

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

É. GENTAZ*, **, L. SPRENGER-CHAROLLES***, ****, P. COLÉ****, A. THEUREL**, M. GURGAND*****, C. HURON*****, T. ROCHER*****, M. LE CAM*****

* Laboratoire de psychologie et neurocognition (UMR CNRS 5105), Université de Grenoble, France.

** Faculté de psychologie et sciences de l'éducation, Université de Genève, Suisse.

*** Laboratoire de psychologie de la perception (UMR CNRS 8158), Paris-Descartes, PRES Sorbonne-Paris-Cité, France.

**** Laboratoire de psychologie cognitive (UMR CNRS 7290), Aix-Marseille Université, France.

***** École d'économie de Paris, France.

***** NeuroSpin CEA, France.

***** DEPP, France.

Correspondance : Pr. Édouard Gentaz, Faculté de psychologie et des sciences de l'éducation, Université de Genève.

E-mail: edouard.gentaz@unige.ch

RÉSUMÉ : 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

Cette recherche a évalué les effets d'un programme combinant des entraînements fondés sur le décodage et la compréhension sur les performances en lecture d'enfants scolarisés en CP dans des réseaux d'éducation prioritaire par rapport à une méthode classique (sans entraînements spécifiques systématiques), et ce à grande échelle et dans des conditions réalistes. Les résultats des analyses quantitatives montrent que ce programme n'a pas permis de faire davantage progresser en lecture les enfants du groupe test que ceux des groupes témoin. Cette étude, unique en son genre par son ampleur et sa méthodologie, permet de souligner que la mise en œuvre d'expérimentations à grande échelle présente des difficultés spécifiques qui font partie des enjeux de recherche et d'évaluation, et qui doivent être résolues pour faire progresser notre système éducatif.

Mots clés : Apprentissage – Lecture – Entraînements – Pédagogie – Évaluations – Éducation prioritaire.

SUMMARY: Quantitative assessment of large-scale reading practice for pupils who start Grade 1 in educational priority networks : advantages and limitations

This research evaluates the effects of a program combining training based on decoding and understanding on the reading performance of children enrolled in grade one in priority education networks compared to a conventional training and the scale and implemented at realistic scale and conditions. Quantitative analysis of the results show that the program has failed to generate further progress in reading for children in the experimental compared to the control group. In conclusion we highlight the contributions and originality of this experiment and the fact that implementing such an innovative teaching method at a large scale raises specific difficulties that should be analyzed in terms of research and evaluation in view of improving our educational system.

Key words: Learning – Reading – Trainings – Teaching – Assessment – Priority education.

RESUMEN: Evaluación cuantitativa de un entrenamiento de lectura a gran escala para niños de 1º de Primaria escolarizados en las Redes de Educación prioritaria: aportaciones y límites

Esta investigación evaluó los efectos de un programa de entrenamiento basado en la combinación de "decodificar" y "entendimiento" en el rendimiento de lectura de los niños matriculados en el primer grado en las redes de educación prioritaria en comparación con un "método convencional" (sin entrenamientos específicos sistemáticos) y, todo ello, a gran escala y en condiciones "realistas".

Los resultados de los análisis cuantitativos muestran que el programa no ha logrado que los niños del grupo experimental avancen más en la lectura que los del grupo control. Este estudio, único en su alcance y metodología, sirve para enfatizar que la realización de experimentos a gran escala tiene dificultades específicas que forman parte de los retos de la investigación y la evaluación, que deben ser tratados para permitir el progreso del sistema educativo.

Palabras clave: Aprendizaje – Lectura – Entrenamiento – Pedagogía – Evaluaciones – Educación prioritaria.

INTRODUCTION

Apprendre à lire est une activité complexe, exigeante, longue et difficile qui implique un enseignement et un accompagnement systématiques et approfondis (Dehaene, Dehaene-Lambert, Gentaz, Huron & Sprenger-Charolles, 2011 ; Sprenger-Charolles & Colé, 2013). En France, si 5 % des enfants des classes ordinaires ont des difficultés de maîtrise de la langue écrite en CP, ce taux peut dépasser 25 % dans certaines zones défavorisées (cf. Bara, Colé, & Gentaz, 2008 ; Billard *et al.*, 2010 ; Gentaz, Sprenger-Charolles, Theurel & Colé, *in press*). Il est donc très important de proposer des entraînements qui favorisent de manière optimale l'apprentissage de la lecture chez les enfants au cours du cycle 2, et plus particulièrement pour ceux scolarisés dans des réseaux d'éducation prioritaire.

De nombreux travaux ont examiné les effets sur la lecture d'entraînements des capacités d'analyse phonémique. Ces entraînements comportent principalement des tâches orales impliquant soit la comparaison d'items qui ne diffèrent que par un phonème, soit la manipulation explicite des phonèmes (comptage, suppression ou fusion de phonèmes, cf. Gentaz & Sprenger-Charolles, 2012). Ils sont le plus souvent effectués en plusieurs séquences, très structurées, de courte durée. La méta-analyse d'Ehri *et al.* (2001) a montré que l'effet de ce type d'entraînement varie en fonction du niveau de lecture (réel ou prédit) des enfants : il est plus important chez ceux à risque de devenir mauvais lecteurs que chez ceux en difficulté de lecture, surtout à long terme. Ces entraînements sont aussi plus efficaces quand ils sont réalisés en petit groupe qu'en classe entière. Enfin, les effets les plus importants sont observés avec une durée totale des entraînements comprise entre 5 à 18 heures et des sessions individuelles de 25 minutes. Ces entraînements ont un effet à la fois sur les capacités de bas niveau en lecture (lecture de mots familiers et de mots inventés) et sur la compréhension écrite. Les entraînements les plus efficaces sont focalisés sur une ou deux tâches, surtout celles directement impliquées dans l'apprentissage de la lecture (par exemple, assembler des phonèmes pour produire un mot). De plus, l'effet le plus notable est obtenu quand, outre les activités d'analyse phonémique, les entraînements portent sur les relations graphème-phonème (Castles & Colheart, 2004). En effet, l'entraînement à l'analyse phonémique seul n'a qu'un effet limité sur les capacités ultérieures de lecture (Bradley & Bryant, 1983). La présence du support visuel de la lettre a également été relevée par Ehri *et al.* (2001) comme ayant un rôle crucial, parce qu'il permettrait d'expliciter le lien entre phonèmes et graphèmes (Bara, Gentaz, & Colé, 2004).

