

Élèves et classes apprennent à lire

Reprise des données d'une recherche en lecture
selon une approche multiniveau

Jean Cardinet
Martine Wirthner



Élèves et classes apprennent à lire

Reprise des données d'une recherche en lecture
selon une approche multiniveau

Jean Cardinet
Martine Wirthner

IRDP
Faubourg de l'Hôpital 43
Case postale 556
CH-2002 Neuchâtel

Tél. ++41 (0) 32 889 86 18
Fax ++41 (0) 32 889 69 71

E-mail documentation@irdp.ch
<http://www.irdp.ch>

Fiche bibliographique :

Cardinet, Jean & Wirthner, Martine. - Elèves et classes apprennent à lire : reprise des données d'une recherche en lecture selon une approche multiniveau / Jean Cardinet et Martine Wirthner. - Neuchâtel : Institut de recherche et de documentation pédagogique (IRD), 2010. - 73 p. ; 30 cm. - (10.3). - Bibliogr. p. 59-61

CHF 11.--

Euro 8.--

Mots-clés: Apprentissage de la lecture, Compréhension en lecture, Enseignement primaire, Neuchâtel, Suisse romande, Analyse de variance, Analyse factorielle, Traitement des données, Résultat de recherche, Test, Epistémologie, Méthode de lecture, Analyse comparative, Prérequis, Vitesse de lecture, Intérêt de lecture, Motivation, Influence, Comportement de l'enseignant, Style pédagogique, Réussite scolaire, Elève, Classe

Cette publication est également disponible sur le site IRDP:

<http://www.irdp.ch/>

La reproduction totale ou partielle des publications de l'IRD est en principe autorisée, à condition que leur(s) auteur(s) en ai(en)t été informé(s) au préalable et que les références soient mentionnées.

RÉSUMÉ

Pourquoi se pencher à nouveau sur des données datant de 1973 et touchant à l'enseignement / apprentissage de la lecture ? Essentiellement parce qu'elles peuvent être réexaminées à l'aide d'un outil statistique nouveau et plus fin que celui utilisé à l'époque, une analyse multiniveau permettant la mise en interaction des variables individuelles et de celles de la classe. En effet, cette analyse parvient à estimer les sources de la variance vraie du critère, ici la réussite en lecture, et tente d'attribuer à divers facteurs la part qui leur revient, d'abord au niveau des individus, puis à celui des groupes. Ainsi, de nouveaux résultats apparaissent, analysés ensuite à la lumière des connaissances actuelles en didactique de la lecture. Risque-t-on alors de voir se rallumer une guerre des méthodes de lecture pour apprendre à lire ? En aucun cas. En réalité, l'analyse statistique montre un grand nombre d'interactions et prouve que la bonne méthode n'est pas la même selon le type d'élève ou selon le contexte de la classe.

Les variables majeures mises en évidence par l'analyse statistique en raison de leur influence sur les capacités de lecture des élèves sont discutées d'un point de vue didactique ; que disent les chercheurs à leur propos ? Quelle actualité prennent-elles dans l'école en Suisse romande aujourd'hui ? Ces variables sont les suivantes : le décodage, l'apprentissage des lettres et l'approche écrite ; la motivation interne, les objectifs affectifs ; la vitesse d'avancement de l'enseignement/apprentissage. Cette mise en perspective se prolonge par un regard sur les principales recherches processus-produit menées dès les années 1970, du même type finalement que celle de l'IRDP de 1973. Là, certaines interactions entre variables pédagogiques et personnelles des élèves sont prises en considération, en lien avec la question de l'efficacité de l'enseignement. Si des pistes telles que la différenciation de l'enseignement sont citées, la question de l'efficacité de l'enseignement de la lecture reste largement ouverte aujourd'hui encore.

Enfin, une discussion méthodologique termine le document, sous la forme d'un plaidoyer en faveur d'une « alliance » entre recherches quantitatives et qualitatives, vues comme complémentaires et favorables à un travail en équipe. Il ressort en effet en conclusion que l'affinement des outils d'analyse devient de plus en plus nécessaire pour répondre aux questions complexes que pose l'école actuelle.

TABLE DES MATIÈRES

Origine de l'étude	7
Méthode d'analyse multiniveau	9
Préparation de l'étude	15
Résultats de l'étude	21
Discussion des résultats: Apports de la recherche quantitative	39
Discussion des résultats à la lumière de recherches récentes sur la lecture	41
Perspectives méthodologiques	55
Annexes	
Annexe A: effets des divers types d'enseignement	66
Annexe B: réussites moyennes de 4 méthodes de lecture	68
Annexe C: exemple de rapport imprimé par le logiciel HLM	69
Annexe D: échelles des variables utilisées	73

ORIGINE DE L'ÉTUDE

La recherche de 1973

Rarement une étude a mobilisé autant d'enseignant-e-s et d'élèves en Suisse romande, et aussi intensément, puisque environ 2400 élèves et 140 classes (pour l'ensemble du canton de Neuchâtel) ont été suivis pendant deux ans, par rapport à plus de 126 variables par élève et 160 par enseignant-e.

Un livre avait été publié dans l'urgence en 1976, pour répondre aux questions pratiques des responsables de la Coordination Scolaire romande qui avaient demandé l'étude: (Cardinet et Weiss, 1976). La qualité de l'apprentissage se révélait dépendre moins des méthodes de lecture utilisées que de l'expérience des maîtresses et de leur capacité à enrichir leur enseignement (motiver leurs élèves et répondre à leurs besoins personnels par une approche diversifiée).

Une partie des données recueillies n'avait pourtant pas pu être analysée, faute de temps et faute de moyens de calcul adéquats à l'époque, et avait été simplement mise de côté en vue d'un réexamen futur.

Les analyses d'André Parisod

L'abondance des informations engrangées par les chercheurs était motivée par le désir de découvrir d'éventuelles synergies entre les traits psychologiques personnels des élèves et les options pédagogiques de leurs enseignant-e-s. Aux États-Unis, Lee Cronbach venait de proposer ces « interactions aptitudes-traitements » comme le fondement scientifique de l'individualisation de l'enseignement, objectif lointain des réformes pédagogiques. On pouvait espérer appliquer un jour à chaque élève la méthode qui s'adapterait le mieux à ses aptitudes.

C'est dans cette perspective que, durant les années 1980, la banque de données neuchâteloise a été retraitée par André Parisod, alors assistant et doctorant à l'Université de Lausanne.

Mais à l'époque, ni les modèles statistiques, ni les logiciels informatiques n'avaient encore été développés pour de telles recherches. André Parisod put dégager empiriquement certaines interactions, grâce à une utilisation intensive de l'ordinateur de l'EPFL, mais il n'a pas pu mener ses travaux jusqu'à une publication (ses résultats sont présentés à l'annexe A, en honneur à sa mémoire).

MÉTHODE D'ANALYSE MULTINIVEAU

Une nouvelle approche

C'est seulement vers la fin des années 1980 que des modèles appelés « hiérarchiques », ou « mixtes », ou « multiniveau » ont vu le jour. Ils ont commencé à être appliqués dans les années 1990.

Leur but est de réunir dans un même cadre théorique des analyses qui ne pouvaient auparavant être réalisées que séparément. Par exemple, au moment de la recherche neuchâteloise, on avait étudié la valeur prédictive des tests individuels d'aptitude scolaire, mais on devait alors, pour étudier cette validité, faire abstraction des caractéristiques des classes et des enseignantes. En fait, on avait surtout cherché à étudier l'effet des méthodes pédagogiques, mais en considérant alors uniquement les moyennes des classes, et en laissant de côté les différences entre élèves dans les classes. On ne pouvait pas étudier en même temps l'effet des variables de niveau 1, caractérisant les élèves, et celui des variables de niveau 2, caractérisant les classes. En conséquence, du fait du recouvrement de ces deux types d'influence, on ne pouvait pas comparer leur importance relative.

Travailler séparément ainsi sur les résultats des élèves et des classes peut rendre très difficile leur interprétation. Supposons, par exemple, que toute une population scolaire ait passé une même épreuve objective de compréhension de lecture. On peut s'attendre à une corrélation positive entre les scores des élèves à cette épreuve et leurs notes en français. Mais si l'on calcule la corrélation sur les moyennes de classes pour ces mêmes deux variables, elle pourra être au contraire négative, parce que les enseignant-e-s exigeants notent plus sévèrement leurs élèves, alors même qu'ils sont meilleurs comparativement à la population d'ensemble. Par conséquent, il faut pouvoir traiter des deux niveaux dans un modèle conjoint, pour évaluer correctement la part de variance qui provient de chacun d'eux.

Son principe

On peut résumer la méthode multiniveau en disant qu'elle cherche à estimer les sources de la variance vraie du critère, ici la réussite en lecture, et tente d'attribuer à divers facteurs la part qui leur revient, d'abord au niveau des individus, puis à celui des groupes.

Au niveau 1, celui des élèves, elle rend compte de leur réussite individuelle en lecture à l'aide d'équations de régression intraclasses pondérant au mieux les prédicteurs individuels (variables identitaires ou tests psychologiques). Ce faisant, elle réduit considérablement la variance incontrôlée dans les résultats des classes, celle qui est liée à l'échantillonnage aléatoire des élèves, en soustrayant de ces résultats toute la variance prédictible. Seule la variance résiduelle des scores des élèves au critère reste ensuite prise en compte. Cette première étape des analyses multiniveau rend plus puissants tous les tests statistiques ultérieurs relatifs aux classes, en réduisant la grandeur de ce qui est pris comme étalon d'erreur.

Au niveau 2, celui des classes, la méthode multiniveau estime d'abord par régression linéaire la moyenne de classe attendue au critère, à partir de toutes les informations disponibles dans l'espace des classes : moyennes aux variables individuelles de niveau 1, caractéristiques de la classe (effectif, niveau socioculturel moyen, etc.), caractéristiques de la pédagogie de l'enseignant-e, contexte scolaire, résultats antérieurs, etc.

La méthode multiniveau traite ensuite de façon parallèle le coefficient attribué dans chaque classe à chaque variable individuelle (pour prédire au mieux les résultats). Ces coefficients de pente deviennent eux aussi chacun une fonction multilinéaire de toutes les variables connues dans l'espace des classes.

Mais à la différence des modèles de régression classiques antérieurs, la méthode multiniveau fournit une estimation de la variance vraie des moyennes de classes (ou des pentes des prédicteurs) qui est purifiée de l'effet de l'échantillonnage des élèves (une classe d'effectif faible, par exemple, aura ainsi moins de poids qu'une classe nombreuse dans l'estimation de cette variance vraie entre classes). De ce fait, les tests statistiques pour dire s'il existe des différences significatives entre moyennes de classes, ou entre coefficients de pondération, seront plus fiables, parce que non biaisés.

Les coefficients estimés à l'aide des variables de niveau 2 (que ce soient la moyenne ou les pentes valables pour une classe donnée), servent à estimer l'équation de régression qui est applicable, dans cette classe, aux variables individuelles de niveau 1. On voit donc que le problème de mettre en relation des variables individuelles et des variables de groupe est résolu par l'approche multiniveau. Une même équation permet de calculer de combien augmentera le critère pour une augmentation d'une unité d'un prédicteur, qu'il soit de niveau 1 ou 2, et donc de dire que tel prédicteur de niveau 1 a une importance double, par exemple, de celle de tel prédicteur de niveau 2, parce que le critère bénéficiera avec lui d'un effet double.

Du coup, on peut quantifier et comparer les variances attribuables aux caractéristiques individuelles des élèves et les variances dépendant plutôt des classes auxquelles ces derniers appartiennent. Ceci permet de déterminer des effets de contexte, par exemple des bonus moyens de classe, ou valeurs ajoutées pédagogiques, ou apports de la didactique, indépendants de la composition du groupe d'élèves, de leurs connaissances de départ, de leur niveau socioculturel, etc. Parallèlement, le poids donné à ces dernières caractéristiques (individuelles) dans la fabrication de la réussite des élèves dans chaque classe peut être analysé et rattaché aux caractéristiques des classes et de leurs enseignant-e-s. On peut voir ainsi ce qui différencie les classes qui sont plus favorables à l'apprentissage des garçons, ou qui compensent mieux les obstacles à l'apprentissage dus à l'origine sociale ou à la langue maternelle des élèves.

La méthode multiniveau permet enfin de comparer le degré d'adéquation aux données de deux modèles successifs et donc de progresser pas à pas dans la construction d'un modèle descriptif optimum. Elle réduit ainsi l'arbitraire d'une trop grande indétermination des représentations possibles de la réalité, qui avait gêné l'utilisation antérieure des modèles linéaires.

Ses étapes

On comprendra mieux la méthode multiniveau si l'on peut suivre son utilisation dans un cas réel. Voyons donc comment on l'a appliquée lors d'une première analyse de la banque de données sur l'apprentissage de la lecture dans le canton de Neuchâtel.

Il s'agissait, évidemment, d'étudier les déterminants de la compréhension de lecture. Mais nous soulignons que cet exemple ne va servir qu'à illustrer les étapes de la méthode multiniveau, pour faire comprendre moins abstraitement comment elle procède et ce qu'elle peut apporter. Les résultats de la recherche principale seront présentés plus loin. Nous n'expliquerons donc pas ici les variables utilisées et nous ne citerons que quelques résultats au passage à titre d'exemples.

Analyse de variance simple des différences entre classes au critère

Le modèle, sous sa forme la plus simple, se ramène à une analyse de la variance des différences entre classes. Le critère (un test collectif de compréhension de texte narratif, passé en fin de première année) est à disposition pour environ 1700 élèves appartenant à environ 90 classes. Ces scores ont été standardisés sur la population totale. Ils sont donc de moyenne égale à 0 et de variance égale à 1. On peut sur cette base calculer la variance intraclasse de ces scores au critère. Elle est nécessairement inférieure à 1. Elle vaut 0,82945. La variance interclasse vaut 0,17456, soit approximativement le complément à 1.

On voit ainsi que 17 % de la variance de la réussite en lecture est attribuable aux différences entre classes (c'est certainement beaucoup, mais en fin de scolarité, l'hétérogénéité est encore nettement plus grande). Notons aussi que l'analyse de la variance estime des parts de variance vraie, épurées de l'effet de l'échantillonnage aléatoire des élèves.

Le modèle multiniveau va permettre de déterminer la part que différents facteurs jouent dans ces résultats de fin d'année.

Soustraction de l'effet de la maturité scolaire personnelle des enfants entrant à l'école

On peut demander au logiciel utilisé (HLM) d'estimer ces deux variances (interclasse et intraclasse) en retirant du critère l'effet prédictible du niveau d'aptitude initial des élèves (mesuré par une série de tests psychologiques, résumés par leur moyenne, appelée Niveau de départ). La réussite probable est estimée pour chaque enfant à l'aide d'une droite de régression linéaire intraclasse, centrée sur la moyenne de chaque classe au prédicteur et au critère, avec d'abord une pente identique pour toutes les classes. Ce pronostic pour chaque élève est ensuite soustrait de sa réussite observée. La variance intraclasse résiduelle au critère se trouve alors ramenée à 0,59305, soit une baisse de 29 %, ce qui démontre l'effet majeur de ce facteur personnel dans l'apprentissage de la lecture, en tout cas dans la différenciation des résultats à l'intérieur d'une classe.

La variance interclasse est peu affectée. Elle est encore de 0,18986. Elle est même un peu plus grande qu'avant cette correction, mais il s'agit sans doute d'un effet aléatoire, parce qu'il n'y a pas de raison que la soustraction de valeurs centrées sur la moyenne de chaque classe modifie ces moyennes, ni donc leur variance.

Soustraction de l'effet de toutes les caractéristiques personnelles des élèves

Au lieu d'utiliser une seule variable prédictrice, le niveau de départ, on peut établir un pronostic à partir de toutes les informations personnelles dont on dispose, soit dans cet exemple, en plus de l'aptitude scolaire : le genre, le milieu socioprofessionnel d'origine, la facilité d'expression en français, l'absence de redoublement, et la connaissance préalable de la lecture. Variances et covariances intra-groupe servent de base au calcul d'un hyperplan de régression.

Les résultats confirment ce que l'on pouvait attendre, soit un effet positif des facteurs énumérés ci-dessus. Tous les poids Bêta se révèlent significatifs, à l'exclusion du milieu social, mais sont de valeur faible, comprise entre 0,04 et 0,08, de sorte que l'amélioration de la prédiction reste très faible. La variance résiduelle du critère descend seulement de 0,59305 à 0,58121. Ainsi, les caractéristiques personnelles des élèves, en dehors de leur niveau de maturité scolaire, n'expliquent que 1 % de la variance des résultats, du fait que ces différents facteurs se recouvrent et que leur influence se manifestait déjà dans les tests de maturité scolaire exploités à l'étape précédente.

De son côté, l'estimation de la variance des moyennes de classes n'est pas non plus modifiée sensiblement par cette augmentation du nombre de prédicteurs. Sa valeur passe de 0,18986 à 0,19064.

Cette stabilité était à prévoir puisqu'on n'a introduit que des variables de niveau 1, centrées sur les moyennes de classe.

On peut en conclure à l'existence, pour les variables de niveau 1, d'un facteur général de maturité scolaire, ce qui n'est pas contradictoire avec l'effet de nombreux facteurs psychologiques et socioculturels, si ceux-ci se retrouvent englobés dans cet effet d'ensemble. Il explique 25 % ($0,82945 - 0,58121 = 0,24824$) de la variance totale de la réussite en lecture et 30 % ($0,24824 / 0,82945 = 0,29928$) de la variance intraclasse du même critère.

Estimation des moyennes de classes attendues au critère à partir des variables de niveau 2

Dans une étude multiniveau, il est nécessaire de construire progressivement un modèle décrivant au mieux la situation, mais un grand nombre de chemins sont possibles pour y parvenir. On aurait pu exploiter davantage les variables de niveau 1 qui viennent d'être introduites, en ajustant au mieux leurs coefficients Bêta classe par classe. Nous le ferons au point suivant. Pour l'instant, une autre approche va être préférée. Il s'agit de compléter l'équation de régression pour chaque classe en estimant les moyennes attendues au critère, ce qui fixera la droite, ou l'hyperplan, de régression pour chaque classe.

Pour cela, on peut demander au logiciel de prédire au mieux le résultat moyen de chaque classe à partir de toutes les variables de niveau 2 qui la caractérisent. On disposait, dans la banque de données, de 28 variables de ce type, dont les 6 moyennes de variables de niveau 1, exprimées cette fois en écarts à la moyenne générale.

Très peu de ces 28 variables se sont révélées significatives : le nombre de redoublants dans la classe (Bêta = - 0,37), le niveau moyen de départ des élèves (Bêta = 0,53), et la rapidité d'avancement dans le programme (Bêta = 0,16). Deux méthodes d'enseignement particulières ont des effets plutôt négatifs (- 0,05 et - 0,11).

Là encore le facteur général de maturité scolaire des élèves semble intervenir de façon décisive, par l'effet du niveau moyen de la classe, effet renforcé par l'absence de redoublants. Mais les enseignants désireux de recevoir des conseils sur la bonne façon de conduire leur classe ont des raisons d'être déçus : seule la variable didactique de rapidité d'avancement répond à leur désir (confirmée par les mauvais résultats de méthodes s'attardant trop aux préliminaires). Encore faut-il remarquer que le niveau de départ de la classe a environ trois fois plus d'effet que la meilleure variable didactique. Les enseignants ne doivent donc pas se faire d'illusion : leur pédagogie ne peut pas faire de miracles dans certaines classes...

Pourtant, l'introduction dans le modèle multiniveau de ces variables de niveau 2 change radicalement l'estimation de la variance interclasse, qui passe de 0,19064 à 0,10627, une diminution de 46 %. On voit que les variables utilisées pour prédire les moyennes de classe, celles donc que l'on connaît et sur lesquelles on pourrait agir, expliquent près de la moitié de la variance interclasse, ce qui semble encourageant.

Le malheur, pour les éducateurs généreux qui rêvent d'une méthode infaillible, c'est que cette moitié de la variance interclasse qu'ils pourraient maîtriser ne représente pas plus que 8 % de la variance totale du critère ! L'essentiel dépend de l'élève lui-même.

Ce genre de résultat n'aurait pas pu être obtenu par les méthodes existant lors de l'étude initiale. Il montre certes l'influence possible du contexte qui entoure l'enfant sur sa réussite scolaire, mais cette quantification montre en même temps les limites de ces influences. Si l'on a pu maîtriser dans cette étude la moitié de la variance interclasse, on est cependant loin d'avoir maîtrisé le dixième des facteurs de réussite de l'apprentissage. L'influence des méthodes d'enseignement de la lecture dans cet

ensemble n'est, au plus, que de quelques pourcents. Les querelles que ces méthodes ont soulevées se révèlent ainsi bien dérisoires...

A la fin de cette étape de la recherche, la variance intraclasse, au contraire de celle qu'on vient d'examiner, n'a pas pu changer, puisqu'on a travaillé avec des variables collectives de niveau 2. Elle reste égale à 0,58060, contre 0,58121 à l'étape précédente.

Adaptation des coefficients Bêta appliqués aux variables de niveau 1.

Au lieu d'estimer les moyennes attendues pour chaque classe, on aurait pu chercher à adapter pour chaque classe les coefficients multiplicateurs à appliquer aux variables individuelles. Il faut, pour réaliser cette adaptation, ajouter ou soustraire à la pente commune (calculée plus haut pour chaque variable) une constante de pente propre à la classe. Comme pour les moyennes de classes, cette estimation de pente est faite par une équation de régression pouvant utiliser toutes les variables disponibles au niveau 2. Ce calcul est effectué par le logiciel HLM, qui donne ensuite les variances et covariances résiduelles de ces coefficients (comme il fournissait à l'étape précédente la variance intergroupe résiduelle).

Au départ, l'existence de variances résiduelles de pente significatives est logiquement nécessaire, si l'on veut rendre compte de ces différences de pente à partir de variables de niveau 2, de même qu'il faut des différences significatives entre les moyennes des classes pour qu'on puisse chercher à en rendre compte.

Dans le cas présent, ces variances résiduelles sont significatives pour 3 des 6 variables individuelles : facilité d'expression, scolarité régulière, et niveau de maturité scolaire. C'est donc pour adapter les pentes de ces variables qu'il faudra chercher des variables de niveau 2.

Le fait d'ajuster les pentes pour chaque classe a déjà un effet sur la variance intragroupe. La prédiction devenant meilleure dans chaque classe, la variance résiduelle doit diminuer. Elle passe de 0,58060 à 0,54448. Puisque cette modification affecte des variables de niveau 1, elle n'a pas de raison, par contre, de modifier la variance intergroupe et n'a donc pas de portée pédagogique directe.

Recherche d'interactions entre variables de niveau 1 et de niveau 2

On peut montrer, en écrivant l'équation d'ensemble du modèle, qui combine les relations linéaires des variables des deux niveaux, qu'elle comporte des produits d'une variable individuelle par une variable collective. Comme on travaille en scores centrés sur zéro, ces produits seront positifs à condition que les deux variables soient positives, ou les deux négatives. Ils seront négatifs si les deux variables du produit sont de signe contraire. C'est l'origine d'effets non linéaires, dits d'interaction, parce que l'effet d'une variable est modulé par la valeur de l'autre variable.

La méthode multiniveau permet de déterminer si certaines de ces interactions sont significatives. Dans de tels cas, il faudra, pour prédire l'effet d'une variable individuelle sur la réussite d'un élève, moduler la pente moyenne attribuée à cette variable individuelle en fonction de la valeur de la variable de niveau 2, c'est-à-dire en fonction de la caractéristique de la classe à laquelle l'élève appartient.

Par exemple, le score en expression orale est en moyenne affecté du coefficient positif 0,06 pour prédire la réussite en lecture. Mais comme une interaction significative a été trouvée de cette variable individuelle d'expression avec la proportion de filles dans la classe, affectée du coefficient $-0,24$, il faudra ajouter au coefficient 0,06 le produit de cette proportion de filles par $(-0,24)$. C'est le nouveau coefficient total qui multipliera le score en expression orale pour prédire la réussite de cet élève (parmi d'autres prédicteurs). Ainsi, une expression orale facile ne contribuera à établir un pronostic favorable que s'il s'agit d'une classe où les garçons sont majoritaires.

Tenir compte des interactions revient à adapter les pentes des variables individuelles en fonction de ce que l'on sait de la classe de l'élève. Par exemple, dans une classe utilisant une méthode de lecture traditionnelle, centrée sur le déchiffrage des lettres, le handicap relatif des enfants non francophones est compensé, alors qu'à l'inverse des méthodes de lecture centrées sur l'analyse de la langue orale augmentent les difficultés des enfants étrangers. (Le coefficient pour l'interaction entre la centration sur l'écrit et la facilité d'expression en français est de $-0,07$. Il compense donc le coefficient de $0,06$ exprimant l'avantage des enfants francophones. On voit cependant que ces effets sont minimes.)

La prise en compte des variables de niveau 2 justifiant des adaptations de pente doit réduire la variance résiduelle de pente de la variable de niveau 1 considérée et finalement la rendre non significative si toutes les interactions sont prises en compte.

Dans l'exemple que l'on suit dans ces pages, la facilité d'expression avait une variance résiduelle de pente au départ (c'est-à-dire sans prise en compte de variables en interaction) égale à $0,00434$. La recherche d'interactions avec des variables de niveau 2 a montré que trois d'entre elles étaient significatives: l'interaction avec (i) la proportion de filles dans la classe, (ii) le taux d'élèves réguliers (non redoublants) et (iii) le degré de centration sur l'écrit de la méthode de lecture suivie. Après la prise en compte de tous les effets d'interaction, la variance résiduelle de pente n'était plus que de $0,00007$ et elle n'était plus significative à $0,05$.

On voit ainsi comment se construit progressivement la définition du modèle rendant compte de la réussite en lecture selon les caractéristiques des élèves et des classes.

Corrélations entre composantes résiduelles

La méthode multiniveau apporte encore un autre type d'information original. On a vu qu'elle calculait pour chaque classe sa moyenne attendue, puis après soustraction de ce pronostic de la moyenne effective, elle donnait une moyenne résiduelle. Les calculs sont analogues pour les coefficients de pente. En retirant la pente prédictible de la pente effective, on obtient une composante résiduelle de pente, pour chaque variable dans chaque classe.

On obtient donc des valeurs résiduelles pour toutes les variables de niveau 2, que l'on peut mettre en corrélation dans l'espace des classes. Une corrélation élevée entre deux variables de niveau 2 résiduelles signifie que si l'une est élevée dans certaines classes, l'autre variable le sera aussi dans ces classes. Inversement ces deux variables seront de valeur basse toutes les deux dans d'autres classes. Elles évolueront donc conjointement.

