



Les pratiques pédagogiques efficaces

Conclusions de recherches récentes

Pierre-Yves Cusset

Document de travail n°2014-01

Sommaire

Résumé

Introduction

1. Évaluation de l'efficacité pédagogique : principes et limites méthodologiques

- 1.1. Que mesure-t-on et comment le mesure-t-on ?
- 1.2. Limites méthodologiques et première discussion

2. Les études anglo-saxonnes

- 2.1. L'apprentissage de la lecture
- 2.2. L'apprentissage des mathématiques
- 2.3. L'apprentissage des sciences
- 2.4. Le projet *Follow Through*

3. Les études en contexte francophone

- 3.1. Efficacité comparée des méthodes d'apprentissage de la lecture en Belgique
- 3.2. L'impact du manuel utilisé : une étude en zone d'éducation prioritaire
- 3.3. Les limites de l'approche idéovisuelle dans l'apprentissage de la lecture
- 3.4. L'entraînement précoce à la compréhension
- 3.5. Une expérience contrôlée de soutien scolaire pour l'apprentissage de la lecture
- 3.6. Évaluation d'un programme expérimental d'enseignement structuré de la lecture

Conclusion

Résumé

Les études cherchant à évaluer l'efficacité des pratiques pédagogiques, plus nombreuses dans le contexte anglo-saxon qu'en France, se développent tant en quantité qu'en qualité. Ce document de travail présente d'abord la méthodologie et les limites de ces travaux, avant de synthétiser leurs résultats. Pour l'apprentissage de la lecture, ces recherches mettent en évidence l'intérêt, pour le déchiffrage, d'un enseignement systématique de la relation phonème-graphème mais démontrent aussi qu'il est possible de faire progresser les enfants en compréhension de texte en recourant à une approche structurée et explicite. Les approches de type apprentissage coopératif entre pairs sont également bénéfiques pour la lecture, mais uniquement dans la dimension compréhension (pas déchiffrage). Pour l'apprentissage des mathématiques, l'enseignement explicite et l'apprentissage coopératif sont aussi des stratégies d'enseignement particulièrement efficaces. Pour l'enseignement des sciences, les études disponibles sont moins nombreuses et leurs conclusions encore assez contradictoires. Des méthodes d'apprentissage par investigation (*inquiry-based learning*) présentent un réel intérêt mais supposent que l'enseignant parvienne à guider efficacement les élèves. Les effets comparés des différentes approches sur le niveau des élèves restent cependant relativement modestes au regard d'autres facteurs déterminants qui concernent l'élève lui-même : niveau initial, environnement, capacités.

Mots-clefs : pédagogie ; efficacité pédagogique ; évaluation ; école.

Les pratiques pédagogiques efficaces

Conclusions de recherches récentes

Pierre-Yves Cusset¹

Introduction

L'évaluation de l'efficacité des méthodes pédagogiques fait l'objet d'une littérature assez peu abondante en France, et plus largement dans les pays francophones, mais est nettement plus développée dans les pays anglo-saxons et tout particulièrement aux États-Unis.

Le contexte américain est marqué en effet par une prise de conscience assez précoce des difficultés importantes que rencontrent de nombreux élèves dans les apprentissages de base, difficultés touchant spécifiquement les enfants issus des minorités noires, hispaniques et amérindiennes (rapport Coleman en 1966 puis rapport *A Nation at Risk* en 1983)². Cette prise de conscience a abouti au financement par le ministère de l'Éducation américain d'importants programmes de recherche sur l'efficacité des méthodes pédagogiques, pour les élèves qui rencontrent des difficultés particulières comme pour l'ensemble des élèves, puis à la réalisation de méta-analyses de grande ampleur synthétisant les résultats de ces travaux. Au niveau international, John Hattie a ainsi recensé, en 2008, 800 méta-analyses couvrant en tout environ 50 000 études consacrées aux facteurs qui favorisent l'efficacité de l'enseignement (Hattie, 2008). Mais il faut garder à l'esprit que ces études tout comme les méta-analyses censées les synthétiser sont de qualité très inégale.

Au fil des années, l'exigence méthodologique des études est devenue de plus en plus grande. Aux batailles de chapelles pédagogiques et aux affrontements théoriques entre « pédagogies traditionnelles » et « pédagogies progressistes » se sont substituées des évaluations empiriques bâties sur le modèle de ce qui était pratiqué depuis déjà un certain temps dans le domaine médical : constitution d'un groupe expérimental et d'un groupe de contrôle ; affectation aléatoire des élèves ou des classes aux deux groupes (étude

¹ Pierre-Yves Cusset, chargé de mission, département Questions sociales, France Stratégie (pierre-yves.cusset@strategie.gouv.fr).

² En 2007, les écarts restaient très importants. Lors de l'évaluation nationale des élèves américains scolarisés en 4^e année d'école primaire (équivalent du CM1), 43 % des enfants blancs atteignaient le niveau de maîtrise le plus élevé, contre 17 % des élèves hispaniques, 14 % des élèves afro-américains et seulement 8 % des élèves indiens (« American Indian children »).

« randomisée ») ; évaluation de la progression des élèves sur la base de tests standardisés auxquels les élèves sont soumis avant et après le traitement ; estimation de l'ampleur de l'effet du traitement en termes de capacité à faire progresser les élèves dans la hiérarchie des résultats... Aux États-Unis, le ministère de l'Éducation a créé en 2002 une structure, la *What Works Clearing House*³, spécifiquement chargée de répertorier, selon des standards précis⁴, les évaluations des interventions pédagogiques. Les programmes éducatifs et pratiques pédagogiques répertoriés font alors l'objet d'une notation synthétique, qui donne des informations tant sur l'ampleur moyenne de l'effet enregistré sur les résultats des élèves que sur la robustesse des études scientifiques mobilisées pour estimer cet effet.

Les données accumulées ne trancheront sans doute jamais totalement les controverses qui opposent les tenants des diverses approches pédagogiques. Elles permettent néanmoins de les éclairer.

³ <http://ies.ed.gov/ncee/wwc/>. « Clearing House » : chambre de compensation.

⁴ Voir *Procedures and standards Handbook*, actuellement dans sa version 3.0.

1. Évaluation de l'efficacité pédagogique : principes et limites méthodologiques

Pour évaluer l'efficacité des méthodes pédagogiques, les chercheurs ont recours à des méthodologies variées. Ce choix est souvent contraint par des considérations pratiques de coût, d'opportunité, de temps disponible ou d'acceptabilité de la part des parents et des enseignants. On présente ici les caractéristiques d'une évaluation « idéale » qui correspondrait aux critères de rigueur scientifique les plus exigeants, critères qui guident notamment le choix des études prises en compte dans les méta-analyses que l'on mobilisera. Certaines des études mentionnées dans ce document de travail pourront s'éloigner de ce modèle canonique. Même les évaluations les plus rigoureuses présentent un certain nombre de limites qui seront discutées.

1.1. Que mesure-t-on et comment le mesure-t-on ?

- **Une évaluation qui repose sur une mesure des progrès des élèves**

Le premier facteur d'explication du niveau de maîtrise atteint par un élève une année donnée reste, et de loin, le niveau qu'il avait atteint l'année précédente. Aussi, ce qui compte réellement, c'est la capacité d'une méthode pédagogique à faire progresser les élèves à partir de leur niveau initial. Une évaluation rigoureuse nécessite donc d'avoir pour chaque élève au moins deux prises d'information sur ce niveau : un « pré-test », réalisé en début d'année ou à tout le moins avant la mise en pratique de la méthode que l'on cherche à évaluer, et un « post-test », réalisé en fin d'année ou en tout cas après une certaine durée d'exposition de l'élève à cette méthode. Les évaluations les plus élaborées, qui reposent sur la constitution de panels, peuvent aller plus loin et chercher à estimer dans quelle mesure les progrès constatés sont durables ou bien au contraire tendent à s'estomper voire à disparaître avec le temps.

- **Une mesure relative du rythme de progression**

Le but de tout enseignant est bien d'amener chaque élève à atteindre le niveau de maîtrise le plus élevé possible de la matière qu'il enseigne. Mais lorsque l'on souhaite comparer l'efficacité de différentes méthodes pédagogiques, il est plus aisé de raisonner en termes relatifs. Aussi, la performance de chaque élève est rapportée au niveau de performance des autres élèves, que ce soit au moment du pré-test ou du post-test. Ainsi, on ne mesure pas simplement les progrès de l'élève, mais on cherche à estimer dans quelle mesure un élève exposé à une méthode pédagogique donnée progresse plus ou moins vite qu'un autre élève exposé à une autre méthode. À cette fin, le score de chaque élève est traduit en termes de position dans la hiérarchie des performances des élèves. Si un élève se situe au cinquantième centile de cette hiérarchie avant et après exposition à la méthode pédagogique dont on cherche à estimer l'efficacité, cela ne veut pas dire qu'il n'a pas progressé : cela veut dire qu'il a progressé au même rythme que les autres élèves et que, si ce résultat est vrai en moyenne pour tous les élèves exposés à la méthode, cette méthode pédagogique ne se distingue pas des autres méthodes. Au contraire, une méthode pédagogique qui permet à un élève de s'élever dans cette hiérarchie de performance est celle qui facilite une progression plus rapide. C'est donc une méthode plus efficace que les autres.

- **Le modèle « canonique » de l'évaluation expérimentale**

Dans l'idéal, un exercice d'évaluation d'une méthode pédagogique se déroule de la façon suivante. On définit un groupe expérimental (aussi appelé groupe de traitement) et un groupe témoin (ou groupe de contrôle). Les élèves du premier groupe seront exposés au « traitement » (ici, la méthode pédagogique dont on cherche à évaluer l'efficacité), tandis que les élèves du second groupe seront exposés à la méthode pédagogique « habituelle ». Les élèves (ou les classes) qui participent à l'évaluation sont alors affectés de façon aléatoire au groupe expérimental et au groupe témoin (on parle d'étude randomisée). Ils sont testés une première fois avant le début de l'expérimentation (pré-test), mais, en principe, après leur affectation à l'un des deux groupes. À la fin de l'expérimentation, ils sont à nouveau testés (post-test). Ils peuvent éventuellement être testés à nouveau plusieurs mois voire plusieurs années après la fin de l'expérimentation, si l'on s'intéresse au caractère durable des effets observés.

Pour que soit garantie la qualité de la démarche d'évaluation, plusieurs conditions doivent être remplies. Tout d'abord, on doit s'assurer que les résultats des deux groupes au pré-test sont suffisamment proches. C'est généralement le cas du fait de l'affectation aléatoire des élèves. Ensuite, dans l'idéal, ce ne sont pas les enseignants qui font passer les tests et les corrigent, même si ce qui est jugé n'est pas leur qualité d'enseignant mais l'efficacité de la méthode qu'ils mettent en application. Enfin et surtout, l'instrument de mesure doit être indépendant du traitement. Cela veut dire par exemple que si une méthode met l'accent sur la conscience phonologique pour faciliter l'apprentissage de la lecture, ce qui est mesuré *in fine*, c'est bien la maîtrise de la lecture et non la seule conscience phonologique.

Pour estimer la taille de l'effet, du fait de l'affectation aléatoire des élèves aux deux groupes, on pourrait simplement comparer les résultats moyens des deux groupes au post-test. Dans la pratique, on conserve assez souvent des variables de contrôle sur les élèves, en commençant par leurs résultats au pré-test, mais en incluant éventuellement des informations sur leur âge, leur sexe, leur milieu social, leur parcours scolaire antérieur, etc. On peut de même inclure dans l'analyse des variables de contrôle relatives à l'enseignant (ancienneté, formation, adhésion à la méthode, etc.) ou à la classe (nombre d'élèves, niveau moyen, milieu social « moyen », etc.).

La taille de l'effet est alors obtenue en faisant la différence entre la moyenne « ajustée » obtenue par le groupe traité et celle obtenue par le groupe de contrôle au post-test, « ajustée » signifiant que l'on a pris en compte les résultats des pré-tests et l'influence éventuelle des autres variables de contrôle utilisées. Cette différence est ensuite divisée par l'écart-type du score non ajusté obtenu par le groupe de contrôle au post-test. L'effet est donc exprimé en pourcentage d'écart-type. Rappelons que l'écart-type est une mesure de la dispersion des résultats autour de la moyenne. En résumé, la construction de cet indicateur permet d'expliquer quelle part des différences de progression constatées entre élèves peut être attribuée à la méthode pédagogique employée.

