sur le site de l'IFÉ, toutes les données nominatives sont codées et anonymées si bien que plus aucun nom de circonscription, d'école, d'enseignant ou d'élève n'apparaît nulle part. Autrement dit, excepté l'évaluateur des élèves qui s'engage à respecter la confidentialité, nul ne peut associer un résultat à une personne physique. Ce sont les comparaisons entre groupes expérimentaux qui nous intéressent, pas entre écoles ou classes.

En outre, conformément à l'article 9 du Code civil touchant le droit à l'image et à l'enregistrement de la voix, l'ensemble des participants à la recherche s'est engagé à ne pas diffuser et ne pas utiliser en public les enregistrements sonores réalisés. Ceux-ci ne sont traités que par les chercheurs impliqués dans le projet.

3° De la nature des évaluations : les épreuves que nous avons retenues sont classiquement utilisées en contexte scolaire et évaluent des connaissances et des compétences

qui sont explicitement inscrites au programme de l'école maternelle (MEN, 2015). Les évaluateurs sont tous personnels (actifs ou retraités) de l'Éducation nationale. Nous avons recommandé aux équipes d'organiser une réunion d'école pour expliquer aux parents ce que leurs enfants allaient apprendre et les informer que ces derniers seraient régulièrement évalués.

4° Du document qu'ils auraient à renseigner systématiquement : un « carnet de bord » (nous y reviendrons).

Seuls les enseignants des groupes 1 et 2 ont été informés que les scénarios *Narramus* respectent rigoureusement les prescriptions du programme de l'école maternelle et les pistes proposées pour leur mise en œuvre dans les « ressources d'accompagnement » (MEN, 2015).

1.4.2. Les engagements des enseignants des groupes 1 et 2

Pour ne pas compromettre l'étude, les enseignants des groupes G1 et G2 s'engagent 1° à ne pas diffuser les documents fournis par les chercheurs et les accompagnateurs et 2° à étudier les différents albums en suivant le plus fidèlement possible les scénarios décrits dans le guide du maître et 3° à utiliser les supports numériques créés par les auteurs.

Ils acceptent qu'une partie seulement de leurs élèves (6) soit évaluée à plusieurs reprises au cours de l'année et de rédiger un carnet de bord.

Au-delà de ces consignes très précises, les enseignants restent les seuls responsables de leur classe : ils sont donc libres d'ajuster les tâches ou les modalités de groupement au niveau des élèves et à leur contexte d'exercice. Ils peuvent tout à fait, par exemple, confier un petit groupe au maître supplémentaire, au maître E ou à tout autre adulte qui travaille avec eux.

C'est à dessein, précisent les auteurs, que la durée des modules n'est jamais indiquée : seul l'enseignant connaît le rythme d'apprentissage de ses élèves et leurs capacités d'attention. Ils peuvent donc, s'ils le jugent pertinents, scinder certains modules en deux voire en trois périodes.

Les enseignants du G1 s'engagent enfin à participer aux différentes rencontres organisées par l'équipe d'accompagnateurs.

Nous avons invité lesdites équipes à garder une mémoire des enseignants absents lors de ces rencontres. Nous avons décidé, en amont, que ceux qui, à la fin de l'année, auraient manqué plus d'une fois seraient considérés comme faisant partie du G2. Aucun professeur n'a été amené à changer de groupe : les enseignants du G1 ont tous suivi les accompagnements proposés par les équipes de circonscription.

1.4.3. Les engagements des enseignants du groupe 3

Le contrat qui nous lie avec les maîtres du groupe 3, qui constitue notre groupe contrôle, est le plus complexe. En effet, nous avons besoin d'eux pour savoir si des pratiques différentes produisent des effets contrastés sur les apprentissages des élèves. Aussi était-il nécessaire de leur présenter la recherche pour qu'ils puissent identifier le rôle déterminant qu'ils jouent dans la réussite de notre entreprise :

1° sans eux, nous ne serions pas en mesure de savoir si les propositions des auteurs sont bien innovantes et donc différentes des pratiques pédagogiques habituelles (c'est-à-dire les pratiques qui préexistent à l'outil *Narramus*) et, si oui, sur quels aspects.

2° sans eux, nous ne pourrions pas savoir si les scénarios didactiques de l'outil apportent aux élèves de G1 et G2 un réel bénéfice qu'une simple mobilisation sur la compétence à raconter (par exemple une recommandation dans le cadre d'une conférence pédagogique) ne suffirait pas à provoquer.

Nous avons toutefois attiré l'attention des formateurs sur l'importance de ne pas trop en dire et de ne surtout pas diffuser les principes didactiques et pédagogiques présents dans *Narramus*.

Dès le début de la recherche, les enseignants du groupe 3 savaient qu'ils seraient destinataires des outils *Narramus* (dont ils ignoraient tout) à la fin de l'année et qu'ils pourraient bénéficier d'un accompagnement en année 2, en fonction des possibilités et de l'organisation de chaque équipe de circonscription.

1.4.4. Les caractéristiques des enseignants

L'étude 1 concerne 250 classes de petite, moyenne et grande sections de 79 écoles. 90,1 % des 250 professeurs des écoles sont des femmes. En moyenne, leur ancienneté professionnelle est de 16,3 ans (écart-type = 9,7) et leur expérience en école maternelle de 11,1 ans (écart-type = 8,1).

Nous avons sollicité les équipes de circonscription pour répartir les écoles dans les différents groupes en leur demandant de :

- veiller à ce que l'expérience et la motivation des maîtres concernés s'équilibrent au sein d'une même circonscription ;
- recruter au moins 3 écoles, 1 par groupe pour assurer une diversité territoriale de bon aloi.

L'ensemble donne une étude qui n'est pas randomisée au niveau des écoles mais qui s'en approche par la pluralité des critères de choix. : la diversité intra-groupe est aussi grande dans chacun des sous-groupes.

Après l'affectation des écoles dans trois groupes, nous avons vérifié qu'ils étaient équivalents pour le critère de l'ancienneté en maternelle : 11,0 ans pour le groupe 1 (écart-type = 8,4), 11,1 ans pour le groupe 2 (écart-type = 7,6) et 11,3 ans pour le groupe 3 (écart-type = 8,1). Les différences ne sont pas significatives ($p < 0,001$). Ce sont donc bien des professeurs expérimentés qui contribuent à l'évaluation de *Narramus*.

1.4.5. Le contrat établi avec les enseignants de chaque groupe

Les enseignants du groupe 1 ont reçu les trois albums et les trois scénarios *Narramus* (un chaque trimestre), ils les ont mis en œuvre en bénéficiant de temps de formation et d'accompagnements assurés 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 s'engageaient, eux-aussi, à les faire étudier à leurs élèves durant les quatre semaines qui séparaient le pré-test du post-test et, comme les autres, étaient informés de la nature des évaluations. Tous savaient, en effet, que leurs élèves seraient interrogés sur le lexique de l'histoire puis qu'ils devraient la raconter seuls et sans support.

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 maitres de maternelle. Le fait d'être enrôlés dans une recherche, de savoir que les élèves vont être évalués, de connaître la nature des épreuves, et de disposer de quatre semaines pour travailler sur l'album modifie sans nul doute les manières de faire (*cf.* chapitre 2).

Mais, nous assumons ce choix: dans la mesure où nous voulions évaluer les effets de l'utilisation d'un outil didactique spécifique sur la compréhension, nous devons créer les conditions permettant une comparaison inter-groupes valide, et donc garantir que les élèves des trois groupes recevraient la même quantité d'enseignement de la compréhension et que leurs maitres étaient au clair sur la nature des compétences évaluées. La création d'un groupe contrôle motivé (son aventure continuera l'année suivante avec l'accès aux scénarios *Narramus*), appelé à faire valoir son savoir-faire professionnel, disposant d'un temps conséquent pour enseigner et averti des normes évaluatives ne fait que renforcer l'exigence de la comparaison. Si les résultats des groupes expérimentaux sont positifs, ils n'en auront que plus de valeur.

1.5. Les élèves

Étant donné le nombre de classes engagées dans la recherche (N = 250) et la nature des évaluations (rappel de récit enregistré, questionnaires de compréhension et de lexique), nous ne pouvions tester individuellement tous les élèves (N = 6 250). Nous devons 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 de moyenne et grande sections nés entre les mois de mars et septembre et nous les avons rangés en trois terciles (fort, intermédiaire, faible). 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 (sur les 1 500 élèves qui constituent notre échantillon final, seuls 3 ont été exclus au vu de la discordance entre le jugement de l'enseignant et le résultat au test et donc remplacés par 3 autres présentant un profil cohérent).

Le résultat à l'épreuve É.Co.S.Se. – mis en lien avec les avis donnés par l'enseignant – a donc permis de choisir 6 élèves dont les performances ont été évaluées tout au long de l'année.

En résumé, nous avons choisi, comme il est d'usage dans ce type d'étude (Cain, Jiang, & Logan, 2019; Neuman & Kaefer, 2018; Zucker, Cabell, Justice, Pentimonti, & Kaderavek, 2013), de ne 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.

Compte tenu de l'âge des élèves de Petite Section et de leurs compétences langagières, nous n'avons pas pu recourir au test É.Co.S.Se. Nous avons donc demandé à chaque enseignant de choisir lui-même six élèves en respectant les mêmes critères (âge, parité garçons-filles, niveau).

On trouvera, dans le Tableau 9, la répartition des élèves selon le niveau scolaire et le groupe d'appartenance.

Tableau 9. Nombre de classes et d'élèves évalués dans chaque groupe en PS, MS et GS

Groupe	Nombre de classes				Nombre d'élèves testés			
	PS	MS	GS	Total	PS	MS	GS	Total
G1	40	40	35	115	240	240	210	690
G2	26	28	25	79	156	168	150	474
G3	19	19	18	56	114	114	108	336
Total	85	87	78	250	510	522	468	1500

L'échantillon final est composé de 1 500 élèves (49,9 % de filles et 50,1 % de garçons) ; 510 viennent de 85 classes de PS ; 522 sont issus de 87 classes de MS et 468 proviennent de 78 classes de GS.

62,6 % d'élèves parlent uniquement le français à la maison contre 6,9 % qui ne le parlent pas. Les autres, 30,5 %, parlent au moins deux langues à la maison dont le français.

690 élèves font partie du groupe 1 (G1 = *Narramus* avec accompagnement), 474 du groupe 2 (G2 = *Narramus* sans accompagnement) et 336 du groupe 3 (G3 = « pratiques habituelles »).

1.6. Les albums

Les élèves de chaque niveau ont étudié trois albums de littérature de jeunesse au cours de l'année scolaire (un par trimestre).

On trouve, dans le Tableau 10 la liste des albums exploités par les enseignants des trois sections des trois groupes lors de la première année d'expérimentation. Pour le premier trimestre, les auteurs ont proposé le même album aux élèves des trois sections de maternelle et ce, dans le but de favoriser les échanges entre les enseignants lors de la prise en main de l'outil innovant. Basé sur une histoire relativement facile à comprendre et présentant peu de difficultés lexicales et syntaxiques, cet ouvrage permet de familiariser tous les élèves avec le déroulement des modules ainsi qu'avec les tâches et les activités qui seront reprises dans tous les autres scénarios. L'histoire s'organise autour d'une succession d'épisodes comparables, son étude peut en outre initier efficacement aux enjeux esthétiques des contes sériels.

Tableau 10. Albums étudiés (par niveau) dans les trois groupes expérimentaux en 2016/2017

	Titres et auteurs		
	PS	MS	GS
Trimestre 1	La sieste de Moussa ⁴⁰ Zémanel	La sieste de Moussa Zémanel	La sieste de Moussa Zémanel
Trimestre 2	Le Machin Stéphane Servant et Cécile Bonbon	Le gros chagrin d'Edgar Xavier Gorce et Yves Calarnou	Gruffalo Julia Donaldson et Alex Scheffler
Trimestre 3	Un poussin de mauvais poil Sylvie Poillevé et Nadine Rouvière	La sorcière dans les airs Julia Donaldson et Alex Scheffler	Les deniers de Compère lapin Michèle Simonsen et Magali Le Huche

⁴⁰ Toutes les séances ne sont pas à mener avec les élèves de petite section : par exemple, celles portant sur les états mentaux des personnages sont réservées à ceux des moyennes et grandes sections.

1.6.1. Les instruments de recueil des données

Nous présentons ici les outils utilisés pour recueillir nos données et observer les effets sur les apprentissages des élèves.

1.6.2. Les épreuves d'évaluation

Les épreuves d'évaluation ont permis 1° de sélectionner les élèves évalués par classe en Moyenne et Grande section et 2° de mesurer les performances des élèves de petite, moyenne et grande sections en compréhension.

1.6.2.1. L'organisation temporelle des épreuves

Les évaluations étaient toutes individuelles et nécessitaient plusieurs déplacements dans les écoles. Le pré-test de lexique était toujours réalisé avant la première écoute de l'album alors que les pré-tests de rappel et de compréhension étaient administrés quelques jours plus tard, après les deux écoutes de l'histoire dans la classe. Cette deuxième partie du pré-test était systématiquement organisée immédiatement après la deuxième écoute de l'histoire. Le Tableau 11 présente une « organisation temporelle-type » du pré-test au post-test.

Tableau 11. Organisation temporelle du pré-test au post-test

Pré-test (partie 1)	2 écoutes de l'album (1 narration et 1 lecture)	Pré-test (Partie 2)	Enseignement avec album (4 semaines)	Post-test
Lexique		Rappel Questionnaire		Lexique Rappel Questionnaire

Les procédures, les conditions de passation, les critères de correction et les barèmes ont été définis très précisément dans un cahier des charges. Il est indispensable que les modalités de passation des évaluations soient identiques dans tous les lieux et dans toutes les classes si l'on veut éviter un biais et pouvoir comparer les données.

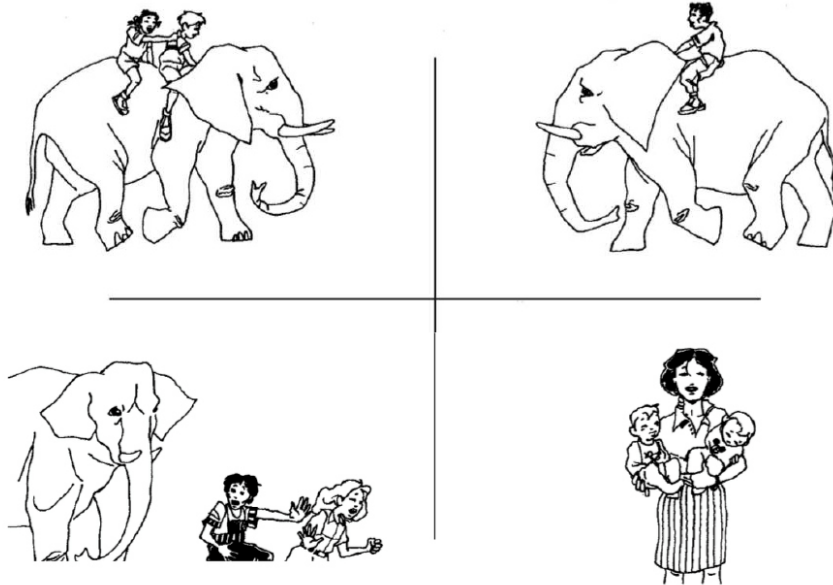
1.6.2.2. Épreuve de sélection des élèves de MS et GS : compréhension de phrases entendues

L'épreuve épreuve de compréhension syntaxico-sémantique (É.Co.S.Se.⁴¹, Lecocq, 1996) vise à évaluer la compréhension de phrases courtes, simples sur le plan lexical, mais dont la complexité syntaxique va croissante.

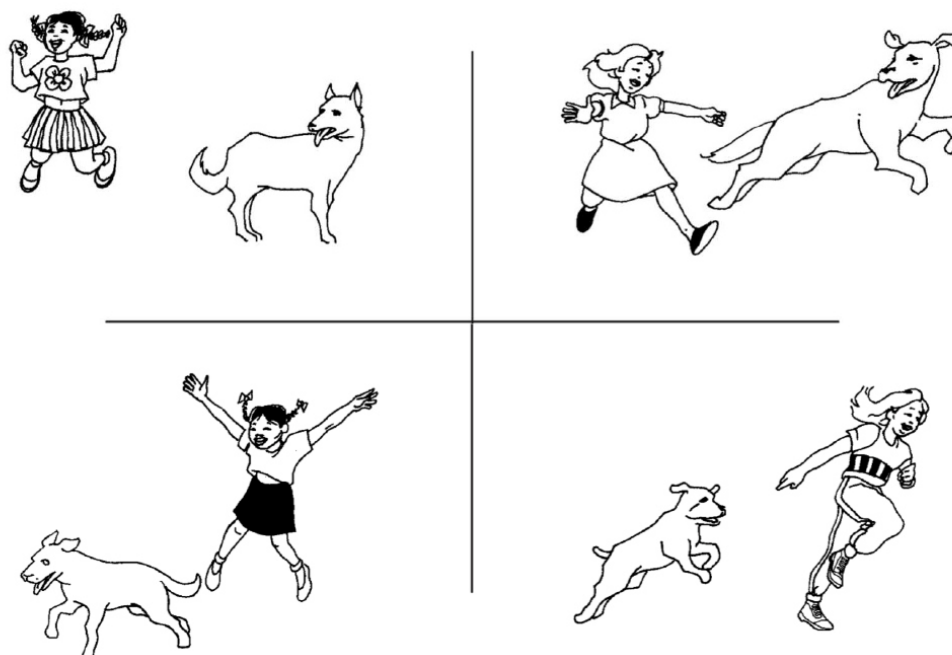
⁴¹ Version 10 planches pour le *Test d'évaluation Compréhension MS* et celui de *GS/CP* réédités et diffusés en ligne par La Cigale [https : //www.editions-cigale.com/424224cc](https://www.editions-cigale.com/424224cc)

L'épreuve est constituée de onze planches, les supports étant différents pour les élèves de MS et de GS. La première sert d'exemple et vise à garantir la compréhension de la consigne par l'enfant. La passation est individuelle. 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. Un point étant alloué pour chaque réponse correcte, le score maximum est de 10.

Exemple *Item 2* Grande section
L'éléphant les porte.



Exemple *Item 2* Moyenne section
La fille poursuit le chien qui saute.



1.6.2.3. Épreuves d'évaluation des apprentissages

Nous avons conçu trois épreuves originales - rappel de récit, lexique et questionnaire - à partir des albums étudiés. Elles ont été proposées, à l'identique, avant enseignement (pré-test) et après enseignement (post-test). Les élèves ont donc été testés six fois à propos de trois albums différents comme l'indique le Tableau 12.

Les passations étaient individuelles et prenaient environ 20 minutes pour le pré-test (5 minutes pour la partie 1 et 15 minutes pour la partie 2, cf. tableau 11 page 127) et 30 minutes pour le post-test. Les rappels de récit ont tous été enregistrés sur support numérique. Après enregistrement, les récits des élèves ont tous été déposés sur la plateforme dédiée.

Tableau 12. 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	Juin 2017
Tests	É.Co.S.Se.	Pré-test et post-test album 1 : rappel ⁴² , lexique	Pré-test et post-test album 2 : rappel, lexique, questionnaire	Pré-test et post-test album 3 : rappel, lexique, questionnaire	Post-test différé de lexique albums 1 & 2

⁴² L'épreuve « questionnaire » n'a pas été proposée pour le premier scénario commun aux trois sections de maternelle. Nous en exposons les raisons plus loin dans la présentation de l'épreuve questionnaire.

1.6.2.3.1. Rappel de l'histoire

La tâche de rappel de récit est une tâche classique en psycholinguistique (Blanc, 2010) qui vise à mesurer la compréhension qu'ont les élèves des éléments de l'histoire. C'est une épreuve exigeante mais plus fiable que le questionnaire même si les résultats entre ces deux épreuves sont fortement corrélés (Lever & Sénéchal, 2011).

Pour le pré-test, les élèves sont évalués après deux écoutes de l'histoire. L'enseignant raconte une première fois l'histoire sans montrer les illustrations (jour 1). Puis il 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.

Pour tous les élèves, qu'ils soient en petite, moyenne ou grande section, cette épreuve est nouvelle. Ils peuvent donc être déstabilisés par la consigne (raconter) et par les conditions de passation (raconter à quelqu'un qu'ils ne connaissent pas). L'évaluateur avait donc à soutenir l'enfant grâce à un certain nombre de commentaires de « facilitation ». Ces derniers pouvaient prendre la forme de relance « et puis, et alors, et après ? », de répétitions des propos de l'enfant ou de reformulations. Rappelons que tous les évaluateurs ont été préalablement formés et savaient que si un enfant ne dit pas un mot ou refuse de répondre (cela arrive, même en GS) une seule petite aide (« alors c'est l'histoire de quoi ? C'est l'histoire d'un garçon ? ») suffit parfois à faire démarrer ou poursuivre le rappel. Si malgré tout, l'enfant disait ne pas savoir, se taisait ou refusait de raconter, il avait été convenu de ne pas insister et de se contenter alors de coder 0 chacune des propositions.

Nous avons demandé aux évaluateurs de ne pas coder en présence des enfants pour qu'ils puissent accorder toute leur attention au récit de ces derniers. Ils devaient donc le faire plus tard en réécoutant l'enregistrement. Si cette consigne est coûteuse pour eux, elle garantit, pour nous, une plus grande précision des codages puisque, en cas de doute, ils pouvaient réécouter l'enfant autant de fois que nécessaire (ou nous demander d'assurer le codage).

Le rappel a été introduit en absence des images ou de tout autre support pour :

1° ne pas proposer un support connu des uns et inconnu des autres en fonction de l'appartenance au groupe expérimental,

2° diminuer la tendance qu'ont certains enfants à se contenter de décrire les images qu'ils ont sous les yeux (Veneziano, 2010).

La tâche ainsi définie est très exigeante sur le plan cognitif (recherche en mémoire et verbalisation) mais elle l'est pour tous de la même manière.

La qualité des rappels a été évaluée en comptant le nombre d'éléments rappelés. Chaque rappel est codé deux fois comme il est de rigueur en psycholinguistique (Hughes, McGillivray, & Schmidek, 1997; Justice et al., 2006). Le premier codage porte sur la macrostructure, soit le nombre d'idées essentielles rappelées (variable « M »), le second, sur la microstructure soit le nombre de propositions sémantiques relatées par l'élève (variable « mi »). « L'analyse macrostructurale examine les capacités narratives des enfants en termes d'organisation hiérarchique d'ordre supérieur. Elle met généralement l'accent sur l'inclusion, par les enfants, des composantes grammaticales des histoires et sur la complexité de la structure des épisodes. [...] L'analyse microstructurale, en revanche, tient compte des structures linguistiques internes utilisées dans la construction narrative, comme les conjonctions, les groupes nominaux et les propositions dépendantes » (Justice et al., 2006, p. 178). Ainsi, les variables macro et micro structurelles représentent deux éléments distincts de la composante narrative, les deux codages donnent des informations différentes mais complémentaires.

Les travaux de Kintsch et Van Dijk (1975) nous ont également servi d'appui pour élaborer nos grilles d'analyse des rappels des élèves. Pour ces auteurs, l'évaluation des récits questionne leur structure sémantique et leur structure « logique » sous-jacente. La représentation sémantique du discours est définie comme un « n-tuple de propositions ». Une proposition est composée d'un prédicat suivi d'une séquence d'arguments. Les propositions peuvent être « composées » (deux propositions sont liées par paire au moyen d'un connecteur) ou « complexes » (une proposition équivaut à un de ses arguments). Le texte correspond à la structure formelle, grammaticale, d'un discours et la base du texte à la structure sémantique sous-jacente au texte. La base du texte est donc un « n-tuple de propositions ». La base de texte implicite se distingue de la base de texte explicite. La première correspond à une séquence de propositions qui composent l'entrée des règles d'expression et évince toutes les propositions qui dénotent des faits généraux ou particuliers supposés connus de l'auditeur par le locuteur. Dans la seconde, toutes ces propositions sont présentes afin de définir la bonne forme théorique de la base du texte, c'est à dire sa cohérence. Les règles de cohérence sont définies à deux niveaux : la base de texte linéaire (propositionnelle) et la macrostructure. La microstructure correspond à la structure locale, inter-phrases et la macrostructure à la structure de signification globale d'un texte et donc à un ensemble de propositions. « Il n'est pas nécessaire qu'une macrostructure (ni le résumé qui l'exprime) soit constituée de propositions apparaissant dans le texte : une macrostructure peut résumer une séquence entière de propositions du texte. Il s'ensuit que tout modèle cognitif qui définirait les macrostructures et les résumés par simple omission de propositions serait inadéquat. » (Kintsch & Van Dijk, 1975, p. 102). Un même texte peut

disposer de plusieurs macrostructures de même niveau. Les règles qui agissent sur les microstructures et qui produisent les macrostructures respectent un principe de « pertinence relative », c'est-à-dire qu'on ne peut supprimer une proposition si celle-ci est une présupposition d'une autre macro-proposition. Au-delà de ces principes sémantiques généraux et afin de déterminer le contenu d'une macro-proposition et l'organisation interne de la macrostructure d'un récit, Kintsch et Van Dijk ont établi des règles déterminant quelle information d'un récit on tend à se rappeler et comment l'information est organisée. Différents éléments fixent ainsi la structure du récit : « une situation initiale doit être décrite avec la caractérisation des agents, des propriétés, du lieu, du moment, et des circonstances physiques et socio-culturelles » (idem, p. 104). Viennent ensuite la « complication » et la « résolution » de l'histoire qui sont parfois suivies « par une évaluation qui spécifie les réactions (mentales) de l'agent/narrateur à l'épisode, et par une morale tirant les conséquences possibles de toute l'histoire pour la situation présente » (idem, p. 104).

Lorsque nous avons élaboré les 14 grilles de codage (7 macrostructures et 7 microstructures, cf. annexe 2) pour les 7 albums étudiés en 2016-2017⁴³, nous avons pris en compte ces caractéristiques. Pour Kintsch et Van Dijk (1975), lors d'un rappel différé dans le temps, les propositions de la macrostructure rappelées sont plus nombreuses que celles de la microstructure. À l'instar de ces auteurs, nous n'avons pas retenu les propositions ajoutées qui faisaient l'objet d'inférences.

Dans notre étude, lors du codage, les récits ont reçu un point pour chaque élément correctement relaté. Un masque composé de la liste des propositions sémantiques était placé sur la plateforme : l'évaluateur⁴⁴ devait cocher 1 si l'enfant avait relaté la proposition, 0 dans le cas contraire. Toute proposition de sens proche (paraphrase, reformulation...) était codée 1. Par exemple, pour la proposition « *Moussa est très fatigué* », étaient acceptées « *Moussa il a trop sommeil* », « *Moussa i veut dormir* », « *Moussa i veut faire la sieste* ».

Les scores bruts pour chaque épreuve variaient en fonction du texte de l'album, comme le montre le Tableau 13.

⁴³ Les macrostructures et microstructures des albums étudiés en année 2 et 3 ont également été élaborées par nos soins. Nous avons inclus celles de la deuxième année dans les annexes.

⁴⁴ Des entraînements au codage avaient été préalablement effectués avec les formateurs et des consignes précises avaient été données.

Tableau 13. Nombre de propositions macrostructurale et microstructurale pour chaque album étudié

	PS		MS		GS	
	Macro	Micro	Macro	Micro	Macro	Micro
Album 1	20	72	20	72	20	72
Album 2	24	63	23	75	30	82
Album 3	19	37	42	97	48	109

Excepté pour les histoires étudiées en PS, le nombre de macro-propositions et micro-propositions augmentent pour les albums proposés aux élèves de MS et GS : le contenu des albums se complexifie avec un texte plus conséquent à chaque fois. Prenons ici l'exemple des GS pour lesquels les trois textes étudiés sont des contes en randonnée. Dans le premier album⁴⁵, les principaux obstacles à la compréhension viennent de l'ordre d'apparition des personnages et de l'agacement que ceux-ci provoquent chez le personnage principal Moussa. Dans le second⁴⁶, les personnages successifs sont moins nombreux : ce sont leurs états mentaux et leurs intentions qui constituent des obstacles importants à la compréhension. En outre, le texte et le lexique sont plus denses. Dans le dernier album⁴⁷, les difficultés résident à la fois dans l'importance de l'ordre d'apparition des personnages, des états mentaux et des intentions des protagonistes et du lexique sophistiqué.

1.6.2.3.2. Lexique

Cette épreuve vise à mesurer 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). Les auteurs du *National Reading Panel* (2000) soutiennent que l'augmentation spécifique du vocabulaire est mieux évaluée par des épreuves développées par les chercheurs (ou les maîtres) que par des tests standardisés qui évaluent des mots souvent non enseignés et qui ne permettent donc pas de rendre compte des apprentissages réalisés. C'est pourquoi nous avons dû créer une épreuve de lexique pour chaque album qui intègre plusieurs modalités d'évaluation, identiques dans chaque test (nous y reviendrons). Nous avons, par conséquent, créé 7 épreuves (cf. annexe 3)⁴⁸ différentes pour la première année de la recherche (une par album étudié). Comment avons-nous sélectionné les mots évalués ? La partie suivante répond à cette question.

⁴⁵ *La sieste de Moussa*

⁴⁶ *Gruffalo*

⁴⁷ *Les deniers de Compère lapin*

⁴⁸ Les grilles de codage de l'épreuve de lexique de l'année 2 sont incluses dans les annexes.

1.6.2.3.2.1. Les mots évalués

Les mots que nous avons évalués étaient tous présents dans les albums étudiés. Pour chaque test lexical, les mots ont été sélectionnés pour représenter, d'une part, un vocabulaire qui ne serait probablement pas familier des élèves tout en étant important pour l'histoire (Coyne et al., 2004) et, d'autre part, des mots courants utilisés dans les conversations entre adultes (Beck, McKeown, & Kucan, 2002). Si ces études nous ont donné une première directive, une question persistait : comment juger de l'importance ou de l'utilité des mots ? Le modèle en trois niveaux de Beck et McKeown (Beck, McKeown, & Kucan, 2013; McKeown & Beck, 2004) nous a aidée à choisir. En effet, le niveau 1 correspond aux mots que l'on trouve habituellement dans les conversations quotidiennes, le niveau 2 aux mots plus sophistiqués, fréquemment employés dans les conversations entre adultes et les textes écrits, et le niveau 3 comprend des mots rarement utilisés par les adultes ou attachés à des domaines très spécifiques (technologies, sciences...). Ce sont donc les mots des niveaux 1 et 2 que nous avons introduits dans toutes nos épreuves.

On pourrait s'étonner que nous ayons conservé des mots de niveau 1 puisqu'ils risquent d'être bien connus des jeunes enfants et, partant, peu sensibles aux progrès. Mais nous assumons ce choix compte tenu de la population d'élèves que nous étudions : il s'agit d'enfants issus, dans leur grande majorité de milieux populaires et dont, pour certains, le français n'est pas la langue première⁴⁹. Nous avons donc supposé que « les mots même simples ne peuvent rien évoquer » parce qu'ils « n'ont jamais été entendus, utilisés, et ce d'autant plus que les situations scolaires sont inédites. » (IGEN, 2011, p. 129). Enfin, nous avons utilisé la liste de fréquence lexicale disponible sur le site Éduscol⁵⁰.

Beck et McKeown ajoutent que les mots choisis doivent évidemment être compréhensibles pour de jeunes élèves et que, si les mots de niveau 2 sont nombreux, le choix se porte sur des mots pouvant être expliqués dans un langage très simple avec des termes connus. S'il est nécessaire d'utiliser des mots probablement inconnus des élèves pour expliquer le mot cible alors le mot choisi est trop difficile. En résumé, si le mot peut être expliqué avec des mots connus et s'il est susceptible d'être utilisé dans les conversations courantes ou dans les livres de littérature de jeunesse, c'est un mot approprié.

⁴⁹ 6,9 % des élèves de nos groupes expérimentaux ne parlent pas le français à la maison et 30,5 % parlent au moins deux langues.

⁵⁰ Cette liste, rassemblant près de 1500 mots les plus fréquents de la langue française, a été constituée par le lexicologue Étienne Brunet.

Diverses questions ont ainsi guidé notre réflexion : dans quelle mesure le mot est-il important ? S'agit-il d'un mot que les élèves sont amenés à rencontrer dans d'autres textes ? Sera-t-il utile aux élèves pour décrire leurs propres expériences ? Peut-il se rapporter à d'autres mots, à des idées que les élèves connaissent déjà ?

Pour chaque album, nous avons veillé à proposer des verbes, des noms, des adjectifs et des expressions comme le préconisent les programmes de l'école maternelle : « Les verbes [...] sont essentiels, en lien avec les consignes et les activités de la classe mais aussi les histoires et les lectures documentaires. Ils constituent le noyau de toute phrase. Autour d'eux, s'articulent des sujets et des compléments. [...] Or, les noms sont fortement prégnants dans les choix pédagogiques et occupent une place centrale dans les imagiers : ils ne doivent pourtant pas faire perdre de vue que les autres classes sont nécessaires pour améliorer sa maîtrise de la langue orale » (MEN, 2015, p. 5).

1.6.2.3.2.2. Le déroulement de l'épreuve

Le pré-test est toujours réalisé avant la première narration de l'histoire. En effet, pour certains élèves au moins, le contexte de l'histoire fournit un bon exemple de l'utilisation des mots et pourrait leur permettre d'apprendre le sens d'un mot évalué. Aussi avons-nous proposé le pré-test avec toute écoute de l'histoire en classe. Les post-tests ont été administrés après la dernière séance d'enseignement.

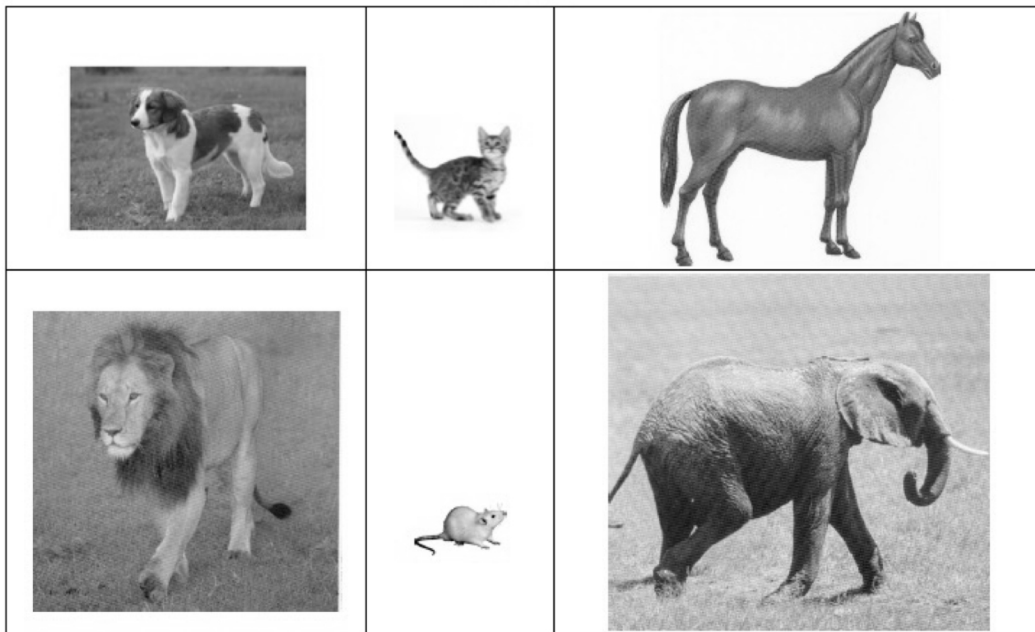
L'évaluation du lexique en réception et en production (Hiebert & Kamil, 2005) permet de construire un score unique (Variable « el ») basé sur 25 items. Les premiers servent à familiariser l'élève avec la tâche demandée, les réponses ne sont donc pas codées. Quatre modalités d'évaluation sont proposées aux élèves que nous illustrons par des extraits de la grille de codage du premier album placés ci-dessous.

1° une tâche de dénomination d'images (vocabulaire en production)

Cette tâche très accessible pour de jeunes élèves permet d'évaluer la connaissance de mots en production.

Une planche d'images (en noir et blanc) est montrée aux élèves qui disposent d'un temps d'observation. Pour ne pas créer de biais entre les élèves des trois groupes et montrer des images déjà présentes dans l'outil, nous avons veillé à ne pas évaluer les élèves avec des images déjà présentes dans les annexes numériques. Partant, aucune image ne correspond à celles retenues par les auteurs de *Narramus* pour enseigner le lexique.

Planche 1



La première ligne sert d'exemple, l'enquêteur ne code que les réponses de la seconde ligne d'images.

Consignes		Code
Montrer la planche 1 (6 animaux)		
Pointer tour à tour les trois animaux de la première ligne et demander à l'élève de les nommer : chien, chat et cheval . S'il ne connaît pas le nom d'un animal, le lui donner et passer au suivant.		
1	Pointer le lion et demander : « et cet animal, qu'est-ce que c'est, c'est un... ? »	1 — 0
2	Pointer la souris et demander : « et cet animal, qu'est-ce que c'est ? C'est une... ? » (la réponse « rat » est acceptée)	1 — 0
3	Pointer l'éléphant et demander : « et cet animal, qu'est-ce que c'est, c'est un... » ?	1 — 0

L'élève obtient 1 point s'il produit le mot correct, 0 dans le cas contraire ou s'il ne dit rien.

2° une tâche de désignation d'image (vocabulaire en réception)

Cette tâche, souvent proposée dans les tests standardisés comme l'échelle de vocabulaire en image Peabody (Dunn, Dunn, & Thériault-Whalen, 1993) ou le TVAP (Deltour & Hupkens, 1980) permet d'évaluer la connaissance réceptive des mots.

Remonter la planche 1 (6 animaux)		
7	Montre-moi l'animal qui peut aboyer .	1 — 0
8	Montre-moi l'animal qui peut rugir .	1 — 0
9	Montre-moi l'animal qui peut ronronner .	1 — 0
10	Montre-moi l'animal qui peut barrir .	1 — 0

L'élève doit choisir une image parmi les six présentées. Il obtient 1 point s'il montre la bonne image correcte, 0 dans le cas contraire ou s'il ne montre rien.

3° une tâche d'explication verbale (vocabulaire en production)

Cette tâche permet de vérifier si l'élève connaît le sens du mot.

12	C'est quoi un matelas ? Si l'élève ne répond pas, demander : à quoi ça sert un matelas ? Où est-ce qu'on peut trouver un matelas	1 — 0
13	Qu'est-ce que ça veut dire refuser ? Si l'élève ne répond pas, demander : quand est-ce que tu peux refuser ? Qu'est-ce que tu as envie de refuser toi des fois ?	1 — 0

L'évaluateur pose la première question : « c'est quoi un matelas ? ». Si l'élève propose une réponse correcte (c'est sur un lit, c'est pour dormir, c'est pour se coucher, c'est pour la sieste...), l'évaluateur code 1. S'il ne répond pas, une relance est proposée : « à quoi ça sert un matelas ? Où est-ce qu'on peut trouver un matelas ? ». L'élève obtient 1 point quelle que soit la manière dont il manifeste sa connaissance du mot ou de l'expression que la réponse soit spontanée ou après une relance (0 dans le cas contraire). Si l'élève préfère mimer, la proposition est acceptée.

4° une tâche de mime (vocabulaire en production)

Ce type de tâche s'avère pertinente pour de jeunes élèves, particulièrement pour l'évaluation des verbes caractérisant des actions ou d'adjectifs qualifiant des propriétés physiques.

22	Fais semblant de t'étirer .	1 — 0
23	Fais semblant de mordiller .	1 — 0

Après lecture de la consigne par l'évaluateur, l'élève est invité à mimer. S'il préfère fournir une explication verbale, la proposition est acceptée. Un point est accordé quelle que soit

la manière dont l'élève manifeste sa connaissance du mot ou de l'expression. Dans le cas contraire, l'enquêteur code 0.

Nous nous sommes aussi interrogée sur le maintien à long terme des mots et des expressions enseignées. Les scores des élèves seraient-ils les mêmes quelques mois plus tard ? Qu'en est-il du maintien du lexique ?

1.6.2.3.2.3. Le maintien du lexique

Pour vérifier que l'intervention produit un apprentissage durable, une épreuve différée (cf. annexe 4²) reprenant des mots et des expressions des albums 1 et 2, a été proposée aux élèves en juin 2017, soit entre 3 et 6 mois après l'enseignement.

Deux autres épreuves de lexique différé ont été également conduites en juin 2018 (sur les mots et expressions des albums 4 et 5, cf.annexe 4) et en juin 2019 (sur des mots des albums 1 à 8), comme l'indique le Tableau 14.

Tableau 14. Organisation temporelle des épreuves de lexique différé

Période	Juin 2017	Juin 2018	Juin 2019
Tests	Post-test différé de lexique évaluant des mots et expressions des albums 1 et 2 (élèves de PS, MS et GS)	Post-test différé de lexique évaluant des mots et expressions des albums 4 et 5 (élèves de MS et GS)	Post-test différé de lexique évaluant des mots et expressions des albums 1 à 8 (élèves de GS)

Nous avons créé une épreuve de lexique différé visant à évaluer la stabilité du lexique. Notre questionnaire portait sur les effets de l'enseignement lexical proposé dans *Narramus* et était formulée ainsi : un enseignement explicite et systématique du lexique permet-il aux élèves d'améliorer leurs performances de manière durable ? Si tel est le cas, les résultats des élèves du groupe 1 et 2 devraient être supérieurs à ceux des élèves du groupe 3 pour ladite épreuve.

1.6.2.3.3. Questionnaire

En GS, au second trimestre et en PS et MS, au troisième trimestre, nous avons introduit une épreuve complémentaire d'évaluation de la compréhension : un questionnaire, tâche plus habituelle que le rappel de récit dans l'univers scolaire. Nous avons ainsi pu renforcer l'administration de la preuve des effets de *Narramus* en compréhension. Ceci nous a permis également d'argumenter avec certains professeurs qui croyaient que les élèves pouvaient raconter presque par cœur sans comprendre.

La plupart des enfants d'âge préscolaire sont capables de répondre à des questions littérales (Goldtsein, 2016) mais les questions portant sur l'implicite des textes posent davantage de difficultés (nous l'avons déjà évoqué, pour des élèves plus âgés, avec la présentation des résultats aux évaluations internationales dans le chapitre 1). En outre, il est attesté que le langage inférentiel contribue à la compréhension en lecture (Cain, Oakhill, & Bryant, 2004). Nous avons donc introduit un questionnaire proposant les deux types de questions (cf. annexe 5). Les questions inférentielles comprenaient des items sur les émotions des personnages (Pourquoi le chien remue-t-il la queue quand il monte sur le balai ?⁵¹), leurs intentions (Pourquoi le renard invite-il la souris à manger dans son terrier ?⁵²) et leurs actions (Au début de l'histoire, pourquoi la chevette poursuit-elle Tipoilou ?⁵³).

Des questions littérales et inférentielles ont ainsi été posées aux élèves après leur rappel de récit (variable « Q »). Cette épreuve nous a permis d'analyser les liens entre la qualité du rappel et les réponses aux questions.

2. Les pratiques des enseignants

Pour chaque album étudié, nous avons conçu une trame de carnet de bord et l'avons communiquée aux enseignants des trois groupes. Les carnets de bords étaient transmis au chercheur chaque trimestre.

2.1. Les carnets de bord

Les enseignants des trois groupes expérimentaux ont renseigné un carnet de bord nous permettant d'obtenir des informations très précieuses sur la manière dont les enseignants du G1 et du G2 ont mis en œuvre l'outil et sur les pratiques d'enseignement de la compréhension adoptées par le G3.

2.1.1. Les carnets de bord des enseignants des groupes 1 et 2

Pour chaque module réalisé, les enseignants des groupes 1 et 2 ont déclaré le degré de fidélité avec lequel ils ont suivi les prescriptions des auteurs, la durée de l'enseignement et le mode de

⁵¹ Extrait du questionnaire MS Album *La sorcière dans les airs*

⁵² Extrait du questionnaire GS Album *Gruffalo*

⁵³ Extrait du questionnaire PS *Un poussin de mauvais poil*

groupement des élèves retenu (collectif, petit groupe, travail individuel...), comme l'indique le Tableau 15.

Tableau 15. Carnet de bord des enseignants G1 et G2

Module	Degré de fidélité aux activités proposées*	Durée en minutes (moyenne par élève)	Modalités de travail			
			Mettre une croix dans les cases correspondantes Plusieurs choix sont possibles pour une même séance			
			collectif	petits groupes	par 2	autres (préciser)
1.	... %					
2.	... %					
...						

2.1.2. Les carnets de bord des enseignants du groupe 3

Les enseignants du groupe 3 devaient, eux, renseigner précisément les différentes activités menées et leur durée.

2.1.2.1. Les caractéristiques des carnets de bord des enseignants du groupe 3

Un carnet de bord « type » était proposé comme le montre le Tableau 16 mais la forme de ce document était donnée à titre indicatif : les enseignants pouvaient utiliser leur propre présentation pour le carnet de bord, la seule exigence étant de respecter les informations demandées. Il leur était également possible de communiquer des productions d'élèves, des plans de séquence, des fiches de préparation... ; en bref, tout support constituant une trace du travail de compréhension mené sur l'album avec leurs élèves.

Ces indications sont très précieuses pour mieux comprendre ce qui influence les progrès des élèves : les pratiques ordinaires nous intéressent dans leur inventivité, bien au-delà de l'outil développé par les chercheurs.

Tableau 16. Carnet de bord des enseignants G3

Séance n°	Titres des activités proposées au cours de la séance	Durée (par élève)	Modalités de travail (collectif, petits groupes, travail individuel...)	Observations, remarques, commentaires éventuels
1.	- ...	- ...	- ...	
...				

2.1.2.2. Synthèse des contenus des carnets de bord du G3

Comme nous l'avons écrit dans le chapitre 2, il existe peu d'observations et d'études de grande ampleur décrivant les pratiques ordinaires d'enseignement de la compréhension à l'école maternelle. Les enseignants du groupe 3 nous ont donc fourni donc des données intéressantes sur ce sujet.

Les maitres devaient, a minima, indiquer le temps passé à enseigner la compréhension avec l'album, un titre pour chacune des activités proposées ou un bref descriptif du contenu, et le mode de groupement des élèves retenu (collectif, petit groupe, travail individuel...). Ils ont, de fait, été beaucoup plus explicites dans leurs renseignements, nous permettant de visualiser et comprendre plus finement à travers leurs descriptifs le contenu des séances proposées, les compétences travaillées et les activités mises en œuvre pour les atteindre.

2.1.2.2.1. Description des carnets de bord renseignés lors du travail sur l'album 1 par les enseignants du groupe 3

Pour le travail sur le premier album *La sieste de Moussa*, 20 carnets de bord détaillés nous ont été communiqués par les enseignants. Des documents supplémentaires étaient parfois fournis telles que les fiches individuelles données aux élèves ou les traces d'un texte rédigé en dictée à l'adulte.

Classiquement, les modalités de mise en œuvre alternent des temps collectifs, des moments de travail par petits groupes et des tâches individuelles. Le travail écrit est systématiquement proposé individuellement (souvent sous forme de fiches) alors que le travail de découverte du texte est assuré en collectif. Les propositions en petits groupes correspondent notamment aux activités de mise en scène (théâtre), et d'ateliers de découverte (travail sur la couverture, sur les noms des personnages...).

En résumé, les propositions sur la compréhension sont les suivantes :

- de multiples lectures,
- des questionnements oraux sur le texte (les élèves répondent aux questions du maître),
- une découverte des noms, des cris et des habitats des animaux, voire plus exceptionnellement de leurs actions,
- une explication du vocabulaire en contexte, lors de la découverte du texte,
- une remise en ordre chronologique de l'introduction des personnages de l'histoire,
- des mises en scène (théâtre ou marottes),
- un travail sur la couverture (titre, illustrateur, auteur, image),
- des fiches autour de l'écrit : écrire les noms d'animaux, associer les noms des animaux et l'image correcte,

Plus rarement, les enseignants ont déclaré raconter l'histoire (une enseignante), utiliser un vidéoprojecteur (une enseignante) ou faire réviser le lexique appris (le nom et le cri des animaux) avec des jeux de lotos ou de memory.

2.1.2.2.2. Description des carnets de bord renseignés par le groupe 3 lors de l'enseignement sur l'album 2

Trente-six enseignants ont communiqué un carnet de bord détaillant l'enseignement mené en compréhension sur l'album 2.

En PS, les activités conduites à partir de l'ouvrage *Le Machin* sont proches de celles proposées pour *La sieste de Moussa* : des lectures répétées, des mises en scène, un travail sur la chronologie. Plusieurs enseignants ont, en outre, instauré un travail spécifique ponctuel sur le vocabulaire (les vêtements, les injures, les émotions des personnages) avec des jeux d'appariement et de loto. Plus rarement (1 enseignant à chaque fois), les maîtres ont évoqué raconter l'histoire avec des marottes, travailler la causalité ou dédier un temps de langage autour de la compréhension de l'album (personnages, lieux, états mentaux « *il rit, il est fâché, il boude, il est dégoûté* »).

En MS, les carnets de bord portent sur la mise en œuvre de l'album *Le gros chagrin d'Edgar*. Les enseignants ont majoritairement proposé des activités autour du lexique (émotions, animaux de la ferme), une remise en ordre chronologique des illustrations (fréquemment cité), des échanges oraux avec des questions-réponses sur le texte, des lectures à plusieurs voix.

Dans de nombreuses classes, des temps étaient spécifiquement dédiés au travail de l'écrit : remise en ordre du titre de l'histoire, discrimination visuelle (les mots de l'histoire),

écriture de mots (les animaux), dictée à l'adulte pour raconter l'histoire, écriture individuelle tâtonnée de l'histoire ou invention d'une autre histoire pour réinvestir le vocabulaire sur les émotions.

Comme en PS, un travail spécifique sur la couverture est régulièrement effectué (titre, image, puzzle) et l'album a servi de support pour un travail plus large en découverte du monde autour de la ferme (male, femelle, bébé, animaux de la basse-cour).

Enfin, les activités de création plastique sont fréquemment mises en œuvre : peinture, animaux en pâte à modeler, dessin de chardon, coloriage...

En GS, les carnets de bord détaillent le travail conduit par les enseignants sur l'album *Gruffalo*. Comme dans les autres niveaux, la majorité de l'enseignement comporte des activités autour de : la découverte du lexique généralement réalisé lors de la lecture de l'histoire (lieux, personnages), la remise en ordre chronologique (des personnages ou des images séquentielles à décrire et ordonner), la mise en scène (théâtralisation, utilisation de masques) et des temps de questions-réponses sur la compréhension de l'histoire. Tous les enseignants ont dédié un moment à la description physique du Gruffalo et des autres personnages.

Comme en PS ou MS, certains maîtres ont proposé un travail en arts plastiques ou en découverte du monde (déplacements, habitats, régimes alimentaires des animaux).

D'autres, moins nombreux, précisent avoir travaillé plus spécifiquement les pensées des personnages ou les passages implicites du texte (qui mange qui ? Qui a peur de qui ? Qui est malin ? Qui est sot ?), proposé des rituels de mémorisation (la comptine du *Gruffalo*), des activités écrites (écritures tâtonnée...) et des temps dédiés à la narration par un élève.

Enfin, un enseignant déclare avoir filmé les mises en scène de ses élèves et avoir effectué une analyse collective. Un autre a fait visionner le film d'animation à la fin de l'enseignement. L'écoute de la version audio de l'histoire a également été citée une seule fois.

2.1.2.2.3. Description des carnets de bord renseignés par le groupe 3 lors de l'enseignement sur l'album 3

Trente-six carnets de bord ont été renseignés de manière détaillée pour le troisième album par les maîtres du groupe 3.

En PS, les activités conduites sur *Un poussin de mauvais poil* se rapprochent de celles décrites pour les deux premiers albums (vocabulaire des animaux de l'histoire, des lieux de l'histoire, mime des animaux, loto des émotions, mise en scène, marottes pour raconter) mais s'en écartent aussi avec un travail important déclaré en arts visuels (personnages à peindre ou colorier, création d'un décor pour faire déplacer les personnages). La plupart des enseignants

ont, en outre, travaillé sur la ferme (tri des animaux, lectures en réseau, recherches documentaires sur chaque animal : les femelles, les mâles, les bébés, les cris, l'alimentation).

Sur ces troisièmes carnets de bord, certains enseignants rédigent des commentaires qui dépassent le simple descriptif des séances et déplorent le manque de résultats visibles chez leurs élèves avec une « *participation pauvre* » ou « *encore beaucoup d'ignorance sur le vocabulaire de l'histoire à la fin de l'enseignement* ».

En MS, les contenus de l'enseignement proposé pour *La sorcière dans les airs* comprennent principalement des remises en ordre chronologique, des questions orales de compréhension, des jeux pour mémoriser les noms des animaux, leur cri et leur déplacement (devinettes, memory), des activités d'arts visuels (dessin, collage sur le personnage de la sorcière ; fabrication d'un monstre en terre, en peinture...).

Plus ponctuellement, les enseignants ont mis en œuvre des activités en phonologie, (rimes avec la formule magique, rimes avec sorcière), en production écrite (inventer des formules magiques, des potions), en musique (comptines et chants sur la sorcière) et en écriture inventée.

Plus rarement, des fiches individuelles portaient sur les personnages et leurs actions (activités de reliage, puzzles des personnages ou de la couverture), la numération (compter le nombre de sorcières, donner un balai à chaque sorcière) ou sur l'écrit (associer le mot écrit et l'image des personnages de l'album).

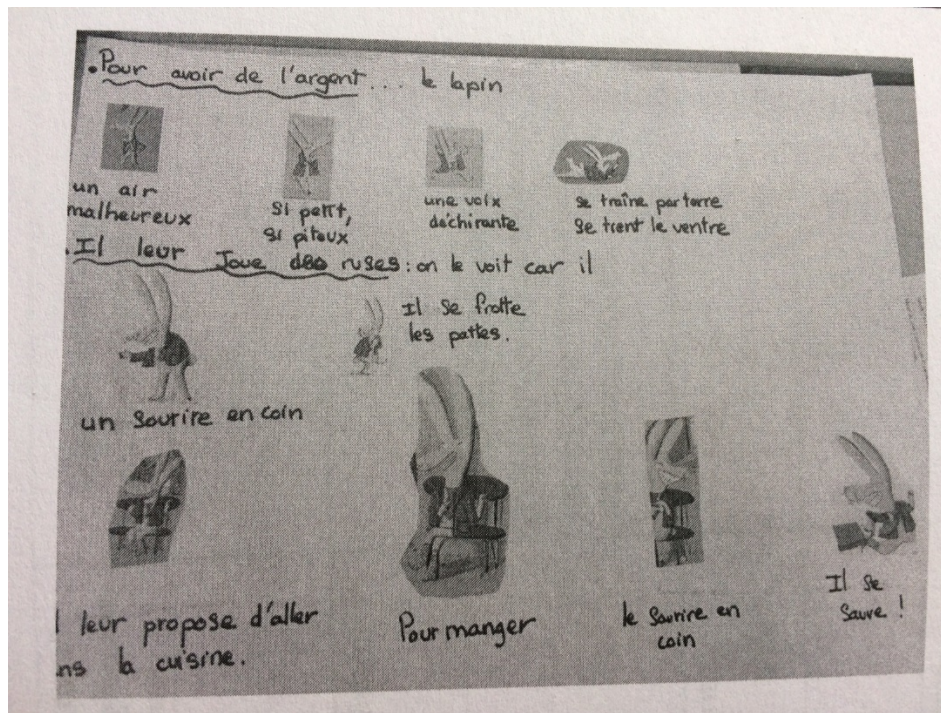
Une seule enseignante a proposé un visionnage du film en entier à la fin et une autre a fait écouter plusieurs fois la version lue de l'histoire.

Les enseignants soulignent l'intérêt des élèves pour cette histoire mais restent plus mitigés sur la progression en compréhension, comme l'illustre cet extrait : « *Les enfants sont entrés très facilement dans l'histoire. Quelques enfants ont progressé mais pas suffisamment pour raconter la totalité de l'histoire. Il convient alors de s'interroger sur les dispositifs mis en place qui ne sont pas suffisants [...]* ».

En GS, le dernier album étudié s'intitule *Les deniers de Compère lapin*. Les enseignants ont principalement proposé les activités suivantes : une explicitation des mots et des expressions en contexte, de multiples relectures, des temps de questionnement oral sur le texte, un rappel avec des marottes, une remise en ordre chronologique des images de l'histoire ou des personnages de l'histoire et une mémorisation des structures répétitives de l'histoire

D'autres ont effectué un travail de repérage sur la couverture (nombre de mots, auteur, illustrateur), de la production écrite (inventer la suite de l'histoire), des écrits individuels sur les mots du texte et une recherche documentaire sur les animaux de l'histoire.

Plus rarement, quelques maîtres ont proposé une analyse des attitudes de compère lapin et des descriptions sur les habitats et les états d'esprit des autres personnages, comme le montre l'image ci-dessous.



L'écrit est aussi travaillé à cette occasion : reconstitution du titre dans les trois écritures, calligraphie ou reconnaissance des noms des animaux.

Deux enseignants ont mis en œuvre un temps d'échanges oraux sur la moralité de l'histoire et l'attitude du lapin, une autre a fait écouter une version enregistrée (par ses soins) du récit avec des bruitages et de la musique pour favoriser la mémorisation des personnages, de leurs intentions, de l'histoire...

2.1.2.3. Conclusion

En résumé, d'un album à l'autre, les propositions sont assez hétérogènes sur le nombre de séances conduites (5 à 20 séances) mais plutôt consensuelles dans le contenu. Elles restent en effet très proches d'une classe à l'autre pour la plupart des activités.

On retrouve majoritairement des lectures nombreuses, des remises en ordre chronologique des images ou des personnages de l'histoire, des questionnements oraux, des activités en arts plastiques ou en découverte du monde, des fiches autour de l'écrit et un enseignement du vocabulaire.

Le lexique est principalement découvert en contexte ou expliqué directement par l'adulte au fil de l'histoire et n'inclut pas de révision ni de réinvestissement. Les traces écrites

servant de mémoire des mots travaillés n'apparaissent dans aucun carnet. Sur l'ensemble des 92 carnets analysés, une seule enseignante déclare s'appuyer sur des images imprimées (la lande, d'une femme rousse...) pour expliquer le vocabulaire de l'album *La sorcière dans les airs*. On voit ici le souci des enseignants de proposer un apprentissage lexical mais celui-ci est très différent de celui proposé dans *Narramus* notamment sur le plan de la mémorisation du vocabulaire étudié.

Le constat est le même sur les compétences narratives en réception et en production : très peu de maîtres font raconter les histoires aux élèves ou dédient un temps aux états mentaux et intentions des personnages. Un rappel de l'histoire est souvent proposé en début de séance pour se remémorer les passages antérieurs mais rarement en fin de séance pour lier le passage découvert au reste de l'histoire et jamais porté par un seul élève.

2.2. Les questionnaires

Le modèle de développement professionnel envisagé par les auteurs de *Narramus* diffère du modèle conventionnel qui repose sur un simple apport connaissances aux enseignants, en postulant que celles-ci transformeront leurs pratiques et, par conséquent, l'apprentissage des élèves (Kennedy, 2016). Pour les auteurs de *Narramus*, c'est la modification des pratiques professionnelles provoquée par l'outil qui apporte des connaissances aux enseignants et participe à leur développement professionnel, permettant d'améliorer l'enseignement et les performances des élèves. C'est pourquoi nous avons voulu savoir comment les enseignants s'approprièrent leur outil et si celui-ci générerait des modifications durables de pratiques.

Pour répondre à nos interrogations sur les liens entre l'outil et le développement professionnel, nous avons élaboré un premier questionnaire (cf. annexe 6) à destination des enseignants des groupes 1 et 2. Celui-ci nous a permis d'obtenir des informations qualitatives à l'issue de la première année de recherche (juin 2017). Il comportait 5 items, volontairement très larges, renvoyant à leurs perceptions touchant de possibles transformations professionnelles liées à l'usage de l'outil didactique. Nous leur avons demandé de renseigner les thématiques suivantes :

- 1° les points de réussite,
- 2° les gestes professionnels modifiés,
- 3° ce qu'ils retiennent de l'année,
- 4° les inventions,
- 5° l'activité qu'ils détestent.

Les enseignants étaient libres de compléter toutes les rubriques, pouvaient exemplifier leurs propos et insérer des documents s'ils le souhaitaient. Si leurs réponses nous informent sur les changements des pratiques professionnelles provoqués par la mise en œuvre de l'outil, elles ne suffisent pas pour nous permettre de distinguer les évolutions liées à l'outil lui-même de celles qui dépendent de l'accompagnement.

Nous avons donc proposé un deuxième questionnaire (cf. annexe 6) en juin 2018 à tous les enseignants ayant utilisé l'outil. Celui-ci compte 17 items répartis en 2 volets : un premier sur l'outil et un deuxième sur l'accompagnement.

Les résultats recueillis, que nous détaillerons dans le chapitre 7, offrent des données sur l'usage de l'outil, les modifications des pratiques professionnelles, la manière dont les enseignants ont perçu l'accompagnement... Ils nous permettent donc de répondre aux questions de recherche touchant :

1° le développement professionnel des enseignants : dans quelle mesure l'utilisation de *Narramus* transforme-t-elle les pratiques professionnelles ?

2° l'accompagnement : quelle est, selon les usagers, la plus-value à être accompagnés dans la mise en œuvre d'un outil ? Quels sont les effets de cet accompagnement sur le développement professionnel des enseignants ?

Pour répondre à nos hypothèses sur les effets « élèves » et « accompagnement » (hypothèses 1 et 2), nous avons eu besoin de connaître la nature exacte des dispositifs d'accompagnement : à quelle fréquence les rencontres ont-elles eu lieu ? Quels ont été les choix effectués par les formateurs ? Quelles sont les modalités retenues ? Quels sont les contenus proposés ? La partie suivante apporte des réponses à ces questions.

3. Le recueil des données sur l'accompagnement

Les accompagnateurs étaient libres de choisir leurs modalités d'accompagnement. Nous voulions qu'ils puissent rester dans leur domaine de compétences et choisir un mode de formation qui leur convienne. En contrepartie, nous souhaitions savoir ce qu'ils faisaient.

Les journaux de bord des accompagnateurs étaient déposés sur la plateforme au moins trois fois par an, parfois davantage, certains formateurs préférant communiquer les rencontres et leur contenu au fur et à mesure de leur réalisation. Lorsque le contenu des journaux des accompagnateurs n'étaient pas clairs, nous prenions directement contact avec eux pour leur demander des précisions complémentaires.

Des échanges par mails, des entretiens informels (lors des séminaires de formation par exemple) ainsi que des vidéos transmises par les formateurs ont permis de compléter les données sur la nature des accompagnements. Nous avons aussi régulièrement assisté à des temps d'accompagnement assurés par un conseiller pédagogique dans l'Allier⁵⁴.