Contrairement à d'autres pays industrialisés, il existe relativement peu de recherches en France ou dans des pays francophones qui ont examiné les effets d'entraînement sur l'apprentissage de la lecture. Par exemple, Bianco *et al.* (2010) et Bianco, Pellenq, Lambert, Lima, & Doyen (2012) montrent que seuls des entraînements sur deux ans (moyenne et grande section de maternelle, MSM et GSM) ont un effet sur le niveau de lecture en CP : l'un à la compréhension MSM-GSM sur le niveau de compréhens-

sion en lecture ; l'autre phonologique MSM-GSM sur les capacités phonologiques de lecture-écriture. L'absence d'effet de l'entraînement phonologique effectué uniquement en GSM sur le niveau en lecture en CP est surprenante car il correspondait largement aux conditions optimums préconisées (petits groupes de 4 à 7 enfants, durée de 7 à 8 heures, exercices de segmentation et de discrimination phonémique et de mise en correspondance entre graphème et phonème). Il est possible que cette absence d'effet provienne du manque de formation préalable des enseignants. En effet, Blachman, Ball, Black et Tangel (1994) montrent des différences entre les recherches réalisées dans la situation ordinaire de la classe et celles effectuées par des chercheurs, le point capital étant que les enseignants impliqués doivent être bien formés (cf. Bara, Gentaz & Colé, 2007).

En 2007, une expérimentation (intitulée « Parler ») ayant pour objectif l'évaluation de trois entraînements centrés respectivement sur la compréhension (Bianco, Coda, Gourgue-Giollito, & Robert, 2005), la phonologie (Jacquier-Roux & Zorman, 1998) et la fluence de lecture (Pourchet & Zorman, 2008) a été effectuée dans huit classes (de la GSM au CE1) situées dans des zones d'éducation prioritaire. La mise en œuvre de ces entraînements sur une année scolaire a nécessité une réorganisation du travail de classe. Les élèves, répartis en petits groupes de besoin composés de 4 à 6 enfants de niveau assez homogène, ont bénéficié d'un entraînement structuré de 30 minutes effectué par leur enseignant. Pendant ces séances, les autres enfants, en petits groupes aussi, ont appris à travailler en autonomie. Pour dédoubler toutes les séances (celles dirigées comme celles en autonomie), cette étude a bénéficié d'un enseignant supplémentaire. Les résultats (non publiés, cf. Bressoux & Zorman, 2009) semblent encourageants : le niveau de lecture des enfants des classes expérimentales est supérieur à la moyenne nationale. Cependant, cette recherche présente un certain nombre de limites, parmi lesquelles 1) un faible effectif par niveau ainsi qu'une absence 2) de « randomisation », 3) d'entraînement basé sur le décodage, 4) de vérification des compétences dans d'autres domaines tels que les mathématiques, et 5) de contrôle des effets psychosociaux de l'expérimentation (effets des attentes, placebo, comparaisons sociales, etc.). En outre, les résultats significatifs apparaissent tardivement.

NOTRE ÉTUDE

L'objectif principal de notre étude est d'évaluer avec l'application la plus rigoureuse possible de la méthode expérimentale, à grande échelle et dans des conditions réalistes dictées par les contraintes des différents acteurs de l'Éducation nationale (enseignants, conseillers pédagogiques, inspecteurs de l'Éducation nationale, inspecteurs d'académie, recteur, direction générale de l'enseignement scolaire), les effets d'un programme combinant des entraînements centrés sur le décodage et la compréhension sur les performances en lecture d'enfants scolarisés en REP par rapport à une méthode classique (sans entraînements spécifiques). Le décodage est une compétence qui implique la capacité d'utiliser les correspondances

graphèmes-phonèmes (et donc celle de pouvoir isoler les phonèmes), qui doit devenir progressivement automatique. La compréhension est une compétence qui permet d'accéder au sens d'un énoncé, qu'il soit oral ou écrit. Il s'agit de la finalité de l'apprentissage de la lecture.

L'expérience s'est déroulée sur l'année scolaire du CP. Pour évaluer les progrès en lecture, les performances de tous les élèves des classes tests de l'académie de Lyon ont été mesurées avant et après les séances d'entraînement et ont été comparées à celles des élèves des classes témoin. Par ailleurs, les résultats des enfants des classes tests observés au début de l'année scolaire ont permis d'identifier leurs besoins afin d'adapter à chacun, qualitativement et quantitativement, les exercices proposés dans les différents entraînements. En parallèle, la direction de l'évaluation, de la prospective et de la performance (DEPP) du MEN a conduit sa propre évaluation (lecture et mathématiques) dans les mêmes classes ainsi que dans d'autres classes similaires issues de toute la France pour identifier les effets potentiels inhérents à toute participation à une expérimentation originale.

Au niveau éthique, cette étude a été conduite selon les règles internationales habituelles qui régissent les recherches comportementales non interventionnelles définies par la Déclaration d'Helsinki et le Comité d'éthique du LPNC-CNRS-UPMF, contrairement à ce suggère un rapport de l'Inspection générale (2012). Tous les enseignants ont été informés des conditions de l'expérimentation avant de décider d'y participer volontairement. Tous les parents des enfants des classes tests et témoins ont reçu une lettre présentant l'expérimentation signée par chacun des inspecteurs d'académie et par le responsable scientifique (EG). Seuls les enfants de ceux qui ont donné un accord signé ont participé aux deux phases d'évaluation CNRS. Enfin, l'expérimentation a fait l'objet d'une déclaration préalable à la CNIL déposée en juillet 2010.