Donnons quelques exemples pour comprendre les résultats qui seront présentés dans ce document de travail :

- en France, en 2009, l'écart de résultats en compétences générales en fin de collège entre élèves du public hors éducation prioritaire et élèves en éducation prioritaire était de 0,44 écart-type⁵ ; il était de 0,78 écart-type entre élèves en retard

⁵ Enquête CEDRE (Cycle d'enquêtes Disciplinaire Réalisées sur Echantillon, MENJVA-DEPP), 2009.

- scolaire et élèves « à l'heure » ;
- aux États-Unis, au stade enseignement primaire, dans le cadre d'une étude randomisée de grande ampleur (projet STAR *Student Teacher Achievement Ratio*), on a pu montrer que le fait d'être dans une petite classe (13 à 17 élèves) plutôt que dans une classe de taille normale (22 à 26 élèves) se traduisait, au bout de quatre ans, par une augmentation des résultats des élèves d'un quart d'écart-type (Nye & al., 2001) ;
 - si l'effet d'une pratique pédagogique est évalué à 0,4 écart-type, cela signifie que cette pratique devrait permettre à un élève moyen, se situant au milieu de la distribution des performances des élèves (50^e centile), de progresser de 16 places sur 100 (66^e centile) au sein de cette distribution.
- **Des études dont la méthodologie s'écarte de ce modèle, avec un impact plus ou moins important sur la robustesse des résultats**

De nombreuses études ne remplissent pas ce cahier des charges exigeant, pour des raisons de temps, de moyens ou pour des raisons éthiques soulevées notamment par l'affectation aléatoire des élèves au groupe expérimental et au groupe témoin. Les « écarts » à cette norme scientifique remettent plus ou moins en cause la validité des résultats.

Ainsi, dans le cas d'une étude randomisée et si l'échantillon est suffisamment important, l'existence d'un pré-test n'est pas absolument nécessaire. En effet, la loi des grands nombres implique en principe que les élèves des deux groupes sont similaires du point de vue de leurs caractéristiques clés, y compris leur niveau de départ.

Certaines études ne reposent pas sur une affectation aléatoire des élèves, mais sur une technique d'appariement (« *matching* »). Cette méthode suppose de comparer, un à un, les élèves du groupe expérimental aux élèves du groupe de contrôle, et d'apparier chaque élève du groupe expérimental à un ou plusieurs élèves du groupe de contrôle qui se révèlent très proches de lui du point de vue de caractéristiques clés. Cela impose en l'occurrence de bien disposer d'un pré-test, en plus des informations classiques relatives aux caractéristiques socioéconomiques de l'élève. Cette méthodologie est *a priori* moins rigoureuse que les méthodologies qui reposent sur une affectation aléatoire. En effet, l'affectation aléatoire permet de se prémunir d'éventuels biais de sélection qui peuvent surgir lorsque l'affectation à un groupe se fait sur la base du volontariat. Il se peut dans ce cas que l'enseignant ou les élèves du groupe expérimental soient davantage motivés que l'enseignant ou les élèves du groupe de contrôle, à caractéristiques observables similaires, et que ce soit ce surplus de motivation qui explique les différences de résultats, et non la méthode pédagogique elle-même.

Dans les faits, les méta-analyses réalisées aux États-Unis qui seront présentées *infra* montrent que les études randomisées et les études qui reposent sur un appariement, à condition que celui-ci soit effectué avant le début de l'expérimentation, donnent des résultats similaires, ce qui laisse supposer qu'un bon appariement est suffisant pour obtenir des résultats robustes. En revanche, lorsque l'appariement se fait *a posteriori* (« *post hoc matching* »), les résultats peuvent être faussés car on ne conserve dans l'échantillon traité que les « survivants », c'est-à-dire les classes ou les écoles où l'expérimentation a été menée à son terme (donc en général, là où elle a donné de bons résultats).

En revanche, ces méta-analyses mettent en évidence des différences de résultats importants, dans le sens d'une surévaluation de l'effet, lorsque :

- la méthodologie repose uniquement sur une comparaison des scores au pré-test et

- des scores au post-test en l'absence de groupe de contrôle ;
- l'analyse repose sur la comparaison des résultats d'un groupe expérimental et d'un groupe de contrôle sans affectation aléatoire et sans que les élèves aient été testés avant l'expérimentation.

1.2. Limites méthodologiques et première discussion

Quand bien même une recherche répondrait à ce cahier des charges exigeant, ses conclusions pourraient être discutées. De fait, ces recherches parfois dénommées « processus-produits » font l'objet de critiques nombreuses. Les plus insistantes concernent trois points principaux : la réalité des pratiques mises en œuvre par les enseignants, l'étroitesse du champ disciplinaire évalué, une focalisation sur les compétences de base au détriment des compétences de plus haut niveau.

- **Pratiques déclarées et pratiques effectives**

Un premier reproche adressé aux études sur l'efficacité des méthodes pédagogiques concerne le fait que les pratiques pédagogiques mises en œuvre par les enseignants ne font pas toujours l'objet d'une observation directe par les chercheurs. Cela est exact : l'observation directe par les chercheurs des pratiques mises en œuvre par les enseignants est plus l'exception que la règle. Lorsque l'étude est expérimentale, on *suppose* le plus souvent que l'enseignant met en œuvre (ou tente de le faire) la pratique que l'on cherche à évaluer. Mais dans quelle mesure l'enseignant du groupe expérimental adhère-t-il à la méthode pédagogique qu'il est censé utiliser ? La maîtrise-t-il réellement ? En règle générale, lorsqu'une méthode pédagogique innovante est expérimentée, les enseignants sont formés par une équipe de professionnels. Mais lorsque la méthode pédagogique est très éloignée de la pratique habituelle, ou bien lorsque les enseignants sont diversement convaincus de son efficacité, il se peut que les pratiques pédagogiques réellement mises en œuvre diffèrent d'un enseignant à l'autre.

Lorsque l'étude est corrélationnelle, c'est-à-dire lorsqu'on ne parle plus de « méthode pédagogique » mais que l'on cherche à mettre en relation d'une part les pratiques variées des enseignants en termes de feed-back, de temps passé à exposer les concepts, d'organisation du travail en groupes, etc., et d'autre part les résultats des élèves, une description fine de ces pratiques est indispensable. Dans ce cas, s'en tenir aux déclarations des enseignants, comme c'est le cas dans nombre d'études, peut être trompeur, car ceux-ci, de bonne foi, ne se représentent pas toujours très bien la réalité de leurs pratiques. Si l'on ne peut observer directement les pratiques des enseignants, ce qui prend du temps car on ne peut s'en tenir à une observation ponctuelle, il vaut généralement mieux demander aux élèves de décrire les pratiques de leurs enseignants à partir d'une grille de description fournie par les chercheurs.

Notons enfin que dans les groupes de contrôle, les pratiques pédagogiques effectivement mises en œuvre peuvent également être variables, d'un enseignant à l'autre, mais aussi d'un contexte à l'autre. Cela peut expliquer les différences de résultats constatées entre études menées dans des pays différents. Par exemple, dans le cas des méthodes de tutorat destinées aux enfants en difficulté d'apprentissage de la lecture, les études américaines enregistrent un effet positif relativement limité, alors que les études australiennes et britanniques mettent en évidence un effet beaucoup plus important. Comment expliquer cette différence ? La raison tient sans doute au fait que, d'un pays à l'autre, la méthode « traditionnellement » appliquée est différente. Aux États-Unis, les enfants en difficulté

d'apprentissage bénéficient souvent déjà d'une instruction en petits groupes qui repose sur l'approche phonologique, ce qui est moins fréquemment le cas en Australie et en Grande-Bretagne. Du coup, les méthodes pédagogiques mobilisées dans les groupes de contrôle sont plus proches de celles mobilisées dans le groupe expérimental aux États-Unis qu'elles ne le sont au Royaume-Uni ou en Australie. L'effet est donc d'ampleur plus limitée (Slavin & al., 2009).

- **Un champ de compétences étudiées souvent limité**

La quasi-totalité des évaluations disponibles porte sur les compétences de base des élèves, c'est-à-dire sur la maîtrise de la lecture et de la compréhension des textes et sur les mathématiques, le plus souvent l'arithmétique élémentaire et la résolution de problèmes. Il s'agit d'une limite évidente de cette littérature. Cette focalisation des évaluations sur les compétences de base s'explique néanmoins. D'une part, parce que ces compétences sont utiles à l'ensemble des apprentissages réalisés dans les autres matières. D'autre part, parce que, pour ces compétences, il existe de nombreux tests standardisés et, bien souvent, des évaluations régulières réalisées au niveau national, ce qui rend beaucoup plus aisé l'étalonnage de la performance des élèves inclus dans les échantillons, qu'il s'agisse des élèves du groupe expérimental ou de ceux du groupe témoin.

Plus généralement, certains aimeraient que l'on ne s'en tienne pas à une évaluation du savoir-faire, mais que l'on prenne en compte également l'effet des méthodes pédagogiques sur le bien-être des élèves et sur leur « savoir-être ». Certaines recherches mobilisent déjà des questionnaires permettant d'évaluer l'impact des méthodes sur le bien-être ou l'estime de soi des élèves. Mais une évaluation des savoir-être est difficile car elle supposerait des prises d'informations répétées, effectuées qui plus est dans des situations non contraintes dans la mesure où un savoir-être exprime l'intériorisation d'une norme de comportement (Ketele, 2009) qui n'a plus besoin d'être sollicitée.

- **Une focalisation sur des tâches simples au détriment des compétences de plus haut niveau ?**

Il est également reproché aux recherches sur les méthodes pédagogiques efficaces de ne s'intéresser qu'à la réussite des élèves sur des tâches simples ou relativement mécaniques, au détriment d'une réelle capacité à mobiliser des compétences diverses face à des situations complexes et contextualisées (Carette, 2008 ; Ketele, *op. cit.*). Certains auteurs font ainsi remarquer que les tests utilisés sont parfois assez pauvres du point de vue de la qualité du savoir qui est évalué : il n'est de fait pas équivalent d'évaluer un « savoir-restituer » (redonner une définition par exemple), un savoir-faire de base de type application (par exemple, être capable de multiplier deux entiers à deux chiffres) et un savoir-faire complexe de type résolution de problème. Le type de question utilisé peut également influencer les résultats : s'agit-il d'une question à choix multiple ou d'une question ouverte ? Dans le premier cas, la question est susceptible d'introduire dans le répertoire cognitif de l'élève des connaissances qui ne s'y trouvaient pas ou n'étaient pas mobilisables immédiatement.

Vincent Carette rend ainsi compte d'une étude qu'il a menée avec Bernard Rey, Anne Defrance et Sabine Kahn en communauté belge francophone (Carette, 2008, *op. cit.*). Ce travail repose sur un modèle d'évaluation conçu pour tester la capacité des élèves à mobiliser effectivement, et en contexte, les procédures apprises. Dans une première phase, une situation complexe et inédite est présentée aux élèves. L'élève se trouve devant une feuille blanche et doit choisir et organiser lui-même les procédures nécessaires à la

résolution du problème. Dans une deuxième phase, les élèves sont confrontés à la même situation complexe et aux mêmes tâches qu'en première phase, mais la tâche complexe est décomposée en tâches élémentaires dont les consignes sont explicites. Dans une troisième phase, on présente aux élèves des items plus classiques qui évaluent la maîtrise des procédures de base nécessaires à la résolution des tâches proposées en phases un et deux. Il en ressort que de nombreux élèves sont en difficulté lorsqu'on leur présente des tâches complexes et que la maîtrise des procédures (phase 3) est une condition nécessaire mais non suffisante à la résolution des tâches complexes.

Cette critique essentielle appelle cependant deux remarques. D'abord, tous les élèves qui réussissent les tâches complexes maîtrisent les tâches simples. S'assurer que la maîtrise de ces tâches simples est acquise par tous les élèves n'est donc pas totalement illégitime, d'autant plus que c'est encore souvent loin d'être le cas. Ensuite, les recherches les plus récentes s'attachent généralement à tester les élèves sur des exercices de niveaux de difficultés variés et croissants. Dans de nombreux cas, elles mobilisent des tests mis au point par les services du gouvernement pour mesurer les compétences des élèves attendues au niveau national, compétences dont on peut supposer qu'elles ne se limitent pas à la maîtrise des tâches simples. Et parfois, elles vont plus loin, en proposant aux élèves des exercices dont le niveau d'exigence intellectuelle est supérieur à ce qui est attendu par la tutelle. Dans tous les cas, les méta-analyses réalisées dans le contexte anglo-saxon spécifient bien, pour chaque étude recensée, les tests utilisés par les chercheurs. Il reste vrai que le développement chez les élèves des habiletés intellectuelles supérieures (conceptualisation, inférence, estimation) constitue un défi plus difficile à relever que celui de la maîtrise des compétences de base.