Dans l'étude que nous suivons, nous trouvons que la moyenne résiduelle de classe, cette sorte de valeur ajoutée par rapport au résultat prédictible pour cette classe, grâce notamment à l'enseignant-e, est en corrélation négative importante ($-0,610$; $-0,749$; et $-0,574$) avec la valorisation donnée dans cette classe (mesurée par le coefficient de pente ajouté) au fait d'être fille, d'être originaire d'un milieu social élevé et d'avoir une maturité scolaire évoluée (respectivement). La même corrélation peut être formulée de façon inverse. L'efficacité particulière de la classe est en corrélation positive (apparaît conjointement) avec le fait de valoriser dans cette classe les garçons, les niveaux sociaux culturellement défavorisés et les enfants au développement plutôt lent.

Il faut noter qu'il s'agit là non pas de la formulation d'une politique égalitariste, mais d'une simple constatation objective. Etant donné les handicaps connus des garçons, des milieux sociaux défavorisés et des enfants en retard, les enseignant-e-s qui font réussir spécialement bien l'ensemble de leur classe le font en obtenant un résultat particulièrement bon de ces groupes particuliers.

Cette évidence ne surprendra personne, mais elle peut justifier une politique compensatoire consciente et organisée.

PRÉPARATION DE L'ÉTUDE

Campagnes successives.

Les résultats présentés ci-dessus sont réels et bien établis. Leur publication aurait été justifiée. Mais ces analyses font partie d'un ensemble d'études trop redondantes pour être toutes présentées séparément, du fait notamment qu'elles reposent sur le même ensemble de données au départ. Il était préférable de considérer comme définitive et de publier la série d'analyses qui obtenait le plus de résultats significatifs et dont la validité conceptuelle paraissait la meilleure.

Ce qui distingue ces diverses études, c'est d'abord qu'elles ont porté sur des sous-ensembles différents de la banque de données totale recueillie en 1973. C'est ensuite que plusieurs critères de réussite en lecture ont été choisis successivement. Les paragraphes qui suivent décrivent ces jeux de données partiels et les campagnes d'analyses auxquelles ces observations ont donné lieu.

Puisque ces recherches préalables ne « méritent » pas une publication, pourquoi en parler ? – dira-t-on. C'est d'une part parce que nous y faisons souvent référence, comme dans les pages précédentes ; c'est aussi pour rendre accessibles des documents et des jeux de données qui pourraient servir à d'autres études ; enfin, c'est parce que les raisons qui nous ont obligés à reprendre des analyses successives sont en elles-mêmes un résultat scientifique, porteur d'information pour ceux qui voudraient poursuivre ce genre de recherche. Mais le lecteur pressé pourra sauter directement aux résultats définitifs, un peu plus bas.

Etude A – Sur données incomplètes

Données :

- Fichier EL3 (2278 élèves x 37 variables)
- Fichier CL3 (117 classes et 59 variables) d'avril 2004.

Rapport :

Application du logiciel HLM aux variables identitaires des données « Lecture » (ApplicHLM.doc du 25.5.2004).

Problèmes rencontrés :

Le logiciel oblige à abandonner tout élève dont les résultats sont incomplets. Or au total, 18 % des données manquaient, réparties de façon très aléatoire. Selon les variables étudiées, c'étaient donc d'autres élèves qui pouvaient être pris en compte ou rejetés. Plus le nombre de variables étudiées augmentait, moins il restait de cas à disposition. Il en était de même pour les caractéristiques des classes. Ainsi, l'échantillon des enseignant-e-s englobés dans l'étude variait du début à la fin de l'analyse.

En conséquence, la puissance des tests de signification diminuait et des résultats d'abord significatifs ne l'étaient plus quand l'étude s'élargissait. Le fait de prendre en compte des classes différentes selon les variables étudiées créait aussi une grande instabilité des résultats. On ne pouvait vraiment pas faire confiance aux conclusions.

Se limiter aux sujets et aux classes sans données manquantes aurait été impossible, car il n'y en avait pratiquement pas, vu le grand nombre de variables. Devant cette absence d'alternative, on a décidé d'exploiter la richesse des données de base et leur redondance partielle en complétant les observations manquantes par le moyen d'équations de régression.

Étude B – Sur données complétées à 2 ans :

Données :

- Fichier EL5 (1839 élèves x 38 variables)
- Fichier CL5 ((93 classes x 60 variables) du 24 mai 2005

Problèmes rencontrés :

Les études effectuées pour pouvoir compléter les données manquantes ont mis en lumière la faiblesse de certaines intercorrélations, marque soit de la spécificité, soit du manque de fidélité de ces épreuves. Dans ces cas, on a cru bon de totaliser des groupes de tests, du côté des prérequis comme du côté des critères, pour disposer de mesures plus significatives et plus fiables. D'où la préparation d'autres jeux de données (ci-dessous), calculés sur la base d'analyses factorielles des épreuves initiales.

Étude C Sur données factorisées complétées à 2 ans :

Données :

- Fichier EL7 (1839 élèves x 43 variables)
- Fichier CL7 (93 classes x 65 variables) du 7 septembre 2005

Rapports :

- *Premier rapport sur l'analyse multiniveau des données « Lecture Neuchâtel »* (Premier rapport. doc du 7.12.2005)
- *Stratégie d'exploration avec HLM – Exemple : Données Lecture Neuchâtel* (StratégieMultiniveau. pdf, 7 avril 2006)

Il était normal que les résultats des élèves et les rapports des enseignantes soient presque complets en début d'année. Il semblait donc peu problématique de partir de ces données initiales pour compléter les absences du mois d'après. La situation se retrouvait après qu'on ait complété le second relevé, et ainsi de suite, d'une période de relevé à la suivante. Mais il est clair que les effets des absences se cumulaient dans cette procédure déroulante et que les données devenaient de moins en moins valides, et de plus en plus affectées par le niveau général d'aptitude de l'élève mesuré au départ.

Par ailleurs, pour mieux assurer la fidélité de la mesure du critère en fin de 2^e année primaire, les différentes épreuves de compréhension de lecture en fin de 1^e et en cours de 2^e année ont fait l'objet d'une analyse factorielle et c'est un score factoriel qui les a résumées et qui est devenu le critère. Des résultats complétés par la procédure déroulante mentionnée ci-dessus ont donc contribué à l'estimation de cette note en facteur de Compréhension de 2^e année. Ceci renforçait la part de facteur général d'aptitude scolaire dans la mesure du résultat final.

Pour ne pas recréer des élèves fictifs, on a tout de même abandonné les élèves qui n'avaient pas eu de résultats en 2^e année. Cette absence de données pouvait résulter d'un problème administratif, lorsque les élèves avaient été dispersés, suite à la fermeture d'une classe, mais dans la plupart des cas il s'agissait simplement de redoublements : ces élèves n'avaient pas été admis en 2^e année.

Problèmes rencontrés :

Voir *L'effet des données manquantes sur la validité prédictive de tests d'aptitudes* (TraitementsStat.doc du 5.9. 06).

Les résultats de l'étude C paraissaient à première vue satisfaisants, mais les validités prédictives semblaient pourtant un peu « trop belles pour être vraies ». Un réexamen critique de la procédure utilisée révéla les erreurs de méthode suivantes :

a) Contamination des critères. La valeur prédictive des tests de prérequis a été comparée selon que le critère avait été observé directement, ou bien estimé par régression.

Les élèves dont la réussite avait été seulement estimée étaient parfaitement prédictibles, puisqu'on avait construit leur résultat. Ils représentaient le quart des cas. Si on les mélangeait avec les cas complets, le pourcentage de variance expliqué par le niveau de maturité initial était globalement de 28 %. Il n'était que de 24 % dans la situation normale où le critère avait été réellement observé. Le biais introduit par la procédure d'estimation était donc déjà de 4 % de la variance de la réussite, soit une part relativement importante de ce que l'on croyait pouvoir prédire. Pourtant ce mode de calcul du biais ne révélait encore qu'une partie du problème.

L'utilisation comme critère d'une note en facteur englobant des résultats obtenus par régression sur les deux années accentuait encore le biais, le portant à 10 % de la variance de la réussite, un résultat tout à fait inacceptable. La validité de la mesure avait été sacrifiée à l'augmentation de sa fidélité.

b) Changement de problématique. Le fait d'exclure de l'étude les sujets qui n'avaient pas de résultats en 2^e année modifiait subrepticement la nature de la recherche, puisqu'on ne considérait plus que les élèves « réguliers » et qu'on négligeait par conséquent les facteurs pouvant expliquer l'échec des autres. L'objectif de la recherche initiale avait été perdu de vue.

Il n'est pas sans intérêt de pouvoir analyser ainsi les mécanismes de ce glissement sémantique des objets d'étude, lié aux moyens de calcul utilisés.

*Étude D Sur données complétées à 1 an :***Données :**

- Fichier ELEVES8 (1605 élèves x 13 variables)
- Fichier CLASSES8 (91 classes x 33 variables) du 15 novembre 06

Pour ne pas avoir à abandonner les élèves redoublants, il fallait choisir comme critère la réussite aux épreuves de lecture de fin de 1^{re} année. Mais pour ne pas avoir à compléter des critères par régression, seuls les élèves qui ont passé ces épreuves ont été conservés. Ceci a amené à modifier sensiblement la composition de l'échantillon étudié par rapport aux analyses précédentes. Des classes ont été rejetées parce que leur effectif était devenu trop faible, mais d'autres qui avaient été abandonnées parce que dispersées en 2^e année ont pu être récupérées.

Problèmes rencontrés :

Sur la base d'un ensemble d'enseignant-e-s en partie nouveau, on arrive à des résultats partiellement différents. La question de la représentativité de ce groupe ne se pose pas puisqu'il représente toutes les classes qui pouvaient participer à l'étude ; mais un tiers des enseignantes neuchâtelaises ont dû être exclues pour diverses raisons, et cette sélection jette certains doutes sur l'effet d'une telle réduction. Une conséquence au moins est évidente : avec un échantillon d'effectif plus petit, les tests statistiques sont moins puissants et certaines des relations apparues antérieurement vont disparaître.

Étude E Sur la moyenne des critères

Données :

- Fichier ELEVES8N (1605 élèves x 13 variables)
- Fichier CLASSES8N (91 classes x 34 variables) du 15 novembre 06

Rapport: Le présent texte

Les trois critères globaux de la réussite en lecture dégagés par analyse factorielle des tests de fin d'année (Compréhension, Vitesse et Exactitude orale) ont conduit à trois études aux résultats convergents (dont l'Étude D citée ci-dessus). On a donc jugé préférable de faire la moyenne de ces trois critères pour pouvoir parler de la réussite dans un sens large et couvrir tout ce que l'on entend d'habitude par « savoir lire ».

La liste des variables individuelles est présentée dans le tableau 1 ; celle des variables de classe apparaît dans le tableau 2.

Tableau 1 – Variables relatives aux élèves

N° des variables	VARIABLES ET SIGNIFICATION DES CODES
	IDENTIFICATION
1	N° de district et de classe
2	N° de l'élève dans la classe
	VARIABLES PERSONNELLES
3	Sexe (01 = Masculin ; 02 = Féminin)
4	Âge calculé en mois
5	Profession du parent responsable (1 ^{er} chiffre de la codification genevoise à 3 chiffres), mais 0 transformé en 1 : niveau ouvrier.
6	Expression facile en français (1 = non ; 2 = plus ou moins ; 3 = oui)
7	Scolarité régulière (Redoublants = 1 ; réguliers = 3 ; incertains = 2)
8	Sait Lire ? (1 = non ; 2 = certains mots ; 3 = textes simples ; 4 = oui)
	PRÉREQUIS
9	Donné de départ : Total pondéré des 4 sous-tests d'Inizan + Maturité scolaire + Discrimination auditive + Suite de sons + Rappel d'histoires, avec pondération par équation de régression pour prédire les résultats de fin d'année. Standardisé ensuite.
	CRITÈRES
10	Compréhension de textes : total de 4 scores standardisés (Classes différenciées, Identification de graphèmes, Structures approuchantes et Nadine-Total des points), pondérés après analyse factorielle pour mesurer le facteur Compréhension. Score ensuite restandardisé.
11	Technique de lecture : total pondéré selon la saturation en facteur Technique des 4 mêmes tests, pour mesurer la compétence de déchiffrement. Score ensuite restandardisé.
12	Vitesse de lecture : mesurée par le temps mis pour répondre à la première page du test de Burion : Nadine. Son logarithme a été soustrait d'une constante, pour mesurer le facteur Vitesse. Score restandardisé ensuite.
13	Critère moyen standardisé : vu les résultats analogues obtenus avec les 3 critères précédents, ils ont été totalisés et standardisés.

Influence sur les conclusions

Comme on peut l'imaginer, ce choix d'un critère moyen a des avantages et des inconvénients. L'intérêt qui a paru décisif, c'est que le nombre de résultats significatifs était un peu plus élevé ; mais on perdait la différenciation des différents aspects du Savoir-Lire. En réalité, les différences entre les listes de variables significatives avec l'un ou l'autre des trois critères étaient minimales et non vraiment interprétables. Elles semblaient plutôt aléatoires.

Cette relative instabilité des résultats a dû être réduite par le choix d'un critère moyen, mais non éliminée. C'est pour cette raison que l'étude a été considérée comme plutôt exploratoire. Ses résultats sont indicatifs, mais certainement pas définitifs. Le seuil de signification a pour cette raison été choisi à $p = 0,10$, dans l'idée qu'il fallait surtout rechercher toutes les relations intéressantes, et accepter le risque correspondant d'être peu rigoureux en présentant des observations qui pourraient facilement être dues au hasard.

Les résultats qui vont être donnés maintenant ne seront donc pas tous commentés, un dixième d'entre eux étant certainement de nature aléatoire. C'est leur validité conceptuelle qui pourra aider à faire la différence.

Tableau 2 >>>

Tableau 2 – Variables relatives aux classes

N° des variables	VARIABLES ET SIGNIFICATION DES CODES
	IDENTIFICATION
1	N° de district et de classe
2	Nombre d'élèves dans la classe
	VARIABLES PERSONNELLES
	MOYENNES DE CHAQUE CLASSE POUR LES VARIABLES PERSONNELLES SUIVANTES :
3	Genre
4	Âge calculé en mois
5	Profession du parent responsable
6	Expression orale en français (classe rurale)
7	Scolarité régulière
8	Savoir lire initial
	PRÉREQUIS
	MOYENNES DE CHAQUE CLASSE POUR LA VARIABLE INDIVIDUELLE SUIVANTE :
9	Donné de départ
	CRITÈRES
	MOYENNES DE CHAQUE CLASSE POUR LES CRITÈRES INDIVIDUELS DE RÉUSSITE EN LECTURE SUIVANTS :
10	Compréhension de textes : total de 4 scores standardisés, (Classes différenciées, Identification de graphèmes, Structures appochantes et Nadine-Total des points), pondérés après analyse factorielle pour mesurer le facteur Compréhension
11	Technique de lecture : total pondéré selon la saturation en facteur Technique des 4 mêmes tests, pour mesurer la compétence de déchiffrage
12	Vitesse de lecture : mesurée par le temps mis pour répondre à la première page du test de Burion : Nadine. Son logarithme a été soustrait d'une constante, pour mesurer le facteur Vitesse
13	Critère moyen standardisé
	CARACTÉRISTIQUES DE LA CLASSE
	LES VARIABLES SUIVANTES ONT ÉTÉ STANDARDISÉES :
14	Nombre de degrés dans la classe
15	Option pédagogique : Importance du décodage
16	Idem : Objectifs éducatifs de niveau élevé
17	Idem : Étude des lettres
18	Idem : Motivation externe
19	Idem : Exercices préparatoires à la lecture
20	Idem : Motivation interne
21	Idem : Objectifs affectifs
22	Idem : Rapidité d'avancement
23	Idem : Bibliothèque dans la classe
24	Idem : Matériel de lecture dans la classe
25	Approche Oral (score factoriel)
26	Approche Écrit (score factoriel)
27	Approche Exercices de prérequis (score factoriel)
28	Approche Enrichissement des activités (score factoriel)
29	Emploie la Méthode S'Exprimer-Lire
30	Emploie la Méthode Gestuelle
31	Emploie la Méthode du Sablier
32	Emploie la Méthode Syllabique
33	Nombre d'années de pratique de la méthode utilisée
34	Jugement sur la qualité de la méthode utilisée

RÉSULTATS DE L'ÉTUDE

Comme on l'a vu au chapitre précédent, une analyse multiniveau est complexe et présente des résultats de différentes natures, relatifs aux élèves, aux classes, et à l'interaction de ces deux niveaux, son intérêt spécifique étant de quantifier ces influences par rapport à une échelle commune, celle de la variance du critère dont on veut rendre compte. Les résultats vont être présentés dans l'ordre qui vient d'être indiqué.

Facteurs individuels (effets dépendant des élèves)

Le tableau 3 montre que toutes les variables recueillies concernant chaque élève ont un effet positif sur son apprentissage, dans le sens attendu (sens qui avait d'ailleurs déterminé le codage, puisqu'on souhaitait rendre les coefficients positifs pour simplifier leur emploi).

Tableau 3 – Variables personnelles influençant l'apprentissage de la lecture

FACTEUR CONSIDÉRÉ	COEFFICIENT*	SIGNIFICATION
Genre (Fille = +; Garçon = -)	0,07	p = 0,000
Niveau socioprofessionnel	0,02	N. S. (0,44)
Expression orale facile	0,02	N. S. (0,30)
Scolarité régulière	0,13	p = 0,000
Connaissance préalable de la lecture	0,11	p = 0,000
Donné de départ (tests d'aptitude scolaire)	0,52	P = 0,000

* Il s'agit de poids Bêta dans l'équation de régression multiple prédisant la réussite en lecture.

Ces résultats ne surprendront personne. Les filles confirment ici un avantage scolaire connu dans tous les pays et qui se marque surtout dans les branches à composante verbale, comme l'étude de la langue maternelle. On peut se demander plutôt la raison du si faible poids du niveau socioprofessionnel et de la langue maternelle française dans la réussite en lecture, alors qu'on connaît bien leur importance pour l'école, d'après toutes les recherches antérieures.

En fait, les cinq premières variables personnelles ont toutes à peu près la même valeur prédictive, avec des corrélations autour de 0,10. Parmi elles, c'est justement le niveau socioprofessionnel et la langue maternelle française qui ont les validités les plus hautes, respectivement égales à 0,14 et 0,17 (toutes classes mélangées).

La raison de leur faible poids Bêta, c'est que le modèle multiniveau calcule une équation de régression multiple pour prédire au mieux la réussite, et qu'alors les recouvrements entre prédicteurs peuvent aboutir à donner plus d'importance à une variable qu'à une autre, même si leur validité directe (univariée) est la même.

Ici, c'est le Donné de départ qui reçoit tout le poids dans la prédiction, parce qu'il est de loin le plus corrélé avec la réussite de fin d'année (0,53). Le niveau socioprofessionnel et la langue maternelle sont des variables corrélées avec le niveau d'aptitude donné au départ (respectivement à 0,17 et 0,30). Elles sont donc redondantes, et c'est pourquoi leur poids est si faible, dans le système formé par toutes les variables identitaires.

La variable Scolarité régulière, en cette première année d'école, oppose les élèves d'âge normal aux élèves qui répètent leur première année. On voit que la réussite de ces derniers, après leur redoublement, reste en fin d'année inférieure à la moyenne, de façon très significative.

La connaissance préalable de la lecture a une signification ambiguë. Elle peut être le fait d'enfants au développement précoce, qui ont appris à lire par eux-mêmes, ou avec l'aide de leur famille. Dans ce cas, cette avance peut les avantager. On voit d'après le signe du coefficient Bêta que c'est le cas en général. Mais il peut s'agir aussi de redoublants, qui ont acquis certains rudiments de ce savoir-faire au cours de la première année, sans avoir passé le cap du savoir-lire véritable. Pour ces derniers élèves, leurs bribes de savoir révèlent plutôt un handicap, comme on l'a dit au paragraphe précédent. La combinaison de ces deux influences opposées peut expliquer la faible valeur prédictive du savoir-lire initial.

Le résultat le plus frappant, dans le tableau 3, c'est le fait que le niveau de départ explique à lui tout seul $(0.52)^2 = 27\%$ de la réussite en lecture, alors que l'ensemble des 6 variables individuelles ont une corrélation multiple (toutes classes confondues pour l'instant) égale à 0,546, expliquant ainsi à peine 30 % du critère de lecture. On retrouve la conclusion évoquée précédemment à partir d'une des analyses préalables : les nombreux facteurs qui ont un poids dans la réussite révèlent déjà leur influence dans la réussite aux tests d'aptitude scolaire de début d'année. Ainsi la mesure d'aptitude individuelle ne doit pas être conçue comme la cause génétique du succès scolaire, mais comme le révélateur de l'ensemble des facteurs (externes et internes) favorisant l'apprentissage scolaire.

Les résultats précédents découlent de l'application de méthodes corrélationnelles classiques (équations de régression appliquées aux variables individuelles). Les résultats qui vont suivre peuvent aussi être obtenus par des méthodes classiques, à condition d'appliquer les équations de régression à la population des classes, en prédisant les moyennes de classe au critère à partir des moyennes de classe aux prédicteurs. Mais l'approche multiniveau peut renseigner à la fois sur ces deux niveaux, plus sur leur interaction, qu'on présentera plus loin.

Facteurs collectifs (effets généraux sur les classes)

L'étude des effets généraux intervenant sur la réussite des classes va être effectuée en deux fois, pour une raison technique sur laquelle on va revenir. Le premier groupe de variables sera constitué des caractéristiques du groupe d'élèves formant la classe. Il pourrait s'agir, par exemple, de leur âge moyen. Le second groupe de variables correspondra plutôt aux caractéristiques de l'enseignante, comme par exemple ses options pédagogiques.

Tableau 4a – Influence des caractéristiques moyennes des élèves de la classe

(Coefficient brut = Poids Bêta de l'équation de régression multiple calculée sur les moyennes de classes)

(Coefficient ajusté = Poids précédent multiplié par l'écart type des moyennes de classes correspondantes)

Coefficient brut	Niveau de signification	Variable considérée	Coefficient ajusté
0.08	0.76	Proportion de filles dans la classe	0.015
0.03	0.83	Niveau socioprofessionnel moyen	0.011
- 0.10	0.48	Classe d'expression majoritairement française (= rurale)	- 0.036
0.33	0.10	Classe d'élèves réguliers	0.079
- 0.30	0.01	Classe avec élèves sachant un peu lire (= redoublants)	- 0.130
0.49	0.00	Classe de haut niveau d'aptitude au départ	0.224

La méthode multiniveau fournit une grande équation linéaire où toutes les variables prédictives reçoivent chacune un coefficient multiplicateur optimal pour prédire le critère, ici la performance en

lecture de chaque élève en fin de première année. Ce coefficient, directement utilisable, est celui qui est donné dans la première colonne du tableau ci-dessus.

On voit que le « Niveau de départ » moyen de la classe domine toutes les autres sources d'influence. Son coefficient est moitié plus grand que pour la composition de la classe (du point de vue de sa proportion d'élèves réguliers ou d'élèves sachant déjà un peu lire). Il est au moins cinq fois plus grand que pour les trois autres caractéristiques moyennes des élèves.

Il est intéressant de se rappeler que, à l'intérieur des classes, le poids de cette même variable était de 0.52, pour prédire la réussite des élèves les uns par rapport aux autres (Tableau 3). Cette validité prédictive individuelle est peu différente de 0.49, le coefficient s'appliquant aux moyennes de classes. On voit ainsi que l'effet de l'aptitude est direct et indépendant de la classe. L'avance ou le retard personnel de l'élève aura le même effet, quelle que soit la classe dans laquelle il se trouvera placé. En effet, si l'élève appartient à une classe de moyenne élevée, sa position par rapport à la moyenne de sa classe sera dépréciée d'autant, mais son pronostic de réussite sera rétabli par l'effet de l'autre droite de régression dont la pente est la même, celle relative aux classes, où sa classe bénéficiera du même décalage, cette fois pour relever son pronostic.

A côté de cette variable essentielle du niveau de départ, les autres moyennes de classe font pâle figure. Elles seront commentées dans l'ordre que propose le tableau ci-dessus.

La proportion de filles dans la classe n'a pas d'effet significatif, mais son coefficient est pratiquement le même (0.08) que celui donné plus haut, dans le tableau 3, à la variable individuelle correspondante (0.07). La discussion du niveau d'aptitude au paragraphe précédent vient de souligner la signification de cette égalité de pente du prédicteur individuel et de la moyenne de classe correspondante. Sa conclusion (les effets s'ajoutent) peut s'appliquer à l'effet Fille/Garçon. Dans une classe majoritairement féminine, une fille aura peu d'écart à la moyenne et semblerait relativement désavantagée par rapport à ce qu'elle aurait pu connaître dans d'autres classes. Mais elle retrouve son avantage du fait que la moyenne de sa classe sera plus élevée sur cette variable et que cela augmentera d'autant son pronostic, puisque les deux pentes, intra- et inter-classes, sont les mêmes et positives toutes les deux.

Le niveau socioprofessionnel moyen n'est pas plus significatif au niveau des moyennes de classe qu'au niveau individuel. Faut-il accuser des relevés ou des codages défectueux ? On peut aussi se demander si on a eu raison d'utiliser la classification genevoise qui rassemblait au niveau 1 tous les travailleurs de l'industrie et au niveau 2 les agriculteurs et les employés. On mélangeait ainsi des niveaux bien différents sur le plan culturel. Mais en dépit de ces suspicions, on verra plus bas qu'on peut expliquer l'absence d'effet du niveau social par des interactions avec les autres variables individuelles, qui ont annulé son effet moyen général.

La langue maternelle moyenne ne semble pas avoir d'effet marqué, et en tout cas pas dans la direction que l'on pouvait attendre, puisque son coefficient est négatif pour les classes où la langue française est majoritaire. On trouve là un exemple de la dissociation annoncée plus haut comme possible entre les pentes au niveau individuel et au niveau des classes. Cette influence apparemment négative de la langue française quand il n'y a pas d'étrangers dans la classe provient sans doute de ce que le facteur linguistique est compensé par l'habitat citadin des enfants étrangers, milieu plus stimulant, surtout à cette époque. Ces enfants avaient eu notamment la possibilité de fréquenter un jardin d'enfants, ce qui était rare dans les campagnes neuchâteloises à ce moment-là. Au niveau 2, celui des classes, les moyennes peuvent donc correspondre à des sources de variation réelles différentes de ce que la variable de niveau 1 laissait attendre. Au lieu de la langue, peut-être s'agit-il en fait de l'opposition entre classes rurales et classes urbaines, à l'avantage des villes. L'effet n'est pas significatif de toute façon.