3.1. Description des accompagnements

Les accompagnements ont été réalisés par les conseillers pédagogiques, les formateurs REP+ et des inspecteurs dans le cadre des horaires dévolus à la formation continue en Éducation prioritaire. Les modalités organisationnelles ont été dépendantes de la situation de l'école (REP+, REP ou relevant de la politique de la ville) et des calendriers de formations départementales déjà établis, ce qui a généré une grande disparité dans les possibilités de dégager du temps scolaire pour l'accompagnement. Certaines circonscriptions ont centré toutes les heures de formation autour de l'expérimentation alors que d'autres les ont conciliées avec les formations déjà engagées. En outre, les temps d'accompagnement se déroulaient soit sur le temps scolaire dans les écoles situées en REP+ qui disposaient de temps dégagés spécifiques dévolus à la formation, soit après la sortie des classes ou sur la pause méridienne pour les autres écoles.

3.1.1. Descriptif général de la nature des accompagnements

Les formateurs étaient libres de leurs modalités d'accompagnement en fonction de leurs compétences et de leurs préférences mais tous ont bénéficié, la première année, de quatre journées de formation à l'Institut français de l'éducation (ENS Lyon) animées par les auteurs de *Narramus*.

3.1.1.1. Présentation synthétique des accompagnements

Comme nous l'avons écrit plus haut, les formateurs devaient 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. L'analyse des journaux nous a permis de synthétiser les accompagnements réalisés.

⁵⁴ Nous remercions vivement les enseignants de l'école Marie Noël, le conseiller pédagogique et l'Inspectrice de l'Éducation nationale du secteur de Montluçon qui ont accepté notre présence tout au long de l'année 2016-2017.

3.1.1.1.1. Les objectifs ciblés par les accompagnateurs

Les formateurs attribuent différents objectifs aux accompagnements que nous avons regroupés en quatre catégories. Celles-ci sont ici présentées selon leur fréquence (de la plus citée à la moins citée) :

1° théoriser et faire évoluer les pratiques

- revenir sur les principes pédagogiques de l'outil *Narramus*,
- outiller les équipes avec les données de la recherche dans le but de faire réussir tous les élèves,
- donner du sens aux pratiques enseignantes,
- former les enseignants à la didactique de la compréhension,
- fournir des justifications théoriques,
- donner les fondements théoriques de la recherche,
- expliciter les choix didactiques des concepteurs,
- former les enseignants au racontage,
- s'appropriier le scénario,
- s'appropriier les activités phares,
- donner des apports didactiques,
- expliciter certains passages des scénarios,
- apporter des éléments théoriques sur les états mentaux, la théorie de l'esprit,
- accompagner les changements de pratiques,
- donner les moyens aux enseignants de réfléchir et de mettre en lien leur pratique avec la recherche ;

2° guider la pratique par le partage d'expérience entre les enseignants et/ou avec un formateur

- organiser des échanges de pratiques entre les enseignants,
- ajuster les gestes professionnels des enseignants,
- identifier la modification des pratiques pédagogiques (bilan réflexif) ;

3° aider à la mise en œuvre

- résoudre les problèmes matériels (vidéoprojecteurs, organisation dans les locaux, photocopies...),
- observer des séances en classe,
- organiser des échanges par niveau sur la mise en œuvre de l'outil,
- accompagner les enseignants en classe ;

4° inciter à persévérer

- aider les enseignants dans les classes pour conserver leur engagement : entendre les difficultés, essayer de comprendre les problèmes de métier souvent liés aux conceptions des enseignants sur les enseignements à mener et sur la manière de les mener,
- rassurer,
- soutenir les équipes par la disponibilité et la présence,
- entretenir la dynamique.

Ces objectifs ont orienté le contenu et les modalités des accompagnements mis en œuvre par les formateurs.

3.1.1.1.2. Les modalités et les contenus des accompagnements

Malgré des objectifs déclarés très proches et ce, quel que soit le lieu des accompagnements, force est de reconnaître que ces derniers ont revêtu différentes formes lors de leur mise en œuvre sur le terrain.

L'accompagnement proposé aux enseignants du groupe 1 a duré 15,7 heures en moyenne (écart-type 7,4), avec un minimum de 4 heures et un maximum de 36 heures.

Lorsqu'ils étaient proposés, les apports théoriques s'organisaient principalement autour de la présentation de la recherche, des principes didactiques de l'outil, des temps de formation spécifiquement dédiés à l'enseignement du lexique, des états mentaux et du racontage de l'enseignant.

Dans la majorité des secteurs, tous les enseignants recevaient le même contenu de formation. Dans d'autres, les formations étaient plus individualisées avec soit des moments personnalisés en fonction des besoins et des questions des enseignants, soit des propositions « à la carte ». Par exemple, sur Lyon 8 et Lyon Meyzieu, pour le premier temps de regroupement, des enseignants choisissaient entre cinq thèmes :

- thème 1 : l'architecture du scénario
- thème 2 : l'enseignement explicite
- thème 3 : le rappel de récit par l'élève
- thème 4 : comprendre les états mentaux des personnages
- thème 5 : apprendre et pouvoir réutiliser le lexique.

Pour l'ensemble des accompagnements organisés autour des principes didactiques de l'outil, le matériel support utilisé contenait généralement :

- les powerpoints présentés lors des séminaires et transmis aux formateurs par les chercheurs,

- le matériel nécessaire à l'enseignement en classe : les scénarios, les annexes numériques, l'album, les enregistrements audio des versions lues et racontées des albums étudiés,
- les carnets de bord des enseignants,
- des documents préparatoires demandés aux enseignants comprenant leurs questions, les réussites et leurs difficultés constatées dans la mise en œuvre, les changements constatés dans leur pratique, les nouvelles connaissances didactiques produites, les relations aux familles. Ces documents peuvent être envisagés comme des représentations du moment d'apprentissage et du changement vécu par les enseignants. Ces écrits temporaires assurent le rôle des « outils de passage » décrits par Charlier & Pereya (2003).

En fonction des secteurs et du nombre de classes engagées dans la recherche, l'organisation et la mise en œuvre des accompagnements variaient dans leur aspect temporel (durée et placement des temps dédiés), leur contenu, leur matériel mais également dans les personnes assurant l'accompagnement. Quand plusieurs formateurs travaillaient ensemble à l'élaboration des temps d'accompagnement, d'autres les organisaient seuls.

3.1.2. Du travail collectif à un travail plus solitaire

Dans certaines académies qui avaient un nombre conséquent d'écoles engagées dans la recherche comme Lyon ou Versailles, ce sont des équipes de formateurs qui sont intervenues avec un accompagnement pouvant concerner jusqu'à une quarantaine de classes. Dans d'autres secteurs, un conseiller pédagogique œuvrait seul avec un accompagnement généralement assuré pour trois classes⁵⁵.

3.1.2.1. Des accompagnements réalisés en « réseau » de formateurs

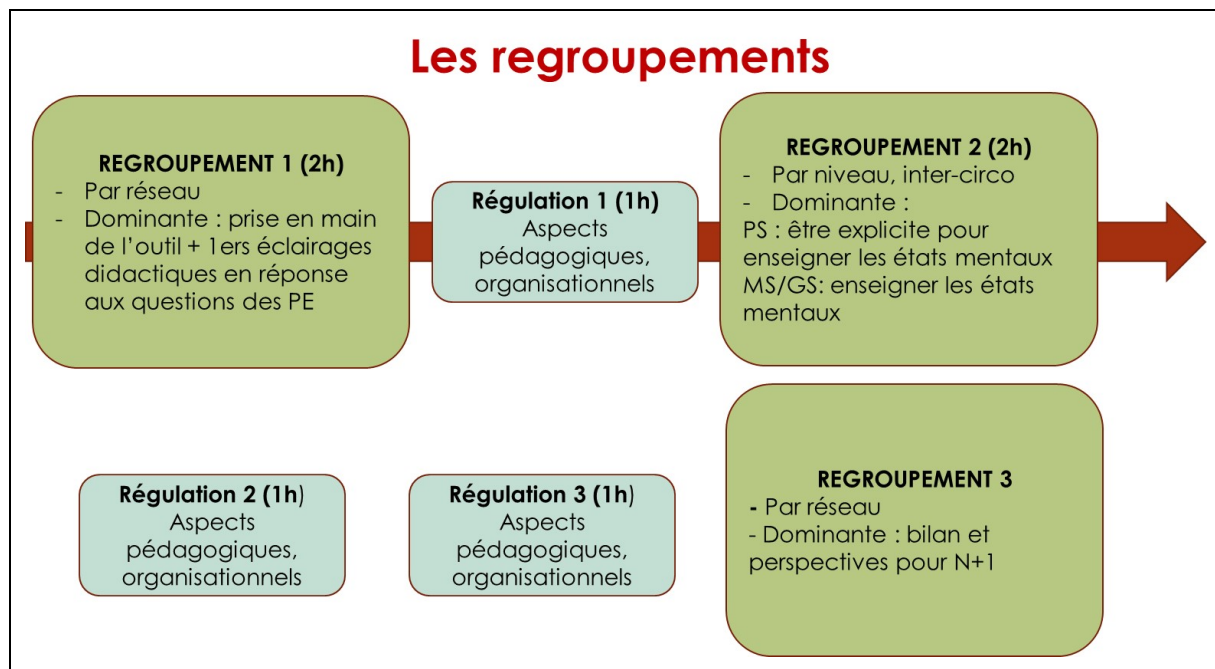
Des équipes de formateurs ont parfois choisi d'organiser des rencontres entre enseignants de plusieurs écoles et de différentes circonscriptions, assurant ainsi des échanges en trois temps : un premier regroupant tous les enseignants des différentes écoles (par réseau), un second réunissant les maîtres exerçant dans une même école (par école) et un troisième rassemblant les enseignants par niveau d'enseignement (par niveau, inter-circonscriptions). Ces modalités ont permis des échanges de pratiques répondant aux besoins spécifiques des enseignants souvent liés au niveau dans lequel ils exerçaient.

⁵⁵ Pour rappel, les équipes de circonscription s'étaient engagées à travailler avec 3 enseignants par école.

L'exemple de Lyon

Dans l'académie de Lyon, les conseillers pédagogiques ont fait le choix de regrouper deux circonscriptions et d'accompagner 40 classes (dont 9 engagées dans la recherche⁵⁶) dans la mise en œuvre de *Narramus*. Une alternance entre des temps de formation avec des apports théoriques et des temps de régulation au milieu des scénarios a été proposée. Neuf heures de regroupement ont ainsi été assurées pendant la première année, comme l'illustre la Figure 5.

Figure 5. Synthèse des regroupements proposés en 2016-2017 autour de la recherche *Narramus*



Document réalisé par F. Delay-Goyet et F. Jarre (2017), académie de Lyon

Le premier regroupement propose une formation de deux heures en présentiel pour présenter l'outil, son utilisation et les enjeux didactiques. Une heure de régulation et d'échanges de pratiques autour de l'utilisation de l'outil (ce que m'a fait l'outil, ce que j'ai fait à l'outil...) a eu lieu pendant la mise en œuvre du premier scénario.

Le second regroupement (2 heures de formation en présentiel) est organisé autour d'apports théoriques sur les compétences inférentielles et l'étude des états mentaux des personnages (« *l'étude des états mentaux des personnages avait été présentée comme posant des difficultés lors de la première régulation sur Moussa* [album support du premier scénario],

⁵⁶ On note ici la volonté affirmée par les équipes de circonscription : impulser des dynamiques locales en proposant à tous les enseignants de travailler ensemble sur *Narramus* même si toutes les classes n'ont pas des élèves évaluées pour la recherche.

et c'est à ce niveau que l'écart entre les pratiques enseignantes et celles induites par l'outil était le plus fort »). Une heure de régulation a ensuite été mise en œuvre, comme pour le scénario 1.

Les deux derniers temps d'accompagnement ont été regroupés : les formateurs ont débuté par une heure de régulation (réussites, difficultés, choix) puis ont terminé l'année avec deux heures de « *formation et des apports théoriques sur le rappel de récit, son évaluation et la reformulation en cascade*.⁵⁷ » Des enregistrements d'élèves du scénario 1 et du scénario 3 ont été présentés « *afin que chaque enseignant visualise concrètement les progrès des élèves et prenne des repères en fonction de l'âge des élèves*⁵⁸ ».

En sus de cet accompagnement collectif et commun à tous les enseignants, un accompagnement individualisé et personnalisé a été mis en œuvre « lorsqu'il était demandé par des enseignants. »

L'exemple de Versailles

Les formateurs du Val d'Oise ont accompagné 25 classes engagées dans la recherche. Ils ont opté pour préparer, organiser et mettre en œuvre des formations communes tout au long de l'année. Celles-ci ont eu lieu sur le temps de travail des enseignants qui étaient alors remplacés. Les 18 heures d'animation pédagogique obligatoires ont été banalisées pour des concertations avec les écoles sur le projet autour de *Narramus*.

Comme à Lyon, les formations alternent des temps de travail avec tous les enseignants des différentes écoles et des temps plus spécifiques organisés par niveau de classe.

Trente-six heures d'accompagnement (12 heures par trimestre) ont été réalisées avec :

- une journée initiale, en octobre, pour présenter les enjeux, les objectifs et la mise en œuvre de la recherche.
- Une demi-journée systématiquement consacrée à la découverte du scénario. Les enseignants prenaient connaissance au préalable du scénario qui était ensuite lu et détaillé en collectif.
- Une journée bilan organisée en fin de scénario.
- Des visites en classe effectuées par les conseillers pédagogiques avec des temps d'analyse de pratique puis de concertations avec l'équipe d'enseignants de l'école.
- Des formations théoriques spécifiques sur le lexique et « apprendre à raconter ».

⁵⁷ Extrait du journal de bord rédigé par les conseillers pédagogiques de Lyon.

⁵⁸ Idem.

Conclusion

Les accompagnements dans ces deux secteurs, bien que très différents dans la durée (du simple au triple), convergent dans les modalités organisationnelles retenues (regroupements inter et intra écoles) et le contenu (thématiques des formations théoriques autour des principes didactiques de l'outil, moments de bilan). Le fait de disposer de davantage de temps semble avoir permis une appropriation collective du scénario, des visites de classe et des échanges de pratiques plus fréquents.

3.1.2.2. Des accompagnements assurés par un seul et même formateur

Dans l'Allier ou les Pyrénées-Atlantiques où une seule école (appartenant au G1) était engagée dans l'expérimentation, c'est un conseiller pédagogique qui assurait seul l'accompagnement, portant ainsi une voix unique.

L'exemple de l'Allier

Trois classes accompagnées d'une même école participaient à la recherche sur *Narramus* avec 18 élèves évalués (6 en PS, 6 en MS et 6 en GS). L'accompagnement s'est déroulé avec tous les enseignants de l'école (4 classes) incluant donc une enseignante qui ne participait pas à la recherche.

Vingt-sept heures d'accompagnement sur le temps scolaire, soit neuf demies-journées ont été spécifiquement dédiées à la recherche. La répartition dans le temps s'est organisée en concertation avec l'équipe enseignante : parfois en amont de la mise en œuvre des scénarios, mais également après ces temps. Les rencontres avaient lieu dans l'école où exerçaient les enseignants. Les échanges portaient sur les contenus théoriques en réponse aux questionnements et préoccupations des enseignants.

Au début de chaque rencontre, le conseiller pédagogique revenait sur le document de travail rempli par les enseignants : les points de réussite, les transformations produites par cette nouvelle manière d'enseigner la compréhension, les difficultés repérées. Avec les enseignants, il identifiait leurs besoins et leurs questions puis effectuait un travail de catégorisation collective mettant en évidence les points forts et les points faibles dans la mise en œuvre de l'outil. Le formateur proposait ponctuellement des apports théoriques en réponse au diagnostic et aux besoins identifiés. Il pouvait également différer ces apports à la rencontre suivante en fonction des éléments disponibles dans « sa boîte à outils ».

Les enseignants échangeaient également sur les « trucs et astuces » ajoutés au scénario. Différents conseils d'ordre pratique (comment s'organiser pour assurer un rappel individuel des élèves avec une classe de 23 élèves ? Comment organiser un module pour proposer des temps d'enseignement d'environ 15-20 minutes en PS ?), technique (comment détourner les problèmes de compatibilité informatique avec les annexes numériques ? Quel format de maquette individuelle retenir pour faciliter la manipulation et le transport ?) ou plus didactique (l'intérêt de proposer les dictées de mime même si la gestion des élèves peut, en premier lieu, effrayer) étaient alors échangés. Ce moment pouvait aussi donner lieu à une présentation des différents maquettes, masques ou marottes réalisées pour raconter en classe ou à la maison.

Un temps spécifique était conservé pour présenter les éléments saillants ou novateurs du nouveau scénario.

Les résultats des élèves étaient aussi présentés individuellement à chaque enseignant.

À chaque fin de regroupement, le conseiller pédagogique proposait un déroulé succinct de la future rencontre et demandait aux enseignants s'ils avaient des attentes particulières pour réajuster la proposition.

L'exemple des Pyrénées Atlantiques

Les modalités de formation se déroulaient dans des conditions proches de celles présentées précédemment : trois classes dans une même école.

L'organisation temporelle était cependant différente puisqu'aucun temps sur l'horaire scolaire ne pouvait être dégagé pour assurer les rencontres. Les enseignants ont alors été regroupés sur le moment de la pause méridienne à raison d'une fois par période (soit 5 fois dans l'année).

Les échanges étaient organisés dans le but de « lever un maximum de difficultés⁵⁹ » rencontrées par les enseignants dans la mise en œuvre de l'outil. Les obstacles soulevés portaient sur la narration portée par les maîtres (manière de faire nouvelle et non maîtrisée), la narration réitérée assurée par les élèves (difficulté à mener cette activité de manière répétée), la modification de l'organisation spatiale (utilisation des lieux et du matériel) et temporelle habituelle (temps consacré à l'enseignement de la compréhension, changement de l'emploi du temps), le manque d'information sur les progrès « réels » des élèves et une impression d'ennui parfois présente (pour l'enseignant et les élèves).

⁵⁹ Extrait du journal de bord rédigé par la conseillère pédagogique.

Pour la conseillère pédagogique, l'accompagnement devait permettre de surmonter ces obstacles. Toutes les situations visaient à aider les enseignants :

- l'utilisation des powerpoints transmis par les chercheurs lors des séminaires de formation « donne du sens à des tâches qui sont nouvelles pour les enseignants, et qui posent problème (lexique : répétition, mémorisation, réinvestissement ; narration réalisée par l'élève)⁶⁰ »,
- un travail spécifique en équipe sur les emplois du temps et la participation du maître supplémentaire permet de répondre aux questionnements relatifs à l'organisation,
- la transcription d'un rappel de récit d'un élève et la comparaison avec le texte source de l'album lève la question du « par cœur »,
- la présentation d'histogrammes des résultats des élèves aux pré-tests et aux post-tests « donne à voir les progrès des élèves et l'impact de l'enseignement réalisé⁶¹ »,
- l'observation directe en classe et la réalisation de films permettent à chaque enseignant de visionner le réel et de porter une attention particulière aux productions des élèves. Cette modalité permet de « faire plus attention à ce qu'ils [les élèves] savent faire, et non à un idéal attendu⁶² ».

Conclusion

Ces deux exemples permettent d'identifier plus en détails le contenu des accompagnements proposés et montrent que, malgré une organisation temporelle différente, le souci des formateurs reste le même : aider à analyser les situations, envisager ensemble des solutions possibles et proposer des retours théoriques ajustés aux besoins et demandes des enseignants.

3.2. Synthèse

Avec des modalités logistiques propres à chaque secteur, les accompagnements ont toujours été conçus autour de trois axes :

1° une aide organisationnelle et logistique à la mise en œuvre dans les classes. Dans certaines écoles, les temps d'accompagnement ont permis aux enseignants d'exprimer les difficultés liées à la mise en œuvre du protocole ou à des questions matérielles. Pour des équipes plus « en souffrance », les échanges se sont focalisés sur ces difficultés mais cette étape a été identifiée comme nécessaire pour s'engager correctement dans la suite de la recherche.

⁶⁰ Idem.

⁶¹ Idem.

⁶² Idem.

2° Des apports théoriques sur les principes qui sous-tendent l'outil. Les contenus identifiés correspondaient généralement aux cibles de l'enseignement de *Narramus* : les principes généraux sur la compréhension chez les jeunes élèves, les compétences lexicales, les compétences narratives en production et les compétences inférentielles. Des temps spécifiquement centrés sur l'architecture des scénarios, l'enseignement explicite et des focus sur les états mentaux des personnages, la narration de l'enseignant ou le rappel de récit des élèves ont aussi été proposés dans ces apports didactiques.

3° Des échanges de pratiques avec un retour des enseignants sur la mise en œuvre des scénarios, les difficultés rencontrées... et la présentation des résultats des élèves par les formateurs-évaluateurs.

Tous les accompagnements ont prévu un bilan de la mise en œuvre d'un ou des scénarios pour répondre aux questions des enseignants, partager les choix pédagogiques effectués par chacun, mettre en commun le matériel utilisé, rechercher des solutions communes aux problèmes rencontrés. Ces temps d'échanges ont parfois été conduits individuellement après un temps observation de l'enseignant en classe. Dans ces temps de mutualisation et d'interactions avec les autres, l'accompagnement assure des échanges organisés autour des pratiques, des expériences positives et des difficultés rencontrées lors de l'utilisation de l'outil.

En résumé, malgré une diversité dans l'organisation temporelle (sur le temps solaire ou sur du temps scolaire dégagé, avant et après chaque scénario ou de manière plus régulière) et les regroupements (inter ou intra-écoles), les contenus des accompagnements sont assez proches. Reste à savoir quels sont les liens entre l'accompagnement et les pratiques des maitres ? L'accompagnement permet-il un développement professionnel plus efficace ? Aide-t-il les enseignants à conceptualiser davantage la mise en œuvre de l'outil ?

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 maitres 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. On verra dans le chapitre suivant qu'il n'en est rien.

Chapitre 4

Résultats de l'étude 1 :
année 1 (2016-2017)

Chapitre 4 Résultats de l'étude 1 année 1 (2016-2017)

Dans cette partie, nous présentons d'abord les résultats des élèves de MS et GS, puis ceux de PS pour les trois dimensions évaluées : narration, compréhension et lexique.

1. Rappel des hypothèses et des questions de recherche

Selon notre **hypothèse 1**, les élèves bénéficiant de *Narramus* progresseront plus que les autres.

Cinq questions de recherche sont associées à cette hypothèse.

1. L'intervention produit-elle des effets positifs sur les apprentissages des élèves ?
2. Sur quels aspects spécifiques les effets sont-ils les plus forts ?
3. Les effets sont-ils différents selon le niveau initial des élèves ?
4. Les effets sont-ils cumulatifs ? Quels sont les résultats après deux scénarios ? Après trois scénarios ?)
5. Le vocabulaire enseigné au premier et au deuxième trimestre est-il encore connu des élèves à la fin de l'année ?

L'**hypothèse 2** est ainsi formulée : conçu sur des bases scientifiques attestées et en partenariat avec des enseignants, *Narramus* produira des effets similaires avec ou sans accompagnement.

Deux questions de recherche découlent de cette hypothèse.

1. L'usage de l'outil didactique est-il différent selon que les enseignants bénéficient ou non d'un accompagnement régulier ?
2. Quels sont les effets de cet accompagnement sur les apprentissages des élèves ?

La nature des accompagnements a déjà été décrite dans la partie précédente. Il était, en effet, nécessaire de connaître les dispositifs mise en œuvre avant de présenter les résultats des élèves.

1.1. Variables de contrôle des pratiques pédagogiques

À partir des données communiquées dans les carnets de bord renseignés par les enseignants, nous avons établi des variables de contrôle (fidélité et temps) que nous utiliserons ensuite dans l'analyse des effets.

1.1.1. 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. Pour rappel, 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 du G1 et du 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 moyenne constitue notre variable *Fidélité* (cf. Tableau 17).

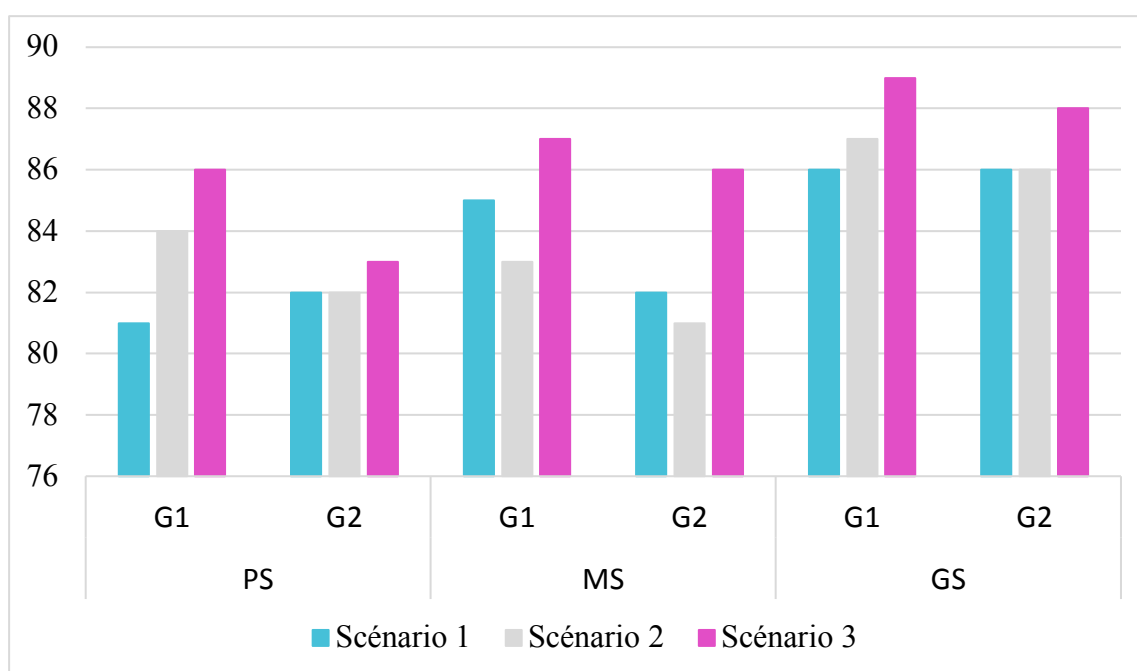
Tableau 17. Données descriptives pour la variable Fidélité exprimée en pourcentages de tâches prescrites réalisées

	Fidélité (en %)								
	PS			MS			GS		
	G1	G2	Test	G1	G2	Test	G1	G2	Test
	Effectif Moyenne (Écart-type)		Test de Student p	Effectif Moyenne (Écart-type)		Test de Student p	Effectif Moyenne (Écart-type)		Test de Student p
Album 1	N = 35 81 (12)	N = 22 82 (14)	0,918 NS	N = 37 85 (9)	N = 25 82 (11)	0,286 NS	N = 34 86 (15)	N = 22 86 (15)	0,884 NS
Album 2	N = 35 84 (16)	N = 23 82 (14)	0,558 NS	N = 34 83 (8)	N = 23 81 (14)	0,568 NS	N = 31 88 (8)	N = 22 86 (14)	0,451 NS
Album 3	N = 33 86 (10)	N = 23 83 (12)	0,339 NS	N = 33 87 (7)	N = 23 86 (9)	0,704 NS	N = 28 89 (8)	N = 23 88 (8)	0,697 NS

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

La lecture du Tableau 17 fait apparaître que, pour chaque album, les enseignants des deux groupes expérimentaux ont été fidèles au protocole expérimental : l'indice calculé supérieur à 81 % en PS et MS et à 86 % en GS est bien au-dessus du seuil minimal exigé dans la littérature scientifique.

Figure 6. Histogramme de la fidélité déclarée par les enseignants des groupes G1 et G2 pour chacun des trois scénarios (en %)



Comme l'indique la Figure 6, la fidélité déclarée par les enseignants de PS et GS augmente légèrement au fil des scénarios, de la même manière pour le groupe 1 et le groupe 2.

En MS, la fidélité déclarée pour le scénario 3 (MG1 = 87 % et MG2 = 86 %) est légèrement supérieure à celle donnée pour le scénario 1 (MG1 = 85 % et MG2 = 82 %) mais on relève une légère baisse pour la fidélité du scénario 2 (MG1 = 83 % et MG2 = 81 %). Ce résultat s'explique probablement par la complexité de l'album n° 2 (Le gros chagrin d'Edgar) proposé aux enseignants : l'histoire n'est pas un conte en randonnée, les états mentaux sont nombreux, ce qui constitue certainement un obstacle à la compréhension pour des élèves en milieu d'année de MS. Nous reviendrons plus tard sur la particularité de l'album n° 2 proposé au second trimestre aux élèves de MS.

Malgré ces légères différences, les taux de fidélité restent élevés et ne présentent aucune différence significative entre le groupe 1 et le groupe 2 : l'accompagnement n'a pas d'effet sur la fidélité. Nous considérons donc que la fidélité déclarée est suffisante pour juger que le protocole expérimental a été respecté par les deux groupes.

1.1.2. 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 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 de l'allongement 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 quotidiennement, 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 les pré- des post-tests. Cette somme, exprimée en minutes, constitue la variable *Temps* (cf. Tableau 18 et Tableau 19).

Pour rappel, un scénario *Narramus* représente 7 à 8 modules pouvant être scindés en plusieurs moments de travail : les enseignants du G1 et G2 ont déclaré consacrer, en moyenne, 10 à 20 séances d'enseignement à chaque scénario (certains ont systématiquement scindé un module en 2 voire 3 séances). Les enseignants du G3 disposaient également de quatre semaines pour travailler l'album. Dans cet intervalle, certains ont proposé cinq séances quand d'autres en organisaient une vingtaine.

Tableau 18. Données descriptives pour la variable Temps des classes de MS et GS

	Temps (en minutes)											
	GS						MS					
	G1	G2	G3	Tests post Anova			G1	G2	G3	Tests post Anova		
	Effectif Moyenne (Écart-type)			G1 vs G3	G2 vs G3	G1 vs G2	Effectif Moyenne (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	7h 31	7h 04	3h 40				6h 52	5h 48	3h 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

Dans tous les scénarios, la durée déclarée par les enseignants du groupe 1 est légèrement supérieure à celle du groupe 2 mais cette différence n'est pas significative, 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 durées d'enseignement sont significativement différentes (presque du simple au double) entre le groupe 3 d'une part et les groupes 1 et 2 d'autre part. Or les enseignants du groupe 3 ont majoritairement déclaré que la recherche les a incités à consacrer plus de temps à l'étude des albums que d'ordinaire. Les utilisateurs de *Narramus* (G1 et G2) sont donc comparés à des maitres plus mobilisés que des enseignants lambda, ce qui rend la comparaison plus exigeante et l'hypothèse 1 plus difficile à valider.

On observe la même chose pour les enseignants de PS, à l'exception du premier album de l'année, comme l'indique le Tableau 19.

Tableau 19. Données descriptives pour la variable Temps des classes de PS

	Temps (en minutes)					
	PS					
	G1	G2	G3	Tests post Anova p		
	Effectif Moyenne (Écart-type)			G1 vs G3	G2 vs G3	G1 vs G2
Album 1	N= 34 307 (164)	N= 23 355 (248)	N= 13 164 (92)	0.070 NS	0.016 *	0.633 NS
Album 2	N= 34 395 (178)	N= 23 317 (103)	N= 11 334 (440)	0.735 NS	0.976 NS	0.428 NS
Album 3	N= 32 309 (120)	N= 23 245 (92)	N= 10 328 (425)	0.964 NS	0.519 NS	0.519 NS
Moyenne par album	5h 37	5h 06	4h 35			

Pour le premier album, le temps alloué à l'enseignement par les maitres des groupes 1 et 2 est presque deux fois supérieur à celui des enseignants du groupe 3 (307 min et 355 min vs 164 min). Seules les différences G2 vs G3 sont significatives.

Pour l'album 2, les enseignants du groupe 1 accordent davantage de temps (395 min) à l'enseignement que ceux des deux autres groupes (317 min pour le G2 et 334 min pour le G3) alors que pour l'album 3, ce sont les enseignants du groupe 3 qui déclarent passer plus de temps sur l'album (328 min vs 309 min pour le G1 et 245 min pour le G2).

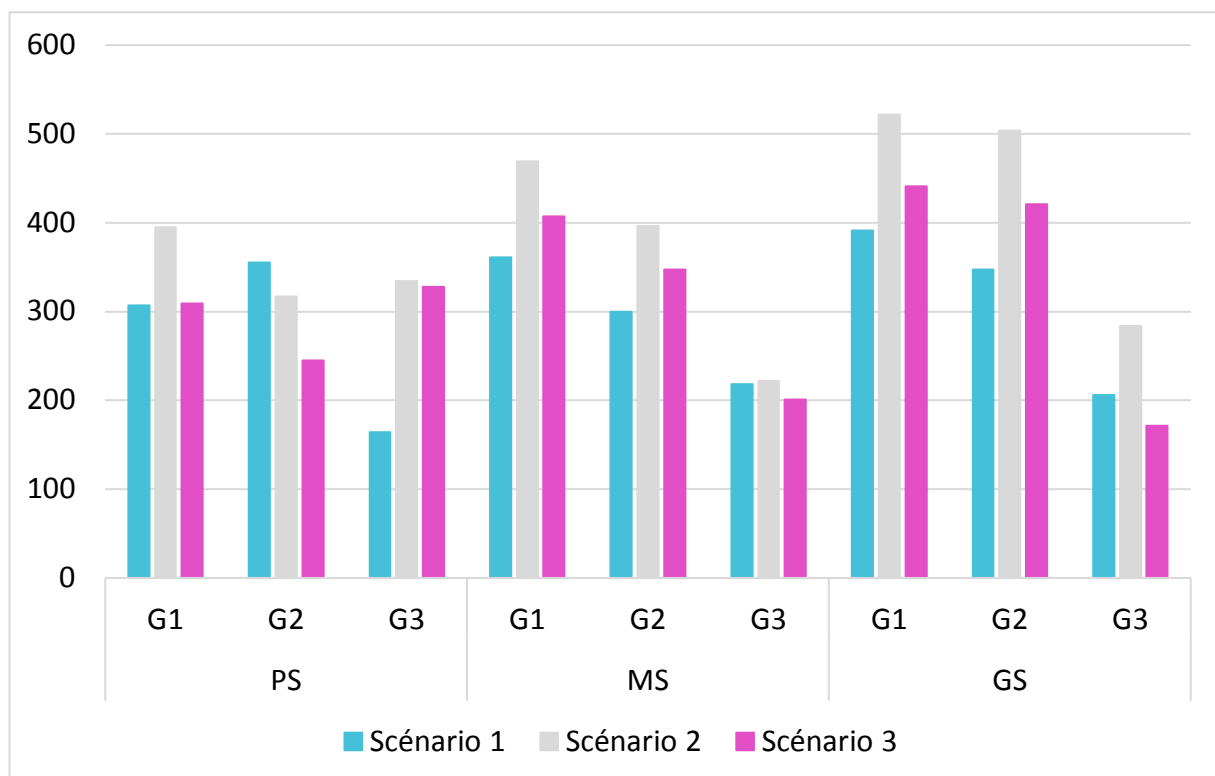
En résumé, en PS, les différences de durée d'enseignement entre les trois groupes ne varient pas toujours dans le même sens et, à l'exception du travail sur l'album 1 G2 vs G3, les écarts ne sont pas statistiquement significatifs.

Conclusion

Comme on s'y attendait, l'outillage provoque un accroissement du temps consacré à l'enseignement de la compréhension et de la narration. La Figure 7 illustre ce constat valable pour les trois niveaux (à un degré moindre pour les PS pour lesquels la tendance s'inverse pour le scénario 3).

Il sera donc nécessaire d'introduire la variable *Temps* comme variable de contrôle dans les analyses ultérieures pour tester l'hypothèse 1 et pondérer ainsi les effets constatés.

Figure 7. Histogramme de la durée d'enseignement déclarée par les enseignants des 3 groupes pour les trois scénarios (en minutes)



1.2. Constitution des trois groupes

Pour tester nos hypothèses 1 et 2, nous avons formé trois groupes. Nous avons d'abord vérifié que le niveau initial des élèves de ces groupes était équivalent. Pour cela, nous avons comparé leurs résultats au premier pré-test en lexique et en rappel de récit (macro et micro-propositions).

1.3. Équivalence initiale des trois groupes

Compte-tenu de l'âge des élèves de Petite Section et de leur développement langagier, nous ne leur avons pas proposé d'épreuve lexicale en pré-test. Nous sommes toutefois consciente de la limite introduite par ce choix puisque nous ne disposons pas de point de comparaison pour analyser les progrès en lexique. Les élèves de Moyenne et Grande section ont, quant à eux, passé une épreuve lexicale en pré-test et en post-test pour les trois scénarios. Les scores de rappel (macrostructure et microstructure) nous servent de repère pour comparer le niveau initial des groupes en PS, MS et GS. Que nous apprennent ces épreuves ?

Le Tableau 20 contient les résultats initiaux des tests passés par les élèves de PS, MS et GS et ce, pour les trois groupes.

Tableau 20. Résultats initiaux des élèves de PS, MS et GS (pré-test T1)

Épreuves	Tous les élèves		G1	G2	G3	Test
	Min-max	Moyenne (Écart-type)	Moyenne (Écart-type)			p (Anova)
Macro PS	0-17	N=454 1,6 (2,4)	N=217 1,3 (2,2)	N=136 1,6 (2,2)	N=101 2,3 (3,0)	0,002 **
Micro PS	0-22	N=454 2,5 (3,5)	N=217 2,4 (3,4)	N=136 2,2 (3,1)	N=101 3,2 (4,2)	0,08 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
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

En MS comme en GS, les tests statistiques ne révèlent aucune différence significative entre les groupes sur les variables lexique et rappel (macrostructure et microstructure). La comparaison des progrès des trois groupes sera donc légitime.

En PS, les différences ne sont pas non plus significatives sauf en macrostructure à l'avantage du G3 (MG1 = 1,3 ; MG2 = 1,6 ; MG3 = 2,3). Malgré ce résultat, nous considérons que nous sommes en mesure de tester nos hypothèses puisque la différence relevée est à l'avantage des élèves du groupe contrôle.

En lexique, on observe que les élèves réussissent environ 50 % de l'épreuve dans les trois groupes de MS et de GS, ce qui valide la construction du test et le choix des items : tous les mots et les expressions testées ne font pas tous partie du répertoire lexical des élèves mais ces derniers en maîtrisent certains. De plus, les scores moyens obtenus par les élèves de MS sont très proches pour les trois groupes : 10,5 (/25) pour le G1 ; 10,6 pour le G2 et 11,4 pour le G3. En GS, le constat est le même : 13,9 pour le G1 ; 14,2 pour le G2 et 13,8 pour le G3.

En conclusion, les trois 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 dont les résultats sont présentés dans le Tableau 21.

Tableau 21. Résultats des élèves de MS et GS au test É.Co.S.Se.

	Tous les élèves		G1	G2	G3	Test
	Min - max	Moyenne (Écart-type)	Moyenne (Écart-type)			p (Anova)
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
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

En MS et en GS, les scores obtenus au test É.Co.S.Se sont très proches et les différences observées ne sont pas significatives.

Dans la mesure où les données ne font apparaître aucune différence significative entre les groupes, en MS et GS, et ce pour les trois épreuves et une seule différence significative (à la faveur du G3) pour les PS en macrostructure, nous pouvons tester notre première hypothèse.

2. Les résultats des élèves de Moyenne et Grande Section en 2016-2017

La partie suivante présente les résultats que nous avons recueillis pour ces deux niveaux lors de la première année de l'étude. Les épreuves de petite section étant légèrement différente (pas de pré-test lexical), nous avons axé prioritairement l'analyse sur les résultats des élèves de moyenne et grande section.

2.1. 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).

2.1.1. Résultats en compréhension

La compréhension a été évaluée par une épreuve de rappel et par une autre de questionnaire (à partir du second trimestre pour les GS et du troisième trimestre pour les MS). La passation et le codage des épreuves ont été réalisés par les formateurs-évaluateurs préalablement formés par les chercheurs. Les enregistrements des récits enfantins (près de 5 200 narrations), leurs cotations et les réponses aux questions de compréhension ont été déposés sur la plateforme de l'ENS-Lyon⁶³.

2.1.1.1. Accord inter-juges

Afin de vérifier la fiabilité des codages, nous avons fait recoder de manière aléatoire 278 enregistrements de récits réalisés en MS et en GS, soit 5,3 % de l'ensemble. Un étudiant de Master 2 a effectué cette tâche rémunérée grâce au financement alloué par l'institut Carnot de l'éducation en Auvergne-Rhône-Alpes⁶⁴.

On lira, dans le Tableau 22, que l'accord inter-juge est compris entre 0,88 et 0,98, score considéré comme excellent dans la littérature (Landis & Koch, 1977).

⁶³ Nous remercions vivement Yvonnick Fesselier (ENS-Lyon) qui a développé et fait vivre cette plateforme 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.

⁶⁴ Merci à Christophe Albert.

Tableau 22. Données descriptives pour l'accord inter-juge en rappel de récit

M_T5 $\rho = (\text{IdC } 95 \%)$	mi_T5 $\rho = (\text{IdC } 95 \%)$	M_T6 $\rho = (\text{IdC } 95 \%)$	mi_T6 $\rho = (\text{IdC } 95 \%)$
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

2.1.1.2. Corrélations des épreuves de compréhension

Nous avons ensuite effectué les corrélations pour les épreuves de compréhension. En MS, les scores obtenus en post-test aux questions de compréhension sont fortement corrélés aux scores de rappel de récit comme l'indiquent les Tableau 23 à 25.

Tableau 23. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de MS du groupe 1

	Mi_T2	M_T2	Mi_T4	M_T4	Mi_T6	M_T6	Q_T6
Mi_T2	1	0,893***					
M_T2		1					
Mi_T4			1	0,931***			
M_T4				1			
Mi_T6					1	0,960***	0,588***
M_T6						1	0,564***
Q_T6							1

*** $p < 0,001$

Tableau 24. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de MS du groupe 2

	Mi_T2	M_T2	Mi_T4	M_T4	Mi_T6	M_T6	Q_T6
Mi_T2	1	0,892***					
M_T2		1					
Mi_T4			1	0,913***			
M_T4				1			
Mi_T6					1	0,963***	0,617***
M_T6						1	0,630***
Q_T6							1

*** p < 0,001

Tableau 25. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de MS du groupe 3

	Mi_T2	M_T2	Mi_T4	M_T4	Mi_T6	M_T6	Q_T6
Mi_T2	1	0,890***					
M_T2		1					
Mi_T4			1	0,938***			
M_T4				1			
Mi_T6					1	0,932***	0,471***
M_T6						1	0,527***
Q_T6							1

*** p < 0,001

L'analyse, basée sur les propositions relatées dans les épreuves de rappel (microstructure et macrostructure), fait apparaître une corrélation très forte ($r > .89$; $p < .001$), que ce soit pour le groupe 1, le groupe 2 ou le groupe 3.

Dans tous les groupes, l'analyse effectuée à partir des réponses exactes au questionnaire et de la qualité des rappels (microstructure et macrostructure) indique une corrélation forte ($0.47 < r < 0.63$; $p < 0.001$) entre ces les épreuves. Autrement dit, les élèves qui rappellent bien sont aussi, dans l'ensemble, ceux qui savent répondre aux questions qu'on leur pose. Ce résultat rejoint celui obtenu dans d'autres études (Suggate, Schaughency, McAnally, & Reese, 2018).

Les

Tableau 26 à

Tableau 28 indiquent que, en GS, les scores obtenus en post-test aux questionnaires sont également corrélés aux scores de rappel.

Tableau 26. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de GS du groupe 1

	Mi_T2	M_T2	Mi_T4	M_T4	Q_T4	Mi_T6	M_T6	Q_T6
Mi_T2	1	0,851***						
M_T2		1						
Mi_T4			1	0,882***	0,464***			
M_T4				1	0,480***			
Q_T4					1			
Mi_T6						1	0,679***	0,516***
M_T6							1	0,538***
Q_T6								1

*** $p < 0,001$

Tableau 27. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de GS du groupe 2

	Mi_T2	M_T2	Mi_T4	M_T4	Q_T4	Mi_T6	M_T6	Q_T6
Mi_T2	1	0,868***						
M_T2		1						
Mi_T4			1	0,923***	0,630***			
M_T4				1	0,649***			
Q_T4					1			
Mi_T6						1	0,947***	0,572***
M_T6							1	0,624***
Q_T6								1

*** p < 0,001

Tableau 28. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de GS du groupe 3

	Mi_T2	M_T2	Mi_T4	M_T4	Q_T4	Mi_T6	M_T6	Q_T6
Mi_T2	1	0,818***						
M_T2		1						
Mi_T4			1	0,922***	0,539***			
M_T4				1	0,529			
Q_T4					1			
Mi_T6						1	0,932***	0,626***
M_T6							1	0,641***
Q_T6								1

*** p < 0,001

En GS, la corrélation entre les scores de la microstructure et de la macrostructure est très forte ($r > 0.679$; $p < 0.001$) tout comme l'est celle qui lie les résultats au questionnaire et la qualité du rappel ($0.464 < r < 0.641$; $p < 0.001$). Ces deux constats sont valables pour les trois groupes expérimentaux.

Conclusion

Pour tous les groupes (G1, G2 et G3) et dans les deux niveaux (MS et GS), les corrélations entre les variables du rappel de récit (mi et M) et la variable questionnaire sont fortes, ce qui indique un lien important entre les compétences de compréhension et de narration que nous avons évaluées. Les corrélations très fortes entre les variables de micro- et de macrostructure indiquent que les deux mesures sont étroitement liées : les élèves qui rapportent le plus de détails sont aussi ceux qui expriment les idées essentielles du récit.

2.1.1.3. Rappel de récit et questionnaire : résultats des élèves de Grande Section

En Grande Section, les résultats de la comparaison entre les groupes expérimentaux (G1 et G2) et le groupe contrôle sont exposés dans le Tableau 29. Dans toutes les analyses, la valeur de la significativité retenue est ≤ 0.05 .

Tableau 29. Comparaison des progrès des élèves de Grande Section en rappel de récit

Albums	Tous les élèves			G1	G2	G3	Tests			d de Cohen	
	Épreuves	Min-Max	Effectifs Moyenne (é-t)	Moyenne (é-t)			Anova (p =) Test post Anova			G1vsG3	G2vsG3
Album 1	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				
	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	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		
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***			1,11	0,93	
						G1vsG3 ***	G2vsG3 ***	G1vsG2 NS			
Album 3	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***			0,62	0,48
							G1vsG3 ***	G2vsG3 NS	G1vsG2 NS		
	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***			0,65	0,48
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS		
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***			0,93	0,7	
						G1vsG3 ***	G2vsG3 ***	G1vsG2 NS			
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***			1,16	0,89	
						G1vsG3 ***	G2vsG3 ***	G1vsG2 *			

Légende : * p<0,05 ; ** p<0,01 ; ***p<0,001
NS Non Significatif

Si tous les groupes progressent entre les pré- et post-tests, les résultats du G1 et G2 sont significativement supérieurs à ceux du G3 à chaque post-test (T2, T4 et T6), dans les deux épreuves et pour tous les scores (M et mi).

Les différences G1 vs G3 et G2 vs G3 sont toutes significatives. L'hypothèse 1 selon laquelle les élèves travaillant avec l'outil *Narramus* progressent plus que les autres est donc vérifiée en GS.

Pour évaluer l'ampleur de l'effet, nous avons calculé le coefficient d de Cohen⁶⁵ : 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 (*cf.* tableau 29).

Au premier trimestre, les effets sont faibles à l'exception de la micro-structure (mi_T2) pour laquelle l'effet est moyen pour le G1 ($d = 0,73$).

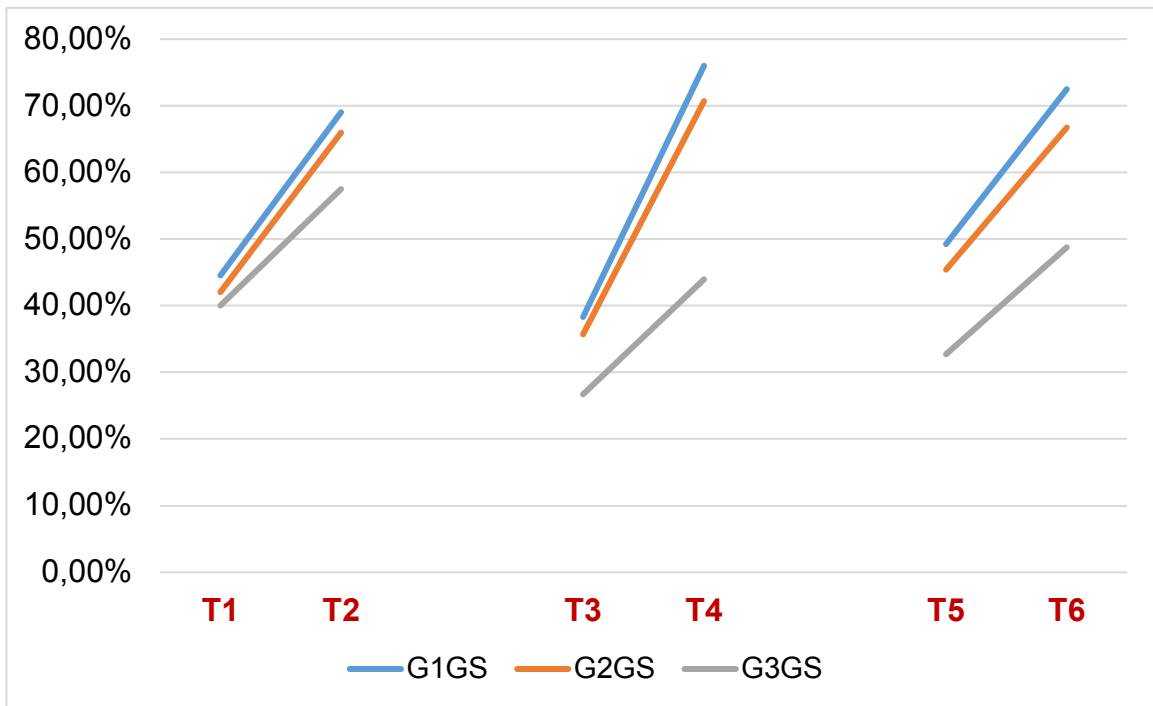
Dès le deuxième trimestre, l'écart entre G1 et G2 d'une part et G3 de l'autre s'accroît et les effets sont particulièrement forts pour l'épreuve de rappel de récit. En macrostructure, le coefficient d est égal à 1,04 pour G1 vs G3 et à 0,87 pour G2 vs G3. En microstructure, il s'élève à 1,11 pour G1 vs G3 et à 0,93 G2 vs G3.

Au troisième trimestre, les effets sont forts et s'amplifient encore.

La figure 8, à titre d'exemple, illustre les progrès des élèves de GS en rappel de récit pour les trois albums. Les performances des trois groupes sont mesurées en pourcentage de macro- puis de micro-propositions rappelées.

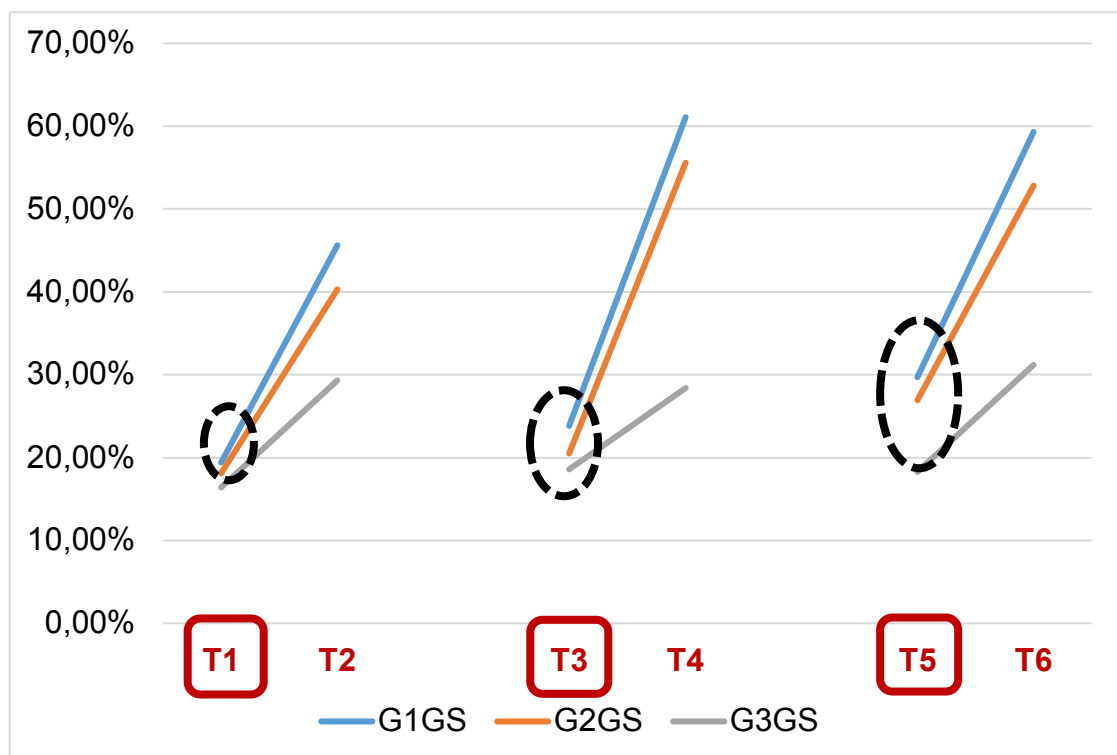
⁶⁵ La taille d'un effet correspond à la différence entre deux moyennes, divisée par l'écart-type de l'ensemble de la distribution : $(m_1 - m_2) / \sigma$.

Figure 8. Performances des élèves de GS en rappel de récit (macrostructure) : comparaison des trois groupes aux pré- et post-tests des trois scénarios *Narramus* (un par trimestre)



Les progrès des élèves en microstructure sont présentés dans la Figure 9.

Figure 9. Performances des élèves de GS en rappel de récit (microstructure) : comparaison des trois groupes aux pré- et post-tests des trois scénarios *Narramus* (un par trimestre)



Les deux graphiques placés ci-dessus (réalisés à partir du tableau 29) indiquent 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 que les différences n'étaient pas significatives au premier pré-test (T1). Les groupes équivalents au début de l'année ne le sont plus : avant l'étude du troisième album, la taille de l'effet est égale à 0,48 pour le G2 et à 0,65 pour le G1. Autrement dit, au troisième trimestre, les élèves des groupes expérimentaux sont devenus nettement plus compétents que leurs camarades du groupe contrôle pour comprendre et raconter les histoires qu'un adulte vient de leur lire et de leur raconter. Les apprentissages réalisés au premier puis au deuxième trimestre se sont transférés dans la situation d'écoute d'une nouvelle histoire : 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 croissante au fil de l'année. Elle est jugée forte dès le second scénario, c'est-à-dire après huit semaines d'enseignement.

Qu'en est-il pour les scores au questionnaire de compréhension ? Les résultats à l'épreuve questionnaire sont présentés dans le Tableau 30.

Tableau 30. Comparaison des progrès des élèves de Grande Section au questionnaire

Albums	Tous les élèves			G1	G2	G3	Tests			d de Cohen G1vsG3 G2vsG3	
	Épreuves	Min- Max	Effectifs Moyenne (é-t)				Moyenne (é-t)				
Album 1											
Album 2	Questions T3 (Q_T3) /5	0-5	N=429 2,5 (1,6)	N=192 2,5 (1,7)	N=141 2,5 (1,6)	N=96 2,4 (1,6)	0,915 NS				
	Questions T4 (Q_T4) /5	0-5	N=428 3,7 (1,4)	N=196 3,9 (1,3)	N=136 3,8 (1,4)	N=96 3,2 (1,5)	<0,001 ***			0,50	0,43
							G1vsG3 ***	G2vsG3 *	G1vsG2 NS		
Album 3	Questions T5 (Q_T5) /5	0-5	N=416 1,7 (1,3)	N=197 1,8 (1,3)	N=128 1,7 (1,4)	N=91 1,5 (1,3)	0,169 NS				
	Questions T6 (Q_T6) /5	0-5	N=413 3,1 (1,5)	N=193 3,4 (1,5)	N=130 3,2 (1,5)	N=90 2,5 (1,6)	<0,001 ***			0,60	0,47
							G1vsG3 ***	G2vsG3 **	G1vsG2 NS		

Légende : * p<0,05 ; ** p<0,01 ; ***p<0,001 NS Non Significatif

Les élèves des trois groupes progressent entre les pré- et les post-tests. Toutefois, comme pour l'épreuve de rappel de récit (M et mi), à chaque post-test (T2, T4 et T6), les scores (Q) des élèves du G1 et G2 sont significativement supérieurs à ceux du G3 pour les trois albums. Cette épreuve valide également l'hypothèse 1.

Pour évaluer l'ampleur de l'effet, nous avons de nouveau utilisé le coefficient *d* de Cohen qui nous permet de répondre à notre questionnement initial sur la force des effets en fonction des épreuves.

Au second et au troisième trimestre, les effets restent faibles à moyens ($0.43 < d < 0.60$) comparativement à ceux observés en rappel de récit. On peut raisonnablement penser que ce type d'épreuve – un questionnaire comprenant seulement 4 ou 5 questions – est beaucoup moins discriminant qu'un rappel portant sur l'intégralité du récit.

2.1.1.4. Rappel de récit et questionnaire : résultats des élèves en Moyenne section

Pour les résultats des élèves de MS, nous avons procédé exactement de la même manière et les données, recensées dans le Tableau 31, valident aussi l'hypothèse 1.

Tableau 31. Comparaison des progrès des élèves de Moyenne Section en rappel de récit

	Tous les élèves			G1	G2	G3	Test			Test	
	Épreuves	Min-Max	Effectifs Moyenne (é-t)	Effectifs Moyenne (é-t)			Anova (p=)			d de Cohen G1vsG3 G2vsG3	
Album 1	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				
	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	Macro T3 (M_T3) /23	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) /75	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				
	Macro T4 (M_T4) /23	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) /75	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***			0,52 0,64		
						G1vsG3 ***	G2vsG3 ***	G1vsG2 NS			
Album 3	Macro T5 (M_T5) /42	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) /97	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*			0,35	
							G1vsG3 NS	G2vsG3 *	G1vsG2 NS		
Macro T6 (M_T6) /42	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***			0,51 0,55		
						G1vsG3 ***	G2vsG3 ***	G1vsG2 NS			
Micro T6 (mi_T6) /97	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***			0,57 0,63		
						G1vsG3 ***	G2vsG3 ***	G1vsG2 NS			

Légende : * p < 0,05 ; ** p < 0,01 ; ***p < 0,001 ; NS Non Significatif

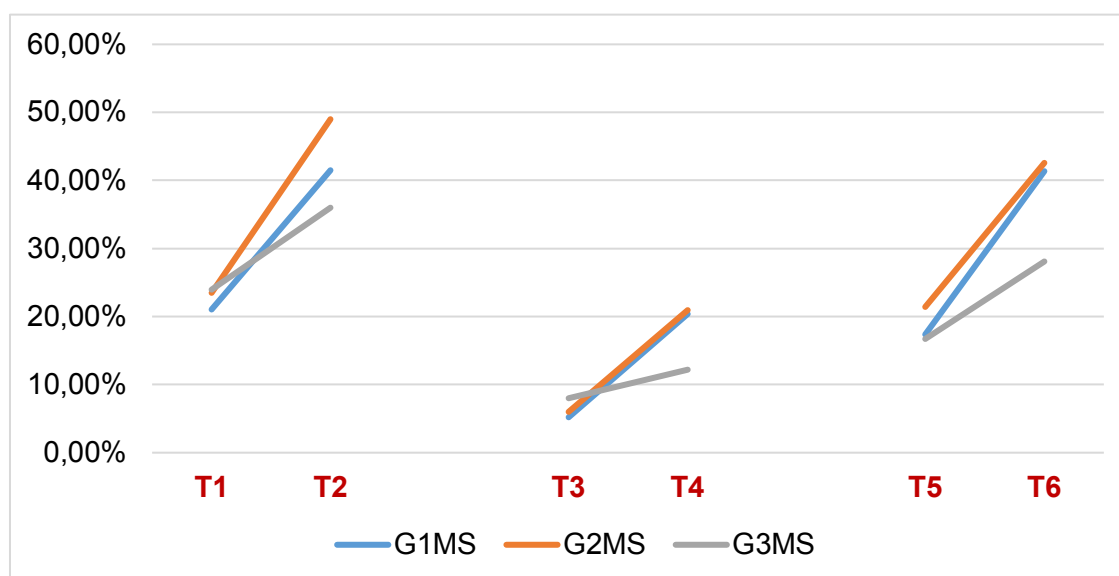
À chaque post-test (T2, T4 et T6), les résultats des deux groupes *Narramus* (G1 et G2) sont significativement supérieurs à ceux du groupe 3 à une exception près : le rappel en T2.

En macrostructure comme en microstructure, on peut noter une particularité dans les résultats obtenus par les élèves des trois groupes pour le second album intitulé *Le gros chagrin d'Edgar*. En effet, ces scores sont particulièrement bas au-pré test (T3) et restent faibles en post-test (T4), même si on observe des différences inter-groupes à l'avantage des groupes expérimentaux.

L'album explique probablement la faiblesse de ces résultats : contrairement aux autres histoires, celle-ci n'est pas un conte en randonnée. Le texte est dense (879 mots contre 467 pour le premier album) et il ne présente pas de ritournelle ou d'épisodes qui se répètent (avec l'arrivée d'un nouveau personnage, par exemple) facilement mémorisables par les élèves : les mots différents sont donc plus nombreux dans cet album que dans les deux autres. En outre, les états mentaux des personnages changent plusieurs fois au fil de l'album. On peut donc supposer que la quantité de mots, la complexité des états mentaux et la structure de l'histoire ont rendu le rappel plus difficile. En choisissant cet album, les auteurs de *Narramus* n'ont-ils pas été trop ambitieux ? Nos résultats pourraient les alerter.