MÉTHODE

Échantillons « Académie de Lyon » et « National »

L'expérimentation concerne un échantillon de 56 écoles de l'académie de Lyon. Cet échantillon est partagé entre 28 écoles qui forment le groupe test, et 28 écoles qui constituent le groupe témoin (*table 1*). Au total, ces écoles accueillent 2 398 élèves de CP mais les classes n'étaient pas toutes concernées par l'expérimentation car seules les classes avec des enseignants volontaires ont participé à cette étude. Par ailleurs, la DEPP a échantillonné, dans la base nationale, des écoles primaires accueillant des élèves de CP¹. Ces écoles constituent un groupe non-traité qui n'était pas affecté par le fait de participer à une expérimentation. Cet échantillon n'a pas été constitué selon le même protocole et n'est donc pas directement comparable à l'échantillon expérimental des 56 écoles. Cependant, on peut le comparer au groupe de

contrôle expérimental tenant compte des compétences initiales des élèves, de manière à corriger une partie des biais potentiels. Cet échantillon comporte 2 375 élèves dans 74 écoles. Pour des raisons de comparabilité, nous avons exclu les écoles privées de l'analyse, ce qui laisse 1 806 élèves dans 52 écoles.

In fine, les enfants des deux échantillons sont répartis en trois strates (ceux du privé étant exclus) : public hors éducation prioritaire (EP), réseaux ambition réussite-RAR, réseaux de réussite scolaire-RRS (ces réseaux remplaçant les réseaux d'éducation prioritaire, cf. <http://www.educationprioritaire.education.fr/echanger/questions-reponses.html>)

La comparaison des scores obtenus en français et mathématiques aux évaluations nationales CE1 et CM2 en 2011 par les élèves des écoles visées (soit 28 écoles tests, 28 écoles témoins et 52 écoles de l'échantillon national) ne révèle pas de différences significatives entre les différents groupes d'écoles lorsque l'on contrôle la structure des quatre strates (tous les $p > .05$). Les échantillons de cette étude sont donc comparables.

Participation aux différentes sessions de tests

Les élèves de l'échantillon de l'académie de Lyon ont passé deux fois dans l'année (vers octobre 2010, session 1 ou pré-test, et vers mai 2011, session 2 ou post-test) deux évaluations : d'une part, des évaluations élaborées par la DEPP (conçues en 1997 pour le panel CP, puis reprises pour le plan Illettrisme en 2003-2004) et effectuées par les enseignants en collectif (appelées par la suite « évaluations DEPP ») ; d'autre part, des évaluations conçues par l'équipe de recherche qui a dirigé l'expérimentation et effectuées le plus souvent en individuel (appelées par la suite « évaluations CNRS »). Les élèves de l'échantillon national n'ont été soumis qu'aux évaluations de la DEPP (*table 2*). Il est à noter que l'attrition (c'est-à-dire la perte d'élèves, de classes ou d'écoles) a tendance à être surestimée si l'on considère les effectifs d'élèves, car parmi les 2 398 élèves de CP des écoles engagées, les classes n'étaient pas toutes concernées par le dispositif (seuls les enseignants volontaires ont participé à l'étude). Il est également à noter que les évaluations DEPP ont perdu 10 écoles tests et 8 écoles témoins et les évaluations CNRS 4 écoles témoins.

Procédure générale (4 phases)

Préparation du projet, constitution de l'échantillon et formation des participants (phase 1)

Le projet de cette étude a été proposé à la DGESCO en février et accepté en mai 2010. L'expérimentation a débuté en mai 2010 par une journée de formation destinée à environ 80 cadres (inspecteurs, conseillers pédagogiques, etc.) de l'Éducation nationale de l'académie de Lyon. Les chercheurs ont présenté un état des lieux des recherches sur l'apprentissage de la lecture et le projet d'expérimentation (objectifs et contraintes). Sur cette base, un échantillon de 89 classes qui appartenaient pour la plupart aux réseaux d'éducation prioritaires (situés dans des zones de classe

¹ Hors DOM-TOM, écoles hors contrat, écoles à l'étranger, écoles spécialisées, écoles avec moins de 10 élèves de CP, écoles déjà présentes dans l'échantillon de Lyon.

Table 1. Caractéristiques des échantillons « Lyon » et « National ».

	Académie de Lyon						National	
	Classes « Test »		Classes « Témoin »		Total		Total	
Strate	Écoles	Élèves	Écoles	Élèves	Écoles	Élèves	Écoles	Élèves
Public hors éducation prioritaire	12	476	10	367	22	843	21	598
RAR (réseaux ambition réussite)	10	374	11	526	21	900	16	599
RRS (réseaux de réussite scolaire)	6	296	7	359	13	655	15	609
Privé	0	0	0	0	0	0	22	569
<i>Ensemble</i>	28	1 146	28	1 252	56	2 398	74	2 375

Table 2. Participation aux deux sessions des évaluations « Lyon » et « National ».

	Académie de Lyon						National		Ensemble	
	TEST		TÉMOIN		TOTAL		(hors privé)		(hors privé)	
	Écoles	Élèves	Écoles	Élèves	Écoles	Élèves	Écoles	Élèves	Écoles	Élèves
Échantillon	28	1 146	28	1 252	56	2 398	52	1 806	108	4 204
Évaluations DEPP										
Participation session 1 (pré-test)	24	890	24	869	48	1 759	51	1 682	99	3 441
Participation session 2 (post-test)	21	771	23	830	44	1 601	49	1 580	93	3 181
Participation sessions 1 et 2	20	703	22	729	42	1 432	49	1 526	91	2 958
Correspondance sessions 1 et 2	18	613	20	605	38	1 218	39	1 179	77	2 397
Évaluations Lyon										
Participation session 1 (pré-test)	28	891	24	691	52	1 582				
Participation session 2 (post-test)	28	841	25	711	53	1 552				
Participation sessions 1 et 2	28	824	24	638	52	1 462				
Correspondance DEPP-Lyon										
Session 1 (pré-test)	23	720	21	545	44	1 265				
Session 2 (post-test)	19	571	20	501	39	1 072				
Sessions 1 et 2	19	554	19	456	38	1 010				
Correspondances sessions 1 et 2	18	516	17	399	35	915				

moyenne ou défavorisée) a été constitué et, suite à un tirage aléatoire par strates, 44 classes ont été désignées comme classes tests, les 45 autres constituant les classes témoins. Début juillet, les enseignants et les conseillers pédagogiques des classes tests ont participé à trois journées de formation (soit 18 heures). Ensuite, durant l'année scolaire, ils ont reçu 24 heures supplémentaires de formation. De plus, un site internet leur était accessible à tout instant, et une conseillère pédagogique répondait à leurs questions. Ce site leur permettait d'échanger entre eux et de préparer des enseignements et des exercices adéquats. Pour des raisons éthiques, une formation similaire a été proposée aux enseignants en CP du groupe témoin après la phase de post-test en juin 2011.