Il se peut qu'une influence cachée du même ordre explique l'effet négatif du redoublement (si l'on choisit ce pôle pour définir cette variable). Le coefficient de -0.33 est élevé et statistiquement significatif, au seuil de 0.10 , de sorte que l'on doit admettre que la présence d'enfants redoublants dans une classe est liée à un mauvais pronostic de réussite. Mais une corrélation n'est pas une liaison causale. La raison de ce handicap peut être une caractéristique générale du milieu, sans doute culturellement défavorisé.

C'est la même source d'influence qui explique le poids négatif donné à la variable suivante, la connaissance partielle de la lecture en début d'année. Il doit s'agir des éléments de savoirs que les enfants qui redoublent leur première année ont néanmoins acquis. On observe en effet une corrélation de 0.51 sur les quelque 1600 élèves étudiés entre le fait de redoubler et d'avoir une connaissance partielle de la lecture. Le coefficient de -0.30 , nettement significatif cette fois, ne serait ainsi que l'autre face de l'effet précédent, le milieu social.

Le tableau 4a donne dans sa première colonne les coefficients à appliquer directement aux moyennes de classes, mais ces coefficients ne traduisent pas l'importance réelle accordée à chaque prédicteur. Ils sont en effet affectés par l'écart-type des moyennes qu'ils pondèrent. On aimerait comparer le poids donné à des variables standardisées. Dans ce cas, ce sont les coefficients de la dernière colonne qu'il faut considérer. Ceux-ci sont égaux aux valeurs de la première colonne, multipliées par l'écart type des moyennes correspondantes (données à l'Annexe D).

On voit alors que les trois premières variables reçoivent des pondérations pratiquement nulles, qu'il est donc inutile de discuter. Le niveau moyen de capacité scolaire (en bas du tableau) reste la variable la plus importante, mais son poids (0.22) n'est plus que le double de celui des deux variables liées aux redoublements (-0.08 et -0.13). Ce sont ces dernières pondérations qui vont pouvoir être comparées à celles du tableau 4b suivant, parce que les variables présentées sous 4b sont toutes standardisées, comme dans la dernière colonne du tableau 4a. Elles correspondent à des caractéristiques soit de la classe (comme son équipement), soit de l'enseignante (comme son approche didactique). Ces coefficients permettent de comparer l'importance des variables correspondantes pour la réussite de l'apprentissage (les variables sans relation avec la réussite ont été laissées de côté).

On notera que le niveau de départ de la moyenne de la classe reste, avec son coefficient de 0.22 , le facteur commun de réussite le plus important, mais que son poids n'est plus tellement différent de celui des autres variables pédagogiques étudiées ci-dessous, qui peuvent se situer entre 0.10 et 0.20 .

Tableau 4b – Influence des variables pédagogiques sur l'apprentissage de la lecture

(Poids Bêta directement applicables)

Coefficient ≥ 0.10	Niveau de signification	Variable considérée
0.13	0.193	Option Lettres
0.18	0.000	Motivation interne
0.12	0.025	Rapidité d'avancement
0.10	0.079	Équipement en classe
0.17	0.001	Satisfaction de l'enseignante pour sa méthode

On a vu au tableau 2 que 33 variables caractérisent les classes qui ont été étudiées (le tableau comporte 34 entrées mais ne considère pas le numéro de la classe comme une variable). Les résultats pour 6 d'entre elles (les caractéristiques moyennes des élèves de la classe) viennent d'être présentés au tableau 4a. Parmi les 27 variables restantes, beaucoup sont non significatives et vont être laissées de côté, pour simplifier ce rapport, mais les résultats détaillés sont donnés dans l'Annexe C. Quatre autres, relatives aux méthodes didactiques choisies, ont dû être abandonnées dans l'étude

multiniveau et traitées différemment, pour des raisons qui seront expliquées ci-dessous. Parmi les 23 variables restantes, beaucoup sont non significatives et vont être laissées de côté, pour simplifier ce rapport, mais les résultats détaillés sont donnés dans l'Annexe C.

Le tableau 4b présente les 5 variables qui ont soit un coefficient égal ou supérieur à 0.10, soit un niveau de signification inférieur à 0.10. Ces deux critères sont peu exigeants, mais ils ont été choisis pour diminuer le risque de ne pas apercevoir une relation intéressante.

La variable « Satisfaction de l'enseignante », d'abord, n'est pas à proprement parler une variable pédagogique. C'est pourtant (presque) celle qui a le plus de liaison avec le niveau de réussite des élèves ! Plusieurs interprétations peuvent en être données, qui ne sont pas forcément exclusives l'une de l'autre. Il est clair que si sa classe progresse bien dans l'apprentissage de la lecture, une enseignante s'en aperçoit et est satisfaite. Mais inversement, on peut penser que si une enseignante a confiance dans sa méthode, elle sera plus à l'aise dans son maniement, elle pourra l'exploiter plus complètement et plus efficacement. D'où un cercle de renforcements qui ne peut que profiter aux élèves. Malheureusement, on ne sait pas comment amorcer ce cercle vertueux et la liaison trouvée n'aide guère à choisir une didactique. On peut voir, en recourant aux données de base, que ce sont les maîtresses utilisant la méthode du Sablier qui étaient les plus enthousiastes pour leur approche pédagogique, mais cela ne prouve pas que cette méthode soit la meilleure pour autant. On connaît les effets de mode dans ce domaine.

(Une parenthèse doit être ouverte ici à ce propos. Le questionnaire adressé à toutes les enseignantes comportait la mention des diverses méthodes à disposition à l'époque, y compris des catégories assez larges, comme l'approche syllabique, par exemple. Chaque enseignante avait dû choisir une de ces méthodes comme la plus proche de sa façon d'enseigner. On espérait pouvoir comparer les résultats en lecture des élèves selon les méthodes. Or c'était une erreur méthodologique d'introduire ce genre de variables, car elles n'étaient plus indépendantes. Lorsqu'une méthode avait été choisie, toutes les autres étaient considérées comme refusées. Le coefficient qu'elles recevaient dans l'étude multiniveau devenait en conséquence négatif, sans qu'on puisse vraiment interpréter sa valeur. On trouvera en Annexe B une analyse statistique simple des résultats de quatre méthodes, mais qui ne pourra pas dissocier l'influence des autres sources de variance, comme l'analyse multiniveau cherche à le faire.)

La variable pédagogique qui semble la plus recommandable dans le tableau 4b ci-dessus, est intitulée « Motivation interne ». C'est celle dont le poids est le plus élevé pour prédire le succès, avec un coefficient de 0.18, significatif à $p < 0.001$. Elle avait été définie dans l'étude de 1973 comme le premier facteur issu de l'analyse factorielle du questionnaire rempli par les enseignantes pour décrire leur méthode. L'étude multiniveau montre aujourd'hui que c'est la méthode qui a le plus d'efficacité.

Voici comment ce facteur était décrit : « Les six variables qui saturent ce facteur définissent une conception particulière de la motivation que nous appelons interne dans ce sens que l'intérêt pour la lecture doit être suscité par l'apprentissage même de ce savoir-faire. Selon ce point de vue, propre à certaines enseignantes, l'intérêt pour la lecture dépend des succès de l'apprentissage, du caractère plaisant, significatif et facultatif des exercices, du soutien pédagogique à disposition des élèves, de leur désir d'apprendre, de l'intérêt des textes qu'on leur propose et de la forme ludique des leçons. »

Pour chaque enseignante, un « score résumé » a alors été calculé en totalisant ses réponses aux questions correspondant à cette description. Le même travail a été fait pour d'autres facteurs. Ces nouvelles variables résumant le premier Questionnaire-Méthode ont donné lieu à une seconde analyse factorielle. De nouveau un facteur (le facteur 8) est apparu comme un choix didactique particulier. Il était décrit de la façon suivante :

« Selon ce facteur, la stimulation à la lecture provient de la lecture elle-même et de son apprentissage. Les variables " lecture comme recherche de signification " et " élaboration ", fortement saturées dans

ce facteur de second ordre, clarifient encore la signification du facteur de premier ordre " motivation interne ". Dans ce cas, la motivation pour la lecture est donc provoquée par un enseignement de la lecture significatif pour l'enfant, c'est-à-dire où le déchiffrage est vu comme une technique au service d'objectifs plus élevés comme la compréhension et l'élaboration, soit l'application de la lecture à des domaines variés. »

Après la Motivation interne, c'est l'Option Lettres qui a le poids le plus élevé dans la réussite, égal à 0.13, même si son niveau de signification statistique n'atteint pas le seuil habituellement requis. C'était le deuxième facteur de l'analyse factorielle du Questionnaire-Lecture. Il avait été dénommé « Lecture comme technique de décodage » et décrit ainsi : « Selon cette conception, l'enseignement de la lecture est considéré comme une succession d'étapes précises, bien réglées, par lesquelles l'élève doit passer pour maîtriser l'acte lexicque : introduction progressive des difficultés, apprentissage par synthèses successives des éléments de la phrase (lettres, syllabes, mots), avec support (gestuel ou visuel), puis sans support, et enfin renforcement des acquisitions par répétition. La lecture est considérée essentiellement comme une activité de décodage. »

Dans l'analyse factorielle de second ordre, cette variable devenait le pôle positif du premier facteur bipolaire opposant une approche par les lettres ou par les sons. « Ce pôle accorde la priorité à l'écrit, à l'enseignement direct des lettres et des éléments du mot. C'est une approche rigoureuse, se déroulant selon des étapes bien précises. Les enseignantes qui la choisissent bénéficient d'une très longue expérience de la méthode et considèrent la lecture comme une technique de décodage. Elles n'ont guère l'habitude d'utiliser un matériel d'enseignement. »

La troisième variable pédagogique qui prédit la réussite de fin de première année est la Rapidité d'avancement. Son coefficient est de 0.12 et sa significativité statistique est certaine, avec une probabilité de 0.025. Elle aussi correspond à un facteur de la première analyse factorielle, dit « d'avancement », décrit de la façon suivante :

« Il correspond à un enseignement qui progresse rapidement dans le programme (mesuré par le nombre de lettres et groupes de lettres, ou par le nombre de sons-voyelles et de sons-consonnes, lus à un moment donné de l'année, ou par le nombre de règles " grammaticales " enseignées jusque-là. »

Ce facteur était réapparu dans l'analyse de second ordre avec la même signification, formant le facteur 6, intitulé « Progression dans le programme ».

La dernière des variables retenues selon les critères annoncés plus haut a été « Équipement en classe ». Son coefficient est de 0.10 et sa probabilité d'être due au seul hasard est de 0.079, ce qui satisfait le seuil de signification choisi.

Ce facteur correspondait surtout à la méthode développée à l'École Normale de Neuchâtel, « S'Exprimer-Lire », qui offrait aux enseignantes de nombreuses formes d'exercices de lecture. Mais d'autres méthodes nouvelles, comme le Sablier, proposaient aussi un abondant matériel.

Il est intéressant de noter que les enseignantes qui avaient choisi l'option Lettres se distinguaient justement par le fait de ne pas utiliser de matériel particulier tout en réussissant très bien dans leur enseignement. De même, la première variable, Motivation interne, reposait plus sur le contenu des lectures et leur sens pour les élèves que sur du matériel mis à leur disposition.

Ainsi les différentes variables retenues par l'analyse multiniveau ne correspondent pas du tout à une théorie polarisée idéologiquement de l'enseignement de la lecture, mais au contraire à des conceptions plutôt opposées, qui réussissent néanmoins à contribuer à la réussite. Il semblerait ainsi que des approches complémentaires soient possibles. Ce pourrait être là la conclusion didactique la plus intéressante, parce que paradoxale et imprévue, de cette étude des facteurs collectifs influençant l'apprentissage de la lecture.

Il reste à voir maintenant comment les effets individuels et les effets collectifs étudiés dans les deux parties précédentes peuvent interagir. C'est ce type de questions que l'analyse multiniveau est la seule à pouvoir traiter.

Interactions des niveaux (effets différenciateurs du contexte)

Si le lecteur a pu être surpris au point précédent du petit nombre de résultats significatifs lorsqu'il s'agissait de dire quelle méthode obtenait les meilleures performances en moyenne, il risque d'être surpris à l'inverse de découvrir le grand nombre d'effets significatifs qui apparaissent lorsqu'on recherche quels aspects du contexte modulent l'influence des facteurs individuels déjà étudiés.

Le principe des interactions entre niveau individuel et niveau collectif a déjà été expliqué et illustré au début de ce texte. Ces interactions proviennent du fait qu'un prédicteur individuel, par exemple l'aptitude scolaire, n'a pas forcément la même valeur prédictive dans toutes les classes. Dans certaines, la réussite dépendra beaucoup du niveau de maturité scolaire de l'enfant; dans d'autres, l'enseignante réussira à créer une dynamique d'apprentissage qui suffira à compenser certains des handicaps de départ. On peut donc ajuster la pente (le coefficient d'importance) du niveau de maturité dans chaque classe pour obtenir une prédiction optimale. La variance d'erreur résultant de l'écart entre résultat prédit et résultat effectif va alors diminuer.

On peut tester statistiquement si cette réduction est spécialement importante ou non. Si elle reste dans ce que le hasard peut expliquer, on conclura qu'il n'y a pas d'interaction à analyser, que la pente est la même dans toutes les classes, aux fluctuations aléatoires près. Si elle est très importante, on pourra chercher à l'expliquer par l'effet de variables du contexte (celles qui caractérisent les classes). Un effectif de classe élevé, par exemple, pourrait moduler l'effet des caractéristiques personnelles, en rendant plus difficile à l'enseignant de prêter une attention individualisée à ses élèves.

Une nouvelle équation de régression à ce niveau des classes permettra de prédire la modification de pente (augmentation ou diminution) applicable dans chaque classe pour améliorer la prédiction. De nouveau, on aura une part de la variance d'erreur qui sera explicable par cette équation de régression et une part résiduelle, les deux pouvant être testées par rapport au hasard. La première étape de l'étude des interactions est donc de tester, pour chaque prédicteur individuel, si les pentes de ses droites de régression classe par classe varient significativement autour de la pente moyenne calculée pour toutes les classes. Lorsque ce n'est pas le cas, il n'y a pas de place pour des interactions. Le tableau 5 reproduit les résultats du logiciel HLM sur ce point, pour les six prédicteurs individuels.

Tableau 5 – Composantes de pente résiduelles pour chaque prédicteur individuel

Estimation finale des composantes de variance

Effet aléatoire	Code	Écart-type	Comp.Var.	d. l.	Chi carré	Probabilité
Intercept1	U0	0.32652	0.10661	16	169.12981	0.000
Pente pour Genre	U1	0.05727	0.00328	44	54.97102	0.124
Pente p. NivProf	U2	0.11533	0.01330	44	58.05077	0.076
Pente p. ExpOrale	U3	0.04583	0.00210	44	37.15961	>0.500
Pente p. Régulier	U4	0.07641	0.00584	44	65.70268	0.018
Pente p. Prélecture	U5	0.05171	0.00267	44	47.47388	0.333
Pente p. DonnéDé	U6	0.11908	0.01418	44	71.71162	0.005
Niveau 1	R	0.65968	0.43518			

Les Chi carrés ci-dessus ont été calculés sur les unités qui avaient suffisamment de données (45 sur 91). Les effets fixes et les composantes de variance reposent sur l'ensemble des données.

On voit dans la dernière colonne (donnant la probabilité d'obtenir ce résultat par hasard) que deux des prédicteurs individuels n'ont pas de variance résiduelle de pente significative : U3 l'expression orale en français et U5 la connaissance préalable de la lecture. Les quatre autres prédicteurs individuels, par contre, ont des variances résiduelles de pente qui varient selon les classes d'une façon qu'on peut essayer de prédire (le cas du prédicteur Fille/Garçon est à la limite, mais il paraît intéressant de l'englober dans l'étude). Les tableaux 6 à 9 ci-dessous montrent pour chaque prédicteur quelles caractéristiques du contexte de la classe modifient son influence.

Genre

Tableau 6 – Variables contextuelles modulant l'effet de la caractéristique personnelle Fille/Garçon

Coefficient ≥ 0.10	Niveau de signification	Variable considérée
- 0.04	0.09	Classe nombreuse (effet compensatoire)
0.12	0.08	Classe d'expression majoritairement française (rurale)
- 0.15	0.08	Classe d'élèves réguliers
- 0.05	0.05	Rapidité d'avancement
- 0.05	0.09	Bibliothèque dans la classe
0.04	0.07	Approche par Exercice des prérequis

Les trois variables de classes : Expression majoritairement française, Élèves réguliers, Exercice des prérequis, n'ont pas été standardisées, ce qui triple environ leurs coefficients relativement à ceux des autres variables. Voir l'Annexe D.

Rappelons qu'avec un niveau de signification de 0.10, un dixième des résultats présentés seront dus au hasard. Il est donc permis de ne pas chercher à tout interpréter. Mais certains coefficients peuvent mettre en évidence des effets compensatoires (ou au contraire renforçateurs) importants, dont la réalité semble alors mieux assurée.

Effectif de la classe

Par compensatoire, on désigne un effet qui (à son pôle positif) contrecarre, et par conséquent équilibre, l'effet principal qui avait été annoncé plus haut, à propos de chaque variable individuelle. On avait vu ainsi que le genre féminin de l'élève donnait un avantage systématique de 0.07 sigmas, (on dira de 7 points), aux filles dans le calcul de leur pronostic de réussite en lecture. La première ligne du tableau 6 montre, du fait du signe négatif du coefficient correcteur relatif à l'effectif de la classe, que l'effet du genre de l'élève sera réduit dans une classe nombreuse (et parallèlement augmenté dans une classe de faible effectif).

On le comprend en détaillant les calculs. La pente moyenne de 0.07 sera corrigée dans chaque classe en lui ajoutant le produit de l'effectif de la classe (en score standardisé) par le coefficient indiqué de - 0.04. Pour une classe se situant à 1 écart-type au-dessus de la moyenne du point de vue Nombre d'élèves, avec un score de +1 par conséquent pour la variable Effectif, le coefficient correcteur de la pente du Genre sera de (+1) x (- 0.04) = - 0.04. Ce coefficient sera ajouté pour cette classe à celui de la pente moyenne de l'effet Fille/Garçon (qui est 0.07), donnant un poids de 0.03 pour le Genre dans ce contexte particulier.

Pour la prédiction individuelle de la réussite, on applique ce poids au score standardisé correspondant au genre féminin ou masculin de l'élève et qui se monte à + 1 ou - 1, à quelques décimales près. Il en résulte un avantage de 3 points seulement pour les filles et un handicap correspondant de 3 points seulement pour les garçons, ce qui réduit l'inégalité moyenne et montre l'effet compensateur d'un effectif de classe élevé.

Mais pour une classe moins nombreuse, par exemple à la moyenne, le coefficient correcteur sera nul, puisque $(+0) \times (-0.04) = 0$ et le poids du genre restera le poids moyen de 0.07.

Pour une classe peu nombreuse (à 1 sigma en dessous de la moyenne et donc avec un score de -1 en Effectif), la correction de pente deviendra renforçatrice du fait que $(-1) \times (-0.04) = +0.04$. L'avantage des filles sera alors de $(+1, \text{ le score du genre Fille}) \times (0.07 + 0.04) = 0.11$ sigmas sur l'échelle de réussite en lecture. Le handicap des garçons sera l'opposé, car la pente $(0.07 + 0.04)$ sera multipliée par -1 environ, le score des Garçons sur l'échelle Filles/Garçons. Il vaut donc mieux pour un garçon qu'il apprenne à lire dans une classe nombreuse.

Langue majoritaire

La deuxième ligne du tableau 6 montre que l'avantage des filles dépend aussi de la langue majoritaire dans la classe. Dans le contexte d'une classe essentiellement francophone, les filles verront leur avantage initial (de 0.07) considérablement renforcé, puisqu'elles bénéficieront d'un avantage supplémentaire de 12 points (dans une classe à un sigma au-dessus de la moyenne en taux de francophones). Mais l'effet sera inversé dans une classe comportant beaucoup d'étrangers, avec un score de -1 pour cette variable de classe. En effet, la correction de pente sera alors de -0.12 pour une classe à un sigma en-dessous de la moyenne pour l'expression orale en français. Le coefficient multiplicateur du genre y sera au total $(+0.07) + (-0.12) = (-0.05)$. Du coup, ce sont les garçons qui auront un avantage de 5 points dans cette classe comportant beaucoup d'étrangers, dans un contexte probablement urbain.

Redoublement

Les autres effets d'interaction mis en évidence sont plutôt compensatoires à leur pôle positif. Dans une classe comportant peu de redoublants, l'avantage des filles disparaîtra, le coefficient multiplicateur du score du Genre étant égal à $0.07 - 0.15 = -0.08$. Dans ces conditions, ce sont les garçons qui seront avantagés, par rapport à la situation moyenne dans l'ensemble des classes. Ceci provient sans doute du fait que ce sont surtout les garçons qui redoublent d'habitude. Mais inversement, dans une classe avec beaucoup de redoublants, on peut s'attendre à avoir beaucoup de garçons en difficulté. Leur handicap sera en effet de $(0.07 + 0.15) \times (-1.0) = -0.22$ dans cette classe située à -1 sigma en taux d'élèves réguliers.

Rapidité d'avancement

De même, dans une classe où l'enseignante maintient un rythme de travail rapide, avec par conséquent probablement un niveau d'aspiration exigeant et un temps considérable consacré à la lecture, les garçons ne connaissent presque plus de handicap par rapport aux filles, l'avantage de ces dernières n'étant que de $0.07 - 0.05 = 0.02$. Le handicap attendu des garçons est réduit à -0.02 .

Bibliothèque

La présence d'une bibliothèque dans la classe a un effet compensatoire similaire pour réduire le biais Fille/Garçon. Ceci provient peut-être du fait que ce sont des enseignantes expérimentées qui ont pu constituer progressivement une bibliothèque et que ce sont ces enseignantes qui parviennent à enseigner efficacement et à soutenir effectivement les élèves en difficulté, qui sont surtout des garçons. Le handicap de ces derniers dans les classes avec bibliothèques a une valeur attendue de -0.02 seulement.

Exercice des prérequis

L'inverse peut être dit de la méthode didactique qui veut assurer les bases psychologiques requises pour l'étude de la lecture (latéralisation, reconnaissance des formes, etc.) avant d'entreprendre l'apprentissage proprement dit de la lecture. Le signe positif de son coefficient correcteur est opposé à

celui trouvé pour une variable didactique à effet compensatoire, comme la rapidité d'avancement. Le temps passé à ces exercices ne doit donc pas profiter réellement aux garçons. Le déficit de performance attendu pour eux dans les classes exerçant les prérequis est de -0.11 .

Le lecteur sera peut-être tenté de considérer ces interactions comme mineures et de peu d'importance, puisqu'elles ne modifient les poids bêta qu'à la seconde décimale le plus souvent. Mais en réalité, les poids bêta eux-mêmes sont du même ordre de grandeur, restant aux alentours de 0.10, à part le niveau de maturité scolaire de départ. Il s'agit donc, pour les interactions, d'effets aussi notables et à prendre en considération aussi attentivement que les effets moyens désormais bien connus des variables personnelles (genre, niveau social, etc.).

Les interprétations précédentes sont assez faciles à formuler, parce que la variable Fille/Garçon ne comporte que deux valeurs possibles. Avec des valeurs standardisées, Fille est codé +1 et Garçon est codé -1, à quelques décimales près. C'est le produit par ces +1 ou -1 des coefficients 0.07, 0.04, 0.02, etc. qui donne les avantages ou handicaps systématiques dont on vient de parler. Mais avec les autres variables individuelles qui vont être examinées maintenant, le score de chaque élève sera une valeur pouvant aller de -2 à +2 environ. L'avantage ou le handicap dû à une caractéristique personnelle sera donc plus difficile à calculer, étant le produit du coefficient multiplicateur (pente générale + modulation de pente) par le score de l'individu, allant de -2 à +2.

Pour pouvoir faire comprendre ces effets, il faut bien simplifier leur présentation. On admettra donc dorénavant qu'on parle d'un élève dont le score sur la variable personnelle étudiée est de +1 sigma. Si ce score est en réalité de -1 sigma, l'effet indiqué sera à inverser. Les effets seront toujours proportionnels à la valeur de la variable personnelle envisagée.

Niveau social d'origine

La prochaine variable personnelle qui va être étudiée est le niveau social d'origine de l'élève.

Tableau 7 – Variables contextuelles modulant l'effet du Niveau socioprofessionnel

Coefficient ≥ 0.10	Niveau de signification	Variable considérée
- 0.05	0.07	Classe nombreuse (effet compensatoire)
0.11	0.42	Classe majoritairement de filles
- 0.19	0.02	Classe d'expression majoritairement française (rurale)
- 0.05	0.05	Importance du décodage
0.05	0.07	Objectifs affectifs

Les deux variables de classes: *Proportion de filles* et *Expression majoritairement française*, n'ont pas été standardisées, ce qui triple environ leurs coefficients relativement à ceux des autres variables. Voir l'Annexe D.

On se souvient que l'effet direct de la variable Niveau social d'origine était très faible, son coefficient Bêta étant seulement de 0.02 et non significatif. Les résultats du tableau 7 donnent l'explication de ce phénomène au premier abord surprenant. Les effets d'interaction avec le contexte sont très importants et peuvent ainsi annuler toute tendance générale. L'origine sociale n'en intervient pas moins, mais donc de façon différente selon le contexte de la classe.

Effectif de la classe

L'effet du nombre d'élèves dans la classe sur le biais social qui s'y manifeste a déjà été relevé dans des recherches antérieures. Des classes nombreuses ont un effet compensateur vis-à-vis de l'effet socio-

culturel (ceci ressort également des coefficients du tableau 7. Le coefficient correcteur est négatif, égal à -0.05 . La pente du prédicteur est ainsi de $0.02 + (+1.00)(-0.05) = -0.03$, annulant l'influence classique du niveau socioculturel). La raison avancée pour expliquer ce fait est la moindre visibilité des caractéristiques personnelles dans une classe nombreuse. L'origine sociale des élèves disparaît aux yeux de l'enseignant-e et même aux yeux des camarades, si c'est un grand groupe qui doit se plier à la même discipline collective.

Proportion de filles

Dans notre étude, l'effet du niveau social des parents est renforcé dans la direction habituelle si la classe contient beaucoup de filles. Le succès attendu est surélevé de 13 points. Mais si elle contient beaucoup de garçons, l'effet socioprofessionnel courant s'inverse. Un milieu d'origine élevé ne procure plus alors d'avantage scolaire. Le coefficient correcteur est en effet dans ce cas $(-1) \times 0.11 = -0.11$. La pente est alors $0.02 - 0.11 = -0.09$ et elle contrebalance l'effet social connu. C'est peut-être cette interaction qui expliquerait un phénomène que les enseignant-e-s connaissent bien en parlant de « bonnes » et de « mauvaises » classes. Une proportion élevée de filles contribuerait à constituer une classe où les valeurs culturelles traditionnelles sont respectées, mais avec beaucoup de garçons, l'effet inverse pourrait s'observer.