2. Les études anglo-saxonnes

La littérature anglo-saxonne sur l'efficacité des pratiques pédagogiques étant abondante, elle a pu faire l'objet de nombreuses synthèses. On se restreindra ici à quelques méta-analyses qui se distinguent par leur niveau élevé de rigueur scientifique, notamment dans la sélection des études retenues, afin d'en présenter les résultats les plus saillants.

2.1. L'apprentissage de la lecture

L'apprentissage de la lecture est, de tous les apprentissages, celui qui a fait l'objet du plus grand nombre d'expérimentations et d'évaluations. Le nombre élevé d'études disponibles permet aujourd'hui de se montrer particulièrement sélectif dans le choix de celles prises en considération dans les méta-analyses.

Le *National Reading Panel*, un groupe d'experts réuni sous l'autorité du Congrès des États-Unis, a ainsi eu pour mission à partir de 1998 de faire un bilan des recherches sur l'efficacité des différentes méthodes d'apprentissage de la lecture à destination des enfants. Ses conclusions ont été rendues en 2000 dans un rapport qui fait aujourd'hui référence (National Reading Panel, 2000). Seules les études publiées en anglais dans une revue à comité de lecture ont été retenues dans l'analyse. Surtout, il devait s'agir exclusivement d'études expérimentales ou quasi expérimentales⁶.

Parmi les études répondant aux critères de sélection, 52 études concernaient l'enseignement explicite (voir encadré 3) et systématique de la conscience phonémique (conscience des sons), 38 l'enseignement explicite et systématique de la phonétique (correspondance lettres/sons), 16 la lecture à haute voix, 14 la lecture en silence, 50 le développement du vocabulaire, 205 l'enseignement explicite des stratégies de compréhension, 32 la formation des enseignants et 21 l'enseignement assisté par ordinateur.

Plusieurs enseignements, de portée assez générale, ressortent de cette méta-analyse :

- l'instruction de la conscience phonémique a un effet positif sur l'apprentissage de la lecture et de la capacité à épeler. Elle ne constitue pas un apprentissage complet de la lecture, mais en est une fondation importante ;
- l'enseignement de la phonétique a un effet significatif tant sur le décodage que sur la compréhension. L'effet est particulièrement important pour les enfants qui ont des difficultés d'apprentissage, souffrent de handicaps ou sont issus de milieux sociaux défavorisés ;
- l'entraînement à la lecture à voix haute a une influence positive sur la reconnaissance des mots et l'aisance en lecture tandis que la lecture silencieuse ne présente pas d'effet sur les performances en lecture ;
- les méthodes de développement du vocabulaire ont une certaine efficacité pour améliorer la compréhension, mais la méthode doit être appropriée à l'âge et aux compétences de l'enfant. L'apprentissage du vocabulaire doit se faire à la fois

⁶ Le « quasi » peut renvoyer au fait que l'affectation des élèves au groupe de contrôle et au groupe témoin n'a pas pu se faire de façon aléatoire ou bien qu'un seul groupe d'élèves est suivi dans le temps mais fait l'objet de plusieurs traitements, ou bien encore qu'il n'existe pas à proprement parler de groupe de contrôle mais que l'on compare les résultats de plusieurs groupes expérimentaux, avec une prise d'information avant l'expérimentation et après l'expérimentation.

directement et indirectement. La répétition et l'exposition répétée aux nouveaux mots sont importantes ;

- en ce qui concerne l'enseignement des stratégies de compréhension de textes, plusieurs exercices permettent de faire progresser les élèves : apprendre aux élèves à être conscients de leur niveau de compréhension du texte ; demander aux élèves d'échanger sur leurs stratégies de compréhension (apprentissage coopératif) ; utiliser des représentations graphiques pour aider à la compréhension ; demander aux élèves de répondre à des questions posées par l'enseignant ; leur demander de formuler eux-mêmes des questions sur le texte qu'ils sont en train de lire ; leur apprendre à prendre conscience de la structure du récit ; leur apprendre à résumer un texte (encadré 1) ;
- les formations à destination des enseignants en poste permettent une amélioration des acquisitions des élèves ;
- l'apprentissage assisté par ordinateur semble donner des résultats positifs mais beaucoup de questions restent sans réponse sur la façon d'utiliser efficacement cet outil.

En 2009, Slavin *et al.* (2009, *op. cit.*) ont réalisé une synthèse de 96 études expérimentales évaluant l'efficacité de programmes à destination des enfants connaissant des difficultés de lecture en primaire. Là encore, les critères de sélection des études étaient stricts. L'affectation des élèves au groupe traité et au groupe de contrôle devait être aléatoire ou bien l'étude devait utiliser une méthode d'appariement de qualité reposant sur les résultats aux pré-tests et d'autres variables exogènes (catégorie socioprofessionnelle des parents, âge, sexe, etc.). Les études dans lesquelles on constatait des différences supérieures à 50 % d'un écart-type entre le groupe traité et le groupe de contrôle lors des pré-tests n'étaient pas retenues. Les mesures des performances en lecture devaient être des mesures communément utilisées, les mesures construites par l'expérimentateur n'étant acceptées que si elles étaient indépendantes du traitement (exclusion par exemple des mesures de conscience phonique) et ne désavantageaient pas le groupe de contrôle. Au moins 12 semaines devaient s'écouler entre les pré-tests et les post-tests. Enfin, la taille des groupes traités devait être suffisante (au moins 15 élèves et deux enseignants/tuteurs par groupe).

Les résultats de cette synthèse sont les suivants :

- le tutorat en face-à-face par des enseignants (19 études ont été analysées, dont 5 avec affectation aléatoire) a un effet positif évalué en moyenne pondérée⁷ à 0,38 écart-type ;
- le tutorat par des para-professionnels (11 études analysées, dont 6 avec affectation aléatoire) donne des résultats similaires, avec un effet moyen pondéré de 0,38 écart-type ;
- le tutorat par des bénévoles (7 études analysées, dont 4 avec affectation aléatoire) semble un peu moins efficace, avec un effet moyen pondéré de 0,16 écart-type. Mais l'effet remonte à 0,51 écart-type si l'on exclut les programmes *Experience Corps* et *West Philadelphia Tutoring*, qui ne proposent des séances qu'une à deux fois par semaine ;
- les programmes d'instruction en petits groupes, centrés sur la phonétique (20 études, dont 16 avec affectation aléatoire) donnent un effet moyen pondéré de 0,31 écart-type ;
- les programmes visant à modifier le processus d'apprentissage de la lecture en

⁷ La pondération se fait en fonction de la taille des échantillons. Pour chaque étude, un effet moyen est calculé lorsque la maîtrise de plusieurs dimensions de la lecture est évaluée (reconnaissance de mots, vocabulaire, compréhension, fluence, etc.).

classe entière (16 études dont 1 avec affectation aléatoire) démontrent un effet positif moyen de 0,56 écart-type. Les effets sont plus forts pour les élèves en difficulté que pour l'ensemble des élèves. Dans 8 des 16 études, les programmes concernent la mise en place de modèles coopératifs d'apprentissage, où les élèves s'entraident, avec un fort accent sur la phonétique. Pour les autres programmes, il s'agit essentiellement de la mise en place d'approches systématiques et explicites, avec un fort accent sur la phonétique ;

- les programmes visant à mettre en place une instruction assistée par ordinateur (14 études dont 5 avec affectation aléatoire) semblent avoir un effet positif très limité : + 0,09 écart-type en moyenne.

Les études recensées confirment donc l'importance du travail de la phonétique pour les élèves rencontrant des difficultés de lecture. Il semble aussi efficace, mais nettement moins coûteux, de revoir la méthode d'instruction de la lecture en contexte normal (classe entière) que de proposer aux élèves en difficulté des programmes intensifs de tutorat en petits groupes ou en face-à-face. Les auteurs précisent que pour obtenir des effets durables, il convient d'assurer un suivi au-delà de la première année d'école primaire, car les effets des programmes de tutorat en face-à-face généralement mis en place en première année de primaire s'estompent assez vite avec le temps.

Encadré 1

L'enseignement de la compréhension de textes⁸

Dans les années 1970 aux États-Unis, les recherches de Dolores Durkin avaient montré qu'un temps considérable était consacré dans les classes de primaire à l'enseignement phonologique et à l'analyse du code, mais que très peu de temps était consacré à l'enseignement de la compréhension. Pourtant, la compréhension d'un texte ne va pas de soi. Ainsi, Connor *et al.* (2004) ont montré que des activités implicites d'enseignement de la compréhension, telles que la lecture silencieuse, n'ont aucune incidence sur l'évolution des performances en compréhension de lecture.

Plusieurs travaux indiquent pourtant qu'il est possible d'enseigner aux élèves des stratégies cognitives spécifiques pour leur permettre de résoudre cette tâche *a priori* peu structurée qu'est la compréhension de textes.

Plusieurs stratégies peuvent être utilisées :

- un enseignement explicite fondé sur l'interprétation des unités anaphoriques (pronoms personnels, expressions définies, etc.), l'analyse morphologique et les stratégies contextuelles ;
- un « enseignement transactionnel de stratégies », dans lequel les élèves, guidés par l'enseignant, sont invités à échanger activement autour des interprétations qu'ils fournissent et des procédures à mettre en œuvre pour parvenir à une interprétation consensuelle. En France, Bianco (2003) a montré que l'enseignement transactionnel de stratégies pouvait être efficace pour enseigner à des élèves de CE2 et de CM2 des stratégies d'interprétation des anaphores. Lima *et al.* (2006) ont également mis en évidence l'efficacité de l'enseignement transactionnel pour enseigner des stratégies générales de compréhension ;

⁸ Cet encadré reprend des éléments présentés dans Bianco M. et Bressoux P. (2009).

- un enseignement réciproque, où les élèves apprennent à justifier et à négocier les interprétations qu'ils font d'un texte et s'aident mutuellement dans l'utilisation des stratégies de compréhension. À mesure qu'ils deviennent plus habiles, les élèves réalisent la majeure partie du travail d'interprétation. Une revue de la littérature menée par Rosenshine et Meister (1994) indique que les méthodes d'enseignement réciproque des stratégies de compréhension sont efficaces, mais ne le deviennent que pour les élèves du cours moyen, et qu'elles atteignent leur maximum d'efficacité au collège. Les meilleurs résultats sont obtenus lorsque l'enseignement réciproque est précédé d'une phase de présentation explicite de chaque stratégie.

2.2. L'apprentissage des mathématiques

La littérature, principalement anglo-saxonne, sur l'apprentissage des mathématiques a fait l'objet de trois méta-analyses récentes qui répondent à un haut niveau d'exigence dans la sélection des études prises en considération.

La première, réalisée par Slavin et Lake en 2007, concerne l'enseignement des mathématiques au primaire. Plusieurs stratégies d'intervention sont distinguées. La première consiste à intervenir sur les contenus enseignés (notamment *via* le changement de manuel) sans modifier la pédagogie (13 études dont 2 avec affectation aléatoire et 10 avec appariement *a posteriori*). L'effet moyen de ce type d'intervention reste modeste puisqu'il est d'environ 0,10 écart-type. La deuxième stratégie consiste à mobiliser l'outil informatique en s'appuyant, en plus du cours, sur une instruction assistée par ordinateur (38 études analysées dont 15 expérimentations avec affectation aléatoire). L'effet médian de ce type d'intervention est supérieur au précédent, puisqu'il est évalué en moyenne à 0,19 écart-type. Les effets sont, comme on peut s'y attendre, plus forts sur les tâches de calcul que sur celles relevant de la maîtrise des concepts ou de la résolution de problèmes.