Évaluations DEPP et CNRS : procédure et matériel (phases 2 et 4)

Évaluations DEPP

Chaque enseignant disposait de trois types de document donnés par la DEPP : un guide contenant les consignes de passation, un livret comprenant les planches pour la passation de l'épreuve d'écriture, et pour chaque élève un cahier contenant 8 séquences distinctes (S1 : Reconnaissance des mots, S2 : Lexique, S3 : Traitements sémantiques,

S4 : Écriture, S5 : Phonologie, S6 : Compréhension orale, S7 : Traitements complexes, S8 : Mathématiques). Ces séquences étaient prévues pour durer environ 20 minutes chacune. Il était recommandé d'étaler les passations des épreuves sur 8 demi-journées. À chaque session, la prise d'information (passation des épreuves, envoi des cahiers à la DEPP) était organisée sur une période de 15 jours. La passation des épreuves devait s'effectuer pendant les heures habituelles de classe des élèves. La correction était centralisée et prise en charge par la DEPP selon des procédures standardisées.

Évaluations CNRS

Un groupe de 20 expérimentateurs (psychologues ou étudiants en master 2 de psychologie), différents entre les deux sessions, a fait passer les évaluations individuelles aux élèves de CP des écoles composant l'échantillon de Lyon. Ces expérimentateurs ont reçu une journée de formation avant chaque session. Les épreuves ou tests utilisés étaient partiellement similaires entre la session pré-test et la session post-test. Le décodage et la saisie des résultats de chaque enfant ont été effectués par les expérimentateurs. Ces épreuves sont en parties disponibles (<http://www.unige.ch/fapse/sensori-moteur/ap/apsco>

/aplec/aplectexp.html).

– Pré-tests

Compétences associées à la lecture : 4 tests issus d'EVA-LEC (Sprenger-Charolles, Colé, Béchenec & Kipffer-Piquard, 2005 ; Sprenger-Charolles, Colé, Kipffer-Piquard & Leloup, 2010) ont été utilisés : 3 tests d'analyse phonologique et un test de dénomination sérielle rapide. Les tests d'analyse phonologique impliquaient la suppression du premier élément d'un mot inventé : soit la syllabe, soit le phonème. Le test syllabique comportait 10 trisyllabiques consonne-voyelle (CV) et les tests phonémiques 24 items de 3 phonèmes (12 CVC et 12 CCV). Ces tests ont été administrés avec un ordinateur portable (à noter : les résultats de la session 1 sont à prendre avec précaution, certains des signaux présentés n'étant pas bien audibles en raison d'une mise à jour défectueuse de la carte son des ordinateurs, ce problème, solutionné partiellement pour cette session, l'a été complètement pour la seconde). La rapidité et la précision de la dénomination ont été évaluées par un test de dénomination sérielle rapide : les enfants devaient dénommer le plus rapidement et le plus précisément possible 6 couleurs (rouge, jaune, vert, bleu, blanc, gris) présentées 8 fois en ordre différent.

Compréhension orale (test ECOSSE, Lecocq, 1996) : les élèves devaient désigner, après présentation orale d'une phrase, l'image adéquate parmi 4. Ce test comportait 24 items de 6 structures : phrases actives et passives (1-2) et phrases incluant des pronoms variant en nombre (3), une double négation (4), des termes spatiaux (5) ou une proposition relative (6).

Compréhension lexicale (test TVAP, Deltour & Hupkens, 1980) : après la présentation orale d'un mot, les élèves devaient désigner une image parmi 6, avec des images distractrices contrôlées. Pour une réponse exacte, 2 points étaient attribués et 1 point seulement lorsque l'image choisie était celle proche de celle attendue. 30 items étant présentés, le score maximum était donc de 60.

Pré-lecture (à haute voix) de mots familiers et inventés : les élèves étaient invités à lire à haute voix des mots familiers en 1 minute (35 mots, 5 par ligne). Les mots choisis étaient fréquents (issus des 1 000 premiers mots des manuels de CP, cf. Lété, Sprenger-Charolles & Colé, 2004) et, pour la majorité d'entre eux, faciles à lire : en l'occurrence, courts, réguliers sur le plan des correspondances graphèmes-phonèmes, et ne comportant que peu de groupes consonantiques. De même, les élèves devaient lire des mots inventés en 1 minute (30,5 par ligne). Les items choisis étaient appariés aux mots fréquents en longueur, en structure syllabique et en difficulté orthographique.

– Post-tests

Après les entraînements, une partie des épreuves utilisées en pré-test a été utilisée de nouveau : les épreuves d'analyse phonologique, de dénomination rapide, de vocabulaire, de lecture à haute voix de mots familiers et inventés (avec 60 mots fréquents et 60 mots inventés au lieu d'une trentaine pour la session précédente), ainsi qu'une partie (12 sur les 24) des phrases de l'épreuve de compréhension

orale (ces 12 phrases étant réparties en 6 blocs, les mêmes que pour la session précédente). Les enfants ont également passé de nouvelles épreuves de lecture et compréhension écrite.

Lecture : test de l'Alouette (Lefavrais, 1967) : les enfants devaient lire un texte de 265 mots aussi rapidement et précisément que possible. On relève soit le temps de lecture, s'il est inférieur à 3 minutes, soit le nombre de mots lus en 3 minutes. Le niveau de lecture est calculé en tenant compte de la rapidité et de la précision (nombre d'erreurs).