Langue majoritaire

Si la classe comporte beaucoup d'enfants étrangers, il s'agit d'une classe de ville et l'effet du niveau social est très important, avec un coefficient de $(-1) \times (-0.19) = 0.19$ du niveau social pour prédire la réussite. Si l'on y ajoute la pente moyenne de 0.02, on obtient pour le niveau socioprofessionnel une pente de $+0.21$, la plus haute que l'on ait rencontrée, à part celle de la maturité scolaire. Par contre, dans une classe avec très peu d'étrangers on est en milieu rural. Or les agriculteurs sont codés 2 comme les employés. Il ne serait pas étonnant que dans un contexte social homogène, l'effet socioculturel habituel soit minimisé, voire inversé. C'est ce que suggère la valeur de la correction de pente dans les écoles de campagne. La pente est dans ce cas de $(+1) \times (-0.19)$. Avec une pente moyenne de 0.02, le coefficient de pente du niveau social est de -0.17 dans les classes sans élèves étrangers (au lieu de $+0.19$). Au vu de ces oppositions de tendance, on comprend que l'effet moyen du niveau social soit non significatif dans cette étude.

Importance du décodage

Un coefficient de correction de pente égal à -0.05 s'applique à la variable de classe correspondant à une didactique centrée sur le décodage des graphèmes. Un calcul analogue aux précédents montre que, dans les classes suivant cette didactique, la corrélation du niveau social avec la réussite est plus faible que dans les autres classes. Le rôle compensateur de cette option de l'enseignante qui vise d'abord à assurer chez ses élèves la maîtrise du décodage est ainsi apparu dans cette recherche comme dans d'autres études antérieures. Les milieux socialement défavorisés ont besoin d'une approche systématique avec suffisamment de répétitions pour sécuriser leurs apprentissages scolaires. Les milieux favorisés, par contre, réagissent mal à une méthode enseignant surtout des techniques, qui leur paraît trop fastidieuse. Ainsi le biais social de l'école peut être réduit si la méthode pédagogique suivie dans la classe est nettement centrée sur le décodage.

Objectifs affectifs

Au contraire, on peut imaginer une enseignante qui s'appuie sur les intérêts des enfants pour développer leur amour de la lecture : elle n'emploiera pas les mêmes démarches didactiques que ses collègues qui visent le décodage. On peut comprendre qu'elle favorise du même coup l'apprentissage des

enfants provenant des milieux sociaux supérieurs, comme le laisse penser le coefficient de correction de pente positif de 0.05 que reçoit le niveau social pour prédire la réussite dans ces classes. On peut calculer que cette didactique fait cependant baisser de 7 points le pronostic de réussite du niveau social inférieur alors qu'il serait amélioré de 3 points avec une didactique différente.

Régularité du parcours scolaire

Tableau 8 – Variables contextuelles modulant l'effet d'une scolarité régulière

Coefficient ≥ 0.10	Niveau de signification	Variable considérée
0.19	0.015	Classe d'expression majoritairement française (rurale)
0.05	0.039	Motivation interne
- 0.07	0.006	Objectifs affectifs
- 0.09	0.043	Bibliothèque dans la classe
0.12	0.002	Équipement
0.09	0.21	Approche par l'écrit

La variable de classe : Expression majoritairement française n'a pas été standardisée, ce qui triple environ son coefficient relativement à celui des autres variables. Voir l'Annexe D.

La pente moyenne de + 0.13 (sur l'ensemble des classes) pour le pronostic de réussite des élèves réguliers montre que le redoublement ne suffit pas à ramener les élèves qui ont répété la 1^{ère} année au niveau moyen de leurs camarades. Les élèves réguliers conservent un certain avantage. Les coefficients des interactions présentées au tableau 8, en permettant de corriger l'effet des facteurs personnels, révéleraient-ils au moins certaines caractéristiques des classes qui parviennent à compenser les handicaps des redoublants ?

Langue majoritaire

En tout cas, le facteur aggravant le plus évident est donné à la première ligne de ce tableau : c'est lorsque le redoublement s'est fait dans une classe où presque tous les enfants sont d'expression française. A la pente moyenne de + 0.13 favorisant les élèves réguliers s'ajoute alors le supplément de pente égal à + 0.19 multiplié par la valeur du taux de francophones dans la classe. Le pronostic de réussite en lecture des élèves non redoublants est alors nettement augmenté. A noter : le taux de francophones est une moyenne brute, non standardisée. Son écart type est de 0.36. C'est pourquoi le coefficient de correction de pente (+ 0.19) paraît si grand. Son importance réelle, en comparaison de celle des autres variables de ce tableau est de $(+ 0.19) \times (0.36) = + 0.07$ seulement. Les écarts types des variables utilisées sont généralement de + 1.00. Ils sont donnés à l'annexe D.

Pourquoi cet effet du contexte linguistique ? C'est qu'un redoublement est moins pénalisant s'il est effectué dans une classe comportant beaucoup d'enfants étrangers. La raison en apparaît facilement. Un enfant non francophone, surtout s'il arrive en cours d'année, peut être gêné dans sa lecture par son manque de connaissance du français. En refaisant sa première année, il a le temps d'apprendre la langue française et il peut ensuite rattraper ses camarades et poursuivre une scolarité normale. Son âge plus avancé peut même alors lui donner un certain avantage sur les autres. Au contraire, dans un milieu essentiellement francophone, il y a peu de chances que le redoublement soit simplement lié à la langue. Il est la marque d'autres problèmes qui risquent de subsister.

Motivation interne

La caractéristique contextuelle suivante pouvant modifier l'effet du redoublement est l'option de l'enseignante de s'appuyer sur la motivation interne de l'élève. Son coefficient de correction de pente se révèle positif (+ 0.05). Des calculs semblables à ceux des pages précédentes ne sont plus possibles,

parce qu'il n'existe plus de scores moyens de + 1.00 et - 1.00 pour simplifier les comparaisons. Pourtant une estimation de l'effet de cette option est calculable. Le score standard traduisant ce choix didactique est de + 1.90 et le rejet de cette approche est quantifié par - 3.58. Ces deux moyennes, multipliées par le coefficient de correction de + 0.05, donnent une augmentation de pente de 0.09 dans les classes ayant choisi cette option didactique, et une diminution de pente plus importante encore (- 0.18), pour les classes qui s'en tiennent à une approche plus rigoureuse. Dans ces classes, le total de la pente moyenne de + 0.13 et sa correction de - 0.18 donne une pente de - 0.05 qui peut avoir un effet compensateur pour les redoublants. Pourquoi ce probable échec compensatoire d'une pédagogie fondée sur la motivation interne? Une explication banale serait de dire que les élèves qui n'ont pas réussi à apprendre à lire en une année sont probablement découragés et qu'on ne peut pas compter sur leur seule motivation propre pour les faire dépasser leur échec.

Objectifs affectifs

Faut-il considérer comme différente l'approche qui vise en priorité des objectifs affectifs? En tout cas, elle semble donner plus de chance aux redoublants puisque son coefficient correcteur est de signe opposé. Considérons deux classes dont l'une se situe à 1 sigma au-dessus de la moyenne sur cette dimension affective et l'autre à la même distance en-dessous de la moyenne. Le choix de viser des objectifs affectifs amènera un coefficient correcteur de - 0.07 pour la première, alors que le handicap des élèves redoublants sera accru d'autant dans l'autre classe plus prosaïque. On sait l'importance de l'affectivité dans l'apprentissage. Tout dépend donc de l'atmosphère de travail positive qu'une enseignante expérimentée parvient à établir avec ses élèves en difficulté.

Bibliothèque

On vient d'évoquer le rôle possible de l'expérience de l'enseignante. Un lecteur adepte du « doute méthodique » pensera peut-être que c'est cette caractéristique qui se cache derrière le coefficient important du facteur « Bibliothèque ». Ce sont en effet les enseignantes avancées dans leur carrière qui, à l'époque de cette recherche, avaient eu le temps de constituer, avec leurs groupes d'élèves successifs, des bibliothèques de classe (qui ne figuraient pas dans le matériel fourni par l'État). Mais on peut tout aussi bien accepter ce résultat comme une sorte de contrôle expérimental de l'utilité d'avoir une bibliothèque de classe. Une classe située à + 1 sigma au-dessus de la moyenne sur cette variable de niveau 2 voyait le handicap de 13 points des élèves redoublants réduit à 4 points, ce qui est une amélioration considérable. L'aspect motivant de nombreux livres d'enfants doit notamment augmenter la motivation des élèves en difficulté.

Équipement

Le coefficient élevé, mais de signe opposé au précédent, du facteur Équipement montre bien que ce n'est pas le matériel qui importe, mais ce que l'on en fait. A l'époque de la recherche, seules de jeunes enseignantes sortant de l'École Normale avaient reçu du matériel pour l'initiation à la lecture. Elles pouvaient ne pas savoir encore bien l'utiliser. En tout cas, la preuve est faite ainsi que doter les classes de moyens d'enseignement ne suffit pas à compenser les difficultés des élèves en retard.

Centration sur l'écrit

L'approche classique de l'apprentissage de la lecture par l'écrit, c'est-à-dire par l'étude des lettres (plutôt que par celle des phonèmes, comme le proposent les linguistes) ne semble pas jouer systématiquement le rôle compensateur affirmé par les tenants des méthodes traditionnelles car le coefficient de correction de pente n'est pas du signe désiré. Pourtant bien des résultats des pages précédentes auraient fait attendre un résultat plus favorable. On voit ainsi que la complexité de la didactique de la lecture exclut toute recette, fût-elle soutenue par son usage courant dans le passé.

Niveau d'aptitude au départ

Tableau 9 – Variables modulant l'effet du niveau d'aptitude scolaire donné au départ

Coefficient ≥ 0.10	Niveau de signification	Variable considérée
- 0.14	0.021	Niveau d'aptitude de départ de la classe (effet compensatoire)
0.06	0.079	Classe à degrés multiples (deux degrés)
- 0.05	0.078	Importance du décodage
- 0.06	0.057	Motivation externe
- 0.04	0.063	Motivation interne
0.08	0.012	Présence d'une bibliothèque
- 0.04	0.088	Approche par Exercice des prérequis
- 0.05	0.037	Satisfaction de l'enseignante pour sa méthode

La variable: « Niveau d'aptitude de départ de la classe » n'a pas été standardisée, ce qui double environ son coefficient relativement à celui des autres variables. Voir l'Annexe D.

Le tableau 9 traite du dernier des prédicteurs individuels qui manifestaient au tableau 5 des variances résiduelles de pente significatives, les trois autres ayant été traitées aux tableaux 6, 7 et 8. La longue liste de variables de niveau 2 (relatives à la classe) que présente ce tableau 9 montre que le niveau de départ d'un élève détermine sa réussite de façon très complexe, en interaction avec tout son entourage pédagogique et social.

Niveau de la classe

On sait que le niveau d'aptitude scolaire donné au départ a une forte influence sur la réussite en lecture, puisque le coefficient de pente moyen est de 0.52. Il s'agit là cependant encore d'une relation purement interne à la classe. Le niveau moyen de départ de la classe vient ajouter son influence, et cela de deux façons. D'abord directement, en ajoutant une constante à la moyenne des scores de réussite de la classe, sur la base des poids donnés aux caractéristiques moyennes des classes, selon le tableau 4a. Mais une influence indirecte intervient également, en modifiant la pente de certaines variables individuelles, comme l'ont montré les tableaux précédents.

Le tableau 9 montre d'abord justement comment la variable de niveau 2 Niveau moyen de départ de la classe peut modifier le coefficient appliqué dans chaque classe à la variable individuelle Niveau de départ. Le coefficient de l'interaction est de - 0.14. La modulation de pente qui en résulte peut jouer ainsi le rôle d'un correctif, car elle réduit l'influence du niveau personnel d'aptitude au départ, dans les classes composées d'enfants en moyenne déjà bien avancés.

En effet, imaginons une classe de haut niveau, se situant à + 1 sur l'échelle des moyennes de classe en niveau d'aptitude. Un correctif de pente est à ajouter dans cette classe à la pente moyenne de 0.52 (qui vaut dans toutes les classes) pour pondérer le score individuel d'aptitude. Le terme de correction lié à cette classe est de $(- 0.14) \times (+1)$, donc - 0.14. Le poids du prédicteur individuel « Niveau de départ » dans cette classe est donc de $0.52 + (- 0.14) = 0.38$. La réussite est moins déterminée par le niveau de départ, dans cette « bonne » classe, que s'il s'agissait d'une classe moyenne, où la pente serait de 0.52.

Imaginons maintenant une classe de niveau faible, se situant à -1 sur l'échelle des moyennes de classe en capacité initiale. Le terme de correction lié à cette classe est de $(- 0.14) \times (-1)$, donc + 0.14. Le poids du prédicteur individuel Niveau d'aptitude va être alors encore plus grand que dans une classe moyenne, puisqu'il va être égal à $0.52 + 0.14 = 0.66$.

On voit ainsi que le pronostic de réussite est davantage lié au niveau d'aptitude initial d'un élève si ce dernier est dans une classe faible que s'il est dans une classe forte. Ainsi les classes de haut niveau sont davantage préservées du biais d'aptitude initiale.¹

Les coefficients des autres variables modulant l'effet du niveau d'aptitude personnel sont tous du même ordre de grandeur. Ils seront discutés dans l'ordre où ils apparaissent dans le tableau 9.

Classe à degrés multiples

Les classes véritablement à degrés multiples ont été exclues de cette étude, parce qu'elles introduisaient tant de caractéristiques particulières qu'elles auraient rendu difficilement interprétables plusieurs des relations trouvées (avec l'effectif de la classe, par exemple). Par contre, les classes à deux degrés ont été conservées, du fait de leur fréquence notamment, qui rendait possible d'en apprécier l'effet.

Du fait que cette caractéristique de classe n'est pas apparue au tableau 4b, on a pu voir que cette variable n'a pas d'effet moyen significatif. Les classes à degrés multiples ne sont donc ni favorisées, ni défavorisées en moyenne.

Par contre, le tableau 9 révèle un coefficient positif, qui accentue donc l'importance moyenne dans ces classes du niveau d'aptitude scolaire initial pour la réussite de l'apprentissage de la lecture. Le coefficient de 0.06 multiplie le score sur l'échelle standardisée opposant les classes à degrés multiples aux classes normales. Du fait de la rareté des classes à degrés multiples, ce score est de 2.68 pour les classes à degrés multiples, et de - 0.37 pour les classes à un seul degré (une échelle standardisée doit en effet avoir une moyenne de zéro et un écart type de 1). Le coefficient de pente ajouté pour les premières est donc $2.68 \times 0.06 = 0.16$. Le coefficient ajouté pour les secondes est de $(- 0.37) \times (0.06) = - 0.02$.

On voit ainsi que, si les classes ordinaires sont peu concernées parce qu'elles représentent la grande majorité, le poids du niveau d'aptitude personnel dans les classes à degrés multiples devient très important. A la pente moyenne de 0.52 s'ajoute en effet 0.16, ce qui donne un coefficient prédictif de 0.68 au niveau d'aptitude. Ainsi, dans ces classes, la variance du résultat de l'apprentissage est alors déterminée à 46 % (0.68 au carré) par le niveau de départ personnel de l'élève, ce qui est énorme et laisse peu de place à l'intervention de l'école si l'on tient compte de tous les facteurs incontrôlables qui interviennent dans une carrière scolaire.

L'explication de cette relation est facile à trouver. Lorsque la maîtresse doit se partager entre deux groupes d'élèves, elle doit laisser chacun d'eux travailler de façon autonome la moitié du temps, ou en tout cas beaucoup plus que dans les classes habituelles. Cette autonomie profite à ceux qui ont le profil scolaire au départ et pénalise ceux qui ne l'ont pas, qui ont du mal à travailler indépendamment. On comprend que les écarts entre élèves se creusent dans ces conditions et renforcent leur classement initial.

Apprentissage du code

La variable suivante, « Importance du décodage », a été obtenue en totalisant une série de rubriques du Questionnaire Enseignantes, dans lequel les maîtresses décrivaient leur méthode didactique et les principes qui guidaient leur pratique. Celles qui avaient un score élevé disaient se centrer sur l'apprentissage des lettres et de l'oralisation de groupes de lettres, en passant donc moins de temps à des

¹ Notons toutefois que cet effet automatique d'atténuation n'est pas aussi important qu'il le semble, parce que les moyennes de classe pour lesquelles le coefficient de - 0.14 est valable sont les moyennes brutes et non des moyennes standardisées. Leur écart-type (voir annexe D) est de 0.46, soit moitié moins grand que pour une variable standardisée. Le cas des deux classes hypothétiques ci-dessus serait ainsi tout à fait exceptionnel. Les corrections de pente typiques sont moitié moins grandes que les valeurs indiquées ci-dessus. Le poids relatif du niveau moyen de départ de la classe pour moduler les pentes des variables personnelles est en fait du même ordre de grandeur que celui des autres variables de niveau 2.

exercices préparatoires de latéralisation ou d'analyse de sons, par exemple, et en se souciant moins de motiver leurs élèves ou de développer leurs capacités d'expression.

Cette pratique rigoureuse semble relativement payante pour une politique compensatoire, puisque le coefficient est négatif et égal à -0.05 . Ainsi, dans les classes où la maîtresse a choisi cette politique, le niveau de départ a moins d'importance, devenant $0.52 - 0.05 = 0.47$. Dans les classes où règne une politique opposée, le poids du niveau de départ est de $0.52 + 0.05 = 0.57$. L'objectif généralement admis, et même mis en préambule de certaines lois scolaires, de réduire les inégalités sociales face à l'éducation, oblige à prendre en compte sérieusement cette centration sur le décodage. Mais cette approche du problème n'est pas la seule possible et d'autres pratiques peuvent aussi contribuer à plus d'équité.

Motivation externe et interne

Les variables Motivation externe et Motivation interne, malgré leur caractère assez contradictoire, permettent toutes les deux des effets compensatoires, d'après leurs coefficients respectifs de -0.06 et de -0.04 . Toutes deux contribuent à une implication plus grande des élèves, la première en leur offrant des récompenses pour leur travail, la seconde en cherchant des thèmes qui éveillent leur intérêt. La dimension intellectuelle de l'apprentissage est alors un peu réduite, au profit de ses déterminants affectifs et l'équité en bénéficie.

Bibliothèque

Par contre, la présence d'une bibliothèque dans la classe renforce, assez logiquement, la prééminence de l'aptitude au travail scolaire (coefficient de 0.08), puisque les élèves qui en tireront des livres seront ceux qui pourront les lire, ou qui auront déjà perçu la valeur de l'écrit pour communiquer. Il est difficile dans une classe de ne pas renforcer davantage les enfants qui sont déjà engagés avec succès dans un apprentissage scolaire. Pour obtenir plus d'équité, on ne doit pas les décourager, mais trouver d'autres chemins pour motiver les autres.

Prérequis

Une approche de la lecture par l'Exercice des Prérequis souhaiterait justement assurer des apprentissages pré-scolaires sous une forme ludique mais qui faciliterait ensuite l'acquisition des mécanismes de la lecture. Le niveau d'aptitude de départ, dont on a vu le caractère déterminant, mesure jusqu'à quel degré ces apprentissages généraux sont déjà acquis. Mais rien n'oblige à considérer ce niveau de départ comme donné génétiquement et immuable. Ces compétences doivent au contraire pouvoir être acquises. Elles résultent en effet d'apprentissages cognitifs et affectifs qui devraient pouvoir être réalisés à l'école tout aussi bien que dans la famille. Il s'agirait alors véritablement d'éducation compensatoire. C'est ce que certaines enseignantes ont essayé de réaliser dans leur classe en début d'année, notamment pour compenser l'absence d'école enfantine dans leur région. La recherche montre, par le coefficient de -0.04 , que ces efforts ont eu un résultat. Dans les classes qui ont bénéficié de ces pratiques, le caractère déterminant du niveau de préparation des élèves en début d'année a été assoupli.

Satisfaction de l'enseignante

C'est sans doute aussi une des raisons de la satisfaction des enseignantes pour la méthode qu'elles ont suivie. Le coefficient négatif trouvé montre que dans les classes des maîtresses satisfaites, la hiérarchie des compétences des enfants au départ est moins déterminante que dans les autres classes. C'est donc qu'elle a été modifiée en cours d'année. On peut penser que c'est du fait justement des apprentissages réalisés, puisque la valeur prédictive du niveau initial était supérieure dans les classes n'appliquant pas ces méthodes.

Le grand nombre d'effets d'interaction statistiquement significatifs que l'on a trouvés ci-dessus explique en partie qu'aucune méthode d'enseignement de la lecture ne se soit imposée de façon générale. Des groupes d'élèves aux caractéristiques différentes profitent davantage d'une méthode ou d'une autre. Il restera à voir si la connaissance de ces effets peut être mise à profit didactiquement.

Quantification des effets

Agir oblige toujours à choisir des priorités, en comparant avantages et inconvénients. Il faut savoir si tel effet l'emporte sur tel autre. Quantifier n'est donc pas seulement un souci de théoricien. Cela peut être essentiel pour la pratique, y compris le choix de didactiques.

L'avantage de la méthode multiniveau est d'apporter justement beaucoup d'informations comparatives, surtout en termes de parts de la variance totale des résultats au critère qui peuvent être attribuées à telle ou telle source d'influence.

Il n'est pas possible dans un rapport global de détailler toutes les sources de variance. Mais on peut estimer l'apport de groupes de variables en décomposant la variance totale du critère.

Celle-ci vaut 1.00 puisque cette variable est standardisée. Si l'on fait une analyse de la variance simple, on obtient les estimations de deux composantes de variance essentielles :

- Variance Intraclasse = $R = 0.7310$
- Variance Interclasse = $U0 = 0.2664$

On voit que les différences de rendement moyen entre classes rendent compte de 26 % de la variance totale. C'est énorme, surtout si l'on tient compte du fait que i) les classes à degrés multiples ou les classes trop peu nombreuses ont été exclues de l'étude et qu'il s'agit donc de classes normales, *a priori* homogènes; ii) l'estimation de la variance interclasse a déjà été corrigée de l'effet des fluctuations aléatoires provenant de l'échantillonnage des élèves.

Mais on sait que les élèves peuvent être très différents et provenir de milieux sociaux très contrastés. Pour égaliser en partie le recrutement, on peut prendre en compte chaque caractéristique personnelle des enfants entrant à l'école et, à l'aide d'une équation de régression multiple, retirer de sa réussite au sein de la classe en fin d'année ce qui provient de lui-même et de son milieu familial. Ceci ne peut pas changer la variance entre les classes, puisqu'on travaille classe par classe, mais ceci réduit de beaucoup la variance attendue des résultats à l'intérieur des classes. De 0.7310 cette variance tombe à 0.4660. Ainsi, la variance des résultats de fin d'année dans chaque classe est déjà explicable en moyenne à 36 % $(0.7310 - 0.4660)/(0.7310)$ par les caractéristiques personnelles des élèves en début d'année.

Si l'on cherche maintenant à expliquer les différences de résultats entre classes en fin d'année par ce que l'on sait des caractéristiques de ces classes (moyennes des caractéristiques personnelles de leurs élèves) et des caractéristiques de leurs enseignant-e-s (conceptions pédagogiques et didactiques), on peut réduire la variance entre classes de 0.2664 à 0.1374. On explique alors $(0.2664 - 0.1374) / 0.2664 = 48 \%$ de la variance des résultats entre classes.

Ce n'est pas tout, puisqu'on peut encore ajuster dans chaque classe les pentes des droites de régression pour les variables personnelles des élèves. On arrive (voir l'annexe C) à une estimation de la variance intraclasse du critère égale à 0.4352 et à une estimation de la variance des moyennes de classes égale à 0.1066. On explique alors 40 % $(0.7310 - 0.4352)/(0.7310)$ des différences de résultats à l'intérieur des classes et $(0.2664 - 0.1066) / 0.2664 = 60 \%$ de la variance de résultats entre classes. Il semble qu'avec ce degré de maîtrise d'un phénomène, on devrait pouvoir agir...

DISCUSSION DES RÉSULTATS : APPORTS DE LA RECHERCHE QUANTITATIVE

La phrase conclusive du précédent chapitre se veut provocatrice. Est-elle justifiée ? Les résultats présentés sont-ils réellement probants ? Que montrent-ils au fond ? Il faut prendre un peu de recul pour discuter cette recherche.

La valeur prédictive des variables personnelles est frappante, mais c'est un fait bien connu que dans un échantillon représentatif de la population générale, toutes les mesures d'aptitude et de performance sont très corrélées. Dans le domaine scolaire, en particulier, on peut vraiment parler d'un facteur général, puisque c'est sur ce type d'épreuves que le psychologue anglais Spearman a démontré l'existence de cette source d'influence commune. La stabilité du classement des élèves n'est même que trop connue, puisqu'elle crée les échecs scolaires à répétition et les problèmes sociaux qui en découlent. L'intérêt de la présente recherche serait de montrer, comme au tableau 9, que certains contextes scolaires ont un effet compensateur du niveau général. Dans certaines conditions, les capacités cachées d'un élève peuvent se révéler. Le fait que les coefficients de pente correspondants soient tous inférieurs ou égaux à 0.06 montre cependant l'étroitesse de cette fenêtre d'espoir.

La valeur prédictive des particularités de la classe et des orientations pédagogiques des enseignantes est elle aussi frappante. Mais elle provient autant des caractéristiques moyennes du groupe classe, sur lesquelles l'école ne peut pas agir, que des choix de méthode de la maîtresse. On a vu que le nombre de variables qui prédisaient un bonus ou surplus de réussite pour la moyenne de la classe était étonnamment faible, ruinant l'espoir de pouvoir définir la « bonne » façon d'enseigner la lecture. Tout au plus a-t-on vu que diverses approches pouvaient avoir chacune leur intérêt.

Le résultat nouveau de cette recherche est de mettre le doigt sur la cause probable de ces échecs. Un grand nombre d'interactions sont en effet apparues, montrant que la bonne méthode n'était pas la même selon le type d'élève et selon le contexte de la classe, ce qui a pour effet d'empêcher de parvenir aux généralisations que l'on souhaitait trouver.