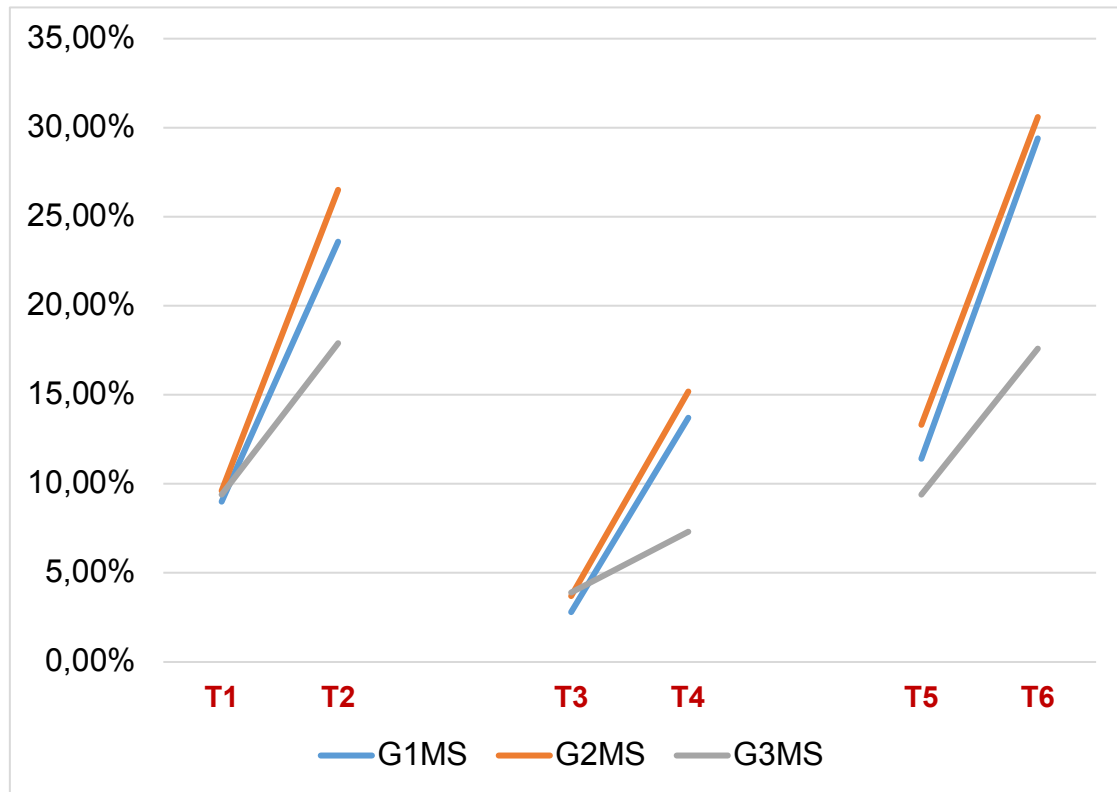
Les Figure 10 et Figure 11 illustrent les progrès des élèves de chaque groupe.

Figure 10. Performances des élèves de MS en rappel de récit (macrostructure) : comparaison des trois groupes aux pré- et post-tests des trois scénarios *Narramus* (un par trimestre)



Sur ces graphiques, la pente de progression des élèves des G1 et G2 apparaît nettement plus forte que celle des élèves de G3. L'effet constaté est de plus en plus fort au fil des scénarios.

Figure 11. Performances des élèves de MS en rappel de récit (microstructure) : comparaison des trois groupes aux pré- et post-tests des trois scénarios *Narramus* (un par trimestre)



Comme en GS, les effets sont modestes au premier trimestre : ils ne deviennent importants qu'après deux modules de quatre semaines. De manière générale, les effets en MS 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).

Des effets positifs sont également visibles à l'épreuve questionnaire, même s'ils restent moins marqués que pour les autres épreuves, comme le montre le Tableau 32.

Tableau 32. Comparaison des progrès des élèves de Moyenne Section au questionnaire

Albums	Tous les élèves			G1	G2	G3	Tests			d de Cohen G1vsG3 G2vsG3
	Épreuves	Min-Max	Effectifs Moyenne (é-t)				Moyenne (é-t)			
Album 3	Questions T5 (Q_T5) /5	0-5	N=457 1,6 (1,5)	N=224 1,5 (1,4)	N=138 1,7 (1,5)	N=95 1,6 (1,5)	0,653 NS			
	Questions T6 (Q_T6) /5	0-5	N=452 2,9 (1,6)	N=220 2,9 (1,5)	N=135 3,0 (1,7)	N=97 2,5 (1,6)	0,038*			0,31
							G1vsG3 NS	G2vsG3 *	G1vsG2 NS	

Légende : * p < 0,05 ; ** p < 0,01 ; ***p < 0,001 ; NS Non Significatif

Pour cette épreuve, les élèves des groupes 1 et 2 obtiennent, en moyenne, un score légèrement supérieur à celui des élèves du G3 avec un écart significatif entre les groupes 2 et 3.

2.1.2. Résultats en lexique

Deux épreuves lexicales ont été proposées aux élèves : l'une immédiatement après les quatre semaines d'enseignement (à la suite des épreuves de rappel et du questionnaire) et l'autre, en fin d'année (post-test différé de lexique), soit 3 à 6 mois après l'enseignement. Nous souhaitons ainsi évaluer le maintien du lexique (Pullen et al., 2010).

Les résultats de la variable lexique sont donc présentés à deux reprises et séparément en fonction du niveau de classe (GS et MS).

2.1.3. Le post-test immédiat : résultats des élèves de Grande Section

En lexique, les résultats présentés dans le tTableau 33 valident également l'hypothèse 1.

Tableau 33. Comparaison des progrès des élèves de Grande Section en lexique

	Tous les élèves			G1	G2	G3	Tests			d de Cohen	
	Épreuves	Min-Max	Effectifs Moyenne (Écart-type)				Moyenne (Écart-type)			p (Tests post Anova)	
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				
	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		
Album 2	Lexique T3 (el_T3) /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				
	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 **		
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				
	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***			0,90	0,68
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS		

Légende : * p < 0,05 ; ** p < 0,01 ; ***p < 0,001 ; NS Non Significatif

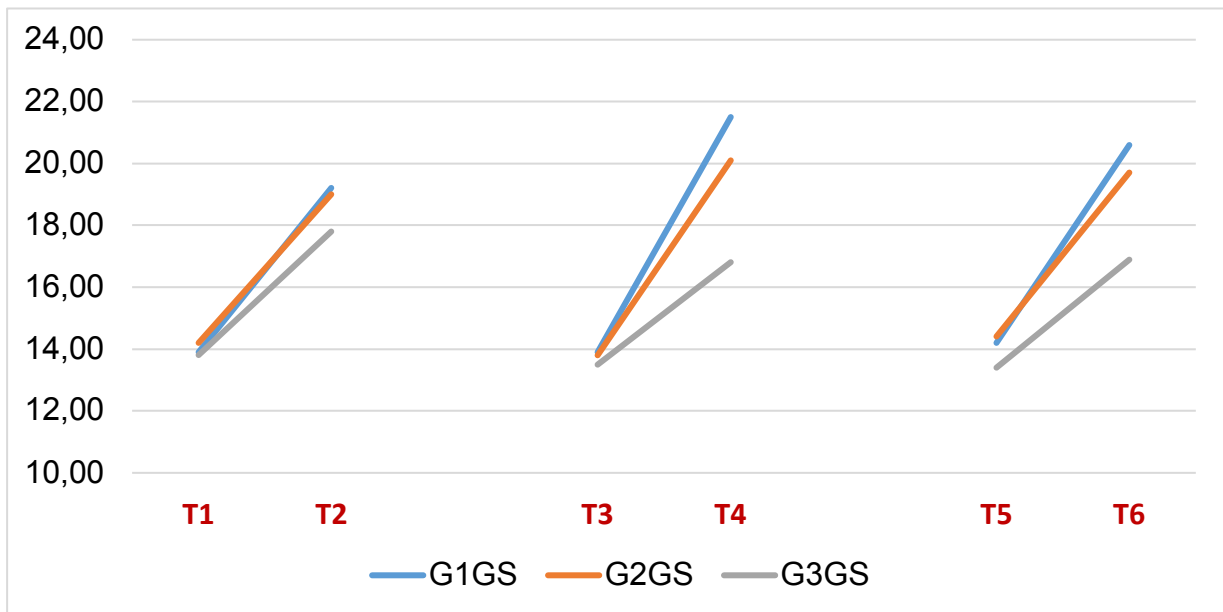
Tous les groupes progressent entre les épreuves avant et après enseignement mais les performances des élèves des G1 et G2 sont systématiquement supérieures à celles des élèves de G3. En post-test, les différences G1 vs G3 et G2 vs G3 sont toutes significatives, à l'exception du post-test lexique du premier album pour G2 vs G3. Cette exception s'explique sans doute par la nature et la moindre importance accordée à l'enseignement du lexique dans le scénario 1 commun aux trois niveaux de scolarité. En effet, pour l'album 1, *Narramus* propose un enseignement sur deux séances décrochées alors que, pour l'album 2 (et les suivants), l'enseignement est ritualisé dans chaque module avec des révisions systématiques et une utilisation plus intensive du vidéoprojecteur (images animées, vidéo...).

Notons, qu'à chaque pré-test, les différences ne sont pas significatives. Cela signifie que, contrairement au rappel de récit, il n'y a pas de transfert des connaissances lexicales d'un scénario à l'autre. Ce résultat n'est pas surprenant dans la mesure où le vocabulaire change à chaque album et que son apprentissage est spécifique. En revanche, en rappel de récit, les

attitudes, les stratégies attentionnelles d'écoute et de traitement des inférences peuvent se transférer d'un scénario à l'autre même si les histoires changent.

La Figure 12 met nettement en évidence l'équivalence des groupes en pré-test et des progrès croissants avec des écarts qui se creusent entre les élèves travaillant avec *Narramus* (G1 et G2) et les autres (G3) au fil de l'année.

Figure 12. Comparaison des progrès des élèves de Grande Section en lexique



2.1.3.1. Le post-test immédiat : résultats des élèves de Moyenne section

Les résultats obtenus en lexique pour les MS valident aussi l'hypothèse 1 (Tableau 34).

Tableau 34. Comparaison des progrès des élèves de Moyenne Section en lexique

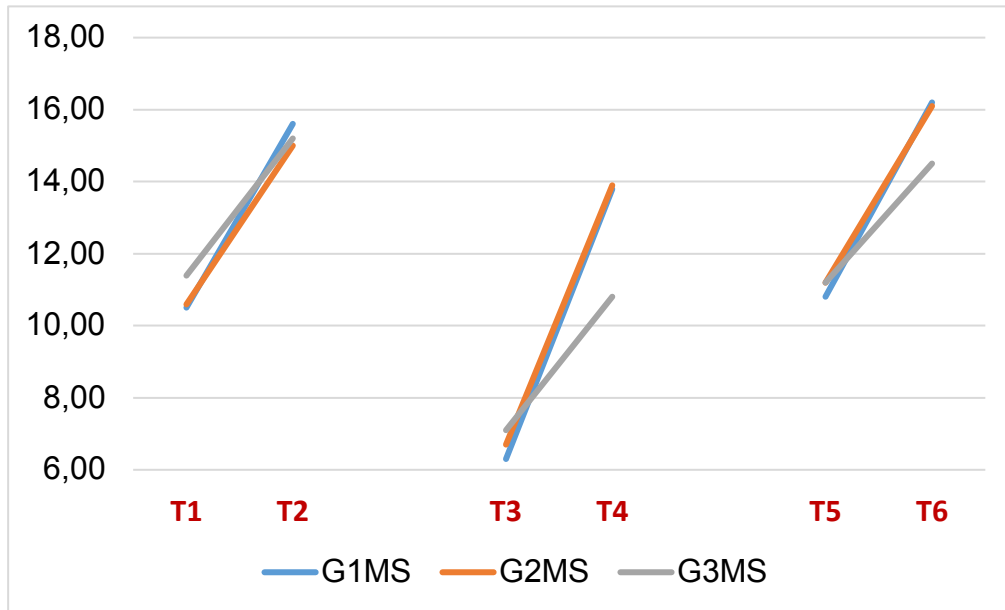
	Tous les élèves			G1	G2	G3	Tests			d de Cohen	
	Épreuves	Min-Max	Effectifs Moyenne (Écart-type)	Moyenne (Écart-type)			p (Tests post Anova)			G1vsG3	G2vsG3
Album 1	Lexique T1 (el T1) /25	0-22	N=490 10,7 (4,6)	N=229 10,5 (4,5)	N=156 10,6 (4,5)	N=105 11,4 (5,0)	0,211				
	Lexique T2 (el T2) /25	0-25	N=465 15,3 (5,5)	N=219 15,6 (5,7)	N=145 15,0 (5,5)	N=101 15,2 (5,2)	0,551				
Album 2	Lexique T3 (el T3) /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				
	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*** G1vsG3*** G2vsG3*** G1vsG2 NS			0,50	0,52
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				
	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

Légende : * p < 0,05 ; ** p < 0,01 ; ***p < 0,001 ; NS Non Significatif

Pour les post-tests T4 et T6, les scores des élèves des groupes 1 et 2 sont significativement supérieurs à ceux des élèves du groupe 3.

La Figure 13 illustre les progrès des élèves et l'écart qui se creuse dès l'album 2, à l'avantage des élèves du G1 et G2. Comme en GS, et pour les mêmes raisons, les effets obtenus sont moins forts ($0,33 < d < 0,52$) que ceux observés en rappel de récit.

Figure 13. Comparaison des progrès des élèves de Moyenne Section en lexique



2.1.3.2. Le post-test différé : résultats des élèves de Moyenne et Grande sections

La première série d'analyses était centrée sur les résultats obtenus à partir des données recueillies immédiatement après l'enseignement de quatre semaines. Que nous apprennent les résultats du post-test différé qui a eu lieu entre 3 et 6 mois après l'enseignement ?

Tableau 35. Résultats des élèves de Moyenne et Grande Section au post-test différé de lexique

Niveau	Tous les élèves		G1	G2	G3	Test		
	Min-Max	Effectifs Moyenne (Écart-type)				Moyenne (Écart-type)		
GS	2-20	N=397 14,5 (4,2)	N=193 15,2 (3,9)	N=120 14,60 (4,2)	N=84 12,9 (4,93)	< 0,001***		
						G1vsG3 ***	G2vsG3 *	G1vsG2 NS
MS	0-20	N=423 11,9 (5,0)	N=206 12,2 (4,9)	N=119 12,2 (5,2)	N=98 10,8 (5,0)	0,05		
						G1vsG3 0,064	G2vsG3 0,116	G1vsG2 1,0

Légende : * p < 0,05 ; ** p < 0,01 ; ***p < 0,001 ; NS Non Significatif

En GS et en MS, les résultats des élèves des groupes 1 et 2 sont supérieurs à ceux du groupe 3. En moyenne, ils ont mémorisé 2 à 3 mots (sur 20) de plus en GS et 1 à 2 mots en MS. Si les différences observées sont significatives en GS, elles ne sont que tendanciennes en MS.

Trois à six mois après l'enseignement, les effets sont plus forts chez les élèves qui ont travaillé avec *Narramus*. On peut noter, qu'en moyenne, les élèves de Grande Section se souviennent de 15 mots sur 20, ce qui nous renseigne sur la solidité des apprentissages effectués. En Moyenne Section, des effets sont également visibles mais restent plus modestes. Rappelons toutefois que le groupe contrôle (G3) n'est pas un groupe témoin ordinaire : les enseignants avaient 4 semaines pour travailler les mêmes albums et savaient que leurs élèves auraient à passer une évaluation centrée sur le lexique contenu dans l'album.

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

Les maîtres du G3 allouent moins de temps à l'étude des albums que ceux du G1 et du G2, nous l'avons indiqué plus haut (*cf.* Tableau 18). 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 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 ($S_f = S_i \times T_p \times G$).

Nous avons reporté ci-dessous les résultats de l'analyse de régression multiple en grande section (*cf.* Tableau 36). 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.

Tableau 36. Analyses de régression multiple en GS (modèle Sf = Si x Tps x G)

Évaluations		Macro	Micro	Questions	Lexique
Variables Album 1	Effectifs	382	382		398
	R ²	0,405	0,406		0,603
	Score initial	b = 0,583 ***	b = 0,538 ***		b = 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	Effectifs	329	329	352	367
	R ²	0,471	0,518	0,405	0,603
	Score initial	b = 0,444 ***	b = 0,473 ***	b = 0,581 ***	b = 0,627 ***
	Temps	0,258 ***	0,274 ***	0,069 NS	0,172 ***
	G1	0,285 ***	0,281 ***	0,233 ***	0,378 ***
	G2	0,241 ***	0,264 ***	0,175 **	0,206 ***
Variables Album 3	Effectifs	314	309	324	334
	R ²	0,548	0,604	0,286	0,581
	Score initial	b = 0,603 ***	b = 0,600 ***	b = 0,440 ***	b = 0,627 ***
	Temps	0,174 ***	0,158 ***	0,190 **	0,343 ***
	G1	0,177 **	0,268 ***	0,101 NS	0,137 ***
	G2	0,129 *	0,214 ***	0,108 NS	0,035 NS

Légende : * p < 0,05 ; ** p < 0,01 ; *** p < 0,001
Les coefficients b sont standardisés.

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 dans le tableau 36), 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 (Jarlegan, Piquée, Fontanieu, & Goigoux, 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 pour le rappel de récit (M et mi), entre 0,63 et 0,75 pour le lexique et entre 0,44 et 0,58 pour le questionnaire.

La durée de l'enseignement et l'outillage ont également une influence bénéfique, à peu près équivalente, sans négliger le fait que l'allongement du temps est lui-même une conséquence de l'outillage.

Pour le rappel de récit, la variable *Temps* a un coefficient compris entre 0,16 et 0,27 et la variable *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 et au questionnaire pour G1 et G2).

En 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.

En moyenne section, nous avons procédé de la même manière et avons obtenu des résultats voisins (*cf.* Tableau 37) même si le temps et l'appartenance groupale expliquent une moindre part de la variance qu'en GS.

Tableau 37. Analyses de régression multiple en MS (modèle Sf = Si x Tps x G)

Compétences		Macro	Micro	Questions	Lexique
Variables Album 1	Effectifs	386	386		407
	R ²	0,398	0,300		0,674
	Score initial	b = 0,606 ***	b = 0,506 ***		b = 0,817 ***
	Temps	0,138 **	0,178 ***		0,094 **
	G1	0,059 NS	0,047 NS		0,079 *
	G2	0,193 ***	0,167 **		0,023 NS
Variables Album 2	Effectifs	348	348		372
	R ²	0,293	0,339		0,539
	Score initial	b = 0,449 ***	b = 0,472 ***		b = 0,694 ***
	Temps	0,222 ***	0,209 ***		0,129 **
	G1	0,152 *	0,189 *		0,234 ***
	G2	0,135 *	0,179 **		0,205 ***
Variables Album 3	Effectifs	336	336	348	358
	R ²	0,424	0,461	0,354	0,554
	Score initial	b = 0,580 ***	b = 0,605 ***	b = 0,655 ***	b = 0,703 ***
	Temps	0,165 **	0,153 **	0,041 NS	0,163 ***
	G1	0,105 NS	0,092 NS	0,176 **	0,106 NS
	G2	0,074 NS	0,088 NS	0,166 **	0,071 NS

Légende : * p < 0,05 ; ** p < 0,01 ; ***p < 0,001

Les coefficients b sont standardisés.

Le pouvoir explicatif du modèle est fort en lexique (entre 54 % et 67 %) mais plus faible en rappel de récit (entre 29 % et 46 %) et au questionnaire (18 % et 17 %). 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. 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. En d'autres termes, l'effet *Narramus* semble plus consistant en GS qu'en MS : l'accroissement de la durée d'enseignement générée par l'outil joue un rôle moindre en GS, au profit des caractéristiques strictement didactiques des scénarios proposés.

On note enfin que la durée d'enseignement n'a pas d'influence significative, en MS, sur les scores au questionnaire.

2.2. Test de l'hypothèse 2 : G1 versus G2

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 déjà vérifié (cf. tableau 18) que ce dernier 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 (cf. tableaux 31, 32 et 34), les performances des élèves de G1 et de G2 ne sont pas significativement différentes, en T2 comme en T4 et T6.
- En GS (cf. tableaux 29, 30 et 33), 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 maitres 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 entièrement validée.

2.2.1. Les résultats en fonction du niveau initial de l'élève

En moyenne, le fait d'utiliser *Narramus* produit des effets positifs sur les apprentissages des élèves. Dans la perspective de démocratisation de l'école que nous avons exposée en première partie, il nous paraît nécessaire d'aller plus loin et de savoir si cet enseignement profite à tous les élèves, même aux plus faibles parmi les plus défavorisés. Autrement dit, est-ce que *Narramus* permet aux plus faibles de progresser ou profite-t-il seulement aux meilleurs des élèves de l'Éducation prioritaire ?

Puisque nous observons un effet moyen, nous avons voulu savoir si les progrès apparaissaient de manière différenciée selon le niveau initial (NI) des élèves. Pour cela, nous avons catégorisé leurs résultats en fonction des frontières constituées par $+ \frac{1}{2}$ ou $- \frac{1}{2}$ écart-type. Les analyses statistiques ont d'abord porté sur trois sous-groupes constitués selon les niveaux de performance initiaux : les élèves initialement faibles qui obtenaient des scores inférieurs ou égaux à $- \frac{1}{2}$ écart-type à chaque épreuve (macrostructure, microstructure et lexique) au T1 (pré-test), les élèves aux scores intermédiaires compris entre $- \frac{1}{2}$ et $+ \frac{1}{2}$ écart-type et les élèves initialement forts qui obtenaient des scores supérieurs ou égaux à $+ \frac{1}{2}$ écart-type. Comme on peut le voir dans le Tableau 38, chaque groupe (G1, G2 et G3) comprend des élèves des trois sous-groupes de niveaux, même si ce n'est pas exactement dans les mêmes proportions.

Tableau 38. Répartition des élèves des trois groupes en fonction de leur niveau initial au T1 en macrostructure, en microstructure et en lexique

Niveau	Épreuve	Niveau au T1	Effectif total	Effectif G1	Effectif G2	Effectif G3
MS	Macrostructure	faible	200	104 ⁶⁶	57	39
		intermédiaire	143	58	49	36
		fort	129	60	40	29
	Microstructure	faible	191	96	56	39
		intermédiaire	145	63	49	33
		fort	136	63	41	32
	Lexique	faible	155	74	51	30
		intermédiaire	192	95	60	37
		fort	143	60	45	38
GS	Macrostructure	faible	152	66	48	38
		intermédiaire	148	68	41	39
		fort	146	68	50	28
	Microstructure	faible	159	67	50	42
		intermédiaire	149	66	45	38
		fort	138	69	44	25
	Lexique	faible	128	63	34	31
		intermédiaire	176	75	60	41
		fort	145	69	45	31

⁶⁶ Ce score « anormal » est dû sans doute au nombre de MS qui ne racontent pas du tout au début de l'année.

Reste à savoir si l'enseignement de *Narramus* s'exerce différemment selon les performances initiales des élèves.

2.2.1.1. *Narramus* : un effet différencié en fonction du niveau initial ?

Pour réaliser ces analyses, nous avons constitué un groupe unique en réunissant tous les élèves qui bénéficiaient de *Narramus* (G1+G2). Nous avons ensuite comparé leurs résultats avec ceux du groupe contrôle (G3). Notre questionnement principal portant sur la présence ou non d'un effet *Narramus* chez les élèves faibles, les analyses seront principalement centrées sur les scores de ce sous-groupe. Les deux groupes n'étant pas de composition équivalente en terme d'effectifs, les résultats doivent toutefois être interprétés avec précaution.

2.2.1.1.1. Un effet *Narramus* pour les élèves faibles de l'Éducation prioritaire ?

Les Tableau 39 et Tableau 40 présentent les résultats des élèves de Grande et de Moyenne Section en rappel de récit et en lexique en fonction du niveau initial des élèves. Compte tenu de la très forte corrélation entre la macrostructure et la microstructure et pour alléger la lecture, nous avons choisi de centrer notre analyse sur la première qui correspond au nombre d'idées essentielles rappelées. Tous les traitements ont été effectués en double (Macro et micro) mais nous ne présentons ici que les plus exigeants sur le plan statistique (la variable Macro), les progrès des élèves étant plus amples en microstructure qu'en macrostructure (*cf.* annexe 8).

Les cases colorées indiquent les épreuves pour lesquelles les différences G1&G2 vs G3 sont significatives.

Tableau 39. Comparaison des progrès des élèves de Grande Section G1&G2 vs G3 en fonction du niveau initial de l'élève (rappel de récit et lexique)

Albums	Épreuves	Élèves de NI faible			Élèves de NI intermédiaire			Élèves de NI fort		
		G1&G2	G3	T Test	G1&G2	G3	T Test	G1&G2	G3	T Test
Album 1	Macro T1 (M_T1) /20	N= 114 2,4 (1,8)	N= 38 2,4 (1,7)	0,978 NS	N= 109 8,4 (1,8)	N= 39 8,7 (1,8)	0,326 NS	N= 118 14,9 (2,1)	N= 28 14,5 (1,9)	0,378 NS
	Lexique T1 (el_T1) /25	N= 97 8,1 (2,6)	N= 31 8,5 (2,2)	0,358 NS	N=135 14,1 (1,4)	N= 41 13,9 (1,6)	0,463 NS	N=114 19,1 (1,8)	N= 31 19,0 (1,7)	0,842 NS
	Macro T2 (M_T2) /20	N= 107 9,4 (6,0)	N= 33 8,3 (4,8)	0,275 NS	N= 100 14,5 (3,6)	N= 33 12,0 (4,3)	0,005 **	N= 114 16,9 (2,6)	N= 26 15,2 (2,6)	0,005 **
	Lexique T2 (el_T2) /25	N=96 15,2 (4,5)	N= 28 13,7 (3,6)	0,081 NS	N=129 19,2 (3,2)	N= 38 17,9 (3,3)	0,033 *	N=109 22,7 (2,0)	N= 26 21,9 (2,1)	0,075 NS
Album 2	Macro T3 (M_T3) /30	N= 100 5,0 (5,7)	N= 30 4,1 (5,2)	0,418 NS	N= 100 11,4 (7,7)	N= 37 8,7 (7,0)	0,063 NS	N=105 16,5 (7,8)	N= 26 11,9 (6,8)	0,004 *
	Lexique T3 (el_T3) /25	N=96 9,3 (3,6)	N= 30 10,4 (3,8)	0,142 NS	N=125 14,0 (3,8)	N= 39 12,7 (3,5)	0,07 NS	N=108 18,1 (3,0)	N= 30 17,7 (4,3)	0,682 NS
	Macro T4 (M_T4) /30	N= 90 16,9 (9,9)	N= 31 9,0 (8,2)	< 0,001 ***	N= 98 24,0 (6,5)	N= 35 14,5 (6,7)	< 0,001 ***	N= 104 25,7 (5,8)	N= 25 17,4 (6,8)	< 0,001 ***
	Lexique T4 (el_T4) /25	N=92 18,0 (4,5)	N= 28 12,8 (5,2)	< 0,001 ***	N=126 21,2 (3,2)	N= 40 17,3 (3,8)	< 0,001 ***	N=103 23,4 (1,9)	N= 26 20,6 (3,2)	< 0,001 ***
Album 3	Macro T5 (M_T5) /48	N= 94 14,7 (11,1)	N= 35 9,6 (10,6)	0,02 *	N= 97 23,7 (11,7)	N= 37 16,0 (10,8)	0,041 *	N=110 29,9 (9,2)	N= 26 25,15 (10,3)	0,041 *
	Lexique T5 (el_T5) /25	N=92 11,0 (3,4)	N= 30 10,7 (3,2)	0,709 NS	N=120 14,6 (3,2)	N= 39 13,6 (3,0)	0,083 NS	N=105 17,0 (2,6)	N= 25 16,4 (1,9)	0,218 NS
	Macro T6 (M_T6) /48	N= 96 27,4 (12,7)	N= 29 18,1 (11,7)	0,001 ***	N= 97 34,4 (9,3)	N= 31 25,2 (11,3)	0,001 **	N=105 39,7 (6,9)	N= 19 31,6 (8,3)	0,001 **
	Lexique T6 (el_T6) /25	N=91 18,0 (4,2)	N= 29 14,7 (4,4)	0,001 ***	N=120 20,3 (3,0)	N= 33 16,9 (3,8)	< 0,001 ***	N=107 22,2 (2,5)	N= 23 20,1 (2,4)	0,001 **

En rappel de récit, les différences G1 vs G3 et G2 vs G3 sont significatives dès le premier album étudié (post-test T2) pour les élèves de NI intermédiaire et fort. Pour les élèves

de NI faible, elles ne le sont pas. Pour eux, les écarts entre le groupe expérimental et le groupe contrôle ne sont significatifs qu'à partir du post-test (T4). L'effet est donc plus long à s'installer.

À partir de l'album 2, les différences sont significatives dans les deux épreuves (rappel et lexique en T4 et en T6).

Si les effets sont plus tardifs que chez les élèves initialement forts, on peut tout de même conclure qu'en grande section l'outil permet aux élèves les plus faibles de développer des compétences langagières leur permettant de mieux comprendre et raconter les histoires étudiées. Il leur permet aussi de surpasser les élèves faibles ne bénéficiant pas de *Narramus* dès le pré-test T5, c'est-à-dire avant l'étude du troisième album. Les compétences construites sur les deux premiers albums semblent donc se transférer : ils tirent plus profit des lectures qu'on leur fait, racontent mieux et comprennent davantage de mots.

Le Tableau 40 présente les progrès des élèves de MS : observe-t-on les mêmes effets qu'en GS ? Les différences significatives G1&G2 vs G3 sont également indiquées par les cases colorées.

Tableau 40. Comparaison des progrès des élèves de Moyenne Section G1&G2 vs G3 en fonction du niveau initial de l'élève (rappel de récit et lexique)

Albums	Épreuves	Élèves de NI faible			Élèves de NI intermédiaire			Élèves de NI fort		
		G1&G2	G3	T Test	G1&G2	G3	T Test	G1&G2	G3	T Test
Album 1	Macro T1 (M_T1) /20	N= 161 0,6 (0,8)	N= 39 0,8 (0,8)	0,168 NS	N= 107 4,5 (1,1)	N= 36 4,5 (1,2)	0,846 NS	N= 100 10,5 (3,0)	N= 29 10,7 (3,3)	0,180 NS
	Lexique T1 (el_T1) /25	N= 125 5,5 (2,4)	N= 30 5,1 (2,5)	0,490 NS	N= 155 10,9 (1,4)	N= 37 11,2 (1,5)	0,225 NS	N= 105 16,0 (1,9)	N= 38 16,6 (2,3)	0,358 NS
	Macro T2 (M_T2) /20	N= 150 5,2 (5,0)	N= 35 3,6 (4,3)	0,080 NS	N= 95 10,3 (5,1)	N= 31 8,3 (4,8)	0,05 *	N= 89 14,0 (4,4)	N= 27 11,6 (3,8)	0,013 *
	Lexique T2 (el_T2) /25	N= 115 10,5 (5,2)	N= 25 9,3 (3,7)	0,174 NS	N= 135 16,0 (3,7)	N= 33 15,7 (3,1)	0,708 NS	N= 98 20,4 (3,0)	N= 36 19,6 (3,2)	0,199 NS
Album 2	Macro T3 (M_T3) /23	N= 132 0,4 (1,1)	N= 32 0,4 (0,9)	0,933 NS	N= 94 1,5 (2,0)	N= 30 2,9 (2,9)	0,004 **	N= 86 2,6 (2,7)	N= 29 2,7 (2,9)	0,853 NS
	Lexique T3 (el_T3) /25	N= 114 3,9 (2,6)	N= 28 3,6 (2,5)	0,624 NS	N=149 6,3 (3,1)	N= 31 6,2 (3,0)	0,916 NS	N= 100 9,6 (3,3)	N= 36 10,5 (3,4)	0,174 NS
	Macro T4 (M_T4) /23	N= 136 2,7 (3,3)	N= 34 1,3 (2,4)	0,006 **	N= 89 5,4 (4,2)	N= 34 3,9 (3,7)	0,061 NS	N= 78 7,9 (4,3)	N= 27 4,0 (4,2)	< 0,001 ***
	Lexique T4 (el_T4) /25	N= 106 8,9 (4,6)	N= 26 7,4 (4,2)	0,108 NS	N= 133 14,2 (5,0)	N= 32 9,3 (5,2)	< 0,001 ***	N= 94 18,8 (4,0)	N= 35 15,3 (4,1)	< 0,001 ***
Album 3	Macro T5 (M_T5) /42	N= 130 4,3 (5,0)	N= 36 3,4 (3,5)	0,239 NS	N= 93 9,3 (7,5)	N= 35 7,9 (4,4)	0,203 NS	N= 92 12,6 (8,2)	N= 27 10,6 (6,0)	0,179 NS
	Lexique T5 (el_T5) /25	N= 112 7,2 (12,8)	N= 27 7,3 (3,5)	0,943 NS	N= 142 11,1 (3,4)	N= 32 10,7 (3,7)	0,524 NS	N= 99 15,1 (2,9)	N= 35 14,8 (3,3)	0,702 NS
	Macro T6 (M_T6) /42	N= 131 12,3 (10,2)	N= 32 6,7 (6,2)	< 0,001 ***	N= 90 19,4 (10,2)	N= 28 13,4 (6,5)	0,072 NS	N= 90 25,0 (9,7)	N= 28 13,4 (6,5)	< 0,001 ***
	Lexique T6 (el_T6) /25	N= 106 12,8 (4,8)	N= 26 11,2 (4,8)	0,126 NS	N= 141 16,4 (3,8)	N= 35 14,6 (4,1)	0,024 *	N= 93 19,9 (3,1)	N= 34 17,3 (3,0)	< 0,001 ***

Les effets obtenus en MS sont assez proches de ceux des GS mais de manière moins marquée. Pour les élèves de NI faible, comme en GS, les écarts ne sont significatifs en rappel de récit qu'à partir du second trimestre (T4 et T6) : deux modules de quatre semaines

d'enseignement sont nécessaires pour que les progrès des élèves les plus fragiles soient sensibles. Les écarts ne sont pas significatifs en lexique.

Pour les élèves de NI fort en revanche, les écarts G1&G2 vs G3 sont significatifs pour tous les post-tests, donc dès le premier module, en macrostructure ; ils le deviennent à partir du second (T4 et T6) en lexique.

Pour les élèves de NI intermédiaire, en rappel de récit, les différences sont significatives en T2 mais plus en T4 et T6. Elles le sont en revanche en lexique en T4 et T6.

Conclusion

En MS et en GS, on observe un effet *Narramus* quel que soit le niveau initial des élèves, mais celui-ci est moins marqué en MS qu'en GS et apparaît à différentes échéances temporelles selon les épreuves (rappel ou lexique) et les groupes (faible, intermédiaire, fort).

Nous pouvons toutefois répondre à notre questionnement initial : l'outil ne profite pas seulement aux élèves les plus performants au départ, il permet aux plus faibles de mieux comprendre les histoires à condition que l'intervention perdure. Les progrès deviennent significatifs à partir du second trimestre et sont plus forts en GS qu'en MS : un temps d'enseignement suffisamment long semble donc nécessaire pour ces élèves.

2.2.1.1.2. *Narramus* : jusqu'où va la démocratisation ?

Nous venons de voir que les trois sous-groupes (NI faible, intermédiaire et fort) profitent de l'enseignement avec *Narramus*. Mais ce bénéfice est-il le même selon les performances initiales des élèves ? Pour le savoir, nous avons étudié l'évolution des écarts entre ces trois sous-groupes pour les élèves ayant utilisé *Narramus* (seulement G1 et G2). Augmentent-ils ? (Solution 1) Sont-ils stables ? (Solution 2) Ou bien les élèves initialement faibles rattrapent-ils une partie de leur retard sur les élèves initialement intermédiaires ou forts ? (Solution 3)

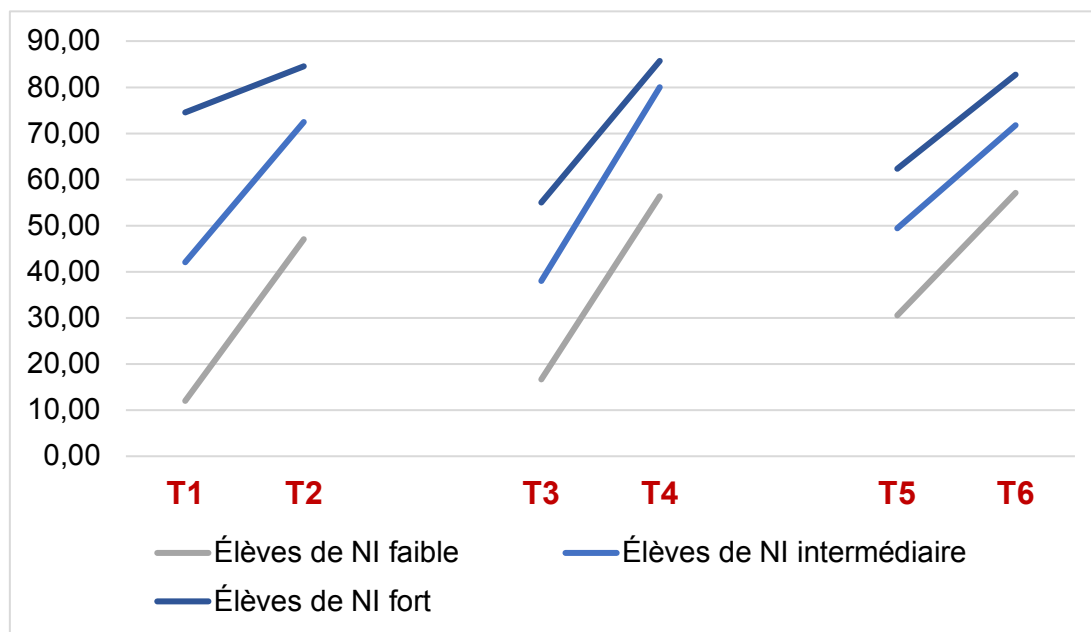
La solution 3 attesterait un rôle compensatoire même si celui-ci devrait être relativisé : le niveau initial des élèves étant beaucoup plus bas que celui des autres, leur marge de progrès est plus grande.

Les progrès ont été calculés en comparant des proportions d'items réussis aux 6 tests. Nous les avons analysés en rappel (macrostructure) et en lexique pour les élèves de Grande Section et de Moyenne Section.

2.2.1.1.2.1. Progrès des élèves de Grande Section

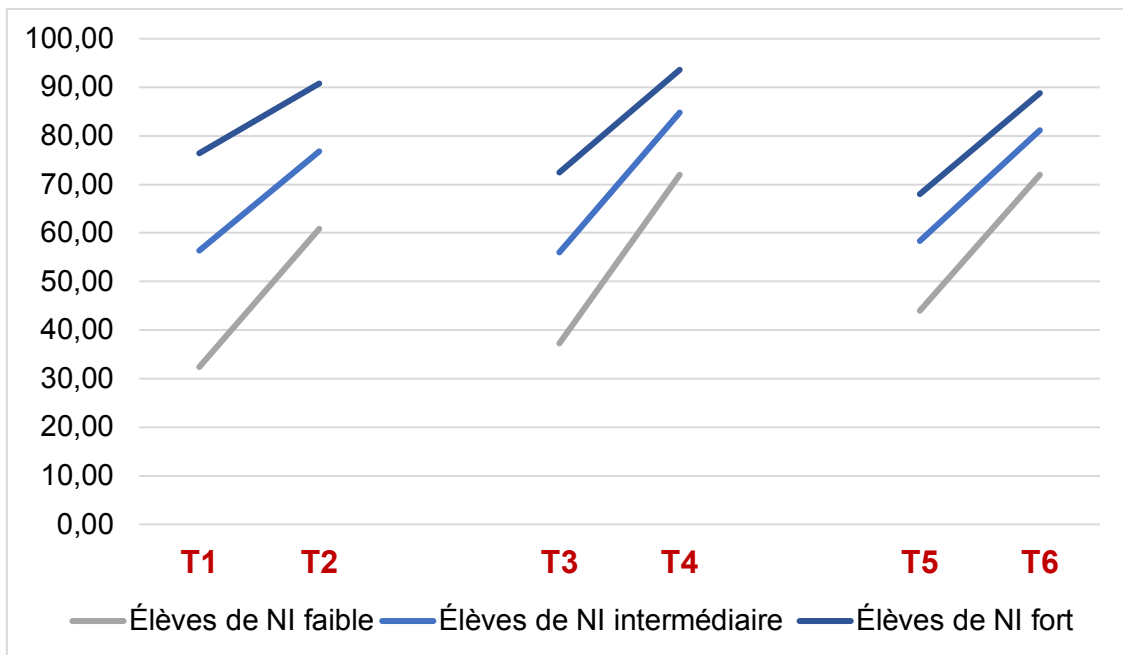
La Figure 14 indique les progrès des élèves de GS en macrostructure sur l'année ; les performances des trois sous-groupes sont mesurées en pourcentage de macro-propositions rappelées. L'observation des pentes des droites portant les segments reliant pré et post-tests laisse entrevoir que les écarts se réduisent.

Figure 14. Progrès des élèves de GS en rappel de récit en fonction de leur niveau initial



Dès le pré-test T1, les élèves de NI fort rappellent 74,5 % des macro-propositions (vs 42 % pour ceux de NI intermédiaire et 12 % pour ceux de NI faible), ce qui indique qu'ils disposaient de grandes habiletés de compréhension et de narration en début de Grande Section. En lexique, le constat est le même : ces élèves connaissent davantage de mots et d'expressions que leurs camarades, comme le montre la Figure 15. Le point de départ des élèves des trois sous-groupes sont si différents que leur marge de progrès en valeur absolue ne peut être similaire.

Figure 15. Progrès des élèves de GS en lexique en fonction de leur niveau initial (en % de réponses correctes)



Les résultats des élèves de Grande Section en rappel de récit et en lexique en fonction de leur niveau initial (présentés dans la Figure 15 ci-dessus) sont indiqués dans le Tableau 41 dans lequel nous avons ajouté les pourcentages d'items réussis (indiqués en gras).

Tableau 41. Comparaison des progrès des élèves de Grande Section travaillant avec *Narramus* en rappel de récit et lexique en fonction de leur niveau initial (NI)

Albums	Tous les élèves			Élèves de NI faible	Élèves de NI intermédiaire	Élèves de NI fort
	Épreuves	Min-Max	Effectifs N= Moyenne (% d'items réussis) (é-t)	Effectifs N = Moyenne (% d'items réussis) (é-t)		
Album 1	Macro T1 (M_T1) /20	0-19	N=341 8,6 (43 %) (5,5)	N= 114 2,4 (12 %) (1,8)	N= 109 8,4 (42 %) (1,8)	N= 118 14,9 (74,5 %) (2,1)
	Lexique T1 (el_T1) /25	1-23	N=346 14,1 (56,4 %) (4,7)	N= 97 8,1 (32,4 %) (2,6)	N=135 14,1 (56,4 %) (1,4)	N=114 19,1 (76,4 %) (1,8)
	Macro T2 (M_T2) /20	0-20	N=335 13,6 (68 %) (5,3)	N= 107 9,4 (47 %) (6,0)	N= 100 14,5 (72,5 %) (3,6)	N= 114 16,9 (84,5 %) (2,6)
	Lexique T2 (el_T2) /25	3-25	N=340 19,1 (76,4 %) (4,4)	N=96 15,2 (60,8 %) (4,5)	N=129 19,2 (76,8 %) (3,2)	N=109 22,7 (90,8 %) (2,0)
Album 2	Macro T3 (M_T3) /30	0-29	N= 318 11,4 (38 %) (8,5)	N= 100 5,0 (16,7 %) (5,7)	N= 100 11,4 (38 %) (7,7)	N=105 16,5 (55 %) (7,8)
	Lexique T3 (el_T3) /25	2-25	N=343 13,8 (55,2 %) (4,9)	N=96 9,3 (37,2 %) (3,6)	N=125 14,0 (56 %) (3,8)	N=108 18,1 (72,4 %) (3,0)
	Macro T4 (M_T4) /30	0-30	N=303 22,2 (74 %) (8,4)	N= 90 16,9 (56,3 %) (9,9)	N= 98 24,0 (80 %) (6,5)	N= 104 25,7 (85,7 %) (5,8)
	Lexique T4 (el_T4) /25	6-25	N=333 20,9 (83,6 %) (4,0)	N=92 18,0 (72 %) (4,5)	N=126 21,2 (84,8 %) (3,2)	N=103 23,4 (93,6 %) (1,9)
Album 3	Macro T5 (M_T5) /48	0-46	N=315 22,9 (47,7 %) (12,3)	N= 94 14,7 (30,6 %) (11,1)	N= 97 23,7 (49,4 %) (11,7)	N=110 29,9 (62,3 %) (9,2)
	Lexique T5 (el_T5) /25	2-23	N=324 14,3 (57,2 %) (3,9)	N=92 11,0 (44 %) (3,4)	N=120 14,6 (58,4 %) (3,2)	N=105 17,0 (68 %) (2,6)
	Macro T6 (M_T6) /48	0-48	N= 312 33,7 (70,2 %) (11,2)	N= 96 27,4 (57,1 %) (12,7)	N= 97 34,4 (71,7 %) (9,3)	N=105 39,7 (82,7 %) (6,9)
	Lexique T6 (el_T6) /25	7-25	N=325 20,2 (80,8 %) (3,7)	N=91 18,0 (72 %) (4,2)	N=120 20,3 (81,2 %) (3,0)	N=107 22,2 (88,8 %) (2,5)

Selon les albums, en post-test, les élèves de NI faible restituent entre 47 % et 57,1 % de la macrostructure. Pour les élèves de NI intermédiaire, ce taux varie entre 71,7 % et 80 % ; pour les élèves de NI fort, il oscille entre 82,7 % et 85,7 %.

Les scores des élèves de NI fort sont toujours supérieurs à ceux des élèves de NI intermédiaire et faible, ceux des seconds toujours supérieurs à ceux des troisièmes. Nous pouvons toutefois remarquer, par exemple, que le niveau final des élèves de NI faible en lexique (el_T6 = 18/25) a quasiment rejoint celui atteint par les élèves de NI intermédiaire en T2 (el_T2 = 19,2/25). L'enseignement lexical leur a donc permis d'apprendre plus de mots et de se rapprocher progressivement des résultats des élèves de NI intermédiaire.

Nous avons comparé et testé sur le plan statistique les progrès des trois sous-groupes, progrès calculés ici par la différence de scores entre les post-tests et les pré-tests. Ces résultats sont présentés dans le Tableau 42 ; les cases colorées indiquent les écarts significatifs.

Tableau 42. Analyse des progrès des élèves de GS travaillant avec *Narramus* entre les pré-tests et les post-tests en rappel et lexique en fonction du niveau initial

Différence des scores entre le post-test et le pré-test	Effectif			Anova		
	Progrès des élèves en score brut			Test post Anova		
	faible	intermédiaire	fort	faible vs interm.	faible vs fort	interm. vs fort
M_T2 - M_T1	N = 107 7,0	N = 100 6,1	N = 114 2	< 0,001 ***		
				0,351 NS	< 0,001 ***	< 0,001 ***
M_T4 - M_T3	N = 85 11,6	N = 93 12,8	N = 94 9,7	0,044 *		
				0,986 NS	0,434 NS	0,039 *
M_T6 - M_T5	N = 89 13,1	N = 92 10,7	N = 103 10,3	0,122 NS		
el_T2 - el_T1	N = 96 7,0	N = 129 5,1	N = 109 3,6	< 0,001 ***		
				< 0,001 ***	< 0,001 ***	< 0,001 ***
el_T4 - el_T3	N = 92 8,8	N = 118 7,1	N = 100 5,4	< 0,001 ***		
				0,001 **	< 0,001 ***	0,001 **
el_T6 - el_T5	N = 96 6,8	N = 129 5,8	N = 109 5,3	0,006 **		
				0,085 NS	0,005 **	0,817 NS

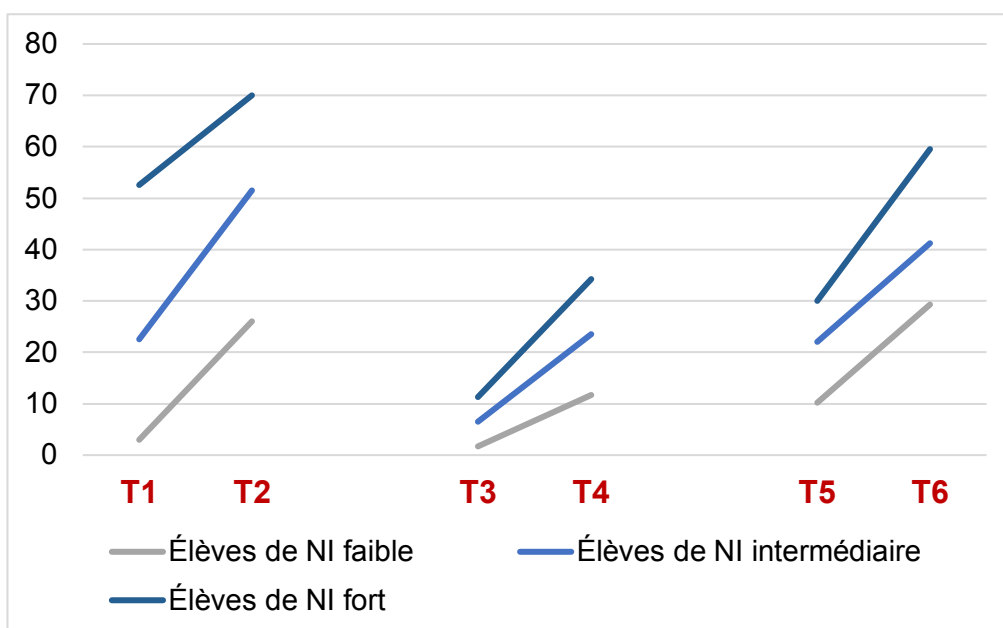
En rappel de récit (variable Macrostructure), on peut noter que les différences entre les élèves de NI faible et ceux de NI intermédiaire ne sont jamais significatives. En d'autres termes, leurs progrès sont comparables (sur la figure 14, les deux segments sont quasi parallèles). Au troisième trimestre, les différences de progrès des trois sous-groupes ne sont pas significatives : les trois segments sont parallèles, les trois sous-groupes progressent de manière similaire. C'est donc la solution 2 qui apparaît comme la réponse à nos interrogations pour la tâche de rappel de récit. Tous les sous-groupes bénéficient également de *Narramus*.

En lexique, les progrès des élèves initialement faibles sont plus importants que ceux des deux autres sous-groupes. Partis de très bas, ils tirent plus profit de l'enseignement dispensé. Cette réduction des écarts n'est donc pas surprenante. Elle constitue toutefois un indice qui nous permet d'affirmer que les progrès des élèves les plus faibles sont à la hauteur de ceux des autres élèves : eux aussi profitent positivement de l'enseignement avec *Narramus* et l'effet observé n'est pas seulement un effet moyen tiré par les plus forts.

2.2.1.1.2.2. Progrès des élèves de Moyenne Section

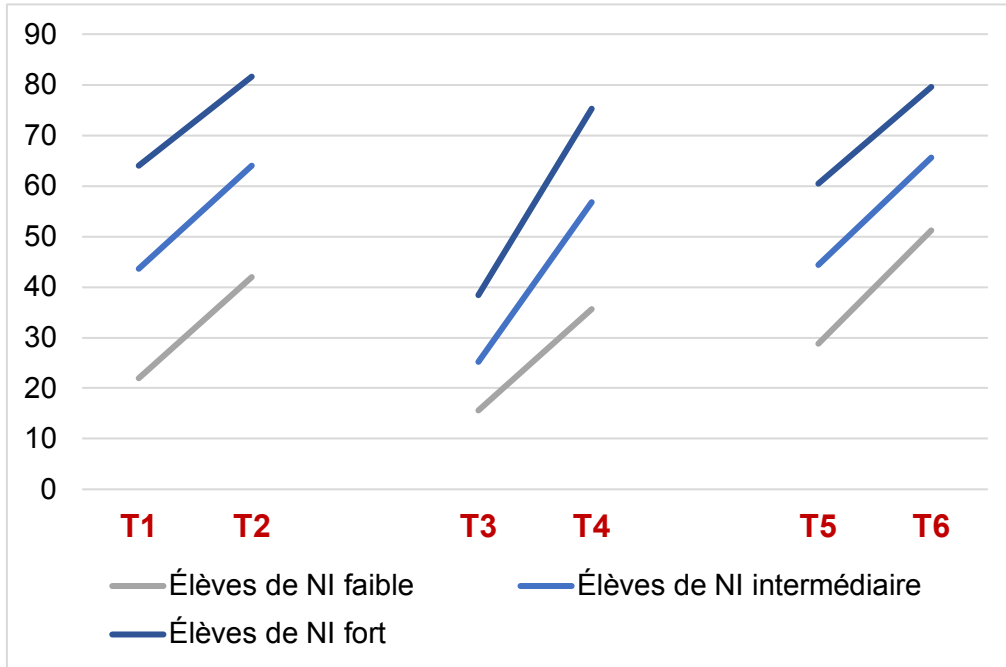
En rappel de récit, la tendance est assez proche de celle observée en Grande Section sauf pour l'album du second semestre qui, comme nous l'avons exposé plus haut, génère des résultats atypiques. Il est si difficile que tous échouent en pré-test et que les plus forts progressent plus que les plus faibles.

Figure 16. Progrès des élèves de MS en rappel de récit en fonction de leur niveau initial



En lexique, les tendances sont les mêmes qu'en GS, sauf pour l'album du second trimestre qui creuse les écarts comme l'illustre la Figure 17.

Figure 17. Progrès des élèves de MS en lexique en fonction de leur niveau initial (en % de réponses correctes)



Les résultats des élèves de Moyenne Section sont présentés dans le Tableau 43.

Tableau 43. Comparaison des progrès des élèves de Moyenne Section travaillant avec *Narramus* en rappel de récit et lexique en fonction de leur niveau initial (NI)

Albums	Tous les élèves			Élèves de NI faible scores ≤ - 0,5	Élèves de NI intermédiaire - 0,5 < scores < 0,5	Élèves de NI fort scores ≥ + 0,5
	Épreuves	Min-Max	Effectifs Moyenne (é-t)	Effectifs Moyenne (é-t)		
Album 1	Macro T1 (M_T1) /20	0-18	N= 368 4,4 (22 %) (4,4)	N= 161 0,6 (3 %) (0,8)	N= 107 4,5 (22,5 %) (1,1)	N= 100 10,5 (52,5 %) (3,0)
	Lexique T1 (el_T1) /25	0-22	N= 385 10,5 (42 %) (4,5)	N= 125 5,5 (22 %) (2,4)	N= 155 10,9 (43,6 %) (1,4)	N= 105 16,0 (64 %) (1,9)
	Macro T2 (M_T2) /20	0-20	N=364 8,9 (44,5 %) (6,1)	N= 150 5,2 (26 %) (5,0)	N= 95 10,3 (51,5 %) (5,1)	N= 89 14,0 (70 %) (4,4)
	Lexique T2 (el_T2) /25	0-25	N= 364 15,4 (61,6 %) (5,6)	N= 115 10,5 (42 %) (5,2)	N= 135 16,0 (64 %) (3,7)	N= 98 20,4 (81,6 %) (3,0)
Album 2	Macro T3 (M_T3) /23	0-10	N=337 1,3 (5,2 %) (2,1)	N= 132 0,4 (1,7 %) (1,1)	N= 94 1,5 (6,5 %) (2,0)	N= 86 2,6 (11,3 %) (2,7)
	Lexique T3 (el_T3) /25	0-19	N= 380 6,5 (26 %) (3,8)	N= 114 3,9 (15,6 %) (2,6)	N=149 6,3 (25,2 %) (3,1)	N= 100 9,6 (8,7 %) (3,3)
	Macro T4 (M_T4) /23	0-19	N=329 4,8 (20,9 %) (4,4)	N= 136 2,7 (11,7 %) (3,3)	N= 89 5,4 (23,5 %) (4,2)	N= 78 7,9 (34,3 %) (4,3)
	Lexique T4 (el_T4) /25	0-25	N=345 13,8 (55,2 %) (6,0)	N= 106 8,9 (35,6 %) (4,6)	N= 133 14,2 (56,8 %) (5,0)	N= 94 18,8 (75,2 %) (4,0)
Album 3	Macro T5 (M_T5) /42	0-34	N=344 8,0 (19,0 %) (7,5)	N= 130 4,3 (10,2 %) (5,0)	N= 93 9,3 (22,1 %) (7,5)	N= 92 12,6 (30 %) (8,2)
	Lexique T5 (el_T5) /25	0-22	N=371 10,9 (43,6 %) (4,6)	N= 112 7,2 (28,8 %) (12,8)	N= 142 11,1 (44,4 %) (3,4)	N= 99 15,1 (60,4 %) (2,9)
	Macro T6 (M_T6) /42	0-42	N= 336 17,6 (41,9 %) (11,2)	N= 131 12,3 (29,3 %) (10,2)	N= 90 19,4 (41,2 %) (10,2)	N= 90 25,0 (59,5 %) (9,7)
	Lexique T6 (el_T6) /25	0-25	N=355 16,2 (64,8 %) (4,8)	N= 106 12,8 (51,2 %) (4,8)	N= 141 16,4 (65,6 %) (3,8)	N= 93 19,9 (79,6 %) (3,1)

En T1, les élèves de NI faible racontent 3 % de la macrostructure, en T2, ils en rappellent 26 %. Pour les élèves de NI intermédiaire, ce taux varie entre 22,5 % et 51,5 % ; pour les élèves

de NI fort, il oscille entre 52,5 % et 70 %. On voit donc que les habiletés de narration varient beaucoup d'un groupe à l'autre dès le début de l'année.

En lexique, on peut noter une progression très importante pour les élèves de NI intermédiaire et de NI fort pour l'album 2. On peut supposer que les choix d'enseignement effectués par les auteurs de *Narramus* (des séances de découverte et de mémorisation quotidienne au second trimestre *versus* deux séances décrochées pour le premier trimestre) contribuent à cette forte progression.

Ces observations sont confortées par les tests statistiques reportés Tableau 44.

Tableau 44. Analyse des progrès des élèves de MS travaillant avec *Narramus* entre les pré-tests et les post-tests en rappel et lexique en fonction du niveau initial

Différence des scores entre le post-test et le pré-test	Effectif			Anova		
	Progrès des élèves en score brut			Test post Anova		
	faible	intermédiaire	fort	faible vs interm.	faible vs fort	interm. vs fort
M_T2 - M_T1	N = 150 4,6	N = 95 5,8	N = 89 3,5	0,007 **		
				0,233 NS	0,237 NS	0,005 **
M_T4 - M_T3	N = 118 2,4	N = 81 4,0	N = 74 5,4	< 0,001 ***		
				0,009 **	< 0,001 ***	0,067 NS
M_T6 - M_T5	N = 114 8,1	N = 84 9,7	N = 87 12,5	0,004 **		
				0,635 NS	0,003 **	0,158 NS
el_T2 - el_T1	N = 115 5,0	N = 135 5,0	N = 98 4,4	0,369 NS		
el_T4 - el_T3	N = 99 5,2	N = 130 8,0	N = 92 9,0	< 0,001 ***		
				< 0,001 ***	< 0,001 ***	0,231 NS
el_T6 - el_T5	N = 101 5,8	N = 132 5,4	N = 91 4,9	0,181 NS		

Il faut attendre le dernier trimestre pour que les écarts de progrès entre le groupe des élèves de NI faible et celui de NI intermédiaire ne soient plus significatifs, en rappel et en lexique. La solution 2 prédomine ici aussi.

Conclusion

Comme en GS, les élèves de NI faible en MS profitent de l'enseignement avec *Narramus* et progressent tout autant que les autres. Dans une perspective de démocratisation et d'amélioration de l'École, ces résultats sont encourageants.

2.2.1.2. Analyses inférentielles

Pour mieux comprendre l'effet *Narramus* en fonction du niveau initial des élèves et distinguer la part des progrès imputable à ses caractéristiques didactiques (choix des tâches et planification) de celle qui dépend de l'accroissement du temps d'enseignement qu'il produit, nous avons procédé à une analyse de régression multiple.

Comme précédemment, nous avons évalué l'influence de trois facteurs sur les performances des élèves en post-test : leur niveau initial au pré-test, la durée de l'enseignement dont ils ont bénéficié et leur groupe d'appartenance. Le G3 sert de groupe de référence (G1 vs G3 et G2 vs G3). Les analyses ont été effectuées par épreuve (lexique et rappel - macro -) en prenant en compte les compétences initiales de chaque élève (au T1) après avoir centré et réduit les variables. Dans tous les cas, les classes des trois groupes sont représentées dans le modèle qui porte sur les groupes d'élèves faibles, intermédiaires ou forts.

2.2.1.2.1. Rappel de récit en Grande Section et en Moyenne Section

Les résultats des analyses de régression à l'épreuve de rappel en macrostructure en GS sont présentés dans le Tableau 45. Les cases colorées indiquent que $p < 0,05$.

Tableau 45. Effets du score initial en macrostructure, du temps et du groupe sur les performances finales en macrostructure selon le score initial de l'élève de GS

		Tous les élèves	Élèves de NI faible	Élèves de NI intermédiaire	Élèves de NI fort
	Effectifs	N = 382	N = 130	N = 123	N = 129
Album 1	R ²	0,405	0,211	0,112	0,110
	M_T1z (coef. β)	0,583 (p < 0.001)	0,335 (p < 0.001)	0,154 (NS)	0,244 (p < 0.01)
	tps_total_S1z (coef. β)	0,162 (p < 0.001)	0,296 (p < 0.01)	0,090 (NS)	0,089 (NS)
	GPE1 (coef. β)	0,124 (p < 0.05)	-0,019 (NS)	0,326 (p < 0.01)	0,275 (p < 0.05)
	GPE2 (coef. β)	0,077 (NS)	-0,033 (NS)	0,224 (p < 0.05)	0,231 (NS)
Album 2	Effectifs	N = 329	N = 96	N = 120	N = 102
	R ²	0,471	0,472	0,400	0,455
	M_T3z (coef. β)	0,444 (p < 0.001)	0,431 (p < 0.001)	0,292 (p < 0.001)	0,364 (p < 0.001)
	tps_total_S2z (coef. β)	0,258 (p < 0.001)	0,433 (p < 0.001)	0,176 (p < 0.05)	0,159 (NS)
	GPE1 (coef. β)	0,285 (p < 0.001)	0,078 (NS)	0,453 (p < 0.001)	0,532 (p < 0.001)
	GPE2 (coef. β)	0,241 (p < 0.001)	0,074 (NS)	0,418 (p < 0.001)	0,518 (p < 0.001)
Album 3	Effectifs	N = 314	N = 96	N = 108	N = 95
	R ²	0,548	0,558	0,462	0,318
	M_T5z (coef. β)	0,603 (p < 0.001)	0,601 (p < 0.001)	0,528 (p < 0.001)	0,338 (p < 0.001)
	tps_total_S3z (coef. β)	0,174 (p < 0.001)	0,236 (p < 0.05)	0,150 (NS)	0,204 (NS)
	GPE1 (coef. β)	0,177 (p < 0.01)	0,157 (NS)	0,212 (NS)	0,326 (p < 0.05)
	GPE2 (coef. β)	0,129 (p < 0.05)	0,076 (NS)	0,168 (NS)	0,200 (NS)

Légende : Les coefficients b sont standardisés.

M_T1z (coef. β) signifie le niveau initial des élèves en macrostructure au pré-test T1

tps_total-S1z signifie la durée totale de l'enseignement dont les élèves ont bénéficié pour le scénario 1

GPE1 (coef. β) signifie l'appartenance au groupe expérimental 1

GPE2 (coef. β) signifie l'appartenance au groupe expérimental 2

Les modèles de régression permettent d'expliquer entre 41 % et 55 % de la variance du score final en macrostructure. Pour les élèves de NI faible, la part de variance est comprise entre 21 % et 56 % ; pour les élèves de NI intermédiaire, elle fluctue entre 11 % et 46 % et pour les élèves de NI fort, elle se situe entre 11 % et 46 %.