Le troisième type d'intervention repose sur une modification de la pédagogie à contenu enseigné inchangé (36 études dont 19 avec affectation aléatoire). Les programmes étudiés sont classés en sept catégories : apprentissage coopératif (encadré 2) ; apprentissage coopératif accompagné d'un suivi individualisé des élèves ; instruction directe ; pédagogie de la maîtrise (*Mastery learning*) ; formation des enseignants centrée sur les mathématiques ; formation des enseignants centrée sur la gestion de classe et la motivation des élèves ; cours supplémentaires. Les études sont en général de bien meilleure qualité que celles relatives aux deux types d'intervention précédents. L'effet médian pour les études expérimentales avec affectation aléatoire est relativement important : + 0,33 écart-type. Pour l'ensemble des études, l'effet médian est aussi de + 0,33 écart-type. Plusieurs types d'intervention dans le domaine de la pédagogie semblent faire leurs preuves, dont l'apprentissage coopératif (9 études, effet médian de + 0,29 écart-type), l'instruction directe (4 études, effet supérieur à 0,3 écart-type mais échantillons réduits), mais aussi la formation des enseignants, qu'il s'agisse d'un enseignement centré sur les mathématiques ou plus généralement sur la gestion de classe et la motivation des élèves.

Encadré 2

L'apprentissage coopératif

L'apprentissage coopératif repose sur le travail en petits groupes hétérogènes, généralement de deux ou quatre élèves. Au sein du groupe, les élèves ne sont pas laissés à l'abandon : le travail est structuré de façon à s'assurer que chaque élève participe effectivement à l'accomplissement de la tâche proposée. La coopération peut être obtenue par un encouragement à la discussion des points de vue ou bien par un partage des rôles au sein du groupe qui rend réellement les élèves dépendants les uns des autres.

L'intérêt de l'apprentissage coopératif réside dans le fait qu'il met les élèves en interaction et les incite ainsi à verbaliser et à reformuler leurs idées, à comparer leurs stratégies et leurs façons d'apprendre. Il nécessite néanmoins un rôle actif de l'enseignant, qui doit s'assurer que l'élève le plus à l'aise aide effectivement celui qui a le plus de difficulté, que ce dernier accepte cette aide, qu'un élève ne prend pas tout simplement en charge, seul et pour l'ensemble du groupe, la résolution du problème, etc. Il s'agit donc d'une pratique qui demande un haut niveau d'attention de la part de l'enseignant en même temps qu'une grande technicité.

Une seconde synthèse, consacrée cette fois à l'enseignement des mathématiques dans le secondaire, a été réalisée par Slavin, Lake et Groff en 2008, avec les mêmes critères d'exigence pour la sélection des études. On retrouve les trois types d'intervention évoqués *supra*. Pour l'intervention sur les contenus (40 études dont 8 avec affectation aléatoire), les résultats sont souvent non significatifs (effet médian de + 0,03 écart-type). Seul un manuel (*Saxon Textbooks*) permet un faible gain pour les élèves, évalué à 0,13 écart-type. L'apport de l'instruction assistée par ordinateur (40 études) paraît lui aussi assez faible, avec un effet médian estimé à + 0,08 écart-type. Aucun programme ne sort vraiment du lot.

L'effet médian des interventions sur la pédagogie (22 études) est plus significatif puisqu'il est estimé à 0,18 écart-type. Mais les résultats varient fortement selon le programme. Deux programmes semblent faire preuve d'une particulière efficacité :

- le programme *Student Teams-Achievement Divisions (STAD)* qui repose sur un travail en équipe hétérogène de quatre élèves. Le temps d'enseignement est partagé entre enseignement en classe complète, travail en groupe et travail individuel. Les équipes reçoivent des gratifications symboliques (certificats) sur la base des scores moyens réalisés par chacun de leurs membres lors des quizz hebdomadaires. Pour ce programme, quatre études sont disponibles dont trois avec affectation aléatoire, avec un effet moyen pondéré de + 0,38 écart-type.
- le programme *Improve*, un programme israélien qui repose à la fois sur l'apprentissage coopératif et sur la mise en œuvre de stratégies métacognitives, pour lequel on dispose de trois études, avec un effet moyen pondéré de + 0,52 écart-type. Les stratégies métacognitives désignent des méthodes par lesquelles on apprend aux élèves à se poser à haute voix des questions de compréhension, à rechercher des similarités et des différences avec d'autres problèmes abordés précédemment, à s'approprier des stratégies de résolution.

Une autre méta-analyse, centrée sur les élèves en difficulté en mathématiques, a été publiée en 2007 par Gersten et Clarke. Ses résultats, qui sont résumés dans le tableau 1, confirment l'intérêt de **l'instruction explicite et de l'apprentissage coopératif**. La verbalisation du

raisonnement a également un impact très positif sur les résultats des élèves. Elle constitue un moyen, pour les élèves, de rendre explicites les stratégies mises en œuvre pour résoudre les problèmes et se rapproche ainsi de l'objectif de l'instruction explicite.

Tableau 1
Effets de certaines stratégies d'instruction pour les élèves en difficulté en mathématiques ou en éducation spécialisée (écart-type)

Stratégie d'instruction	Ampleur de l'effet pour les élèves en éducation spécialisée	Ampleur de l'effet pour les élèves en difficulté
Description visuelle et graphique des problèmes	0,50	Pas de données
Instruction systématique et explicite	1,19	0,58
Raisonnement à voix haute	0,98	Pas de données
Apprentissage coopératif entre pairs dans des groupes hétérogènes	0,42	0,62
Mise à disposition des enseignants d'évaluations formatives	0,32	0,51
Mise à disposition des élèves d'évaluations formatives	0,33	0,57

Encadré 3
L'enseignement explicite⁹

Formalisé tout d'abord dans les années 1960 par Siegfried Engelmann, créateur de la méthode de *Direct Instruction*, l'enseignement explicite repose sur l'idée générale que pour aborder des tâches nouvelles et/ou complexes, il faut avoir automatisé un certain nombre de procédures de base, et qu'il convient donc, d'une part, d'aller du simple au complexe, et d'autre part, de décomposer autant que possible les tâches complexes en tâches élémentaires.

Concrètement, l'enseignement explicite se divise en trois grandes étapes :

- le modelage : l'information est présentée en petites unités, dans une séquence graduée, généralement du simple au complexe. La présentation d'une trop grande quantité d'informations nuit à la compréhension en surchargeant la mémoire de travail de l'élève ;
- la pratique guidée ou dirigée : l'enseignant propose aux élèves des tâches semblables à celles qui ont été effectuées à l'étape du modelage. Durant cette phase, il les questionne de façon à établir une rétroaction régulière. Cette étape peut être favorisée par le travail en équipe qui permet aux élèves de vérifier leur compréhension en échangeant des idées entre eux ;
- la pratique autonome : l'enseignant ne délaisse la pratique guidée pour la pratique autonome que lorsqu'il s'est assuré que les élèves ont atteint un niveau de maîtrise élevé de la matière à apprendre.

⁹ Cet encadré reprend des éléments présentés dans Bissonnette *et al.* (2005).

L'enseignement explicite ne doit pas être confondu avec le cours magistral. C'est dans la deuxième étape de sa démarche, c'est-à-dire la pratique guidée, que l'enseignement explicite s'en distingue fondamentalement.

2.3. L'apprentissage des sciences

Les études qui cherchent à évaluer l'effet de diverses approches pédagogiques pour l'enseignement des sciences sont beaucoup plus rares que celles consacrées à l'apprentissage de la lecture ou des mathématiques. Par exemple, sur le site *What Works Clearing House*, du ministère de l'Éducation américain, 77 « interventions » (programme, pratique pédagogique, politique éducative) concernent l'apprentissage de la lecture et de l'écriture, 39 les mathématiques et seulement 5 les sciences.

Une rapide consultation des revues de la littérature suggère que la question la plus débattue concerne les mérites de l'approche fondée sur l'investigation par les élèves (*inquiry-based learning*), vis-à-vis de méthodes d'instruction plus directes. Mais les études disponibles relatent souvent des expériences menées sur de très courtes périodes (une semaine par exemple), et sur de petits échantillons d'élèves. Du coup, la qualité des méta-analyses ne peut atteindre le niveau de celles que l'on a présentées dans le cas de l'apprentissage de la lecture et des mathématiques. Leurs résultats sont souvent contradictoires¹⁰.

Ces contradictions tiennent en partie au fait qu'il est difficile de définir précisément de quoi l'on parle. Le vocabulaire employé est souvent différent d'une étude à l'autre, et surtout, les pratiques sous-jacentes peuvent beaucoup varier. La méthode d'apprentissage par investigation peut ainsi être utilisée pour enseigner aux élèves ce qu'est la nature du travail scientifique ; pour enseigner aux élèves comment mener un travail scientifique (apprendre à investiguer soi-même) ; pour aider les élèves à mieux comprendre et s'approprier les résultats de la science (apprendre les connaissances scientifiques par l'expérimentation).

Pour éclairer le débat, les auteurs tendent de plus en plus à distinguer plusieurs niveaux d'investigation, selon la latitude laissée aux élèves dans le choix des questions à traiter et des méthodes à mobiliser. Ces niveaux sont résumés dans le tableau 2, tiré de Blanchard *et al.* (2010). Dans la réalité des pratiques, bien sûr, les choses ne sont pas aussi tranchées.

¹⁰ Pour des évaluations contrastées, on pourra consulter Barron et Darling-Hammond (2008) ; Bruder et Prescott (2013) ; Kirschner *et al.* (2006).

Tableau 2
Niveaux d'investigation

	Source des questions scientifiques	Méthodes de collecte de données	Interprétation des résultats
Niveau 0 : vérification	Données par l'enseignant	Données par l'enseignant	Donnée par l'enseignant
Niveau 1 : investigation structurée	Données par l'enseignant	Données par l'enseignant	Laissée aux élèves
Niveau 2 : investigation guidée	Données par l'enseignant	Laissées aux élèves	Laissée aux élèves
Niveau 3 : investigation ouverte	Laissées aux élèves	Laissées aux élèves	Laissée aux élèves

Source : Blanchard *et al.* (2010).

Les conclusions que l'on peut tirer de ces analyses contradictoires sont nécessairement nuancées. Les méthodes d'investigation présentent un intérêt pour l'enseignement des sciences, à condition que les élèves ne soient pas livrés à eux-mêmes et que l'enseignant soit en mesure de les guider efficacement dans leur questionnement.

Par exemple, Blanchard *et al.* (2010) ont cherché à comparer les effets d'une méthode d'investigation guidée (niveau 2 de l'investigation dans le tableau précédent) à ceux d'une méthode de simple vérification (niveau 0) dans le cadre d'une série de séances de laboratoire s'étendant sur une semaine. Les enseignants assignés à la méthode d'investigation guidée recevaient une formation *ad hoc* de six semaines. Les enseignants étaient filmés et leurs pratiques évaluées par deux évaluateurs externes. Ce recours à l'observation permettait d'une part de s'assurer que les enseignants des deux groupes travaillaient effectivement différemment avec les élèves, mais permettait aussi, au sein de chaque groupe, de voir si le fait d'être plus ou moins fidèle aux pratiques attendues avait un impact sur les résultats des élèves.

Les meilleurs résultats, mais aussi les moins bons, étaient obtenus par des enseignants affectés à la méthode d'investigation guidée. Celle-ci se révélait donc plus sensible que la méthode de simple vérification aux pratiques effectives des enseignants. Une méthode d'investigation guidée peut donc être plus efficace qu'une méthode de simple vérification, mais sa bonne mise en œuvre semble aussi plus délicate.

La plupart des études s'accordent pour reconnaître que les démarches qui reposent sur l'investigation sont plus appréciées des élèves que les méthodes d'instruction plus traditionnelles. Elles constituent un moyen efficace d'engager les élèves dans l'apprentissage. Même si leur efficacité pour acquérir des connaissances scientifiques (lois universelles, faits, concepts) ne se révélait pas supérieure aux méthodes d'instruction directe, elles conserveraient leur intérêt pour faire comprendre aux élèves la nature du travail scientifique et pour leur faire acquérir la maîtrise des méthodes expérimentales.

2.4. Le projet *Follow Through*

Les études recensées dans les méta-analyses évoquées jusqu'ici sont des expériences contrôlées, qui s'appliquent à des échantillons de taille forcément limitée, et qui visent à tester l'efficacité de pratiques pédagogiques sur des objets bien définis comme l'apprentissage de la lecture/compréhension. Il est intéressant de mentionner une expérimentation de beaucoup plus grande envergure, menée aux États-Unis à la fin des

années 1960 : le projet *Follow Through*.