Peut-être une certaine forme de généralisation s'impose-t-elle néanmoins : c'est de voir que la prise en compte des diverses particularités de chaque élève peut conduire, pour ce qui le concerne, à des améliorations de pronostic dépassant largement l'avantage attendu d'une méthode d'enseignement particulière. Reformulée en termes simples, cette généralisation revient à dire que les facteurs psychologiques individuels jouent un rôle plus grand dans la réussite que les procédures didactiques collectives.

Limites de l'interprétation

Une telle assertion est bien hardie sur la base des chiffres présentés plus haut.

D'abord elle saute à une conclusion générale sur la base de quelques relations qui étaient apparues comme statistiquement significatives. Elle ne représente en fait qu'une intuition. Seuls les résultats détaillés qui semblent la fonder peuvent faire l'objet d'une évaluation. Que peut-on penser d'eux ?

Le lecteur aura sans doute objecté, dès la présentation des données, que les conditions sociales qui prévalaient dans le canton de Neuchâtel en 1973 ne sont plus les mêmes aujourd'hui. Cela est particulièrement vrai en ce qui concerne l'immigration étrangère et aussi l'école enfantine, qui est

maintenant généralisée. L'interprétation des résultats a tenté d'en tenir compte, mais la portée des conclusions reste incertaine.

Une autre source de questionnement est le choix du critère. On a vu que la recherche avait été d'abord menée sur deux ans, puis qu'il avait semblé préférable de s'en tenir aux résultats de fin de première année, pour ne pas perdre de vue les élèves qui n'avaient pas suivi leur classe en deuxième année. Ce choix avait son coût : c'est qu'on mesurait des résultats à court terme, alors que beaucoup de méthodes visent au contraire à développer des compétences générales aux bénéfices plus lointains. Ce sera au lecteur d'apprécier les nuances à apporter aux résultats de ce point de vue.

Enfin, il a été rappelé plusieurs fois dans les pages précédentes que les seuils de signification choisis laissaient place à un nombre appréciable de fluctuations purement aléatoires. Puisque plusieurs analyses statistiques successives ont été menées, par rapport à des critères distincts (Compréhension, Exactitude, et Rapidité de la lecture), il est apparu de façon manifeste que des variables qui étaient en vedette dans un cas n'apparaissaient plus comme importantes dans une autre analyse. On voit donc que certains résultats sont difficilement généralisables.

Cohérence globale rassurante

Cependant, si on accepte de ne pas s'attacher à une variable particulière, mais que l'on suit des orientations plus larges, on retrouve bien un effet commun identique pour les différentes variables caractérisant une approche pédagogique donnée. Par exemple, on a vu apparaître la valeur de l'option Lettres, ou celle de l'Importance du décodage, ou bien celle de l'Approche écrite, mais il s'agissait toujours d'une méthode syllabique, qui assurait un apprentissage plus rapide à court terme, compensatoire des inégalités sociales, mais sans qu'on puisse se prononcer sur sa valeur à long terme ou pour l'ensemble des élèves.

De même les maîtresses qui optaient pour développer une motivation interne, viser des objectifs affectifs et un développement intellectuel global (objectif élevés) avaient des résultats favorables. Ces méthodes jouaient aussi un rôle compensatoire vis-à-vis du niveau d'aptitude initial. D'une certaine façon, les résultats de cet autre faisceau de variables pédagogiques se confirmaient dans leur tendance générale. C'est ce qui peut donner confiance dans les résultats présentés.

On pourrait aussi parler d'une confirmation approximative semblable si l'on considère les résultats obtenus avec les trois critères Compréhension, Exactitude et Vitesse. C'est en voyant la similitude de leurs résultats qu'il a été décidé de les réunir et d'en faire la moyenne pour évaluer la compétence globale des élèves en lecture.

Enfin, on peut rappeler que les phases antérieures de cette recherche, rappelées au début de ce texte, ont porté sur des classes en partie différentes, sans que des contradictions apparaissent d'un échantillon à l'autre. Il s'agit, là encore, de confirmations partielles, même si on ne peut parler vraiment de contre validation.

Les résultats des premières analyses dépendent des données utilisées, elles-mêmes influencées par les procédures de remplacement choisies pour les données manquantes. Pour permettre à des chercheurs de procéder à d'éventuelles réanalyses fondées sur d'autres options de remplacement, les divers fichiers de données ont été conservés, avec les rapports correspondants, cités dans les pages précédentes. Ils ont été mis sur le cédérom « DonnéesLecture73 » qui reste disponible à la Documentation de l'IRD à Neuchâtel.

DISCUSSION DES RÉSULTATS À LA LUMIÈRE DE RECHERCHES RÉCENTES SUR LA LECTURE

Une réanalyse de données vieilles de plus de 35 ans pose des problèmes particuliers. Bien entendu, d'une part les recherches effectuées durant ce temps dans ce domaine ont apporté des connaissances supplémentaires, d'autre part les programmes, les moyens d'enseignement de lecture, les conditions de l'enseignement/apprentissage et les pratiques en classe ont changé.

Nous allons donc nous demander, dans ce qui suit, dans quelle mesure les résultats obtenus par l'analyse multiniveau gardent un sens en regard de ce que les recherches en didactique de la lecture nous ont appris aujourd'hui. Mais nous devons choisir un cadre de référence plus large, car les études multiniveau s'inscrivent dans l'ensemble des recherches de type processus-produit. Nous nous appuyerons donc dans notre évaluation sur les descriptions et les bilans publiés par Crahay (2006 ; 2007) concernant ce type de recherche. Nous verrons plus clairement ainsi leurs apports et leurs limites. Nous tenterons en conclusion de proposer une approche scientifique plus englobante.

Rappelons au départ le cadre méthodologique qu'il s'agit de discuter. L'analyse multiniveau se situe clairement dans les recherches processus-produit, qui ont connu leur point culminant précisément dans les années 1970, aux États-Unis en particulier. C'est à ce moment-là qu'a eu lieu la récolte des premières données de la recherche de l'IRDIP sur la lecture. Ces recherches sont une tentative de déterminer de façon scientifiquement fondée ce qui fait qu'un enseignement est efficace pour l'apprentissage. Dans ce sens, elles accordent une importance non négligeable à l'enseignant-e et à ses caractéristiques personnelles, desquelles dépendrait, en grande partie du moins, la réussite des apprentissages de ses élèves. Elles ont mis en évidence des observables tels que la quantité et la nature des *feedbacks*, le temps accordé par l'enseignant-e aux élèves les plus faibles, le degré de structuration de la leçon, la clarté des objectifs visés par les activités, les modalités de travail en classe, etc. En lecture, la place donnée au travail sur le code, la progression des activités, la part du travail collectif et du travail en groupes figurent parmi les variables prises en considération.

Comme indiqué précédemment, l'apport certainement le plus important de l'analyse multiniveau moderne par rapport aux recherches des années 1970 est le traitement de l'interaction entre effets individuels et effets collectifs définis dans l'étude. Dans notre recherche, plusieurs faisceaux de variables d'enseignement ont pu être couplés avec des variables liées aux élèves (genre, âge, milieu socioéconomique, langue parlée à la maison, etc.). Ce traitement de l'interaction a entre autres pour avantage de nuancer l'apport des variables pédagogiques et contextuelles prises en considération, comme celles que l'on vient d'énumérer, selon les liens qu'elles peuvent avoir entre elles.

Un des résultats principaux mis en évidence est que les configurations ainsi obtenues ne font pas apparaître « la bonne façon » d'enseigner la lecture ou « la bonne méthode », mais laissent voir des approches diverses à même de favoriser à différents degrés l'apprentissage de la lecture chez de jeunes élèves.

Nous allons discuter d'abord l'effet des variables prises isolément, laissant pour la partie suivante la discussion des interactions entre ces variables.

Des variables pédagogiques majeures

En un premier temps, en partant du résultat le plus significatif sur un plan didactique, montrant l'influence des variables pédagogiques sur l'apprentissage de la lecture (tableau 4b), nous allons discuter les points suivants :

- le décodage, l'apprentissage des lettres et l'approche écrite ;
- la motivation interne, les objectifs affectifs ;
- la vitesse d'avancement de l'enseignement/apprentissage ;
- l'interprétation de ces variables en termes de processus généraux.

En effet, pour rappel, le tableau 4b mettait en avant tout particulièrement une relation significative entre la motivation interne (dont nous redirons ce qu'elle recouvre), la rapidité d'avancement de l'apprentissage, mais aussi, dans une certaine mesure, l'option lettres (impliquant une suite d'étapes structurée privilégiant un enseignement de l'écrit par apprentissage des lettres, des mots puis des phrases) et la réussite de l'apprentissage de la lecture. A cela s'ajoute la satisfaction de l'enseignante pour sa méthode, dont nous dirons également quelques mots.

Ces variables sont précisément des éléments de l'enseignement devant garantir un meilleur apprentissage des élèves, donc des constituants de l'efficacité de l'enseignement. Que sait-on à ce sujet aujourd'hui ?

Décodage, apprentissage des lettres et approche écrite

L'analyse multiniveau a révélé que « l'option lettres » était un facteur pédagogique prédictif important pour la réussite en lecture. Dans le contexte du début des années 1970, ce facteur signifiait une approche progressive des difficultés, par synthèse successive des petites unités (lettre, syllabe, mot), centrée donc essentiellement sur le décodage et prioritairement axée sur l'écrit.

Les recherches sur la lecture, sur ce qu'elle est, avec ses composantes, mais aussi sur son enseignement et son apprentissage sont nombreuses, menées depuis longtemps, et se sont poursuivies durant ces dernières décennies. Investies majoritairement par les psychologies cognitives, elles ont surtout permis de mettre en avant les composantes cognitives de la lecture. Il existe toutefois des recherches qui touchent aux aspects affectifs, sociaux et culturels de la lecture et de son enseignement/apprentissage.

Il est certain, au-delà des divergences qui subsistent encore entre les chercheurs se réclamant de courants épistémologiques différents, que, au cours de ces dernières décennies, les conceptions de la lecture et de son enseignement/apprentissage ont changé, aussi bien dans le monde francophone que dans les pays anglo-saxons. La lecture est vue comme un processus complexe et multidimensionnel qui vise la compréhension, l'interprétation de textes. Grossièrement dit, d'un modèle considérant la compréhension écrite comme essentiellement la capacité d'extraire et de reconstruire l'information contenue dans un texte (généralement narratif), on a passé à des modèles plus complexes faisant place à l'interaction entre le lecteur et le texte, la compréhension étant un processus dynamique de construction de représentations, d'inférences à de multiples niveaux du texte, mais aussi du contexte. Avec le temps, la conception de la compréhension de la lecture s'est élargie aussi sous l'effet de la multiplication des textes (de genres différents) et de la diversité des situations de lecture².

Etant donné la complexité de cet acte qu'est lire et comprendre des textes, comment envisager son apprentissage ? Si le choix de la bonne méthode n'est plus à l'ordre du jour aujourd'hui (Goigoux,

² Pour une vue d'ensemble bien documentée et récente, se reporter au dossier élaboré par Feyfant et Gausse, 2007, accessible sur le site de l'INRP.

2006; Rieben, 2004; 2005; Chauveau, 2007³), cela ne signifie cependant pas qu'on ne peut rien dire sur la façon de procéder en classe, lors des premiers apprentissages en particulier. Selon Rieben (2004), depuis les années 1970 il a été démontré que l'automatisation de la reconnaissance des mots est nécessaire pour la compréhension (l'inverse n'étant pas vrai). Cette auteure décrit trois composantes permettant à l'enfant d'identifier des mots: la compréhension du principe alphabétique, la connaissance des lettres et la conscience phonologique. Là encore, Rieben affirme que les recherches entreprises depuis plus de vingt ans attestent de l'importance de la conscience phonologique dans l'apprentissage de la lecture. Mais elle n'en fait pas un prérequis de la lecture; un travail sur la connaissance des lettres peut en effet se faire parallèlement. De plus, conscience phonologique et connaissance des lettres peuvent aller de pair avec des activités en situation complexe de lecture et d'écriture, dans une perspective communicative. Cependant, ces constats remettent en question le « devinement » qui a connu un certain succès dans les années 1980 lorsqu'il a été jusqu'à constituer parfois un objectif pédagogique. S'il est tout au plus possible de le considérer comme une stratégie que peuvent utiliser les débutants, au vu des erreurs auxquelles il conduit souvent il est indispensable de ne pas construire l'apprentissage sur cette base, mais de donner aux enfants le moyen de savoir déchiffrer les mots.

Remarquons ici qu'il s'agit d'une entrée dans l'écrit couplant lecture et écriture, décrite finement par Saada-Robert et ses collaboratrices (2003) par exemple, qui sont parties d'observations faites dans des classes où avaient lieu les premiers apprentissages de l'écrit. Elles ont finalement montré une évolution quasi parallèle dans les résultats en lecture et en écriture. De même, un travail alliant décodage et compréhension de textes dans la perspective communicative citée par Rieben était déjà recommandé par les auteurs de *Maîtrise du français* (Besson et al., 1979), qui, par ailleurs, y avaient développé toute une série d'ateliers sur la reconnaissance phonologique. Goigoux et Cèbe (2006) ne voient pas non plus d'antagonisme entre déchiffrement et compréhension, l'un étant la condition de l'autre, condition nécessaire mais non suffisante. Selon ces auteurs, il s'agit de trouver un équilibre entre plusieurs composantes: l'accès à la culture écrite, la compréhension de textes, la compréhension de phrases, la lecture de mots, l'écriture de mots et la production de textes. On retrouve donc ici le couple lecture-écriture.

Selon Giasson (2007), et à l'instar de ce qui précède, on a passé d'une appréhension traditionnelle de la compréhension de la lecture, envisagée comme un ensemble de sous-habiletés à acquérir les unes après les autres de façon hiérarchique, à une approche plus globale qui met en interaction ces habiletés, chacune étant influencée par les autres tout en les influençant à son tour. Aujourd'hui, on est d'accord pour dire que le lecteur est créateur du sens du texte. Pour cela, il s'appuie sur le texte, sur ses propres connaissances et sur son intention de lecture. S'ajoute encore la variable du contexte de lecture.

L'entrée dans l'écrit est déclinée comme telle dans le plan d'études romand actuel (PER). Selon les propositions de Saada-Robert et de ses collaboratrices, la progression est définie comme spiralaire, puisque le point de départ de l'apprentissage est d'emblée ancré dans des situations qui mettent en réseau un ensemble de composantes d'un même savoir, en évitant ainsi de décortiquer celui-ci en unités présentées de façon juxtaposée, comme apparemment simples et additives. Si bien que les élèves, d'entrée, travaillent les petites unités que sont la lettre et le son, la syllabe et le mot, et les grandes unités que sont les textes (à lire et à écrire).

Il semblerait donc qu'aujourd'hui les exigences de l'enseignement des premiers pas dans la lecture soient plus élevées, dans la mesure où les enseignant-e-s doivent mener en parallèle des activités portant à la fois sur les petites et les grandes unités, ce qui rend la question de la progression des

3 Chauveau insiste sur le fait que de nombreux facteurs externes agissent sur l'efficacité des méthodes et que le rôle des maîtres et de leur pédagogie au quotidien est primordial.

apprentissages plus complexe. Il ne suffit donc plus de suivre une méthode préconisant d'aller du simple au complexe, mais de maîtriser plusieurs approches, à moduler en fonction des élèves et de leurs difficultés. Cela implique aussi de la part des enseignant-e-s de donner aux jeunes élèves entrant dans l'écrit un accès aux livres, non seulement en leur laissant le plaisir de les découvrir, mais surtout en leur donnant des outils pour en saisir le sens.

La motivation interne, les objectifs affectifs

Dans les résultats de l'analyse multiniveau, la variable pédagogique appelée Motivation interne a été mise en avant comme bonne prédictrice de la réussite en lecture. En fait, c'est en particulier ce qui touche à la lecture elle-même, à l'intérêt de son apprentissage, qui entre surtout en ligne de compte dans ce facteur. Un enseignement de la lecture significatif pour les enfants, le déchiffrement appréhendé comme une technique qui va servir pour pouvoir lire des textes et les comprendre, accroîtraient la motivation des élèves.

Si nous avons insisté sur les aspects cognitifs de la lecture et de la compréhension de textes, certes au cœur des travaux de recherche dans ce domaine, nul ne défend l'idée que la lecture se limiterait à ces aspects. Les dimensions affective, sociale et culturelle entrent également en jeu dans l'acte de lire.

Même si la recherche PISA concerne les élèves de fin de scolarité, elle montre, pour la littératie, l'importance de ces variables sur les résultats des élèves de 9^e année en Suisse romande (Nidegger (Coord.), 2001).

« On constate que le plaisir de lire dépend dans une certaine mesure de caractéristiques individuelles ou de contexte, comme la pratique de la langue (langue parlée à la maison), le contexte socio-économique et le genre des élèves. Les filles ont en moyenne davantage de plaisir à lire que les garçons et les élèves des catégories sociales les plus favorisées davantage que les élèves des autres catégories » (p.146).

Terwagne, Vanhulle et Lafontaine (2001) soulignent, de leur côté, que nombre de recherches ont mis en évidence les discussions avec les pairs et l'enseignant-e comme un facteur important de motivation des élèves à la lecture. Citant Gambrell (1996), ils relèvent plus précisément six facteurs de motivation à la lecture : un-e enseignant-e modèle explicite de lecteur ; un environnement de classe riche en livres ; l'occasion donnée aux élèves de choisir un livre à lire ; des occasions d'interagir socialement avec les autres ; des occasions de se familiariser avec de nombreux livres ; des incitants appropriés liés à la lecture (recevoir un livre en cadeau ou en prix par exemple). Plus encore, ils montrent que certaines attitudes des enseignant-e-s favorisent la motivation de leurs élèves : l'organisation des apprentissages à travers des interactions ; l'explicitation des objectifs poursuivis, la sollicitation des élèves à mettre en avant les connaissances acquises et les stratégies indispensables à la compréhension ; la proposition de défis, de tâches complexes nécessitant la réflexion et les échanges dans la classe. Les cercles de lecture⁴ qu'ils ont développés prennent en compte l'ensemble de ces facteurs. D'autres dispositifs se développent, dans ce même souci de favoriser la motivation à lire des enfants : la Bataille des livres, le Rallye-lecture, le Sac à histoires, par exemple.

Les auteurs de *Maîtrise du français* n'avaient pas négligé l'importance de motiver les élèves à la lecture. Ce sont eux qui ont prôné par exemple la mise en place dans la classe d'un coin-bibliothèque permettant aux élèves d'accéder à des livres de leur choix pour une lecture libre, à divers moments de la journée. Ils proposent et illustrent des activités de correspondance scolaire et de composition de

4 Le cercle de lecture est un dispositif didactique structuré proposant un travail en groupes d'élèves hétérogènes d'interprétation et de construction de connaissances à partir de textes littéraires ou d'idées. Les élèves, au fil des interactions en classe, s'approprient des stratégies fines de compréhension et d'interprétation. Le rôle de l'enseignant-e est primordial : organiser, gérer, animer, étayer les propositions des enfants, est une autre façon d'enseigner.

livres. Pour eux, il s'agit de mettre le plus souvent possible les élèves en contact avec les livres, dès les débuts de l'apprentissage de la lecture.

La question de la motivation est délicate. A voir les exemples ci-dessus, ce serait en proposant aux élèves des activités significatives, proches des pratiques sociales de référence (Fijalkow (2000) ou, selon Mauffrey et Cohen (1995) en envisageant la lecture dans son fonctionnement authentique, et en suscitant les échanges entre élèves et avec l'enseignant-e, voire en plaçant les élèves face à des tâches complexes, que l'école développerait la motivation de ces derniers. Il reste cependant à la fois à définir ce qu'on entend par motivation (plaisir de lire, envie d'apprendre, envie de réussir les tests?), à l'opérationnaliser et à l'évaluer, sachant que la motivation ne prend pas le même sens selon la culture et l'origine sociale de l'enfant.

Vitesse d'avancement de l'enseignement/apprentissage

La question du rythme de l'enseignement/apprentissage de la lecture est peu abordée généralement alors qu'elle semble jouer un rôle pour la qualité des compétences acquises dans ce domaine. Il est intéressant de constater que cette variable est prise en considération dans la recherche de 1973 (rapidité d'avancement dans le programme). Il semblerait, selon cette étude, qu'une progression rapide dans le décodage serait un facteur de réussite accrue en lecture en fin de 1P. De plus, ce rythme soutenu serait particulièrement favorable aux garçons et réduirait l'écart de leurs résultats (en leur défaveur) avec les filles.

Nous avons trouvé quelques résultats de recherche sur cette question qui nous paraît encore largement à explorer. Renard (2007), par exemple, a comparé la manière de gérer le temps chez des enseignant-e-s experts et des enseignant-e-s novices de classes de CP, en France. S'il ne se dégage pas des résultats obtenus une manière de faire propre aux enseignant-e-s expert-e-s, il ressort toutefois que pour ces enseignant-e-s la gestion du temps est un instrument de régulation de leur pratique. Très vite au début de l'année scolaire ils se mobilisent fortement pour l'enseignement de la lecture et maintiennent ce degré de mobilisation tout au long de l'année. Dès le début, ils mettent au point des routines et des contrats de travail avec leurs élèves qui leur permettent de se plonger rapidement dans les activités d'apprentissage. De plus, celles-ci sont plus denses le matin que l'après-midi.

Pour prévenir les difficultés en lecture en CP toujours, Bishop et Sraïki (2007) montent un dispositif dont un des éléments centraux est la gestion et l'organisation du temps par la mise en place d'unités cohérentes et inscrites dans la continuité du cycle 2 (dispositif mené en CP). Deux questions liées au temps sont ainsi posées : comment organiser l'horaire de la classe pour assurer le déroulement des activités projetées ? Quel temps accorder aux différentes dimensions de l'enseignement de la lecture ? Une première évaluation a montré, en effet, que les dimensions de la production écrite et de la compréhension (travaillée spécifiquement) étaient plus laissées de côté que l'identification des mots, qui, elle, prenait la plus large part du temps. Un meilleur équilibre entre ces activités a alors été construit.

Il reste que d'importantes différences dans la gestion du temps sont constatées chez les enseignant-e-s au travers des recherches qui existent à ce sujet.

« On voit donc bien qu'il est impossible, sur la base d'une observation à un moment donné, d'en conclure à la méthode ou l'expertise d'un maître ». Crahay a ainsi mis en évidence que la variabilité des pratiques du même enseignant à contextes différents et à moments différents est aussi, sinon plus importante que la variabilité entre enseignants à conditions égales » (Nonnon & Goigoux, 2007, pp.16 et 17).

Toujours selon Crahay, s'appuyant sur une étude de 1976 (Brophy et Evertson),

« l'enseignant efficace se caractérise par une capacité à gérer la succession des apprentissages de façon progressive et souple. Dit autrement, dans les classes de ces enseignants,

la progression se ferait de façon rapide, mais par étapes et les élèves y expérimentent régulièrement le succès, ce qui a des effets boule-de-neige» (Crahay, 2007, p.63).

Cet-te enseignant-e efficace parvient aussi à s'adapter à ses élèves; si sa classe est composée d'une majorité d'élèves défavorisés (sur le plan socioéconomique), il saura organiser son enseignement en étapes plus petites, en donnant un plus grand nombre d'explications et d'exercices à faire que dans une classe composée en majorité d'élèves favorisés socio-économiquement, mais aussi en multipliant les répétitions.

Nous verrons plus loin dans quelle mesure ces facteurs marquant un enseignement efficace conviennent également à des populations d'élèves différentes, en se référant à la partie de l'analyse multiniveau qui traite des interactions entre les variables pédagogiques et les variables personnelles des élèves.

Interprétation des variables en termes de processus généraux

L'étude multiniveau a tout spécialement fait ressortir les trois variables que nous venons de présenter plus en détail: le décodage et l'approche écrite, la motivation interne et la vitesse d'avancement de l'enseignement/apprentissage comme des éléments d'une démarche d'enseignement efficace. D'un point de vue plus général, ces trois variables mettent l'accent sur le bien-fondé d'un enseignement structuré, progressif, permettant aux élèves d'avancer rapidement dans le programme, centré sur une approche où le décodage, basé sur la connaissance des lettres, tient un rôle important, d'autant plus motivant que les objectifs sont clairs, les activités significatives, qu'il est au service de la lecture de textes, de livres auxquels les élèves ont accès. Les élèves sont suivis et encouragés par la maîtresse, attentive aux plus faibles.

Des recherches processus-produit des années 1970 corroborent ces constats, là encore en termes généraux, décrivant largement les processus d'enseignement les plus efficaces. Ainsi, dans le cadre de la recherche BTES (de 1972 à 1977; voir à ce propos Crahay, 2006, mais aussi Crahay, 2000), organisée en quatre phases alliant des méthodologies quantitatives et qualitatives, les chercheurs montrent que les enseignant-e-s considérés comme efficaces (2P et 5P) sont bien organisés, maximisent le temps alloué aux matières principales (dont la lecture) et réduisent dans la mesure du possible le temps consacré à la préparation des activités, à la discipline ou à d'autres tâches connexes. Les *feedbacks* sont positifs et les encouragements liés positivement au rendement de l'élève. En 2P, pour la lecture, les enseignant-e-s utilisent un matériel diversifié. Faisant une synthèse des traits positifs liés à l'efficacité de l'enseignement, Crahay (2006) souligne trois constats généraux: l'enseignement doit avoir un lien direct avec les objectifs poursuivis; le temps d'implication des élèves dans les tâches et les occasions qui leur sont offertes d'apprendre, doivent être abondants; le travail collectif est particulièrement profitable dans les travaux individuels ou en groupes, il importe que l'enseignant-e assure un contrôle serré du travail des élèves. Il importe encore qu'il tienne compte des élèves forts et faibles, en plaçant les premiers dans des situations problèmes exigeantes et en adaptant les activités aux seconds pour éviter qu'ils se trouvent en situation d'échec. Il convient encore de leur préciser les objectifs visés, de veiller au rythme de progression de l'apprentissage et de leur fournir des explications claires.

Il faut donc bien admettre une certaine cohérence des résultats des recherches processus-produit: l'enseignement efficace tend à créer les conditions organisationnelles (centration sur les activités en lien avec les objectifs visés, contrôle strict des tâches et de la progression des activités) et affectives (manifestation positive de la confiance et de la conviction que tous les élèves peuvent réussir) favorables à une implication forte des élèves dans les apprentissages scolaires.

Sachant les apports de ces recherches, l'analyse entreprise par nos soins vient-elle confirmer ces résultats? Et existe-t-il des données en didactique qui, elles aussi, appuieraient de tels constats? Ce paradigme de l'efficacité garde-t-il toute sa pertinence, voire se renforce-t-il?