Le Tableau 46 indique les résultats des analyses de régression à l'épreuve de rappel de récit des élèves de Moyenne Section.

Tableau 46. Effets du score initial en macrostructure, du temps et du groupe sur les performances finales en macrostructure selon le score initial de l'élève de MS

		Tous les élèves	Élèves de NI faible	Élèves de NI intermédiaire	Élèves de NI fort
Album 1	Effectifs	N = 386	N = 167	N = 117	N = 102
	R ²	0,398	0,116	0,150	0,157
	M_T1z (coef. β)	0,606 (p < 0.001)	0,278 (p < 0.001)	0,156 (NS)	0,290 (p < 0.01)
	tps_total_S1z (coef. β)	0,138 (p < 0.01)	0,154 (p < 0.05)	0,179 (NS)	0,151 (NS)
	GPE1 (coef. β)	0,059 (NS)	0,139 (NS)	-0,030 (NS)	0,178 (NS)
	GPE2 (coef. β)	0,193 (p < 0.001)	0,205 (p < 0.05)	0,305 (p < 0.01)	0,279 (p < 0.05)
Album 2	Effectifs	N = 348	N = 135	N = 100	N = 92
	R ²	0,293	0,190	0,151	0,325
	M_T3z (coef. β)	0,449 (p < 0.001)	0,344 (p < 0.001)	0,313 (p < 0.01)	0,389 (p < 0.001)
	tps_total_S2z (coef. β)	0,222 (p < 0.001)	0,282 (p < 0.01)	0,160 (NS)	0,152 (NS)
	GPE1 (coef. β)	0,152 (p < 0.05)	-0,029 (NS)	0,205 (NS)	0,385 (p < 0.05)
	GPE2 (coef. β)	0,135 (p < 0.05)	0,038 (NS)	0,199 (NS)	0,217 (NS)
Album 3	Effectifs	N = 336	N = 129	N = 101	N = 85
	R ²	0,424	0,278	0,320	0,521
	M_T5z (coef. β)	0,580 (p < 0.001)	0,470 (p < 0.001)	0,495 (p < 0.001)	0,454 (p < 0.001)
	tps_total_S3z (coef. β)	0,165 (p < 0.01)	0,098 (NS)	0,208 (p < 0.05)	0,290 (p < 0.01)
	GPE1 (coef. β)	0,105 (NS)	0,205 (NS)	0,016 (NS)	0,191 (NS)
	GPE2 (coef. β)	0,074 (NS)	0,088 (NS)	-0,070 (NS)	0,298 (p < 0.05)

Les modèles de régression expliquent entre 29 % et 42 % de la variance du score final pour le rappel de récit en macrostructure. Pour les élèves de NI faible, la part de variance est comprise entre 12 % et 28 % ; pour les élèves de NI intermédiaire, elle fluctue entre 15 % et 32 % et pour les élèves de NI fort, elle se situe entre 16 % et 52 %.

Analyse

En MS comme en GS, le facteur qui a le coefficient le plus élevé dans l'estimation du score final est le niveau initial : entre 0,44 et 0,60 pour les GS et entre 0,45 et 0,61 pour les MS. Les coefficients sont toujours significatifs.

Le principal résultat à retenir est que le niveau initial des élèves pèse davantage dans le modèle des élèves de NI faible que dans celui des élèves de NI intermédiaire ou de NI fort. En

outre, pour les trois albums, le temps pèse davantage dans le modèle des élèves initialement faibles que dans celui des deux autres groupes. S'il existe toujours un effet bénéfique du temps sur les performances des élèves, celui-ci est plus marqué pour les élèves initialement faibles qui semblent donc avoir besoin de plus de temps pour apprendre.

2.2.1.2.2. Lexique en Grande Section et en Moyenne Section

Les résultats des analyses de régression à l'épreuve de lexique en GS sont présentés dans le Tableau 47.

Tableau 47. Effets du score initial en lexique, du temps et du groupe sur les performances finales en lexique selon le score initial de l'élève de GS

		Tous les élèves	Élèves de NI faible	Élèves de NI intermédiaire	Élèves de NI fort
Album 1	Effectifs	N = 398	N = 118	N = 155	N = 125
	R ²	0,603	0,415	0,199	0,263
	el_T1z (coef. β)	0,747 (p < 0.001)	0,573 (p < 0.001)	0,362 (p < 0.001)	0,486 (p < 0.001)
	tps_total_S1z (coef. β)	0,095 (p < 0.01)	0,244 (p < 0.01)	0,128 (NS)	- 0,078 (NS)
	GPE1 (coef. β)	0,138 (p < 0.01)	0,169 (NS)	0,168 (NS)	0,294 (p < 0.05)
	GPE2 (coef. β)	0,117 (p < 0.01)	0,096 (NS)	0,169 (NS)	0,293 (p < 0.01)
Album 2	Effectifs	N = 356	N = 104	N = 142	N = 111
	R ²	0,603	0,516	0,602	0,439
	el_T3z (coef. β)	0,627 (p < 0.001)	0,432 (p < 0.001)	0,570 (p < 0.001)	0,428 (p < 0.001)
	tps_total_S2z (coef. β)	0,172 (p < 0.001)	0,289 (p < 0.01)	0,156 (p < 0.05)	0,129 (NS)
	GPE1 (coef. β)	0,378 (p < 0.001)	0,506 (p < 0.001)	0,400 (p < 0.001)	0,513 (p < 0.001)
	GPE2 (coef. β)	0,206 (p < 0.001)	0,259 (p < 0.01)	0,254 (p < 0.01)	0,339 (p < 0.01)
Album 3	Effectifs	N = 334	N = 100	N = 122	N = 104
	Modèle 3 R ²	0,581	0,516	0,547	0,302
	el_T5z (coef. β)	0,627 (p < 0.001)	0,562 (p < 0.001)	0,514 (p < 0.001)	0,412 (p < 0.001)
	tps_total_S3z (coef. β)	0,343 (p < 0.001)	0,560 (p < 0.001)	0,329 (p < 0.001)	0,240 (p < 0.05)
	GPE1 (coef. β)	0,137 (<0,05)	-0,025 (NS)	0,246 (p < 0.01)	0,220 (NS)
	GPE2 (coef. β)	0,035 (NS)	-0,146 (NS)	0,198 (p < 0.05)	0,056 (NS)

Les coefficients b sont standardisés.

Le niveau initial des élèves (à T1, T3 et T5) est le facteur qui influence le plus les performances finales en lexique. Dans 2 cas sur 3 (albums 1 et 3), le coefficient est plus important dans le modèle des élèves initialement faibles que dans les deux autres modèles.

Comme en rappel de récit, pour les trois albums, le temps pèse davantage dans le modèle des élèves initialement faibles que dans les deux autres modèles. Ce constat est valable pour le modèle des élèves intermédiaires comparativement à celui des élèves initialement forts.

Pour l'album 1, le groupe n'influence pas significativement les performances en lexique des élèves initialement faibles. Le fait d'appartenir au groupe 1 ou 2 a des effets significatifs et positifs sur les performances en lexique des élèves initialement forts.

Pour l'album 2, on observe des effets significatifs et positifs de l'appartenance aux groupes 1 et 2 pour tous les élèves, quel que soit leur niveau initial.

Pour l'album 3, les effets positifs constatés pour élèves intermédiaires ne sont pas visibles chez les élèves initialement faibles ou forts. L'allongement du temps prend le pas sur le groupe. Le poids du temps est d'ailleurs équivalent à celui du niveau initial pour les élèves faibles.

Comme en rappel de récit, l'effet bénéfique du temps est plus marqué pour les élèves initialement faibles qui semblent donc avoir besoin de plusieurs scénarios pour apprendre à mémoriser des mots.

Le Tableau 48 indique les résultats des analyses de régression à l'épreuve de lexique en MS.

Tableau 48. Effets du score initial en lexique, du temps et du groupe sur les performances finales en lexique selon le score initial de l'élève de MS

		Tous les élèves	Élèves de NI faible	Élèves de NI intermédiaire	Élèves de NI fort
Album 1	Effectifs	N = 407	N = 128	N = 156	N = 123
	R ²	0,674	0,559	0,188	0,227
	el_T1z (coef. β)	0,817 (p < 0.001)	0,724 (p < 0.001)	0,341 (p < 0.001)	0,398 (p < 0.001)
	tps_total_S1z (coef. β)	0,094 (p < 0.01)	0,050 (NS)	0,222 (p < 0.01)	0,132 (NS)
	GPE1 (coef. β)	0,079 (p < 0.05)	0,105 (NS)	0,013 (NS)	0,225 (p < 0.05)
	GPE2 (coef. β)	0,023 (NS)	-0,015 (NS)	0,021 (NS)	0,064 (NS)
Album 2	Effectifs	N = 372	N = 110	N = 142	N = 110
	R ²	0,539	0,269	0,456	0,402
	el_T3z (coef. β)	0,694 (p < 0.001)	0,483 (p < 0.001)	0,540 (p < 0.001)	0,502 (p < 0.001)
	tps_total_S2z (coef. β)	0,129 (p < 0.01)	0,028 (NS)	0,198 (p < 0.01)	0,213 (p < 0.05)
	GPE1 (coef. β)	0,234 (p < 0.001)	0,177 (NS)	0,355 (p < 0.001)	0,341 (p < 0.01)
	GPE2 (coef. β)	0,205 (p < 0.001)	0,175 (NS)	0,305 (p < 0.01)	0,268 (p < 0.05)
Album 3	Effectifs	N = 358	N = 109	N = 137	N = 101
	R ²	0,554	0,408	0,348	0,542
	el_T5z (coef. β)	0,703 (p < 0.001)	0,619 (p < 0.001)	0,474 (p < 0.001)	0,594 (p < 0.001)
	tps_total_S3z (coef. β)	0,163 (p < 0.001)	0,044 (NS)	0,321 (p < 0.001)	0,182 (NS)
	GPE1 (coef. β)	0,106 (NS)	0,236 (p < 0.05)	0,013 (NS)	0,206 (NS)
	GPE2 (coef. β)	0,071 (NS)	0,141 (NS)	-0,047 (NS)	0,247 (p < 0,5)

Pour chaque album et dans tous les modèles, le niveau initial a le coefficient le plus important.

Le temps pèse davantage dans le modèle des élèves initialement de niveau intermédiaire. Il n'influence pas significativement les résultats des élèves de NI faible et des élèves de NI fort, à l'exception du second album pour ces derniers.

En résumé, en MS comme en GS (à l'exception des élèves de NI faible en MS), on note un effet particulièrement positif à partir de l'album 2 sur les performances lexicales : le fait d'appartenir aux groupes *Narramus* (G1 et G2) a un effet significatif en sus de l'effet-temps.

Ces résultats s'expliquent probablement par les choix didactiques des concepteurs de l'outil pour le second album⁶⁷. C'est à ce moment-là, rappelons-le, qu'ils ont modifié leur enseignement lexical en remplaçant les deux séances décrochées (album 1) par un enseignement ritualisé et quotidien (albums 2 et 3)

Synthèse des résultats des MS et des GS

Les analyses effectuées permettent de répondre à notre questionnement initial « à qui profite *Narramus* ? » : les trois sous-groupes (faible, intermédiaire et fort) progressent. On observe un effet du temps ou de l'outillage sur les performances finales.

En GS, l'effet *Narramus* est particulièrement sensible à l'allongement de la durée d'enseignement pour les élèves de NI faible. Le temps explique l'essentiel de la variance. L'effet du temps est plus modéré et fluctuant pour les élèves de NI intermédiaire ou fort, plus sensibles à l'effet outillage.

En MS, l'influence des facteurs *temps* et *appartenance groupale* est plus modeste et moins fréquente pour les élèves de NI faible. Pour eux, l'effet *Narramus* est moindre. Ils montrent aussi qu'il faut du temps pour qu'un effet apparaisse et que les bénéfices diffèrent selon l'âge des élèves.

2.3. Les résultats des élèves de Petite Section

Nous présentons ici les résultats des élèves de Petite Section en rappel de récit, questionnaire (pour l'album 3) et lexique. Cette partie se distingue de la précédente pour des raisons méthodologiques liées à l'âge des enfants et à la manière dont nous avons constitué notre échantillon. Rappelons que nous n'avons pas administré l'épreuve É.Co.S.Se et que nous n'avons pas proposé de pré-test pour le lexique. Ces contraintes méthodologiques contribuent à une analyse moins approfondie des données. Le traitement sera moins développé que les précédents, mais comme nous avons recueilli un matériau très conséquent, nous avons tenu à présenter nos résultats mêmes si leur validité est moindre.

Quels sont les résultats obtenus par les élèves de Petite Section et vont-ils dans le même sens que les analyses réalisées précédemment ? Précisons que, au regard des compétences

⁶⁷ Pour plus de détails, nous renvoyons le lecteur à la page 182 de ce même chapitre.

langagières des élèves de cet âge, nous nous attendions à ce que les effets de *Narramus* sur les élèves de PS soient plus modestes que ceux observés en MS ou GS.

2.3.1. Test de l'hypothèse 1

Comme en GS et MS, les épreuves de rappel des PS ont été administrées en pré-test et en post-test, après quatre semaines d'enseignement.

2.3.1.1. Rappel de récit et questionnaire : résultats des élèves de Petite Section

Le Tableau 49 présente les résultats obtenus par les élèves de PS. Comme nous l'avons déjà écrit, le groupe 3 était initialement plus fort que le groupe 1 (en T1, $M_T1G1 = 1,3$ versus $M_T1G3 = 2,3$; $p = 0,02$). On ne dispose toutefois pas du test É.Co.S.Se pour le confirmer.

Tableau 49. Comparaison des progrès des élèves de Petite Section en rappel de récit

Albums	Tous les élèves			G1	G2	G3	Tests			d de Cohen	
	Épreuves	Min-Max	Effectifs Moyenne (é-t)	Moyenne (é-t)			Anova (p =) Test post Anova			G1vsG3	G2vsG3
Album 1	Macro T1 (M_T1) /20	0-17	N=454 1,6 (2,4)	N=217 1,3 (2,2)	N=136 1,6 (2,2)	N=101 2,3 (3,0)	0,02			- 0,42	
							G1vsG3 0,001	G2vsG3 NS	G1vsG2 NS		
	Micro T1 (mi_T2) /72	0-22	N=454 2,5 (3,5)	N=217 2,4 (3,4)	N=136 2,2 (3,1)	N=101 3,2 (4,2)	0,78				
	Macro T2 (M_T2) /20	0-17	N=441 3,8 (4,0)	N=217 3,9 (4,2)	N=134 4,1 (3,9)	N=90 3,1 (3,7)	0,162				
	Micro T2 (mi_T2) /72	0-45	N=442 6,1 (7,1)	N=217 6,4 (7,4)	N=135 6,5 (6,8)	N=90 4,9 (7,0)	0,207				
Album 2	Macro T3 (M_T3) /24	0-23	N=459 2,2 (3,4)	N=215 2,2 (3,5)	N=141 2,3 (3,4)	N=103 2,2 (3,3)	0,895				
	Micro T3 (mi_T3) /63	0-49	N=457 3,8 (6,3)	N=213 4,0 (6,9)	N=141 3,6 (5,6)	N=103 3,6 (6,0)	0,869				
	Macro T4 (M_T4) /24	0-24	N=442 7,0 (7,1)	N=220 8,0 (7,4)	N=131 7,0 (6,9)	N=91 4,5 (5,7)	0,001			0,49	0,35
							G1vsG3 0,001	G2vsG3 0,015	G1vsG2 NS		
	Micro T4 (mi_T4) /63	0-58	N=440 14,3 (15,4)	N=219 16,5 (16,2)	N=130 14,8 (15,5)	N=91 8,2 (11)	<0,001			0,54	0,43
						G1vsG3 <0,001	G2vsG3 0,004	G1vsG2 NS			
Album 3	Macro T5 (M_T5) /19	0-15	N=471 2,4 (3,5)	N=227 2,4 (3,5)	N=143 2,2 (3,2)	N=101 2,4 (3,5)	0,634				
	Micro T5 (mi_T5) /37	0-28	N=471 3,9 (5,6)	N=227 4,1 (5,7)	N=143 3,7 (5,3)	N=101 4,0 (5,7)	0,607				
	Macro T6 (M_T6) /19	0-19	N=443 6,0 (5,6)	N=220 6,8 (5,8)	N=141 5,6 (5,5)	N=82 4,3 (5,0)	0,002			0,45	
							G1vsG3 0,001	G2vsG3 NS	G1vsG2 NS		
	Micro T6 (mi_T6) /37	0-34	N=443 10,6 (10,0)	N=220 12,1 (10,1)	N=141 10,1 (9,8)	N=82 7,5 (8,9)	< 0,001			0,46	
						G1vsG3 0,001	G2vsG3 NS	G1vsG2 NS			

Légende : * p < 0,05 ; ** p < 0,01 ; ***p < 0,001 ; NS Non Significatif

Le tableau 49 fait apparaître que les trois groupes progressent entre les pré-tests T1, T3 et T5 et les post-tests T2, T4 et T6.

En macrostructure au T1, les résultats des élèves du G3 sont supérieurs à ceux du G1 mais, à la fin du scénario 1, cette avance a fondu : la différence, en sens inverse, n'est plus significative. Pour les albums 2 et 3, en post-tests (T4 et T6), les scores du G1 sont significativement supérieurs à ceux du G3 en macro et en microstructure. Pour l'album 2, cet effet est également visible pour le G2. On peut donc conclure qu'au fur et à mesure des

scénarios, les élèves des groupes outillés racontent mieux que ceux du groupe contrôle avec des effets dont l'amplitude peut être jugée faible, voire moyenne ($0,35 < d < 0,54$).

Pour l'épreuve questionnaire aux T5 et T6, la tendance est la même, comme on peut le voir dans le Tableau 50.

Tableau 50. Comparaison des progrès des élèves de Petite Section au questionnaire

Albums	Tous les élèves			G1	G2	G3	Tests			d de Cohen G2vsG3
	Épreuves	Min-Max	Effectifs Moyenne (é-t)				Moyenne (é-t)			
Album 3	Questions T5 (Q_T5) /5	0-5	N=441 1,5 (1,7)	N=213 1,6 (1,6)	N=143 1,3 (1,5)	N=85 1,9 (1,9)	0,023			0,35
							G1vsG3 NS	G2vsG3 0,021	G1vsG2 NS	
	Questions T6 (Q_T6) /5	0-5	N=435 2,6 (2,0)	N=210 2,6 (1,9)	N=142 2,5 (1,9)	N=83 2,5 (2,0)	0,816			

Légende : * p < 0,05 ; ** p < 0,01 ; ***p < 0,001 ; NS Non Significatif

Si, au pré-test, les élèves du G3 surpassent ceux des deux autres groupes (avec une différence significative G2 vs G3) en post-test, les résultats des trois groupes sont équivalents (avec un léger avantage pour les G1). On peut donc conclure que les élèves des groupes 1 et 2 progressent davantage que ceux du G3. L'hypothèse 1 est vérifiée pour les compétences de rappel de récit.

2.3.1.2. Résultats en lexique

En PS, les épreuves de lexique n'ont été administrées qu'en post-test immédiat et différé.

2.3.1.2.1. Les post-tests immédiats : résultats des élèves de Petite Section

Comme en GS et MS, les résultats présentés dans le Tableau 51 permettent de valider l'hypothèse 1.

Tableau 51. Comparaison des scores des élèves de Petite Section en post-test immédiat de lexique

	Tous les élèves			G1	G2	G3	Tests			d de Cohen	
	Épreuves	Min-Max	Effectifs Moyenne (Écart-type)	Effectifs Moyenne (Écart-type)			p (Tests post Anova)			G1vsG3	G2vsG3
Album 1	Lexique T2 (el_T2) /25	0-24	N=441 9,9 (5,4)	N=214 10 (5,6)	N=136 10,5 (5,1)	N=91 8,9 (5,4)	0,098 NS				
Album 2	Lexique T4 (el_T4) /25	0-25	N=446 16,3 (6,3)	N=221 17,3 (6,2)	N=128 16,6 (5,8)	N=97 13,7 (6,5)	<0,001***			0,57	0,46
							G1vsG3 ***	G2vsG3 **	G1vsG2 NS		
Album 3	Lexique T6 (el_T6) /25	0-25	N=451 15,1 (6,6)	N=222 15,9 (6,4)	N=142 15,6 (6,2)	N=87 12,0 (6,8)	<0,001***			0,59	0,55
							G1vsG3 ***	G2vsG3 ***	G1vsG2 NS		

Légende : * p < 0,05 ; ** p < 0,01 ; ***p < 0,001 ; NS Non Significatif

Les performances des élèves des groupes 1 et 2 sont supérieures à celles du groupe 3 : ils ont mémorisé plus de mots et d'expressions présents dans les albums étudiés (3 à 4 sur 20). Les différences sont significatives pour les albums 2 et 3 avec un effet moyen ($0,46 < d < 0,59$).

2.3.1.2.2. Le post-test différé

En juin 2017, tous les élèves de PS ont été soumis à un post-test différé de lexique comprenant 20 mots des albums travaillés au premier et au second trimestre.

Tableau 52. Résultats des élèves de Petite Section au post-test différé de lexique

Niveau	Tous les élèves		G1	G2	G3	Test
	Min-Max	Effectifs Moyenne (Écart-type)	Moyenne (Écart-type)			Anova p test post anova
MS	0-20	N=423 11,9 (5,0)	N=206 12,2 (4,9)	N=119 12,2 (5,2)	N=98 10,8 (5,0)	0,679

Les résultats des trois groupes ne présentent pas de différence significative (tableau 52). Il n'y a plus d'effet *Narramus* en lexique 3 et 6 mois après enseignement.

2.3.1.3. Synthèse des résultats des élèves de PS

Les progrès des élèves de PS sont plus modestes en rappel et en lexique. L'effet observé est moins fort qu'en MS et GS même si, au fil de l'année, l'écart se creuse entre les élèves des trois groupes à l'avantage de ceux du G1 et du G2. À partir du deuxième scénario, il existe un effet *Narramus* ($0,35 < d < 0,59$). Il semble que les habitudes acquises en petite section au fil de l'année (participer à des temps d'enseignement collectifs, écouter les autres, attendre son tour) deviennent plus compatibles avec les caractéristiques des activités proposées dans *Narramus*. Mais on peut aussi raisonnablement penser que le choix d'un scénario commun (*La sieste de Moussa*) pour lancer l'opération *Narramus*, s'il a bien permis de créer un collectif de travail, n'a vraiment pas servi les enseignants de PS et leurs élèves (texte trop difficile, rythme des activités trop rapide...). La grande variabilité intra-groupe (G1 et G2) observée sur la variable *temps* et l'analyse qualitative (cf. chapitre 7) confirment cette hypothèse.

Si les progrès des trois groupes plaident en faveur d'un enseignement précoce et durable de la compréhension, l'avantage pris par les ceux du G1 et G2 au fil de l'année valide la plus grande efficacité des activités proposées dans *Narramus*.

2.3.1.4. Contrôle de l'influence de la durée de l'enseignement en PS

Dans 3 cas sur 6 (comparaison des durées de G1 et G2 vs G3 sur les 3 albums), les maitres du G3 accordent moins de temps à l'étude des albums que ceux du G1 et du G2 (cf. tableau 19). La durée moyenne par album est comprise entre 245 et 395 minutes pour les groupes expérimentaux, entre 164 à 334 minutes pour le groupe contrôle.

Afin de discerner les variables qui influencent les progrès, nous avons donc distingué la part des progrès imputable aux caractéristiques didactiques de l'outil de celle qui dépend de l'accroissement du temps d'enseignement qu'il produit.

Comme en MS et GS, nous avons procédé à une analyse de régression multiple prenant en compte les compétences initiales de chaque élève après avoir centré et réduit les variables. Nous avons évalué l'influence de trois facteurs sur les performances des élèves en post-test : leur niveau initial au pré-test, la durée de l'enseignement dont ils ont bénéficié et leur groupe d'appartenance. Le G3 sert de groupe de référence (G1 vs G3 et G2 vs G3).

Les résultats de cette analyse sont présentés dans le Tableau 53.

Comme nous l'avons observé dans les analyses des MS et des GS, le niveau initial est le facteur qui a le coefficient le plus élevé dans l'estimation du score final : entre 0,44 et 0,60. En rappel (M et mi), l'outil exerce une influence bénéfique avec un coefficient compris entre 0,14 et 0,27. En revanche, la durée de l'enseignement n'influence jamais significativement les résultats des élèves (en rappel et au questionnaire) en post-test, contrairement à ce que nous avons observé en MS et GS.

Tableau 53. Analyse⁶⁸ de régression multiple en PS (modèle Sf = Si x Tps x G)

Évaluations		Macro	Micro	Questions
Variables Album 1	Effectifs	342	343	
	R ²	0,214	0,226	
	Score initial	b = 0,441 ***	b = 0,452 ***	
	Temps	-0,012 NS	-0,037 NS	
	G1	0,251 ***	0,219 **	
	G2	0,223 **	0,213 **	
Variables Album 2	Effectifs	303	302	
	R ²	0,348	0,375	
	Score initial	b = 0,576 ***	b = 0,590 ***	
	Temps	0,043 NS	0,06 NS	
	G1	0,212 **	0,250 ***	
	G2	0,115 NS	0,184**	
Variables Album 3	Effectifs	317	317	313
	R ²	0,380	0,366	0,355
	Score initial	b = 0,589 ***	b = 0,573 ***	b = 0,596 ***
	Temps	0,038 NS	0,062 NS	0,060 NS
	G1	0,238 **	0,269 ***	0,074 NS
	G2	0,135 *	0,118 *	0,079 NS

Légende : * p<0,05 ; ** p<0,01 ; ***p<0,001
Les coefficients b sont standardisés.

Les modèles de régression permettent d'expliquer entre 21 % et 48 % de la variance du score final pour le rappel de récit (M et mi) et le questionnaire. Au troisième trimestre, par exemple (case grisée sur le tableau 53), le modèle permet d'expliquer 38 % de la variance des macro-propositions (R² = 0,380). Les coefficients qui expriment le poids de l'effet de chacune des variables sont égaux à 0,59 pour le niveau initial, et à 0,24 ou 0,14 pour l'appartenance aux groupes 1 ou 2. Le temps n'influence pas les performances finales des élèves à niveau initial contrôlé alors que le fait d'appartenir aux groupes *Narramus* (G1 et G2) exerce un effet

⁶⁸ Ces analyses n'ont pas été effectuées sur les scores en lexique, les élèves de PS n'ayant pas passé de pré-test.

significatif. C'est donc l'introduction de tâches développementalement pertinentes pour des élèves de cet âge (3-4 ans) – par exemple leur donner la parole souvent et de manière continue, leur faire apprendre du vocabulaire... – qui provoque l'effet le plus sensible, et non l'allongement de la durée de l'enseignement.

Autrement dit, l'effet *Narramus* en PS peut être attribué aux principes didactiques et pédagogiques de l'outil, fort différents des pratiques habituellement mises en œuvre basées sur des « lectures offertes » (IGEN, 2011).

2.3.2. Test de l'hypothèse 2

Rappelons que notre seconde hypothèse porte sur les effets produits par l'accompagnement (G1 vs G2) et qu'en MS et GS (*cf.* tableaux 17 et 18) ce dernier n'influçait ni la fidélité ni la durée d'utilisation de *Narramus* : on observe le même résultat en PS. Nous avons ensuite comparé les progrès des deux groupes : là encore, aucun écart significatif, les performances des élèves de G1 et de G2 n'étant pas différentes, en T2 comme en T4 et T6 (*cf.* tableaux 49, 50 et 51).

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 en PS, comme en MS et GS

Chapitre 5

Résultats de l'étude 1 : années 2 et 3 (2017-2019)

Chapitre 5 Résultats de l'étude 1 : années 2 et 3 (2017-2019)

Cette seconde partie porte sur les effets de cumul et de transfert chez les élèves travaillant avec *Narramus*. Nous souhaitons, en effet, poursuivre notre enquête pour analyser les progrès enfantins au fil des années (à N+2 et N+3).

Comme nous l'avons écrit dans la partie méthode, en deuxième année, l'échantillon des élèves évalués est moins important et l'organisation des groupes diffère de celle mise en place la première année :

- les enseignants du G3 ne pouvaient pas être maintenus dans l'état de frustration qui était le leur la première année. La plupart ont rejoint les G1 et G2 pour utiliser *Narramus*. Dans une étude portant sur les effets à long terme de l'outil, leurs élèves ne pouvaient être testés puisqu'ils n'avaient pas travaillé avec *Narramus* la première année.
- En deuxième année, nous n'avons plus distingué les groupes expérimentaux G1 et G2 dans la mesure où tous étaient traités de la même manière par l'encadrement de l'Éducation nationale. Nous les avons réunis en un seul groupe qualifié de groupe expérimental ($G = G1 + G2$), les équipes de circonscription ayant pour consigne de tous les accompagner à minima. Signalons que la majorité des écoles a bénéficié de temps d'accompagnement principalement organisés autour de l'analyse de pratiques.
- En deuxième année, les élèves testés en MS et GS sont les mêmes que ceux qui ont été évalués en première année : ils ont tous bénéficié préalablement d'une année d'enseignement avec *Narramus* (en PS ou en MS).

Nous nous intéressons ici au suivi longitudinal des élèves : est-ce que le fait d'utiliser *Narramus* deux ans de suite est profitable ? Y a-t-il un effet de cumul ? Nous supposons que les élèves de MS qui avaient déjà utilisé *Narramus* en PS seraient plus performants que ceux qui l'avaient découvert en MS. Le raisonnement était le même pour les résultats des élèves de GS ayant bénéficié de *Narramus* en MS.

Cette partie de l'étude est toutefois exploratoire puisqu'elle ne respecte pas des critères évaluatifs aussi exigeants que ceux utilisés pour l'analyse des résultats de la première année que nous avons définis dans notre chapitre 1.

Nous allons, en effet, comparer des groupes composés d'élèves différents (ceux de MS en 2017-2018 ne sont pas les mêmes que ceux de 2016-2017) travaillant sur d'autres albums. Les résultats des élèves de MS qui ont bénéficié de deux ans d'enseignement avec *Narramus* (PS et MS) seront comparés avec ceux des élèves de MS qui n'ont travaillé qu'un an avec *Narramus* (MS). Les comparaisons sont effectuées sur des albums différents : plusieurs enseignants exerçant dans des cours à double niveau (PS et MS ; MS et GS...), il n'était pas envisageable de leur proposer le même support avec les mêmes élèves, deux ans de suite.

Nous sommes consciente que ces différentes contraintes méthodologiques fragilisent les résultats obtenus. Ce chapitre nous permet toutefois d'explorer de nouvelles perspectives de recherche en allant au-delà des effets obtenus sur un an. Il permettra également d'introduire le questionnement de l'étude numéro 2.

1. Résultats année 2 : un effet de cumul ?

En 2017-2018, les élèves de Moyenne Section qui étaient en Petite Section en 2016-2017 et ceux de Grande Section qui étaient en Moyenne Section en 2016-2017 ont continué à bénéficier de *Narramus*. Ils ont de nouveau été évalués en pré- et post-tests (tests T7 et T8 au premier trimestre, T9 et T10 au deuxième, T11 et T12 au troisième) sur trois nouveaux albums (différents pour chaque niveau : PS, MS et GS). Nous avons cherché à savoir si les élèves continuaient de progresser et si ces progrès variaient selon leur niveau initial.

Pour effectuer les comparaisons, nous avons calculé des taux de réussite pour chaque test. Puisque, par exemple, les albums ne comportaient pas le même nombre de micro-propositions, nous avons comparé les pourcentages de micro-propositions rappelées, au même âge et au même trimestre de l'année scolaire, par les élèves ayant bénéficié de *Narramus* un an ou deux ans.

1.1. Comparaison des progrès des élèves : 1 an de *Narramus* vs 2 ans de *Narramus*

Trois albums ont été étudiés chaque année (un par trimestre) : toutes les histoires étaient illustrées, elles comportaient un lexique varié et conséquent ; le récit reposait sur de nombreux implicites.

En outre, nous cherchons à évaluer les effets généraux de *Narramus*, pas ceux d'un scénario d'un album en particulier. Cet outil – nous l'avons vu dans le chapitre 2 – est construit

sur un modèle identique d'un scénario à l'autre. Les cibles et les activités sont stabilisées d'un opus à l'autre, même si les auteurs les adaptent aux spécificités de chaque texte.

Les différences entre albums se traduisent par des nombres de macro-propositions et de micro-propositions variables selon les albums, comme on peut le voir dans le Tableau 54.

Tableau 54. Nombre de macro-propositions et de micro-propositions présentes dans les albums proposés aux élèves de MS et GS en fonction de l'année

	MS de l'année 1 (2016-2017) (1 an de <i>Narramus</i>)			MS de l'année 2 (2017-2018) (2 ans de <i>Narramus</i>)	
	Macro	Micro		Macro	Micro
Album 1	20	72	Album 4	12	52
Album 2	23	75	Album 5	11	63
Album 3	42	97	Album 6	13	52
	GS de l'année 1 (2016-2017) (1 an de <i>Narramus</i>)			GS de l'année 2 (2017-2018) (2 ans de <i>Narramus</i>)	
Album 1	20	72	Album 4	13	75
Album 2	30	82	Album 5	18	93
Album 3	48	109	Album 6	12	105

1.1.1. Comparaison des progrès des élèves de MS

Nous comparons les scores des élèves ayant pratiqué *Narramus* un an avec ceux qui l'ont pratiqué deux ans. Pour comparer les réussites, nous avons ramené tous les scores en pourcentages d'items réussis. Nous avons utilisé un test de comparaison des proportions de bonnes réponses (test de Student appliqué à des pourcentages).

Les résultats des élèves de MS sont présentés dans le tableau 55. La lecture du tableau se fait ligne par ligne : sur la première, le test indique que la différence entre les résultats au M_T1 et en M_T7 est significative ($p = 0,0004$).

Nous avons indiqué les comparaisons pour l'album 2 vs l'album 5, mais nous ne les analyserons pas car l'album n°2 (*Le gros chagrin d'Edgar*), de l'aveu même des auteurs de *Narramus*, était trop complexe pour des enfants de MS. En année 1, les élèves avaient obtenu des scores très faibles sur cet album (cf. Tableau 55). Par conséquent, en année 2, les performances des élèves sur un album plus adapté ne pouvaient être que meilleures.

Tableau 55. Comparaison des pourcentages de réussite des élèves de MS (G1 + G2) de l'année 1 et des élèves de MS (G) de l'année 2 en macrostructure et en lexique

Résultats des MS de l'année 1 (2016-2017) (1 an de <i>Narramus</i>)				Résultats des MS de l'année 2 (2017-2018) (2 ans de <i>Narramus</i>)				Test T test
	Variables Score /	Effectifs Moyenne (é-t)	% de réussite		Variables Score /	Effectifs Moyenne (é-t)	% de réussite	p =
Album 1	M_T1 /20	N= 368 4,4 (4,4)	22,1 (22,0)	Album 4	M_T7 /12	N =377 3,4 (3,0)	28,2 (25,0)	0.0004
	M_T2 /20	N=364 8,9 (6,1)	44,4 (30,7)		M_T8 /12	N =350 7,6 (3,2)	63,2 (26,3)	< 0,0001
	el_T1 /25	N= 385 10,5 (4,5)	42,1 (18,0)		el_T7 /25	N =393 10,9 (4,6)	43,8 (18,3)	NS
	el_T2 /25	N= 364 15,4 (5,6)	61,5 (22,6)		el_T8 /25	N =353 19,4 (4,8)	77,7 (19,1)	< 0,001
Album 2	M_T3 /23	N=337 1,3 (2,1)	5,6 (9,1)	Album 5	M_T9 /11	N =343 3,3 (2,7)	30,5 (24,7)	<i>Comparaison non fiable</i>
	M_T4 /23	N=329 4,8 (4,4)	20,7 (15,0)		M_T10 /11	N =315 7,1 (3,3)	64,4 (29,9)	
	el_T3 /25	N= 380 6,5 (3,8)	26,1 (15,0)		el_T9 /25	N =363 10,8 (4,6)	43,3 (18,3)	
	el_T4 /25	N=345 13,8 (6,0)	55,3 (23,9)		el_T10 /25	N =316 18,7 (4,8)	74,9 (19,1)	
Album 3	M_T5 /42	N=344 8,0 (7,5)	19,0 (17,8)	Album 6	M_T11 /13	N = 326 2,9 (2,9)	22,4 (22,2)	0.0295
	M_T6 /42	N= 336 17,6 (11,3)	41,9 (26,9)		M_T12 /13	N = 309 5,8 (3,8)	45,0 (21,0)	NS
	el_T5 /25	N=371 10,9 (4,6)	43,8 (18,3)		el_T11 /25	N = 341 10,8 (5,3)	43,3 (21,0)	NS
	el_T6 /25	N=355 16,2 (4,8)	64,8 (19,3)		el_T12 /25	N = 307 18,8 (5,5)	75,4 (22,0)	< 0,001

Si l'on compare les pourcentages de réussite ligne par ligne pour le premier trimestre, les scores des élèves de Moyenne Section ayant travaillé deux ans avec *Narramus* sont significativement supérieurs à ceux obtenus par les élèves travaillant *Narramus* pour la première année en macrostructure (pré-test et post-test) et en lexique (post-test).

Pour le troisième trimestre, les écarts observés sont significatifs en pré-test de rappel et en post-test lexical.

Ces résultats, bien que fragiles méthodologiquement, indiquent une tendance intéressante :

1° les élèves de MS ayant déjà travaillé avec *Narramus* démarrent l'année en racontant mieux que ceux qui le découvrent. Pour les deux pré-tests de rappel, ces élèves sont meilleurs que ceux qui découvrent *Narramus*.

2° l'enseignement lexical sur une deuxième année permet aux jeunes élèves d'obtenir des résultats significativement supérieurs à chaque post-test. Ils mémorisent davantage de mots et d'expressions au fil des scénarios. On peut supposer que l'apprentissage cumulé sur plusieurs albums est bénéfique. Mais on peut aussi penser que les maitres utilisent l'outil différemment. L'analyse qualitative (chapitre 7) corrobore cette dernière hypothèse même s'il est impossible de trancher.

1.1.2. Comparaison des progrès des élèves de GS

Comme en MS, en pré-test et en post-test, les scores des élèves de Grande Section ayant travaillé deux ans avec *Narramus* sont supérieurs (ou quasiment identiques comme en M_T4 et M_10) à ceux obtenus par les élèves travaillant *Narramus* pour la première année. Ces résultats sont présentés dans le Tableau 56.

Tableau 56. Comparaison des résultats des élèves de GS (G1 et G2) de l'année 1 et des élèves de GS (=G) de l'année 2

Résultats des GS de l'année 1 (2016-2017) (1 an de <i>Narramus</i>)				Résultats des GS de l'année 2 (2017-2018) (2 ans de <i>Narramus</i>)				Test
	Variables Score /	Effectifs Moyenne (é-t)	% de réussite		Variables Score /	Effectifs Moyenne (é-t)	% de réussite	p =
Album 1	M_T1 /20	N=341 8,6 (5,5)	43,2 (27,4)	Album 4	M_T7 /13	N =387 6,4 (4,2)	49,0 (32,5)	0.0098
	M_T2 /20	N=335 13,6 (5,3)	67,9 (26,7)		M_T8 /13	N =383 10,1 (3,4)	77,9 (26,0)	< 0,001
	el_T1 /25	N=346 14,1 (4,7)	56,2 (18,8)		el_T7 /25	N =397 15,2 (4,7)	60,7 (18,8)	0,0012
	el_T2 /25	N=340 19,1 (4,4)	76,5 (17,8)		el_T8 /25	N =390 21,5 (3,9)	85,8 (15,6)	< 0,001
Album 2	M_T3 /30	N= 318 11,1 (8,5)	37,1 (28,3)	Album 5	M_T9 /18	N =355 7,9 (5,2)	44,0 (29,0)	0,0019
	M_T4 /30	N=303 22,2 (8,4)	74,0 (28,1)		M_T10 /18	N =340 13,3 (4,1)	73,8 (22,7)	NS
	el_T3 /25	N=343 13,8 (4,9)	55,4 (19,7)		el_T9 /25	N =378 14,3 (4,3)	57,1 (17,3)	NS
	el_T4 /25	N=333 20,9 (4,0)	83,7 (15,9)		el_T10 /25	N =355 21,0 (3,9)	84,1 (15,6)	NS
Album 3	M_T5 /48	N=315 22,9 (12,3)	47,7 (25,7)	Album 6	M_T11 /12	N = 341 6,5 (3,2)	54,4 (27,0)	0,0012
	M_T6 /48	N= 312 33,7 (11,2)	70,2 (23,4)		M_T12 /12	N = 311 8,4 (2,8)	70,0 (23,6)	NS
	el_T5 /25	N=324 14,3 (3,9)	57,1 (15,5)		el_T11 /25	N = 370 13,9 (4,8)	55,7 (19,2)	NS
	el_T6 /25	N=324 20,2 (3,7)	80,9 (14,9)		el_T12 /25	N = 349 20,8 (4,1)	83,3 (16,5)	NS

En GS, les élèves ayant déjà travaillé avec l'outil obtiennent des scores significativement supérieurs en rappel de récit et en lexique en pré- et en post-test sur le premier album (album 4). Ils démarrent ainsi l'année scolaire avec des habiletés de narration qui leur permettent de mieux raconter une histoire.

Pour les albums étudiés au second et au troisième trimestre, les différences significatives sont présentes aux pré-tests de rappel M_T3 vs M_T9 et M_T5 vs M_T11. On observe donc que les élèves sont systématiquement meilleurs sur la première narration.

Les effets observés « 1 an avec *Narramus* » versus « 2 ans » en GS sont donc assez proches de ceux des MS.

Conclusion

Nous supposons que les élèves de MS qui avaient déjà utilisé *Narramus* en PS seraient plus performants que ceux qui n'en avaient bénéficié qu'en MS. Nous avons le même raisonnement pour les résultats des élèves de GS ayant bénéficié d'un an de *Narramus* en MS. Nous avons échoué à établir ce fait, même si nos résultats indiquent une tendance positive sur le premier trimestre de chaque année en lexique en MS et sur tous les pré-tests en rappel en MS et GS. Les scores obtenus aux pré-tests en rappel constituent toutefois des indices d'un possible cumul qui se voit par transfert, c'est à dire avant enseignement.

Notre méthode de comparaison de pourcentages d'items réussis à des tests de même nature (mais composés d'éléments différents) fragilise nos conclusions. On peut seulement dire que les élèves semblent avoir développé des habiletés leur permettant de mieux raconter un album avant enseignement. L'enseignement lexical permet également aux plus jeunes élèves (MS) d'apprendre plus de mots et d'expressions après une deuxième année d'enseignement.

1.2. Suivi des résultats des élèves sur deux ans

Les résultats précédents indiquent une tendance selon laquelle les élèves ayant travaillé deux ans avec l'outil pourraient être plus performants que ceux l'ayant pratiqué un an. Pour tenter d'éprouver ce résultat, nous avons procédé à une étude longitudinale et nous avons analysé les progrès des mêmes élèves d'une année sur l'autre.

1.2.1. Suivi des élèves de PS vers MS

Le Tableau 57 présente les résultats des élèves de T1 à T12 en macrostructure et en lexique. Nous n'avons conservé que les scores des élèves de Petite Section des G1 et G2 et avons exclu ceux des G3 puisque ceux-ci ne participaient plus à l'étude l'année suivante.

Les cases en gris clair indiquent les résultats des élèves en post-tests ; les cases vides correspondent aux pré-tests lexicaux de l'année 1 qui n'avaient pas été proposés aux élèves de Petite Section.

Tableau 57. Scores des élèves de PS puis de MS en rappel de récit et en lexique de T1 à T12

Épreuves	Année 1 Elèves de G1+G2 2016-2017 (PS)		Année 2 Elèves de G1+G2 2017-2018 (élèves passés en MS)	
	Effectifs N = Moyenne % de réussite		Effectifs N = Moyenne % de réussite	
Rappel Macrostructure (score sur)	M_T1 /20	N= 353 1,41 7,05 %	M_T7 /12	N = 377 3,39 28,25 %
	M_T2 /20	N = 351 3,94 19,7 %	M_T8 /12	N = 350 7,58 63,17 %
	M_T3 /24	N = 337 2,26 9,42 %	M_T9 /11	N = 343 3,35 30,45 %
	M_T4 /24	N = 329 7,65 31,88 %	M_T10 /11	N = 315 7,08 64,36 %
	M_T5 /19	N = 352 2,46 12,94 %	M_T11 /13	N = 326 2,91 22,38 %
	M_T6 /19	N = 347 6,34 33,37 %	M_T12 /13	N = 309 5,85 45 %
Lexique (score sur)	el_T1		el_T7 /25	N = 393 10,94 43,76 %
	el_T2 /25	N = 350 10,2 40,8 %	el_T8 /25	N = 353 19,43 77,72 %
	el_T3 /25		el_T9 /25	N = 363 10,82 43,28 %
	el_T4 /25	N = 327 17,05 68,2 %	el_T10 /25	N = 316 18,72 74,88 %
	el_T5 /25		el_T11 /25	N = 341 10,82 43,28 %
	el_T6 /25	N = 350 15,71 62,84 %	el_T12 /25	N = 307 18,84 75,36 %

En macrostructure, en pré-tests, les élèves progressent de T1 à T9 (sauf en T11). Autrement dit, avant enseignement, ils racontent un peu mieux à chaque fois au fil des deux années. En post-tests T8 et T10 de l'année 2 (MS), les élèves obtiennent des scores très supérieurs (> 63 %) à ceux qu'ils avaient obtenus en première année (PS, < 34 %).

En lexique, les résultats sont également supérieurs aux trois post-tests de l'année 2.

On observe donc une progression continue des élèves d'une année sur l'autre, avec un effet particulièrement fort en rappel.

1.2.2. Suivi des élèves en MS vers GS

Dans le chapitre précédent, nous avons recensé des indices indiquant que les élèves de NI faible progressaient, les effets moyens observés n'étant pas dus à la réussite des élèves les plus performants. Le suivi longitudinal va-t-il dans le même sens ? Comment évoluent les élèves de NI faible, de NI intermédiaire et de NI fort ? Les élèves continuent-ils de progresser d'un scénario à l'autre ? Leurs progrès sont-ils les mêmes dans les différentes épreuves (rappel et lexique) ?

Nous avons comparé les résultats des élèves aux différentes épreuves sur deux ans (MS puis GS). Nous avons ajouté un point de comparaison en nous référant à leur niveau initial au T1. Le Tableau 58 présente les résultats des élèves de T1 à T12 en macrostructure et en lexique. La colonne colorée en rose correspond aux résultats de l'ensemble des élèves. Nous présentons dans les autres colonnes les scores des élèves en fonction de leur niveau initial.

Tableau 58. Score des élèves de MS puis de GS en rappel de récit et en lexique de T1 à T12

Année 1 - 2016-2017 (MS)					Année 2 - 2017-2018 (élèves passés en GS)				
Variable (score sur)	Effectifs Moyenne % de réussite				Variable (score sur)	Effectifs Moyenne % de réussite			
	Tous les élèves	Élèves NI faible	Élèves NI intermédiaire	Élèves NI fort		Tous les élèves	Élèves NI faible	Élèves NI intermédiaire	Élèves NI fort
M_T1 /20	N= 368 4,4 22 %	N = 161 0,63 3,15 %	N = 107 4,51 22,55 %	N = 100 10,45 52,25 %	M_T7 /13	N = 387 6,37 49,0 %	N = 107 4,66 35,85 %	N = 88 7,35 56,54 %	N = 73 8,19 63 %
M_T2 /20	N = 364 8,9 44,5 %	N = 150 5,24 26,2 %	N = 95 10,31 51,55 %	N = 89 13,96 69,80 %	M_T8 /13	N = 383 10,13 77,9 %	N = 106 9,11 70,08 %	N = 90 10,48 80,61 %	N = 78 11,17 85,92 %
M_T3 /23	N = 337 1,3 5,2 %	N = 132 0,42 1,83 %	N = 94 1,47 6,39 %	N = 86 2,55 11,09 %	M_T9 /18	N = 355 7,93 44,1 %	N = 103 5,47 30,39 %	N = 84 8,23 45,72 %	N = 66 11,29 62,72 %
M_T4 /23	N = 329 4,8 20,9 %	N = 136 2,69 11,70 %	N = 89 5,42 23,57 %	N = 78 7,85 34,13 %	M_T10 /18	N = 340 13,28 73,8 %	N = 99 12,81 71,17 %	N = 81 13,21 73,39 %	N = 64 15,09 83,83 %
M_T5 /42	N = 344 8,0 19 %	N = 130 4,26 10,14 %	N = 93 9,31 22,17 %	N = 92 12,58 29,95 %	M_T11 /12	N = 341 6,53 54,4 %	N = 94 5,40 45 %	N = 78 6,91 57,58 %	N = 64 8,44 70,33 %
M_T6 /42	N = 336 17,6 41,9 %	N = 131 12,26 29,19 %	N = 90 19,39 46,17 %	N = 90 24,99 59,5 %	M_T12 /12	N = 311 8,40 70 %	N = 85 7,22 60,17 %	N = 73 8,59 71,58 %	N = 64 9,80 81,67 %
el_T1	N = 385 10,53 42,12 %	N = 125 5,48 21,92 %	N = 155 10,91 43,64 %	N = 105 15,99 63,96 %	el_T7 /25	N = 397 15,17 60,68 %	N = 90 12,12 48,48 %	N = 118 16,03 64,12 %	N = 80 19,00 76,0 %
el_T2 /25	N = 364 15,37 61,48 %	N = 115 10,51 42,04 %	N = 135 15,96 63,84 %	N = 98 20,41 81,64 %	el_T8 /25	N = 390 21,45 85,80 %	N = 88 18,92 51,68 %	N = 122 22,42 89,68 %	N = 80 23,73 94,92 %
el_T3 /25	N = 380 6,51 26,04 %	N = 114 3,90 15,60 %	N = 149 6,29 25,16 %	N = 100 9,63 38,52 %	el_T9 /25	N = 378 14,27 57,08 %	N = 85 11,60 46,40 %	N = 114 14,77 59,08 %	N = 78 17,85 71,44 %
el_T4 /25	N = 345 13,83 55,32 %	N = 106 8,92 35,68 %	N = 133 14,18 56,72 %	N = 94 18,84 75,36 %	el_T10 /25	N = 355 21,02 84,08 %	N = 79 18,73 74,92 %	N = 111 21,50 86,0 %	N = 71 23,30 93,20 %
el_T5 /25	N = 371 10,94 43,76 %	N = 112 7,24 28,96 %	N = 142 11,11 44,4 %	N = 99 15,07 60,28 %	el_T11 /25	N = 370 13,91 55,64 %	N = 77 10,64 42,56 %	N = 114 14,24 56,56 %	N = 79 17,76 71,04 %
el_T6 /25	N = 355 16,20 64,80 %	N = 106 12,84 51,36 %	N = 141 16,37 65,48 %	N = 93 19,94 79,76 %	el_T12 /25	N = 349 20,81 83,24 %	N = 78 18,49 73,96 %	N = 113 21,44 85,76 %	N = 71 23,37 93,48 %

Que nous apprend ce tableau sur les performances des élèves les plus faibles en macrostructure ? Pour l'essentiel, d'un pré-test à l'autre, les élèves progressent puisque les scores au pré-test suivant sont presque toujours supérieurs à ceux du précédent (M_T3 = 1,83 % ; M_T5 = 10,14 % ; M_T7 = 35,85 % ; M_T9 = 30,39 % et M_T11 = 45 %).

Fait remarquable, après enseignement, ces élèves ne se contentent pas seulement de progresser mais les apprentissages effectués leur permettent d'atteindre des seuils de réussite très satisfaisants. À partir du T8, ils racontent plus de 60 % de la macrostructure et ce, trois fois de suite (T8, T10, T12). On peut raisonnablement conclure qu'ils restituent suffisamment bien les idées essentielles et qu'ils ont effectué des progrès conséquents sur des albums de littérature de jeunesse de leur âge. Petit à petit, ils atteignent un niveau qui se rapproche de celui des élèves de NI intermédiaire (M_12 = 60,17 % vs 71,58 %).

En lexique, les résultats obtenus par l'ensemble des élèves après enseignement en année 2 sont élevés (> 83 %), presque identiques chaque trimestre (entre 83 % et 86 % de réussite) et nettement supérieurs à ceux obtenus en année 1. On peut ici identifier un effet plafond : en Grande Section, les enfants se souviennent très bien des mots. Est-ce dû à leur maturité physiologique ? Ont-ils appris à faire plus attention aux sens des mots grâce à *Narramus* ? Est-ce lié à l'habileté développée par leur maître dans l'enseignement ? Il est impossible de statuer.

On peut toutefois noter que la progression des élèves de NI faible est continue sur les deux années. En effet, en post-tests, ces élèves obtiennent des résultats quasiment toujours supérieurs d'un album à l'autre. Les scores obtenus par les élèves des deux autres sous-groupes sont toujours supérieurs mais l'écart entre les trois groupes se réduit.

Conclusion

Si, au début de l'année 1, les élèves de NI faible obtenaient des scores vraiment très bas, on observe qu'au fil des scénarios, ils ont beaucoup progressé. Cette évolution s'exerce certes sous l'effet cumulé de leur développement langagier, de leur appropriation de la démarche proposée dans les scénarios, de la maîtrise de l'outil par leur enseignant..., mais le résultat n'en reste pas moins positif. Aux trimestres 2 et 3 de GS, les plus faibles parviennent à mémoriser 75 % des mots étudiés et, en rappel, ils restituent plus de 70% des macrostructures aux T8 et T10.

Un autre résultat nous paraît particulièrement intéressant en lexique : lors de la deuxième année, les élèves de NI intermédiaire ou fort ont des résultats très proches en post-

tests, quel que soit l'album étudié. Pour les élèves de niveau intermédiaire, les pourcentages d'items réussis sont compris entre 85,76 % et 89,68 % ; pour les élèves de NI fort, ils se situent entre 93,20 % et 94,92 %. La progression est différente pour les élèves de NI faible : on voit un saut entre les résultats obtenus sur le lexique de l'album 1 (51,68 %) et ceux des deux autres albums (74,92 % et 73,96 %). On peut donc en conclure que les objectifs sont atteints avec les plus faibles.

2. Résultats année 3

Dans la logique précédente des indices d'un effet de cumul, nous avons analysé les résultats des élèves sur l'année 3, en limitant notre présentation aux résultats de l'album étudié au deuxième trimestre. En effet, pour répondre à la faiblesse méthodologique de l'année 2 qui compare des taux de réussite sur des épreuves dissemblables, nous avons proposé, en année 3, les mêmes tests sur un album déjà étudié la première année (*Gruffalo*). Les comparaisons s'effectuent toujours sur des élèves différents mais de même niveau scolaire (GS) avec un texte et des épreuves identiques.

2.1. Un même test sur un même album en 2017 et 2019

Lors de la première année de l'étude, les élèves de GS (G1 et G2) avaient travaillé cette histoire au second trimestre (janvier 2017) : c'était le deuxième scénario mis en œuvre dans la classe par l'enseignant. Il a été de nouveau proposé aux enseignants au second trimestre de la troisième année d'étude (janvier 2019) : il s'agissait donc du huitième scénario pour les élèves qui avaient déjà bénéficié de *Narramus* en PS et en MS.

Notre hypothèse porte sur les effets de l'enseignement après plus de deux années de travail avec *Narramus* : si l'effet de cumul existe (comme le laisse supposer les indices obtenus en pré-tests en année 2), les résultats obtenus en année 3 devraient être supérieurs à ceux obtenus en année 1.

Pour tester cette hypothèse, nous nous sommes assurée de proposer :

- 1° le même scénario (le guide pédagogique n'a pas été modifié, l'ordre et la succession des tâches proposées aux élèves sont identiques) à deux années d'intervalle,
- 2° à la même époque de l'année de GS, c'est-à-dire au second trimestre (en janvier 2017 puis en janvier 2019),
- 3° les mêmes épreuves (dans le contenu et dans l'ordre de passation)
- 4° les mêmes conditions de correction (les évaluateurs sont majoritairement restés les mêmes entre les deux années).

2.2. Les résultats

Comme pour les autres scénarios, trois épreuves ont été administrées aux élèves en pré- et en post-tests : lexique, rappel et questionnaire (cf. annexes 2, 3 et 5). Nous avons également contrôlé les variables fidélité et temps.

2.2.1. Variables temps et fidélité

Nous avons comparé la fidélité déclarée par les enseignants et la durée de l'enseignement entre les deux années. Le fait d'utiliser le même scénario une deuxième fois⁶⁹ modifie-t-il un de ces deux paramètres ? On pouvait, par exemple, supposer que les maîtres maîtriseraient davantage l'outil et s'en détacheraient en délaissant certaines tâches et en proposant les leurs.

Les données sont indiquées dans le Tableau 59.

Tableau 59. Données descriptives pour les variables Fidélité et Temps du scénario *Gruffalo* GS

	Fidélité (en %)		Temps (en minutes)	
	Effectif Moyenne (Écart-type)	T Test p =	Effectif Moyenne (Écart-type)	T Test p =
Janvier 2017 (année 1)	N = 53 87 (10,6)	NS 0.8824	N = 52 514 (176)	NS 0,300
Janvier 2019 (année 3)	N = 32 88 (13)		N = 32 560 (223)	

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

La fidélité à la mise en œuvre déclarée par les enseignants est identique entre la première et la troisième année. Autrement dit, après plus de deux ans d'utilisation et un enseignement de sept scénarios, les maîtres continuent à être très fidèles aux prescriptions des auteurs. On notera que la différence de durée n'est pas non plus significative : le fait de renouveler l'usage d'un scénario ne modifie pas le temps consacré à l'enseignement.

La comparaison à laquelle nous allons procéder ci-dessous n'en est que plus valide.

⁶⁹ La majorité des enseignants de GS en 2017 ont conservé le même niveau en 2019.

2.2.2. Comparaison des résultats

Pour répondre à notre question de recherche, nous avons conservé les résultats des élèves de GS du G1 et G2 de l'année 1 (janvier 2017) à l'épreuve *Gruffalo* et les avons comparés (en janvier 2019) avec ceux des élèves qui avaient débuté l'étude en 2016, scolarisés en Petite Section. Rappelons que ces derniers, dorénavant en GS, ont, au moment des passations, bénéficié d'un enseignement avec *Narramus* en PS (2016-2017 : 3 scénarios), en MS (2017-2018 : 3 scénarios) et en GS (1 scénario).

Les résultats sont présentés dans le Tableau 60. La partie gauche du tableau compare les scores des pré-tests, la droite celle des post-tests.

Tableau 60. Comparaison des résultats des élèves de Grande Section aux épreuves du scénario Gruffalo (année 1 vs année 3)

	Pré-tests Gruffalo			Post-tests Gruffalo		
	T3 Janvier 2017	T15 Janvier 2019	Test statistique	T4 Février 2017	T16 Février 2019	Test statistique
Niveau	Effectifs Moyenne (Écart-type) Min-max		T test p	Effectifs Moyenne (Écart-type) Min-max		T test p
Macro /30	N = 318 11,1 (8,5) 0-29	N = 216 14,0 (9,5) 0-30	0,003 **	N = 303 22,2 (8,4) 0-30	N = 205 23,0 (7,5) 0-30	0,2818 NS
Micro /82	N = 318 18,3 (15,1) 0-65	N = 216 24,5 (18,0) 0-64	< 0,001 ***	N = 303 48,5 (23,0) 0-82	N = 204 49,8 (20,2) 0-81	0,5185 NS
Questions /5	N = 333 2,5 (1,6) 0-5	N = 224 2,7 (1,7) 0-5	0,1613 NS	N = 332 3,8 (1,4) 0-5	N = 216 3,9 (1,3) 0-5	0,7694 NS
Lexique /25	N = 343 13,8 (4,9) 2-25	N = 246 15,0 (4,7) 2-25	0,003 **	N = 333 20,9 (4,0) 0-25	N = 218 21,2 (4,0) 0-25	0,3850 NS

Légende : * p<0,05 ; ** p<0,01 ; ***p<0,001

Aucune différence significative n'apparaît en post-test (T4 vs T16). Il faut noter que les scores moyens de la première année étaient déjà élevés. Bref, notre hypothèse ne se vérifie pas : il n'y a pas d'effet de cumul si l'on examine les performances finales des élèves.

En revanche, cet effet apparaît si l'on examine les données des pré-tests. Au T15, les performances en rappel (macro et micro) et lexique sont significativement supérieures à celles

observées au T3. Autrement dit, avant enseignement, les élèves obtiennent de meilleurs résultats après avoir travaillé 7 scénarios *Narramus* qu'après en avoir étudié un seul.

Au fil des scénarios, ils ont probablement développé des compétences qu'ils mobilisent spontanément pour mieux comprendre un texte nouveau (lu et raconté). En moyenne, les élèves testés en T15, restituent 3 macro et 6 micro-propositions de plus que ceux testés en T3, ils connaissent entre 1 et 2 mots de plus.

Ce constat n'est pas valable pour l'épreuve de questionnaire, ce qui était prévisible au regard de la nature des questions posées. Cette épreuve comporte des questions inférentielles nécessitant le guidage du maître, une prise d'indices complexe et des connaissances du monde que les élèves ne maîtrisent pas nécessairement avant enseignement.

2.2.3. Conclusion

Les résultats rapportés ci-dessus nous autorisent à faire l'hypothèse qu'un enseignement continu sur plusieurs années (PS à GS) avec *Narramus* participe au développement de compétences que les élèves sont capables de mobiliser, sans aide, au service de la compréhension et de la narration d'un texte inconnu.

Toutefois, notre analyse souffre de limites méthodologiques puisque notre dispositif expérimental ne nous a pas permis de faire des comparaisons fiables.

Aussi avons-nous mené une seconde étude, basée sur un protocole expérimental rigoureux, pour tester cette hypothèse d'un transfert des compétences. Elle sera présentée dans le chapitre 6.

3. Discussion des résultats de la première étude

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 supposons que les progrès de ceux qui en bénéficieraient seraient supérieurs à ceux qui en seraient privés (hypothèse 1 : l'effet *Narramus*). Nous présumons 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).

Nous avons également mené une analyse exploratoire sur un éventuel effet de cumul. Les résultats obtenus en deuxième année par les élèves les plus faibles plaident en faveur d'un usage prolongé de l'outil sur plusieurs années d'école maternelle. En outre, de manière

générale, les élèves paraissent avoir développé des compétences qu'ils sont capables de mobiliser spontanément comme l'attestent les résultats du pré-test 15 sur l'album *Gruffalo*.

Compte tenu des limites méthodologiques, ces derniers résultats ont une portée relativement réduite mais suggèrent des pistes à suivre pour de futures recherches. Il serait, en effet, pertinent d'observer les effets sur plusieurs années en contrôlant notamment le niveau initial des élèves et en proposant les mêmes albums.