- **L'expérimentation**

L'idée originale de ce projet était de proposer aux enfants issus de milieux défavorisés des services éducatifs qui prennent le relais de ceux proposés par le programme *Head Start*, lancé en 1965 et destiné pour sa part aux enfants en bas âge avant leur entrée à l'école. Par manque de moyens, le projet initial a été réorienté pour devenir une expérience de grande échelle permettant de tester l'efficacité de plusieurs modèles éducatifs. Plus de 75 000 enfants pauvres, répartis dans 170 villes, ont participé à cette expérimentation lancée à partir de 1969 et autorisée par un amendement à l'*Economic Opportunity Act* de 1967. Le coût du projet, mené sur dix ans entre 1969 et 1978, était évalué en 1978 à 500 millions de dollars, dont 30 à 50 millions de dollars pour l'évaluation de ses effets.

Les écoles étaient sélectionnées parce qu'elles accueillaient des enfants pauvres, dont les résultats étaient très inférieurs aux moyennes nationales (aux alentours du 20^e centile de la distribution des scores en moyenne). Dans chacune des écoles sélectionnées, les parents d'élèves choisissaient le modèle éducatif qui allait être mis en pratique par les enseignants, ceux-ci étant appuyés par des équipes de professionnels spécialistes du modèle retenu. Les enfants étaient exposés au « traitement » entre l'équivalent de la dernière année de l'école maternelle et la troisième année d'école élémentaire. Après appel à candidature et sélection, 9 modèles ont ainsi pu être évalués.

- **Les modèles éducatifs évalués**

De manière nécessairement réductrice, on peut classer les 9 modèles en trois catégories : les modèles centrés sur les compétences de base, les modèles visant à développer les compétences conceptuelles, et enfin les modèles visant un développement affectif de l'enfant.

Dans la première catégorie, on trouvait :

- le modèle *Direct Instruction* (testé sur 10 sites) : centré sur les apprentissages de base et reposant sur un enseignement très structuré et incrémental, il fonctionne sur un rythme assez élevé de questions-réponses et de *feed-backs*. Les matières ou sujets qui ne font pas partie des compétences prioritaires ne sont abordés qu'une fois acquise la maîtrise de ces compétences. Le travail peut se faire en petits groupes homogènes ;
- le modèle *Behavior Analysis* (testé dans 8 sites) : assez similaire au *Direct Instruction Model*, il s'en distingue notamment par la mise en place d'un système de récompenses pour les enfants ;
- le modèle *Language Development* : modèle centré sur les apprentissages de base mais s'adressant à des enfants bilingues et se centrant sur l'apprentissage des langues.

Dans la deuxième catégorie, on peut rassembler :

- le modèle *Cognitively Oriented Curriculum* (testé dans 6 sites) : centré sur le développement de l'enfant, en référence aux travaux de Piaget, notamment à sa capacité à raisonner, il consiste à donner aux enfants le choix de leurs activités. Le travail avec l'enseignant se fait en petits groupes. Les enseignants sont formés pour devenir des catalyseurs et des « motivateurs » de l'apprentissage des enfants plutôt que des entraîneurs ou des fournisseurs d'information ;
- le modèle *Florida Parent Education* (testé dans 9 sites) : comme son nom l'indique,

ce modèle donne la priorité à l'éducation par les parents. D'inspiration piagétienne, il vise à motiver les parents et à les former pour qu'ils soient les premiers éducateurs de leurs enfants. Tout est ainsi fait pour les associer le plus étroitement possible à l'éducation de leurs enfants, que ce soit en classe ou à la maison ;

- le modèle *Tucson Early Education* (testé sur 12 sites) : centré lui aussi sur le développement de l'enfant, il repose sur un travail en petits groupes hétérogènes et vise à favoriser les expériences d'apprentissage.

Enfin, dans la troisième catégorie, on peut regrouper :

- le modèle *EDC Open Education* (testé dans 8 sites) : d'inspiration piagétienne, il est centré sur l'enfant, son développement et ses besoins affectifs. Les enfants ne sont pas comparés les uns aux autres et ne sont pas évalués. La classe est divisée en « aires d'intérêt », où l'enfant peut trouver du matériel pour expérimenter et apprendre seul. L'enseignant doit être capable de diagnostiquer les forces, le potentiel et les centres d'intérêt de chaque enfant, et de lui fournir, en conséquence, des expériences d'apprentissage pertinentes, au moment adéquat ;
- le modèle *Responsive Education* (testé dans 11 sites) : centré sur les expériences de l'enfant, il entend s'adosser aux expériences vécues en dehors du cadre scolaire ;
- le modèle *Bank Street* (testé dans 8 sites) : lui aussi centré sur le développement affectif, il propose aux enfants des expériences variées qui doivent être autant d'occasions d'apprendre.

- **L'évaluation et ses résultats**

Du fait du caractère non aléatoire du design expérimental, il a fallu constituer un groupe de contrôle se rapprochant le plus possible des caractéristiques du groupe expérimental. Une école, aux caractéristiques proches et située dans les environs, devait ainsi être identifiée pour servir de groupe de contrôle. Malheureusement, dans la plupart des cas, les écoles du groupe de contrôle accueillait une population moins désavantagée que la population accueillie par les écoles du groupe expérimental. Du coup, un traitement économétrique (analyse de la covariance) a été effectué pour contrôler la différence de recrutement social des élèves entre groupes expérimentaux et groupes de contrôle.

L'échantillon qui sert à évaluer les effets des différentes approches comprend 9 255 enfants dans le groupe expérimental et 6 485 enfants dans le groupe témoin. Tous les enfants ont été testés à l'entrée dans le programme et ensuite au printemps de chaque année, jusqu'en troisième année d'école élémentaire.

Les tests devaient permettre de mesurer trois types de compétences :

- les compétences de base : relations phonèmes-graphèmes, vocabulaire, reconnaissance de mots, arithmétique, capacité à épeler correctement, ponctuation, reconnaissance de nombres, lecture de mots, résolution orale et écrite de problèmes de mathématique, etc. ;
- les compétences cognitives/conceptuelles : compréhension de texte, concepts mathématiques (principes mathématiques et relations), résolution de problèmes mathématique, tests de logique (matrices progressives de Raven) ;
- les compétences affectives : ces compétences étaient évaluées *via* des tests psychologiques permettant de juger dans quelle mesure l'enfant attribue sa réussite ou ses échecs à lui-même ou à des forces extérieures et d'évaluer son estime de soi, la façon dont il pense que les autres le perçoivent et son sentiment vis-à-vis de l'école.

Le premier résultat de l'évaluation est un résultat assez décourageant : alors que les

enseignants bénéficiaient de l'aide de deux assistants et d'une aide financière de 350 dollars par élève, la plupart des approches se sont révélées incapables de faire réaliser aux élèves des gains d'apprentissage significativement plus élevés que dans les classes des groupes témoin. Une méthode innovante n'est donc pas nécessairement plus efficace qu'une méthode qui ne l'est pas.

Parmi les modèles testés, seul le modèle d'instruction directe tirait son épingle du jeu, enregistrant les meilleurs résultats en moyenne sur les trois dimensions (compétences de base, compétences cognitives, compétences affectives). Les modèles mettant en avant l'importance des compétences affectives furent ceux qui eurent les moins bons résultats dans ce domaine. Ceux qui mettaient l'accent sur les compétences cognitives se sont révélées incapables de faire progresser davantage les élèves sur ce type de compétences et ont obtenu de moins bons résultats dans ce domaine que les approches centrées sur les compétences de base.

- **Une évaluation critiquée**

Lors de la parution du rapport d'évaluation, la plupart des approches testées ayant fait l'objet d'une évaluation négative, les critiques furent nombreuses.

Une commission (Commission House), financée par la Fondation Ford, fut donc diligentée pour ré-analyser les résultats. Elle rendit son rapport en 1978¹¹, critiquant une méthodologie souffrant de trop nombreuses faiblesses : validité scientifique discutable de l'analyse de la covariance, insuffisance des mesures destinées à évaluer les compétences affectives et cognitives, au détriment des approches centrées sur ce type de compétences, non significativité des résultats dès lors que l'on raisonne au niveau site et non au niveau élève, etc. Chacun des points de critique fit lui-même l'objet de réponses circonstanciées, tant de la part des responsables de l'évaluation initiale (Anderson *et al.*, 1978) que des représentants du ministère américain de l'éducation (Wisler *et al.*, 1978) ou de chercheurs favorables au modèle d'instruction directe (Bereiter et Kurland, 1981). Critiques et contre-arguments sont synthétisés dans le tableau 3.

Quelques leçons peuvent être tirées de cette expérience coûteuse : i) il est très difficile d'apporter la preuve irréfutable qu'une méthode ou une approche est plus efficace qu'une autre, et ce d'autant plus que les critères à l'aune desquels on juge l'efficacité font eux-mêmes l'objet de controverses ; ii) le caractère innovant d'une méthode pédagogique ne garantit en rien sa supériorité par rapport à une méthode « classique » ou « traditionnelle » ; iii) l'effet d'une méthode ou d'une approche pédagogique sur les résultats des élèves est en moyenne d'ampleur relativement limitée par rapport à d'autres facteurs déterminants, à commencer par les caractéristiques individuelles des élèves.

¹¹ Pour une présentation, voir House *et al.* (1978).

Tableau 3
**Le projet *Follow Through* :
synthèse des controverses**

Critiques : House, Glass, McLean et Walker	Réponses et contre-arguments
<p>Les groupes de contrôle sont trop différents des groupes expérimentaux.</p> <p>Du coup, l'évaluation repose sur un ajustement des scores bruts par une analyse de la covariance dont la validité scientifique fait débat.</p>	<p>Réponse de Anderson, St. Pierre, Proper, Stebbins (responsables de l'évaluation initiale).</p> <p>Il ne faut pas exagérer les écarts : dans 47 % des 138 sites, les différences aux pré-tests sont inférieures à un quart d'écart-type et dans 77 % des sites, ils sont inférieurs à un demi écart-type. Par ailleurs, ces écarts sont en faveur du groupe expérimental dans 44 % des cas et en faveur du groupe témoin dans 56 % des cas, donc il n'y a pas de biais systématique.</p> <p>Réponse de Wisler, Burns et Iwamoto (représentants du ministère de l'éducation américain).</p> <p>Les expérimentations randomisées posent aussi problème, car il est difficile de s'assurer d'une adhésion des équipes éducatives et des parents à une méthode qui leur est totalement imposée. Or le niveau de cette adhésion influence inéluctablement les résultats.</p> <p>Réponse de Bereiter et Kurland (chercheurs favorables au modèle d'instruction directe).</p> <p>L'absence de groupes de contrôle vraiment adéquats interdit peut-être de comparer les classes expérimentales aux classes témoins, mais n'interdit pas de comparer les performances des classes expérimentales entre elles. Lorsqu'on fait ce travail (en incluant des variables de contrôle au niveau classe et non individus), les différences de performance entre les modèles redeviennent significatives, et la supériorité des approches directes ou explicites sur les approches cognitives et surtout affectives ne fait plus de doute. Deux modèles au contraire sous-performent presque systématiquement : <i>EDC Open Education</i> et <i>Responsive Education</i>.</p>
<p>L'évaluation est injuste car les tests utilisés n'évaluaient qu'un spectre étroit des compétences développées chez les enfants et ne permettaient donc pas de mettre en évidence tous les effets des interventions. En particulier, l'atteinte des objectifs affectifs, priorités de certains modèles, est insuffisamment mesurée.</p>	<p>Réponse de Anderson, St. Pierre, Proper, Stebbins.</p> <p>L'évaluation ne mesure pas tout, mais ce qu'elle mesure, elle le fait avec le même instrument pour tout le monde, donc elle n'est pas injuste. On peut toujours prétendre que l'atteinte de certains objectifs importants n'est pas évaluée, encore faut-il qu'elle soit évaluable. Les personnes qui ont critiqué les tests n'ont pas proposé d'instruments alternatifs permettant d'évaluer l'atteinte de ces objectifs.</p> <p>Réponse de Wisler, Burns et Iwamoto.</p> <p>Le champ des compétences évaluées est jugé trop étroit, mais l'évaluation réalisée nécessitait déjà de mobiliser les élèves durant 35 heures de test. On ne mesure pas tout, mais tout le monde s'accorde à dire que les compétences évaluées sont des compétences cruciales pour la réussite future des élèves, et que le test utilisé (<i>Metropolitan Achievement Test</i>) est un test pertinent. Donc il n'y a pas d'injustice. On mesure ici des différences d'efficacité, pas des différences d'effets.</p> <p>Réponse de Bereiter et Kurland.</p> <p>Toutes les études indiquent une forte corrélation entre les compétences « affectives » et les compétences évaluées dans les tests utilisés.</p>
<p>Les méthodes qui reposent sur l'apprentissage par cœur</p>	<p>Réponse de Wisler, Burns et Iwamoto.</p>