Compréhension écrite (test ECOSSE) : les élèves devaient désigner, après la lecture à haute voix d'une phrase, l'image adéquate (parmi 4). Ce test incluait 12 phrases réparties en 6 blocs de même structure que celle utilisée pour le test de compréhension orale, mais avec quelques modifications principalement dans le choix des mots.

Compréhension de texte (Les deux poules) : les élèves devaient lire à voix haute une histoire courte puis répondre à 8 questions posées par l'expérimentateur sur ce qu'ils avaient lu, sans avoir la possibilité de relire le texte.

Les entraînements (phase 3)

Après la phase 2, les classes « test » ont reçu, semaine après semaine (de novembre 2010 à avril 2011, phase 3), un enseignement conçu pour développer deux types de compétences : le décodage et la compréhension de l'écrit. Plus précisément, 4 entraînements ont été utilisés : un sur la phonologie (Jacquier-Roux & Zorman, 1998), un sur le décodage (<http://www.unige.ch/fapse/sensori-moteur/ap/apsco/aplec/aplectexp.html>), un sur la compréhension (Bianco, Coda, Gourgue-Giollito & Robert, 2005) et un sur la fluence de lecture (Pourchet & Zorman, 2008). À partir de février, l'entraînement centré sur la fluence a été proposé aux enfants devenus de bons décodeurs alors que, pour ceux qui présentaient toujours des difficultés pour réaliser les tâches d'analyse phonémique, l'entraînement phonologique a été poursuivi. Fait important, ces différents entraînements venaient en complément de l'utilisation du manuel de lecture, dont le choix incombait à chaque école. Il est à noter que le groupe de recherche a pu proposer une progression pédagogique rationnelle de l'ordre d'apprentissage des correspondances graphèmes-phonèmes (cf. Dehaene *et al.*, 2011) mais qui n'a pas pu être utilisée par les enseignants, ces derniers suivant l'ordre proposé par le manuel.

Le temps imparti chaque semaine aux différents entraînements ne devait pas dépasser celui dédié à l'étude de la langue dans les programmes officiels (10-12 heures), afin de ne pas défavoriser les autres disciplines. Pour le vérifier, les enseignants devaient compléter chaque semaine un emploi du temps des contenus abordés. Toutefois, l'analyse des informations recueillies n'a pas permis d'obtenir de résultats exploitables.

Pour les entraînements centrés sur le décodage et la phonologie, les enfants de chaque classe ont été répartis en 4 sous-groupes de besoin, échelonnés sur la base de leurs résultats aux évaluations CNRS du début de l'année. Par la

suite, deux fois par semaine, pendant 1 h 30, ces sous-groupes étaient pris en charge par leur enseignant et un autre adulte disponible au sein ou autour de l'école. Chaque encadrant prenait en charge la moitié de la classe (10 enfants) et proposait, pendant les premières 30 minutes, des séances dirigées à 5 élèves, les 5 autres enfants travaillant en autonomie. Pendant la seconde période, les sous-groupes étaient inversés. Durant la troisième période, les enfants des deux sous-groupes qui éprouvaient encore des difficultés en décodage et en phonologie étaient pris en charge, tandis que les autres travaillaient en autonomie. Pour le travail sur la compréhension, les enfants de chaque classe ont été répartis en deux sous-groupes de besoin et ont bénéficié, deux fois par semaine, d'une séance d'1 h, comprenant 30 minutes de séances de travail dirigé et 30 minutes de travail autonome. En tout, 5 heures ont été dévolues à cet enseignement spécifique. Pour arriver aux 10-12 heures hebdomadaires d'enseignements du français prévues par le programme, l'enseignant complétait ses leçons en approfondissant certains points ou/et en abordant d'autres points (un travail sur le manuel, sur des albums, etc.). Les séances de travail dirigé étaient très structurées. Elles se déroulaient généralement dans un espace calme pour optimiser l'attention des enfants qui étaient assis en petits groupes autour d'une table pour favoriser la discussion. Les exercices proposés sollicitaient la participation active de chacun d'eux. Les séances de travail autonome étaient préparées comme de véritables séances de travail et posaient rarement des difficultés aux enfants habitués à travailler de cette manière en maternelle.

RÉSULTATS

Il est à noter que la majorité des classes appartenant à des réseaux d'éducation prioritaire (REP), la plupart des enfants étaient issus de milieux socio-économiques et culturels en grande difficulté. Ce fait est reflété par le niveau moyen de lecture évalué avec le test de l'Alouette (Lefavrais, 1967) passé par les enfants de l'académie de Lyon à la fin de l'année scolaire, test qui montre que 30 % de cet échantillon était significativement au-dessous (à plus d'un écart-type) du niveau moyen (contre 15 % attendus normalement).

Pour estimer les effets moyens de l'expérimentation, une modélisation par régression linéaire a été effectuée à l'aide du logiciel SAS. Le modèle de régression porte sur chacun des scores observés au post-test en fonction de ceux observés au pré-test, ainsi que des 3 strates (hors éducation prioritaire, ZEP, RAR). Comme l'équilibre entre les échantillons initialement produits par le tirage au sort a pu être altéré par la non-réponse, la prise en compte des pré-tests permet de contrôler les éventuels biais sur la composition des élèves (mais pas ceux liés aux caractéristiques des enseignants). Elle permet également de rendre l'estimation plus précise. Par ailleurs, l'estimation tient compte de la structure en grappes de l'échantillon (niveau école). Cette analyse a été conduite pour chacune des évaluations (évaluation CNRS, évaluation DEPP et évaluation sur les enfants qui ont passé ces deux évaluations). Les scores ont été standardisés (centrés-réduits) pour effectuer les comparaisons statistiques entre les performances des classes tests

et celles des classes témoins. Une attrition est observée entre les pré-tests et les post-tests. Cependant, les analyses montrent qu'elle n'a pas eu d'impact sur la structure des échantillons et n'a pas introduit de différences significatives entre les groupes tests et témoins.