3.1. L'effet *Narramus*

Nos résultats montrent que, dès la première année de l'expérimentation, l'effet est positif sur les trois dimensions évaluées (la qualité du rappel, les réponses au questionnaire et la mémorisation du lexique) en PS, 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 activités et de modalités pédagogiques) qui se combinent pour favoriser les apprentissages. Cet effet est plus fort en GS qu'en MS et PS.

On observe enfin que les élèves des trois groupes de GS, qui avaient le même niveau langagier 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, ceux qui ont bénéficié de l'outil *Narramus* restituent davantage et racontent mieux que les autres avant l'étude de l'album : il y a eu transfert d'apprentissage. Cet effet est confirmé par les résultats des élèves de Grande section au pré-test T15 de la troisième année.

Pourquoi observe-t-on de tels effets chez les élèves, dans tous les niveaux et dans les trois épreuves (rappel, questionnaire et lexique) ? Pour expliquer ces progrès, il faut revenir sur les caractéristiques de *Narramus* présentées dans le chapitre 2.

3.1.1. Un enseignement spécifique du lexique

Les résultats aux épreuves lexicales montrent que, dans les trois niveaux (PS, MS et MS), les mots et expressions présents dans les albums étudiés sont davantage mémorisés par les élèves de G1 et G2 que par ceux du G3.

On peut penser que les activités proposées pour apprendre aux élèves à encoder les mots et les expressions lors de la phase d'enseignement et que la révision systématique ont permis une mémorisation plus efficace et plus solide. L'enseignement proposé intègre des modalités pédagogiques et didactiques qui ont fait la preuve de leur efficacité dans de nombreuses recherches. Celles-ci nous aident à expliquer l'origine de nos résultats.

Du côté des pratiques centrées sur l'encodage

Nous pouvons d'abord attribuer ces effets aux consignes données aux enseignants qui utilisent *Narramus* quand ils ont à faire découvrir un nouveau mot ou d'une nouvelle expression. On leur enjoint, en effet, de multiplier les liens sémantiques. Dans le scénario 3⁷⁰ destiné aux élèves de PS, les auteurs font, par exemple, découvrir le mot poussin de manière très explicite : « le poussin est le petit de la poule et du coq ; quand il sera grand, il sera une poule ou un coq ; le poussin est jaune comme... mais pas comme... ; le poussin a des plumes, comme... mais pas comme... ; le poussin a un bec comme... mais pas comme... ; il vit dans une basse-cour ; il mange des graines... ». Les enseignants sont invités à procéder de cette manière pour chaque nouveau mot ou nouvelle expression.

Les résultats obtenus aux post-tests lexicaux immédiats et différés corroborent les résultats d'études antérieures qui ont prouvé que l'accès aux connaissances stockées en mémoire est étroitement lié à la nature et à la qualité de leur encodage, c'est-à-dire la manière dont elles ont été enseignées, mémorisées et stockées (Beck, McKeown, & Kucan, 2013; Lane & Allen, 2010; Nagy, Herman, & Anderson, 1985). On a montré qu'on ne peut avoir accès à une information en mémoire qu'à partir des indices ou des liens sémantiques utilisés lors de son encodage (Cellier, 2017; Chi & Koeske, 1983; Chi, Hutchinson, & Robin, 1989).

Du côté des pratiques dédiées à l'enseignement, la révision et à l'entraînement

Dans tous les scénarios, les maîtres font découvrir puis réviser les mots pendant quatre semaines. Les élèves participent quotidiennement à des activités leur permettant d'utiliser le lexique nouvellement appris dans différentes situations (jeux de cartes, jeux de loto, tâches de catégorisation, jeux de dénomination rapide). Le vocabulaire est aussi systématiquement mimé en classe ou en salle de jeux.

L'intensité de l'enseignement et les rétroactions nombreuses donnent aux élèves de nombreuses occasions d'apprendre. Or, il est avéré que ces stratégies pédagogiques appuient

⁷⁰ Album *Un poussin de mauvais poil*.

l'apprentissage des mots, particulièrement chez les élèves en difficulté (Loftus, Coyne, McCoach, Zipoli, & Pullen, 2010). Ceci permet d'expliquer, au moins en partie, l'origine des progrès lexicaux des élèves de niveau initial faible observés dans notre étude.

L'enseignement du lexique dans *Narramus* repose, de surcroît, sur des supports visuels et parfois auditifs variés (vidéos, gifs animés, images fixes, enregistrements sonores pour le cri des animaux, par exemple). Si l'on en croit Wasik, Hindman, et Snell (2016) ainsi que Zipoli, Coyne, et McCoach (2011) ; ces manières de faire participent à rendre les apprentissages plus solides.

À la fin de la séance de vocabulaire, les auteurs de l'outil proposent à tous les élèves de redire les mots nouveaux à haute voix à partir des images vidéoprojetées. Cette activité participe sans doute à l'amélioration de la mémorisation à long terme et de la compréhension si l'on se réfère à l'étude menée par Rosenthal et Ehri (2011).

En MS et GS, le post-test différé de lexique montre que le maintien est meilleur chez les élèves de G1 et G2. Le fait de réviser le vocabulaire, de l'utiliser dans différents contextes et de proposer des entraînements réguliers et divers (Beck et al., 2013) a probablement participé à assurer le maintien du vocabulaire en mémoire (Pullen et al., 2010). On a également montré que les progrès des élèves sont particulièrement sensibles aux activités de définitions explicites, d'intégration des mots dans des contextes significatifs, d'entraînements continus et de suivis réguliers des progrès (Beck & McKeown, 2007; Marulis & Neuman, 2010). Ces différentes activités se retrouvent dans *Narramus* et nous supposons donc qu'elles participent, elles aussi, aux effets obtenus.

Enfin, les auteurs de *Narramus* font une proposition nouvelle, peu mise en œuvre dans les classes maternelles⁷¹ : l'enseignement du vocabulaire avant la découverte du texte de l'histoire.

Cette stratégie pédagogique vise à offrir un contexte sur lequel les élèves peuvent s'appuyer pour construire une première représentation mentale du texte. On sait, en effet, que plus le lecteur (ou l'auditeur) connaît le sens des mots d'un texte, mieux il comprend (Ouellette, 2006; Stahl & Nagy, 2007).

La capacité accrue pour récupérer un mot en mémoire peut, en outre, accélérer le processus de compréhension (Perfetti, 2007). Ce lien entre connaissance du vocabulaire et

⁷¹ Nous effectuons ce constat suite 1° à notre expérience de professeure des écoles et de formatrice et 2° à l'analyse des retours des questionnaires enseignants. Nous y reviendrons dans le chapitre 7.

compréhension en lecture peut aussi expliquer la relation entre le vocabulaire et la compréhension auditive (DeBruin-Parecki, Van Kleeck, & Gear, 2015). Les élèves comprennent les mots connus mais peuvent aussi tirer profit de leur compréhension du nouveau vocabulaire pour déduire le sens de nouveaux mots entendus conjointement avec d'autres mots qu'ils connaissent. Un des effets plus distancés de cet apprentissage cumulatif est donc que les élèves élargissent leur vocabulaire au-delà des seuls mots enseignés ce qui a eu sans doute une influence sur le vocabulaire en général et la compréhension.

Le type d'enseignement intensif et prolongé proposé dans *Narramus* semble avoir aidé les élèves à développer une curiosité pour les mots (Beck, McKeown, & Kucan, 2013; Coyne, Neugebauer, Ware, McCoach, & Madura, 2015) ce qui pourrait expliquer les résultats observés en lexique au T15. Au pré-test du 8^e scénario, les élèves de GS expliquent en moyenne 1 à 2 mots de plus qu'au pré-test du 2^e scénario (T3).

Nos résultats rejoignent ceux déjà établis par Penno, Wilkinson et Moore (2002) et Justice, Meier et Walpole (2005) : un enseignement direct favorise l'apprentissage des mots nouveaux rencontrés dans les histoires. Dans l'étude conduite par Justice *et al.* (2005), cinquante-sept élèves de maternelle ont été exposés à soixante mots cibles extraits d'une histoire. Le livre a été lu à tous les élèves une première fois et trente mots ont été expliqués. L'enseignant a donné le sens du mot et a utilisé ce dernier dans une phrase au cours de la lecture. Les trente autres mots n'ont pas été l'objet d'un échange et l'enseignant les a simplement prononcés lors de la lecture. Les résultats montrent que la seconde condition est plus efficace que la première.

L'enseignement direct de la signification des mots s'avère, de manière évidente, plus efficace que l'exposition accidentelle sans explication.

Une explication rapide du sens des mots lors de la lecture du texte ne semble pas davantage suffire pour une mémorisation efficace. En effet, on observe dans les carnets de bord des enseignants du G3 qu'ils sont nombreux à déclarer expliquer les mots complexes ou inconnus lors de la découverte du texte. Les résultats obtenus par leurs élèves montrent que cet enseignement est moins efficace. Ce constat n'est pas nouveau : « un mot expliqué de façon hâtive, donné une fois ou deux seulement, non travaillé spécifiquement ne sera pas mémorisé, ni utilisé. Il faut donc prévoir quelques dispositifs pour son appropriation et sa réactivation, pendant ou après l'exploitation de l'album. Il faut aménager les lectures pour les rendre vraiment productives et pour faire de ces rencontres incidentes avec les mots des moments d'acquisition du vocabulaire » (Cellier, 2017, p. 34).

Nous avons vu, dans le chapitre 2, que les mots pouvaient être appris de manière incidente mais que tous les enfants n'étaient pas égaux. Les résultats des élèves ayant bénéficié de *Narramus* paraissent d'autant plus forts qu'on les met en relation avec les conclusions établies par d'autres chercheurs : le vocabulaire est une habileté plutôt résistante aux dispositifs scolaires censés le favoriser (Apthorp et al., 2012; Elleman, Lindo, Morphy, & Compton, 2009; Zorman et al., 2015). De ce point de vue nos résultats sont majeurs s'ils incitent les enseignants de maternelle REP à intégrer ces pratiques efficaces dans leur classe.

3.1.2. Un enseignement innovant des compétences narratives

Outre un enseignement direct et spécifique du vocabulaire, le rappel des histoires permet d'acquérir une connaissance plus approfondie des mots (Karweit, 1989; Leung, 2008) et une meilleure compréhension de l'histoire (Newman, Dickinson, Hirsh-Pasek, & Michnick-Golinkoff, 2015). C'est pourquoi, dans chaque opus de *Narramus*, les auteurs invitent les enseignants à apprendre aux élèves à raconter. Ces derniers passent alors d'interactions collectives à des conduites langagières individuelles, étayées par le professeur, puis progressivement autonomes.

Dans tous les scénarios, le rappel de récit donne un but intégrateur aux activités mises en œuvre et finalise la tâche d'écoute de la lecture oralisée. « Comprendre un texte, c'est s'en faire une représentation mentale cohérente en combinant les informations explicites, implicites qu'il contient à ses propres connaissances. Cette représentation est dynamique et cyclique. Elle se transforme et se complexifie au fur et à mesure de la lecture. Chaque fois que le lecteur rencontre un nouvel élément dans le texte, il doit décider de quelle façon l'intégrer à sa représentation du texte. » (Giasson, 2012, p. 227). Pour pouvoir raconter, sans aide, l'histoire étudiée en classe, les élèves doivent « prendre ensemble » et réunir toutes les idées du texte. Le Avec l'activité de rappel de récit, 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.). Elle permet également de centrer leur attention sur le lien qui unit compréhension et mémorisation, et sur l'effort qu'il convient de faire pour ne pas oublier de relater des informations essentielles » (Cèbe & Goigoux, 2017, p. 13). Cette activité accompagnée par les enseignants a probablement participé à coconstruire le sens du texte (Bishop, Boiron, Schmehl-Postaï, & Royer, 2017).

Dans leurs méta-analyse, Swanson et ses collaboratrices (2011) ont recensé trois études de lecture dialogique dont l'objectif était de permettre aux enfants d'acquérir une compréhension approfondie de l'histoire au moyen de plusieurs stratégies, dont la narration. Les preuves les plus solides proviennent des interventions de lecture dialogique intègrent des interactions prolongées entre enfants et adultes et un questionnement autour des livres d'histoires. La méta-analyse ne présente pas de manière détaillée les dispositifs mis en œuvre mais on peut supposer que ces dialogues sont plus intensifs et laissent davantage la parole aux élèves que ce qui est d'usage dans les classes françaises. Dans le rapport des inspecteurs généraux de 2011, on peut lire que « l'activité collective est très encadrée par les questions du maître – souvent de plus en plus fermées – auxquelles les élèves apportent des réponses brèves : l'enseignant, par sa façon de guider, amène les enfants à ordonner un discours collectif dont lui seul a conscience, chaque enfant étant engagé pour une bribe seulement. Pour le maître, l'objectif visé est bien conforme au programme mais le profit pour les enfants est faible du fait de cette pédagogie invisible qui ne dit pas ce qu'elle cherche à obtenir, ce qu'elle a obtenu et pourquoi c'est satisfaisant ou cela ne l'est pas. » (p.128).

En combinant un apprentissage de la narration et des questionnements nombreux sur les passages implicites du texte, sur les émotions et les intentions des personnages..., l'usage de *Narramus* a permis aux élèves de construire des compétences solides en compréhension. Aux post-tests rappel et questionnaire, les résultats des élèves du G1 et du G2 sont systématiquement supérieurs à ceux du G3 dans les trois niveaux et les différences sont presque toujours significatives en MS et en GS.

On peut supposer, en outre, que l'attention portée à l'implicite du texte et notamment à la théorie de l'esprit participant au développement des compétences de compréhension. Rappelons que, dans *Narramus*, les enseignants amènent régulièrement les élèves à s'interroger sur les états mentaux des personnages. Ils leur demandent, par exemple, de se mettre provisoirement à la place d'un protagoniste pour imaginer ce qu'il pense, ressent, croit... Ils les invitent à puiser dans leurs expériences personnelles et dans leurs connaissances sur la psychologie humaine pour interpréter les états mentaux des différents personnages. À travers ces interactions, les enseignants aident les élèves à argumenter et raisonner collectivement mais aussi à développer une théorie de l'esprit. Bautier (2018) réaffirme la nécessité de « faire référence aux travaux théoriques et empiriques qui soulignent l'importance de travailler sur ce que les élèves pensent de ce que les gens pensent, des intentions des textes et des énoncés, domaine d'acquisition très socialement différenciateurs des élèves ». En effet, tous les enfants

ne mettent pas spontanément en œuvre ce type de raisonnement. Leur apprendre à produire des inférences et à s'interroger sur les états mentaux des personnages semble crucial si on veut réduire les écarts entre les élèves de milieux sociaux contrastés. Kim (2016) et Newman *et al.*(2015) ont montré que, quand on le fait, cela influe de façon importante sur la qualité de la compréhension.

Toutefois, il ne suffit pas d'avoir mémorisé et organisé la suite des événements ni d'en avoir construit une représentation mentale cohérente pour savoir raconter. L'activité de narration produite sous forme de monologue est une activité complexe qui nécessite la maîtrise du langage d'évocation et la mobilisation simultanée de plusieurs compétences : « Au niveau de la planification et de la structuration, le récit doit constituer un tout présentant un minimum d'intelligibilité thématique et généralement une certaine cohérence d'ensemble par laquelle les événements sont expliqués et prennent en cela leur signification. Au niveau de la mise en mots, les récits doivent présenter une certaine cohésion discursive qui laisse transparaître la cohérence d'ensemble » (Veneziano, 2010, p. 4).

La narration monogérée se développe ainsi sur plusieurs années scolaires, les élèves étant, en outre, différemment familiarisés avec ce type de discours avant leur entrée à l'école maternelle. Nous attribuons les différences de qualité de rappel observées entre les élèves de Petite Section et ceux de Moyenne et Grande Section aux complexités que nous venons d'évoquer. Certains élèves, en sus de l'inquiétude provoquée par l'épreuve (raconter seul, sans support à un adulte qui ne fait pas partie de notre environnement quotidien), ont besoin de temps pour apprendre à raconter. Les résultats obtenus par les élèves en pré-test T15 nous incitent à penser qu'un enseignement continu sur toutes les années d'école maternelle est particulièrement bénéfique et améliore la qualité des rappels et de la compréhension.

3.1.3. Une mise en scène des histoires

Pour faciliter la compréhension et le rappel de récit, les auteurs de *Narramus* proposent aux enseignants de faire jouer les histoires étudiées. Une maquette et des figurines sont aussi proposées aux élèves pour qu'ils puissent s'entraîner à raconter l'histoire. En fin de scénario, les enfants rentrent chez eux avec une reproduction miniature de ce matériel pour raconter l'histoire.

« Plusieurs auteurs [...] ont prouvé que l'implication du corps dans la mise en scène du texte améliore la qualité de la compréhension. Tout ce qui – comme les marionnettes – anime


dans l'espace et le temps ce qui n'est que mots sur une page constitue donc une aide précieuse. » (Cèbe & Goigoux, 2017, p. 15).

La mise en scène a pu aider les élèves à se souvenir et à mieux comprendre l'histoire. Plusieurs études montrent l'impact de ce type d'activités sur l'utilisation du vocabulaire lié aux états mentaux des personnages (Biazak, Marley, & Levin, 2010 ; Snell, Hindman, & Wasik, 2015). Vu la nature complexe et cognitivement coûteuse de l'activité de rappel, ces auteurs suggèrent de proposer des accessoires aux élèves pour les aider à raconter. Les enquêteurs de notre académie nous ont rapporté que, pendant l'épreuve évaluative de rappel de récit, les élèves du G1 et G2 mimaient et jouaient certains passages de l'histoire. Privés de support ou d'accessoires fortement utilisés dans *Narramus*, ces élèves semblent avoir intériorisé et mémorisé des passages grâce aux mises en scènes effectuées en classe. Pendant l'enseignement, les événements de l'histoire ont été vécus par le mouvement et en relation avec les accessoires, ce qui entraîne une amélioration de ce qu'on appelle une représentation incarnée qui a des effets positifs sur la récupération en mémoire (Glenberg, 1997).

3.1.4. « Un temps d'attente » fréquemment proposé

On peut également supposer que les modalités « d'attente » régulièrement proposées aux élèves participent aux progrès. Dans *Narramus*, on invite les enseignants à laisser systématiquement aux élèves le temps de réfléchir avant de leur demander leurs réponses. On trouvera ci-dessous quelques exemples :



1° lors de la découverte du vocabulaire : différentes propositions permettent aux élèves de retrouver le mot dans leur tête et de le prononcer lors du signal de la maitresse. L'extrait issu du guide du maître *La sorcière dans les airs* (MS, 3^e album, p. 12) illustre cette modalité.

	<p>Afficher D46 – Rappeler la consigne : je vais afficher les images des expressions et les mots que vous avez appris depuis le début du travail, mais vous vous souvenez ? Vous n'avez pas le droit de dire leur nom avant que je donne le signal. Quand je le donnerai, vous devrez tous dire le mot, ensemble, le plus doucement possible, en chuchotant. Vous êtes prêts ?</p> <p>Poursuivre : laissez bien votre boîte <i>La sorcière dans les airs</i> ouverte parce que vous allez apprendre plusieurs mots nouveaux qu'il faudra y ranger aussi parce qu'ils vont vous aider à bien comprendre l'histoire.</p>
---	--


2° Lors de la prévision des images avec l'affichage du logo « point d'interrogation⁷² » : après la découverte du texte, les élèves disposent d'un temps pour imaginer l'illustration à venir puis

⁷² Le logo « point d'interrogation » invite les élèves à prévoir l'illustration.

d'un autre temps pour comparer ce qu'ils avaient imaginé et l'image de l'album. L'extrait du guide pédagogique *Le jour où loup gris est devenu bleu* (GS, 3^e scénario, année 2, p. 8) exemplifie cette proposition.

	<p>Dire : je vais lire le début de l'histoire, mais je ne vous montrerai pas l'image en même temps. Vous devez essayer de la fabriquer en transformant les mots en images, comme si vous fabriquiez un dessin animé. Ensuite, je vous montrerai l'image de l'album et vous pourrez la comparer avec la vôtre.</p>
<p>D14</p>	<p>Lire le paragraphe en montrant clairement aux élèves ce qui est lu et en attirant leur attention sur la taille des caractères qui augmente montrant visuellement que le loup a une faim de loup de plus en plus grande.</p>
<p>Loup gris avait faim, très faim... Depuis des jours et des jours, il chassait dans la forêt, rôdait sur les chemins, mais n'attrapait rien.</p>	
<p>Pas un mouton, pas un lapin.</p>	
<p>Cliquer pour faire disparaître le texte et pour afficher le pictogramme <i>l'enseignant-e raconte</i>.</p>	
<p>Reformuler le texte dans une langue adaptée aux élèves : reformuler les mots nouveaux ajouter des informations, expliciter l'explicite, mimer, jouer sur l'intonation...</p>	
<p>Cliquer pour faire disparaître le texte et afficher le pictogramme <i>point d'interrogation</i>.</p>	
<p>Inviter les élèves à prévoir l'illustration. Leur laisser quelques instants de réflexion et demander à l'un d'entre eux de dire ce qu'il s'attend à voir sur l'image qui va être affichée.</p>	
<p>Cette activité qui consiste à exprimer ce qu'on s'attend à voir sur l'illustration est une bonne manière de commencer à apprendre à raconter (sans réciter par cœur). Elle permet aussi aux élèves d'utiliser le lexique enseigné.</p>	
	<p>Afficher l'illustration, laisser aux élèves le temps de l'observer et la leur faire commenter. S'ils n'en parlent pas spontanément, attirer leur attention sur le fait que :</p>
<p>D15</p>	<ol style="list-style-type: none"> 1. contrairement à leurs attentes, l'image n'illustre pas la seconde partie du texte relatant la quête éperdue du loup les jours qui précèdent ; 2. l'image ne représente que la première phrase du texte : « <i>Loup gris avait faim, très faim</i> » (leur faire observer que le loup qui a très faim n'a pas l'air bien « méchant ») ; 3. il semble à bout de force tant il a faim, prêt, sans doute, à manger un ver de terre.

3° Lors de la prévision de la suite de l'histoire avec le logo « trois points de suspension » : les élèves sont parfois invités à anticiper la suite de l'histoire. Outillés avec un apport préalable du vocabulaire et une étude approfondie des épisodes précédents, les élèves peuvent être amenés à raconter ou à dessiner la suite de l'histoire, comme c'est le cas dans l'extrait ci-dessous pris dans le scénario *Les deniers de Compère lapin* (GS, 3^e scénario année 1, p. 28).

	<p>Expliquer : vous allez essayer de prévoir la suite de l'histoire, en dessinant ce qui va se passer après, maintenant que la poule a mangé le ver de terre. Laisser aux élèves le temps de dessiner et ramasser les dessins sans les commenter.</p>
---	--

Les élèves dessinent et le module se termine. Il faut attendre le suivant, soit le lendemain pour mettre en commun les prévisions et lire la suite du texte.

Cette stratégie « d'attente » présente à l'évidence des contraintes importantes à prendre en considération, mais il semble que les avantages potentiels pourraient l'emporter sur les inconvénients (Wasik & Hindman, 2018). Les ressources pour l'école maternelle disponibles sur Éduscol (Mobiliser le langage dans toutes ses dimensions) vont d'ailleurs dans ce sens.

L'un des principaux bénéfices attendus des temps d'attente est d'offrir aux jeunes enfants plus d'occasions de parler et d'utiliser la langue. Les données de la littérature scientifique suggèrent que 1° les élèves parlent plus quand l'enseignant leur laisse du temps pour réfléchir après qu'il a posé une question et 2° l'exactitude et la complexité de leurs réponses sont plus grandes (idem, 2018).

Le temps d'attente permet aux élèves de réfléchir avant de répondre, établissant ainsi une habitude de conversation positive. En outre, lorsque l'enseignant attend la réponse d'un élève et la lui fait compléter, il participe à améliorer le niveau de langage. Au lieu d'aller solliciter immédiatement un autre élève quand la réponse est lacunaire ou incomplète, il l'invite à poursuivre : cette stratégie pédagogique donne à n'en pas douter de la valeur au raisonnement de l'enfant et participe à la construction de compétences langagières qui vont bien au-delà du vocabulaire, comme l'atteste la recherche menée par Wasik et Hindman (2018).

3.1.5. Conclusion

Le focus porté sur l'enseignement de la narration et du lexique ainsi que les modalités particulières mises en œuvre expliquent en partie les résultats obtenus. Nous ne reviendrons pas ici sur les caractéristiques fondamentales de *Narramus* qui ont probablement également contribué à l'amélioration de l'apprentissage : le caractère explicite et ritualisé de l'enseignement, l'engagement actif des élèves et l'attention conjointe, les multiples supports utilisés...

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 : 250 classes et 1 500 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 (pré-tests 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 Réseau 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. (« modulaire » vs « intégrative »). Les auteurs de *Narramus*, plaident 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 *et al.* (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

3.2. 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. Il est important de signaler que cette absence d'effet porte sur la comparaison entre les deux groupes G1 et G2 mais cela n'exclut pas que, dans certaines écoles, cet accompagnement ait pu être influent. Notre méthodologie ne nous permet cependant pas d'identifier ces « déviants positifs » (Bryk, 2015) car trop peu d'élèves ont été évalués dans chaque classe.

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 al.* (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 provoque 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 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 & 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 est l'objet d'une analyse (chapitre 7) réalisée à partir des questionnaires que nous avons recueillis auprès des enseignants impliqués dans l'étude.

3.3. L'étude 2 : un prolongement de ces premiers résultats

Les résultats présentés portent essentiellement sur la première année de l'expérimentation d'une étude longitudinale qui en compte trois. En GS, les 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 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. Les données de la troisième année d'expérimentation tendent à confirmer cette analyse : il est raisonnable de conclure que bénéficier d'une deuxième année d'intervention puis d'une troisième est à l'origine de progrès plus importants encore.

Les résultats de la première année (confirmés ensuite par ceux de la troisième) nous avaient alertée sur le transfert des compétences. C'est pourquoi nous avons proposé une deuxième étude expérimentale présentée dans le chapitre 6.

3.4. Les limites de l'étude

Nous ne pouvons conclure sans mentionner quelques-unes des limites de cette étude pilotée par nous-même qui travaillons 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 analyses lexicales ont été effectuées sur des tests de vocabulaire passif et non sur le vocabulaire actif. Il serait intéressant d'étudier les enregistrements des élèves et de comptabiliser le nombre de mots et expressions du texte présents dans les récits enfantins. Les élèves qui obtiennent un bon score à l'épreuve lexicale réinvestissent-ils le vocabulaire dans leur rappel ? Quels sont les mots les plus employés ? A-t-on des traces des mots des histoires 1 et 2 présents dans le rappel de l'histoire 3 ? Nous pensons ici, par exemple, aux récits enfantins du troisième album étudié en GS (*Les deniers de Compère Lapin*). Lorsque le lapin part en ville « mener la belle vie », plusieurs élèves ont raconté qu'il partait festoyer. Ce verbe était présent dans l'histoire précédente *Gruffalo* étudiée au second trimestre. Certains élèves ont ainsi su l'utiliser à bon escient dans un autre contexte.

Les analyses des rappels de récit ont été effectuées à partir de deux grilles de codage (macrostructure et microstructure). Nous avons évalué la quantité d'items relatés par les élèves pour juger de la qualité des récits. Les ajouts des élèves, dont nous avons constaté qu'ils étaient

conséquents, tels que les références aux états mentaux des personnages (pensées, émotions...) n'ont pas été pris en compte.

Dans l'épreuve lexicale, l'élève devait parfois expliquer le mot demandé (5 à 8 mots sur 25). S'il ne répondait pas, l'enquêteur le relançait avec deux questions prédéfinies sur la grille de codage. Dans nos analyses, nous n'avons pas distingué si la réponse était donnée avant ou après relance. Or, lorsqu'on mémorise une information, cette dernière passe par trois phases : l'encodage, le stockage et la récupération (Cellier, 2017). La récupération permet d'accéder à l'information recherchée. Retrouver cette information peut être plus ou moins facile et peut prendre différentes formes :

- 1° si le mot est redonné de manière libre et spontanée, on parle de rappel libre ;
- 2° s'il est nécessaire de procéder par association et de s'appuyer sur des indices de récupération (évocation du contexte, du moment de découverte du mot...), le rappel est alors indicé ;
- 3° si le mot est reconnu parmi une série de termes donnés, c'est de la reconnaissance.

Nous aurions pu distinguer les scores des élèves en fonction de la nature du rappel : spontané ou indicé, voire ajouter une troisième étape et proposer une série de termes parmi lesquels les élèves pouvaient choisir.

Nous sommes aussi consciente que les épreuves proposées ne ciblent que certains aspects de la connaissance du mot. Elles n'ont pas permis de vérifier si les élèves avaient intégré les mots étudiés au point de les utiliser de manière spontanée et correcte en contexte de communication. Partageons ici un exemple communiqué aux enseignants par les parents: Lorsque pour le goûter, une mère propose un autre gâteau à son fils, celui-ci répond « Volontiers, moi les biscuits, j'en raffole ». Il faisait référence à l'album *Un peu perdu* travaillé en classe !

Les principes de traitement des données énoncés au départ et le grain d'analyse que nous avons retenu ne nous permettent pas de mesurer l'influence de l'accompagnement sur les progrès des élèves en fonction des contextes locaux. D'autres investigations seraient donc à conduire pour discuter les résultats obtenus en fonction des caractéristiques des accompagnements mis en œuvre (notamment la durée, le contenu et les modalités organisationnelles).

Ce travail ouvre ainsi de nouvelles perspectives de recherche. La mise en œuvre effective d'un scénario *Narramus* diffère-t-elle selon la nature de l'accompagnement ? Les

gestes professionnels sont-ils les mêmes en fonction de la durée de l'accompagnement ? La généralisation dépend-elle des modalités d'accompagnement proposées ?

Partie 3

Un transfert de compétences ?

Une réduction des inégalités ?

Chapitre 6

Méthode, résultats, discussion de l'étude 2

Chapitre 6

Méthode, résultats,
discussion de l'étude 2

Chapitre 6 Méthode, résultats, discussion de l'étude 2

« [...] Il est à la fois légitime et nécessaire de pouvoir apprécier la contribution de l'école maternelle à la réalisation des objectifs assignés à la scolarité primaire, et donc les acquis des élèves à la sortie de l'école maternelle, notamment du point de vue de la maîtrise du langage et de la numération. Cela doit certes être fait « selon des conditions méthodologiques prudentes et appropriées (distinctes des modalités d'évaluation des élèves à l'école élémentaire), de sorte que l'on n'institue pas, pour des enfants de cet âge, un dispositif d'« évaluation » [qui serait] mal conçu [...].

Quelle que puisse être la difficulté de l'exercice, il conviendrait de pouvoir disposer d'éléments d'évaluation de quelques acquis fondamentaux des élèves à l'école maternelle. Étant donné que les apprentissages premiers peuvent être déterminants pour la réussite scolaire, il serait également important de pouvoir apprécier la contribution de l'école maternelle à la réduction des inégalités scolaires dans le domaine de la maîtrise du langage et du vocabulaire et, par conséquent, son apport à la prévention de l'illettrisme. »

(IGEN, 2011, p. 129)

Dans le prolongement des résultats de la première étude qui montrent un effet bénéfique de *Narramus* sur la compréhension et la narration de récits enseignés en classe, on peut se demander si les compétences langagières construites sur les albums étudiés peuvent se transférer à d'autres textes entendus. En mobilisant le langage dans des activités systématisées et finalisées, les élèves ont appris à mieux raconter les textes enseignés avec *Narramus*, mais ont-ils développé des connaissances plus générales qui leur permettent de mieux comprendre et de raconter un texte nouveau ? La première étude ne nous permettant pas d'évaluer cette généralisation, nous en avons mené une seconde. Celle-ci vise à mesurer le rôle compensatoire des inégalités sociales que peut jouer l'école maternelle sur le développement des compétences langagières chez de jeunes élèves. Elle a été proposée à des élèves de fin de Grande Section de maternelle.

1. Le dispositif expérimental de l'étude 2

Pour atteindre nos objectifs, nous avons donc testé deux nouvelles hypothèses (en sus des deux déjà testées dans l'étude 1 et de celle qui sera traitée dans le chapitre 7) : une sur le transfert, une sur la réduction des inégalités.

L'hypothèse 4 stipule que, dans nos groupes expérimentaux, les compétences construites en Moyenne et en Grande Section de maternelle se transfèrent sur d'autres récits que ceux étudiés en classe. Pour tester cette hypothèse, nous avons choisi un album nouveau et nous avons comparé les performances langagières des élèves ayant bénéficié d'un enseignement explicite de la compréhension avec *Narramus* (groupe A) avec celles d'autres élèves de REP ne l'ayant pas utilisé (groupe B).

L'hypothèse 5 stipule qu'un usage prolongé de *Narramus* (six modules de quatre semaines chacun, répartis sur deux années consécutives) permet de réduire les écarts entre élèves de milieux sociaux contrastés. Pour tester cette hypothèse, nous avons comparé les performances langagières des élèves de REP ayant travaillé pendant deux ans avec *Narramus* (groupe A) avec celles d'élèves issus de milieu favorisé dont les maîtres n'utilisaient pas *Narramus* (groupe C).

1.1. Composition de l'échantillon

Pour appréhender la question de la réduction des inégalités, nous avons recherché des terrains contrastés tant socialement que scolairement : les uns situés en Réseau d'Éducation Prioritaire, les autres en secteur urbain, accueillant un public socialement favorisé.

1.1.1. Les groupes

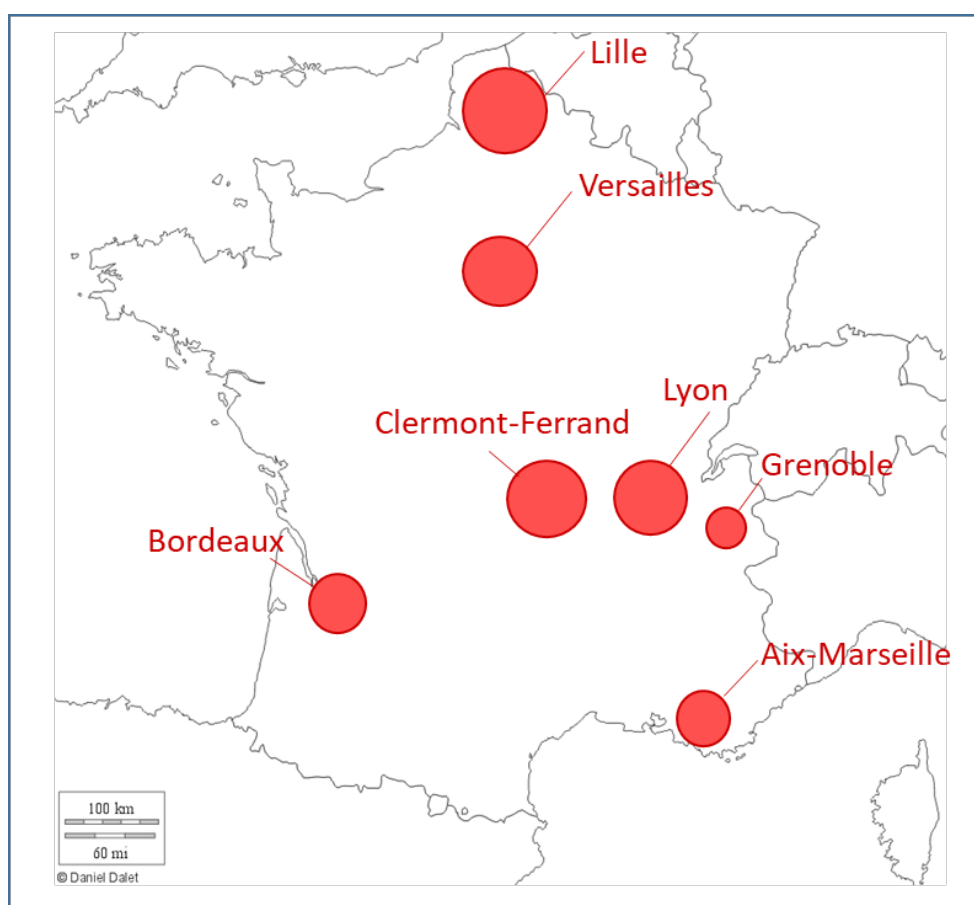
Pour mettre à l'épreuve nos deux hypothèses, nous avons choisi de comparer trois groupes d'élèves provenant de sept académies (Aix-Marseille, Bordeaux, Clermont-Ferrand, Grenoble, Lille, Lyon, Versailles). Le premier (groupe A) rassemble les élèves qui ont suivi deux années d'enseignement basées sur *Narramus*. Ils sont scolarisés dans une école classée en REP ou REP+ et sont majoritairement issus de milieux défavorisés : c'est notre groupe expérimental. Nous l'avons comparé à deux groupes dont les enseignants n'utilisent pas *Narramus* mais qui se distinguent par l'origine sociale des élèves : les uns (groupe B) fréquentent, comme ceux du groupe A, une école classée en REP ou REP+, les autres (groupe C) sont issus d'un milieu

social favorisé⁷³ (cf. tableau 61). Les élèves des trois groupes ont bénéficié d'au moins trois ans d'école maternelle en France.

Nous avons comparé les résultats du groupe A avec ceux du groupe B pour tester l'hypothèse 4 puis avec ceux du groupe C pour l'hypothèse 5.

La Figure 18 présente la répartition géographique des académies avec lesquelles nous avons travaillé. La variété des territoires impliqués dans l'étude a favorisé la diversité des publics de l'échantillon et l'autonomie professionnelle des enseignants fort éloignés des chercheurs.

Figure 18. Répartition géographique des académies engagées dans l'étude 2



1.1.2. Les élèves de Grande section

L'échantillon est composé de 522 élèves (46 % de filles et 54 % de garçons) de grande section de maternelle. On trouve, dans le Tableau 61, le nombre d'élèves qui constituent les groupes.

⁷³ Toutes les classes du groupe C sont situées à Chamalières, ville qui se trouve dans la banlieue favorisée de Clermont-Ferrand.

Tableau 61. Nombre d'élèves dans chaque groupe

Groupe	Nombre d'élèves testés
A. REP avec <i>Narramus</i>	160
B. REP sans <i>Narramus</i>	205
C. Milieu favorisé sans <i>Narramus</i>	157
Total	522

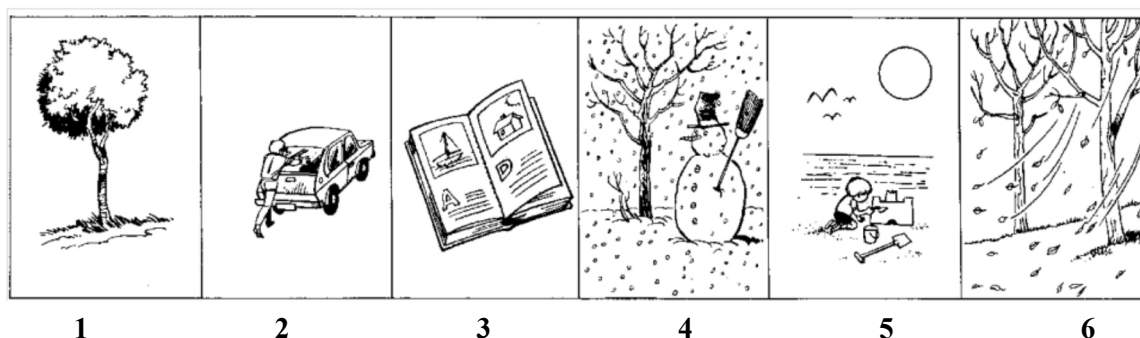
1.2. Constitution des trois groupes

Nous avons d'abord vérifié la constitution de notre échantillon pour savoir si les compétences des groupes A et B étaient équivalentes et si celles du groupe C correspondaient bien à celles qu'obtiennent classiquement des populations favorisées dans les tests standardisés.

Pour cela, nous avons comparé les résultats des trois groupes à l'épreuve TVAP⁷⁴ (Deltour & Hupkens, 1980), un test évaluant une compétence clé et socialement très marquée : le lexique. Nous avons utilisé la version dite « passive » du TVAP dans laquelle l'élève doit choisir, parmi six images, celle qui correspond au mot prononcé par l'évaluateur.

L'épreuve comporte trente mots notés sur une échelle de 0 à 2 (2 points pour une réponse correcte, 1 point pour une réponse approchante). Le score maximal est de 60 points.

La Figure 19 fournit un exemple des images proposées pour le mot « hiver ».

Figure 19. Extrait du TVAP

Le Tableau 62 recense les résultats obtenus par chacun des groupes.

⁷⁴ Test de vocabulaire actif et passif pour enfants de 5 à 8 ans (TVAP 5-8)

Tableau 62. Résultats des élèves des trois groupes au TVAP

	Groupe A REP avec <i>Narramus</i>	Groupe B REP sans <i>Narramus</i>	Groupe C Milieu favorisé sans <i>Narramus</i>	Test		
	N= 160	N= 205	N= 157	Tests post Anova (p =)		
	Moyenne (Écart-type) Min-max			GA vs GB	GA vs GC	GB vs GC
TVAP	35,3 (7,7) 16-55	34,1 (7,5) 17-52	42,3 (5,4) 27-53	0,284 NS	***	***

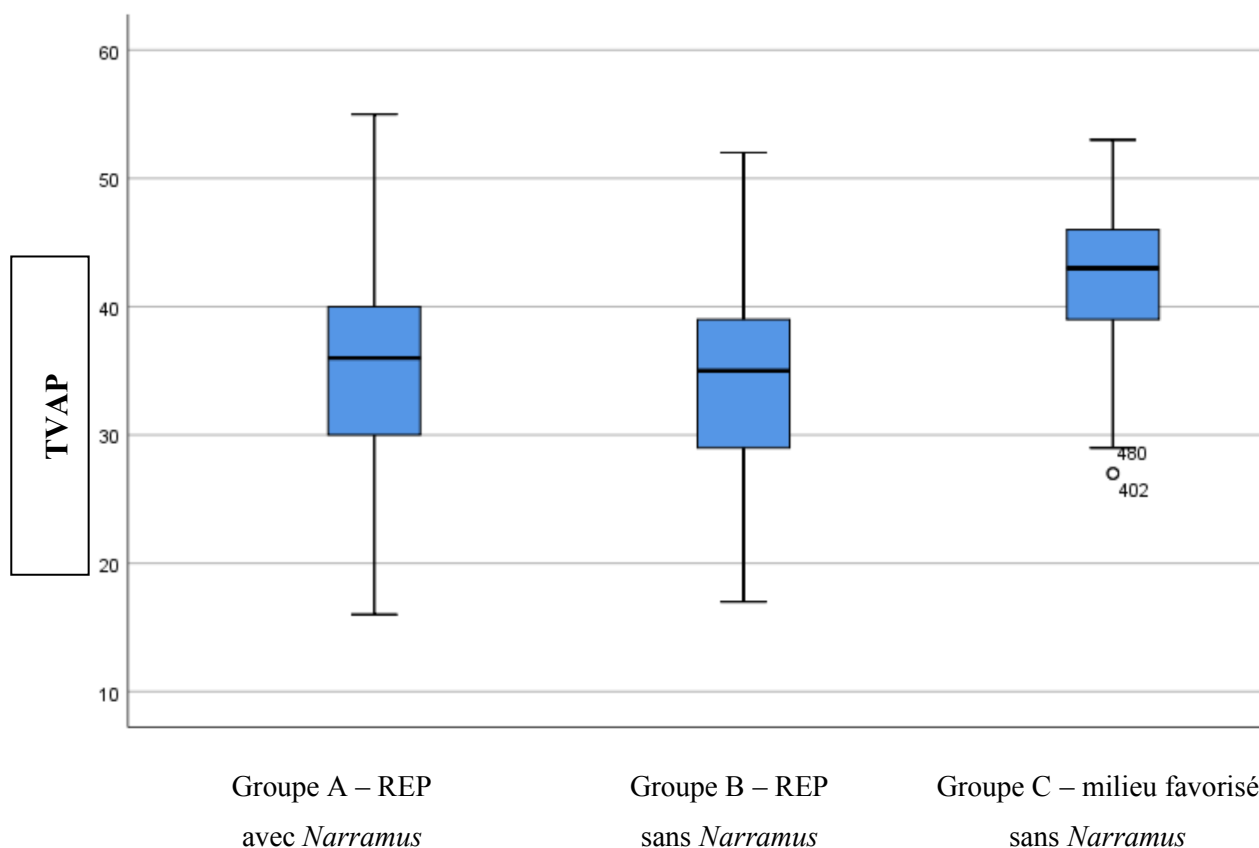
Légende : * p < 0,05 ** p < 0,01 *** p < 0,001 NS : Non significatif

Les résultats montrent que la différence de moyenne des groupes A et B n'est pas significative (p = 0,284) : leurs performances en compréhension pourront donc être comparées en postulant leur équivalence.

Comme on s'y attendait, les élèves favorisés surpassent nettement les élèves socialement défavorisés, ce qui n'empêche pas de trouver chez ces derniers, quelques enfants très performants : les scores maximaux des élèves des groupes A et B (55 et 52 points) sont voisins de ceux du groupe C (53 points). Le constat est différent pour les scores minimaux : les plus faibles des élèves de REP n'obtiennent que 16 (groupe A) ou 17 points (groupe B) alors que ceux de milieu favorisé atteignent 27 points sur 60. L'écart type de ce dernier groupe est également plus faible (5,4) que ceux des groupes A et B (7,7 et 7,5) : les scores des élèves de milieu favorisé sont donc moins contrastés et moins dispersés que ceux des élèves de REP.

L'écart de dispersion des scores entre les trois groupes apparaît nettement dans la Figure 20 : le premier quartile du groupe C correspond aux deuxièmes et troisièmes quartiles des groupes A et B.

Figure 20. Dispersion des scores des élèves des trois groupes au TVAP (/60)



Pour affiner la caractérisation sociale de notre échantillon, nous l'avons situé par rapport à une population de référence. Nous avons consulté l'étude Lire-Écrire CP (Goigoux, 2016) dont les résultats, plus récents que ceux de Deltour et Hupkens (1980), constituent un étalonnage actualisé du TVAP (*cf.* Tableau 63) pour des enfants de six ans. Cette étude portait sur un échantillon représentatif de 2 507 élèves testés au début du CP (notre test a eu lieu en juin de GS).

Tableau 63. Résultats des élèves en début de CP au TVAP

(Recherche Lire-Écrire CP, épreuve de septembre 2013)

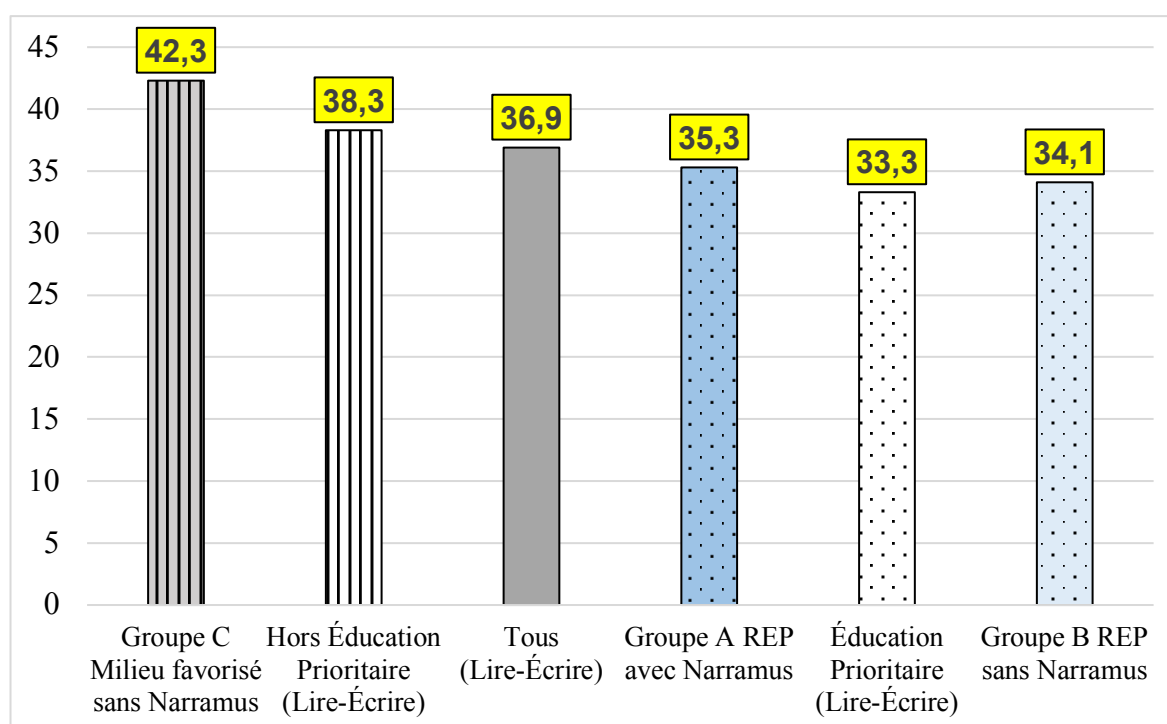
	Hors Éducation prioritaire N=1816	Tous N=2507	Éducation prioritaire N=691
	Moyenne (écart-type)		
TVAP Septembre 2013	38,3 (7,6)	36,9 (7,9)	33,3 (7,6)

On peut constater que les résultats des élèves de nos groupes A ($m = 35,3$) et B ($m = 34,1$) sont très proches de ceux obtenus par les élèves de l'éducation prioritaire ($m = 33,3$)

dans l'étude Lire-Écrire CP (cf. figure 21). Nos groupes A et B, issus d'écoles situées en REP ou REP+, ont donc bien les caractéristiques initiales de l'éducation prioritaire.

Le groupe C, quant à lui, peut être considéré comme favorisé dans la mesure où les performances des élèves qui le constituent ($m = 42,3$) sont nettement supérieures à celles des élèves scolarisés hors éducation prioritaire ($m = 38,3$) dans l'étude précitée (cf. Figure 21).

Figure 21. Comparaison des scores au test TVAP des groupes A, B et C de l'étude 2 et du pré-test de la recherche Lire-Écrire en 2013 (/60)



Les caractéristiques de notre échantillon permettent donc de répondre à nos deux questions de recherche.

1.3. Le choix de l'album support de l'évaluation

Nous avons choisi *La petite marmite qui tiptopait* (Bigot & Gaudrat, 2012), un album issu de la littérature de jeunesse, mais peu diffusé et rarement présent dans les écoles maternelles. Constitué d'un texte autonome (cf. annexe 9) accompagné d'illustrations, il appartient à la catégorie des contes dits « en randonnée ». Il raconte l'histoire d'une vieille dame qui, n'ayant plus rien à manger, range sa marmite devenue inutile sur une étagère. Mais la marmite ne veut pas rester sans rien faire et part se promener (l'auteur invente le néologisme : « tiptoper » pour évoquer la déambulation de la marmite). Elle rencontre une femme qui ramasse des fraises. La

marmite les lui vole et les rapporte à la vieille dame qui les mange toutes, avant de reposer la marmite sur l'étagère redevenue inutile. Mais la marmite repart tiptoper, rencontre un voleur, accapare son trésor et le rapporte à la vieille qui s'achète un tas de bonnes choses. Et la marmite retrouve son vrai « métier » : cuire la soupe !

Pourquoi ce choix ? Tout d'abord, il s'agit d'un album écrit dans une langue relativement simple (syntaxe et lexique) et qui comporte peu d'implicites. Il s'agit donc d'un texte qui ne pose pas de problème de compréhension majeure à des élèves de 5-6 ans.

Sa longueur (31 phrases et 457 mots) permet d'être lu ou écouté en une seule fois. En outre, les histoires en randonnée représentent un bon support pour évaluer les compétences narratives de jeunes élèves. « Leurs structures où des chaînes de personnages ou d'événements se répètent jusqu'au dénouement sont identifiables par les enfants⁷⁵ » (Éduscol, 2016, p. 16). Les enseignants impliqués dans l'étude ont tous déclaré ne pas avoir lu cet ouvrage à leurs élèves avant l'évaluation et ne pas l'avoir dans leur bibliothèque de classe.

1.4. Les épreuves d'évaluation

Pour évaluer la compréhension, nous avons conçu trois épreuves originales et identiques pour les trois groupes – rappel de récit, questionnaire et lexique – à partir de l'album *La petite marmite qui tiptopait*. Les élèves ont été testés après deux écoutes de l'histoire : ils ont tous entendu la même narration orale de 7 minutes 33 enregistrée par une conteuse, puis une lecture à haute voix. La narration orale est une version élargie de l'histoire qui comprend des ajouts (mots de liaison, reprises...), des reformulations lexicales et syntaxiques, et des jeux de voix. Elle vise à enrichir la qualité de la compréhension.

1.4.1. Déroulement de l'épreuve

L'épreuve de rappel de récit est proposée après que les élèves ont entendu deux fois l'histoire. Le premier jour, ils ont écouté un enregistrement de la version élargie sans voir les illustrations. Le lendemain, l'enseignant leur a lu l'album, cette fois en leur montrant et en commentant les images. Le même jour, un expérimentateur⁷⁶ a évalué tous les élèves en leur

⁷⁵ Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, http://cache.media.eduscol.education.fr/file/Langage/41/7/Ress_c1_langage_oral1.2_456417.pdf

⁷⁶ Les expérimentateurs avaient tous été préalablement formés par les chercheurs.

proposant trois épreuves. Les passations, d'une durée moyenne de quinze minutes, étaient individuelles.

1.4.1.1. Épreuve 1 : rappel de l'histoire

L'épreuve de rappel vise à évaluer la mémorisation des éléments importants du récit (macrostructure) et la quantité d'informations relatées (microstructure). L'élève raconte l'histoire sans support, l'expérimentateur peut le relancer en répétant les derniers mots prononcés ou en le questionnant « Et après, qu'est-ce qui se passe ? ».

Les rappels de récit sont enregistrés⁷⁷ sur support numérique et sont codés deux fois (*cf. Erreur ! Source du renvoi introuvable.*). Le premier codage porte sur la microstructure (score maximal = 48) et permet de savoir combien de propositions sémantiques du récit sont relatées (variable « mi »). Le second concerne la macrostructure (score maximal = 9), soit le nombre d'idées essentielles rappelées (variable « MA »).

Afin de nous assurer de la fiabilité des codages, nous avons procédé à un double codage en aveugle de tous les rappels : l'un est réalisé par nous, l'autre par un étudiant de master 2 formé par un membre de l'équipe de recherche. Le Tableau 64 présente les corrélations entre les scores proposés par les deux codeurs.

Tableau 64. Coefficients de corrélation intra classe

MA $\rho = (\text{IdC } 95 \%)$	mi $\rho = (\text{IdC } 95 \%)$
0,95 (0,94-0,96)	0,96 (0,95-0,97)

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

L'accord inter-juge s'élève à 0,95 en macrostructure et 0,96 en microstructure, seuil qui, dans la littérature statistique, est considéré comme excellent (Landis & Koch, 1977).

⁷⁷ Les enregistrements ont été déposés sur la plate-forme de l'ENS-Lyon.

1.4.1.1.1. Épreuve 2 : questionnaire

Cette épreuve (*cf.* annexe 11) a pour objectif d'évaluer la compréhension fine du texte entendu. Elle permet aussi d'analyser les liens entre la qualité du rappel et la compréhension de l'histoire. Immédiatement après le rappel, l'expérimentateur pose six questions à l'élève (variable « Q »).

Les réponses attendues étaient indiquées aux enquêteurs pour unifier au maximum le codage. Comme pour l'étude 1, en cas de doute, nous assurions nous-même le codage. Le questionnaire était toujours administré après le rappel de récit :

1. Au début de l'histoire, pourquoi la vieille range-t-elle la marmite sur l'étagère ? (Parce qu'elle n'a plus rien à manger)
2. Pourquoi la marmite prend-elle les fraises à la dame dans le champ ? (Pour les rapporter à la vieille, pour donner à manger à la vieille...)
3. Pourquoi la marmite attend-elle que le voleur ait le dos tourné pour retourner chez la vieille ? (Pour ne pas se faire rattraper par le voleur, pour ne pas être vue...)
4. Que fait la vieille avec les sous ? (Elle s'achète de bonnes choses à manger)
5. Quel est le vrai métier de la marmite ? (Faire la soupe)
6. Qui vole quelque chose dans cette histoire ? (1 point : le voleur ; 2 points : le voleur et la marmite)

Les cinq premières questions valent un point, la dernière, plus complexe, deux points, soit un score maximum de 7 points.

1.4.1.1.2. Épreuve 3 : lexique

Avec cette épreuve (*cf.* annexe 12), on cherche à mesurer les connaissances dont disposent les élèves sur le vocabulaire de l'album après les deux écoutes et selon deux modalités : en réception (désignation) et en production (dénomination, explication, mime).

Elle comprend 16 items (soit un score maximal de 16 points) : cinq noms (étagère, fraise, métier, route, voleur), cinq verbes (croquer, dégringoler, lécher, nettoyer, ranger), un pronom indéfini (rien), un adjectif (plein), une préposition (sous), deux expressions (ne rien faire, tourner le dos) et un mot inventé par l'auteur (le verbe *tiptoper*).

L'objectif est d'évaluer si des élèves de niveau lexical équivalent (évalué au TVAP) profitent différemment de l'écoute du récit lu puis raconté. Nous faisons, en effet, l'hypothèse que, compte tenu des caractéristiques de l'enseignement qu'ils ont reçu, les élèves du groupe A devraient être plus capables que ceux du groupe B de tirer parti du contexte pour donner un sens aux mots inconnus d'eux avant ces deux écoutes. Autrement dit, nous supposons qu'ils

réussiront mieux le texte lexical malgré un niveau lexical général équivalent. Les mots proposés dans notre test ont été sélectionnés selon différents critères :

1° nous nous sommes basée sur le modèle à 3 niveaux (Beck, McKeown, & Kucan, 2013; McKeown & Beck, 2004) en privilégiant des mots de niveau 1 et 2 ;

2° nous avons veillé à évaluer des mots de nature grammaticale différente (*cf. supra*) ;

3° nous avons sélectionné certains mots en nous assurant de leur présence dans la liste de fréquence proposée pour le cycle 2 dans les ressources Éduscol⁷⁸ (MEN). Cette liste constituée par le lexicologue Étienne Brunet rassemble près de 1500 mots les plus fréquents de la langue française et rend compte de la langue que lisent les élèves francophones. Les mots, extraits de textes littéraires ou non, ont été ramenés à leur base lexicale. Sept des mots de notre corpus appartiennent à cette liste (étagère, ranger, voleur, métier, plein, rien, route). En bref, la moitié des mots testés figure dans ce répertoire (on ne peut évidemment ni compter les deux expressions *ne rien faire* et *tourner le dos*, ni le néologisme *tiptoper*).

Nous avons deux hypothèses majeures sur les mots testés.

La première portait sur le néologisme *tiptoper*. Dans les deux épisodes de l'histoire, la marmite déclare : « *Bon, puisque c'est comme ça, je m'en vais tiptoper !* ». Cette phrase, répétée deux fois dans le texte, est chaque fois suivie de la ritournelle « *Et hop ! La voilà partie sur la route. Et tip top et... tip top et... tip top !* ». Afin de mesurer les interprétations personnelles des élèves, nous avons demandé à la conteuse et aux enseignants de ne pas expliquer ce mot (même si les élèves les questionnaient). Nous pensions que les élèves du groupe A pourraient mettre à profit les multiples activités lexicales dont ils avaient bénéficié avec *Narramus* pendant deux ans pour donner un sens plausible à ce mot, nécessairement inconnu.

La seconde touche le verbe *dégringoler*. Ce terme, peu utilisé dans le langage quotidien, revient deux fois dans l'histoire : « *Elle a sauté de l'étagère, elle a dégringolé par terre [...]* ». Dans la narration, nous avons demandé à la conteuse de produire une expansion du moment où la marmite dégringole.

Pour l'épisode 1, elle a raconté de la manière suivante : « *Elle commence à bouger un tout petit peu et puis de plus en plus et tout d'un coup elle saute carrément de l'étagère, elle tombe d'abord sur la table et après zou, elle dégringole jusqu'en bas. Heureusement elle ne*

⁷⁸ <https://eduscol.education.fr/cid50486/liste-de-frequence-lexicale.html>

s'est pas fait mal en tombant par terre ! ». Nous avons supposé que cette reformulation pouvait aider les élèves à se représenter mentalement la scène même s'ils ignoraient, au départ, le sens du mot *dégringoler*.

Pour l'épisode 2, voici ce qu'elle raconte : « *Elle recommence à bouger d'abord un tout petit peu et puis de plus en plus et comme la dernière fois elle finit par sauter de l'étagère et elle tombe d'abord sur la table et après elle dégringole jusque sur le sol. Heureusement qu'elle est solide et qu'elle ne s'est pas cassée en tombant* ».

Avec *Narramus*, les élèves ont régulièrement écouté des narrations faites par l'enseignant, une conteuse ou leurs camarades. Ils ont, en outre, appris à s'interroger sur le sens des mots inconnus puisque le texte était toujours présenté avant l'image. C'est pourquoi, selon nous, les élèves du groupe A devraient être en mesure de mieux le comprendre que ceux du groupe B.

2. Résultats

Notre première hypothèse porte sur les compétences construites par les élèves d'éducation prioritaire ayant travaillé avec *Narramus*. Nous cherchons à savoir si leurs apprentissages sont limités aux textes étudiés en classe (*cf.* Roux-Baron, Cèbe & Goigoux, 2017) ou s'ils ont une portée plus générale. Autrement dit, nous voulons savoir s'ils ont profité de l'enseignement pour développer des attitudes et des habitudes de traitement qui les rendraient plus efficaces que leurs camarades de même milieu, dans une situation inédite ? Si cette hypothèse est valide, ils devraient mieux raconter, mieux répondre aux questions et comprendre plus de mots.

2.1. Hypothèse 1 : vers un transfert des compétences ?

Compte tenu de notre hypothèse 1, nous ne comparerons ici que les résultats du GA et du GB. Le Tableau 65 rassemble les résultats des élèves scolarisés en éducation prioritaire, aux trois épreuves.