Confirmation par les études d'efficacité de l'enseignement

Le questionnement à la base de notre étude porte en effet sur l'efficacité des pratiques d'enseignement; qu'est-ce qui, dans les approches des enseignant-e-s contribue favorablement à l'apprentissage de la lecture chez de jeunes élèves? Il convient de rappeler tout d'abord un des constats majeurs de la première partie de notre étude :

« Ainsi les différentes variables retenues par l'analyse multiniveau ne correspondent pas du tout à une théorie polarisée idéologiquement de l'enseignement de la lecture, mais au contraire à des conceptions plutôt opposées, qui réussissent néanmoins à contribuer à la réussite. Il semblerait donc que des approches complémentaires soient possibles ».
(p. 24)

Rappelons que, même si des querelles pour savoir quelle est la bonne méthode d'enseignement de la lecture sont revenues à la surface des débats de ces dernières années, nombreuses sont les recherches qui montrent – de la même manière – qu'il n'existe pas de méthode idéale, encore moins unique pour garantir la réussite de l'apprentissage de la lecture (cf. ci-dessus). Étant donné la complexité de ce qu'est la lecture, le grand nombre de variables qui interviennent dans le processus de son apprentissage, l'enseignement ne peut se limiter à une seule approche, aussi bonne soit-elle.

Au-delà des propositions didactiques qui foisonnent aujourd'hui, nous voulons souligner quelques recherches qui se sont attachées à l'observation de pratiques enseignantes pour comprendre ce qui en garantissait l'efficacité. Elles nous semblent en effet intéressantes dans la mesure où elles partent de ce qui se fait en classe et lient ainsi non seulement les approches privilégiées par l'enseignant-e, mais aussi des façons d'organiser les activités, des interactions entre élèves et enseignant-e, ou encore la prise en considération des difficultés rencontrées par ces derniers. De telles études de terrain attestent la complexité de l'enseignement de la lecture, en particulier dans les premiers apprentissages, tout en la prenant en compte.

« Pour comprendre ce qui différencie véritablement les choix pédagogiques opérés par les maîtres et leur effet sur les apprentissages des élèves, il est nécessaire de substituer à cette approche en terme de " méthode " une analyse reposant sur l'examen simultané de nombreux indicateurs et de dépasser les déclarations de principes pour entrer dans le détail des pratiques concrètes ». (Goigoux, 2004, pp.42-43)

Goigoux (2002), par exemple, a analysé la pratique d'une enseignante expérimentée du CP français et a mené un entretien d'auto-confrontation simple, pour permettre la mise en évidence des caractéristiques de cet enseignement expert. Il résume l'intervention de cette enseignante en six buts principaux :

- faciliter l'identification de mots
- s'assurer de la compréhension des phrases
- donner accès à la compréhension des textes
- faire en sorte que chaque élève participe à l'activité collective de lecture
- organiser le dialogue et maintenir l'orientation de l'attention des élèves sur cette activité
- assurer les conditions sociales propices au travail (régulation des prises de paroles, maintien d'un calme relatif).

Il relève encore la capacité de cette enseignante de repérer les erreurs des élèves, qu'elles soient graves, nécessitant une remédiation par un guidage procédural pertinent, prévisibles, appelant la construction de compétences spécifiques nécessaires à telle ou telle étape, ou marginales qui ne demandent qu'un traitement rapide. Parce que l'enseignante a une bonne maîtrise de son ensei-

gnement, elle peut consacrer son énergie à ce qui se déroule en classe, aux élèves, aux situations imprévues ou nouvelles. Pour savoir distinguer les types d'erreurs chez ses élèves, cette enseignante possède une bonne connaissance de l'activité de lecture et des étapes de son apprentissage. Elle maîtrise les procédures d'enseignement pertinentes pour guider l'activité de la classe et sait organiser le travail collectif.

Ces constats vérifient l'hypothèse que c'est la capacité de gérer l'ensemble de ces connaissances et savoir-faire qui fait l'enseignement efficace de cette enseignante expérimentée. Par ailleurs, ils corroborent l'idée d'approches plurielles de l'enseignement/apprentissage de la lecture, qui portent de front sur les mots, les phrases et les textes.

Cèbe et Goigoux (2006) confirment à la fois l'existence d'une grande diversité de pratiques enseignantes et l'impossibilité de dégager une manière de faire meilleure que les autres pour apprendre à lire. Toutefois, ils relèvent – à titre d'hypothèses – quelques pratiques appartenant à des enseignant-e-s dont les élèves obtiennent de bons résultats en lecture : enseigner minutieusement toutes les compétences requises pour les apprentissages ultérieurs ; accorder une attention particulière aux élèves des milieux populaires n'ayant pas bénéficié des apports familiaux dont profitent généralement les autres élèves ; mettre en oeuvre une pratique cohérente et réfléchie ; vouloir que tous les élèves réussissent, et y croire. Il s'agirait donc de consacrer en classe une attention particulière à ces élèves, dès les premiers apprentissages, ce que relèvent par exemple des auteurs tels que Bruphy et Evertson (1976) qui, parmi les comportements efficaces de l'enseignant-e, citent sa capacité de contrôler les activités menées et de les aménager pour aider les élèves les plus en difficulté.

Il est ainsi intéressant de constater que, au début des années 1970, des chercheurs américains avaient mis en avant plusieurs des traits relevés par Goigoux et Cèbe pour caractériser un enseignement efficace dans les premiers apprentissages de la lecture (début du primaire). Ainsi, Stalling et Kasowitz (1974) ont montré l'importance d'un enseignement structuré, aux objectifs clairs, centré sur la lecture (sans digressions), comportant des *feedbacks* correctifs, pour expliquer les bons résultats des élèves.

La question de l'efficacité de l'enseignement est donc récurrente et se pose aujourd'hui encore de manière aiguë.

Mais d'autres variables interagissent aussi

Le comportement de l'enseignant-e, pourtant, n'explique pas tout et d'autres chercheurs remettent en cause cette première façon de poser le problème. Existe-t-il réellement des « enseignant-e-s efficaces », demandent-ils ? Crahay n'éluide pas la question et montre que celle-ci s'est posée pratiquement en même temps que foisonnaient les études visant à définir les traits enseignants de l'efficacité. Ainsi, Doyle (1977 ; 1978 ; 1986) rend attentif au fait que ces recherches tendent à neutraliser les variables du contexte alors que ces conditions peuvent avoir des effets sur l'enseignement et donc sur les apprentissages.

Ces effets ont d'ailleurs été démontrés (Bayer, 1979 ; Crahay, 1988, entre autres). Par conséquent d'autres variables devraient également entrer en ligne de compte dans les études : le contenu d'enseignement, les caractéristiques des élèves, le cadre institutionnel, entre autres. La question de la nature de la mesure des produits de l'enseignement/apprentissage doit, elle aussi, être posée. Quelles compétences sont-elles mesurées ? Quel est leur lien avec les contenus d'enseignement ?

En outre et d'une manière générale, les études processus-produit tendent à considérer un-e enseignant-e face à sa classe, celle-ci vue comme un milieu naturel, tributaire de ses lois propres, indépendamment de l'établissement et de son lieu d'implantation. Or, un-e enseignant-e est lui-même marqué par son histoire, par sa formation, par les habitudes du métier. Sa pratique est le plus souvent

un produit sédimenté reflétant les grandes lignes de l'histoire de l'école, de ses structures, de ses objets, etc. Il n'est donc pas si « simple » de distinguer ce qui fait son efficacité, compte tenu de tous ces paramètres contextuels.

Par ailleurs, il n'a pas été démontré que ce qui fait l'efficacité d'un-e enseignant-e était transposable à ses collègues. Ou encore, un-e même enseignant-e, efficace une ou deux années de suite, peut, avec une autre classe, ou en enseignant une autre matière, ne plus s'avérer aussi efficace. Dans ce sens, Bressoux (1994) s'interroge sur la part de l'effet-classe dans celle de l'effet-maître.

De plus, cet ensemble de caractéristiques d'un enseignement efficace valent-elles pour tous les élèves ? Nous venons de voir que l'enseignement diffère quelque peu selon que l'enseignant-e s'adresse à des élèves forts ou faibles. On peut encore se demander si ces constats généraux pourraient s'appliquer de la même manière à des élèves d'âge différent, provenant du primaire ou du secondaire, ou encore d'enfants de grandes villes ou de campagnes, etc.

Les résultats cités par Crahay donnent déjà quelques réponses, le plus souvent allant dans le sens d'une différence de traitement en fonction des groupes d'élèves. Prenant en considération des variables relatives à la classe, aux types d'élèves. L'analyse multiniveau propose, elle aussi, des constats allant dans le même sens.

Interactions entre variables pédagogiques et variables personnelles des élèves

Illustration par l'étude multiniveau

Les principaux résultats concernant ces interactions se trouvent illustrés dans plusieurs des tableaux présentés dans la première partie de ce texte.

Le tableau 3 fait ressortir quatre variables individuelles (élèves) jouant un rôle sur l'apprentissage de la lecture : le genre (les filles réussissent mieux que les garçons), le fait de suivre une scolarité régulière (pas de retard scolaire), une bonne connaissance préalable de la lecture (obtenue par le biais de la famille le plus souvent, mais acquise aussi par la répétition chez un redoublant) et le donné de départ (mesuré par les tests d'aptitude scolaire).

A propos de l'effet du genre sur l'apprentissage (tableau 6), la prise en considération de certaines variables contextuelles amène à y apporter des nuances. Ainsi, l'effectif de la classe peut jouer en faveur des filles ou des garçons selon qu'il est bas ou élevé ; plus précisément, les garçons ont intérêt à fréquenter une classe nombreuse, un gros effectif atténuant en le neutralisant l'effet de la réussite plus grande des filles (en fait, les caractéristiques individuelles perdent de leur visibilité dans une classe nombreuse). De même, ils sont moins désavantagés dans une classe qui compte des élèves allophones et redoublants. Les garçons bénéficieraient encore d'être dans une classe où l'enseignant-e maintient un rythme de travail rapide (favorisant les élèves en difficulté), ou dans une classe comportant une bibliothèque (l'enseignant-e peut travailler avec des groupes d'élèves qui peinent et laisser les autres lire un livre de la bibliothèque).

Corroborant certains de ces constats, des recherches (Nidegger, 2001) montrent que le plus souvent ce sont des élèves provenant de milieux socioéconomiques peu élevés, mais aussi plutôt des garçons, qui ont davantage de difficultés en lecture que les autres élèves.

De même, certaines variables contextuelles influencent l'effet du redoublement sur l'apprentissage de la lecture (tableau 8). Ainsi, redoubler dans une classe majoritairement francophone est plus pénalisant que dans une classe comportant aussi des allophones. En revanche, une atmosphère positive dans la classe, un-e enseignant-e sensible aux objectifs affectifs, sont à même d'aider les redoublants à réussir leur apprentissage de la lecture. Comme nous l'avons vu plus haut, un-e enseignant-e efficace sait donner

des *feedbacks* positifs et encourager tous les élèves, persuadé que tous, même les plus faibles, peuvent réussir.

Enfin, l'équipement jouerait un rôle mitigé sur l'apprentissage de la lecture des redoublants. Il faut préciser que, à l'époque de la recherche en lecture dans les années 1970, seuls les jeunes enseignant-e-s sortant de leur formation avaient reçu un matériel pour cet apprentissage. Cela plaiderait en faveur de l'idée que doter les enseignant-e-s d'un matériel, aussi bon puisse-t-il être, ne suffit pas, notamment, à compenser les difficultés des élèves en retard, l'essentiel se situant davantage dans l'usage que l'enseignant-e fait de son matériel.

Qu'en est-il du milieu socio-professionnel de l'élève ? Dans PISA, entre autres, cette variable se révèle importante; elle est associée de près aux résultats obtenus (Moreau, 2001). Les élèves (il s'agit d'élèves de 9^e année) issus d'un milieu socio-professionnel élevé obtiennent généralement de meilleurs résultats que ceux provenant d'un milieu modeste. Or, dans la recherche qui nous occupe, sur la lecture en 1P, un tel constat n'apparaît pas de manière si évidente. En fait, cette variable est très sensible à l'interaction avec le contexte. Ainsi elle prend de moindres proportions lorsque l'effectif de classe est élevé, que les filles n'y sont pas trop nombreuses, que la classe est citadine, avec un nombre substantiel d'élèves étrangers, et que la méthode pédagogique est centrée sur le décodage.

Finalement, les résultats montrent que les effets d'interaction de certaines variables individuelles avec le contexte sont très importants et peuvent ainsi remettre en question les tendances générales constatées par ailleurs. Il convient de souligner ici l'intérêt de l'analyse multiniveau qui met en avant les effets du contexte, c'est-à-dire des variables pédagogiques, sur les variables individuelles des élèves.

Mêmes constats dans d'autres recherches

Dans ce sens, certaines recherches mentionnées par Crahay (2006) laissent entrevoir cette modulation. Par exemple, Bruphy & Evertson (1976) mentionnent que la variable socioéconomique se conjugue à des traitements pédagogiques différenciés (des défis, des situations-problèmes aux élèves de milieu socioéconomique élevé, des activités plus simples et moins risquées pour les élèves de milieu socioéconomique bas). Dans l'ensemble, ces auteurs, ainsi que Turner (1971), montrent qu'il est préférable de proposer un enseignement basé sur la répétition, avec des explications claires et un rythme de progression modéré, c'est-à-dire dans des conditions qui effacent tout risque que les élèves en difficulté puissent perdre pied dans leur apprentissage (essentiellement en lecture et en mathématiques).

Comme on le voit par ces exemples, si les variables pédagogiques agissent différemment selon les groupes d'élèves, cela pourrait signifier qu'il est important, voire nécessaire, de promouvoir un enseignement différencié à l'école, au primaire en tout cas.

Mais concrètement, qu'est-ce que cela signifie pour les premiers apprentissages de la lecture ? Est-ce envisageable ?

La différenciation de l'enseignement voulue, mais difficile et exigeante

Parmi les facteurs d'efficacité figure la capacité de l'enseignant-e d'interpréter les actions des élèves et d'adapter le soutien à leurs caractéristiques individuelles. Il reste que, au vu des conditions générales de travail dans les écoles, la gestion de la part individuelle et de la part collective de l'enseignement est extrêmement difficile. Des solutions sont encore à rechercher pour différencier autant que faire se peut l'enseignement, avec les difficultés que cela suppose (Simon, 2007).

A titre d'exemple, Bishop et Sraïki (2007) ont introduit, dans leur dispositif de recherche, un volet différenciateur en lien avec l'évaluation tout d'abord, mais aussi avec une organisation des activités

dans le temps. Un adulte supplémentaire a été introduit dans la classe (de ZEP⁵) pour précisément se pencher plus particulièrement sur certains élèves et permettre une différenciation de l'enseignement. Le but de ce dispositif était d'amener tous les élèves à faire des progrès, ce qui a finalement été observé. Il reste que de tels dispositifs sont lourds et difficiles à généraliser.

D'autres recherches tentent de cerner les approches didactiques qui favorisent tel ou tel type d'élève. Piquée et Sensevy (2007), par exemple, mettant en relation des données didactiques et des résultats d'élèves en lecture, montrent que des progrès sont observés chez les élèves, plus ou moins importants selon leur niveau et selon les approches privilégiées.

Ainsi, sur le plan micro (travail sur les mots), un travail de discrimination auditive, s'il est fréquent, est profitable aux élèves les plus faibles ; de manière générale, les enseignant-e-s dont la pratique est essentiellement axée sur ce niveau micro ne varient guère les activités en classe. Cette variété apparaît en revanche plus importante pour un enseignement axé sur les plans méso (travail sur les phrases) et macro (travail sur les textes). Un travail sur les phrases conduit à des progrès en lecture de tous les élèves, quel que soit leur niveau, mais davantage pour les bons. Cependant, un travail sur la syntaxe s'avère particulièrement bénéfique aux plus faibles. Quant au plan macro, il engendre des activités sur les textes permettant également des progrès de tous les élèves, en particulier s'il s'agit d'un travail sur la morphologie et l'identification des caractéristiques des personnages. En revanche, le travail fréquent sur les familles de mots et sur les caractéristiques des genres de texte est plus profitable aux élèves initialement de bon niveau. Les auteurs affirment finalement que, pour les élèves faibles, des pratiques ancrées aux trois plans, micro, méso et macro, se révèlent nécessaires à leurs progrès.

Ces constats rejoignent ceux d'autres études (ONL, 2005 ; Fijalkow, 2000). De plus, la variété des activités menées en classe aurait là encore un impact bénéfique sur les apprentissages et les progrès. Ceci serait d'autant plus vrai pour les élèves faibles qui progressent davantage si les dimensions travaillées sont nombreuses et variées.

Cain (2009), étudiant les difficultés des mauvais lecteurs, va dans le même sens en affirmant que, vu le grand nombre de variables langagières et liées à la littératie, impliquées dans les difficultés de lecture, les mesures de remédiation des faibles lecteurs, pour être efficaces, ne doivent pas porter isolément sur le mot, la phrase ou le texte, mais tenir compte de ces trois plans ensemble.

Pour Fijalkow (2000), il n'est pas évident pour des élèves en difficulté de comprendre le lien entre le décodage et la compréhension si ce lien ne leur est pas révélé. Il est donc nécessaire de trouver une articulation des activités portant sur le code et des activités mettant les enfants dans des bains d'écrits, au contact avec les livres, la culture de l'écrit, des adultes lisant et écrivant. Les enfants doivent avant tout comprendre le sens de l'écrit pour accéder à l'apprentissage de l'écrit.

C'est par un enseignement explicite de stratégies de lecture que les élèves les plus en difficulté (mais les autres aussi) peuvent améliorer leurs compétences en lecture, selon Cèbe et Goigoux (2007). Les activités que ces auteurs ont mises au point dans cette perspective doivent également conduire les élèves à développer des capacités métacognitives et de contrôle de l'activité. Ils insistent aussi sur la nécessité de remobiliser de manière systématique les habiletés travaillées, de les exercer dans des contextes et sur des contenus différents.

A ce propos, Mauffrey et Cohen (1995) voient également la nécessité de proposer aux enfants en apprentissage des activités systématiques et contraignantes, afin de mettre en place et d'exercer des compétences et des stratégies.

5 En France, les Zones d'éducation prioritaire (ZEP) sont des lieux dans lesquels se trouvent des établissements scolaires dotés de moyens supplémentaires et d'une plus grande autonomie pour faire face à des difficultés d'ordre scolaire et social.

En Suisse romande, dès les années 1980 la difficulté de différencier l'enseignement de la lecture a été constatée, avec pour principale conséquence un accroissement des écarts entre élèves faibles et élèves forts (Weiss, 1987, Hutin, 1985). En même temps, des études sur les types de difficultés des élèves et sur la façon d'y remédier ont été également conduites et mises au service des enseignant-e-s de la scolarité obligatoire (Favre, 1983 ; Hutin et Lurin, 1984 ; Bain, Erard et Séchaud, 1994 ; Leu et Soussi, 1995 ; Aeby, 2004).

Il faut dire que parmi les principes qui ont présidé au développement de l'enseignement de la lecture en Suisse romande dès la fin des années 1970 figuraient les deux points suivants : « dans la mesure du possible, l'enseignement sera individualisé pour consolider l'apprentissage de la lecture de l'élève ; l'enfant bénéficie de deux ans pour apprendre à lire ». Cette deuxième proposition a suscité de nombreuses critiques, car il est assez rapidement apparu que les élèves qui ne savaient pas lire (décoder) à la fin de la 1P étaient défavorisés en 2P, faute de pouvoir lire suffisamment bien pour accéder aux apprentissages d'autres disciplines scolaires. Il était donc illusoire de penser laisser deux ans pour l'apprentissage de la lecture. Dans ce contexte, l'enseignement/apprentissage de la lecture en 1P-2P est tout à fait important, parfois déjà amorcé à l'école enfantine (même si d'aucuns craignent une scolarisation de ces années avant la 1P alors que l'école enfantine va devenir obligatoire en Suisse, en tout cas dans les cantons qui ont signé le concordat HarmoS).

Il reste que les enseignant-e-s continuent de pratiquer la différenciation de leur enseignement dans leur classe. C'est en tout cas ce qui ressort d'une étude faite dans le canton de Vaud par Gilliéron Giroud et Pulzer-Graf (2009). Pour ce faire, les enseignantes interrogées disent, pour la majorité d'entre elles, utiliser une palette variée de moyens d'enseignement, alternant moyens issus du commerce et moyens officiels ; adopter une approche pédagogique particulière (par exemple un travail en ateliers ou en niveaux, un décroisement interclasses) ; adapter les tâches demandées aux élèves soit en quantité, soit en niveau de difficulté, en tenant compte des capacités et du rythme d'apprentissage des élèves.

Dans leur rapport, les auteures de cette étude vaudoise font un constat mitigé sur l'ampleur et les effets de la différenciation. Elles se montrent d'ailleurs également critiques envers les mesures externes prises actuellement pour mesurer l'efficacité de l'école par le biais des épreuves administrées à niveau cantonal, régional ou international, qui mettent en évidence les résultats contrastés des élèves, mais sont incapables de suggérer des démarches utiles à l'enseignement.

Le problème de l'écart des acquis entre les élèves demeure donc préoccupant et nécessite certainement une plus grande mobilisation à la fois de la diversité des ressources didactiques à disposition et des partenaires engagés auprès des enfants, selon des formes nouvelles de collaboration par exemple. Il reste que, au-delà des dispositifs nombreux et variés mis en place dans bien des cas et dans les petits degrés pour différencier l'enseignement, la question de la disparité des acquis entre les élèves reste posée. Cette question est d'ailleurs récurrente ; on peut faire l'hypothèse qu'elle concerne finalement non seulement les dispositifs d'enseignement mais aussi le contexte scolaire, son organisation (en cycles, effectifs de classe, conditions de promotion, mesures d'appui à l'intérieur ou à l'extérieur de la classe, etc.).

Une alliance entre recherches quantitatives et qualitatives

Comme cela a été dit à plusieurs reprises, le principal apport d'une étude s'appuyant sur une méthode d'analyse multiniveau est la mise en interaction de facteurs individuels et de facteurs collectifs (caractéristiques des classes et des enseignant-e-s) pour expliquer les résultats des élèves, ici en lecture.

Cependant l'interprétation à donner à ces interactions est délicate, car elles remettent en cause la causalité simple que les chercheurs en pédagogie avaient cru d'abord pouvoir affirmer. Ainsi, les

constatations faites dans notre étude multiniveau nous ont obligés à apporter bien des nuances aux conclusions de la première recherche de 1973 sur la portée réelle d'un facteur ou d'un autre, ou d'un faisceau de facteurs, sur les résultats obtenus. Les liens de cause à effet apparaissent complexes et leur existence bien délicate à affirmer. De ce fait-là, au-delà de certaines conclusions à traiter avec prudence, le travail entrepris dans cette étude amène à entrevoir de nouvelles investigations.

A notre avis, et pour aller dans le sens de Crahay (2006), il vaudrait la peine d'envisager des travaux alliant la mesure des résultats en fonction d'un certain nombre de facteurs (individuels et collectifs) et la prise en considération d'informations sur les pratiques en classe : « Bref, c'est sans doute en jonglant avec les approches corrélationnelles, ethnographiques et quasi-expérimentales qu'il sera possible de bâtir une base de connaissances relative au processus enseignement-apprentissage » (p.138). Pour notre part, nous ajouterions la recherche didactique dans les balles du jongleur.

Pourquoi une telle proposition ? Parce que l'approche corrélationnelle des études processus-produit comporte des limites que Crahay a bien mises en évidence. Par exemple, une corrélation ne signifie pas forcément une relation de cause à effet. Ou encore, si deux variables sont corrélées entre elles et corrélées au rendement, il se peut que ce ne soit en fait qu'une des deux qui soit véritablement déterminante pour le résultat, l'autre lui étant co-dépendante. De ce fait, il apparaît utile d'accompagner les résultats quantitatifs d'une telle approche d'observations fines apportant des informations complémentaires et spécifiques. Des auteurs tels que Rosenshine et Furst (1973) avaient déjà fait les mêmes constats et construit un schéma de recherche en plusieurs étapes, dont la description et l'analyse du processus d'enseignement. Des études de ce genre (il y en a d'autres) affinent donc les constats, tout en restant dans la logique des recherches sur l'efficacité de l'enseignement.

Des recherches relatives aux variables pédagogiques majeures

En recherchant les variables pédagogiques susceptibles d'expliquer les meilleurs résultats des élèves à des épreuves de lecture, nous avons poursuivi une logique d'efficacité de l'enseignement. Si les enseignant-e-s des premiers degrés organisent un enseignement centré sur le décodage par l'apprentissage des lettres, sur la motivation, sur un avancement des activités relativement rapide, avec le souci de promouvoir la différenciation, l'apprentissage des élèves devrait s'en trouver meilleur. Mais, même décrites avec minutie, ces variables ne disent pas encore ce qui se passe réellement dans la classe, comment ces options pédagogiques agissent selon le contexte observé, ni si elles conviennent pour toute classe, tout-e enseignant-e, tout élève. De plus, comme le précise Crahay, ce que l'enseignant-e met en place comme modalité d'intervention dans sa classe n'a pas seulement un effet sur les acquis des élèves, mais peut avoir une répercussion aussi sur le rapport que l'élève instaure à l'objet de savoir. Or c'est seulement par une observation en classe qu'il apparaît possible de comprendre comment se construit ce rapport.

Prenons l'exemple des livres de lecture. A travers la question du matériel d'enseignement utilisé par les enseignant-e-s, qui n'est pas traitée en tant que telle dans notre étude mais indirectement au travers des méthodes de lecture à la disposition des enseignant-e-s, des problèmes similaires se font jour. Le matériel en usage dans les classes déborde généralement largement celui d'une méthode de lecture particulière et il a été montré qu'il n'est pas pertinent de se demander laquelle serait la plus efficace. Il reste que le choix des textes travaillés en classe et les activités organisées à leur propos sont cruciaux pour l'apprentissage des élèves (Thévenaz-Christen et al., à paraître). Cependant, nous connaissons encore peu les modalités d'usage des matériels choisis par les enseignant-e-s pour leur enseignement de la lecture, et il serait intéressant de savoir dans quelle mesure tel ou tel matériel et les modalités de son utilisation contribuent à une meilleure réussite des élèves. Ces moyens d'enseignement fonctionnent comme médiateurs de l'objet d'enseignement/apprentissage : que disent-ils de cet objet ? Qu'en retiennent les élèves ? Quel rapport au savoir instaurent-ils auprès des élèves ?

Est-il différent selon les types d'élèves (garçons, filles, élèves en difficulté, élèves allophones, etc.)? Là encore, ressort le bien-fondé de faire interagir des recherches compréhensives, descriptives, couplées à des études statistiques de type multiniveau.