Évaluations CNRS

Les caractéristiques de l'échantillon de l'évaluation CNRS et les scores bruts obtenus à chaque épreuve aux pré et post-tests par les élèves des deux groupes sont présentés dans la *table 3*. En début d'année, les élèves du groupe test ont obtenu des scores inférieurs à ceux du groupe témoin aux épreuves de lecture de mots inventés ($p=.05$) et familiers ($p=.05$). En fin d'année, les différences entre les deux groupes ne sont pas significatives (cf. *table 4*, tous les $p>.05$) et ce, quel que soit le type d'analyse (avec ou sans contrôle du niveau initial des élèves et des strates).

Évaluations DEPP

En début d'année, les différences entre le groupe test et le groupe témoin ne sont pas significatives. Par contre, en fin d'année, contrairement aux attentes, les élèves du groupe test ont tendance à moins bien réussir les épreuves de lexique, d'écriture et de mathématiques que ceux du groupe témoin ($p<.10$). En contrôlant le niveau initial des élèves et la structure par strates, ce n'est que pour les épreuves de lexique et de mathématiques que les scores du groupe test se révèlent inférieurs à ceux du groupe témoin (respectivement $p = .03$ et $.01$). Les résultats des élèves de l'échantillon national ne sont pas significativement différents de ceux des élèves du groupe témoin.

Évaluations DEPP et CNRS

En début d'année, il n'y a pas de différence de résultats entre les élèves des classes tests et ceux des classes témoins, que ce soit aux évaluations de la DEPP ou aux évaluations du CNRS. En fin d'année, il apparaît un certain nombre de différences en défaveur des élèves du groupe test qui, sans contrôle du niveau initial des élèves et des strates, ont des résultats inférieurs à ceux du groupe témoin aux épreuves de lexique, d'écriture et de mathématiques des évaluations de la DEPP, ainsi qu'à toutes les épreuves de lecture des évaluations CNRS. Avec le contrôle du niveau initial des élèves et des strates, ces différences ne demeurent significatives que pour les épreuves de lexique et de mathématiques des évaluations de la DEPP et pour deux des épreuves de lecture (Alouette et ECOSSE écrit) des évaluations CNRS.

Ces résultats sont néanmoins à tempérer au regard de l'attrition importante qui affecte l'échantillon ayant participé aux évaluations DEPP et CNRS. Cette attrition semble toucher de manière différente les écoles tests et témoins, en termes de résultats aux évaluations nationales de CE1 : les écoles tests non testées apparaissent plus performantes que les écoles témoins non testées. En revanche, cette attrition différentielle est moins nette si l'on observe les différences de résultats des écoles perdues au démarrage de l'expéri-

Table 3. Évaluations CNRS : caractéristiques de l'échantillon et moyenne (et écart type) des groupes tests et témoins aux épreuves utilisées en pré-tests et post-tests ou uniquement en post-tests.

	Classes tests		Classes témoins	
Effectif total « Lyon »	824		638	
Département : Ain / Loire / Rhône	132/177/515		177/130/331	
Age (mois)	75,9 (4,3)		75,4 (4,3)	
Genre : Garçons / Filles	410/414		306/332	
Situation scolaire : À l'heure / En retard / En avance	777/45/2		604/30/4	
Contexte école				
RRS (réseaux de réussite scolaire)	392		334	
RAR (réseaux ambition réussite)	171		156	
CUCS (contrats urbains de cohésion sociale)*	43		0	
Milieu défavorisé	93		105	
Classes ordinaires	125		43	
Année d'expérience de maternelle : 0 / 1 / 2 / 3	4/46/73/700		5/33/50/550	
Épreuves Lyon Pré et post-test	Groupe test : Moyenne (Écart-type)		Groupe témoin : Moyenne (Écart-type)	
	Pré-test	Post-test	Pré-test	Post-test
Vocabulaire (Score total /60)	35,72 (7,16)	38,95 (6,99)	36,10 (7,38)	38,97 (7,58)
Dénomination sérielle rapide (Temps en secondes)	71,57 (28,46)	59,21 (18,62)	69,89 (32,07)	59,52 (24,56)
Phonologie (analyse syllabique et phonémique) (% de réponses correctes)	35,3 (23,4)	66,5 (33,9)	35,7 (25,5)	67,5 (32,6)
Lecture Mots familiers (Mots lus par minute)	6,33 (7,63)	34,94 (16,88)	7,77 (9,65)	37,47 (17,73)
Lecture Mots inventés (Mots lus par minute)	5,88 (6,03)	25,37 (10,91)	6,99 (7,93)	26,75 (11,78)
Lecture (Alouette) (Mots lus correctement en 3 minutes)		68,81 (37,95)		77,47 (42,66)
Lecture de mots en contexte (Poule) (Mots lus correctement en 1 minute)		39,74 (24,17)		43,87 (26,35)
Compréhension écrite de texte (Poule) (% de réponses correctes)		47,40 (25,61)		47,71 (26,39)
Compréhension écrite d'énoncés (Ecosse) (% de réponses correctes)		66,14 (20,93)		67,67 (19,72)
Compréhension orale d'énoncés (Ecosse) (% de réponses correctes)	78,48 (15,76)	82,80 (16,05)	76,42 (17,12)	82,50 (15,13)

* Contrat urbains de cohésion sociale (CUCS) : contrat passé entre l'État et les collectivités territoriales qui engage chacun des partenaires à mettre en œuvre des actions concertées pour améliorer la vie quotidienne des habitants dans les quartiers connaissant des difficultés (chômage, violence...).

Table 4. Évaluations CNRS : effets des entraînements sur les mesures en post-test avec ou sans contrôle du niveau initial des élèves et de la structure des strates. Une différence positive signifie que les classes tests ont de meilleures performances que les classes témoins.

« LYON » - Session 2	Sans contrôle		Avec contrôle	
	Différence	Probt	Différence	Probt
Vocabulaire	0.00	0.98	-0.02	0.75
Dénomination sérielle rapide	-0.01	0.84	-0.07	0.30
Lecture Mots familiers	-0.14	0.25	-0.05	0.58
Lecture Mots inventés	-0.12	0.33	-0.04	0.67
Lecture de texte (Alouette)	-0.20	0.07	-0.10	0.29
Lecture de texte (Poule)	-0.16	0.20	-0.05	0.60
Compréhension écrite (Poule)	-0.01	0.91	-0.01	0.89
Compréhension écrite (Ecosse)	-0.08	0.48	-0.06	0.45
Compréhension orale (Ecosse)	0.02	0.84	-0.05	0.30

mentation (traitement des résultats de la première session des évaluations DEPP et Lyon). Les effets négatifs observés doivent donc être relativisés en raison du phénomène d'attrition.