<p>sont favorisées puisqu'on évalue surtout des compétences mécaniques (lecture, écriture, arithmétique).</p>	<p>Les compétences évaluées ne s'acquièrent pas par apprentissage par cœur. On ne comprend pas ce qu'on lit en apprenant par cœur des mots. Par ailleurs, la question est de savoir si les compétences évaluées sont importantes à acquérir. La réponse est oui. Donc ces compétences doivent être acquises, d'une façon ou d'une autre. Une méthode qui ne permet pas de les acquérir n'est pas une méthode que l'on peut recommander.</p>
<p>Le classement des modèles en trois grands types est impropre, de même que le classement des différents tests dans les trois grandes catégories de compétences, qui recouvrent d'ailleurs les trois types de modèles : compétences de base ; compétences cognitives et conceptuelles ; compétences affectives. Ces catégorisations jouent un rôle crucial dans l'interprétation des résultats.</p>	<p>Réponse de Anderson, St. Pierre, Proper, Stebbins.</p> <p>Le rapport d'évaluation synthétise les résultats en catégorisant les résultats aux tests par grandes catégories de compétences et les modèles par grands types d'approches. Mais rien n'interdit de consulter les résultats détaillés, soit 2 000 réponses à des questions du type « dans ce site, pour cette cohorte, et sur ce sous-test, dans quelle mesure les résultats des élèves du groupe expérimental diffèrent-ils de ceux des élèves du groupe témoin ? »</p> <p>Réponse de Wisler, Burns et Iwamoto.</p> <p>Les résultats sont disponibles de façon désagrégée. La catégorisation des modèles ou des compétences n'a pas d'importance pour qui veut se donner la peine de lire le rapport.</p>
<p>L'instrument destiné à mesurer les compétences scolaires est bon. Mais les compétences cognitives sont en partie mesurées par les matrices progressives de Raven, or ce n'est pas leur objet car elles sont destinées à mesurer le QI. Par ailleurs, le choix des tests utilisés pour mesurer les compétences affectives est discutable.</p>	<p>Réponse de Wisler, Burns et Iwamoto.</p> <p>Les matrices de Raven ne sont effectivement pas pertinentes pour mesurer un apprentissage, puisque, mesurant un QI, elles sont justement faites pour être insensibles à un entraînement de la part de celui qui passe le test. En revanche, les tests <i>Coopersmith Self-Esteem Inventory</i> et <i>Intellectual Achievement Responsibility Scale</i>, utilisés pour mesurer les compétences affectives, sont deux tests qui ont été recommandés par les panels d'experts auditionnés par le ministère.</p> <p>Les critiques ne proposent par ailleurs rien comme épreuve de remplacement.</p>
<p>Le niveau élève, utilisé dans l'évaluation, n'est pas le bon niveau. En menant l'analyse statistique au niveau des sites, les écarts entre modèles ne sont plus significatifs, même si le classement des modèles du point de vue de leur efficacité n'est pas fondamentalement bouleversé.</p>	<p>Réponse de Anderson, St. Pierre, Proper, Stebbins.</p> <p>Si l'analyse n'a pas été menée au niveau site, c'est justement parce qu'une telle analyse donnerait des résultats non significatifs. Par ailleurs, si l'analyse menée par House <i>et al.</i> donne des résultats non significatifs, le classement entre les modèles est presque le même.</p> <p>Réponse de Wisler, Burns et Iwamoto.</p> <p>Le niveau classe était de toute façon inutilisable, car le programme dure 4 ans pour un élève et que l'élève change de classe d'une année sur l'autre.</p>
<p>La variance inter-sites est plus forte que la variance entre les modèles. C'est LE résultat valide que l'on peut tirer de l'évaluation. Les effets des modèles pédagogiques sont faibles et le contexte local joue plus que le modèle pédagogique adopté.</p>	<p>Réponse de Anderson, St. Pierre, Proper, Stebbins.</p> <p>Ce n'est pas le seul résultat, mais c'est un résultat important car cela veut dire que des stratégies de remédiation coûteuses, mobilisant des esprits brillants, des équipes pédagogiques et des parents motivés, se révèlent incapables (à quelques exceptions) de changer vraiment la donne en matière de résultats scolaires pour les enfants pauvres.</p> <p>Réponse de Wisler, Burns et Iwamoto.</p> <p>Oui, les résultats sont en général décevants. Mais le modèle de l'instruction directe, même s'il n'a pas montré son efficacité sur tous les sites où il a été testé, s'est révélé efficace sur la majorité de ces sites.</p> <p>Réponse de Bereiter et Kurland.</p>

Les caractéristiques sociodémographiques ont une influence connue sur les résultats des élèves. Ce n'est pas parce que l'effet de ces caractéristiques est plus grand que celui de l'approche pédagogique adoptée que celle-ci n'a pas d'importance. L'important, c'est de savoir quelle part de la variance non expliquée par ces facteurs sociodémographiques peut être expliquée par l'approche pédagogique adoptée. La réponse est : entre 17 % et 55 %.

Pour l'apprentissage de la lecture, les études anglo-saxonnes mettent donc en évidence l'intérêt pour le déchiffrage d'un enseignement systématique de la relation phonème-graphème mais montrent aussi qu'il est possible de faire progresser les enfants en compréhension de texte en mobilisant une approche structurée et explicite. Le développement de cette compétence bénéficie également d'approches de type apprentissage coopératif entre pairs, visant à faciliter l'échange entre élèves de stratégies de compréhension et la discussion de leurs interprétations. Pour l'apprentissage des mathématiques, l'enseignement explicite et l'apprentissage coopératif semblent également faire preuve d'une efficacité supérieure à la moyenne. L'apport des technologies informatiques est moins consensuel, sauf pour l'arithmétique. Enfin, si les études rigoureuses consacrées à l'enseignement des sciences sont plus rares et contradictoires, elles laissent penser que des méthodes d'investigation ont un réel potentiel, à condition que les élèves bénéficient d'un guidage efficace de la part des enseignants.

3. Les études en contexte francophone

S'il existe en langue française de nombreux articles discutant *théoriquement* des avantages supposés de telle ou telle approche pédagogique, on trouve encore peu d'études empiriques, et très peu d'expériences contrôlées. Les choses sont néanmoins en train d'évoluer puisque les travaux d'évaluation publiés ces dernières années sont plus nombreux et mobilisent des méthodologies de plus en plus rigoureuses. Ces études, essentiellement consacrées à l'apprentissage de la lecture, viennent plutôt confirmer les résultats des recherches anglo-saxonnes. Nous en présentons quelques-unes, en terminant par celles qui nous semblent les plus robustes d'un point de vue méthodologique.

3.1. Efficacité comparée des méthodes d'apprentissage de la lecture en Belgique

En 1996, Braibant et Gérard ont publié une étude réalisée dans l'agglomération bruxelloise auprès de 450 élèves scolarisés en deuxième année d'école primaire dans 25 classes appartenant à 12 écoles. L'étude visait à mesurer l'efficacité comparée, en décodage et en compréhension, de trois méthodes distinctes : une méthode « gestuelle », forme de méthode synthétique mettant l'accent sur l'enseignement explicite du lien graphèmes-phonèmes ; une méthode « fonctionnelle », forme de méthode « globale » ou « analytique » qui privilégie les activités de compréhension en partant des phrases et de leur contexte pour isoler progressivement des éléments plus petits, et enfin une méthode mixte empruntant aux deux modèles précédents.

Malheureusement, cette étude n'est pas une expérience contrôlée et souffre de deux importantes lacunes méthodologiques : la collecte des informations relatives à la méthode utilisée repose sur un questionnaire adressé aux enseignants ; surtout, on ne dispose que d'une seule prise d'information sur le niveau des élèves, en milieu de deuxième année d'école primaire. Les résultats des élèves font néanmoins l'objet d'une analyse économétrique permettant d'isoler l'effet de la méthode utilisée de celui d'un certain nombre de variables relatives aux élèves (âge, sexe, nationalité, langue parlée à la maison, profession du père ou de la mère, retard scolaire), aux enseignants (âge et ancienneté), aux classes (nombre d'élèves et recrutement social classé en trois niveaux en fonction du pourcentage d'élèves étrangers et de l'indice socioéconomique moyen des élèves) et aux autres pratiques pédagogiques (nombre d'heures consacrées à la lecture, fréquence d'évaluation de la lecture à haute voix, utilisation par l'enseignant d'un guide méthodologique, utilisation quotidienne par les élèves d'un manuel scolaire).

Malgré ces lacunes importantes, l'étude conclut à la plus grande efficacité d'une méthode d'apprentissage centrée sur le décodage par rapport à une pédagogie « fonctionnelle » de la lecture. Les différences concernent tant le niveau de performance moyen que le nombre d'élèves présentant un retard en lecture ou que l'hétérogénéité des élèves au sein de la classe. Le modèle complet (avec toutes les variables) n'explique pas plus de 34 % de la variance en décodage et 36 % de la variance en compréhension. Les variables décrivant la méthode de lecture (méthode utilisée en première année d'école primaire, et pratiques pédagogiques mises en œuvre en deuxième année de primaire) expliquent ensemble 27 % de la variance des résultats en décodage et 26 % de la variance en compréhension. Les auteurs constatent par ailleurs qu'aucune méthode de lecture n'empêche les bons élèves d'apprendre à lire.

3.2. L'impact du manuel utilisé : une étude en zone d'éducation prioritaire

Deauvieu, Espinoza et Bruno (2013) ont cherché à mettre en évidence l'impact du manuel utilisé pour apprendre à lire en menant une enquête en 2013 auprès de 446 élèves scolarisés en CP en zone d'éducation prioritaire à Paris ou en petite couronne parisienne. Les manuels utilisés étaient identifiés et classés parmi deux catégories : i) méthode syllabique, c'est-à-dire centrée sur l'apprentissage systématique et progressif du code des correspondances graphophonologiques, et ii) méthode mixte, c'est-à-dire empruntant à des degrés divers à la fois à la méthode « globale » ou « analytique » et à la méthode « syllabique » ou « synthétique » (encadré 4).

Là encore, malheureusement, il ne s'agit pas d'une expérience contrôlée. Les manuels utilisés sont choisis par les enseignants (pas d'affectation aléatoire), il n'y a pas de groupe de contrôle à proprement parler, et l'on ne connaît pas le niveau des élèves en lecture à l'entrée en CP. La variable à expliquer est un indicateur synthétique de maîtrise de la langue écrite, construite à partir de quatre tests passés en fin de CP : une épreuve de déchiffrage (« test de fluence »), passée individuellement, et trois épreuves écrites individuelles passées collectivement, permettant de tester la compréhension, la dictée et la syntaxe. Les résultats font l'objet d'une analyse économétrique permettant de contrôler l'effet de plusieurs variables caractéristiques des élèves (origine culturelle des parents, pratiques de lecture autonome à la maison, aides des parents pour la lecture, sexe) et des enseignants (ancienneté dans le métier, adhésion au manuel utilisé).

Les chercheurs constatent que les élèves qui ont travaillé avec le manuel le plus fidèle à la méthode syllabique obtiennent, après prise en compte de l'impact des variables de contrôle, un score moyen supérieur de près de 19 points sur 100 à celui obtenu par les élèves qui ont utilisé le manuel mixte le plus proche de la méthode globale. À titre de comparaison, l'impact du niveau de diplôme des parents n'est que de 17 points. Autre résultat : plus le manuel utilisé est efficace, moins les acquisitions des élèves sont dépendantes de la famille. Par ailleurs, si l'on réduit l'analyse aux seuls 340 élèves dont le maître s'est déclaré satisfait du manuel qu'il utilise, la différence de résultats moyens entre les deux manuels les plus éloignés du point de vue de la méthode passe de 19 points à 25,5 points sur 100. Enfin, l'ouvrage qui insiste le plus sur la rigueur du déchiffrage est celui qui obtient le meilleur résultat au test de compréhension.