Pour prendre un autre exemple, la question de la progression des apprentissages ouvre elle aussi la porte à de multiples investigations. Elle est, didactiquement, difficile à résoudre, d'autant plus que les psychologues ne sont pas tous d'accord sur le type de progression qu'ils recommandent : progression pas à pas, du plus simple au plus complexe, du mot au texte en passant par la phrase, ou progression spiralaire (Saada-Robert et al., 2003), partant du complexe pour aller à des éléments spécifiques et retourner au complexe? Que comprendre sous l'appellation « période sensible », reprise de Montessori et développée par Brossard (2004) dans une perspective vygotkienne (zone proche de développement)? Il y a là un champ énorme à investiguer et à mesurer pour comprendre comment progressent les apprentissages, quelles sont les activités en classe qui favorisent cette progression, et quelles en sont les retombées sur les enfants.

Des recherches relatives aux interactions

Nous avons souligné plus haut (p. 47) qu'un grand nombre d'interactions avaient été mises en évidence entre variables pédagogiques et variables contextuelles, une même intervention pédagogique ayant des conséquences différentes selon le type de classe, d'élèves par exemple. Dans la même perspective, Crahay, parmi les conclusions de son ouvrage de 2006, montre que les apports des études réalisées pour le primaire ne peuvent vraisemblablement pas être généralisés au secondaire.

Une telle conclusion justifierait une diversité des approches d'enseignement, tout en conférant à l'enseignant-e une responsabilité particulière, celle de connaître suffisamment les spécificités de ses élèves pour mettre en place des activités appropriées.

On a vu que la différenciation est très souvent mise en avant comme un moyen de favoriser les apprentissages des élèves; or les modalités de différenciation sont très diverses, pouvant porter sur les modes de travail, les contenus et les rythmes d'apprentissage, l'usage du matériel, etc. Dans le sens de la recherche de Piquée et Sensevy, vue plus haut, il devient alors intéressant de comprendre de manière plus approfondie quelles peuvent être ces modalités et leurs effets sur l'apprentissage. Cela justifierait par exemple d'introduire dans l'étude statistique des variables didactiques minutieusement définies, ici en terme de différenciation, issues de recherches en didactique. A quels types d'élèves ou de classes conviendrait telle ou telle mesure différenciatrice? Qu'impliquent à leur tour ces mesures sur un plan didactique (y a-t-il des mesures convenant à certain-e-s enseignant-e-s qui ne conviendraient pas à d'autres, selon les formations, les méthodes, les matériels, les styles de chacun)? Comment expliquer leurs effets dans le contexte de l'enseignement/apprentissage de la lecture? Une mesure différenciatrice qui s'avérerait positive en 1P est-elle généralisable ensuite, en 2P par exemple?

PERSPECTIVES MÉTHODOLOGIQUES

En conclusion, nous aimerions nous dégager un peu du problème que nous avons étudié, celui de l'enseignement/apprentissage de la lecture, pour développer ce que cette étude nous a appris sur le plan de la méthodologie de la recherche. Car, paradoxalement, c'est en poussant plus loin encore une recherche statistique déjà très élaborée (avec analyses factorielles de second ordre et tests statistiques multivariés, par exemple) que nous avons dégagé des interactions entre variables personnelles et situationnelles qui ont été interprétables qualitativement, qui ont « donné sens » à ce qui se passe à l'école. Or un tel aboutissement ne correspondait aux attentes ni des « qualitatistes », ni à celles des « quantitativistes ».

Pas de recette pour faire apprendre les élèves, selon les qualitatistes

Nos résultats sembleront paradoxaux d'abord aux tenants d'une approche essentiellement subjective, à tous ceux qui nient qu'on puisse appliquer les méthodes de la science objective à la réalité scolaire. Car pour eux, le codage des situations que nécessite leur traitement statistique fausse nécessairement notre appréhension des choses. Les moyennes que nous calculons se réfèrent à des « dimensions » qui n'existent que dans notre esprit. Elles ne nous apprennent rien sur ce qui se fait « réellement » dans les classes, sur ce qui s'y enseigne et ce qui s'y apprend, d'après les personnes qui vivent ces événements.

L'idée même d'efficacité de l'enseignement peut même être rejetée par ceux qui n'acceptent pas qu'on quantifie le savoir des élèves, ou qu'on catégorise le comportement des enseignant-e-s. Il s'agit dans les deux cas pour eux d'abstractions, qui faussent la perception de la réalité humaine des personnes en contact dans les classes. Ces abstractions, diront-ils, sont dangereuses, car elles visent, bien sûr, à amener le plus d'enseignant-e-s possible à acquérir les comportements jugés, décrétés sur une base scientifique comme les plus efficaces pour les apprentissages. De telles recherches supposent que plus l'enseignement correspond aux processus idéaux relevés statistiquement, plus l'apprentissage des élèves sera optimal. Or, une telle idée comporte de sérieuses limites, disent les qualitatistes, par rapport à ce que sont l'enseignement, l'apprentissage et les objets d'enseignement/apprentissage.

L'enseignement d'abord ne saurait se réduire à une série de comportements de l'enseignant-e, indépendamment du contexte scolaire dans lequel celui-ci exerce. L'enseignant-e, tout en ayant forgé son propre style, est un professionnel qui travaille selon la tradition qui demeure présente dans les pratiques de ses collègues. La même insistance sur l'ordre et la discipline, par exemple, sera acceptée comme nécessaire et légitime dans un contexte social, mais jugée étouffante et révoltante dans un autre.

« Par conséquent, il semble nécessaire de penser les méthodes de recherche sur les effets de l'enseignement, en prenant pour acquis que la seule fréquence de certains comportements de l'enseignant ne peut rendre compte de la complexité des relations processus-produit qui se tissent au sein de la classe » (Crahay, 2006, p.129).

L'apprentissage des élèves ensuite n'est pas non plus modulable de l'extérieur comme une réaction chimique. L'élève n'est pas passif dans son étude. Il est finalement responsable de ses progrès (qui dépendent de sa motivation, de son taux d'engagement dans la tâche,...). Lui aussi a des représentations et des conceptions qui lui sont propres et qui l'amènent à réagir de telle ou telle façon à l'enseignement prodigué dans sa classe.

Enfin, dans une perspective didactique, nous ajouterons encore l'effet de l'objet d'enseignement sur l'enseignement comme sur les acquis de l'apprentissage. La plupart des études relatées par Crahay portaient sur la lecture et les mathématiques; quelques-unes ont montré des différences entre ces deux disciplines (voir la recherche américaine BTES). Un-e enseignant-e à l'aise et efficace dans une discipline pourra l'être beaucoup moins dans une autre pour des raisons diverses.

Et pourtant l'école peut faire mieux apprendre

Même si l'on doit donner raison à tous ceux qui affirment qu'il n'y a pas de recette applicable de façon générale dans l'enseignement et que chaque situation scolaire a sa spécificité, il ne s'agit pourtant pas de jeter le bébé avec l'eau du bain. Les tests statistiques que nous avons appliqués dans notre étude ne laissent pas de doute, dans certains cas extrêmes notamment, sur la réalité des relations observées qui ne peuvent être dues au hasard. D'une façon générale, les recherches processus-produit ont mis en évidence des constats présentant une régularité telle qu'on les retrouve pour différentes disciplines ou différents pays. Corrélation n'est pas raison, sans doute, mais il est impossible d'ignorer des co-occurrences aussi démontrées.

Le thème de l'efficacité est fondamental et a même connu déjà plusieurs étapes de développement, qui nous amènent à la présente discussion. L'idée d'une pédagogie expérimentale enthousiasmait bien des enseignant-e-s avant la seconde guerre mondiale déjà. Ils créèrent une association de langue française après la guerre et publièrent de nombreux articles. Mais c'est surtout en anglais que toute une discipline scientifique s'est développée, s'appuyant sur des collaborations au niveau international. L'importance des résultats obtenus et leur applicabilité dans les classes ne sont plus à démontrer.

Pourtant en 1957, un article de Cronbach : *The two disciplines of scientific psychology* fit date dans ce développement. L'auteur y mettait en évidence une absence de recherche regrettable à l'articulation des études expérimentales et des études fondées sur l'observation, notamment celles étudiant les corrélations entre les enseignements donnés et leurs résultats. Curieusement, ce qui faisait l'objet d'étude dans la discipline expérimentale, c'est-à-dire la réussite moyenne des élèves suivant un type de curriculum ou un autre, n'était pas observable par l'autre discipline. Les corrélations ne traitent en effet que les écarts à la moyenne. Inversement, ces écarts ne sont considérés par les expérimentalistes que comme des exceptions, du « bruit » gênant, qu'ils doivent réduire au maximum. Cronbach défendait l'idée au contraire que ces deux approches méthodologiques étaient complémentaires.

Il illustre sa thèse en prenant l'exemple de la comparaison expérimentale de trois méthodes d'enseignement, dont les résultats moyens pouvaient naturellement conduire à recommander l'emploi de l'une d'elles, celle dont la moyenne était la plus haute. La méthode corrélationnelle, de son côté, étudiait pour chaque méthode la relation entre les scores d'aptitude des élèves et leurs résultats à l'épreuve commune de fin d'étude. Cette relation était donnée par trois droites de régression qui montraient le résultat à attendre d'un même élève selon la méthode qui lui était appliquée. Si l'on plaçait chaque élève dans la condition optimale pour lui, la réussite d'ensemble était bien supérieure à ce qu'aurait permis la meilleure méthode, appliquée à tous les élèves. On voyait bien ainsi les limites de chacune des deux approches méthodologiques et la nécessité de les dépasser en les combinant.

Généralisant cet exemple, Cronbach recommandait par la suite l'étude des interactions entre variables personnelles des élèves et variables pédagogiques de la situation d'enseignement. Mais la prise en compte simultanée de plusieurs variables relatives à l'enseignant-e, à sa classe, au milieu environnant, etc. posait des problèmes techniques qui n'ont pu être maîtrisés que par l'introduction des méthodes multiniveau.

Le problème principal de l'étude des interactions demeurait néanmoins, et Cronbach ne l'a pas caché dans la suite de sa réflexion, publiée en 1975 : *Beyond the two disciplines of scientific psychology*. C'est que si le chercheur peut appréhender le sens d'interactions reliant deux ou trois variables, il perd rapidement pied pour des combinaisons entre quatre variables et davantage. Or la réalité en comporte bien davantage.

Les quantitativistes se trouvent donc dépassés par la complexité qu'ils prétendaient maîtriser... Cronbach proposait une approche qualitative pour aller plus loin, s'appuyant sur ce que l'on peut comprendre de la situation, sur les hypothèses « sensées » que l'on peut formuler, à partir des relations objectivement établies et qu'on a pu interpréter.

Il est frappant de voir l'approche scientifique toucher ses limites et rencontrer les faits que ses adversaires lui opposaient au départ. La différence est cependant notable. Après une étude mult niveau, par exemple, on sait mieux l'importance relative des variables dont on invoque l'influence. Les hypothèses que l'on doit faire ne sont plus formulées « en l'air ». Les explications que l'on se risque à avancer sont cadrées par des parts de variance objectivement déterminées.

L'opposition vaine de deux épistémologies

Nous sommes partis d'un conflit de tendances, de nature assez idéologique. Les qualitativistes rejettent par principe les recherches processus-produit. Pour eux, la complexité des processus d'enseignement/apprentissage rend fragile l'idée même d'efficacité. Ce qui conviendra dans un contexte donné, avec telle ou telle personne, avec telle ou telle classe peut ne pas avoir du tout le même effet dans un contexte différent, avec une autre classe, pour un-e autre enseignant-e.

Pour d'autres chercheurs en éducation, la croyance en la possibilité de recherches expérimentales au contraire va de soi. Si certains principes pédagogiques, comme ceux qui ont été abondamment cités plus haut, se révèlent en général efficaces, pourquoi ne les donnerait-on pas en exemple aux enseignants ? Si des faits établis prouvent que certaines façons d'enseigner portent plus de fruit que d'autres procédures, appliquons-les. Comme une pression sociale de plus en plus forte est exercée sur l'école pour élever le niveau de performance des élèves, il est évident que la même pression va se répercuter sur les enseignant-e-s pour qu'ils suivent ces principes d'efficacité pédagogique. La recherche doit dégager les principes à suivre.

Les deux écoles de pensée auxquelles nous venons de donner la parole ont une origine philosophique fort lointaine. Sans remonter au Déluge, rappelons que la première se rattacherait à l'épistémologie historique et la seconde à l'épistémologie expérimentale décrites par Piaget (1967). Il s'agit de deux façons opposées de chercher à comprendre le réel. L'historien veut trouver des raisons spécifiques. Le scientifique veut trouver des lois générales. Le premier cherchera les particularités de la situation qui font que les événements se sont passés de telle façon et pas d'une autre. Le scientifique cherchera au contraire les traits généraux des situations étudiées qui lui permettront de faire abstraction des contingences locales et temporelles.

Selon leur orientation philosophique, les chercheurs valoriseront l'une ou l'autre approche. Aucune des deux n'est fautive, et chacune peut même légitimement affirmer qu'elle rend compte de faits que l'autre ne peut pas expliquer. Le pessimiste ajoutera que, de ce fait, chacune laisse échapper l'autre versant du réel. Voilà pourquoi les historiens continuent à analyser le passé sans parvenir à mieux prédire l'avenir. C'est aussi pourquoi les chercheurs en éducation ne rendent compte que d'une petite partie de la variance des observations qu'ils analysent. Aucune méthodologie à elle seule ne peut épuiser le réel.

Leur complémentarité nécessaire

Puisque chaque type de recherche ne peut expliquer que partiellement la réalité scolaire, il est logique de penser qu'il faut exploiter leur complémentarité.

D'où la proposition que nous faisons d'une alliance entre les études de type multiniveau et de type didactique, les recherches quantitatives et qualitatives. Les unes et les autres à la fois se justifient et comportent des limites. Les unes et les autres voient leurs méthodes s'affiner et leurs constats, tout discutables qu'ils puissent être, se compléter et ouvrir de nouvelles perspectives.

Mais de telles affirmations, à ce niveau de généralité, restent banales et difficiles à appliquer. Peut-on décrire un peu plus précisément, à partir d'exemples, comment ces recherches pourraient s'articuler entre elles ?

Exemple de la médecine

L'évolution de la recherche médicale au XX^e siècle est très parlante. Les anciens médecins, formés au siècle précédent, étaient des littéraires. Ils apprenaient à écouter leurs malades, à chercher leurs particularités, leurs modes de réaction au stress, par exemple. Ils en tenaient compte pour formuler leurs ordonnances, en dosant individuellement le contenu des médicaments qu'ils prescrivaient.

La méthode expérimentale et statistique a transformé la recherche médicale au XX^e siècle. On a vu que les malades réagissaient différemment à un même médicament, ou à un même traitement de base. En cherchant à différencier par des tests objectifs les groupes qui se distinguaient ainsi, on a pu affiner les diagnostics. Sur cette base, de nouveaux médicaments et de nouveaux traitements ont pu être essayés, appropriés à chaque groupe, permettant ainsi de recommencer le cycle de recherche.

Par rapport à la recherche en éducation, cet exemple est très parlant. Si les enseignant-e-s font de leur mieux pour individualiser leur façon de traiter chaque élève, ils ne disposent pas de suffisamment de résultats d'analyses multiniveau leur permettant de déterminer les interactions entre les variables individuelles des élèves et les variables pédagogiques de la forme d'enseignement. Il leur manque ensuite les tests diagnostics s'appuyant sur ces analyses multiniveau, qui les conduiraient à exploiter ces interactions en prédisant l'approche didactique la plus efficace pour chaque type d'élève.

Les médecins, pourtant, n'en restent pas à ces études statistiques. D'autres recherches de laboratoire leur permettent de comprendre toujours mieux les dysfonctionnements qu'ils soignent. La recherche quantitative est donc complétée par une autre recherche qui lui donne sens, qui permet de répondre au Pourquoi de l'efficacité plus grande de tel traitement dans tel groupe. Et c'est sur cette réponse au Pourquoi qu'est finalement basé le traitement. Le médecin a besoin de comprendre le cas pour pouvoir le traiter. En pédagogie aussi, c'est l'intuition de l'enseignant-e qui devra avoir le dernier mot, mais fondée sur toute l'information objective qu'il aura été possible de recueillir sur les particularités de chaque élève en matière d'apprentissage.

Exemples en didactique : dans les domaines de l'éveil aux langues et de la production écrite

Une recherche de grande envergure a été menée dans plusieurs pays pour tenter de mesurer les effets d'une démarche d'éveil aux langues (Candelier, 2003). L'une des principales hypothèses formulées était que les effets du programme Evlang, à savoir développer des aptitudes métalinguistiques et des attitudes d'ouverture à la diversité linguistique et culturelle chez les élèves, devaient être plus importantes dans les classes où les enseignant-e-s mettaient en œuvre une pédagogie axée sur l'implication et l'activité des élèves. Pour ce faire, un dispositif prévoyant une évaluation quantitative et

une évaluation qualitative a été mis au point. Si la première devait permettre des mesures de l'impact des activités d'Evlang sur les attitudes et les aptitudes des élèves (tests initiaux et finaux), la seconde portait un regard interne sur le fonctionnement des interactions dans les activités Evlang mises en œuvre dans les classes. L'articulation des deux évaluations s'est concrétisée tout particulièrement dans la mise en évidence de typologies relatives à l'implication des élèves dans les activités et aux modalités de communication lors du travail en groupe, pour expliquer les résultats par classe. S'il s'est avéré en effet que les classes des enseignant-e-s parvenant à impliquer fortement les élèves dans le travail ont obtenu des résultats meilleurs que les autres, cette différence a porté essentiellement sur les compétences métalinguistiques des élèves. Au-delà des résultats, qui montrent certaines limites de l'idée d'efficacité de l'enseignement (« ... il ne suffit pas que tous les indicateurs des dimensions pédagogiques soient " au vert " pour que l'efficacité maximale soit atteinte » (p.127), deux des chercheurs du projet (Genelot et Tupin, 2003) soulignent l'intérêt de connecter des approches qualitatives d'observation des pratiques de classes avec des approches quantitatives d'évaluation de l'efficacité, dans la mesure où – dans leur cas – elles ont permis de revisiter le paradigme processus-produit (le regard interne a par exemple mis en évidence le rôle de l'élève dans la construction des connaissances), et de passer progressivement du versant enseignement et de ses effets à un versant plus large, celui de l'enseignement/apprentissage, intégrant la dimension interactive que ce dernier implique.

En Suisse romande cette fois, voulant comprendre comment, en l'absence de moyen d'enseignement romand en expression, les enseignant-e-s de 6^e enseignaient la production écrite dans leur classe, une équipe de chercheurs (Béatrix et al., 1999) a mis au point un dispositif alliant des investigations de nature qualitative et une prise d'informations quantitative. D'une part, ils ont analysé le contenu des plans d'études relatif aux objectifs dans ce domaine et ils ont observé l'enseignement d'une dizaine d'enseignant-e-s provenant de cantons différents, leur demandant d'aborder un objet particulier, le même pour tous (un résumé incitatif de conte); d'autre part, ils ont construit un questionnaire destiné aux enseignant-e-s en vue de cerner ce qu'ils disaient de leurs pratiques, et ils ont procédé, au début et au terme de la période d'observation, à l'analyse de productions écrites des élèves. Les constats conclusifs, prenant en compte l'ensemble des résultats obtenus, ont mis en lumière une grande diversité des pratiques, favorisée à la fois par le degré de généralité des objectifs et l'absence ou la quasi absence de moyens d'enseignement. L'observation dans les classes a confirmé ce résultat. De plus, grâce à l'analyse critériée et quantifiée des textes des élèves, certains éléments de l'enseignement ont pu être mis en avant pour expliquer la qualité de l'apprentissage. De tels éléments ont confirmé l'utilité de mettre des outils didactiques à la disposition des enseignant-e-s, en particulier pour l'évaluation des textes des élèves, mais aussi dans le but de proposer des démarches cohérentes d'enseignement de la production écrite.

A la suite de ces exemples, le souhait exprimé par Genelot et Tupin nous paraît pertinent dans la mesure où ces chercheurs délimitent ainsi une possible articulation entre les recherches qualitatives et quantitatives : « On ne peut donc que souhaiter que les nombreux travaux d'observation des classes actuellement engagés par diverses équipes de recherche, permettent dans les années qui viennent de décrire des observables simples mais discriminants pouvant être intégrés ensuite dans des approches quantitatives » (p.128).

Les exigences de cette complémentarité

Les exemples qui viennent d'être évoqués montrent dans tous les cas que les recherches dont il a été question étaient conduites par des équipes de chercheurs et non par des personnes isolées. Au sein de ces équipes, se trouvent des spécialistes de disciplines différentes, parmi lesquels des statisticiens. Aujourd'hui, la complexité des mesures réalisées sur les données récoltées exige à la fois des outils informatiques performants et des personnes capables de les maîtriser. En fonction des hypothèses et des questions de départ, le statisticien doit pouvoir proposer les outils adéquats, expliquer les ré-

sultats récoltés et accompagner les spécialistes d'autres domaines dans leurs interprétations. Il s'agit donc, pour le moins, que les personnes impliquées dans ce genre de recherche soient ouvertes aux deux épistémologies, quantitative et qualitative, tant il est vrai que le quantitativiste doit comprendre l'articulation possible de ses résultats avec ceux provenant d'investigations qualitatives; quant au qualitatif, il doit être capable de lire et d'interpréter les résultats statistiques qui lui sont présentés pour pouvoir ensuite les mettre en relation avec d'autres données, qualitatives.

Le travail ne peut donc s'accomplir séparément, chacun s'occupant de son domaine sans se soucier de celui du voisin. Dès le départ, il s'agit de fixer ensemble les hypothèses, les questions, les bases méthodologiques, les étapes de la recherche, puis d'établir quels seront les unités et les outils d'analyse et comment procéder au traitement des résultats. Les différents partenaires sont associés à un travail commun, en vue de répondre à un problème particulier qu'ils se posent ensemble. C'est précisément la rencontre de points de vue différents, issus de disciplines scientifiques différentes, qui est féconde. Dans une équipe, il peut y avoir des didacticiens, des sociologues, des statisticiens, des enseignant-e-s. Le chercheur didacticien, en outre, est souvent originaire d'une autre discipline, que ce soit la psychologie, la linguistique, ou encore les lettres. On le voit, la recherche en éducation offre la possibilité d'un travail pluridisciplinaire permettant l'échange à propos d'un même problème à partir de champs disciplinaires différents. Encore faut-il que les institutions s'ouvrent à de telles collaborations.

Par le biais des travaux de Crahay, mais aussi au fil de notre texte, nous avons vu que la question de l'efficacité de l'enseignement et de sa mesure s'est posée de façon récurrente depuis des décennies. Cette question est tout à fait d'actualité, tests et enquêtes se multiplient pour mesurer les compétences des élèves de régions, de pays différents afin de comparer les effets des systèmes scolaires. D'aucuns se saisissent des résultats pour critiquer l'enseignement, l'école... Or, le plus souvent, la complexité et la multiplicité des variables susceptibles d'expliquer les résultats ne sont pas prises en considération par ces personnes, ni d'ailleurs les contenus et les conditions d'enseignement. La recherche a ainsi d'autant plus besoin de trouver des outils méthodologiques fins et multiples, lui permettant d'apporter des réponses fondées, concernant la réalité scolaire et les effets de l'enseignement. La réflexion méthodologique doit donc être poursuivie dans toutes les institutions s'occupant de recherche en éducation.

Références bibliographiques

- Aeby, S. (2004). *Récréature : évaluation d'un programme intensif d'enseignement/apprentissage de la lecture destiné à des élèves de 8^e année en grande difficulté*. Neuchâtel : IRDP.
- Bain, D., Erard, S. & Séchaud, M. (1994, éd. provisoire). *Comment aider les lecteurs faibles : propositions d'activités à l'usage des enseignants*. Genève : Cycle d'orientation, Groupe de français.
- Bayer, E. (1979). Sources de variance de l'indice d'influence de Flanders. *Scientia Paedagogica Experimentalis*, 3, 5-21.
- Besson, M.-J., Genoud, M.-R., Lipp, B. & Nussbaum, R. (1979). *Maîtrise du français*. Vevey : Delta.
- Béatrix Köhler, D., Nidegger, Ch., Revaz, N., Riesen, W. & Wirthner, M. (1999). *Et si les pratiques enseignantes en production écrite nous étaient contées...* Neuchâtel : IRDP ; Lausanne : L.E.P.
- Bishop, M.-F. & Sraïki, Ch. (2007). La prévention des difficultés en lecture au CP : un enjeu pour la formation des maitres. *Repères : recherches en didactique du français langue maternelle*, 36, 169-189.
- Bressoux, P. (1994). Les recherches sur les effets-écoles et les effets-maîtres. *Revue française de pédagogie*, 108, 91-137.
- Brophy, J. & Evertson, C. (1976). *Learning from teaching : a developmental perspective*. Boston : Allyn and Bacon.
- Brossard, M. (2004). *Vygotski : lectures et perspectives de recherches en éducation*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.
- Cain, K. (2009). Children's reading comprehension difficulties : a consideration of the precursors and consequences. In C. Wood & V. Connelly (Eds), *Contemporary Perspectives on Reading and Spelling* (pp. 59-75). London : Routledge.
- Candelier, M. (2003). *L'éveil aux langues à l'école primaire. Evlang : bilan d'une innovation européenne*. Bruxelles : De Boeck et Larcier.
- Cardinet, J. & Weiss, J. (1976). *L'enseignement de la lecture et ses résultats*. Berne : Peter Lang.
- Cèbe, S. & Goigoux, R. (2007). Concevoir un instrument didactique pour améliorer l'enseignement de la compréhension de textes. *Repères : recherches en didactique du français langue maternelle*, 35, 185-208.
- Chauveau, G. (2007). *Le savoir-lire aujourd'hui*. Paris : Retz.
- Crahay, M. (1988). Contraintes de situation et interactions maître-élève : changer sa façon d'enseigner, est-ce possible?. *Revue française de pédagogie*, 88, 67-94.
- Crahay, M. (2000). *L'école peut-elle être juste et efficace?*. Bruxelles : De Boeck.
- Crahay, M. (2006). *Un bilan des recherches processus-produit : l'enseignement peut-il contribuer à l'apprentissage des élèves et, si oui, comment?*. Genève : Université, FPSE (Carnets des sciences de l'éducation).
- Crahay, M. (2007). Feedback de l'enseignant et apprentissage des élèves : revue critique de la littérature de recherche. In L. Allal & L. Mottier Lopez (Dir), *Régulation des apprentissages en situation scolaire et en formation* (pp. 45-70). Bruxelles : De Boeck Université.
- Cronbach, L.J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Cronbach, L.J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116-127.
- Doyle, W. (1977). Paradigms for Research on Teacher Effectiveness. *Review of Research in Education*, 5, 163-198.
- Doyle, W. (1986). Paradigmes de recherches sur l'efficacité des enseignants. In M. Crahay & D. Lafontaine (Eds), *L'art de la science de l'enseignement* (pp. 435-481). Bruxelles : Labor.