DISCUSSION

Les analyses quantitatives des résultats aux évaluations conduites par le CNRS et la DEPP montrent de manière similaire que le programme n'a pas permis de faire davantage progresser en lecture les enfants du groupe test (Lyon) que ceux des groupes témoins (Lyon et National). En revanche, les évaluations subjectives réalisées par les différents IEN auprès des enseignants impliqués dans le programme et exposées trimestriellement au cours des réunions du comité de pilotage réunissant tous les acteurs du projet (Recteur, IA, etc.) étaient très positives à ce à différents niveaux : comportements individuels et collectifs des enfants, ambiance de classe, travail de d'équipe... (cf. bilan des comptes rendus subjectifs disponible à l'adresse suivante : <http://www.unige.ch/fapse/sensori-moteur/ap/apsco/aplec/aplectexp.html>).

Il est à souligner que cette recherche est originale sur plusieurs aspects parmi lesquels son ampleur : en effet, il n'existe pas à notre connaissance de publications scientifiques qui relatent une expérimentation impliquant autant d'enfants et d'enseignants avec une telle méthodologie scientifique. Elle est également originale par son souci de contrôler d'éventuels biais expérimentaux comme 1) la désignation aléatoire des écoles tests ou témoins, 2) l'évaluation du niveau en mathématiques (pour mesurer si un enseignement massif de français ne s'est pas fait aux dépens des autres matières comme les mathématiques), 3) la prise en compte de tous les enfants, 4) l'évaluation conjointe et complémentaire de la DEPP et du CNRS et, enfin, 5) la quasi-séparation des acteurs impliqués dans les deux évaluations, les traitements et les analyses et ceux impliqués dans les entraînements (exception faite pour le responsable de l'étude, EG).

Cette recherche présente toutefois plusieurs limites. La première est sa durée dans sa mise en place et sa conduite. En effet, nous n'avons disposé que de peu de temps pour préparer et présenter le projet, puis pour recruter et former les enseignants volontaires avant le début de l'année de l'expérimentation alors que plusieurs semaines auraient été nécessaires ne serait-ce que pour former les enseignants aux entraînements et aux nouvelles organisations de classe à mettre en œuvre à la rentrée. Cette contrainte temporelle a aussi eu pour conséquence de limiter fortement la possibilité de proposer une intégration des contenus des quatre entraînements et de les coordonner avec les autres outils utilisés par les enseignants (manuels de lecture, fichiers, etc.). Elle a également empêché, comme noté dans le rapport de l'Inspection générale (2012), de corriger certaines des imperfections des outils utilisés pour cette expérimentation, outils qui avaient été développés avant elle (par exemple, celui dédié à l'entraînement phonologique). De plus, et c'est la seconde limite de la présente étude, il n'a été possible d'intervenir que sur la première année du pri-

naire (pour des raisons financières et d'organisation). Or, les études montrent l'importance de la préparation à la lecture dès la maternelle et la nécessité d'avoir des entraînements se déroulant sur 2 ou 3 années pour observer des effets bénéfiques significatifs (cf. Bianco *et al.*, 2012). La troisième limite est liée aux conséquences de l'ampleur de l'étude sur le dispositif de formation et de suivi des enseignants : du fait de leur grand nombre et de leur dispersion géographique sur trois départements (contraintes imposées par le rectorat), nous n'avons pu proposer une formation approfondie régulière et un suivi personnalisé pour chacun des enseignants. Au total, chaque enseignant n'a pu recevoir qu'environ 30 heures de formation réparties sur l'année et ce seulement en dehors de la classe, et non in situ. Une quatrième limite pourrait être due à la persistance de l'utilisation du manuel en classe. On ne peut en effet pas exclure qu'un manque de synergie entre les entraînements spécifiques et le manuel de lecture puisse expliquer, en partie, les moins bonnes performances des enfants du groupe test.

En conclusion, cette recherche illustre parfaitement le débat ancien sur l'utilisation des résultats de la recherche (Gentaz & Dessus, 2004). En effet, comme l'a reconnu le célèbre psychologue William James il y a plus d'un siècle, les descriptions psychologiques ne peuvent se transposer directement en prescriptions éducatives spécifiques (Mayer, 1992) : la raison principale en est qu'on ne peut déduire directement de résultats de recherches, spécifiques à une situation, des procédures détaillées, applicables à toutes situations, pour tous publics, même s'il existe une perméabilité entre pédagogie et science, souvent plus importante qu'on l'imagine. C'est ce que signale cette étude, qui montre que, si les résultats sur l'apprentissage de la lecture issus des sciences cognitives sont solides (Dehaene *et al.*, 2011 ; Sprenger-Charolles & Colé, 2013), et si les principes pédagogiques qui en découlent sont aujourd'hui bien connus (Dehaene *et al.*, 2011 ; Ehri *et al.*, 2001), leur mise en application dans les classes demande encore un effort très important de recherche appliquée devant mobiliser conjointement les acteurs de monde de la recherche et ceux du corps éducatif. Nous espérons que la présentation des apports, originalités et limites de la présente expérimentation sera utile aux futurs chercheurs intéressés par ce type d'approche, affreusement compliqué mais absolument indispensable pour faire progresser notre système éducatif.