Ces résultats sont intéressants, mais les lacunes de la méthodologie employée (absence de pré-test alors que l'étude n'est pas une expérience contrôlée avec affectation aléatoire et information sur les pratiques recueillie uniquement *a posteriori*, par voie d'entretien) les rendent très vulnérables à d'éventuelles critiques.

Encadré 4

Méthode globale, mixte, syllabique...

De quoi parle-t-on ?

En matière de méthodes d'apprentissage de la lecture, la confusion est grande. Les experts eux-mêmes ne semblent pas toujours d'accord sur la définition qu'il convient de donner à telle ou telle méthode.

La première distinction que l'on peut faire tient à la place donnée à l'écrit et à l'oral dans

l'apprentissage. Dans l'approche idéovisuelle, promue par Jean Foucambert, on exclut tout déchiffrement et tout détour par l'oral. L'objectif est de développer les procédures de reconnaissance directe des mots par appariements entre une forme visuelle écrite et une représentation abstraite stockée dans un lexique mental. Cette méthode semble avoir totalement disparu.

La méthode globale comme la méthode syllabique font toutes deux partie d'approches que l'on peut qualifier de « grapho-phonologiques », en ce sens qu'elles s'appuient tôt (méthode syllabique) ou tard (méthode globale) sur une mise en correspondance des sons (phonèmes) et des signes qui les codent (graphèmes). Méthode globale et méthode syllabique se distinguent en ce que la méthode syllabique (ou synthétique) part des correspondances entre les lettres et les sons pour arriver au déchiffrement des mots alors que la méthode globale (ou analytique) tend à privilégier un travail à partir de phrases simples, dont on isole des mots. Dans les faits, la plupart des enseignants pratiquent un mélange de ces deux méthodes.

3.3. Les limites de l'approche idéovisuelle dans l'apprentissage de la lecture

L'approche didactique idéovisuelle, promue par Jean Foucambert, vise à développer les procédures de reconnaissance directe des mots par appariement entre une forme visuelle écrite et une représentation abstraite stockée dans un lexique mental. Elle rejette l'enseignement des procédures par assemblage qui impliquent la mise en correspondance systématique de phonèmes et de graphèmes par le biais de règles de conversion (voie phonologique).

Goigoux (2000) a réalisé une enquête permettant de suivre 76 élèves issus de 16 classes différentes localisées dans 11 écoles distinctes entre la troisième année de maternelle et le CE1. Parmi les 76 élèves de l'échantillon, 44 élèves ont suivi une approche purement idéovisuelle et 32 une approche partiellement phonique. Les élèves suivis ont été choisis de manière aléatoire dans les classes de grande section de maternelle associées aux cours préparatoires retenus pour l'étude.

Les pratiques didactiques réellement mises en œuvre par les maîtres ont été étudiées *via* une analyse des contenus d'enseignement en fonction du temps qui leur était consacré (questionnaire auto-administré auprès des enseignants) et une observation directe par l'enquêteur des séquences didactiques conduites en classe de CP. Les compétences des élèves en lecture ont été testées une première fois en grande section de maternelle (quatre épreuves : usages, mots, lettres et conceptualisation) puis à six reprises au cours du CP et du CE1 sur une épreuve d'identification de mots. Enfin, en CE2, l'évaluation nationale conduite par le ministère de l'Éducation nationale a permis d'évaluer la compréhension, la connaissance du code et la production de texte.

Alors qu'à l'issue de l'école maternelle, les élèves scolarisés dans les écoles « sans code » apparaissent plutôt plus à l'aise que leurs homologues scolarisés dans les écoles avec code, en CE2, les élèves du groupe « sans code » obtiennent des scores nettement inférieurs à ceux du groupe « avec code », dont les performances sont similaires à celles de la moyenne nationale. Les élèves issus de conditions sociales défavorisées sont lourdement pénalisés : ils obtiennent un score de réussite à peine supérieur à 50 % lors des évaluations de CE2 (44 points sur 87, soit 23 % de moins que la moyenne nationale établie à 57,3 points). La supériorité des sous-groupes « avec code » sur les sous-groupes « sans code »

apparaît au second trimestre du cours préparatoire et s'accroît par la suite. Ces résultats confirment donc que la découverte du principe alphabétique requiert un enseignement explicite et que l'habileté à identifier et à manipuler les composants phonologiques de la langue est décisive pour l'apprentissage de la lecture.

3.4. L'entraînement précoce à la compréhension

Si l'importance des compétences phonétiques ne fait plus de doute tant les recherches ont été nombreuses pour le confirmer, une bonne maîtrise de la langue maternelle passe aussi par la maîtrise d'autres compétences essentielles, à commencer par la compréhension. Celle-ci est aujourd'hui considérée comme une compétence qui n'est pas spécifique à la lecture. De nombreuses études ont montré qu'il était pourtant possible d'améliorer la compréhension de textes par des méthodes d'instruction spécifiques, focalisées sur les stratégies d'inférence, l'acquisition d'un vocabulaire plus étendu, l'organisation des informations, le résumé des informations, l'analyse de texte, les questions-réponses, etc. (voir encadré *supra*). La plupart des recherches concernent néanmoins des élèves d'âge relativement avancé.

Bianco *et al.* (2010 et 2012) ont cherché à savoir si un entraînement très précoce aux stratégies de compréhension pouvait aussi être efficace. Ils ont donc suivi pendant 3 ans un échantillon de 1 273 élèves âgés de 4 ans au début de l'enquête, scolarisés dans 44 écoles dont la moitié situées en ZEP. Ces enfants ont bénéficié en deuxième et troisième année de maternelle d'un programme parmi trois : un programme explicite focalisé sur la phonétique (PHO) ; un programme explicite focalisé sur la compréhension avec des leçons s'attachant à renforcer les différentes composantes de la compréhension : détection d'incohérences, inférences logiques, modélisation de la situation et structure de l'histoire (CS) ; un autre programme visant à développer la compréhension, mais de façon implicite, *via* la lecture et l'analyse d'histoires (SA).

Chaque école était assignée à un programme, et une seule classe de chaque école (avec l'enseignant habituel de cette classe) suivait le programme. Mais l'école pouvait choisir le type de programme qu'elle mettrait en œuvre (étude non randomisée). Tous les enseignants étaient expérimentés, volontaires et ont suivi une formation spécifique.

Trois groupes d'élèves ont donc été constitués suivant le type de programme suivi (phonétique, compréhension, analyse d'histoire), et chaque groupe était subdivisé en deux sous-groupe selon la durée d'exposition des élèves au programme : un semestre (en troisième année de maternelle) ou deux semestres (un en deuxième année et un en troisième année). Un dernier groupe ne recevait aucune formation (groupe témoin). Au sein des deux premiers groupes, trois sous-groupes étaient constitués selon le type de programme suivi. Les élèves ont d'abord été testés en début de deuxième année de maternelle. Puis ils ont été testés trois fois : en fin de deuxième année de maternelle, en fin de troisième année de maternelle, et en milieu d'année au CP.

À la fin de l'expérience, on constate que :

- l'impact du programme d'entraînement de type lecture et analyse d'histoires (SA) est non significatif tant en ce qui concerne la conscience phonologique que la compréhension, qu'elle soit orale ou écrite ;
- le programme d'entraînement à la phonétique (PHO), à condition qu'il soit suivi deux semestres, permet d'augmenter les performances en conscience phonologique et en reconnaissance des mots, tout particulièrement chez les

- élèves en difficulté, mais n'a pas d'impact durable sur la compréhension orale et écrite ;
- le programme explicite focalisé sur la compréhension orale (CS), lorsqu'il est suivi deux semestres, permet bien d'améliorer les performances des élèves en compréhension, orale, mais aussi écrite et semble aussi avoir un impact positif sur la reconnaissance des mots.

Les auteurs en concluent donc qu'il est possible d'améliorer la compréhension, à l'oral mais aussi à l'écrit, *via* un entraînement précoce (donc forcément sur un support oral en maternelle), reposant sur une approche explicite, à condition qu'il soit suivi suffisamment longtemps. De même, il est possible d'améliorer l'acquisition de la conscience phonologique et de la reconnaissance des mots par un entraînement précoce adapté, suffisamment long. Cette étude confirme aussi le fait que les performances en déchiffrage et en compréhension dépendent de compétences sous-jacentes relativement indépendantes.

3.5. Une expérience contrôlée de soutien scolaire pour l'apprentissage de la lecture

Goux, Gurgand et Maurin (2013) ont évalué un dispositif dénommé *Coup de pouce clé*, consistant à offrir à cinq élèves par classe de CP une activité de soutien pour l'apprentissage de la lecture, activité d'une heure et demie tous les soirs d'école et toute l'année. Les enfants étaient amenés à manipuler les mots de façon ludique et à apprécier la lecture d'histoires. Ce dispositif venait se substituer à des dispositifs existants proposés par l'Éducation nationale pour des élèves présentant des difficultés d'apprentissage de la lecture.

L'évaluation reposait sur une véritable expérience contrôlée, puisqu'il s'agissait de comparer les résultats des élèves des écoles tests et des écoles témoins, tirées au sort parmi 109 écoles des académies de Créteil, Lille et Versailles. Les élèves ont été évalués une première fois en fin de CP, à partir de cahiers de test élaborés par la Direction de l'évaluation, de la prospective et de la performance (DEPP) du ministère de l'Éducation nationale. Ces cahiers contenaient également des questions destinées à mesurer l'intérêt ou le goût des enfants pour la lecture. Les élèves étaient évalués une deuxième fois en fin de CE1 dans le cadre de l'évaluation nationale de CE1 réalisée cette année-là par le ministère de l'Éducation nationale.

Il ressort de cette évaluation que les élèves ayant bénéficié du dispositif *Coup de pouce clé* n'ont pas acquis de compétences en lecture différentes de celles des élèves du groupe témoin, bénéficiant des dispositifs de soutien habituels proposés l'Éducation nationale. Cependant les bénéficiaires du dispositif expriment un goût plus marqué pour la lecture, et généralement pour les matières scolaires.

3.6. Évaluation d'un programme expérimental d'enseignement structuré de la lecture

Le projet *Lecture* est un programme expérimental visant à réduire de moitié le nombre d'élèves en grande difficulté de lecture à l'issue du CE1. Il se situe dans la continuité du programme PARLER expérimenté de 2005 à 2008 dans huit classes du réseau d'éducation prioritaire de Grenoble et de son agglomération. Environ 200 élèves avaient été suivis de la grande section de maternelle au CE1. Le programme, visant à mettre en œuvre des pratiques pédagogiques validées par l'expérimentation (cf. *National Reading Panel*),

consistait en l'enseignement explicite et systématique de la conscience phonologique et du code alphabétique ainsi qu'un enseignement explicite de la compréhension et du vocabulaire de l'écrit oralisé. Il reposait également sur des activités périscolaires et une implication des familles à travers des rencontres en petits groupes. Ce programme pilote avait montré des effets très positifs en compréhension, orthographe, vocabulaire, fluence de lecture et QI (effet maximum en compréhension de l'écrit, évalué à 0,8 écart-type).

Le projet *Lecture* repose de même sur un accompagnement d'enseignants volontaires, de la grande section de maternelle au CE1, pour les aider à mettre en œuvre un enseignement structuré des compétences identifiées comme nécessaires à l'apprentissage de la lecture : conscience phonologique, compréhension orale et code alphabétique. Les conseillers pédagogiques de circonscription assurent le suivi, la formation et l'accompagnement des enseignants engagés dans le dispositif. La méthode promeut le travail des élèves en petits groupes de niveau homogène, chaque séance d'entraînement durant 30 minutes. La fréquence des entraînements est augmentée pour les élèves les plus faibles. L'évaluation de cette expérimentation a été confiée à la DEPP et à deux laboratoires universitaires¹².

L'expérimentation s'est déroulée d'une part en grande section de maternelle, au sein de 59 écoles test et de 59 écoles témoin tirées au sort (3 000 élèves dans les classes expérimentales, 3 000 dans les classes témoin), et d'autre part en CP, au sein de 28 écoles test (1 113 élèves dont 728 ayant déjà suivi le protocole expérimental en grande section) et de 78 écoles témoin tirées au sort. Deux modélisations ont été utilisées pour estimer les effets du dispositif : la première procède par régression linéaire (régression du score de fin d'année sur le score de début d'année, avec contrôle de l'effet du sexe et de l'âge), la seconde par appariement. L'effet est mesuré en pourcentage d'écart-type, comme dans la littérature anglo-saxonne.