- Favre, B. (1983). *L'enseignement de la lecture entre les exigences de la différenciation et les échéances du programme : premier compte-rendu de la recherche sur les pratiques de l'enseignement du français en 1P*. Genève : SRS.
- Feyfant, A. & Gaussel, M. (2007). *Méthodes de lecture et difficultés d'apprentissage* [Page Web]. Lyon : INRP (Dossier d'actualité 31). Accès : <http://www.inrp.fr/vst/LettreVST/31-novembre-2007.php> (Page consultée le 6 avril 2010).
- Fijalkow, J. (2000). *Sur la lecture : perspectives sociocognitives dans le champ de la lecture*. Issy-les-Moulineaux : ESF Editeur.
- Gambrell, L.B. (1996). Creating classroom cultures that foster reading motivation. *The Reading Teacher*, 50, - .
- Genelot, S. & Tupin, F. (2003). Dynamiques de classe et efficacité scolaire. *Les Dossiers des sciences de l'éducation*, 10, 109-130.
- Giasson, J. (2007). *La compréhension en lecture*. Bruxelles : De Boeck.
- Gilliéron Giroud, P. & Pulzer-Graf, P. (2009). *Apprentissage de la lecture et promotion de fin de 2e année : points de vue d'enseignantes sur les nouvelles dispositions institutionnelles*. Lausanne : URSP.
- Goigoux, R. (2002). Analyser l'activité d'enseignement de la lecture : une monographie. *Revue française de pédagogie*, 138, 125-134.
- Goigoux, R. (2006). La guerre des méthodes est finie. *Libération*, 6 novembre et *Les dossiers du Net* [Page Web]. Accès : <http://www.dossiersdunet.com/spip.php?article876> (Page consultée le 6 avril 2010).
- Goigoux, R. & Cèbe, S. (2006). *Apprendre à lire à l'école : tout ce qu'il faut savoir pour accompagner l'enfant*. Paris : Retz.
- Hutin, R. (1985). *Essais d'évaluation de la lecture en 1P-2P : contribution au renouvellement de l'enseignement du français*. Genève : DIP.
- Hutin, R. & Lurin, J. (1984). *Problèmes de lecture entre 8 et 10 ans*. Genève : DIP.
- Leu, N. & Soussi, A. (1995). *Dis-moi comment tu lis... ou Entretiens auprès d'élèves de 6P à propos de leur comportement et leurs stratégies de lecteurs*. Genève : SRP.
- Mauffrey, A. & Cohen, I. (1995). *Lecture : éléments pour une pédagogie différenciée*. Paris : Armand Colin.
- Moreau, J. (2001). Essai d'interprétation des résultats en fonction de l'arrière-fond culturel et social. In Ch. Nidegger (Ed.), *Compétences des jeunes romands : résultats de l'enquête PISA 2000 auprès des élèves de 9^e année* (pp. 139-151). Neuchâtel : IRDP.
- Nidegger, Ch. (Ed.). (2001). *Compétences des jeunes romands : résultats de l'enquête PISA 2000 auprès des élèves de 9^e année*. Neuchâtel : IRDP.
- Nonnon, E. & Goigoux, R. (2007). Travail de l'enseignant, travail de l'élève dans l'apprentissage initial de la lecture. *Repères : recherches en didactique du français langue maternelle*, 36, 5-36.
- Observatoire national de la lecture (ONL). (2005). *Rapport de l'ONL et de l'Inspection générale sur l'apprentissage de la lecture à l'école primaire*. Paris : ONL.
- Piaget, J. (1967). *Logique et connaissance scientifique : les méthodes de l'épistémologie*. Paris : Gallimard (Encyclopédie de la Pléiade).
- Piquée, C. & Sensevy, G. (2007). Lecture au cours préparatoire : une analyse empirique de l'influence des choix pédagogiques et didactiques. *Repères : recherches en didactique du français langue maternelle*, 36, 231-252.
- Renard, P. (2007). Gestion du temps d'enseignement de l'écrit à différentes échelles temporelles dans quatre CP de ZEP. *Repères : recherches en didactique du français langue maternelle*, 36, 37-58.
- Rieben, L. (2004). Le 21^{ème} siècle verra-t-il (enfin) la disparition des polémiques stériles sur l'apprentissage de la lecture?. *Formation et pratiques d'enseignement en questions*, 1, 17-25.
- Rosenshine, B. & Furst, N. (1973). The use of direct observation to study teaching. In R.M.W. Travers (Ed.), *Second Handbook of Research on Teaching* (pp. 122-183). Chicago : Rand Mc Nally.
- Saada-Robert, M., Auvergne, M., Balslev, K., Claret-Girard, V., Marzurczak, K. & Veuthey, C. (2003).

- Ecrire pour lire dès 4 ans : didactique de l'entrée dans l'écrit.* Genève : Université de Genève (Les Cahiers de la Section des sciences de l'éducation 100).
- Simon, F. (2007). Gérer l'hétérogénéité des savoirs des élèves dans une activité de lecture collective de texte au cours préparatoire. *Repères : recherches en didactique du français langue maternelle*, 36, 77-99.
- Stalling, J. & Kaskowitz, D. (1974). *Follow Through Classroom Observation Evaluation 1972-1973 (SRI Project URU-7370)*. Menlo Park : Stanford Research Institute.
- Terwagne, S., Vanhulle, S. & Lafontaine, A. (2001). *Les cercles de lecture : interagir pour développer ensemble des compétences de lecteurs*. Bruxelles : De Boeck et Duculot.
- Thévenaz-Christen, Th. et al. (à paraître). Ruptures et continuités de la lecture enseignée à l'école obligatoire. In R. Goigoux & M.-Ch. Pollet (Eds.), *De la maternelle à l'Université : aspects didactiques de la lecture*. Québec : Les Presses de l'Université Laval : AIRDF.
- Turner, R.L. (1971). Conceptual foundations of research in teacher education. In B.O. Smith (Ed.), *Research in teacher education : a symposium*. Englewood Cliffs : Prentice-Hall.
- Weiss, J. (1987). Quel enseignement pour une réduction des différences entre les élèves?. Neuchâtel : IRDP.

ANNEXES

Annexe A

EFFETS DES DIVERS TYPES D'ENSEIGNEMENT

TRAITEMENTS ET CRITÈRES	Résultats aux critères selon les traitements						TEST « Z » DE LA DIFFÉRENCE
	Peu d'exercice des Prérequis			Beaucoup d'exercice des Prérequis			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	* = Signif.
OPT. PRÉREQ.							
Burion	51,4	9,66	311	48,3	10,71	239	*3,51
Vitesse	51,4	8,91	311	48,5	10,56	233	*3,38
Précision	51,1	8,83	314	48,2	11,17	239	*3,30
Lect. Silencieux	49,6	10,76	336	49,3	9,55	266	0,36
Marmottes	49,7	9,99	257	48,4	9,77	244	1,47

APPR. ORALE	Peu d'analyse de l'oral			Beaucoup d'analyse de l'oral			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	49,6	10,63	315	50,6	9,38	281	-1,22
Vitesse	51,5	9,05	309	51,5	8,70	265	0,00
Précision	52,0	7,95	312	52,0	8,49	286	0,00
Lect. Silencieux	49,6	10,18	311	50,0	9,78	284	-0,49
Marmottes	49,4	9,92	301	50,0	10,10	279	-0,72

APPR. ÉCRITE	Pas de centration sur l'écrit			Priorité donnée à l'écrit			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	50,0	9,70	311	50,9	10,29	298	-1,11
Vitesse	49,2	10,60	301	51,0	9,08	273	*-2,19
Précision	50,5	9,81	305	51,3	8,46	294	-1,07
Lect. Silencieux	49,3	9,84	304	50,4	10,75	295	-1,31
Marmottes	48,6	10,00	302	50,3	10,03	266	*-2,01

ENRICHISSEM.	Peu d'enrichissement de la lecture			Souci d'enrichissement de la lecture			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	48,9	10,93	216	51,2	9,80	301	*-2,46
Vitesse	51,0	9,99	205	50,3	9,30	276	0,78
Précision	51,2	8,84	209	51,0	9,17	297	0,25
Lect. Silencieux	49,2	9,63	293	49,8	10,56	306	-0,73
Marmottes	49,0	10,04	233	50,0	9,80	281	-1,14

MÉTHODE LEC	Méthode S'Exprimer-Lire			Autres méthodes			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	48,9	10,11	802	49,9	10,28	603	-1,82
Vitesse	49,5	10,27	781	50,5	9,60	569	-1,83
Précision	50,3	9,53	798	51,2	8,63	597	-1,84
Lect. Silencieux	49,6	9,85	823	49,3	10,42	599	0,55
Marmottes	49,4	10,09	793	49,3	9,91	545	0,18

EXPÉRIENCE	Maîtresse peu expérimentée			Maîtresse expérimentée			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	49,0	10,25	731	49,5	10,08	608	-0,90
Vitesse	49,5	10,21	710	50,3	9,65	574	-1,44
Précision	50,1	9,49	724	51,5	8,55	605	*-2,82
Lect. Silencieux	49,7	9,89	763	49,0	10,21	600	1,27
Marmottes	49,4	9,93	676	49,2	10,10	562	0,35

Document: André Parisod

TEST DES DIFFÉRENCES DE DÉPART

PRÉDICTEURS ET CRITÈRES	Résultats aux tests d'aptitude de début d'année						TEST « Z » DE LA DIFFÉRENCE
	Peu d'exercice des Prérequis			Beaucoup d'exercice des Prérequis			* = Signif.
OPT. PRÉREQ.	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	49,5	5,32	311	48,9	5,46	239	1,292
Vitesse	49,5	5,43	311	49,0	5,54	233	1,051
Précision	49,5	5,43	314	49,0	5,53	239	1,062
Lect. Silenc.	49,5	5,38	336	49,0	5,38	266	1,132
Marmottes	49,2	5,64	257	48,9	5,43	244	0,607

APPR. ORALE	Peu d'analyse de l'oral			Beaucoup d'analyse de l'oral			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	49,4	5,06	315	49,4	5,57	281	0,000
Vitesse	49,7	5,01	309	49,4	5,49	265	0,679
Précision	49,4	5,04	312	49,5	5,52	286	-0,231
Lect. Silenc.	49,4	5,03	311	49,5	5,34	284	-0,235
Marmottes	49,3	4,97	301	49,5	5,44	279	-0,461

APPR. ÉCRITE	Pas de centration sur l'écrit			Priorité donnée à l'écrit			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	49,4	5,03	311	49,8	6,06	298	-0,884
Vitesse	49,4	4,97	301	50,2	5,98	273	-1,733
Précision	49,3	5,07	305	49,9	5,98	294	-1,322
Lect. Silenc.	49,5	4,77	304	49,7	6,08	295	-0,447
Marmottes	49,4	5,02	302	49,9	6,1	266	-1,058

ENRICHISSEM.	Peu d'enrichissement de la lecture			Souci d'élargir la lecture			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	49,6	4,99	216	49,4	5,34	301	0,436
Vitesse	49,7	4,94	205	49,7	5,27	276	0,000
Précision	49,6	4,98	209	49,5	5,36	297	0,215
Lect. Silenc.	49,5	4,91	293	49,5	5,23	306	0,000
Marmottes	49,5	4,94	233	49,7	5,11	281	-0,450

MÉTHODE LEC	Méthode S'Exprimer-Lire			Autres méthodes			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	49,7	5,15	802	49,1	5,36	603	*2,112
Vitesse	49,8	5,09	781	49,3	5,36	569	1,729
Précision	49,7	5,16	798	49,1	5,39	597	*2,095
Lect. Silenc.	49,7	5,01	823	49,1	5,42	599	*2,127
Marmottes	49,7	5,10	793	49,1	5,44	545	*2,033

EXPÉRIENCE	Maîtresse peu expérimentée			Maîtresse expérimentée			
	Moyennes	Sigmas	Nb. d'Élèves	Moyennes	Sigmas	Nb. d'Élèves	
Burion	49,7	5,12	731	49,5	5,28	608	0,700
Vitesse	49,8	5,08	710	49,7	5,27	574	0,344
Précision	49,7	5,16	724	49,5	5,31	605	0,693
Lect. Silenc.	49,8	4,97	763	49,5	5,34	600	1,061
Marmottes	49,8	5,09	676	49,5	5,33	562	1,006

Original: Communication personnelle d'André Parisod, Université de Lausanne, 1980.

Annexe B

RÉUSSITES MOYENNES DE 4 MÉTHODES DE LECTURE

Quatre méthodes de lecture fréquemment utilisées dans la canton de Neuchâtel en 1973 ont fait l'objet d'analyses spéciales, pour comparer les résultats qu'elles obtenaient des élèves. Cependant, elles ont dû être retirées de l'analyse multiniveau pour une raison de statistique mathématique. Comme chaque enseignante avait reçu un 1 pour une des quatre méthodes et des 0 pour les trois autres, les colonnes étaient linéairement dépendantes. La matrice des corrélations de ces variables était singulière, ce qui posait des problèmes pour le calcul des équations de régression. Or la méthode multiniveau exige beaucoup d'équations de régression. Les résultats relatifs à ces variables étaient en conséquence instables et non interprétables. Ils ont donc été retirés et remplacés par l'analyse ci-dessous qui montre le profil de résultats obtenu par chaque méthode. Les moyennes sont données pour les trois échelles factorielles mesurant respectivement la compréhension de la lecture, la précision de la lecture orale et la vitesse de la lecture. La connaissance de ce spectre de résultats permet des comparaisons plus riches que ce n'est le cas dans l'analyse multiniveau.

MÉTHODE SUIVIE	Score moyen de lecture		
	En Compréhension	En Précision	En Vitesse
S'Exprimer-Lire	- 0.043	- 0.155	- 0.066
Ensemble des autres	+ 0.555	+ 0.137	+ 0.046

MÉTHODE SUIVIE	Score moyen de lecture		
	En Compréhension	En Précision	En Vitesse
Gestuelle	+ 0.085	+ 0.236	+ 0.062
Ensemble des autres	- 0.021	- 0.074	- 0.027

MÉTHODE SUIVIE	Score moyen de lecture		
	En Compréhension	En Précision	En Vitesse
Le Sablier	- 0.182	- 0.107	+ 0.063
Ensemble des autres	+ 0.014	- 0.015	- 0.020

MÉTHODE SUIVIE	Score moyen de lecture		
	En Compréhension	En Précision	En Vitesse
Syllabique	- 0.026	+ 0.223	- 0.250
Ensemble des autres	- 0.001	- 0.044	+ 0.007

On voit ainsi que S'Exprimer-Lire, qui était utilisé par les jeunes enseignantes sortant de l'École Normale, n'a pas de bons résultats comparés aux autres en Compréhension, et surtout en Précision (différence significative), mais ne se distingue presque pas en vitesse.

La méthode Gestuelle l'emporte essentiellement en Précision de la lecture orale ($p < .05$).

Le Sablier a des problèmes en Compréhension et en Précision, mais se rattrape en Vitesse.

La méthode Syllabique est excellente en Précision, mais mauvaise en Vitesse.

Notons toutefois que les deux dernières méthodes n'étaient pratiquées que par 8 et 7 maîtresses et que leurs moyennes sont ainsi sujettes à caution.

Remarque importante: dans ces analyses, on ne peut pas séparer les effets de la méthode et ceux découlant du niveau de départ des élèves qui ont suivi chaque forme d'enseignement. Les résultats de l'Annexe A montrent que les élèves étaient de niveau différent pour certaines comparaisons. La Méthode S'Exprimer-Lire était employée dans de bonnes classes, par exemple.

Annexe C

EXEMPLE DE RAPPORT IMPRIMÉ PAR LE LOGICIEL HLM

Équations de régression des Niveaux 1 et 2 et ajustement des pentes

Problem Title: CONSTANTES ET VARIANCES RESIDUELLES POUR CRITERE MOYEN

Level-1 Model

$$Y = B0 + B1*(SEXE) + B2*(NIVPROF) + B3*(EXPRORAL) + B4*(REGULIER) + B5*(PRELECTU) + B6*(DONNEDED) + R$$

Level-2 Model

$$B0 = G00 + G01*(NBELEVCL) + G02*(SEXECL) + G03*(NIVPROCL) + G04*(EXPORACL) + G05*(REGULICL) + G06*(PRELECCL) + G07*(DONNE_CL) + G08*(NB_DEGCL) + G09*(IMPDECCL) + G010*(OBJELECL) + G011*(OPTLETCL) + G012*(MOTEXTCL) + G013*(PREPARCL) + G014*(MOTINTCL) + G015*(OBJAFFCL) + G016*(RAPIAVCL) + G017*(BIBLIOCL) + G018*(EQUIPECL) + G019*(APORALCL) + G020*(APECRICL) + G021*(APPRERCL) + G022*(APENRICL) + G023*(EXLIRECL) + G024*(GESTULCL) + G025*(SABLIECL) + G026*(SYLLABCL) + G027*(ANNPRACL) + G028*(SATISFCL) + U0$$

$$B1 = G10 + U1$$

$$B2 = G20 + U2$$

$$B3 = G30 + U3$$

$$B4 = G40 + U4$$

$$B5 = G50 + U5$$

$$B6 = G60 + U6$$

ITERATION 693 *****

Sigma_squared = 0.43518

Standard Error of Sigma_squared = 0.01793

Tau

INTRCPT1,	B0					
0.10661	-0.00562	0.00399	-0.00821	-0.00316	-0.00201	-0.00721
SEXE,	B1					
-0.00562	0.00328	0.00142	0.00218	0.00063	-0.00082	-0.00389
NIVPROF,	B2					
0.00399	0.00142	0.01330	0.00047	-0.00440	-0.00340	-0.00126
EXPRORAL,	B3					
-0.00821	0.00218	0.00047	0.00210	0.00081	0.00001	-0.00325
REGULIER,	B4					
-0.00316	0.00063	-0.00440	0.00081	0.00584	0.00281	-0.00222
PRELECTU,	B5					
-0.00201	-0.00082	-0.00340	0.00001	0.00281	0.00267	-0.00151
DONNEDED,	B6					
-0.00721	-0.00389	-0.00126	-0.00325	-0.00222	-0.00151	0.01418

Standard Errors of Tau

INTRCPT1, B0	0.01973	0.00664	0.00819	0.00724	0.00866	0.00916	0.00902
SEXE, B1	0.00664	0.00445	0.00378	0.00326	0.00393	0.00417	0.00415
NIVPROF, B2	0.00819	0.00378	0.00649	0.00409	0.00516	0.00531	0.00521
EXPRORAL, B3	0.00724	0.00326	0.00409	0.00477	0.00419	0.00460	0.00460
REGULIER, B4	0.00866	0.00393	0.00516	0.00419	0.00630	0.00548	0.00536
PRELECTU, B5	0.00916	0.00417	0.00531	0.00460	0.00548	0.00631	0.00585
DONNEDED, B6	0.00902	0.00415	0.00521	0.00460	0.00536	0.00585	0.00777

Tau (as correlations)

INTRCPT1,B0	1.000	-0.301	0.106	-0.549	-0.127	-0.119	-0.185
SEXE,B1	-0.301	1.000	0.215	0.831	0.144	-0.277	-0.570
NIVPROF,B2	0.106	0.215	1.000	0.090	-0.499	-0.571	-0.092
EXPRORAL,B3	-0.549	0.831	0.090	1.000	0.232	0.005	-0.595
REGULIER,B4	-0.127	0.144	-0.499	0.232	1.000	0.712	-0.244
PRELECTU,B5	-0.119	-0.277	-0.571	0.005	0.712	1.000	-0.245
DONNEDED,B6	-0.185	-0.570	-0.092	-0.595	-0.244	-0.245	1.000

Random level-1 coefficient Reliability estimate

INTRCPT1, B0	0.823
SEXE, B1	0.089
NIVPROF, B2	0.247
EXPRORAL, B3	0.050
REGULIER, B4	0.104
PRELECTU, B5	0.040
DONNEDED, B6	0.219

Note: The reliability estimates reported above are based on only 45 of 91 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

 The value of the likelihood function at iteration 693 = -1.725229E+003

The outcome variable is CRITMSEL

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT 1, B0					
INTRCPT2, G00	-0.015912	0.038371	-0.415	62	0.678
NBELEVCL, G01	0.068167	0.051833	1.315	62	0.189
SEXECL, G02	0.078115	0.258752	0.302	62	0.763
NIVPROCL, G03	0.030949	0.146366	0.211	62	0.833
EXPORACL, G04	-0.102183	0.143816	-0.711	62	0.477
REGULICL, G05	0.325626	0.197233	1.651	62	0.098
PRELECCL, G06	-0.297326	0.109328	-2.720	62	0.007
DONNE_CL, G07	0.486235	0.105378	4.614	62	0.000
NB_DEGCL, G08	0.012721	0.050408	0.252	62	0.801
IMPDECCL, G09	0.060757	0.048531	1.252	62	0.211
OBJECL, G010	-0.002738	0.049220	-0.056	62	0.956
OPTLETCL, G011	0.131357	0.100932	1.301	62	0.193
MOTEXTCL, G012	-0.049398	0.054148	-0.912	62	0.362
PREPARCL, G013	-0.053586	0.048639	-1.102	62	0.271
MOTINTCL, G014	0.175688	0.046833	3.751	62	0.000
OBJAFFCL, G015	-0.081973	0.051289	-1.598	62	0.110
RAPIAVCL, G016	0.117268	0.052308	2.242	62	0.025
BIBLIOCL, G017	-0.085531	0.056506	-1.514	62	0.130
EQUIPECL, G018	0.100510	0.057368	1.752	62	0.079
APORALCL, G019	0.023938	0.056324	0.425	62	0.670
APECRICL, G020	-0.060514	0.063219	-0.957	62	0.339
APPRERCL, G021	-0.036678	0.045259	-0.810	62	0.418
APENRICL, G022	0.038854	0.054213	0.717	62	0.473
EXLIRECL, G023	-0.185145	0.068802	-2.691	62	0.008
GESTULCL, G024	-0.129164	0.063489	-2.034	62	0.042
SABLIECL, G025	-0.198374	0.061430	-3.229	62	0.002
SYLLABCL, G026	-0.118950	0.060815	-1.956	62	0.050
ANNPRACL, G027	-0.048372	0.065552	-0.738	62	0.461
SATISFCL, G028	0.173579	0.049942	3.476	62	0.001
For SEXE slope, B1					
INTRCPT2, G10	0.071707	0.018279	3.923	90	0.000
For NIVPROF slope, B2					
INTRCPT2, G20	0.017444	0.022569	0.773	90	0.440
For EXPRORAL slope, B3					
INTRCPT2, G30	0.020581	0.019988	1.030	90	0.304
For REGULIER slope, B4					
INTRCPT2, G40	0.134677	0.024060	5.598	90	0.000
For PRELECTU slope, B5					
INTRCPT2, G50	0.109754	0.025391	4.323	90	0.000
For DONNEDED slope, B6					
INTRCPT2, G60	0.518371	0.024847	20.862	90	0.000

The outcome variable is CRITMSEL

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	0.32652	0.10661	16	169.12981	0.000
SEXE slope,	U1	0.05727	0.00328	44	54.97102	0.124
NIVPROF slope,	U2	0.11533	0.01330	44	58.05077	0.076
EXPRORAL slope,	U3	0.04583	0.00210	44	37.15961	>.500
REGULIER slope,	U4	0.07641	0.00584	44	65.70268	0.018
PRELECTU slope,	U5	0.05171	0.00267	44	47.47388	0.333
DONNEDED slope,	U6	0.11908	0.01418	44	71.71162	0.005
level-1,	R	0.65968	0.43518			

Note: The chi-square statistics reported above are based on only 45 of 91 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Statistics for current covariance components model

Deviance = 3450.457637

Number of estimated parameters = 64

ANNEXE D

ÉCHELLES DES VARIABLES UTILISÉES

LEVEL-1 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
SEXE	1605	-0.00	1.00	-0.97	1.03
NIVPROF	1605	-0.00	1.00	-0.69	2.55
EXPRORAL	1605	-0.00	1.00	-2.37	0.62
REGULIER	1605	-0.00	1.00	-3.96	0.27
PRELECTU	1605	0.00	1.00	-0.45	6.19
DONNEDED	1605	0.00	1.00	-4.26	2.27
CRITMSEL	1605	0.00	1.00	-3.89	3.06

LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
NBELV_CL	91	-0.00	1.00	-3.83	2.31
GENRE_CL	91	-0.00	0.19	-0.43	0.58
AGE_CL	91	-0.01	0.25	-0.59	0.66
NPROF_CL	91	0.03	0.34	-0.57	1.04
ORALF_CL	91	0.03	0.36	-1.14	0.62
REGUL_CL	91	0.02	0.24	-1.02	0.27
PREL_CL	91	-0.02	0.44	-0.45	1.77
DONN_CL	91	0.00	0.46	-1.08	1.60
NBDEG_CL	91	0.00	1.00	-0.37	2.68
DECOD_CL	91	-0.00	1.00	-3.15	2.45
OBELV_CL	91	0.00	1.00	-1.88	2.36
OPLET_CL	91	0.00	1.00	-1.63	2.53
MOEXT_CL	91	-0.00	1.00	-1.88	2.74
EXPRE_CL	91	0.00	1.00	-2.31	3.36
MOINT_CL	91	0.00	1.00	-3.58	1.90
OBJAF_CL	91	-0.00	1.00	-2.26	1.77
METHS_CL	91	0.00	1.00	-2.29	3.00
RAPID_CL	91	0.00	1.00	-2.43	3.15
BIBLI_CL	91	0.00	1.00	-1.61	2.31
EQUIP_CL	91	0.00	1.00	-2.01	2.21
AORAL_CL	91	0.00	1.00	-2.21	2.86
AECRI_CL	91	-0.00	1.00	-1.55	4.49
APRER_CL	91	0.00	1.00	-2.56	2.90
AENRI_CL	91	0.00	1.00	-1.91	3.96
EXLIR_CL	91	-0.00	1.00	-1.16	1.11
GESTU_CL	91	-0.00	1.00	-0.77	2.21
SABLI_CL	91	-0.00	1.00	-0.31	3.20
SYLLA_CL	91	0.00	1.00	-0.64	3.43
PRATI_CL	91	-0.00	1.00	-0.71	5.33
SATIS_CL	91	-0.00	1.00	-3.49	1.09