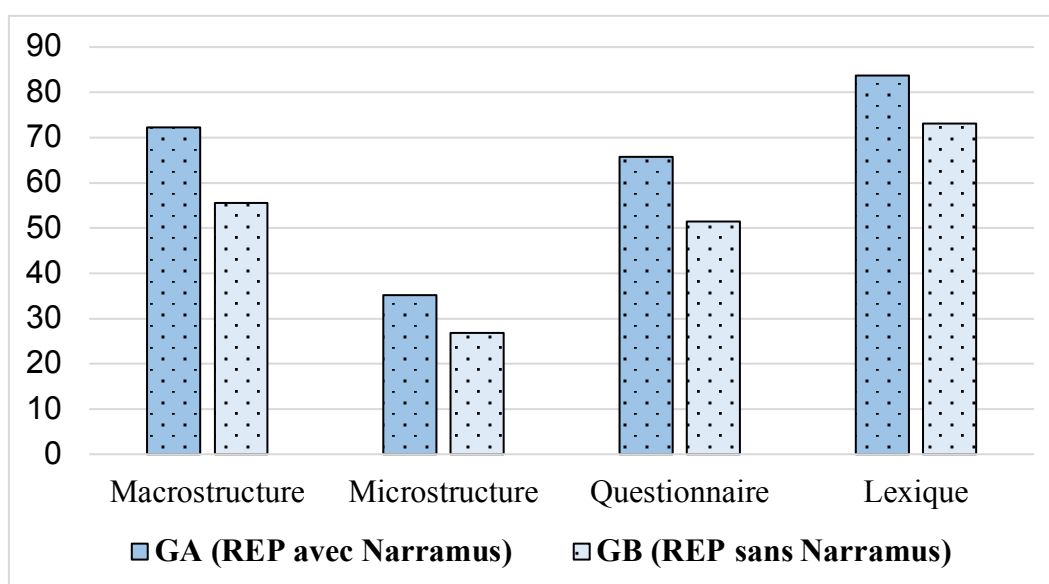
Tableau 65. Comparaison des scores des élèves des groupes A et B aux trois épreuves

	Groupe A REP avec <i>Narramus</i>	Groupe B REP sans <i>Narramus</i>	Tests
	Moyenne (écart-type) Min-max		T test
	N= 160	N= 205	GA vs GB
Rappel : macrostructure	6,5 (2,8) 0-9	5,0 (3,3) 0-9	***
Rappel : microstructure	16,9 (10,2) 0-38	12,9 (10,2) 0-37	***
Questionnaire	4,6 (1,9) 0-7	3,6 (1,9) 0-7	***
Lexique	13,4 (2,2) 2-16	11,7 (3,1) 0-16	***

*** p < 0,001

Dans toutes les épreuves, les élèves du groupe A obtiennent des scores significativement supérieurs à ceux du groupe B⁷⁹ (p < 0,001), comme l'illustre la Figure 22.

Figure 22. Comparaison des scores des élèves de REP aux trois épreuves



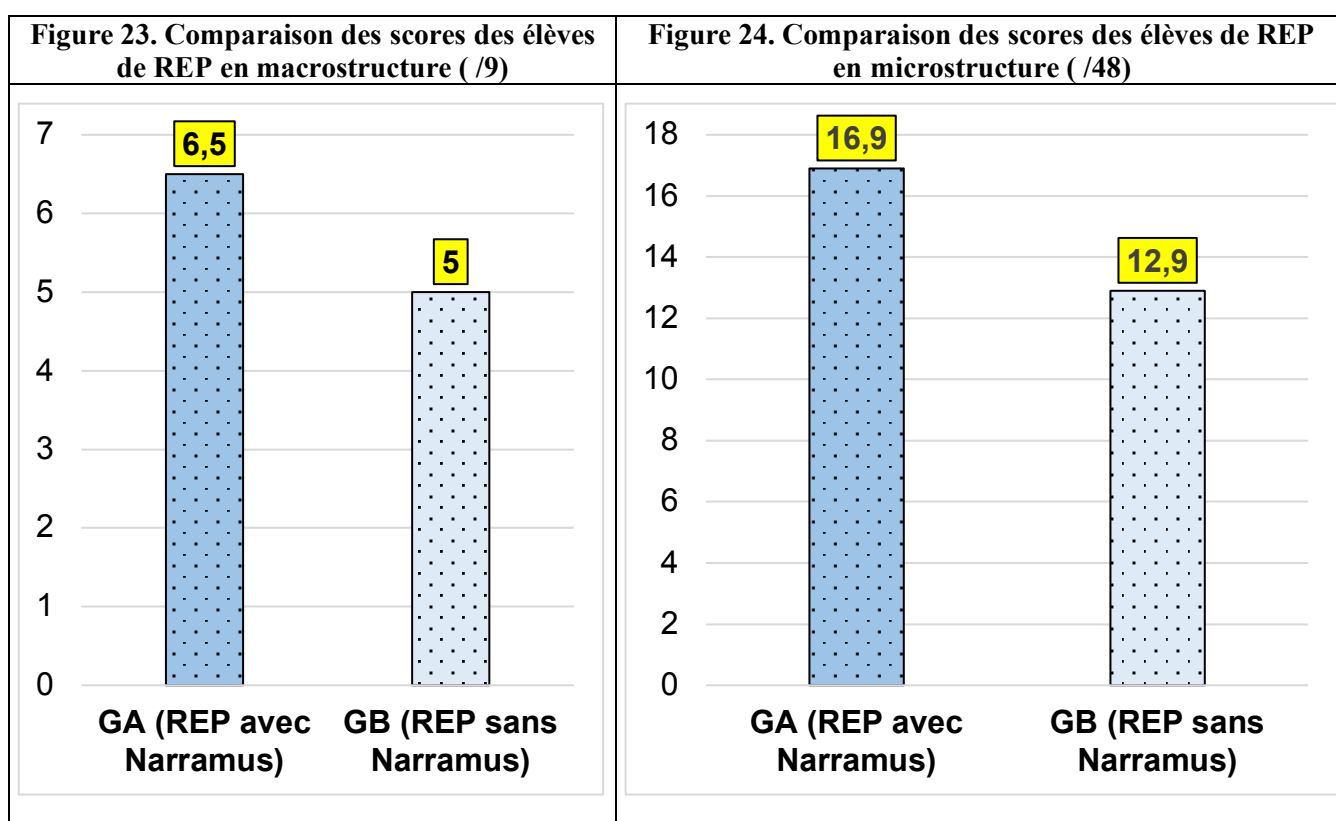
Légende : les scores ont tous été ramenés en base 100

⁷⁹ Afin de valider ces analyses et de tester directement notre hypothèse, nous avons effectué des analyses de contraste comme le suggèrent Brauer et McClelland (2005). Les résultats obtenus confirment que les scores du groupe A sont significativement supérieurs à ceux du groupe B, pour les trois épreuves.

2.1.1. Rappel de récit

En moyenne, les élèves du groupe A rappellent davantage de propositions sémantiques que ceux du groupe B ($m_{GA} = 16,9$; $m_{GB} = 12,9$) et relatent plus d'éléments de la macrostructure ($m_{GA} = 6,5$; $m_{GB} = 5,0$).

Les Figure 24 et Figure 23 donnent une représentation graphique des différences significatives constatées entre les deux groupes.



2.1.2. Questionnaire

À cette épreuve, les élèves du groupe A obtiennent, en moyenne, des scores significativement supérieurs à ceux du groupe B ($GA = 4,6$; $GB = 3,6$). Ces scores sont fortement corrélés aux scores de rappel, comme le montre le Tableau 66.

Tableau 66. Corrélations des variables macrostructure, microstructure et questionnaire pour les groupes A et B

	mi	MA	Q
mi	1	0,874***	0,719***
MA		1	0,708***
Q			1

*** p < 0,001

Les corrélations entre les variables du rappel de récit (mi et MA) et la variable questionnaire sont fortes (0,719 et 0,708), ce qui indique un lien important entre les compétences de compréhension et de narration que nous avons évaluées.

2.1.3. Lexique

Alors que leurs résultats initiaux au TVAP ne les distinguaient pas, les élèves de REP n'obtiennent pas les mêmes scores à l'épreuve de vocabulaire portant sur l'album. Ceux du groupe A (m = 13,4) sont significativement supérieurs à ceux du groupe B (m = 11,7). On peut donc soutenir que l'écoute du récit a permis aux élèves du groupe A de tirer bénéfice du contexte pour comprendre plus de mots.

2.1.4. Conclusion

L'ensemble de ces résultats présentés dans la figure 22 valide notre première hypothèse : les élèves scolarisés en éducation prioritaire ayant travaillé avec *Narramus* ont profité de l'enseignement pour développer des attitudes et des habitudes de traitement qui les rendent plus efficaces que leurs camarades de même milieu, dans une situation inédite. Forte de ces résultats, nous pouvons tester notre seconde hypothèse.

2.2. Hypothèse 2 : vers une réduction des inégalités ?

Pour rappel, nous nous demandions si la mise en œuvre de *Narramus* au cours de deux années scolaires avait permis de réduire les écarts entre élèves de milieux sociaux contrastés sur les trois dimensions testées. Pour répondre à cette question, nous avons comparé les résultats du

groupe A à ceux du groupe C. Les élèves du groupe A ont-ils rejoint ceux du groupe C ? Le Tableau 67 apporte la réponse à cette question.

Tableau 67. Comparaison des scores des élèves des groupes A et C aux trois épreuves

	Groupe A REP avec <i>Narramus</i>	Groupe C Milieu favorisé sans <i>Narramus</i>	Tests
	Moyenne (écart-type) Min-max		T test
	N= 160	N= 157	GAvsGC
Rappel : macrostructure	6,5 (2,8) 0-9	7,4 (2,3) 0-9	**
Rappel : microstructure	16,9 (10,2) 0-38	19,9 (8,8) 0-40	**
Questionnaire	4,6 (1,9) 0-7	5,6 (1,3) 1-7	***
Lexique	13,4 (2,2) 2-16	14,5 (1,6) 9-25	***

Légende : ** $p < 0,01$ *** $p < 0,001$

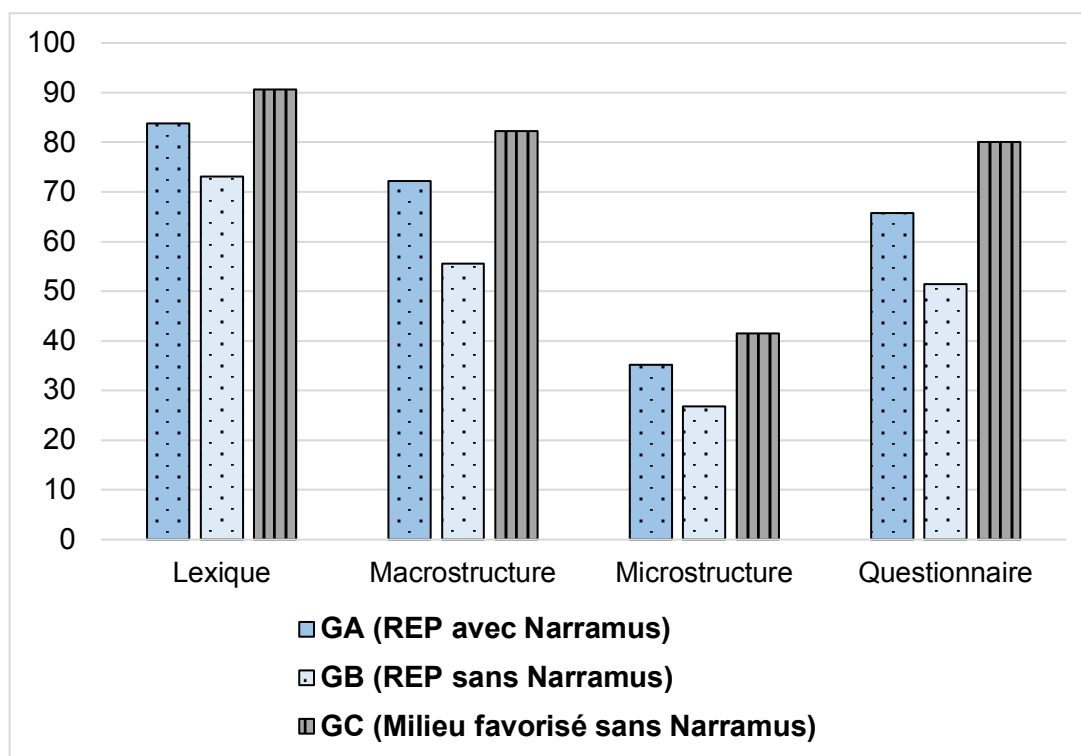
En rappel de récit, les élèves du groupe C obtiennent en moyenne (MA = 7,4 et mi = 19,9) des scores significativement supérieurs ($p < 0,01$) à ceux du groupe A (MA = 6,5 et mi = 16,9). Ils rapportent davantage d'éléments importants de l'histoire et sont plus précis dans la narration.

Pour le questionnaire, les scores des élèves du groupe C sont de 5,6 points contre 4,6 pour ceux du groupe A. Cet écart est significatif ($p < 0,001$).

En lexique, les résultats des élèves du groupe C ($m = 14,5$) sont significativement supérieurs ($p < 0,001$) à ceux du groupe A ($m = 13,4$).

Synthèse des résultats

Les élèves du groupe A n'ont pas rejoint ceux du groupe C mais s'en rapprochent. Les écarts entre les performances du groupe C et celles du groupe A sont moins importants que ceux qui distinguent le groupe A du groupe B, comme le fait apparaître la figure qui suit.

Figure 25. Comparaison des scores des élèves des groupes A, B et C aux trois épreuves

Légende : Les scores ont tous été ramenés en base 100

La Figure 25 illustre que, pour les épreuves de rappel et de lexique, les écarts sont plus forts entre les élèves de même milieu qu'entre ceux du groupe A et du groupe C de milieu très favorisé.

Quand, après deux écoutes de l'album, les élèves du groupe A comprennent 83,75 % des mots, ceux du groupe B en comprennent 73,13 % et ceux du groupe C, 90,63 %.

Pour la macrostructure, la différence GA vs GB est de 4 points, celle GA vs GC est de 3 points ; en microstructure, elle est respectivement de 1,5 et 0,9 points ; en lexique de 1,7 et 1,1 points. Pour l'épreuve questionnaire, la différence est la même : 1 point.

Ces résultats valident notre seconde hypothèse : l'utilisation de l'outil *Narramus* a réduit l'écart qui sépare les élèves de milieux sociaux contrastés sans pour autant le faire disparaître.

- Description des résultats par items

La lecture des résultats obtenus par les élèves des différents groupes à l'épreuve de questionnaire et à celle de lexique nous a amenée à nous interroger sur les réussites et les échecs en fonction des items. Observe-t-on des écarts marqués à certaines questions ? Les différences

de moyenne en lexique se retrouvent-elles sur tous les mots évalués ? Certaines notions marquent-elles plus particulièrement les écarts ?

Pour répondre à ces interrogations, nous avons analysé plus finement les résultats obtenus dans l'académie de Clermont-Ferrand. Le Tableau 68 indique les scores par item à l'épreuve de questionnaire pour les élèves des trois groupes.

Tableau 68. Comparaison de scores (bruts et %) des trois groupes dans l'académie de Clermont-Ferrand à l'épreuve questionnaire

Juin 2018	Groupe A REP avec <i>Narramus</i> N = 92	Groupe B REP sans <i>Narramus</i> N = 78	Groupe C Milieu favorisé sans <i>Narramus</i> N = 101	Anova	
	Nombre de réponses correctes Pourcentage de réponses correctes			Test post Anova p =	
Q1 /1	61 66,3 %	29 37,2 %	89 88,1 %	< 0,001	
				GA vs GB ***	GA vs GC **
Q2 /1	71 77,2 %	29 37,2 %	80 79,2 %	< 0,001	
				GA vs GB ***	NS
Q3 /1	37 40,2 %	10 12,8 %	53 52,5 %	< 0,001	
				GA vs GB ***	NS
Q4 /1	73 79,3 %	41 52,6 %	94 93,1 %	< 0,001	
				GA vs GB ***	GA vs GC *
Q5 /1	69 75 %	41 52,6 %	96 95 %	< 0,001	
				GA vs GB **	GA vs GC **
Q6 /2 1 ou 2 point(s) possible (s)	(1 point) 53 57,6 %	(1 point) 53 67,9 %	(1 point) 59 58,4 %		
	(2 points) 31 33,7 %	(2 points) 10 12,8 %	(2 points) 40 39,6 %	< 0,001	
				GA vs GB **	GA vs GC NS

* p < 0,05 ** p < 0,01 *** p < 0,001 NS : Non Significatif

Pour toutes les questions, les différences entre les élèves de REP sont significatives (p < 0,01 ou p < 0,001) alors que celles qui distinguent les élèves du groupe A et du groupe C ne le sont que pour la moitié des items.

On peut noter que ce sont les questions dont les réponses sont le plus explicites qui donnent lieu aux différences significatives entre les groupes A et C :

– pour la question 1 (*Au début de l'histoire, pourquoi la vieille range-t-elle la marmite sur l'étagère ?*), la réponse est écrite dans le texte mais les élèves doivent tout de même inférer le lien entre les deux premières propositions (*Un jour, la vieille n'eut plus rien à manger. Elle a nettoyé la marmite, elle l'a rangée sur l'étagère, et elle lui a dit : « Tu sais, ma petite marmite, ton métier de faire la soupe est terminé, parce que moi, je n'ai plus rien à te donner »*).

– Les réponses aux questions 4 (*Que fait la vieille avec les sous ?*) et 5 (*Quel est le vrai métier de la marmite ?*) sont explicitement écrites dans le texte : « *Quand la vieille a vu tous les sous, elle ne les a pas mangés, mais non, elle s'est achetée un tas de bonnes choses.* » et « *Et c'était quoi son vrai métier ? C'était de faire la soupe, c'était pas de tiptoper !* »

On observe donc que les élèves du groupe A ont développé des compétences leur permettant de mieux répondre à des questions complexes faisant appel à l'implicite du texte. Ils se rapprochent ainsi des habiletés développées par les élèves de milieu social plus favorisé. Les réponses données à la question 6 illustre bien ce constat : quand 67,9 % des élèves du groupe B obtiennent 1 point en qualifiant de voleur uniquement le personnage facilement identifiable dans l'album (le voleur), ils ne sont plus que 12,8 % à obtenir 2 points en identifiant aussi le second personnage (la marmite). Ces scores sont respectivement de 57,6 % et 33,7 % pour les élèves du groupe A et 58,4 % et 39,6 % pour ceux du groupe C. En résumé, les élèves du groupe *Narramus* approchent, pour l'épreuve questionnaire, des compétences expertes des élèves issus des milieux sociaux défavorisés.

Pour le lexique, on peut dresser sensiblement le même constat en sélectionnant des mots qui marquent davantage les différences entre les groupes, comme l'indique le Tableau 69.

Tableau 69. Comparaison de scores (bruts et %) des trois groupes dans l'académie de Clermont-Ferrand à l'épreuve lexique

Juin 2018	Groupe A REP avec <i>Narramus</i> N = 92	Groupe B REP sans <i>Narramus</i> N = 78	Groupe C Milieu favorisé sans <i>Narramus</i> N = 101	Anova	
	Nombre de réponses correctes Pourcentage de réponses correctes			Test post Anova p =	
el_1 (fraise)	91 99,9 %	77 98,7 %	100 99,0 %	0,983 NS	
el_2 (route)	92 100 %	76 97,4 %	101 100 %	0,083 NS	
el_3 (voleur)	92 100 %	77 98,7 %	100 99,0 %	0,584 NS	
el_4 (étagère)	44 47,8 %	22 28,2 %	34 33,7 %	0,024	
				GA vs GB *	GA vs GC NS
el_5 (sous)	85 92,4 %	56 71,8 %	98 97 %	< 0,001	
				GA vs GB ***	GA vs GC NS
el_6 (rien)	90 97,8 %	72 92,3 %	99 98,0 %	0,085 NS	
el_7 (plein)	90 97,8 %	74 94,9 %	100 99,0 %	0,215 NS	
el_8 (tiptoper)	59 64,1 %	25 32,1 %	68 67,3 %	< 0,001	
				GA vs GB ***	GA vs GC NS
el_9 (ranger)	90 97,8 %	71 91,0 %	92 91,1 %	0,107 NS	
el_10 (métier)	59 64,1 %	35 44,9 %	95 94,1 %	0,008	
				GA vs GB ***	GA vs GC ***
el_11 (lécher)	66 71,7 %	41 52,6 %	101 100 %	< 0,001	
				GA vs GB ***	GA vs GC NS
el_12 (dégringoler)	47 51,1 %	15 19,2 %	67 66,3 %	< 0,001	
				GA vs GB ***	GA vs GC NS
el_13 (croquer)	91 98,9 %	73 93,6 %	100 99 %	0,041	
				GA vs GB NS	GA vs GC NS
el_14 (nettoyer)	88 95,5 %	75 96,2 %	98 97 %	0,206 NS	
el_15 (ne rien faire)	92 100 %	70 89,7 %	101 100 %	***	
				GA vs GB ***	GA vs GC NS
el_16 (tourner le dos)	75 81,5 %	39 50,0 %	82 81,2 %	***	
				GA vs GB ***	GA vs GC NS

* p < 0,05 ** p < 0,01 *** p < 0,001 NS : Non Significatif

Sur l'ensemble de l'épreuve, il n'y a pas de différence significative de réussite entre les groupes pour 7 des 16 mots et expressions évalués. On observe que ces concepts sont les mieux connus des élèves puisque le taux de réussite oscille entre 92,3 % et 100 % pour les trois groupes. Autrement dit, quel que soit leur groupe d'appartenance, les élèves évalués ont une bonne connaissance des mots fraise, route, voleur, rien, plein, ranger, nettoyer. Cinq de ces mots (plein, rien, route, ranger, voleur) figurent sur la liste lexicale des mots fréquents à connaître en fin de cycle 2 (Éduscol).

Les items pour lesquels les écarts sont les plus forts entre les élèves de même milieu social (défavorisé) sont les verbes *tiptoper* (el_8) et *dégringoler* (el_12). Les élèves du groupe A sont deux fois plus nombreux (64,1 %) à expliquer le néologisme que ceux du groupe B (32,1 %) ; pour le verbe dégringoler, la différence est encore plus grande : 51,1 % vs 19,2 %. Ceci est le signe d'une plus grande capacité à tirer profit du contexte linguistique pour élaborer le sens de mots inconnus.

Les expressions *ne rien faire* (el_15) et *tourner le dos* (el_16) distinguent également les élèves de même milieu : les différences sont significatives dans les deux cas. Pour l'item 15, la totalité des élèves du groupe A donnent une explication plausible de l'expression contre 89,7 % des élèves du groupe B. C'est toutefois l'item 16 qui montre l'écart le plus important entre les deux groupes qui obtiennent respectivement 81,5 % et 50 % de réussite. On observe que, pour ces items, les taux de réussite entre les groupes A et C sont quasiment identiques (el_15GC = 100 % et el_16GC = 81,2 %).

Un seul mot présente un écart significatif entre les élèves des trois groupes : *métier*. Le taux de réussite des élèves du groupe A s'élève à 64,1 %, celui du groupe B à 44,9 % et celui du groupe C à 94,1 %. Ce nom commun est le seul item pour lequel la différence est significative entre les groupes A et C. La valeur sociale associée à ce mot n'y est sans doute pas pour rien car les parents de milieux sociaux contrastés n'ont pas le même rapport à leur métier. À Clermont, certains « travaillent » chez Michelin (les ouvriers) alors que d'autres y sont ingénieurs. Emplois et métiers ne sont pas équivalents. Ce constat concorde avec l'expérience des enseignants d'école maternelle qui savent à quel point une partie de leur public ignore le métier de ses parents.

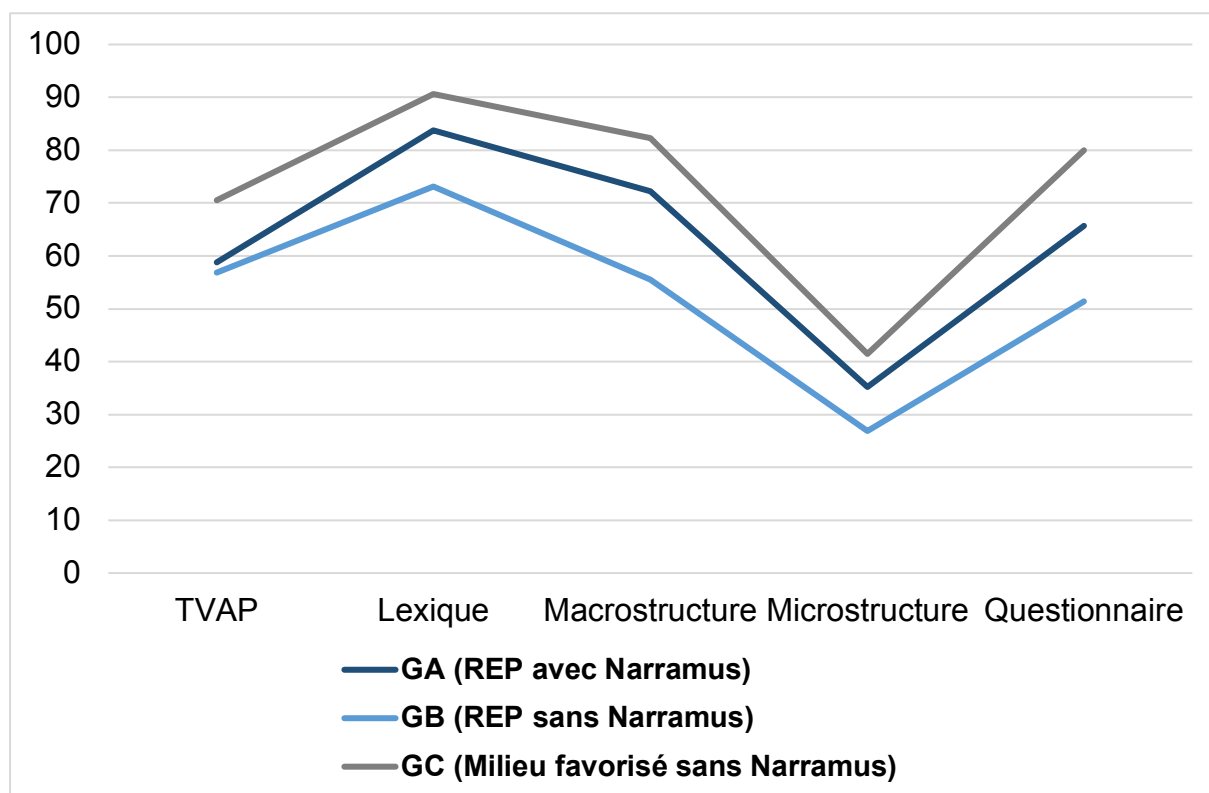
3. Discussion

Avec cette deuxième étude, nous nous attendions à ce que les élèves du groupe A mettent à profit l'enseignement dispensé avec *Narramus* pour mieux comprendre une histoire nouvelle que leurs camarades de même milieu et donc mobiliser des modes de traitement de l'écrit plus efficaces. Au-delà de cet effet de transfert, nous avons émis l'hypothèse que les élèves du même groupe devraient se rapprocher de ceux des élèves de milieu favorisé et obtenir des résultats attestant une réduction des inégalités.

Pour tester ces hypothèses, nous avons évalué la compréhension du langage écrit à partir d'un album de la littérature de jeunesse non étudié en classe. Une épreuve de rappel de récit, une de lexique et un questionnaire de compréhension ont été proposés en fin de Grande Section à trois groupes d'élèves scolarisés dans des écoles classées en REP pour les uns, dans un quartier favorisé pour les autres.

Comme l'illustre la Figure 26, la présente étude montre que les élèves socialement défavorisés ayant bénéficié d'un enseignement particulier de la compréhension au cours de deux années scolaires (A) ont développé des compétences et des connaissances leur permettant, d'une part, d'obtenir des résultats supérieurs à ceux d'élèves de même milieu (B) et, d'autre part, de réduire les écarts avec des élèves de milieu favorisé (C). La ligne A se situe entre les lignes B et C pour toutes les épreuves, sauf pour le TVAP qui est indépendant de l'enseignement dispensé.

Figure 26. Comparaison des scores des élèves des groupes A, B et C aux quatre épreuves



Légende : Les scores ont tous été ramenés en base 100

Cette étude confirme notre hypothèse selon laquelle avoir suivi six scénarios *Narramus* a favorisé le développement de compétences plus générales qui ont permis aux élèves du groupe A de mieux comprendre un texte qu'ils ne connaissaient pas. Elle montre qu'il est possible d'influer positivement sur le parcours scolaire des élèves de milieux populaires en modifiant les conditions d'enseignement qui leur sont faites à l'école maternelle. Or, jusqu'à présent, la scolarisation précoce et la politique des REP n'avaient guère permis de réduire les écarts d'efficience (CNESCO, 2016).

De tels résultats sont rares et peuvent être interprétés à partir des caractéristiques didactiques et pédagogiques de *Narramus*.

- Rappel de récit et questionnaire

La corrélation entre les épreuves de rappel et le questionnaire étant particulièrement forte (*cf.* tableau 66), nous les analysons ensemble. Compte tenu des caractéristiques de l'enseignement proposé, nous faisons l'hypothèse que si les élèves comprennent mieux et racontent mieux que leurs congénères de même milieu c'est parce qu'ils ont appris à traiter l'écrit (les mots, les phrases, le texte) en profondeur. En effet, dans tous les scénarios, le texte

et l'image sont présentés de manière dissociée, un mode de faire souvent absent des classes maternelles – et des pratiques familiales – dans lesquelles les deux supports sont montrés simultanément. Pendant deux ans, les élèves du groupe A ont eu à traiter ce que disait l'enseignant (lecture ou narration d'un court extrait) pour prévoir le contenu de l'illustration. Juste après, ils devaient débattre oralement de ce qu'ils s'attendaient à observer et, une fois la discussion close, l'image était affichée pour valider ou invalider les points de vue. Les élèves devaient ensuite comparer très précisément ce que disait le texte et ce que montrait l'image. Cette manière de traiter les mots avant les images peut expliquer les différences de résultats obtenus entre les élèves du groupe A et ceux du groupe B.

On peut peut-être attribuer les scores en narration à la ritualisation : avec *Narramus*, les élèves ont développé des habitudes de traitement qui ont profondément modifié leur écoute et leur attitude pendant que l'enseignant lit une histoire. Quand ils sont en autonomie, « livrés à eux-mêmes » face à un nouveau texte, ils mobilisent les stratégies habituellement mises en œuvre avec les scénarios. Nous supposons qu'ils reconnaissent la tâche comme faisant partie de leur répertoire et se placent dans une position d'écoute intentionnelle. Ainsi, quand on leur demande de raconter l'histoire de *La petite marmite qui tiptopait*, ils racontent mieux et davantage que leurs camarades de même milieu.

Nous supposons aussi que l'activité de rappel, qui finalise la tâche d'écoute de la lecture oralisée, donne un but intégrateur aux activités menées en classe. Les élèves ont appris à être cognitivement actifs pendant leur écoute. Ils ont compris l'intérêt d'être attentif pour réaliser une tâche qu'ils apprécient (raconter) et ils écoutent le récit avec une claire conscience du but poursuivi. C'est pour mieux raconter qu'ils mettent en mémoire le nom des personnages, les événements, les lieux, les actions et les bonnes raisons que les personnages ont de faire ce qu'ils font.

L'utilisation du vidéoprojecteur facilite le guidage de l'enseignant et le maintien de l'attention sur ce qui fait l'objet de la discussion collective. Les élèves ont appris à écouter et à être attentifs dans des moments d'enseignement collectif alors qu'ils le font peu habituellement et que les apprentissages à l'école maternelle sont très souvent organisés en atelier.

En outre, de multiples activités incitent régulièrement les élèves à raisonner, justifier et à aller au-delà de l'implicite du texte : identifier les états mentaux, produire des inférences causales ou temporelles, expliciter ce que l'auteur ne dit pas... Plusieurs recherches ont montré l'influence de ce type de questionnement sur la compréhension, que ce soit dans le cercle

familial (Reese, Haden, & Fivush, 1993) ou en milieu scolaire (D. K. Dickinson & Smith, 1994). À titre d'exemples, les simples questions de rappel ou de compréhension qui portent sur des éléments précis et explicites du texte, les questionnements fermés entraînant une réponse par « oui » ou par « non » sans justification complémentaire sont des procédés qui engagent moins les enfants dans la compréhension des histoires. Ainsi, les échanges menés avec *Narramus* ont probablement permis de favoriser la construction du sens des histoires et ont pu servir de modèle d'analyse lors de la découverte du texte inédit : les élèves semblent avoir développé des stratégies d'analyse leur ayant permis de mieux comprendre le nouveau récit (de Weck, 2007; Veneziano, 2010).

Enfin, la narration régulière effectuée par l'enseignant permet d'améliorer la compréhension. Les élèves mémorisent et comprennent mieux une histoire quand elle leur est racontée que lorsqu'on la leur lit (Boiron, 2014). En racontant, le maître ajoute des éléments au texte source, reformule, donne la parole à certains personnages, revient sur un passage implicite qu'il faut avoir compris, met en scène le texte... Ces procédés apparaissent dans les stratégies d'adjonction (Frier, 2006) utilisées par les lecteurs pour créer un « espace réflexif » autour du texte. « La lecture et le contage proposent des modalités d'adaptation semblables (intonations montantes, pauses), mais l'activité de contage propose des ajouts (mots de liaison, reprises...) et des marqueurs d'intensité et de dramatisation très accentués qui permettent à un plus grand nombre d'élèves de repérer, de désigner les personnages et de leur attribuer des qualités » (Boiron, 2014, p. 94).

Nous émettons l'hypothèse que les caractéristiques de l'outil décrites ci-dessus ont permis le développement d'une écoute active, d'une attention dirigée, d'un raisonnement et d'un traitement de l'écrit en profondeur favorisant le développement de la compréhension et des compétences narratives.

En outre, pour chaque album, les concepteurs de *Narramus* proposent une intervention relativement longue (un mois) qui peut être à l'origine des effets observés sur la qualité des rappels. En effet, il est possible que le fait d'avoir abordé plusieurs histoires de la même manière ait permis aux élèves de s'approprier, d'intérioriser la démarche pour la réinvestir lors de l'écoute d'autres albums (Martinez & Teale, 1993).

- Lexique

Narramus accorde une place importante à l'enseignement, à la mémorisation, à la révision et à l'utilisation du lexique. On peut faire l'hypothèse raisonnable que cette intervention dense et

originale a permis le développement progressif d'une « conscience lexicale » qui amène les élèves à devenir curieux des mots, à s'y intéresser pour eux-mêmes (leurs formes) et par rapport à d'autres sur le versant de leurs similarités et de leurs différences (Beck, Perfetti, & McKeown, 1982; Graves, 2000; Perfetti, McKeown, & Kucan, 2010). C'est en tout cas ce qui ressort des témoignages d'enseignants. Dans l'étude qualitative que nous avons menée auprès de 102⁸⁰ enseignants en juin 2018, plus de 70 % déclarent avoir profondément modifié leur enseignement du lexique et que ces pratiques nouvelles ont permis de rendre leurs élèves plus attentifs et plus sensibles aux mots du texte⁸¹.

Plus spécifiquement, nous faisons l'hypothèse que les élèves du groupe A ont profité de la lecture et de la narration de l'histoire pour comprendre deux mots qu'ils ne connaissaient pas avant : *tiptoper* et *dégringoler*. Dans l'académie de Clermont-Ferrand, nous avons analysé plus finement les résultats des différents groupes pour ces deux items.

Tableau 70. Comparaison de scores (bruts et %) des trois groupes dans l'académie de Clermont-Ferrand pour les items *tiptoper* et *dégringoler*

Jun 2017	Effectifs	Tiptoper Nombre de réponses correctes % de réponses correctes	Post test Anova p =	Dégringoler Nombre de réponses correctes % de réponses correctes	Post test Anova p =
Groupe A REP avec <i>Narramus</i>	N= 92	59 64,1 %	A vs B ***	47 51,1 %	A vs B ***
Groupe B REP sans <i>Narramus</i>	N = 78	25 32,1 %		15 19,2 %	
Groupe C Milieu favorisé sans <i>Narramus</i>	N = 101	68 67,3 %		A vs C NS	

Pourquoi les élèves du groupe A ont-ils réussi à expliciter le verbe « *tiptoper* » alors que les élèves du groupe B ont échoué ? Nous supposons que les premiers ont su tirer profit de leur

⁸⁰ Il s'agit d'une enquête qualitative menée auprès des enseignants engagés dans l'étude depuis septembre 2016, cf. chapitre 7.

⁸¹ À ces constats, s'ajoute le témoignage de M. remplaçante pour une semaine dans une classe de Grande Section du groupe A : « Les élèves s'intéressent aux mots, ils interrompent ma lecture pour me demander ce que signifie un mot ou m'interrogent sur ce mot à la fin de l'histoire. C'est une découverte pour moi et pourtant j'en vois des élèves ! ».

écoute et ont utilisé le contexte pour donner un sens approximatif à ce mot inventé par l'auteur et donc inconnu (*marcher, se promener, aller quelque part* sont les explications les plus fréquemment données). Ils ont su interagir avec le texte et effectuer des liens qui incombent au lecteur : « Au lecteur revient encore la tâche de tisser des liens entre les mots "treublants", ces mots qui vacillent quand on les rapproche d'autres mots et qui, rassemblés en réseaux pluriels, exhibent à chaque confrontation de nouvelles strates de sens » (Tauveron, 2002, p. 82).

Ces mêmes élèves sont plus nombreux à avoir compris le sens du mot « dégringoler ». Nous interprétons cet écart de réussite par l'importance accordée à l'enseignement des verbes dans *Narramus* ainsi que par les habitudes données aux élèves pour élaborer une représentation mentale des espaces et des déplacements. Des activités variées leur ont appris à représenter mentalement ou physiquement les déplacements : narration avec la maquette, théâtralisation de l'histoire, mimes des mots de l'album en salle de motricité... On connaît l'importance de la mémoire dans le corps qui influence positivement la mémorisation et la récupération des informations (Glenberg, 1997; Glenberg et al., 2004). Nous supposons que la pratique régulière de ces activités les a incités à porter une attention particulière à la narration de l'album et à se représenter mentalement la chute de la marmite.

La version éxpansée⁸² apporte des éléments qui complètent le texte original⁸³ et les élèves se sont emparés de ces informations pour donner un sens approximatif au verbe dégringoler.

Ces deux exemples de réussite sont particulièrement encourageants quand on sait que la quantité de verbes utilisés par de jeunes enfants en langage spontané est plus faible dans les milieux socio-économiques défavorisés (Le Normand et al., 2008) et qu'en 2016, les ressources Éduscol (Mobiliser le langage dans toutes ses dimensions - Lien oral-écrit - Lexique et syntaxe) pour l'école maternelle invitaient les enseignants à s'y intéresser : « Les noms sont fortement prégnants dans les choix pédagogiques [...]. Les verbes et les mots grammaticaux ne doivent surtout pas être négligés : ils apportent une contribution importante à la compréhension des énoncés. » (p. 5).

⁸² « Elle commence à bouger un tout petit peu et puis de plus en plus et tout d'un coup elle saute carrément de l'étagère, elle tombe d'abord sur la table et après zou, elle dégringole jusqu'en bas. Heureusement elle ne s'est pas fait mal en tombant par terre !

⁸³ « Elle a sauté de l'étagère, elle a dégringolé par terre [...]. »

On peut aussi supposer que l'expérience de l'écoute d'histoires a influencé les performances des élèves du groupe A, ceux du groupe B s'occupant peut-être davantage de l'intrigue globale que des nouveaux mots. Les élèves moins expérimentés dans l'écoute et la narration des histoires peuvent également avoir été moins intéressés ou moins motivés à apprendre de nouveaux mots (Robbins & Ehri, 1994).

Activement engagés dans l'apprentissage des mots, les élèves du groupe A semblent avoir développé une conscience des mots et une conscience métalinguistique. Le développement de cette dernière joue sur la manière dont les élèves vont consolider, synthétiser et accéder aux connaissances lexicales sur un mot particulier et être plus actifs dans l'apprentissage des mots (Coyne et al., 2015). Les élèves sauront davantage tirer profit des possibilités d'apprentissage des mots hors contexte d'apprentissage spécifique. Une des conséquences de cette augmentation de la conscience « lexicale » est que les élèves seront plus susceptibles de déduire le sens de mots nouveaux et de mieux comprendre le langage au quotidien. Dickinson et Smith (1994) ont émis l'hypothèse qu'à la suite de discussions sur le langage, les enfants peuvent être plus prompts à remarquer les mots qu'ils ne connaissent pas, plus enclins à chercher leurs significations, et plus en mesure d'utiliser l'information disponible pour calculer pour eux-mêmes. L'enseignement du vocabulaire vise à améliorer les connaissances lexicales des élèves et à développer leur littératie verbale, mais il doit aussi permettre aux élèves de développer des connaissances et des stratégies utilisables pour que les élèves donnent un sens aux mots dans de nouveaux contextes (Stahl & Fairbanks, 1986). On peut supposer que l'enseignement proposé dans *Narramus* (découverte, mémorisation réutilisation dans de multiples contextes...) a permis aux élèves de cumuler des savoirs lexicaux mais aussi de développer un intérêt et des stratégies de raisonnement sur les mots en contexte narratif.

Il nous paraît intéressant de revenir sur la réussite du mot *métier* (el_10). En effet, cet item est le seul qui distingue significativement les élèves du groupe A (64,1 %) de ceux du groupe C (94,1 %). On peut raisonnablement supposer que les élèves issus de milieux socio-économiques différents n'ont pas le même rapport à ce mot puisqu'il est moins fréquemment utilisé que d'autres termes tels que travail ou boulot. Il aurait été intéressant de croiser les résultats obtenus pour ce mot avec les données du test standardisé de lexique TVAP (Deltour & Hupkens, 1980). En effet, le second mot évalué dans cette épreuve est *travailler*. On pourrait

éventuellement observer des différences entre la connaissance de ces deux mots appartenant à des registres de langue différents.

La valorisation et l'intérêt portés au vocabulaire ainsi que les propositions didactiques et pédagogiques mises en œuvre pour enseigner les expressions et les mots dans *Narramus* ont sans doute permis aux élèves de se questionner sur les mots, d'interagir avec le texte et de tirer bénéfice des lectures oralisées pour attribuer un sens aux mots inconnus. Cette « sensibilité lexicale » nous paraît une hypothèse interprétative pertinente à la lecture de nos résultats.

4. Conclusion

Les résultats de cette seconde étude sont prometteurs : l'outil *Narramus*, tel qu'il est conçu, permet d'améliorer les compétences de compréhension d'un texte narratif avant enseignement et participe à la réduction des inégalités sociales.

L'enseignement explicite proposé dans *Narramus* repose sur des scénarios ritualisés et des tâches aux formats stables (Cèbe, 1998). Il a permis aux élèves du groupe A de construire des schèmes de traitement de l'écrit, d'attention et de raisonnement transférables sur un texte nouveau, ce qui a permis de réduire les inégalités sociales sur le versant de la compréhension en lecture et le développement du langage. Nous pouvons nous réjouir de ces résultats au regard d'autres données plus inquiétantes : les inégalités sociales à l'école perdurent et sont multiples (CNESCO, 2016). En effet, la France fait partie des pays de l'OCDE pour lesquels la corrélation des performances avec le milieu socio-économique est la plus forte (20 % en France vs 13 % en moyenne dans l'OCDE) (OCDE, 2016).

Les effets observés avec *Narramus* sont encourageants quand on sait que les compétences précoces en compréhension de textes sont décisives : la plupart des enfants qui sont rangés dans la catégorie des « faibles compreneurs » en fin de grande section comprennent toujours mal les textes qu'ils lisent en fin de CM1 et de 4^e, quelles que soient leurs compétences en décodage (Catts, Adlof, & Weismer, 2006; Elwér et al., 2015).

Les résultats obtenus en lexique semblent particulièrement prometteurs : plusieurs recherches (Landry, Dion, Dupéré, & Barrette, 2017) ont confirmé qu'apprendre le sens d'un mot est complexe et ont souligné le défi considérable pour les enseignants d'aider les enfants à développer leur vocabulaire. D'autres (Tilstra, McMaster, Van den Broek, Kendeou, & Rapp, 2009) ont montré la plus-value d'enseigner le vocabulaire dès l'école maternelle : un travail

précoce peut aider l'élève à interpréter et à extraire le sens d'un texte académique complexe qui se trouve habituellement dans le programme d'études secondaires. L'étude de Tilstra *et al.* (2009) révèle, en outre, un effet direct du vocabulaire de la maternelle sur la compréhension de la lecture en dixième année. Plus récemment, Stanley, Petscher et Catts (2018) ont mis en avant que l'acquisition du vocabulaire semble exercer un effet indirect sur la compréhension de la lecture par le biais de la compréhension de textes entendus et du décodage.

Ces constats de recherche et nos résultats comportent des implications pratiques : ils concourent à valoriser un enseignement explicite et intégré de la compréhension et du vocabulaire auprès de jeunes élèves. Comme Cèbe (1998, p. 102), nous sommes convaincue de la nécessité, pour les élèves de milieu populaire, « de faire l'expérience d'une utilisation rigoureuse du langage qui permet de sortir de l'expérience immédiate, [...], de mémoriser, rappeler, décrire, comparer, partager, faire des hypothèses, argumenter, justifier, résumer, se représenter ».

D'autres prolongements s'avèrent toutefois nécessaires. Les données montrent, par exemple, que les élèves du groupe A se distinguent de ceux du groupe B dans le test de lexique sur l'album. Pour mieux comprendre cet intéressant résultat, il serait bon, à l'instar de Picoche (1992) qui catégorise les mots de la langue en « mots actifs » et « mots passifs », d'examiner l'utilisation de ces mots en réception (compréhension) et en production en analysant les rappels de récit.

En sus d'une étude lexicale approfondie, cette recherche ouvre d'autres perspectives et questionnements. Les compétences développées en Grande Section se maintiennent-elles au Cours Préparatoire ? Sont-elles les mêmes pour les élèves initialement faibles que pour les élèves performants ? La réduction de l'écart entre des élèves de milieux contrastés peut-elle se poursuivre à l'école élémentaire ? Des effets sont-ils visibles en compréhension de texte entendu, en compréhension de texte lu en autonomie, en production écrite ? Autant de questions que de réjouissantes possibilités de recherches nouvelles...

Partie 4

Analyse qualitative

Chapitre 7

La transformation des acteurs : influence de l'utilisation d'un outil innovant et rôle de l'accompagnement

Chapitre 7

La transformation des
pratiques : influence de
l'utilisation d'un outil et rôle
de l'accompagnement

Chapitre 7

La transformation des pratiques : influence de l'utilisation d'un outil et rôle de l'accompagnement

La première étude (cf. chapitres 3 à 5) menée n'a pas révélé de différence sur les progrès des élèves selon que leur enseignant a été accompagné ou pas. Nous avons attribué ce résultat au mode de construction de l'outil *Narramus* qui associe, dès le départ, chercheurs et praticiens dans le processus de conception. Aussi cette étude apporte-t-elle un élément majeur dans le débat actuel touchant le soutien externe à apporter aux innovations : nos données prouvent que l'absence d'accompagnement n'est pas toujours un obstacle à la mise en œuvre fidèle et ne réduit pas l'efficacité de l'outil sur le versant des apprentissages enfantins. Toutefois, nous avons émis une troisième hypothèse : si l'accompagnement n'influence pas directement les résultats des élèves, il pourrait avoir un effet sur le développement professionnel des enseignants.

Pour la tester, nous avons réalisé deux enquêtes basées sur des questionnaires que nous avons proposés aux maîtres impliqués dans l'étude 1. Les données obtenues nous permettent de répondre à deux nouvelles questions de recherche portant sur : 1° le développement professionnel des enseignants : dans quelle mesure l'utilisation de *Narramus* transforme-t-elle les pratiques professionnelles et les connaissances dont les maîtres disposent ? et 2° l'accompagnement : y a-t-il un lien entre accompagnement et développement professionnel, si oui lequel ?

Ce chapitre, principalement descriptif et exploratoire, nous permettra de mieux comprendre les liens qui unissent outillage, accompagnement et développement professionnel. Nous avons voulu y donner la parole à nos compagnons de route - les maîtres impliqués dans notre étude - pour mieux comprendre ce que celle-ci avait produit chez eux. Si les traitements n'ont pas la rigueur des études précédentes, cette enquête (questionnaires, entretiens) apporte des données qualitatives intéressantes pour les formateurs d'enseignants. Nous l'avons organisée pendant deux années consécutives auprès de 194 enseignants et d'accompagnateurs. Elle est pour nous le début prometteur d'un travail à poursuivre.

1. Quelques considérations théoriques

Si l'intention finale – désormais validée – des auteurs de *Narramus*, est bien d'améliorer les performances et de réduire les écarts entre des élèves de milieux sociaux contrastés, c'est par l'outillage (les transformations des pratiques enseignantes) qu'ils espèrent avoir une action. C'est pourquoi nous avons voulu savoir quel est l'effet de l'instrumentation sur le développement professionnel des enseignants et quel est le rôle joué par l'accompagnement dans cette évolution.

1.1. Le développement professionnel

Avant de traiter cette question, nous devons définir ce que nous entendons par « développement professionnel ». Les recherches qui en traitent se distinguent d'abord par leurs finalités (Lefeuvre, Garcia, & Namolovan, 2009). La première, qualifiée d'« heuristique », rassemble les études qui permettent de comprendre et d'expliquer les comportements, les processus mis en œuvre et les ressources des enseignants (Huberman, 1995; Marcel, 2009; Vonk, 1988). La seconde, dite « pragmatique », groupe les études qui visent à transformer les pratiques. Les approches de la didactique professionnelle (Pastré, Mayen, & Vergnaud, 2006) poursuivent cet objectif. Mais il existe aussi, dans ce domaine, deux approches du développement professionnel : l'une « développementale », l'autre « professionnalisante » (Uwamariya & Mukamurera, 2005).

1.1.1. L'approche développementale

Cette approche s'inscrit dans une perspective heuristique. Le développement professionnel y est décrit de manière temporelle et individuelle. D'après les auteurs, il s'agirait d'un processus linéaire, jalonné par différentes étapes. Dans la lignée des travaux de Piaget pour lequel les individus se développent grâce aux expériences qu'ils font sur le monde, on considère que les enseignants transforment leurs pratiques en exerçant leur métier. Toutefois, les tenants de cette position ne s'accordent ni sur le contenu des étapes du développement professionnel ni sur leur ordre d'apparition. Par exemple, Vonk (1988) identifie une phase qu'il intitule « préprofessionnelle » présente dès le début de la formation, quand Huberman (1989) et Dreyfus (2004) situent le point de départ du développement professionnel au moment de l'entrée dans le métier. Cette approche qui, nous l'avons dit, est exclusivement centrée sur l'enseignant et les changements linéaires qu'il connaît, ne tient pas compte du rôle joué par

l'environnement de travail, le contexte et l'entourage professionnels. Aussi certains chercheurs la complètent-ils.

1.1.2. L'approche professionnalisante

Cette approche se situe dans une perspective socio-constructiviste (Uwamariya & Mukamurera, 2005). Le développement professionnel y est envisagé comme un processus d'apprentissage réflexif par lequel l'enseignant transforme ses pratiques, ses conceptions, ses valeurs... en fonction des personnes et des situations – professionnelles et personnelles – qu'il rencontre. Dans cette perspective, l'enseignant peut donc construire des savoirs professionnels, soit dans le cadre de dispositifs de formation initiale ou continue, soit dans l'action.

Dans notre étude, c'est l'introduction de l'outil innovant qui, selon la première approche, transformerait *de facto* l'activité des enseignants en proposant de nouvelles activités et manières de faire et qui amènerait, en retour, le développement de nouvelles connaissances et compétences. Selon la seconde, le collectif que nous avons créé (3 enseignants au moins par école impliquée) et l'accompagnement proposé seraient aussi à l'origine des changements.

1.2. L'accompagnement

Comme nous l'avons écrit dans le chapitre 1, proposer un soutien externe pour aider à la mise en œuvre d'un changement majeur est une idée fréquemment défendue dans la littérature scientifique (Bélanger et al., 2012; Charlier, Biémar, & Jorro, 2012; Kretlow & Bartholomew, 2010; Penneman et al., 2016; Rowan & Miller, 2007).

1.2.1. Considérations théoriques

L'accompagnement peut être envisagé comme un processus qui engage deux personnes, l'une accompagnatrice et l'autre accompagnée. Certains chercheurs (Gather Thurler, 2004; Lafortune, 2008 ; Savoie-Zajc, 2010 ; Wittorski, 2008) postulent qu'il devrait toujours s'adresser à un groupe et non à des individus isolés et sur une période relativement longue pour augmenter l'efficacité de la situation de travail.

Plusieurs principes sous-tendent cette vision :

- le soutien doit être centré sur la construction de connaissances et le développement des compétences des personnes accompagnées ;
- il doit favoriser les interactions, l'émergence et la résolution des conflits sociocognitifs provoqués par le changement demandé (conceptions, valeurs, pratiques antérieures) ;

- il doit être régulier pour favoriser l'action et la réflexion dans et sur l'action ;
- il doit engager l'ensemble des partenaires dans un processus de pratique réflexive ;
- il doit allier la théorie et la pratique dans la formation.

Pour intéressantes qu'elles soient, ces conceptions omettent une dimension importante de l'accompagnement qui sous-tend, selon nous, une partie de l'efficacité de l'accompagnement : les apports des formateurs sur des contenus pédagogiques et didactiques spécifiques qui assurent le lien entre la théorie et la pratique testée et participent, eux aussi, au développement professionnel.

Les dispositifs d'accompagnement effectivement mis en œuvre dans notre étude échappent à cette critique puisque, nous l'avons dit (*cf.* chapitre 3), les formateurs ont tous proposé, lors des rencontres avec les enseignants, des apports théoriques sur les cibles didactiques de l'outil et les principes pédagogiques sur lesquels il repose.

1.2.2. L'accompagnement dans le cadre de notre recherche

Nous avons vu dans le chapitre 3 qu'au cours de l'année 1 les accompagnements ont duré, en moyenne, 15,7 heures (écart-type 7,4) et que l'organisation et la mise en œuvre variaient sur le versant du temps alloué, du matériel support et du nombre d'accompagnateurs. Mais, dans toutes les circonscriptions, on observait sensiblement les mêmes invariants dans les contenus de la formation⁸⁴ :

- 1° une aide organisationnelle et logistique à la mise en œuvre de l'outil dans les classes ;
- 2° des apports théoriques sur les principes pédagogiques et les cibles didactiques de l'outil (le lexique, la théorie de l'esprit, les caractéristiques d'un bon « raconté »...) ;
- 3° des échanges de pratiques sur la mise en œuvre des scénarios, les réussites observées, les changements opérés, les difficultés rencontrées, et la présentation des résultats des élèves au fil des scénarios.

⁸⁴ Rappelons que, si les accompagnateurs étaient laissés libres d'organiser les rencontres, les chercheurs les ont tous préalablement formés.

2. Outiller et accompagner les enseignants : des vecteurs de développement professionnel ?

Pour répondre à cette nouvelle question de recherche, nous avons mené une troisième étude, plus qualitative pour savoir si et comment les enseignants tirent parti de la mise en œuvre de l'outil et de l'accompagnement pour développer de nouvelles pratiques. Le recueil de données a été réalisé, en deux temps, auprès des enseignants des deux groupes utilisant *Narramus* (G1 et G2).

2.1.1. Question de recherche n° 1 : quels sont les liens qui unissent la prise en main d'un outil et le développement professionnel ?

Nous avons reformulé cette question en deux sous-questions : 1° quelle conscience les maîtres ont-ils du fait que la prise en main de l'outil a modifié leurs pratiques ? 2° Ceux du G1 ont-ils une conscience plus aigüe de ce changement ? Si oui, on pourra déduire que l'accompagnement a un effet amplificateur. Dans le cas contraire, on conclura que notre hypothèse était la bonne : la mise en œuvre d'un outil bien conçu est un vecteur de développement professionnel.

2.1.1.1. Méthode

Nous avons élaboré un premier questionnaire titré « Mon retour suite à la mise en place de 3 scénarios » (cf. annexe 6) que nous avons adressé à tous les maîtres⁸⁵ du G1 (N = 115) et du G2 (N = 79) à la fin de la première année de la recherche (juin 2017). Il comprend cinq entrées : 1° mes points de réussite en classe, 2° mes gestes professionnels modifiés, 3° ce que je retiens de cette année, 4° mes belles inventions et 5° l'activité que je déteste. Le nombre de réponses aux questions n'était pas limité.

Nous avons d'abord affecté un numéro à chaque questionnaire et avons analysé les cinq items manuellement, par analyse du contenu avec une lecture flottante et une catégorisation (Bardin, 2007). Nous avons groupé les données par une démarche « ascendante » : il s'est agi d'abord de construire des catégories à partir du matériau brut dont nous disposions pour regrouper les réponses des enseignants. Cette analyse de contenu a été réalisée selon un double codage à l'aveugle⁸⁶. Les désaccords entre les deux codeuses, peu nombreux, ont été discutés et résolus : ils portaient sur les titres donnés à certaines catégories (une codeuse utilisait, par

⁸⁵ Quand deux enseignants se partageaient le temps de classe (temps partiel, décharge de direction...), c'est le titulaire qui était chargé de renseigner le questionnaire.

⁸⁶ Nous remercions chaleureusement Claire Ruffié, étudiante en Master 2 MEEF (formation continue diplômante).

exemple, le terme « engagement », l'autre, le terme « implication ») et sur la distinction de sous-catégories (une codeuse avait groupé dans une seule et même catégorie toutes les réponses évoquant le « plaisir », que celui-ci soit éprouvé par les élèves, les enseignants eux-mêmes ou les parents quand l'autre avait créé trois catégories. Nous avons, ensemble, opté pour la formule la plus brève et la plus explicite.

Nous avons complété ce recueil par le traitement et l'analyse de différents documents émanant de plusieurs sources⁸⁷:

- des enregistrements vidéos (3 heures) réalisés par une conseillère pédagogique de l'académie de Lyon ; sept enseignants ont accepté d'être filmés sur les temps d'accompagnement-bilans mis en œuvre après chaque scénario lors de la première année d'utilisation de l'outil⁸⁸.
- Des enregistrements audio (10 heures) réalisés dans l'académie de Clermont-Ferrand : les temps d'analyse de pratiques proposés à Montluçon (une école) et Clermont-Ferrand (trois écoles) ont été enregistrés.
- Des témoignages filmés (30 minutes) d'enseignants du G1 et du G2 réalisés dans l'académie de Lille⁸⁹ en septembre 2017, soit après un an de mise en œuvre de l'outil dans les classes.
- Des témoignages spontanés des enseignants et des accompagnateurs que nous avons reçus au cours des trois années de la recherche.

2.1.1.2. Résultats

Un taux de réponses très important

Quatre-vingt-quinze questionnaires ont été retournés par les enseignants du G1 (soit un taux de réponse de 82,61 %) et quarante-quatre par ceux du G2 (55,70 %). Les maitres des classes impliquées dans l'étude ont donc massivement répondu à notre sollicitation. Pour expliquer l'écart inter-groupe touchant le nombre de retours, nous pouvons faire deux hypothèses complémentaires. La première a trait au mode de recueil des réponses au questionnaire. Pour les enseignants du G1, ce sont les formateurs qui, lors du bilan réalisé à la fin de la première année, les ont recueillies quand ceux du G2 ont eu à les envoyer par courriel. La seconde est liée aux caractéristiques de l'accompagnement lui-même : au fil des rencontres, les maitres du

⁸⁷ Les enregistrements ont tous été réalisés avec le consentement des formateurs et des enseignants.

⁸⁸ Nous remercions chaleureusement Murielle Da Silva et les enseignants qui ont accepté de nous fournir ce matériau.

⁸⁹ Un grand merci aux enseignantes qui ont participé et à Sylvain Hannebique pour son investissement.

G1 ont établi une relation de confiance avec leur accompagnateur et sans doute ont-ils été soucieux de respecter le contrat qui les liait aux chercheurs et aux formateurs. Ces derniers se sont montrés régulièrement présents et ont répondu aux sollicitations exprimées par les enseignants : on peut donc penser que les maîtres ont pris, en retour, le temps de répondre aux questionnaires.

Les éléments saillants des réponses au premier questionnaire

Le tableau 71 indique les points de réussite identifiés par les enseignants après une année d'utilisation de l'outil. La colonne de gauche comporte les retours du G1 et celle de droite, ceux du G2 ; les catégories présentes dans les deux groupes sont signalées par des couleurs identiques.

Tableau 71. Les points de réussite identifiés par les enseignants du G1 et du G2 (questionnaire 1)

Les points de réussite en classe					
Enseignants du G1 N = 95			Enseignants du G2 N = 44		
Catégories	Nombre de citations dans la catégorie	Nombre de citations en %	Catégories	Nombre de citations dans la catégorie	Nombre de citations en %
1° utilisation de supports variés	N = 77	80,2 %	1° enseignement du lexique	N = 28	63,6 %
2° enseignement du lexique	N = 73	76,04 %	2° utilisation de supports variés	N = 26	59,1 %
3° progrès des élèves en lexique	N = 59	61,5 %	3° engagement des élèves	N = 23	52,3 %
4° engagement des élèves	N = 53	55,2 %	4° progrès des élèves en lexique	N = 19	43,2 %
5° progrès des élèves en compréhension	N = 27	28,1 %	5° progrès des élèves en narration	N = 13	29,5 %
6° progrès des élèves en narration	N = 23	24,0 %	6° progrès des élèves en compréhension	N = 6	13,6 %
7° relations école-familles	N = 16	16,7 %	7° instauration d'habitudes de travail	N = 3	6,8 %
8° instauration d'habitudes de travail	N = 14	14,6 %	8° meilleure écoute, meilleure attention	N = 3	6,8 %
9° projet collectif fédérateur	N = 12	12,5 %	9° projet collectif fédérateur	N = 2	4,5 %
10° mise en place d'un coin écoute des histoires	N = 12	12,5 %	10° relations école-familles	N = 2	4,5 %

La lecture du tableau fait apparaître, qu'après une année d'utilisation, les enseignants dressent une longue liste de réussites ce qui peut, sans doute, être interprété comme la marque du développement d'un sentiment de développement et de compétence professionnels. On observe

aussi que les items utilisés par les enseignants du G1 sont quasiment identiques à ceux identifiés par le G2, quatre catégories étant citées par la moitié d'entre eux (dans les deux groupes) : la variété des supports utilisés, l'enseignement du lexique, l'engagement des élèves et leurs progrès en lexique (dans une moindre mesure pour le G2). Autrement dit, le fait d'être accompagné dans la mise en œuvre n'influe pas sur l'analyse que font les enseignants de leurs réussites. Ils les attribuent aux caractéristiques de l'outil lui-même (l'enseignement du lexique, la variété des supports et des activités) et par les effets que celles-ci produisent sur leurs élèves (engagement et progrès). Pour rendre ces résultats plus parlants, nous avons analysé les verbatim dont nous disposons.

Analyse des verbatim

La plupart des maîtres évoquent une profonde transformation de leurs pratiques sur le versant de l'enseignement du lexique. Ces changements concernent la fréquence de l'enseignement, la nature des supports utilisés (images, gif animés, vidéos, accompagnement audio, boîte « mémoire des mots », coin lexique...), l'utilisation du corps avec le passage par le mime, les activités de réinvestissement systématiquement proposées (mimes, jeux de loto, de *memory*, devinettes...).