En grande section, les deux méthodes d'estimation des effets (régression et appariement) donnent des résultats comparables : les élèves des classes expérimentales progressent davantage sur les dimensions lettres (+ 0,17 et + 0,16 écart-type respectivement), phonologie (+ 0,22 et + 0,25), lecture voie non lexicale¹³ (+ 0,41 dans les deux cas) et mathématiques (+ 0,23 et + 0,31) que les élèves du groupe témoin. Pour les dimensions vocabulaire et lecture voie lexicale, les effets ne sont pas significatifs. En compréhension, l'effet ne ressort vraiment que s'il est estimé par méthode d'appariement (+ 0,15 écart-type). Une analyse par décile de niveau des élèves montre qu'en phonologie, les effets sont plus importants pour les élèves les plus faibles et qu'en compréhension, ils ne sont significatifs que pour les élèves les plus faibles.

En CP, les résultats sont plus contrastés : les élèves du groupe expérimental, ayant suivi l'expérimentation tant en grande section qu'en CP, progressent davantage en phonologie (+ 0,29 et + 0,34), mais moins en compréhension orale (- 0,11 et - 0,14), les autres effets étant non significatifs.

L'effet global du dispositif sur deux années serait caractérisé par une progression significative en phonologie (+ 0,51 écart-type), en lecture voie non lexicale (+ 0,48 écart-type) avec des effets importants pour les élèves les plus faibles, mais par une absence d'effet sur les dimensions compréhension et lecture voie lexicale. On notera que les effets sont différents selon les dimensions cognitives, allant du non significatif au très significatif, ce

¹² Voir DEPP, Laboratoire EMC-Lyon 2, SODAR-IREDU (2014).

¹³ Il s'agit d'une épreuve de lecture de mots inventés, qui respectent les séquences de lettres ou de phonèmes habituelles de la langue. Cet entraînement favorise l'automatisme du déchiffrement.

qui montre que l'effet enregistré n'est pas un artefact lié au volontariat des enseignants du groupe expérimental. Par ailleurs, il est intéressant de noter que les effets sont plus importants pour les élèves les plus faibles, en particulier en phonologie. Il reste à savoir si les effets positifs constatés vont se transférer à des compétences de plus haut niveau en fin de CE1¹⁴

¹⁴ L'évaluation est en cours.

Conclusion

Plusieurs enseignements peuvent être tirés de la recension de cette littérature sur l'efficacité des méthodes ou approches pédagogiques.

Sur les méthodes elles-mêmes tout d'abord, il ressort que les approches structurées, explicites, font preuve d'une particulière efficacité, notamment pour l'acquisition des compétences de base, lorsque les élèves sont jeunes et/ou lorsqu'ils connaissent des difficultés. L'approche structurée n'est pas synonyme d'enseignement magistral. Pour que l'enseignement soit efficace, il faut aussi que les élèves ne soient pas passifs, qu'ils soient réellement engagés dans l'apprentissage. Cela peut être obtenu de différentes manières, mais l'une d'entre elles, l'apprentissage coopératif entre pairs, fait preuve d'une efficacité au-dessus de la moyenne. Guidés par les enseignants, les élèves sont ainsi amenés à clarifier, à formaliser et à expliciter leurs raisonnements.

Cela n'exclut pas d'emblée, pour d'autres compétences, pour des élèves d'âge plus avancé ou déjà suffisamment à l'aise, le recours à d'autres méthodes. Mais cela autorise néanmoins, pour les compétences de base et les publics évoqués *supra*, à recommander « en première intention » un enseignement explicite et structuré, accompagné de pratiques d'apprentissage coopératif.

Il ressort un autre résultat, potentiellement décevant, de ces études : l'ampleur des effets mis en évidence est souvent limitée, et ce d'autant plus que l'échantillon est important et que l'effet propre de l'expérimentation joue moins. La première raison en est que des facteurs plus puissants déterminent prioritairement les résultats des élèves, à commencer par leur niveau initial, leur environnement et leurs capacités. La deuxième raison, qui est rarement évoquée, est tout simplement qu'en moyenne les pratiques des enseignants sont déjà relativement efficaces. Dans ces études, on compare toujours plusieurs méthodes ou bien des approches innovantes à des approches « traditionnelles » ou « habituelles ». Les méthodes innovantes, on l'a dit, ne se distinguent pas toujours positivement des méthodes traditionnelles : l'innovation n'est pas une garantie d'efficacité. Mais même lorsqu'elles démontrent une efficacité particulière, c'est toujours par rapport à des pratiques qui ont-elles-mêmes une certaine efficacité.

Ceci nous amène à insister à nouveau sur un autre point, de vigilance cette fois. On a parlé jusqu'ici de méthode ou d'approche. Dans leurs pratiques, les enseignants empruntent à des méthodes diverses, soit en toute connaissance de cause, soit à la manière de Monsieur Jourdain. De nombreux enseignants pratiquent ainsi un enseignement systématique, explicite, structuré, sans pour autant se référer à une méthode particulière. D'aucuns estiment qu'il ne faudrait absolument pas parler de méthode, car le terme est trop flou, trop englobant ou laisse suggérer un formalisme qui est en fait absent. Utiliser le terme méthode nous permet simplement de regrouper un certain nombre de pratiques, mais on doit bien garder à l'esprit que les frontières entre méthodes ne sont pas toujours précises et que dans les faits, la pratique de chaque enseignant ne se résume évidemment pas à la stricte mise en œuvre d'une méthode.

Sur la méthodologie des études, on doit saluer les efforts réalisés par les équipes de

recherche. La qualité des études est en augmentation, le recours à des expériences contrôlées de plus en plus fréquent. Si l'expérience contrôlée permet de s'assurer de la validité interne de l'étude (comparabilité du groupe expérimental et du groupe traitement, traitement des biais de sélection), la question de la validité externe ne doit pas être laissée de côté. Une méthode qui démontre une particulière efficacité en contexte expérimental, avec des enseignants volontaires, peut se révéler difficilement appropriable ou susciter un faible niveau d'adhésion de la part du corps enseignant. De ce point de vue, les méthodes structurées et explicites pourraient avoir un certain avantage, car elles sont sans doute plus aisément compréhensibles et appropriables que d'autres méthodes plus difficiles à cerner.

Bibliographie

Anderson R. B., St. Pierre R. G., Proper E. C. et Stebbins L. B. (1978), "Pardon Us, But What Was the Question Again? A Response to the Critique of the Follow Through Evaluation", *Harvard Educational Review*, vol. 48, n° 2, mai.

Barron B. et Darling-Hammond L. (2008), *Teaching for meaningful learning: A review of research on inquiry-based and cooperative learning*, George Lucas Educational Foundation.

Bereiter C. et Kurland M. (1981), "A constructive look at Follow Through results", *Interchange*, vol. 12, issue 1.

Bianco M. (2003), « Apprendre à comprendre : l'entraînement à l'utilisation des marques linguistiques », in Gaonac'h D. et Fayol M. (dir.), *Aider les élèves à comprendre, du texte au multimédia*, Paris, Hachette Education.

Bianco M. et Bressoux P. (2009), « Effet-classe et effet-maître dans l'enseignement primaire : vers un enseignement efficace de la compréhension ? », in Dumay X. et Dupriez V. (dir.), *L'efficacité dans l'enseignement. Promesses et zones d'ombre*, Bruxelles, de Boeck.

Bianco M., Bressoux P., Doyen A.-L., Lambert E., Lima L., Pellenq C. et Zorman M. (2010), « Early training in oral comprehension and phonological skills : results of a three-year longitudinal study », *Scientific Studies of Reading*, 14 :3, p. 211-246.

Bianco M., Pellenq C., Lambert E., Bressoux P., Lima L. et Doyer A.-L. (2012), "Impact of early code-skill and oral-comprehension on reading achievement in first grade", *Journal of Research in Reading*, vol. 35 (4), p. 427-455.

Bissonnette S., Richard M. et Gauthier C. (2005), « Interventions pédagogiques efficaces et réussite scolaire des élèves provenant de milieux défavorisés », *Revue française de pédagogie*, n° 150, p. 87-141.

Blanchard M. R., Southerland S. A., Osborne J. W., Sampson V. D., Annetta L. A. et Granger E. M. (2010), "Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction", *Science Education*, vol. 94, issue 4, p. 577-616.

Braibant J.-M. et Gérard F. M. (1996), « Savoir lire : une question de méthodes ? », *Bulletin de psychologie scolaire et d'orientation*, 1, 1996, p. 7-45.

Bruder R. et Prescott A. (2013), "Research evidence on the benefits of IBL", *ZDM Mathematics Education* vol. 45, p. 811-822.

Carette V. (2008), « Les caractéristiques des enseignants efficaces en question », *Revue française de pédagogie*, n° 162, p. 81-93.

Deauvieu J., Espinoza O. et Bruno A.-M. (2013), *Lecture au CP : un effet-manuel considérable. Rapport de recherche*, Laboratoire Printemps, université Versailles Saint-Quentin-en-Yvelines.

DEPP, Laboratoire EMC-Lyon 2, SODAR-IREDU (2014), « *Projet lecture* ». *Rapport d'évaluation*, Fonds d'expérimentation pour la jeunesse.

Gersten R. et Clarke B. J. (2007), "Effective strategies for teaching students with difficulties in mathematics", *Research Brief*, National council of teachers of Mathematics.

Goigoux R. (2000), « Apprendre à lire à l'école : les limites d'une approche idéovisuelle », *Psychologie française*, n° 45-3, p. 233-243.

Goux D., Gurgand M. et Maurin E. (2013), *Évaluation d'impact du dispositif Coup de Pouce clé*, rapport final pour le Fonds d'expérimentation pour la jeunesse.

Hattie J. (2008), *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*, Londres, Routledge.

House E. R., Glass G. V., McLean L. D. et Walker D. F. (1978), "No simple answer: critique of the Follow Through Evaluation", *Harvard Educational Review*, vol. 48, n° 2, mai.

Ketele J.-M. (2009), « Avancées et limites des recherches sur l'efficacité », in Dumay X. et Dupriez V. (dir.), *L'efficacité dans l'enseignement. Promesses et zones d'ombre*, Bruxelles, de Boeck.

Kirschner P. A., Sweller J. et Clark R. E. (2006), "Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching", *Educational Psychologist*, 41(2), p. 75-86.

Lima L., Sylvestre E. et Bianco M. (2006), « Lectures partagées et acquisition de stratégies de compréhension au cycle 3 », in Dessus P et Gentaz E (eds), *Apprentissage et enseignement* (p. 25-42), Paris, Dunod.

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*, US Department of Health and Human Services.

Nye B., Hedges L. V. et Konstantopoulos S. (2001), "Are the effects of small classes cumulative? Evidence from a Tennessee experiment", *The Journal of Educational Research*, 94, p. 336-345.

Rosenshine B. et Meister C. (1994), « Reciprocal teaching: a review of the research », *Review of Educational Research*, 66(2), p. 181-221.

Slavin R. E. et Lake C., (2007), *Effective Programs in Elementary Mathematics: a Best-Evidence Synthesis*, Best evidence encyclopedia.

Slavin R. E., Lake C. et Groff C. (2008), *Effective Programs in Middle and High School Mathematics: a Best-Evidence Synthesis*, Best evidence encyclopedia.

Slavin R. E., Lake C., Davis S. et Madden N. A. (2009), *Effective Programs for Struggling Readers : A Best-Evidence Synthesis*, Baltimore, MD: Johns Hopkins University, Center for Data-Driven Reform in Education, p. 52.

Wisler C. E., Burns G. P et Iwamoto D. (1978), "Follow Through Redux: A response to the critique by House, Glass, McLean and Walker", *Harvard Educational Review*, vol. 48, n° 2, mai.



FRANCE STRATÉGIE

Les documents de travail de France Stratégie sont des études ou des travaux de recherche qui n'engagent que leurs auteurs et ne reflètent pas nécessairement les positions du Commissariat général à la stratégie et à la prospective. L'objet de leur diffusion est de susciter le débat et d'appeler commentaires et critiques.

Directeur de la publication :
Jean Pisani-Ferry,
Commissaire général

Directeur de la rédaction :
Selma Mahfouz,
Commissaire générale adjointe

Dépôt légal : septembre 2014
www.strategie.gouv.fr