Remerciements

Cette expérimentation a bénéficié du soutien financier du MEN (95 %) et de l'association Agir pour l'école, du soutien en personnels et en logistique du CNRS et des universités (en particulier de l'UPMF). Elle a mobilisé les inspecteurs d'académie des départements du Rhône, de la Loire et de l'Ain, les inspecteurs de l'Éducation nationale et les conseillers pédagogiques de circonscription des écoles concernées. Nous avons bénéficié des recommandations scientifiques de M. Bosse, M. Bianco, P. Bressoux, M. Fayol, B. New, R. Peereman, du soutien constant de B. Suchaut, S. Dehaene, A. Hillairet de-Boisferon et L. Cros (haut fonctionnaire, actuellement délégué général de l'associa-

tion Agir pour l'école) et d'un important travail d'É. Pétris pour l'édition d'une partie des nouveaux outils pédagogiques élaborés à cette occasion et disponible à l'adresse suivante : <http://www.unige.ch/fapse/sensori-moteur/ap/apsco/aplec/aplectexp.html>. La formation et l'accompagnement des enseignants durant l'année scolaire 2010-2011 ont été assurés par F. Mirgalet, conseillère pédagogique de l'inspection académique de l'Isère et mise à disposition pendant l'année scolaire, et par une équipe de trois conseillers pédagogiques des inspections académiques du Rhône de l'Ain et de la Loire. Que tous soient ici vivement remerciés.

RÉFÉRENCES

- BARA, F., COLÉ, P. & GENTAZ, E. (2008). Littéracie précoce et apprentissage de la lecture : comparaison entre des enfants à risque, scolarisés en France dans des réseaux d'éducation prioritaire et des enfants de classes régulières. *Revue des Sciences de l'éducation*, 34, pp. 27-45.
- BARA, F., GENTAZ, É. & COLÉ, P. (2004). Les effets des entraînements phonologiques et multisensoriels destinés à favoriser l'apprentissage de la lecture chez les jeunes enfants. *Enfance*, 2-3, pp. 145-167.
- 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, pp. 211-246.
- BIANCO, M., CODA, M., GOURGUE-GIOLLITO, A. & ROBERT, A. (2005). *Compréhension*. Grenoble : La Cigale.
- BIANCO, M., PELLENQ, C., LAMBERT, E., LIMA, L., & DOYEN, A.L. (2012). Impact of early code-skills and oral-comprehension training on reading achievement in first grade. *Journal of Research in Reading*, 35(4), pp. 427-455.
- BILLARD, C., BRICOUT, L., DUCOT, B., RICHARD, G., ZIEGLER, J. & FLUSS, J. (2010). Evolution of competence in reading, spelling and comprehension levels in low socioeconomic environments and impact of cognitive and behavioral factors on outcome in two year. *Revue d'Epidémiologie et de santé publique*, 58, pp. 101-110.
- BLACHMAN, B.A., BALL, E.W., BLACK, R.S. & TANGEL, D.M. (1994). Kindergarten teachers develop phonemic awareness in low-income inner-city classrooms : Does it makes a difference? *Reading and Writing: An Interdisciplinary Journal*, 6, pp. 1-18.
- BRADLEY, L. & BRYANT, P. (1983). Categorizing sounds in learning to read: a causal connection. *Nature*, 301, pp. 419-421.
- BRESSOUX, P. & ZORMAN, M. (2009). *Présentation et évaluation du programme de prévention « Parler »*. ec.europa.eu/education/migration/france2_fr.pdf
- CASTLES, A. & COLTHEART, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, 91, pp. 77-111.
- DEHEANE, S. (Dir), DEHAENE-LAMBERT, G., GENTAZ, É., HURON, C. & SPRENGER-CHAROLLES, L. (2011). *Apprendre à lire*. Paris : Odile Jacob.
- DELTOUR, J. & HUPKENS, D. (1980) *Test de vocabulaire actif et passif pour enfants de 5 à 8 ans (TVAP 5-8)*. Braine-le-Château : éditions de l'Application des techniques modernes.
- EHRI, L.C., NUNES, S.R., WILLOWS, D.M., SCHUSTER, D.M., YAGHOUB-ZADEH, Z. & SHANAHAN, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National reading Panel's meta-analysis. *Reading Research Quarterly*, 36, pp. 250-287.
- GENTAZ, É. & DESSUS, P. (2004). *Comprendre les apprentissages*. Paris : Dunod.
- GENTAZ, É. & SPRENGER-CHAROLLES, L. (2012). *Pourquoi faut-il évaluer et entraîner certaines capacités reliées à la lecture des enfants scolarisés en grande section de maternelle ?* Site du SNUIPP.
- GENTAZ, É., SPRENGER-CHAROLLES, L., THEUREL, A. & COLÉ, P. (in press). Reading comprehension in a large cohort of French first graders from low socio-economic status families: a 7-month longitudinal study. *PLoS One*.
- Inspection générale de l'Éducation nationale (2012). *Évaluation de la mise en œuvre, du fonctionnement et des résultats des dispositifs « P.A.R.L.E.R. » et « R.O.L.L. » : Rapport à monsieur le ministre de l'Éducation nationale*. Rapport 2012-129 (Nov. 2012).
- JACQUIER-ROUX, A. & ZORMAN, M. (1998). *Phonologie*. Grenoble : La Cigale.
- LECOCQ, P. (1996). *L'E.Co.S.Se: une épreuve de compréhension syntactico-sémantique*. Lille : Presses du Septentrion
- LEFAVRAIS, P. (1967). *Test de l'Alouette : manuel*. Paris : ECPA.
- LÉTÉ, B., SPRENGER-CHAROLLES, L. & COLÉ, P. (2004). Manulex: A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, 36, pp. 156-166.
- MAYER, R. E. (1992). Cognition and instruction : Their historic meeting within educational psychology. *Journal of Educational Psychology*, 84, pp. 405-412.
- POURCHET, M. & ZORMAN, M. (2008). *Fluence*. Grenoble : La Cigale.
- SPRENGER-CHAROLLES, L. & COLÉ, P. (2013). *Lecture et dyslexie*. Paris : Dunod.
- SPRENGER-CHAROLLES, L., COLÉ, P., BECHENNEC, D. & KIPFFER-PIQUARD, A. (2005). EVALEC : Batterie d'évaluation diagnostique de la dyslexie. *European Review of Applied Psychology*, 55, pp. 157-186.
- SPRENGER-CHAROLLES, L., COLÉ, P., KIPFFER-PIQUARD, A. & LELOUP, G. (2010). *EVALEC : Batterie informatisée d'évaluation diagnostique des troubles spécifiques d'apprentissage de la lecture*. Ortho-Édition.