Beaucoup soulignent que cet enseignement rigoureux a profité à la plupart de leurs élèves, nos données le confirment (cf. chapitre 4). L'une l'explique ainsi « *les élèves savent ce qu'ils font et apprennent, ils sont rentrés dans une espèce de routine : l'affichage et la boîte mémoire des mots, (même virtuelle) fonctionnent bien* ». Certains disent avoir transformé certaines de leurs pratiques et/ou leurs conceptions de l'apprentissage. Par exemple, touchant le choix des mots à enseigner, une enseignante déclare : « *avant, je proposais des mots plus simples, dans Narramus, c'est plus ambitieux, mais les élèves s'en emparent* » ; une autre : « *le vocabulaire des émotions et des sentiments aident les élèves à comprendre certains passages de l'histoire* ». Ils extraient de leur usage un certain nombre de « règles » sur ce qu'ils considèrent comme étant de bonnes méthodes et justifient leur point de vue : « *il faut expliciter au maximum* » ; « *le côté systématique de l'enseignement du lexique : quand on présente des mots nouveaux et qu'on les révise, les élèves s'engagent plus* » ; « *On voit que les élèves sont heureux de mimer et que ça les aide pour se souvenir* ».

On note également que beaucoup soulignent les progrès faits par les élèves « petits parleurs », peu performants en langage et/ou non francophones comme l'illustre le verbatim du second bilan de l'année 1 réalisé dans une école de l'Ain.

E1⁹⁰ - « Ça a mis en valeur les enfants qui étaient non francophones. Ils n'avaient qu'une envie, c'était vraiment de raconter cette histoire et ils avaient des étoiles dans les yeux.

E2 - Chez les petits parleurs, c'était ça.

E3 - Pas forcément que chez les petits parleurs... Ils se sont pas du tout sentis gênés parce que le vocabulaire n'était quand même pas facile eh bien je les ai pas sentis gênés. Au contraire, ils avaient qu'une envie c'était de raconter, ils étaient toujours les premiers à lever le doigt pour participer, pour jouer au théâtre. J'ai trouvé ça super.

E1- Moi je trouve que les non francophones justement ils ont bien retenu le vocabulaire parce qu'on l'a tellement répété, répété et après même il y en a certains qui ont essayé de raconter avec le vocabulaire parce que c'était nouveau mais ils avaient compris alors...

E4 - Oui, beaucoup d'intérêt ; et au coin écoute ils étaient particulièrement bien sensibles. Parce que l'histoire, ils la maîtrisaient ; parce que généralement quand tu mets les enfants au coin écoute les premiers temps, c'est pas ça ! Tandis que là, ils étaient dedans. »

Les progrès présentés dans le chapitre 4 valident les observations des enseignants : les élèves les moins performants ont également profité de l'enseignement avec *Narramus*.

Les enseignants disent aussi que l'engagement dans un collectif de travail a produit une dynamique nouvelle marquée par de fréquents échanges de pratiques en dehors des temps dévolus à l'accompagnement. Ils apprécient « d'avoir du temps pour travailler en groupe sur le scénario », « la mise en place des stratégies d'équipe pour approfondir, échanger, dynamiser le travail », « le travail avec les maitres E et les ATSEM », « le plaisir de travailler tous ensemble sur un même album ou une même méthode ».

Pour conclure, on peut noter que les exigences et les ambitions des enseignants ainsi que leur regard sur les élèves ont été modifiés par l'utilisation de l'outil. Les enseignants du G1 sont plus nombreux à évoquer les transformations qui concernent l'enseignement du lexique, l'utilisation de supports variés, les progrès et l'implication des élèves. Nous ne pouvons pas affirmer quelles en sont les raisons exactes mais le contenu des accompagnements a pu influencer ces différents points. Nous mettrons à l'épreuve cette interprétation en nous appuyant sur les informations données par le second questionnaire.

⁹⁰ E1, E2... désignent les différents enseignants qui participent à l'échange.

Les gestes professionnels modifiés

Le deuxième item de notre enquête invitait les enseignants du G1 et du G2 à s'interroger sur la modification des gestes professionnels que la mise en œuvre de *Narramus* avait provoquée. Leurs réponses sont synthétisées dans le tableau 72 (le nombre de réponses n'était pas limité).

Tableau 72. Les gestes professionnels modifiés en fonction du groupe (questionnaire 1)

Les gestes professionnels modifiés					
Enseignants du G1 N = 95			Enseignants du G2 N = 44		
Catégories	Nombre de citations dans la catégorie	Nombre de citations en %	Catégories	Nombre de citations dans la catégorie	Nombre de citations en %
1° utilisation de supports variés dans toutes les activités	N = 57	60 %	1° utilisation de supports variés dans toutes les activités	N = 28	63,6%
2° attention accordée à l'enseignement du lexique	N = 46	48,4 %	2° nouvelle approche d'enseignement de la compréhension	N = 22	50 %
3° nouvelle approche d'enseignement de la compréhension	N = 37	38,9 %	3° attention accordée à l'enseignement du lexique	N = 17	38,6 %
4° des activités cognitivement exigeantes	N = 27	28,4 %	4° des activités cognitivement exigeantes	N = 14	31,8 %
5° mettre en scène les histoires étudiées	N = 30	31,6 %	5° mettre en scène les histoires étudiées	N = 13	29,5 %
6° soigner les narrations	N = 24	25,3 %	6° proposer un enseignement explicite	N = 3	6,8 %
7° proposer un enseignement explicite	N = 10	10,5 %	7° soigner les narrations	N = 1	2,3 %
8° meilleure perception des obstacles à la compréhension	N = 2	2,1 %	8° meilleure perception des obstacles à la compréhension	N = 1	2,3 %

Là encore, G1 et G2 se distinguent très peu : les gestes professionnels transformés par l'outillage sont les mêmes d'un groupe à l'autre (*cf.* tableau 72). On peut donc raisonnablement conclure que l'outil, même quand sa mise en œuvre n'est pas accompagnée, 1° transforme les

gestes professionnels et que 2° la plupart des maitres en sont conscients, les G1 comme les G2. Ce résultat nous semble majeur et conforte notre analyse du chapitre 1 : il est souhaitable d'anticiper la future diffusion d'un outil dès le début du processus de conception. Si les gestes ne sont modifiés qu'avec un accompagnement, une diffusion à plus large échelle du dispositif est inenvisageable. Touchant *Narramus*, puisque les enseignants déclarent avoir transformé leurs pratiques habituelles, nous pouvons raisonnablement penser que la diffusion (hors recherche) n'en sera que plus facile.

Plus spécifiquement, on notera que, parmi les trois gestes le plus souvent cités, deux étaient déjà placés en tête sur le premier item (« points de réussites ») : l'utilisation de supports variés dans toutes les activités et l'attention accordée à l'enseignement du lexique. La catégorie la plus saillante correspond à la variété des supports utilisés dans *Narramus* pour enseigner la compréhension, la narration et le lexique (utilisation du numérique, du corps, de maquettes, de boites à mots...). Elle apparaît comme une transformation en profondeur chez plus de 60 % des enseignants des deux groupes.

Dans les verbatim extraits des rencontres avec leurs accompagnateurs, nous avons noté que les maitres du G1 sont très au clair avec l'idée que cette nouvelle méthode leur a permis de tester d'autres manières d'enseigner la compréhension (catégorie « nouvelle approche d'enseignement de la compréhension »). Ils relèvent, par exemple, qu'en faisant étudier moins d'albums et en les travaillant sur une longue durée, ils gagnent en qualité dans l'exploitation du texte. Ils sont aussi plus centrés sur l'activité de compréhension et de narration que sur la mise en lien avec les autres domaines d'apprentissage ou d'autres textes. Ils rapportent aussi avoir compris l'intérêt d'une entrée textuelle sans recours à l'image.

Les belles inventions

La troisième entrée de notre enquête visait à nous informer sur les modifications (« belles inventions ») que les enseignants du G1 et du G2 avaient introduites. Leurs réponses sont synthétisées dans le tableau 73. Là encore, le nombre de réponses attendues n'était pas limité.

Tableau 73. Les belles inventions des enseignants du G1 et du G2 (questionnaire 1)

Les belles inventions					
Enseignants du G1 N = 95			Enseignants du G2 N = 44		
Catégories	Nombre de citations dans la catégorie	Nombre de citations en %	Catégories	Nombre de citations dans la catégorie	Nombre de citations en %
1° fabrication du matériel pédagogique (maquettes, marottes, masques, déguisements...)	N = 43	45,3 %	1° fabrication du matériel pédagogique (maquettes, marottes, masques, déguisements...)	N = 15	34,1 %
2° création de supports collectifs ou individuels	N = 20	21,05 %	2° création de supports collectifs ou individuels	N = 5	11,4 %
3° création de jeux	N = 8	8,4 %	3° prolongements en arts plastiques	N = 5	11,4 %
4° création de scénarios sur d'autres albums	N = 8	8,4 %	4° création de scénarios sur d'autres albums	N = 5	11,4 %
5° prolongements en arts plastiques	N = 6	6,3 %	5° théâtre sur d'autres albums	N = 2	4,5 %
6° création de coins dédiés au projet sur l'album	N = 5	5,3 %	6° travail sur la reconnaissance de mots	N = 1	2,3 %

À la lecture du tableau, on peut faire l'hypothèse que le fait d'utiliser un outil didactique très détaillé libère du temps de préparation que les enseignants utilisent pour créer du matériel (maquettes, marottes...), des outils individuels ou collectifs, des jeux pour imaginer des activités de transfert...

Ce résultat vient apporter un résultat important au débat qui oppose les formateurs touchant l'importance (ou le danger) d'outiller les enseignants en formation. Certains, en effet, soutiennent qu'il ne faut pas donner « de recettes, de trucs, d'outils clés-en-main » au motif qu'ils enfermeraient les enseignants dans un rôle d'exécutants. Nos données semblent indiquer que « l'outil leur donne, au contraire, la liberté d'action en leur offrant une base solide sur laquelle ils peuvent être inventifs. On peut faire l'hypothèse qu'il libère des ressources attentionnelles que les maîtres peuvent consacrer aux élèves, à leur fonctionnement et leurs apprentissages » (Cèbe, 2009, p. 7).

Les activités détestées

Avec la quatrième entrée de notre questionnaire, nous voulions savoir quelles étaient les activités proposées dans la méthode que les enseignants n'approuvaient pas soit qu'elles étaient trop difficiles à guider, soit qu'elles entraient en contradiction avec leurs manières de faire habituelles ou leurs conceptions. Leurs réponses sont classées dans le tableau 74.

Tableau 74. Les activités détestées par les enseignants du G1 et du G2 (questionnaire 1)

Activités que je déteste					
Enseignants du G1 N = 95			Enseignants du G2 N = 44		
Catégories	Nombre de citations dans la catégorie	Nombre de citations en %	Catégories	Nombre de citations dans la catégorie	Nombre de citations en %
1° aucune	N = 34	35,8 %	1° aucune	N = 16	36,4 %
2° utilisation de l'informatique	N = 27	28,4 %	2° lassitude à répéter certaines activités (vocabulaire, narration)	N = 10	22,7 %
3° lassitude à répéter certaines activités (vocabulaire, narration)	N = 20	21,1 %	3° travailler sur les états mentaux difficile en PS	N = 4	9,1 %
4° travailler sur les états mentaux difficile en PS	N = 11	11,6 %	4° fabriquer les maquettes, les décors	N = 3	6,8 %
5° contraintes de temps (outil chronophage)	N = 11	11,6 %	5° contraintes de temps (outil chronophage)	N = 3	6,8 %
6° théâtre	N = 7	7,4 %	6° utilisation de l'informatique	N = 2	4,5 %
			7° théâtre	N° = 2	4,5 %
			8° représenter par le dessin	N° = 1	2,3 %

Comme on peut le voir dans le tableau 74, un tiers des enseignants indique ne pas avoir détesté d'activité, et ce dans les deux groupes. On peut noter que les contraintes liées à l'utilisation de l'informatique, le côté chronophage, les activités théâtrales et la lassitude liée à la répétition de certaines activités se retrouvent dans les deux groupes.

Ce que les enseignants retiennent après une année

La dernière entrée portait sur les éléments marquants retenus après une année d'utilisation de l'outil. Le tableau 75 présente la catégorisation que nous avons faite des réponses.

Tableau 75. Ce qu'ont retenu les enseignants du G1 et du G2 (questionnaire 1)

Ce que je retiens de cette année					
Enseignants du G1 N = 95			Enseignants du G2 N = 44		
Catégories	Nombre de citations dans la catégorie	Nombre de citations en %	Catégories	Nombre de citations dans la catégorie	Nombre de citations en %
1° développement professionnel	N = 34	35,8 %	1° développement professionnel	N = 17	32,8 %
2° utilisation de supports variés	N = 33	34,7 %	2° progression intéressante, méthode efficace	N = 14	31,8 %
3° progression intéressante, méthode efficace	N = 33	34,7 %	3° difficultés de mise en œuvre	N = 12	27,3 %
4° progrès des élèves	N = 32	68,4 %	4° chronophage	N = 12	27,3 %
5° plaisir, envie, motivation des enseignants	N = 25	26,3 %	5° engagement des élèves	N = 11	25 %
6° engagement des élèves	N = 25	26,3 %	6° progrès des élèves	N = 10	22,7 %
7° difficultés de mise en œuvre	N = 17	17,89 %	7° utilisation de supports variés	N° = 9	20,5 %
8° chronophage	N = 11	11,6 %	8° relations positives avec les parents	N° = 5	11,4 %

La catégorie la plus citée par les enseignants est le développement professionnel. Résultat intéressant pour notre propos, on ne note pas de différence dans le taux de réponses en fonction du groupe (G1 : 35,8 % et G2 : 32,8 %). Autrement dit, la méthode est, en elle-même, un vecteur de développement professionnel pour un tiers des enseignants et ils en ont conscience qu'ils soient accompagnés ou pas.

L'intérêt d'utiliser des supports variés, la pertinence de la méthode, les progrès des élèves et l'engagement des élèves sont cités aussi bien par les G1 que les G2. Certains enseignants du G1 (11,6 %) et du G2 (27,3 %) regrettent toutefois l'aspect chronophage et les contraintes (souvent techniques) liées à la mise en œuvre des scénarios (G1 : 17,89 % et G2 : 27,3 %).

Pour distinguer les évolutions liées à l'outil lui-même de celles qui dépendent de l'accompagnement, nous avons élaboré un second questionnaire que les enseignants ont renseigné à la fin de la deuxième année de l'étude (2018).

2.1.2. Question de recherche n° 2 : y a-t-il un lien entre accompagnement et développement professionnel, si oui lequel ?

Dans la mesure où le premier questionnaire (juin 2017) portait seulement sur l'analyse faite par les utilisateurs eux-mêmes de la mise en œuvre de l'outil, nous avons poursuivi notre enquête. Nous voulions savoir 1° sur quels aspects précis de la méthode les enseignants avaient modifié leurs pratiques après deux années d'utilisation (6 scénarios) et 2° comment les enseignants du G1 appréciaient l'accompagnement dont ils avaient bénéficié pendant deux ans comparativement aux G2 qui n'en avaient profité qu'un an.

2.1.2.1. Méthode

Nous avons de nouveau construit un questionnaire (*cf.* annexe 7) que nous avons adressé aux maîtres à la fin de la deuxième année de la recherche (juin 2018). Nous avons traité 102 réponses (60 pour les G1, 42 pour les G2⁹¹) et les enregistrements audio et vidéo communiqués par les formateurs. Les réponses ont été analysées un peu différemment des premières. En effet, dans la mesure où le questionnaire comportait des questions fermées à choix multiples et des questions ouvertes, nous avons utilisé, pour les secondes, l'analyse de contenu citée précédemment et avons complété les traitements avec le logiciel de statistiques textuelles IRaMuteQ (version libre) et logiciel SPSS.

2.1.2.2. Résultats : la modification des pratiques

Nous présentons d'abord les résultats à la première question qui invitait les enseignants à se prononcer sur leur degré d'accord (non ; oui, un peu ; oui, beaucoup) touchant les changements observés dans leurs pratiques sur 17 caractéristiques de la méthode (raconter, faire raconter, prévoir l'illustration, centrer l'attention des élèves sur les états mentaux des personnages...).

On trouve, dans le tableau 76, les réponses de tous les enseignants (G1 et G2) que nous n'avons pas distingués puisque les résultats du premier questionnaire nous avaient montré qu'ils ne se différenciaient pas sur cette question. Les cases grisées correspondent aux réponses formulées par plus de 50 % des enseignants.

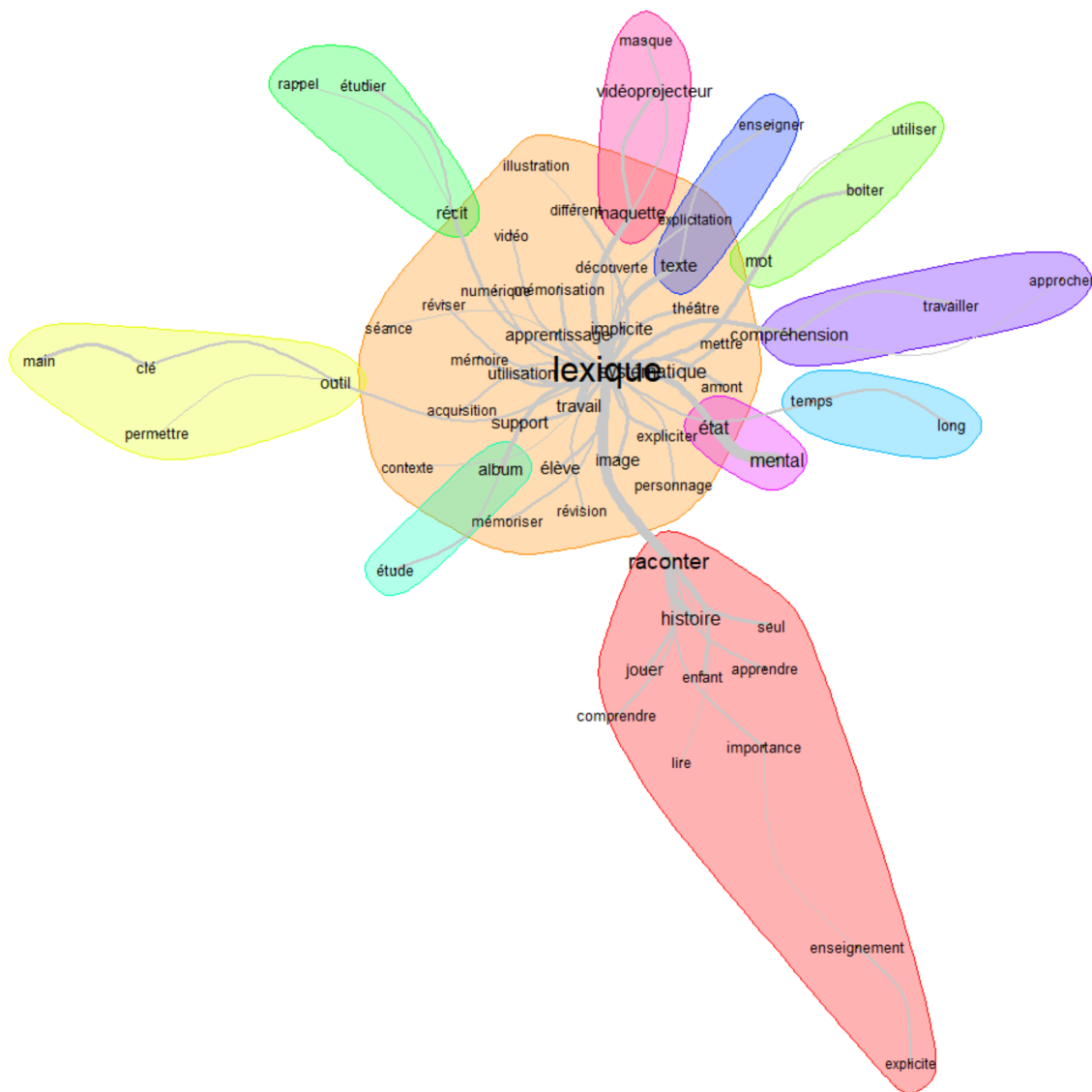
⁹¹ Rappelons que, l'année 2, tous les enseignants (anciens G1 et G2) ont bénéficié de temps d'accompagnement centrés sur l'analyse de pratiques.

**Tableau 76. Les modifications des pratiques liées à l'utilisation de l'outil
(en % de réponses citées)**

	Caractéristiques didactiques ou pédagogiques	non	oui, un peu	oui, beaucoup
1	Raconter (sans le texte et l'image)	8,5 %	46,2 %	45,3 %
2	Faire raconter aux élèves un récit complet	9,4 %	32,5 %	58,1 %
3	Faire raconter à un seul élève et les autres complètent à la fin du récit (1 puis tous)	10,3 %	49,1 %	40,5 %
4	Étudier un récit sur un temps long	20,7 %	33,6 %	45,7 %
5	Lire sans montrer les illustrations (dissocier texte-image)	16,5 %	34,8 %	48,7 %
6	Faire prévoir l'illustration	11,2 %	40,5 %	48,3 %
7	Enseigner le lexique avant la découverte du texte	7,7 %	35,0 %	57,3 %
8	Insister sur la mise en mémoire du vocabulaire	4,3 %	25,00 %	70,7 %
9	Réviser le lexique de manière systématique	5,1 %	29,9 %	65,0 %
10	Utiliser différents supports pour enseigner le lexique (images, vidéos, mimes...)	10,3 %	38,5 %	51,3 %
11	S'intéresser aux états mentaux des personnages	6,0 %	39,3 %	54,7 %
12	Utiliser une maquette	18,8 %	31,6 %	49,6 %
13	Faire jouer l'histoire	9,5 %	47,4 %	43,1 %
14	Utiliser le vidéoprojecteur en compréhension	17,2 %	28,4 %	54,3 %
15	Expliciter davantage	7,8 %	24,1 %	68,1 %
16	Utiliser le même outil (même principes, même déroulement) que les collègues	20,9 %	33,9 %	45,2 %
17	Être attentif.ve aux spécificités des problèmes de compréhension posés par les albums choisis	8,5 %	38,5 %	53,0 %

L'analyse des réponses donne une idée de la manière dont les enseignants évaluent leur développement professionnel. La lecture de ce tableau fait apparaître un résultat majeur : pour la plupart des maîtres, *Narramus* est un outil innovant, puisque sur les 17 items proposés, le pourcentage maximum d'enseignants jugeant ne pas avoir modifié leur pratique s'élève à 20,9 % (item 16). Les aspects sur lesquels les changements sont le plus criants (> à 50 %) sont : faire raconter un récit complet, enseignement du lexique (4 items), s'intéresser aux états

Figure 28. Graphe des mots associés aux caractéristiques de l'outil qui ont profondément influencé les pratiques des enseignants



Cette figure illustre que les enseignants font beaucoup de liens entre les différentes caractéristiques de *Narramus*. On peut en déduire qu'ils traitent l'outil comme un tout et non comme de multiples activités isolées.

Les résultats présentés indiquent que l'outil transforme les pratiques enseignantes, qu'il permet de développer des gestes professionnels et que les maîtres ne l'ignorent pas. Mais ces transformations et cette prise de conscience sont-elles les mêmes dans les deux groupes (G1 et G2) ?

2.1.3. L'accompagnement des enseignants : un accélérateur du développement professionnel ?

Pour discerner les différences, si différences il y a entre G1 et G2, nous nous penchons sur les données recueillies grâce au deuxième volet du second questionnaire. Nous nous limitons ici à traiter deux aspects : 1° les bénéfices de l'accompagnement que les enseignants déclarent et 2° le transfert des pratiques d'enseignement éprouvées dans *Narramus* à d'autres situations ou domaines d'apprentissage. Ici encore, nous donnerons corps aux réponses en les liant aux verbatim dont nous disposons.

2.1.3.1. L'accompagnement : une aide à la prise en main de l'outil ?

La grande majorité des enseignants du G1 (87,5 %) situe la plus forte utilité de l'accompagnement au début du projet, lors de la mise en place du premier scénario : « *Cet accompagnement a permis de comprendre la démarche d'un point de vue théorique tout en décortiquant chaque scénario d'un point de vue méthodologique (la succession des modules et le déroulement des activités)* ».

On découvre, dans les verbatim, que la prise en main de l'outil a contraint les maîtres à repenser le pilotage de leur classe : ils disent avoir réorganisé leur emploi du temps en réservant des plages spécifiquement dédiées à *Narramus* et ils ne l'ont pas fait de la même manière selon leur niveau de classe. Si les enseignants de Grande Section proposent, par exemple, une séance unique dévolue au lexique et à la découverte du texte du jour, les enseignants de Petite Section préfèrent dissocier ces deux activités et les proposer en deux temps distincts sur la journée. Ces choix sont d'autant plus difficiles qu'on participe à une recherche et qu'on n'ose pas changer la prescription des auteurs. Aussi les formateurs ont-ils beaucoup accompagné les équipes du G1 pour qu'ils s'autorisent des aménagements qui tiennent compte des capacités et des besoins de leurs élèves.

2.1.3.2. L'accompagnement : une aide à la compréhension des bonnes raisons que l'on a de faire ce que l'on nous demande de faire

Quand les enseignants sont interrogés sur la plus-value de l'accompagnement, leurs réponses peuvent être groupées en cinq catégories que nous présentons de la plus citée à la moins évoquée. Une fois encore, les verbatim nous servent à illustrer les réponses des maîtres.

Les réponses des enseignants sur les plus-values de l'accompagnement sont illustrées par la figure 29.

Figure 29. Nuage de mots représentant les plus-values attribuées à l'accompagnement par les enseignants



Notre analyse fait apparaître que l'accompagnement a permis aux enseignants :

1° d'être informés des travaux de recherche qui fondent les principes didactiques et pédagogiques de *Narramus* et de mieux comprendre les ambitions des auteurs (43,6 % des réponses) : il fait « *mieux comprendre les fondements théoriques, alerte sur les points emblématiques à ne pas oublier, donne du sens aux activités menées avec les élèves* ». Il aide aussi à « *adopter une attitude réflexive sur ses pratiques* » ; « *la recherche a été l'occasion d'expérimenter de nouvelles pratiques, d'articuler la théorie que nous apportait le conseiller pédagogique et la pratique* » ; « *mieux se saisir de l'outil et y donner du sens, l'explicitation aux enseignants est aussi importante que pour les élèves !* » ; « *la formation théorique que d'habitude je trouve ennuyeuse et qui pour la première fois de ma carrière avait du sens* », « *les apports théoriques (nous en avons eus sur le lexique et la compréhension) permettent de s'approprier beaucoup plus intelligemment l'outil et aident à le maîtriser un peu plus, être plus à l'aise et plus naturel : ça m'a permis d'improviser et de rebondir sur les réactions des élèves, de moins coller au scénario, d'imaginer d'autres tâches* ».

2° De « parler métier » en mettant en commun les expériences partagées (d'où l'intérêt d'un outil commun à l'école) en présence d'un formateur (21,8 % des réponses). Globalement, les échanges permettent de « *revenir sur les modalités de mise en œuvre* », d'« *analyser*

collectivement des pratiques » sur l'enseignement du langage et de la compréhension, de « *créer une culture d'équipe* » ; « *c'est rassurant de confronter nos difficultés et nos réussites aussi* » ; « *mobiliser le collectif d'école dans l'utilisation d'une méthode mais aussi et surtout dans les échanges et les analyses du travail* » ; « *être ensemble, avoir du temps partagé entre collègues met beaucoup d'enthousiasme. Cela devient une affaire collective : cela mobilise et encourage* » ; « *les échanges avec les collègues dédramatisent l'engagement au début* ».

3° D'aider à la prise en main de l'outil (18,2 % des réponses) : le soutien matériel et logistique a été souligné. L'accompagnement a permis de « *régler les problèmes techniques liés au vidéoprojecteur* », de « *réorganiser la journée de classe* ».

4° De rendre persévérant (12,7 % des réponses). : l'accompagnement « *insufflé de l'énergie au démarrage, rassure, mobilise, soutient* ». Certains enseignants insistent sur cet aspect « *nous avons été très chanceux* », « *je ne sais pas si je me serais sentie si impliquée au long cours sans accompagnement* ».

5° Diffuser les résultats de la recherche touchant les progrès des élèves (3,6 % des réponses).

Les données que nous avons extraites de ce questionnaire sont intéressantes à plus d'un titre. On voit d'abord que si l'accompagnement n'avait pas d'effets visibles sur les progrès des élèves, il en a un sur la manière dont les enseignants apprécient de travailler ensemble et avec leurs accompagnateurs autour d'un projet commun ambitieux.

Ces résultats ont donc de quoi rassurer les formateurs sur le rôle qu'ils peuvent jouer dans la transformation des pratiques. Cette analyse rejoint celle d'autres chercheurs : « Les outils pédagogiques, combinés à un travail collaboratif intense au niveau local, agissent comme des médiateurs dans l'appropriation à la fois individuelle et collective de pratiques nouvelles, si toutefois un leadership pédagogique exercé par un acteur légitime oriente leur exploitation. » (Lator et al., 2016, p. 53). Le rôle de l'accompagnateur, souvent médiateur et organisateur des échanges, apparaît ici primordial.

Mais nous avons ici un autre résultat, pour nous, majeur : ce sont les apports théoriques (didactiques et pédagogiques) que les enseignants plébiscitent, ce qui, on l'admettra, n'est pas chose fréquente. Notre expérience de la formation initiale et continue et nos lectures nous ont, en effet, enseigné qu'ils ont plutôt pour habitude de critiquer les formations trop théoriques, trop éloignées de leurs préoccupations et de leurs besoins. Ici au contraire, une formation théorique semble répondre à leur besoin de compréhension de l'action innovante.

Mais cette meilleure connaissance des résultats de recherche et des théories qui sous-tendent un apprentissage efficace est-elle suffisamment solide pour permettre aux enseignants du G1 de transférer leurs pratiques nouvelles dans d'autres domaines ?

2.1.3.3. L'accompagnement favorise-t-il le transfert des pratiques ?

Nous l'avons dit plus haut les enseignants reconnaissent que l'accompagnement (notamment le retour aux théories sous-jacentes) leur a permis de mieux comprendre les pratiques proposées et leur efficacité. On peut donc faire l'hypothèse qu'ils seront plus capables de transférer ces pratiques dans d'autres domaines d'apprentissage. On sait, en effet, que la compréhension est au cœur du processus de transfert : sans elle, tout transfert est impossible.

Pour traiter cette question, nous avons demandé aux enseignants des deux groupes (G1 et G2) de répondre à la question suivante : « avez-vous transposé des caractéristiques de *Narramus* dans d'autres domaines d'apprentissage. Si oui, lesquels ? »

Plus de quarante pour cent (41,2 %) des enseignants déclarent l'avoir fait dans des domaines et des activités variés (manière d'enseigner le lexique dans d'autres domaines ; utilisation d'images, de vidéos et du corps dans l'apprentissage des poésies, des comptines et des chants ; expliquer le lexique avant la découverte de l'histoire ; expliquer le but de l'apprentissage aux élèves ; utiliser des flashcards en mathématiques pour mémoriser les formes géométriques, en sciences pour le nouveau vocabulaire ; insister sur l'effort de mise en mémoire ; expliquer l'implicite ; mettre en place des logos ; expliciter davantage ; utiliser le numérique ; mettre en place la boîte « mémoire des mots » pour d'autres apprentissages). Le transfert apparaît aussi dans les verbatim : « *ce que je fais avec Narramus, je le fais en sciences pour les nouveaux mots. Je les explicite plus et utilise des images ou des vidéos* » ; « *dans l'école, on est beaucoup plus attentive au vocabulaire. On le réutilise dans la cour de récréation, on fait attention aux mots dans les histoires et on essaie de repérer ceux que les élèves ont appris dans d'autres albums* ».

Toutefois, le traitement disjoint des réponses des deux groupes (G1 et G2) apporte des indices corroborant notre hypothèse : l'accompagnement apparaît bien comme un levier amplificateur du transfert puisque 66,7 % des enseignants du G1 déclarent avoir transposé des principes, des activités ou des tâches ; ils ne sont que 33,3 % en G2.

Ce résultat conforte les choix d'ingénierie que nous avons faits :

- 1) offrir aux accompagnateurs impliqués dans notre recherche une formation à l'outil qui intègre les mêmes composants que ceux qui ont présidé à sa conception ;
- 2) confier l'accompagnement à des formateurs proches du terrain et des pratiques des maîtres.

Conclusion

Dans ce chapitre, nous nous sommes centrée sur le point de vue des enseignants impliqués dans notre étude, à la fois sur les pratiques - et leurs transformations - et l'accompagnement. Nos données apportent des résultats intéressants : 1° l'outil a permis de transformer les pratiques sur de nombreux éléments que les enseignants ont eux-mêmes identifiés, 2° l'accompagnement apparaît comme une plus-value dans la compréhension de l'outil et permet aux enseignants de développer des compétences qui vont au-delà de l'enseignement de la compréhension de textes narratifs.

Ces premiers indices de développement professionnel restent, nous l'avons déjà écrit, à approfondir mais ils offrent de belles perspectives pour un travail futur.

Nous ne pouvons pas conclure sans rapporter une donnée extraite des échanges informels que nous avons eus avec les accompagnateurs au cours de ces trois années : beaucoup ont évoqué la pertinence de ce mode de formation continue pour eux innovant, fondé sur la prise en main d'un outil ou d'un dispositif par une équipe d'enseignants. Certains ont déjà répliqué l'expérience avec d'autres outils, dans d'autres disciplines. Autrement dit, ces formateurs donnent raison à Goigoux et Cèbe (2018) qui posent que les outils contribuent à accroître le pouvoir d'agir des professeurs en leur proposant de nouvelles ressources.

Conclusion générale

À l'origine de cette étude, notre questionnement était d'ordre pratique : nous voulions tester la validité d'un outil didactique innovant et nous assurer des effets produits sur les apprentissages des jeunes élèves et sur les pratiques des maîtres dans le domaine de l'enseignement de la compréhension de textes narratifs. Mais il s'est très tôt déplacé vers une ambition méthodologique : parvenir à construire un dispositif d'évaluation qui satisfasse deux exigences trop rarement réunies dans les études antérieures, la rigueur de l'expérimentation et sa validité écologique. Notre objectif était d'apporter une contribution à l'amélioration de l'enseignement *via* l'utilisation d'outils innovants dont l'efficacité pourrait être avérée. Cela nous a conduit à intégrer la question du « passage à l'échelle » au processus de conception de l'outil et à ne pas dissocier les phases de validation et de diffusion. C'est ainsi que nous avons élaboré un plan expérimental complexe pour mesurer les effets de l'enseignement avec *Narramus*. Ce plan nous a aussi permis d'explorer l'influence des dispositifs d'accompagnement de l'innovation et leur impact sur le développement professionnel des enseignants.

Au terme de ce travail, nous pensons avoir atteint une partie de nos objectifs initiaux.

1° Apporter des connaissances sur la place des outils dans la formation

Nous nous demandions s'il était pertinent de recommander *Narramus*. Pour cela, nous devions vérifier si cet outil, élaboré en respectant les critères ergonomiques de conception et sur la base des résultats scientifiques attestés dans d'autres pays (*Evidence based*), produisait les effets escomptés par ses auteurs.

Nous disposons dorénavant de données probantes qui atteste l'efficacité de l'outil. Notre première étude prouve, en effet, que *Narramus* a permis aux élèves qui en ont profité de construire de solides compétences langagières. Leurs résultats sont significativement supérieurs à ceux du groupe contrôle en narration, en compréhension et en lexique. Cet effet est particulièrement fort chez les élèves de Moyenne et Grande Sections, mais il est aussi présent en Petite Section.

Nous pouvons donc conseiller cet outil aux formateurs et aux enseignants, charge à eux de s'en emparer s'ils le souhaitent.

2° Apporter des connaissances sur le rôle de l'outillage didactique dans la réduction des inégalités sociales d'apprentissage

Dans une perspective de démocratisation de l'École, nous nous sommes interrogée sur les effets de la pédagogie proposée par les auteurs de *Narramus* sur les apprentissages des élèves initialement les moins performants.

Les résultats que nous avons obtenus montrent que les progrès de ces élèves sont à la hauteur de ceux de leurs camarades : ils profitent autant qu'eux de l'enseignement dispensé. Autrement dit, l'effet *Narramus* n'est pas dû à un effet moyen tiré par les plus performants. Ce résultat nous semble majeur, d'autant plus que les élèves ayant participé à la recherche sont issus de milieux populaires.

Les résultats de la deuxième étude vont également dans ce sens : celle-ci a mis en évidence une réduction des inégalités sociales et un transfert des compétences construites avec *Narramus* vers d'autres textes que ceux étudiés en classe. Les élèves qui en ont bénéficié comprennent et racontent mieux que leurs camarades de même condition sociale et leurs performances s'approchent de celles d'enfants favorisés. Leur vocabulaire est lui aussi enrichi.

Ces résultats nous paraissent importants si l'on veut améliorer l'École : diffuser un outil novateur de qualité peut être bénéfique aux élèves à grande échelle et jouer un rôle dans la démocratisation. Si nos conclusions valident la pertinence de *Narramus*, ils concourent surtout à valoriser un enseignement explicite et intégré de la compréhension et du vocabulaire à l'école maternelle. Cette approche multi-dimensionnelle, qui repose sur l'étude d'albums de littérature de jeunesse finalisée par une narration, est une alternative crédible aux approches modulaires qui se limitent à la juxtaposition d'entraînements de compétences isolées (inférences, connecteurs, détection d'incohérence... cf. Gourgue, Bianco et Coda, 2014). Il faudra mener d'autres recherches pour savoir si une méthode est plus efficace qu'une autre.

Il reste encore des investigations à mener pour établir plus solidement certains résultats. Nous avons construit des indices sur les effets de cumul, mais nous n'avons pas administré une preuve telle qu'elle est entendue dans l'approche *evidence based*. Il faudra aussi valider la durabilité des effets observés en poursuivant l'étude sur une même cohorte pendant plusieurs années.

3° Formaliser une méthodologie d'évaluation d'un outil didactique basée sur des critères exigeants pour une étude de grande ampleur en contexte ordinaire de travail

Pour élaborer nos protocoles expérimentaux, nous avons recherché un équilibre entre rigueur et validité écologique. Nous voulions éviter, pour contrôler toutes les variables, de nous placer dans des conditions artificielles, éloignées des conditions de travail ordinaires des enseignants ; cela aurait annulé toute ambition de généralisation dans l'école française. Mais nous souhaitions éviter aussi l'écueil des monographies ou des recherches-actions fortement ancrées sur le terrain mais ne prenant pas soin de contrôler les principales sources de variation des performances des élèves. Rappelons que, pour cela, nous avons conçu un protocole expérimental comparant trois groupes de classes et d'élèves :

- chaque groupe était constitué d'un échantillon de classes et d'élèves numériquement important ;
- l'enseignement dispensé était long : trois fois quatre semaines par an pour chaque niveau, une année scolaire pour les élèves de Grande Section, deux années pour les élèves de Moyenne Section et trois années pour les élèves de Petite Section ;
- les tests évaluatifs étaient cohérents avec les apprentissages visés et ils correspondaient à des épreuves classiques de psycholinguistique ;
- une analyse de la taille de l'effet de l'enseignement a été réalisée ;
- nous avons publié ici tous les résultats de nos études, qu'ils soient positifs ou négatifs.

Ce point fait sans doute une part de l'originalité de notre travail doctoral : nous souhaitons à l'avenir poursuivre dans cette voie en recherchant le meilleur compromis possible entre solide rigueur et forte validité.

4° Savoir s'il est indispensable d'accompagner une innovation

Pour *Narramus*, le soutien externe n'est pas apparu comme une plus-value dans l'amélioration des performances des élèves. L'outil, tel qu'il avait été conçu, avait des vertus intrinsèques suffisantes pour générer un effet à lui seul. Ce constat a été vérifié en Petite, Moyenne et Grande Sections. Toutefois, l'accompagnement a joué un rôle non négligeable dans les pratiques professionnelles. Il a agi comme une aide à l'identification de gestes professionnels nouveaux et semble, à ce titre, être un vecteur de développement professionnel. Les enseignants accompagnés déclarent avoir mieux compris les principes didactiques et pédagogiques au point de pouvoir les transférer dans d'autres activités inscrites dans d'autres domaines d'apprentissage.

Nous n'avons toutefois pas été en mesure de mesurer l'influence de l'accompagnement sur les progrès des élèves en fonction de la variété des contextes locaux. Il serait pertinent, dans une autre étude, d'analyser les résultats obtenus en fonction des caractéristiques des accompagnements mis en œuvre.

Les réponses apportées à nos quatre interrogations de départ nous permettent donc de disposer de données valides à communiquer en formation et d'arguments qui pourront être des ressources pour l'action.

Références bibliographiques

Références bibliographiques

- Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, 8(2), 199-209. <https://doi.org/10.1111/j.1467-7687.2005.00406.x>
- Alter, N. (2002). 1. L'innovation : Un processus collectif ambigu. In *Les logiques de l'innovation* (p. 13–40). La Découverte.
- Alter, N. (2010). *L'innovation ordinaire*. Presses Universitaires de France.
- Altet, M., Desjardins, J., Etienne, R., Paquay, L., & Perrenoud, P. (2014). *Travail réel des enseignants et formation : Quelle référence au travail des enseignants dans les objectifs, les dispositifs et les pratiques ?* De Boeck Supérieur.
- Andreu, S., Le Cam, M., & Rocher, T. (2014). Evolution des acquis en début de CE2 entre 1999 et 2013 : Les progrès observés à l'entrée au CP entre 1997 et 2011 ne sont pas confirmés. *Note d'information*, (19).
- Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a Supplemental Vocabulary Program on Word Knowledge and Passage Comprehension. *Journal of Research on Educational Effectiveness*, 5(2), 160-188. <https://doi.org/10.1080/19345747.2012.660240>
- 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*, 164, 225-238. <https://doi.org/10.1016/j.jecp.2017.04.007>
- Aubineau, L.-H., Vandromme, L., & Driant, B. L. (2015). L'attention conjointe, quarante ans d'évaluations et de recherches de modélisations. *L'Année psychologique*, Vol. 115(1), 141-174.
- Aukrust, V. G., & Rydland, V. (2011). Preschool classroom conversations as long-term resources for second language and literacy acquisition. *Journal of Applied Developmental Psychology*, 32(4), 198-207. <https://doi.org/10.1016/j.appdev.2011.01.002>
- Bablet, M. (2015). *Conférence « Évaluer pour faire progresser »*. Consulté à l'adresse <http://centre-alain-savary.ens-lyon.fr/CAS/education-prioritaire/ressources/theme-5-perspectives-relatives-a-la-carte-de-leducation-pilotage/formation-de-formateurs-academiques-rep-synthese-de-trois-temps-forts>
- Baker, S. K., Simmons, D. C., & Kame'enui, E. J. (1995). *Vocabulary Acquisition : Synthesis of the Research*. University of Oregon, National Center to Improve the Tools of Educators.
- 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)
- Barcellini, F., Van Belleghem, L., & Daniellou, F. (2013). Les projets de conception comme opportunité de développement des activités. In P. Falzon, *Ergonomie constructive* (p. 191). <https://doi.org/10.3917/puf.falzo.2013.01.0191>
- Bardin, L. (2007). *L'analyse de contenu* (PUF). Paris.

- Bastide, I., & Joigneaux, C. (2014). Littératie précoce, albums et lectures partagées à l'école maternelle. *Spirale - Revue de recherches en éducation*, 53(1), 9-19. <https://doi.org/10.3406/spira.2014.1045>
- Bautier, É. (2015). *Pratiques scolaires dominantes et inégalités sociales au sein de l'école* (p. 31).
- Bautier, É. (2018). Le langage oral en maternelle, quelles séances de langage pour quels apprentissages ?. Un regard, des questions, des propositions. *Pratiques. Linguistique, littérature, didactique*, (177-178). <https://doi.org/10.4000/pratiques.4055>
- Bautier, É., & Goigoux, R. (2004). Difficultés d'apprentissage, processus de secondarisation et pratiques enseignantes : Une hypothèse relationnelle. *Revue française de pédagogie*, (148), 89-100. Consulté à l'adresse JSTOR.
- Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, 107(3), 251–271.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). Choosing Words to Teach. In *Bringing words to life : Robust vocabulary development* (p. 19-39). New York: Guilford.
- Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74(4), 506-521. <https://doi.org/10.1037/0022-0663.74.4.506>
- Beck, I., McKeown, M., & Kucan, L. (2002). Choosing words to teach. In *Bringing Words to Life : Robust Vocabulary Instruction* (p. 19-39).
- Bédard, D., & Béchar, J.-P. (2009). L'innovation pédagogique dans le supérieur : Un vaste chantier. *Innover dans l'enseignement supérieur*, 29–43.
- Béguin, P., & Cerf, M. (2004). Formes et enjeux de l'analyse de l'activité pour la conception des systèmes de travail. *Activites*, 01(1). <https://doi.org/10.4000/activites.1156>
- Bélanger, J., Bowen, F., Cartier, S., Desbiens, N., Montésinos-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*, 40(1), 56–75.
- Benveniste, C. (2018). Professionnalisation des enseignants et démocratisation scolaire : Une formation initiale pensée à l'aune de la réduction des inégalités scolaires? *Éducation et socialisation. Les Cahiers du CERFEE*, (50).
- Berman, P. (1980). Thinking about programmed and adaptive implementation : Matching strategies to situations. *Why policies succeed or fail*, 8, 205–227.
- Bernardin, J. (2013). *Le rapport à l'école des élèves de milieux populaires : Guide pédagogique*. De Boeck (Pédagogie et Formation).
- Bianco, M., & Bressoux, P. (2009). Chapitre 2. Effet-classe et effet-maître dans l'enseignement primaire : Vers un enseignement efficace de la compréhension. X. Dumay, & V. Dupriez, *L'efficacité dans l'enseignement*, 35–54.
- Bianco, M., Bressoux, P., Doyen, A.-L., Lambert, E., Lima, L., Pellenq, C., & Zorman, M. (2010). Early Training in Oral Comprehension and Phonological Skills : Results of a Three-Year Longitudinal Study. *Scientific Studies of Reading*, 14(3), 211-246. <https://doi.org/10.1080/10888430903117518>

- Biesta, G. (2007). Why “what works” won’t work : Evidence-based practice and the democratic deficit in educational research. *Educational theory*, 57(1), 1–22.
- Bigot, G., & Gaudrat, M.-L. (2012). *La petite marmite qui tiptopait*. Consulté à l’adresse <http://www.librairiegourmande.fr/litterature-jeunesse/4879-la-petite-marmite-qui-tiptopait.html>
- Bishop, M.-F., Boiron, V., Schmehl-Postaï, A., & Royer, C. (2017). Comprendre des histoires en cours préparatoire : L’exemple du « rappel de récit accompagné ». *Repères. Recherches en didactique du français langue maternelle*, (55), 87-107. <https://doi.org/10.4000/reperes.1157>
- Blanc, N. (2010). *Lecture et habiletés de compréhension chez l’enfant*. Paris: Dunod.
- Blanquer, J.-M. (2016). *L’école de demain : Propositions pour une Éducation nationale renouvelée*. Odile Jacob.
- 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*, 101(2), 294.
- Blignières-Légeraud (de), A. (1998). *Pratiques innovantes de formation, clés pour l’analyse de l’innovation et son transfert, Observatoire de l’innovation. Critères et paramètres pour identifier l’innovation dans le domaine des nouvelles technologies appliquées à l’enseignement et à l’apprentissage, étude pour le Cedefop », rapport final*.
- Boiron, V. (2010). Lire des albums de littérature de jeunesse à l’école maternelle : Quelques caractéristiques d’une expertise en actes. *Repères. Recherches en didactique du français langue maternelle*, (42), 105-126. <https://doi.org/10.4000/reperes.254>
- Boiron, V. (2014). Raconter et lire des récits de fiction : Effets comparés sur la compréhension d’élèves de maternelle. *Repères. Recherches en didactique du français langue maternelle*, (50), 83-104. <https://doi.org/10.4000/reperes.778>
- Bonnéry, S. (2009). Scénarisation des dispositifs pédagogiques et inégalités d’apprentissage. *Revue française de pédagogie*, n° 167(2), 13-23.
- Bonnéry, S. (2014). Les livres et les manières de lire à l’école et dans les familles : *Le français aujourd’hui*, n° 185(2), 47-57.
- Bonnéry, S. (2015). *Supports pédagogiques et inégalités scolaires. Études sociologiques (La Dispute)*. Consulté à l’adresse <http://journals.openedition.org/lectures/17140>
- Bonnéry, S., Crinon, J., & Marin, B. (2015). Des inégalités d’usage de la littérature de jeunesse dès les premiers cycles de l’école primaire ? Une enquête par questionnaires. *Spirale - Revue de recherches en éducation*, 55(1), 43-56. <https://doi.org/10.3406/spira.2015.1021>
- Bonnéry, S., Crinon, J., & Simons, G. (2015). Présentation. *Spirale - Revue de recherches en éducation*, 55(1), 3-10.
- Bonnéry, S., & Joigneaux, C. (2015). Des littératies familiales inégalement rentables scolairement. *Le français aujourd’hui*, (3), 23–34.
- Bonnéry, S., & Mamede, M. (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*.

- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement : A meta-analysis. *Review of educational research*, 73(2), 125–230.
- Boyer, A. (2019). *Bilan de l'expérimentation Institut Carnot de l'Éducation en Auvergne-Rhône-Alpes (document de travail provisoire)*. ENS Lyon.
- Bressoux, P. (2017). Practice-based research : Une aporie et des espoirs. Une revue critique de l'article d'Anthony S. Bryk. *Éducation et didactique*, 11(11-3), 123-134.
- Brigaudiot, M. (1993). Quelques remarques à propos du récit et des images à l'école maternelle. *Repères. Recherches en didactique du français langue maternelle*, 7(1), 7–21.
- Brigaudiot, M. (2000). *Apprentissages progressifs de l'écrit à l'école maternelle*. Paris: Hachette éducation.
- Brigaudiot, M., & Ewald, H. (1990). Construction du récit en section de petits. *Repères. Recherches en didactique du français langue maternelle*, 2(2), 87–96.
- Bronckart, J.-P., & Thurler, M. G. (2004). *Transformer l'école*. De Boeck Supérieur.
- Bronfenbrenner, U. (1986). Ecology of the family as a context for human development : Research perspectives. *Developmental psychology*, 22(6), 723.
- Bru, M. (2002). Pratiques enseignantes : Des recherches à conforter et à développer. *Revue française de pédagogie*, (138), 63-73. Consulté à l'adresse JSTOR.
- Bryk, A. S. (2015). 2014 AERA distinguished lecture : Accelerating how we learn to improve. *Educational Researcher*, 44(9), 467–477.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve : How America's schools can get better at getting better*. Harvard Education Press.
- Bryk, A. S., & Schneider, B. (2003). Trust in schools : A core resource for school reform. *Educational leadership*, 60(6), 40–45.
- Cabell, S. Q., Justice, L. M., Logan, J. A. R., & Konold, T. R. (2013). Emergent literacy profiles among prekindergarten children from low-SES backgrounds : Longitudinal considerations. *Early Childhood Research Quarterly*, 28(3), 608-620. <https://doi.org/10.1016/j.ecresq.2013.03.007>
- Cain, K., Jiang, H., & Logan, J. (2019). Improving reading comprehension in the primary grades : Mediated effects of a language-focused classroom intervention. *Journal of Speech, Language, and Hearing Research*.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's Reading Comprehension Ability : Concurrent Prediction by Working Memory, Verbal Ability, and Component Skills. *Journal of Educational Psychology*, 96(1), 31-42. <https://doi.org/10.1037/0022-0663.96.1.31>
- Canut, E., & Vertalier, M. (2012). Lire des albums : Quelle compréhension et quelle appropriation par les élèves de maternelle ? *Le français aujourd'hui*, n°179(4), 51-66.
- Canut E., Masson C...., & Leroy-Collombel M.. (2018). *Accompagner l'enfant dans son apprentissage du langage*. Vanves: Hachette éducation.
- Cartier, S. C., Contant, H., & Janosz, M. J. (2012). Appropriation de pratiques pédagogiques sur l'apprentissage par la lecture en classe de français du secondaire en milieu

- défavorisé au Québec. *Repères. Recherches en didactique du français langue maternelle*, (45), 97–115.
- 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*, 49(2), 278-293. [https://doi.org/10.1044/1092-4388\(2006/023\)](https://doi.org/10.1044/1092-4388(2006/023))
- Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In *The connections between language and reading disabilities* (p. 25-40). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Cèbe, S. (2009). *Des outils pour travailler les compétences oubliées. XYZep n°34*. Centre Alain Savary.
- Cèbe, S. (1998). Une intervention à visée cognitive en Grande section de Maternelle : Ses effets de transfert sur l'apprentissage de la lecture à l'École élémentaire. *Repères. Recherches en didactique du français langue maternelle*, 18(1), 97-112. <https://doi.org/10.3406/reper.1998.2272>
- Cèbe, S., & Goigoux, R. (2013). *Lectorino & Lectorinette CE1-CE2 : Apprendre à comprendre des textes narratifs*. Paris, France: Retz.
- Cèbe, S., & Goigoux, R. (2017). *Narramus : Apprendre à comprendre et à raconter La sieste de Moussa*.
- Cèbe, S., & Goigoux, R. (2018). Lutter contre les inégalités : Outiller pour former les enseignants. *Recherche formation*, (1), 77–96.
- Cellier, M. (2017). *Le vocabulaire à l'école maternelle*. Paris, France: Retz.
- Chapelle, G., & Meuret, D. (2006). *Améliorer l'école*. Paris: Presses universitaires de France.
- Charbonnier, E. (2018). *L'apport des grandes enquêtes internationales : Principaux enseignements de PISA*. Consulté à l'adresse <https://www.dailymotion.com/video/x6e9o8y>
- Charlier, B., & Peraya, D. (2003). *Technologie et innovation en pédagogie : Dispositifs innovants de formation pour l'enseignement supérieur*. De Boeck.
- Charlier, E., Biémar, S., & Jorro, A. (2012). *Accompagner : Un agir professionnel*. De boeck.
- Chartier, A.-M. (2000). Réussite, échec et ambivalence de l'innovation pédagogique : Le cas de l'enseignement de la lecture. *Recherche & formation*, 34(1), 41-56. <https://doi.org/10.3406/refor.2000.1644>
- Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292.
- Chi, M. T. H., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, 19(1), 29-39. <https://doi.org/10.1037/0012-1649.19.1.29>
- Chi, M. T., Hutchinson, J. E., & Robin, A. F. (1989). How inferences about novel domain-related concepts can be constrained by structured knowledge. *Merrill-Palmer Quarterly*, 35(1), 27-62.
- Choplin, H., Audran, J., Cerisier, J.-F., Lemarchand, S., Paquelin, D., Simonian, S., ... Jacquinot-Delaunay, G. (2007). Quelle recherche sur et pour l'innovation pédagogique. *Distances et savoirs, Vol. 5(4)*, 483-505.

- Class, B., & Schneider, D. (2013). La Recherche Design en Education : Vers une nouvelle approche? *Frantice. net*, 7, 5–16.
- Clot, Y. (1999). *La fonction psychologique du travail*. Presses Universitaires de France-PUF.
- Clot, Y., & Faïta, D. (2000). *Genres et styles en analyse du travail : Concepts et méthodes* (Vol. 4). Travailler.
- Coalition for Evidence-Based Policy. (2003). *Identifying and implementing educational practices supported by rigorous evidence. A user friendly guide*. Washington DC: US Department of Education, Institute of Education Sciences, National Center For education Evaluation and Regional Assistance.
- Cobb, P., & Smith, T. (2008). The challenge of scale : Designing schools and districts as learning organizations for instructional improvement in mathematics. *International handbook of mathematics teacher education*, 3, 231–254.
- Cobb, P., Confrey, J., Disessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), 9-13. <https://doi.org/10.3102/0013189X032001009>
- Coburn, C. E. (2003). Rethinking scale : Moving beyond numbers to deep and lasting change. *Educational researcher*, 32(6), 3–12.
- Coburn, C. E., Honig, M. I., & Stein, M. K. (2009). What’s the evidence on districts’ use of evidence? *The Role of Research in Educational Improvement*, 67-86.
- Collin, É. (2012). Politiques scolaires et dispositifs de formation. *Le français aujourd’hui*, (4), 101–112.
- Collins, A. (1992). Toward a Design Science of Education. In E. Scanlon & T. O’Shea (Éd.), *New Directions in Educational Technology* (p. 15-22). Berlin: Springer Heidelberg.
- Compton, D. L., Elleman, A.-M., Olinghouse, N. G., Gilbert, J. K., & Gilbert, G. N. (2009). The influence of in text instruction on declarative knowledge and vocabulary learning in struggling readers : How IQ confounds the story. In *Beyond Decoding : The Behavioral and Biological Foundations of Reading Comprehension*. Guilford Press.
- 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*, 44(2), 220-234. <https://doi.org/10.1006/jmla.2000.2744>
- Coyne, M. D., Neugebauer, S., Ware, S., McCoach, D. B., & Madura, J. (2015). Vocabulary and its role in early comprehension development. *Pre-reader comprehension: An essential building block to becoming a successful reader*, 19–34.
- Coyne, M. D., Simmons, D. C., Kame’enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings : An examination of differential effects. *Exceptionality*, 12(3), 145–162.
- Cros, F. (1997). L’innovation en éducation et en formation. *Revue française de pédagogie*, 118(1), 127–156.
- Cros, F. (1999). L’innovation en éducation et en formation dans tous ses sens. *Recherche & Formation*, 31(1), 127–136.
- Cros, F. (2003). « L’innovation en éducation, imprévue et rebelle ». In *ministère de l’Éducation nationale, direction de l’Enseignement scolaire, bureau de la valorisation des*

innovations pédagogiques, L'innovation, levier de changement dans l'institution éducative [1999]. Paris : CNDP (Pratiques innovantes).

- Cros, F. (2004). *L'innovation scolaire aux risques de son évaluation*. L'Harmattan.
- Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American educator*, 22, 8–17.
- Cusset, P.-Y. (2014). *Les pratiques pédagogiques efficaces*. 42.
- Damhuis, C. M. P., Segers, E., & Verhoeven, L. (2015). Stimulating breadth and depth of vocabulary via repeated storybook readings or tests. *School Effectiveness and School Improvement*, 26(3), 382-396. <https://doi.org/10.1080/09243453.2014.965181>
- Daniellou, F. (2004). L'ergonomie dans la conduite de projets de conception de systèmes de travail. *Ergonomie*, 359–373.
- Daunay, B. (2009). Interroger les implicites de l'évaluation des résultats des élèves lors d'une expérimentation didactique. In C. Cohen-Azria & N. Sayac (Éd.), *Questionner l'implicite* (p. 201-210). <https://doi.org/10.4000/books.septentrion.14688>
- Daussin, J.-M., Keskpaik, S., & Rocher, T. (2011). L'évolution du nombre d'élèves en difficulté face à l'écrit depuis une dizaine d'années. *France, portrait social*, 137–152.
- Davies, P. (1999). What is evidence-based education? *British journal of educational studies*, 47(2), 108–121.
- De Temple, Jeanne M., & Tabor, P. O. (1994, décembre). *Styles of Interaction during a Book Reading Task : Implications for Literacy Intervention with Low-Income Families*. Consulté à l'adresse <https://eric.ed.gov/?id=ED379615>
- De Temple, J.M, & Snow, C. (1996). Styles of parent-child book reading as related to mother's view of literacy and children's literacy outcomes. In J. Shimron (Éd.), *Literacy and education : Essays in memory of dina feitelson* (Hampton Press, p. 49-68). Cresskill, NJ: Hampton Press.
- de Weck, G. (2007). L'acquisition des discours à l'âge préscolaire : L'exemple du récit d'expériences personnelles. *Inter-Actions*, 44.
- Deaudelin, C., Dussault, M., & Brodeur, M. (2002). Impact d'une stratégie d'intégration des TIC sur le sentiment d'autoefficacité d'enseignants du primaire et leur processus d'adoption d'une innovation. *Revue des sciences de l'éducation*, 28(2), 391–410.
- DeBruin-Parecki, A., Van Kleeck, A., & Gear, S. (2015). *Developing early comprehension : Laying the foundation for reading success*. Paul H. Brookes Publishing Co.
- Delahaye, J.-P. (2015). *Grande pauvreté et réussite scolaire. Le choix de la solidarité pour la réussite de tous*. [Rapport à l'Inspection générale de l'Éducation nationale]. Paris.
- Deleau, M. (2007). Le développement de la « théorie de l'esprit ». In J. Lautrey, *Psychologie du développement et de l'éducation* (p. 87-116). Paris, France: Presses universitaires de France.
- Deltour, J. J., & Hupkens, D. (1980). *TVAP 3-5 et 5-8 : Test de vocabulaire actif et passif*. EAP.
- DEPP. (2013). *Forte augmentation du niveau des acquis des élèves à l'entrée au CP entre 1997 et 2011. Note d'information 13.19. DEPP, ministère de l'Éducation nationale*. Consulté à l'adresse

https://cache.media.education.gouv.fr/file/2013/11/2/DEPP_NI_2013_19_forte_augmentation_niveau_acquis_eleves_entree_CP_entre_1997_2011_269112.pdf

- DEPP. (2016). *CEDRE 2003-2009-2015 Maîtrise de la langue en fin d'école : L'écart se creuse entre filles et garçons. Note d'information n°20.*
- DEPP. (2017). *PIRLS 2016 : Évaluation internationale des élèves de CM1 en compréhension de l'écrit. Évolution des performances sur quinze ans. Note d'information n°17-24.*
- DEPP. (2019). *Pratiques de classe, sentiment d'efficacité personnelle et besoins de formation : Une photographie inédite du métier de professeur des écoles début 2018. Note d'information n° 19-22.*
- 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, 29*(2), 104–22. <https://doi.org/10.2307/747807>
- Dickinson, D. K., & Tabors, P. O. (2001). *Beginning literacy with language : Young children learning at home and school.* Baltimore, MD, US: Paul H Brookes Publishing.
- Dickinson, D., & Keebler, R. (1989). Variation in preschool teachers' styles of reading books. *Discourse Processes, 12*(3), 353-375. <https://doi.org/10.1080/01638538909544735>
- Dolle, J. R., Gomez, L. M., Russell, J. L., & Bryk, A. S. (2013). More than a network : Building professional communities for educational improvement. *National Society for the Study of Education Yearbook, 112*(2), 443–463.
- Donovan, M. S., Snow, C., & Daro, P. (2013). The SERP approach to problem-solving research, development, and implementation. *National Society for the Study of Education Yearbook, 112*(2), 400–425.
- Dreyfus, S. E. (2004). The Five-Stage Model of Adult Skill Acquisition. *Bulletin of Science, Technology & Society, 24*(3), 177-181. <https://doi.org/10.1177/0270467604264992>
- Dufflo, E. (2004). Scaling up and evaluation. *Annual World Bank Conference on Development Economics 2004*, 341–369.
- Duke, N., & Pearson, P. (2002). What research has to say about reading instruction. *Effective practices for developing reading comprehension*, 205-242.
- Dumas, J. E., Lynch, A. M., Laughlin, J. E., Smith, E. P., & Prinz, R. J. (2001). Promoting intervention fidelity : Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine, 20*(Suppl1), 38-47. [https://doi.org/10.1016/S0749-3797\(00\)00272-5](https://doi.org/10.1016/S0749-3797(00)00272-5)
- Dunlap, G., Sugai, G., Lewis, T., Goodman, S., & Horner, R. (2009). *Scaling up and sustaining evidenced-based practices.* Consulté à l'adresse www.pbis.org/common/ressources/presentations/Monday_ScalingPBIS_Horner.ppt
- Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young Children's Understanding of Other People's Feelings and Beliefs : Individual Differences and Their Antecedents. *Child Development, 62*(6), 1352-1366. <https://doi.org/10.2307/1130811>
- Dupriez, V. (2015). *Peut-on réformer l'école? : Approches organisationnelle et institutionnelle du changement pédagogique.* De Boeck Supérieur.

- 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*, 41(3-4), 327.
- Duru-Bellat, M. (2002). *Les inégalités sociales à l'école. Genèse et mythes*.
- Duru-Bellat, M., & Mingat, A. (1998). A quelles conditions une démarche de recherche en éducation peut-elle être tenue pour «scientifiquement correcte». *Recherche et éducation. Vers une «nouvelle alliance»*. *La démarche de preuve en dix questions*, 161–190.
- Duru-Bellat, M. (2006). Peut-on diminuer les inégalités sociales à l'école. *G. Chapelle & D. Meuret (éd.), Améliorer l'école, Paris, Presses universitaires de France, coll." Apprendre*, 25–36.
- Duru-Bellat, Marie, Jarousse, J.-P., & Mingat, A. (1993). Les scolarités de la maternelle au lycée. Étapes et processus dans la production des inégalités sociales. *Revue française de sociologie*, 34(1), 43–60.
- Dusenbury, L. (2003). A review of research on fidelity of implementation : Implications for drug abuse prevention in school settings. *Health Education Research*, 18(2), 237-256. <https://doi.org/10.1093/her/18.2.237>
- Ecalte, J., Gomes, C., Auphan, P., Cros, L., & Magnan, A. (2019). Effects of policy and educational interventions intended to reduce difficulties in literacy skills in grade 1. *Studies in Educational Evaluation*, 61, 12-20. <https://doi.org/10.1016/j.stueduc.2019.02.001>
- Ecalte, J., Labat, H., Le Cam, M., Rocher, T., Cros, L., & Magnan, A. (2015). Evidence-based practices to stimulate emergent literacy skills in kindergarten in France : A large-scale study. *Teaching and Teacher Education*, 50, 102–113.
- Elalouf, M.-L., & Bois-Masson, N. (2014). Informer sur les manuels scolaires, une mission délaissée par l'éducation nationale. *Recherches*, (60), 119-135.
- Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The Impact of Vocabulary Instruction on Passage-Level Comprehension of School-Age Children : A Meta-Analysis. *Journal of Research on Educational Effectiveness*, 2(1), 1-44. <https://doi.org/10.1080/19345740802539200>
- Elmore, R. (1996). Getting to scale with good educational practice. *Harvard educational review*, 66(1), 1–27.
- Elwér, Å., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology*, 115(3), 497-516. <https://doi.org/10.1016/j.jecp.2012.12.001>
- Elwér, Å., 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*, 56(2), 157–166.
- Enthoven, S., Letor, C., & Dupriez, V. (2015). Réformes pédagogiques et autonomie professionnelle : Un couple en tension. *Revue française de pédagogie*, n° 192(3), 95-108.

- Evans, M., & Saint-Aubin, J. (2005). What children are looking at during shared storybook reading: Evidence from eye movement monitoring. *Psychological science*, 16(11), 913–920.
- Fayol, M. (2017). Les premiers apprentissages. *Administration Education*, N° 155(3), 103-108.
- FEJ [FONDS D'EXPÉRIMENTATION POUR LA JEUNESSE]. (2014). *Projet lecture, rapport d'évaluation*.
- Felouzis, G. (2014). *Les inégalités scolaires : « Que sais-je ? »* n° 3985. Presses Universitaires de France.
- Felouzis, G., Fouquet-Chauprade, B., & Charmillot, S. (2015). Les descendants d'immigrés à l'école en France : Entre discontinuité culturelle et discrimination systémique. *Revue française de pédagogie*, (2), 11–27.
- Fischer, K. W., Goswami, U., & Geake, J. (2010). The future of educational neuroscience. *Mind, Brain, and Education*, 4(2), 68–80.
- 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*, 61(6), 825-850.
- Fishman, B. J., Penuel, W. R., Allen, A.-R., Cheng, B. H., & Sabelli, N. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *National society for the study of education*, 112(2), 136–156.
- Fixsen, D., Blase, K., Metz, A., & Van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children*, 79(2), 213–230.
- Fixsen, D. L., Blase, K. A., Horner, R., & Sugai, G. (2009). Scaling-up brief. *Chapel Hill: University of North Carolina*.
- Flack, Z. M., Field, A. P., & Horst, J. S. (2018). The effects of shared storybook reading on word learning: A meta-analysis. *Developmental Psychology*, 54(7), 1334-1346. <https://doi.org/10.1037/dev0000512>
- Florin, A. (2010). Le développement du lexique et l'aide aux apprentissages. *Enfances Psy*, n° 47(2), 30-41.
- Florin, A. (2016). *Le développement du langage*. Paris, France: Dunod.
- Frier, C. (2006). *Passeurs de lectures lire ensemble à la maison et à l'école*. Paris: Retz.
- Fullan, M. (1996). *Implementation of innovations*. Dans *LE. Plomb et D. P. Ely (dir.), International Encyclopedia of Educational Technology (2e éd.)* (p. 273-281). Oxford : Pergamon.
- Gaglio, G. (2011). *Sociologie de l'innovation: «Que sais-je?»* n° 3921. Presses universitaires de France.
- Galand, B., & Janosz, M. (2019). *Quels sont les verrous à l'amélioration des pratiques éducatives, et comment les débloquent ? Problématique du symposium*. REF. Toulouse.
- Ganea, P. A., Pickard, M. B., & DeLoache, J. S. (2008). Transfer between picture books and the real world by very young children. *Journal of Cognition and Development*, 9(1), 46-66. <https://doi.org/10.1080/15248370701836592>

- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What Makes Professional Development Effective? Results From a National Sample of Teachers. *American Educational Research Journal*, 38(4), 915-945. <https://doi.org/10.3102/00028312038004915>
- Garner, P. W., Jones, D. C., Gaddy, G., & Rennie, K. M. (1997). Low-income mothers' conversations about emotions and their children's emotional competence. *Social Development*, 6(1), 37-52.
- Gather Thurler, M. (2004). Accompagner l'innovation de l'intérieur: Paradoxes du développement de l'organisation scolaire. *Accompagner les réformes et les innovations en éducation, Paris, L'Harmattan*, 69-100.
- Gentaz, É. (2017). Co-construction de recherches en sciences cognitives interventionnelles : Un outil pour développer la formation initiale et continue des enseignant-e-s. *Approche Neuropsychologique des Apprentissages chez l'Enfant*, (146), 9-10.
- Gentaz, É. (2018). Du labo à l'école : Le délicat passage à l'échelle. Consulté 23 janvier 2019, à l'adresse <https://www.larecherche.fr/sciences-cognitives/du-labo-à-lécole-le-délicat-passage-à-léchelle>
- Gentaz, É., Sprenger-Charolles, L., Colé, P., Theurel, A., & Gurgan, 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 (ANA E.)*, 123, 172-181.
- Giasson, J. (1990). *La compréhension en lecture*. De Boeck Université.
- Giasson, J. (2012). *La lecture : Apprentissage et difficultés*. Bruxelles, Belgique: De Boeck.
- Girolametto, L., Weitzman, E., van Lieshout, R., & Duff, D. (2000). Directiveness in teacher's language input to toddlers and preschoolers in day care. *Journal of Speech, Language, and Hearing Research*, 43(5), 1101-1114. <https://doi.org/10.1044/jslhr.4305.1101>
- Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences*, 20(1), 1-55. <https://doi.org/10.1017/S0140525X97470012>
- Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and Imagined Activity Can Enhance Young Children's Reading Comprehension. *Journal of Educational Psychology*, 96(3), 424-436. <https://doi.org/10.1037/0022-0663.96.3.424>
- Glenberg, A. M. (2011). How reading comprehension is embodied and why that matters. *International Electronic Journal of Elementary Education*, 4(1), 5-18.
- Goigoux, R. (2007). Un modèle d'analyse de l'activité des enseignants. *Éducation et didactique*, 1(3), 47-69.
- Goigoux, R. (2016). *Lire et écrire au CP. Étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des apprentissages*.
- Goigoux, R. (2019). *Evaluations : Faire mentir les chiffres, en pédagogie aussi*. Consulté à l'adresse <http://www.cafepedagogique.net/lexpresso/Pages/2019/05/13052019Article636933706640890148.aspx>

- Goigoux, R., & Cèbe, S. (2008). *Lectrix & lector. Apprendre à comprendre les textes*, Paris: Retz.
- Goigoux, R., Renaud, J., & Roux-Baron, I. (2019). *Conception continuée dans l'usage d'outils didactiques innovants au service de l'amélioration des apprentissages des élèves : Un cadre théorique et deux études empiriques*. REF. Toulouse.
- Goigoux, R. (2017). Associer chercheurs et praticiens à la conception d'outils didactiques ou de dispositifs innovants pour améliorer l'enseignement. *Education didactique*, Vol. 11(3), 135-142.
- Goigoux, R., & Cèbe, S. (2009). Un autre rapport entre recherche, pratique et formation. *Les instruments didactiques comme vecteur de transformation des pratiques des enseignants confrontés aux difficultés d'apprentissage des élèves*. Présenté à Réseau international de Recherche en Education et Formation (REF), Nantes.
- 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*, (3), 67-84.
- Goigoux, R., Jarlégan, A., & Piquée, C. (2015). Évaluer l'influence des pratiques d'enseignement du lire-écrire sur les apprentissages des élèves : Enjeux et choix méthodologiques. *Recherches en didactiques*, N° 19(1), 9-37.
- Goldstein, H., Kelley, E., Greenwood, C., McCune, L., Carta, J., Atwater, J., ... Spencer, T. (2016). Embedded Instruction Improves Vocabulary Learning During Automated Storybook Reading Among High-Risk Preschoolers. *Journal of Speech, Language, and Hearing Research*, 59(3), 484-500. https://doi.org/10.1044/2015_JSLHR-L-15-0227
- Gomel, B., & Serverin, É. (2013). L'expérimentation sociale aléatoire en France en trois questions. *Travail et Emploi*, (135), 57-71. <https://doi.org/10.4000/travailemploi.6081>
- Gorard, S., Rushforth, K., & Taylor, C. (2004). Is there a shortage of quantitative work in education research? *Oxford Review of Education*, 30(3), 371-395.
- Gough, D. (2007). Weight of evidence : A framework for the appraisal of the quality and relevance of evidence. *Research papers in education*, 22(2), 213-228.
- Gourgue, D., Bianco, M., & Coda, M. (2013). *Compréhension, Grande section*. Grenoble: Éditions de la Cigale.
- Graves, M. F. (2000). A vocabulary program to complement and bolster a middle-grade comprehension program. *Reading for meaning: Fostering comprehension in the middle grades*, 116-135.
- Grobon, S., & Panico, L. (2019). *Inégalités socio-économiques dans le développement langagier et moteur des enfants à 2 ans*. 8. <https://afpa.org/content/uploads/2019/02/In%C3%A9galit%C3%A9s-socio-%C3%A9conomiques-dans-le-d%C3%A9veloppement-langagier-et-moteur-des-enfants-%C3%A0-2-ans-BEH-janvier-2019.pdf>
- Grossmann, F. (1996). *Enfances de la lecture Manières de faire, manières de lire à l'école maternelle*. Berne: Peter Lang.
- Grossmann, F. (2001). Pour une approche ethnographique des activités de lecture. L'exemple des lectures partagées. *Pratiques de lecture et cheminements du sens, Cahiers du français contemporain*, 7, 135-160.

- Gurgand, M. (2018). Expérimentation scolaire : Du laboratoire à la classe. Colloque du collège de Franc, le rôle de l'expérimentation dans le domaine éducatif, 1er février 2018. Consulté 21 mars 2018, à l'adresse <https://www.college-de-france.fr/site/stanislas-dehaene/symposium-2018-02-01-14h00.htm>
- Haden, C., Reese, E., & Fivush, R. (1996). Mothers' Extratextual Comments During Storybook Reading : Stylistic Differences over Time and Across Texts. *Discourse Processes - DISCOURSE PROCESS*, 21, 135-169. <https://doi.org/10.1080/01638539609544953>
- Hadji, C. (1991). *Innover pour réussir : Des acteurs parlent aux acteurs*. Esf Editeur.
- Hadji, C., & Baillé, J. (1998). *Recherche et éducation : Vers une " nouvelle alliance " : la démarche de preuve en 10 questions*. De Boeck Supérieur.
- Hallinger, P. (2003). Leading educational change : Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of education*, 33(3), 329–352.
- Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., ... Scott-Little, C. (2012). A course on effective teacher-child interactions : Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, 49(1), 88–123.
- Hargreaves, D. H. (1997). In defence of research for evidence-based teaching : A rejoinder to Martyn Hammersley. *British educational research journal*, 23(4), 405–419.
- Hargreaves, D. H. (1999). Revitalising educational research : Lessons from the past and proposals for the future. *Cambridge journal of education*, 29(2), 239–249.
- Heath, S. B. (1982). What no bedtime story means : Narrative skills at home and school*. *Language in Society*, 11(1), 49-76. <https://doi.org/10.1017/S0047404500009039>
- Hiebert, E. H., & Kamil, M. L. (2005). Teaching and Learning Vocabulary : Perspectives and Persistent Issues. In *Teaching and Learning Vocabulary : Bringing Research to Practice* (p. 1-23). New Jersey: Lawrence Erlbaum Associates.
- 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*, 27(2), 287-313. <https://doi.org/10.1007/s11145-013-9445-4>
- Hindman, A. H., Wasik, B. A., & Bradley, D. E. (2019). How Classroom Conversations Unfold : Exploring Teacher–Child Exchanges During Shared Book Reading. *Early Education and Development*, 1–18.
- Horn, I. S., & Little, J. W. (2010). Attending to problems of practice : Routines and resources for professional learning in teachers' workplace interactions. *American educational research journal*, 47(1), 181–217.
- Huberman, M. (1973). *Comment s'opèrent les changements en éducation*.
- Huberman, M. (1982). De l'innovation scolaire et son marchandage. *Revue européenne des sciences sociales*, 20(63), 59–85.
- Huberman, M. (1989). Les phases de la carrière enseignante : Un essai de description et de prévision. *Revue française de pédagogie*, (86), 5-16. Consulté à l'adresse JSTOR.

- Huberman, M. (1995). *Professional careers and professional development: Some intersections*. TR Guskey & M. Huberman (Eds.), Professional development in education: New
- Hughes, D. L., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment*. Thinking Publications Eau Claire, WI.
- IGEN. (2011). *L'école maternelle (rapport n°2011-108)*. Inspection générale de l'éducation nationale et Inspection générale de l'administration de l'éducation nationale et de la recherche. Consulté à l'adresse http://media.education.gouv.fr/file/2011/54/5/2011-108-IGEN-IGAENR_215545.pdf
- IGEN. (2017). *La scolarisation en petite section de maternelle. Rapport n°2017-032*.
- IGEN, & IGAENR. (2017). *Évaluation de la politique publique sur la formation continue des professeurs du premier degré (rapport n°2017-035)*. Inspection générale de l'éducation nationale et Inspection générale de l'administration de l'éducation nationale et de la recherche.
- Imbert, F. (1986). Innovation et temporalité. *Revue française de pédagogie*, 75(1), 53–59.
- James-Burdumy, S., Deke, J., Gersten, R., Lugo-Gil, J., Newman-Gonchar, R., Dimino, J., ... Liu, A. Y.-H. (2012). Effectiveness of Four Supplemental Reading Comprehension Interventions. *Journal of Research on Educational Effectiveness*, 5(4), 345-383. <https://doi.org/10.1080/19345747.2012.698374>
- Jarlegan, 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*. Consulté à l'adresse <https://hal.archives-ouvertes.fr/hal-01683290>
- 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*, 65, 107-125. <https://doi.org/10.1016/j.econedurev.2018.05.001>
- Joigneaux, C. (2009). La construction de l'inégalité scolaire dès l'école maternelle. *Revue française de pédagogie. Recherches en éducation*, (169), 17-28. <https://doi.org/10.4000/rfp.1301>
- Joigneaux, C. (2013). Littératie, forme et inégalités scolaires : Le cas de la « scolarisation » de l'école maternelle. *Le français aujourd'hui*, n°183(4), 41-50.
- 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*, 15(2), 177-191. [https://doi.org/10.1044/1058-0360\(2006/017\)](https://doi.org/10.1044/1058-0360(2006/017))
- Justice, L. M., & Ezell, H. K. (2002). Use of storybook reading to increase print awareness in at-risk children. *American Journal of Speech-Language Pathology*.
- Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks. *Language, speech, and hearing services in schools*, (36), 17-32.
- Justice, L. M., Pullen, P. C., & Pence, K. (2008). Influence of verbal and nonverbal references to print on preschoolers' visual attention to print during storybook reading. *Developmental Psychology*, 44(3), 855.

- Karweit, N. (1989). The effects of a story-reading program on the vocabulary and story comprehension skills of disadvantaged prekindergarten and kindergarten students. *Early Education and Development, 1*(2), 105–114.
- Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of research in Reading, 31*(3), 259–272.
- Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980.
- Kervyn, B., & Rebière, M. (2014). In *Prendre en compte le point de vue des élèves pour outiller la didactique de l'écriture* (p. 51-67). Consulté à l'adresse <https://www.cairn.info/didactique-du-francais-du-cote-des-eleves--9782804184780-page-51.htm?contenu=resume>
- 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, 141*, 101-120. <https://doi.org/10.1016/j.jecp.2015.08.003>
- Kim, Y.-S., & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly, 49*(3), 269–281.
- Kintsch, W., & Van Dijk, T. A. (1975). Comment on se rappelle et on résume des histoires. *Langages, 9*(40), 98-116. <https://doi.org/10.3406/lgge.1975.2300>
- Klingner, J. K., Boardman, A. G., & McMaster, K. L. (2013). What does it take to scale up and sustain evidence-based practices? *Exceptional Children, 79*(2), 195–211.
- Korat, O., Klein, P., & Segal-Drori, O. (2007). Maternal mediation in book reading, home literacy environment, and children's emergent literacy: A comparison between two social groups. *Reading and Writing, 20*(4), 361-398. <https://doi.org/10.1007/s11145-006-9034-x>
- Kretlow, A. G., & Bartholomew, C. C. (2010). Using Coaching to Improve the Fidelity of Evidence-Based Practices: A Review of Studies. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 33*(4), 279-299. <https://doi.org/10.1177/0888406410371643>
- Labat, H., Farhat, S.-L., Andreu, S., Rocher, T., Cros, L., Magnan, A., & Ecalle, J. (2013). Évaluation des connaissances précoces prédictives de l'apprentissage de la lecture en grande section de maternelle. *Revue française de pédagogie. Recherches en éducation, 184*(4), 41-54. <https://doi.org/10.4000/rfp.4219>
- Labrousse, A., & Zamora, P. (2013). Expérimentations de terrain et politiques publiques du travail et de l'emploi. Apports récents et mises en perspective. Introduction. *Travail et emploi, 135*, 5–13.
- Lafortune, L. (2008a). *Compétences Professionnelles Pour L'Accompagnement D'un Changement : Un Référentiel*. PUQ.
- Lafortune, L. (2008b). *Un Modèle D'Accompagnement Professionnel D'un Changement : Pour Un Leadership Novateur*. Québec: PUQ.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics, 159*–174.

- Landry, D., Dion, E., Dupéré, V., & Barrette, A. (2017). Apprendre le sens de nouveaux mots dans un environnement riche en information. Une étude micro-développementale. *Enfance en difficulté*, 5, 55-71. <https://doi.org/10.7202/1043354ar>
- Lane, H. B., & Allen, S. A. (2010). The Vocabulary-Rich Classroom : Modeling Sophisticated Word Use to Promote Word Consciousness and Vocabulary Growth. *The Reading Teacher*, 63(5), 362-370. <https://doi.org/10.1598/RT.63.5.2>
- Lecocq, P. (1996). L'ECOSSE: *une épreuve de compréhension syntaxico-sémantique*. Lille, France: presses Universitaires du Septentrion.
- Le Normand, M.-T., Parisse, C., & Cohen, H. (2008). Lexical diversity and productivity in French preschoolers : Developmental, gender and sociocultural factors. *Clinical Linguistics & Phonetics*, 22(1), 47–58.
- Le Sourn-Bissaoui, S., & Deleau, M. (2001). Discours maternel et compréhension des états mentaux émotionnels et cognitifs à 3 ans. *Enfance*, 53(4), 329–348.
- Lefevre, G., Garcia, A., & Namolovan, L. (2009). Les indicateurs de développement professionnel. *Questions Vives. Recherches en éducation*, (Vol.5 n°11), 277-314. <https://doi.org/10.4000/questionsvives.627>
- LeMahieu, P. G., Nordstrum, L. E., & Potvin, A. S. (2017). Design-based implementation research. *Quality Assurance in Education*. <https://doi.org/10.1108/QAE-11-2016-0077>
- Leplat, J. (2000). *L'analyse psychologique de l'activité en ergonomie : Aperçu sur son évolution, ses modèles et ses méthodes*. Octarès.
- Lessard, C. (2007). Les usages politiques de la recherche en éducation. In *Enseigner* (Dupriez, V. & Chapelle G., p. 69-79). Paris: Presse Universitaire de France.
- Létor, 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*, (35), 37–55.
- Leung, C. B. (2008). Preschoolers' acquisition of scientific vocabulary through repeated read-aloud events, retellings, and hands-on science activities. *Reading Psychology*, 29(2), 165-193. <https://doi.org/10.1080/02702710801964090>
- Lever, R., & Sénéchal, M. (2011). Discussing stories : On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology*, 108(1), 1-24. <https://doi.org/10.1016/j.jecp.2010.07.002>
- Levin, B., & Fullan, M. (2008). Learning about system renewal. *Educational management administration & leadership*, 36(2), 289–303.
- Little, J. W. (2006). Professional community and professional development in the learningcentered school Washington. DC: *National Education Association*.
- Loftus, S. M., Coyne, M. D., McCoach, D. B., Zipoli, R., & Pullen, P. C. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research & Practice*, 124-136.
- Loftus-Rattan, S. M., Mitchell, A. M., & Coyne, M. D. (2016). Direct Vocabulary Instruction in Preschool : A Comparison of Extended Instruction, Embedded Instruction, and

- Incidental Exposure. *The Elementary School Journal*, 116(3), 391-410.
<https://doi.org/10.1086/684828>
- Marcel, J.-F. (2009). Le Sentiment d'Efficacité Professionnelle, un indicateur pour connaître le développement professionnel des « nouveaux » professeurs de l'enseignement agricole français. *Questions Vives. Recherches en éducation*, (Vol.5 n°11), 161-176.
<https://doi.org/10.4000/questionsvives.564>
- Marks, S. U., & Gersten, R. (1998). Engagement and disengagement between special and general educators: An application of Miles and Huberman's cross-case analysis. *Learning Disability Quarterly*, 21(1), 34–56.
- Martinez, M. G., & Teale, W. H. (1993). Teacher Storybook Reading Style : A Comparison of Six Teachers. *Research in the Teaching of English*, 27(2), 175-199. Consulté à l'adresse JSTOR.
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning : A meta-analysis. *Review of educational research*, 80(3), 300–335.
- Mârz, V. (2014). «*It takes two to tango* » : *Structuring actors and acting structures in the implementation of educational innovations. Thèse de doctorat en Sciences de l'éducation*. Université de Leuven, Faculté de psychologie et des sciences de l'éducation, Centre for Innovation and the Development of Teacher and School., Belgique.
- 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*, 25(1), 136-152.
<https://doi.org/10.1080/1350293X.2016.1266226>
- McKenney, S., & Reeves, T. C. (2014). Educational design research. In *Handbook of research on educational communications and technology* (p. 131–140). New York: Springer.
- McKeown, M. G., & Beck, I. L. (2004). Direct and rich vocabulary instruction. *Vocabulary instruction: Research to practice*, 13–27.
- McLaughlin, M. W., & Mitra, D. (2001). Theory-based change and change-based theory : Going deeper, going broader. *Journal of Educational Change*, 2(4), 301–323.
- 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*, 39(2), 147-160.
- McLeod, A. N., & McDade, H. L. (2011). Preschoolers' incidental learning of novel words during storybook reading. *Communication Disorders Quarterly*, 32(4), 256-266.
<https://doi.org/10.1177/1525740109354777>
- McLuhan, M. (1964). *Understanding Media*. MIT - [c2d6debe1f7abffe2129d3f3ddf385f0ecf1.pdf](https://doi.org/10.2307/2526262).
- MEN. (2013). *Loi n° 2013-595 du 8 juillet 2013 d'orientation et de programmation pour la refondation de l'école de la République*. Consulté à l'adresse [En ligne]<http://www.education.gouv.fr/cid72962/publication-au-journal-officiel-de-la-loi-dorientation-et-de-programmation-pour-la-refondation-de-l-ecole-de-la-republique.html>. (26/06/2017)

- MEN. (2019). *Circulaire de rentrée 2019 Les priorités pour l'école primaire Note de service n° 2019-087 du 28 mai 2019*.
- Meuret, D. (2007). *Gouverner l'école : Une comparaison France/États-Unis*. PUF.
- Min, S., Modeste, M. E., Salisbury, J., & Goff, P. T. (2016). Heeding the CALL (Comprehensive Assessment of Leadership for Learning). *Journal of Educational Administration*. <https://doi.org/10.1108/JEA-07-2014-0075>
- Ministère de l'Éducation nationale. (2008). *Horaires et programmes d'enseignement de l'école primaire. Bulletin officiel spécial n°3, numéro hors-série du 19 juin 2008*.
- Ministère de l'Éducation nationale, de l'enseignement supérieur et de la recherche. (2015). *Programme de l'école maternelle. Bulletin officiel spécial n°2 du 26 mars 2015*.
- Monetti, V. (2002). *Certitudes et paradoxes de l'innovation : État des lieux, états d'esprit*. Institut national de recherche pédagogique.
- Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The Words Children Hear : Picture Books and the Statistics for Language Learning. *Psychological Science*, 26(9), 1489-1496. <https://doi.org/10.1177/0956797615594361>
- 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*, 10(3), 294–313.
- Myrttil, 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*, 47, 771–786. Springer.
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly*, 20(2), 233-253. <https://doi.org/10.2307/747758>
- 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*, 51(9), 1031-1039. <https://doi.org/10.1111/j.1469-7610.2010.02254.x>
- National Reading Panel,. (2000). *Teaching Children To Read : An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction. Reports of the Subgroups*. Web site: <http://www>.
- National Research Council. (2003). *Strategic Education Research Partnership*. Consulté à l'adresse <https://doi.org/10.17226/10670>.
- Neuman, S. B., & Kaefer, T. (2018). Developing low-income children's vocabulary and content knowledge through a shared book reading program. *Contemporary Educational Psychology*, 52, 15-24. <https://doi.org/10.1016/j.cedpsych.2017.12.001>
- 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* (p. 35-52). Baltimore, MA: Paul H. Brookes Publishing.
- Normand, R. (2006). Les qualités de la recherche ou les enjeux du travail de la preuve en éducation. *Éducation et sociétés*, (2), 73–91.

- Oakley, A. (2002). Social science and evidence-based everything : The case of education. *Educational review*, 54(3), 277–286.
- OCDE. (2016). *Résultats du PISA 2015 (volume 1) : L'excellence et l'équité dans l'éducation*.
- Odom, S. L. (2009). The tie that binds : Evidence-based practice, implementation science, and outcomes for children. *Topics in Early Childhood Special Education*, 29(1), 53–61.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, 78(1), 33-84. <https://doi.org/10.3102/0034654307313793>
- Pasquinelli, E. (2011). Knowledge-and Evidence-Based Education : Reasons, Trends, and Contents. *Mind, Brain, and Education*, 5(4), 186–195.
- Pastré, P., Mayen, P., & Vergnaud, G. (2006). La didactique professionnelle. *Revue française de pédagogie. Recherches en éducation*, (154), 145–198.
- Paul, M. (2004). *L'accompagnement : Une posture professionnelle spécifique*. Paris: L'Harmattan.
- Paul, M. (2009). L'accompagnement dans le champ professionnel. *Savoirs*, (2), 11–63.
- Payne, A. A., Gottfredson, D. C., & Gottfredson, G. D. (2006). School Predictors of the Intensity of Implementation of School-Based Prevention Programs : Results from a National Study. *Prevention Science*, 7(2), 225-237. <https://doi.org/10.1007/s11121-006-0029-2>
- Penneman, J. (2018). *Lirécrire pour apprendre : Entre appropriation par les enseignants et effets sur les apprentissages des élèves* (PhD Thesis). UCL-Université Catholique de Louvain.
- Penneman, J., Croix, S. D., Dellisse, S., Dufays, J.-L., Dumay, X., Dupriez, V., ... 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(4), 79-98.
- Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories : Do they overcome the Matthew effect? *Journal of Educational Psychology*, 94(1), 23-33. <https://doi.org/10.1037/0022-0663.94.1.23>
- Perfetti, C. (2007). Reading Ability : Lexical Quality to Comprehension. *Scientific Studies of Reading*, 11(4), 357-383.
- Perfetti, C., McKeown, M. G., & Kucan, L. (2010). Decoding, vocabulary, and comprehension. *Bringing reading research to life*, 291–303.
- Pesco, D., & Gagné, A. (2017). Scaffolding narrative skills: A meta-analysis of instruction in early childhood settings. *Early Education and Development*, 28(7), 773-793.
- Picoche, J. (1992). *Précis de lexicologie française : L'étude et l'enseignement du vocabulaire, éditions*. Nathan Université, Paris.
- Pineau, G., Robin, J.-Y., Boutinet, J.-P., & Denoyel, N. (2007). *Penser l'accompagnement adulte : Ruptures, transitions, rebonds*. Presses Universitaires de France.
- Pinsky, L., & Theureau, J. (1992). *Concevoir pour l'action et la communication : Essais d'ergonomie cognitive*. P. Lang.

- Plane, S., & Schneuwly, B. (2000). Regards sur les outils de l'enseignement du français : Un premier repérage. *Repères. Recherches en didactique du français langue maternelle*, 22(1), 3–17.
- Prost, A. (2001). *Pour un programme stratégique de recherche en éducation, rapport au ministre de l'Éducation nationale*.
- Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. D. (2010). A tiered intervention model for early vocabulary instruction : The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research & Practice*, 25(3), 110–123.
- Rabardel, P. (2005). Instrument, activité et développement du pouvoir d'agir. In P Rabardel & P. Pastré, *Collection Travail & activité humaine. Modèles du sujet pour la conception dialectiques, activités, développement* (p. 11-29). Toulouse: Octarès.
- Rabardel, P., & Pastré, P. (Éd.). (2005). *Modèles du sujet pour la conception* (Octares Editions). Toulouse: Octares.
- Rabardel, Pierre. (1995). *Les hommes et les technologies; approche cognitive des instruments contemporains* (Armand Colin).
- Ramel, S., & Bovey, L. (2017). De l'usage des outils issus de la recherche pour accompagner les établissements scolaires. *Revue française de pédagogie. Recherches en éducation*, (200), 61-67. <https://doi.org/10.4000/rfp.6993>
- Ravitch, D. (1985). *The schools we deserve : Reflections on the educational crises of our times*. Basic Books New York.
- Reardon, S. F., Valentino, R. A., & Shores, K. A. (2012). Patterns of Literacy among U.S. Students. *The Future of Children*, 22(2), 17-37. <https://doi.org/10.1353/foc.2012.0015>
- Recherches, revue de didactique et de pédagogie du français n°60. (2014). *Outils*. Septentrion Éditions.
- Reese, E., Cox, A., Harte, D., & McAnally, H. (2003). Diversity in adults' styles of reading books to children. In *On reading books to children* (p. 49–68). Routledge.
- Reese, E., Haden, C. A., & Fivush, R. (1993). Mother-child conversations about the past : Relationships of style and memory over time. *Cognitive Development*, 8(4), 403-430. [https://doi.org/10.1016/S0885-2014\(05\)80002-4](https://doi.org/10.1016/S0885-2014(05)80002-4)
- Repères. Recherches en didactique du français langue maternelle n°22. (2000). *Les outils d'enseignement du français*. Lyon : ENS Editions.
- Resnicow, K., Davis, M., Smith, M., Lazarus-Yaroch, A., Baranowski, T., Baranowski, J., ... Wang, D. T. (1998). How best to measure implementation of school health curricula : A comparison of three measures. *Health Education Research*, 13(2), 239–250.
- Reuter, Y, Cohen-Azria, C., Daunay, B., Delcambre, I., & Lahanier-Reuter, D. (2013). *Dictionnaire des concepts fondamentaux des didactiques* (3e éd. actualisée). Louvain-la-Neuve: De Boeck.
- Reuter, Yves, Condette, S., & Boulanger, L. (2013). Les expérimentations « article 34 de la loi de 2005 ». Bilan et discussion d'une recherche sur des pratiques scolaires « innovantes ». *Les Sciences de l'éducation - Pour l'Ere nouvelle*, Vol. 46(3), 13-39.

- Rey, A. (1998). Dictionnaire historique de la langue française (1992). Paris, Le Robert, édition brochée.
- Rey, O. (2014). *Entre laboratoire et terrain : Comment la recherche fait ses preuves en éducation*. Dossier de veille de l'IFÉ.
- Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational psychology*, 86(1), 54.
- Rochex, J.-Y. (2013). Des inégalités scolaires, des moyens de les mesurer et d'en étudier les processus de production. *Le français aujourd'hui*, n°183(4), 9-28.
- Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.
- Romian, H. (1979). *Pour une pédagogie scientifique du français*. FeniXX.
- Roquet, P. (2009). L'émergence de l'accompagnement. Une nouvelle dimension de la formation. *Recherche et formation*, (62), 13-24. <https://doi.org/10.4000/rechercheformation.398>
- Rosenthal, J., & Ehri, L. C. (2011). Pronouncing New Words Aloud during the Silent Reading of Text Enhances Fifth Graders' Memory for Vocabulary Words and Their Spellings. *Reading and Writing: An Interdisciplinary Journal*, 24(8), 921-950. <https://doi.org/10.1007/s11145-010-9239-x>
- Roux-Baron, I., Cèbe, S., & Goigoux, R. (2017). Évaluation des premiers effets d'un enseignement fondé sur l'outil didactique *Narramus* à l'école maternelle. *Revue française de pédagogie. Recherches en éducation*, (201), 83-104.
- 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*, 44(2), 252–297.
- Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). *Evidence based medicine: What it is and what it isn't*. British Medical Journal Publishing Group.
- Saussez, F., & Lessard, C. (2009). Entre orthodoxie et pluralisme, les enjeux de l'éducation basée sur la preuve. *Revue française de pédagogie. Recherches en éducation*, (168), 111–136.
- Savoie-Zajc, L. (2010). Les dynamiques d'accompagnement dans la mise en place de communautés d'apprentissage de personnels scolaires. *Revue Éducation & Formation*, 292, 9–20.
- Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language*, 24(1), 123-138. <https://doi.org/10.1017/S0305000996003005>
- Sénéchal, M. (2000). Examen du lien entre la lecture de livres et le développement du vocabulaire chez l'enfant préscolaire. *Enfance*, 53(2), 169-186. <https://doi.org/10.3406/enfan.2000.3175>
- Sénéchal, M., LeFevre, J.-A., Hudson, E., & Lawson, E. P. (1996). Knowledge of storybooks as a predictor of young children's vocabulary. *Journal of Educational Psychology*, 88(3), 520-536. <https://doi.org/10.1037/0022-0663.88.3.520>

- Sénéchal, M., Thomas, E., & Monker, J.-A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology, 87*(2), 218-229. <https://doi.org/10.1037/0022-0663.87.2.218>
- Sénéchal, M., Whissell, J., & Bildfell, A. (2017). Starting from home : Home literacy practices that make a difference. *Theories of Reading Development, 383-408*.
- Séré, A., & Bassy, A.-M. (2010). *Le manuel scolaire à l'heure du numérique—Une « nouvelle donne » de la politique de ressources pour l'enseignement* (rapport public N° 2010-087). Consulté à l'adresse <http://www.ladocumentationfrancaise.fr/rapports-publics/114000048/index.shtml>
- Sigel, I. E. (1982). The relationship between parental distancing strategies and the child's cognitive behavior. In *Families as learning environments for children* (p. 47–86). Springer.
- Slavin, R. E. (1995). Best evidence synthesis : An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology, 48*(1), 9-18. [https://doi.org/10.1016/0895-4356\(94\)00097-A](https://doi.org/10.1016/0895-4356(94)00097-A)
- Slavin, R. E. (2002). Evidence-based education policies : Transforming educational practice and research. *Educational researcher, 31*(7), 15–21.
- Slavin, R. E. (2008). Perspectives on Evidence-Based Research in Education—What Works? Issues in Synthesizing Educational Program Evaluations. *Educational Researcher, 37*(1), 5-14. <https://doi.org/10.3102/0013189X08314117>
- Slavin, R. E. (2019). How Evidence-Based Reform Will Transform Research and Practice in Education. *Educational Psychologist, 0*(0), 1-13. <https://doi.org/10.1080/00461520.2019.1611432>
- 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, 68*(7), 560-571.
- Snow, C. E. (1991). The theoretical basis for relationships between language and literacy in development. *Journal of Research in Childhood education, 6*(1), 5–10.
- Snow, C. E. (2015). 2014 Wallace Foundation distinguished lecture : Rigor and realism: Doing educational science in the real world. *Educational Researcher, 44*(9), 460–466.
- Snow, C. E., & Beals, D. E. (2006). Mealtime talk that supports literacy development. *New Directions for Child and Adolescent Development, 2006*(111), 51-66. <https://doi.org/10.1002/cd.155>
- 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, 20*(1). Consulté à l'adresse <https://journals.uncc.edu/dialog/article/view/543>
- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction : A model-based meta-analysis. *Review of Educational Research, 56*(1), 72-110. <https://doi.org/10.2307/1170287>
- Stahl, S. A., & Nagy, W. E. (2007). *Teaching Word Meanings*. Routledge.
- Stanley, C. T., Petscher, Y., & Catts, H. (2018). A longitudinal investigation of direct and indirect links between reading skills in kindergarten and reading comprehension in tenth

- grade. *Reading and Writing*, 31(1), 133-153. <https://doi.org/10.1007/s11145-017-9777-6>
- Stone, C. A. (1998). Moving validated instructional practices into the classroom : Learning from examples about the rough road to success. *Learning Disabilities Research & Practice*, 13(3), 121-125.
- Stringfield, S., Datnow, A., Ross, S. M., & Snively, F. (1998). Scaling up school restructuring in multicultural, multilingual contexts : Early observations from Sunland County. *Education and Urban Society*, 30(3), 326–357.
- Suchaut, B. (2003). De la nécessité d'évaluer les pratiques enseignantes : Entre enjeux sociaux et obstacles méthodologiques. *Les Dossiers des sciences de l'éducation*, (10), 17–29.
- Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence : The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cognitive Development*, 47, 82-95. <https://doi.org/10.1016/j.cogdev.2018.04.005>
- Swanson, E., Vaughn, S., Wanzek, J., Petscher, Y., Heckert, J., Cavanaugh, C., ... Tackett, K. (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*, 44(3), 258-275. <https://doi.org/10.1177/0022219410378444>
- Symons, D. K., Peterson, C. C., Slaughter, V., Roche, J., & Doyle, E. (2005). Theory of mind and mental state discourse during book reading and story-telling tasks. *British Journal of Developmental Psychology*, 23(1), 81-102. <https://doi.org/10.1348/026151004X21080>
- Tardif, M., & Lessard, C. (1999). *Le travail enseignant au quotidien : Expérience, interactions humaines et dilemmes professionnels*. Presses Université Laval.
- Tauveron, C. (2002). *Lire la littérature à l'école : Pourquoi et comment conduire cet apprentissage spécifique?: de la GS au CM2*. Paris: Hatier.
- Teale, W. H. (2003). Reading Aloud to Young Children as a Classroom Instructional Activity : Insights from Research and Practice. In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Éd.), *On Reading Books to Children : Parents and Teachers* (p. 114-135). Mahwah, NJ: Lawrence Erlbaum Associates.
- Teale, W. H., & Sulzby, E. (1999). Literacy acquisition in early childhood : The roles of access and mediation in storybook reading. In D. A. Wagner (Éd.), *The future of literacy in a changing world* (p. 131–150). Cresskill, NJ: Hampton Press.
- Theureau, J. (2011). Appropriation 1, 2, 3 ou Appropriation, Incorporation & 'Inculturation'. *Conférence journée Ergo-Idf Appropriation & Ergonomie*, 16, 06–11.
- Thurler, M. G. (2000). L'innovation négociée : Une porte étroite. *Revue française de pédagogie*, 29–42.
- Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex : Components of the simple view of reading across grade levels. *Journal of Research in Reading*, 32(4), 383-401. <https://doi.org/10.1111/j.1467-9817.2009.01401.x>
- 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 ? Consulté à l'adresse En ligne : <<http://www.cnesco.fr/fr/lecture/paroles-dexperts/lecture-a-lheure-du-numerique/>>

- Tricot, A. (2017). *L'innovation pédagogique*. Paris: Retz.
- Tricot, A. (2018). De l'utilité des connaissances produites par la recherche en éducation. *Diversité*, (192), 35-40.
- Tricot, A., Plécat-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. *Environnements Informatiques pour l'Apprentissage Humain 2003*, 391–402. ATIEF; INRP.
- Tseng, V. (2012). *The uses of research in policy and practice*. Society for Research in Child Development Washington, DC.
- Tyack, D. B., & Cuban, L. (1995). *Tinkering toward utopia*. Harvard University Press.
- Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond Vocabulary : Exploring Cross-Disciplinary Academic-Language Proficiency and Its Association With Reading Comprehension. *Reading Research Quarterly*, 50(3), 337-356. <https://doi.org/10.1002/rrq.104>
- Uwamariya, A., & Mukamurera, J. (2005). Le concept de «développement professionnel» en enseignement : Approches théoriques. *Revue des sciences de l'éducation*, 31(1), 133–155.
- Vaughn, S., Hughes, M. T., Schumm, J. S., & Klingner, J. (1998). A collaborative effort to enhance reading and writing instruction in inclusion classrooms. *Learning disability quarterly*, 21(1), 57–74.
- Vaughn, S., Klingner, J., & Hughes, M. (2000). Sustainability of research-based practices. *Exceptional Children*, 66(2), 163–171.
- Veneziano, E. (2010). *Peut-on aider l'enfant à mieux raconter ? Les effets de différentes méthodes d'intervention*. <https://halshs.archives-ouvertes.fr/halshs-00413310/document>
- Vinel, É. (2014). Comparaison des conduites de récit d'adultes et d'enfants dans deux situations de lecture d'albums. *Repères. Recherches en didactique du français langue maternelle*, (50), 105-130. <https://doi.org/10.4000/reperes.781>
- Vonk, J. H. C. (1988). L'évolution professionnelle des enseignants débutants et ses répercussions sur la formation initiale et continue. *Recherche & formation*, 3(1), 47–60.
- Vygotski, L. S. (1925/1994). Le problème de la conscience dans la psychologie du comportement (F. Sève, Trad.). *Société française*, 50, 35–47.
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5-23. <https://doi.org/10.1007/BF02504682>
- Wasik, B. A., & Hindman, A. H. (2018). Why Wait? The Importance of Wait Time in Developing Young Students' Language and Vocabulary Skills. *The Reading Teacher*, 72(3), 369-378. <https://doi.org/10.1002/trtr.1730>

- Wasik, B. A., Hindman, A. H., & Snell, E. K. (2016). Book reading and vocabulary development : A systematic review. *Early Childhood Research Quarterly*, 37, 39-57. <https://doi.org/10.1016/j.ecresq.2016.04.003>
- Wasik, B., & Bond, M. (2001). Beyond the Pages of a Book : Interactive Book Reading and Language Development in Preschool Classrooms. *Journal of Educational Psychology*, 93, 243-250. <https://doi.org/10.1037/0022-0663.93.2.243>
- Weimer, A. A., & Guajardo, N. R. (2005). False belief, emotion understanding, and social skills among Head Start and non-Head Start children. *Early Education and Development*, 16(3), 341–366.
- West, M. A., & Altink, W. M. (1996). Innovation at work : Individual, group, organizational, and socio-historical perspectives. *European Journal of Work and Organizational Psychology*, 5(1), 3–11.
- Whitehurst, G., Falco, F., Lonigan, C., Fischel, J., DeBaryshe, B., Valdez-Menchaca, M., & Caulfield, M. (1988). Accelerating Language Development Through Picture Book Reading. *Developmental Psychology*, 24(4), 552-559. <https://doi.org/10.1037/0012-1649.24.4.552>
- Whitty, G. (2006). Education (al) research and education policy making : Is conflict inevitable? *British educational research journal*, 32(2), 159–176.
- Wittorski, R. (2008). La professionnalisation. *Savoirs*, n° 17(2), 9-36.
- Wright, T. S., & Cervetti, G. N. (2017). A Systematic Review of the Research on Vocabulary Instruction That Impacts Text Comprehension. *Reading Research Quarterly*, 52(2), 203-226. <https://doi.org/10.1002/rrq.163>
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., ... Zaslow, M. J. (2013). *Investing in our future : The evidence base on preschool education*. Society for Research in 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*, 65(6), 1523–1545.
- Zipoli, R. P., Coyne, M. D., & McCoach, D. B. (2011). Enhancing Vocabulary Intervention for Kindergarten Students : Strategic Integration of Semantically Related and Embedded Word Review. *Remedial and Special Education*, 32(2), 131-143. <https://doi.org/10.1177/0741932510361262>
- 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. Recherches en éducation*, (193), 57-76. <https://doi.org/10.4000/rfp.4890>
- Zucker, T. A., Cabell, S. Q., Justice, L. M., Pentimonti, J. M., & Kaderavek, J. N. (2013). The role of frequent, interactive prekindergarten shared reading in the longitudinal development of language and literacy skills. *Developmental Psychology*, 49(8), 1425.

Table des figures

Figure 1. Les connaissances utiles à la conception d'un outil didactique	67
Figure 2. Les trois étapes d'une conception continuée dans l'usage	89
Figure 3. Pictogrammes présentés aux élèves de l'outil <i>Narramus</i>	98
Figure 4. Répartition géographique des académies engagées dans la recherche	115
Figure 5. Synthèse des regroupements proposés en 2016-2017 autour de la recherche <i>Narramus</i>	152
Figure 6. Histogramme de la fidélité déclarée par les enseignants des groupes G1 et G2 pour chacun des trois scénarios (en %)	161
Figure 7. Histogramme de la durée d'enseignement déclarée par les enseignants des 3 groupes pour les trois scénarios (en minutes)	164
Figure 8. Performances des élèves de GS en rappel de récit (macrostructure) : comparaison des trois groupes aux pré- et post-tests des trois scénarios <i>Narramus</i> (un par trimestre)	175
Figure 9. Performances des élèves de GS en rappel de récit (microstructure) : comparaison des trois groupes aux pré- et post-tests des trois scénarios <i>Narramus</i> (un par trimestre)	175
Figure 10. Performances des élèves de MS en rappel de récit (macrostructure) : comparaison des trois groupes aux pré- et post-tests des trois scénarios <i>Narramus</i> (un par trimestre)	179
Figure 11. Performances des élèves de MS en rappel de récit (microstructure) : comparaison des trois groupes aux pré- et post-tests des trois scénarios <i>Narramus</i> (un par trimestre)	180
Figure 12. Comparaison des progrès des élèves de Grande Section en lexique	183
Figure 13. Comparaison des progrès des élèves de Moyenne Section en lexique	185
Figure 14. Progrès des élèves de GS en rappel de récit en fonction de leur niveau initial	196
Figure 15. Progrès des élèves de GS en lexique en fonction de leur niveau initial (en % de réponses correctes)	197
Figure 16. Progrès des élèves de MS en rappel de récit en fonction de leur niveau initial....	200
Figure 17. Progrès des élèves de MS en lexique en fonction de leur niveau initial (en % de réponses correctes)	201
Figure 18. Répartition géographique des académies engagées dans l'étude 2	254
Figure 19. Extrait du TVAP	255
Figure 20. Dispersion des scores des élèves des trois groupes au TVAP (/60)	257
Figure 21. Comparaison des scores au test TVAP des groupes A, B et C de l'étude 2 et du pré-test de la recherche Lire-Écrire en 2013 (/60)	258
Figure 22. Comparaison des scores des élèves de REP aux trois épreuves	264
Figure 23. Comparaison des scores des élèves de REP en macrostructure (/9)	265
Figure 24. Comparaison des scores des élèves de REP en microstructure (/48)	265
Figure 25. Comparaison des scores des élèves des groupes A, B et C aux trois épreuves	268
Figure 26. Comparaison des scores des élèves des groupes A, B et C aux quatre épreuves..	274

Figure 27. Nuages de mots représentant la fréquence des deux caractéristiques de l’outil qui ont profondément influencé l’évaluation des pratiques	300
Figure 28. Graphe des mots associés aux caractéristiques de l’outil qui ont profondément influencé leurs pratiques.....	301
Figure 29. Nuage de mots représentant les plus-values attribuées à l’accompagnement par les enseignants	303

Table des tableaux

Tableau 1. Performances des élèves en compréhension de phrases et de textes en fonction du milieu social à l'entrée au CP	24
Tableau 2. Répartition des élèves dans les groupes en compréhension d'écrits oralisés au début du CP et à mi-CP selon le secteur de scolarisation (en %)	26
Tableau 3. Définition des seuils pour chaque épreuve et pourcentage d'élèves les ayant atteints au début du CP	27
Tableau 4. Écarts de performances entre élèves scolarisés dans le secteur public hors EP et élèves scolarisés en EP (différences de proportions et odds ratio)	29
Tableau 5. Répartition des élèves dans les groupes en compréhension d'écrits oralisés au début du CE1 selon le secteur de scolarisation (en %)	30
Tableau 6. Les modalités d'accompagnement du changement Rowan & Miller (2007)	57
Tableau 7. Exemples de commentaires rédigés dans les guides du maître <i>Narramus</i>	96
Tableau 8. Hypothèses et questions de recherche	111
Tableau 9. Nombre de classes et d'élèves évalués dans chaque groupe en PS, MS et GS	125
Tableau 10. Albums étudiés (par niveau) dans les trois groupes expérimentaux en 2016/2017	126
Tableau 11. Organisation temporelle du pré-test au post-test	127
Tableau 12. Répartition des évaluations au cours de l'année scolaire	129
Tableau 13. Nombre de propositions macrostructurale et microstructurale pour chaque album étudié	133
Tableau 14. Organisation temporelle des épreuves de lexique différencié	138
Tableau 15. Carnet de bord des enseignants G1 et G2	140
Tableau 16. Carnet de bord des enseignants G3	141
Tableau 17. Données descriptives pour la variable Fidélité exprimée en pourcentages de tâches prescrites réalisées	160
Tableau 18. Données descriptives pour la variable Temps des classes de MS et GS	162
Tableau 19. Données descriptives pour la variable Temps des classes de PS	163
Tableau 20. Résultats initiaux des élèves de PS, MS et GS (pré-test T1)	165
Tableau 21. Résultats des élèves de MS et GS au test É.Co.S.Se	166
Tableau 22. Données descriptives pour l'accord inter-juge en rappel de récit	168
Tableau 23. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de MS du groupe 1	168
Tableau 24. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de MS du groupe 2	169
Tableau 25. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de MS du groupe 3	169

Tableau 26. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de GS du groupe 1	170
Tableau 27. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de GS du groupe 2	171
Tableau 28. Corrélations entre les variables macrostructure, microstructure et questionnaire (post-test) pour les élèves de GS du groupe 3	171
Tableau 29. Comparaison des progrès des élèves de Grande Section en rappel de récit	173
Tableau 30. Comparaison des progrès des élèves de Grande Section au questionnaire	177
Tableau 31. Comparaison des progrès des élèves de Moyenne Section en rappel de récit....	178
Tableau 32. Comparaison des progrès des élèves de Moyenne Section au questionnaire	181
Tableau 33. Comparaison des progrès des élèves de Grande Section en lexique	182
Tableau 34. Comparaison des progrès des élèves de Moyenne Section en lexique	184
Tableau 35. Résultats des élèves de Moyenne et Grande Section au post-test différé de lexique	185
Tableau 36. Analyses de régression multiple en GS (modèle $S_f = S_i \times T_p \times G$)	187
Tableau 37. Analyses de régression multiple en MS (modèle $S_f = S_i \times T_p \times G$)	188
Tableau 38. Répartition des élèves des trois groupes en fonction de leur niveau initial au T1 en macrostructure, en microstructure et en lexique	190
Tableau 39. Comparaison des progrès des élèves de Grande Section G1&G2 vs G3 en fonction du niveau initial de l'élève (rappel de récit et lexique)	192
Tableau 40. Comparaison des progrès des élèves de Moyenne Section G1&G2 vs G3 en fonction du niveau initial de l'élève (rappel de récit et lexique)	194
Tableau 41. Comparaison des progrès des élèves de Grande Section travaillant avec <i>Narramus</i> en rappel de récit et lexique en fonction de leur niveau initial (NI)	198
Tableau 42. Analyse des progrès des élèves de GS travaillant avec <i>Narramus</i> entre les pré-tests et les post-tests en rappel et lexique en fonction du niveau initial	199
Tableau 43. Comparaison des progrès des élèves de Moyenne Section travaillant avec <i>Narramus</i> en rappel de récit et lexique en fonction de leur niveau initial (NI)	202
Tableau 44. Analyse des progrès des élèves de MS travaillant avec <i>Narramus</i> entre les pré-tests et les post-tests en rappel et lexique en fonction du niveau initial	203
Tableau 45. Effets du score initial en macrostructure, du temps et du groupe sur les performances finales en macrostructure selon le score initial de l'élève de GS	205
Tableau 46. Effets du score initial en macrostructure, du temps et du groupe sur les performances finales en macrostructure selon le score initial de l'élève de MS	206
Tableau 47. Effets du score initial en lexique, du temps et du groupe sur les performances finales en lexique selon le score initial de l'élève de GS	207
Tableau 48. Effets du score initial en lexique, du temps et du groupe sur les performances finales en lexique selon le score initial de l'élève de MS	209
Tableau 49. Comparaison des progrès des élèves de Petite Section en rappel de récit	212

Tableau 50. Comparaison des progrès des élèves de Petite Section au questionnaire	213
Tableau 51. Comparaison des scores des élèves de Petite Section en post-test immédiat de lexique	214
Tableau 52. Résultats des élèves de Petite Section au post-test différé de lexique	214
Tableau 53. Analyse de régression multiple en PS (modèle $S_f = S_i \times T_p \times G$)	216
Tableau 54. Nombre de macro-propositions et de micro-propositions présentes dans les albums proposés aux élèves de MS et GS en fonction de l'année	221
Tableau 55. Comparaison des pourcentages de réussite des élèves de MS (G1 + G2) de l'année 1 et des élèves de MS (G) de l'année 2 en macrostructure et en lexique	222
Tableau 56. Comparaison des résultats des élèves de GS (G1 et G2) de l'année 1 et des élèves de GS (=G) de l'année 2	224
Tableau 57. Scores des élèves de PS puis de MS en rappel de récit et en lexique de T1 à T12	226
Tableau 58. Score des élèves de MS puis de GS en rappel de récit et en lexique de T1 à T12	228
Tableau 59. Données descriptives pour les variables Fidélité et Temps du scénario Gruffalo GS	231
Tableau 60. Comparaison des résultats des élèves de Grande Section aux épreuves du scénario Gruffalo (année 1 vs année 3)	232
Tableau 61. Nombre d'élèves dans chaque groupe	255
Tableau 62. Résultats des élèves des trois groupes au TVAP	256
Tableau 63. Résultats des élèves en début de CP au TVAP	257
Tableau 64. Coefficients de corrélation intra classe.....	260
Tableau 65. Comparaison des scores des élèves des groupes A et B aux trois épreuves	264
Tableau 66. Corrélations des variables macrostructure, microstructure et questionnaire pour les groupes A et B	266
Tableau 67. Comparaison des scores des élèves des groupes A et C aux trois épreuves	267
Tableau 68. Comparaison de scores (bruts et %) des trois groupes dans l'académie de Clermont-Ferrand à l'épreuve questionnaire	269
Tableau 69. Comparaison de scores (bruts et %) des trois groupes dans l'académie de Clermont-Ferrand à l'épreuve lexique	271
Tableau 70. Comparaison de scores (bruts et %) des trois groupes dans l'académie de Clermont-Ferrand pour les items tiptoper et dégringoler	277
Tableau 71. Les points de réussite identifiés par les enseignants du G1 et du G2 (questionnaire 1)	290
Tableau 72. Les gestes professionnels modifiés en fonction du groupe (questionnaire 1)	293
Tableau 73. Les belles inventions des enseignants du G1 et du G2 (questionnaire 1)	295
Tableau 74. Les activités détestées par les enseignants du G1 et du G2 (questionnaire 1)....	296

Tableau 75. Ce qu'ont retenu les enseignants du G1 et du G2 (questionnaire 1)	297
Tableau 76 Les modifications des pratiques liées à l'utilisation de l'outil	299

Table des annexes

Annexe 1 : Exemple d'une fiche de préparation détaillée de Narramus	
Annexe 2 : Grilles de codage : macrostructure et microstructure des albums étudiés	
Annexe 3 : Grilles de codage : lexique des albums étudiés	
Annexe 4 : Grilles de codage : post-tests différés de lexique.....	
Annexe 5 : Grilles de codage : questionnaires des albums étudiés	
Annexe 6 : Questionnaire n°1 pour les enseignants	
Annexe 7 : Questionnaire n°2 pour les enseignants	
Annexe 8 Comparaison des progrès des élèves de GS et MS (G1&G2 vs G3) en fonction du niveau initial de l'élève (rappel de récit et lexique)	
Annexe 9 : Texte de l'album La petite marmite qui tiptopait	
Annexe 10 : Grilles de codage de rappel de l'épreuve sur l'album La petite marmite qui tiptopait : macrostructure et microstructure	
Annexe 11 : Grille de codage de l'épreuve questionnaire	
Annexe 12 : Grille de codage de l'épreuve lexique	

Résumé

Pour que l'école maternelle française puisse jouer un rôle compensatoire des inégalités sociales, 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. C'est dans ce but que l'outil didactique *Narramus* a été conçu.

Dans notre recherche doctorale, nous avons voulu savoir si cet outil produisait les effets escomptés sur le développement des compétences langagières des élèves et sur les pratiques des enseignants. Pour cela nous nous sommes posé trois questions :

- une première d'ordre socio-pédagogique : l'utilisation d'un outil didactique innovant peut-elle favoriser les apprentissages langagiers de tous les élèves et contribuer à réduire les inégalités de réussite à l'école ?
- une seconde d'ordre méthodologique : comment construire et mettre en œuvre un dispositif expérimental d'évaluation présentant des critères rigoureux pour tester l'efficacité d'un outil didactique ?
- une troisième relative au rôle de l'outillage et de l'accompagnement dans l'amélioration des pratiques d'enseignement : quels liens existe-t-il entre l'utilisation de l'outil, l'accompagnement assuré par des équipes de circonscription et le développement professionnel des enseignants ?

Pour répondre à ces questions, nous avons mené deux études quasi-expérimentales (impliquant 250 classes et 1 500 élèves de petite, moyenne et grande section) et élaboré deux protocoles expérimentaux qui respectent les critères méthodologiques exigeants définis par les promoteurs de l'*Evidence Based Education*.

La première étude visait à identifier un éventuel effet *Narramus* sur les apprentissages enfantins et à mesurer l'influence de l'accompagnement de cette innovation sur le développement professionnel des enseignants. Nos données montrent que les élèves qui ont bénéficié de *Narramus* obtiennent de meilleurs résultats que leurs camarades du groupe contrôle sur l'ensemble des dimensions évaluées. De plus, les différences entre les deux groupes outillés (avec ou sans accompagnement) ne sont pas significatives : l'accompagnement n'est pas indispensable pour qu'un outil, s'il est bien conçu, provoque les effets positifs attendus. Nous avons cependant montré que l'accompagnement, s'il n'affecte pas directement les performances des élèves, peut agir comme une aide à l'identification de gestes professionnels nouveaux, mobilisables dans d'autres domaines d'apprentissage et semble, à ce titre, être un vecteur de développement professionnel.

La deuxième étude a mis en évidence une réduction des inégalités sociales et un transfert des compétences construites avec *Narramus* vers d'autres textes que ceux étudiés en classe : les élèves qui en ont bénéficié comprennent et racontent mieux que leurs camarades de même condition sociale et leurs performances s'approchent de celles d'enfants favorisés. Leur vocabulaire est lui-aussi enrichi.

Ces conclusions valident la pertinence de l'outil et concourent à valoriser un enseignement explicite et intégré de la compréhension et du vocabulaire auprès de jeunes élèves.

Mots-clés : évaluation - inégalités sociales - innovation - école maternelle - langage – développement professionnel

Abstract

If French nursery schools want to play a compensatory role for social inequalities, they have to allow young learners to demonstrate skills (at school) that the most privileged among them can already develop within their family surrounding. These skills account for the best viaticum for their future school curriculums: understanding of written narratives, being able to recount them and acquiring a lexicon. The teaching tool *Narramus* was created for that purpose.

As part of our doctoral research, we wanted to know whether this tool has the expected impact on developing students' language skills and on teaching methods.

For this purpose, we were interested in three questions :

- The first one deals with a socio-pedagogical issue: does using an innovative teaching tool make the language learning process of students better and contribute to reduce unequal achievement in learning at school?
- The second one deals with a methodological issue: how can we build an experimental set-up in order to evaluate the efficiency of the teaching tool based on strict criteria ?
- The third one is related to the role played by the equipment and the support for the improvement of teaching methods: what is the connection between the use of the tool, the support provided by school district teams and the professional development of teachers?

In order to answer these questions, we conducted two studies almost experimental within 250 classes, including 1500 students from the first, second and third years of nursery schools. We drew up two experimental protocols abiding by strict methodological criteria defined by the developers of *Evidence Based Education*.

The first study aimed to identify some potential *Narramus* effect on children learning process and to examine the influence of the support for this innovative tool on the professional development of teachers. Our data showed that students who had access to *Narramus* had better results than their peers from the control group in all areas assessed. Besides, the differences between the two equipped groups (with or without support) are not significant: the support is not essential to the tool (if well-conceived) to produce the expected positive effects.

However, we showed that the support, if it does not directly affect the students' performances, can although help teachers to identify new professional methods that they can re-use in other learning settings, and thus seems to be a vector of professional development.

The second study brought to light the reduction of social inequalities and a transfer of skills built through *Narramus* to other texts than those studied in class: students who have implemented it understand and recount better than their peers from the same social background and their proficiency level is close to that of more privileged children. Their vocabulary is also enhanced.

These conclusions validate the relevance of the tool and contribute to promote an integrated and explicit teaching of comprehension and of vocabulary to young students.

Key words : evaluation – social inequalities – innovation – nursery school – language – professional development.