

**THÈSE DE DOCTORAT**

**Psychologie cognitive**

**NORA DITTMANN-DOMENICHINI**

Mise à l'épreuve d'un modèle de lecture  
auprès de jeunes lecteurs monolingues et  
plurilingues en Suisse:  
la „Simple View of Reading Theory“

**Thèse dirigée par: Daniel MARTINS**

**ED 139 Connaissances, Langage, Modélisation**

**Soutenu : Janvier / Février 2013**

Prof. D. Legros	Université Paris 8	Rapporteur
Prof. D. Brouillet	Université Montpellier 3	Rapporteur
Prof. P. Rozencwaig	Université Paris 8	Membre du jury
Prof. R. Müller	HEP Berne	Membre du jury
Prof. E. Ramseier	HEP Berne	Membre du jury

*Remerciement*

Cette thèse est le fruit d'un long travail mené souvent dans la solitude, mais non sans soutien. Je sais gré à un grand nombre de personnes qui m'ont encadré, inspiré, aiguillé ou tout simplement aidé.

Je souhaite tout d'abord remercier Monsieur Daniel Martins pour sa confiance en mon travail. Merci pour toutes les relectures toujours très constructives et rapides qui m'ont beaucoup aidé à progresser.

Un grand merci revient également à l'équipe du projet de recherche. Je tiens en particulier à remercier Romano Müller qui a su me transmettre son enthousiasme pour un travail de recherche ambitieux et rigoureux. Cher Romano j'ai beaucoup appris de toi pendant ces trois années et j'aurais sincèrement aimé que cela puisse continuer. Merci évidemment aussi à Stephan Rösselet et Jeannine Khan-Bol, mes deux co-équipiers. Le soutien que nous avons pu nous donner mutuellement, restera certainement l'une des meilleurs souvenirs de mon parcours professionnel.

Merci à Erich Ramseier, dont les conseils avisés ont beaucoup aidé à ce que je ne me perde pas dans la jungle des analyses statistiques multi-niveaux.

Merci aux enseignants et à leurs élèves qui par leur précieuse collaboration ont permis la construction d'un échantillon conséquent sur lequel repose l'ensemble de ce travail de thèse.

Enfin merci à toi Jérôme, mon mari, pour ton soutien infaillible tant moral que pratique par la relecture de mes pages au fil du temps. Tu sais bien à quel point je t'en suis reconnaissante. Et merci à mes filles Eléa et Mei qui ont toujours assuré le fait que mon travail devait se faire de façon efficace et dans un temps très précis. Je ne pouvais ainsi ni me perdre dans les détails ni me méprendre sur l'importance et la valeur de chaque chose dans ma vie.

*Résumé*

Basé sur une étude empirique, la présente thèse apporte un éclairage sur les compétences de lecture, de décodage et de compréhension orale en allemand chez des élèves plurilingues immigrés et leurs homologues monolingues suisses allemands. Un échantillon d'environ 500 élèves plurilingues et de 160 élèves monolingues constitue la base empirique de ce travail qui nous a servi à examiner certains éléments de controverses ainsi que la capacité prédictive de la théorie de la «*Simple View of Reading*» ( Philipp B. Gough & Tunmer, 1986 p. 85; Hoover & Gough, 1990).

Nous constatons que la compréhension orale et le décodage possèdent une forte valeur prédictive pour les compétences en lecture. Cette valeur prédictive dépend cependant fortement de la façon dont les compétences en lecture sont opérationnalisées et elle est plus élevée dans le groupe des lecteurs débutants (en CE1/CE2). La modélisation des compétences de lecture d'après la «*Simple View of Reading Theory*» réussit mieux pour des tests de lecture courts, à choix multiples et sous contrainte de temps que pour des tests à base de textes plus longs et authentiques, qui proposent différents formats de réponses. L'intégration du facteur de l'intelligence dans le modèle de prédiction est plus bénéfique pour le groupe des lecteurs plus expérimentés (CM2/6<sup>ème</sup>). L'utilisation de la méthode des analyses de régressions multiples à plusieurs niveaux (MLwiN) a mis en évidence que l'apport du QI est plus important lorsqu'on examine les compétences au niveau des classes qu'au niveau des élèves pris individuellement. Nous écartons donc la proposition d'une substitution du facteur compréhension orale au profit du QI comme Tiu, Thompson et Lewis (2003) l'ont proposé.

mots-clés : «*Simple View of Reading Theory*», lecture, décodage, compréhension orale, intelligence analyses à niveaux multiples

*Abstract*

This work aimed to examine controversial elements as well as the explanatory strength of the “*Simple View of Reading Theory*”. It has been based on empirical data analyses of language abilities of about 500 multilingual and 160 monolingual school graders in the German speaking part of Switzerland (Philipp B. Gough & Tunmer, 1986; Hoover & Gough, 1990).

We found evidence in both groups (monolingual and multilingual pupils) for a strong impact of listening comprehension and decoding in explaining reading comprehension abilities. The degree of the explained variance was determined by the test used to evaluate reading comprehension and by the age of pupils, showing a bigger effect with young readers (2<sup>nd</sup>/3<sup>rd</sup> graders) than with more experienced readers (5<sup>th</sup> / 6<sup>th</sup> graders). Modeling reading comprehension abilities based on the «*Simple View of Reading Theory*» proved to be more successful when using short, time limited, multiple choice reading tests than using a test with longer, authentic texts and offering a mixture of different response formats. The integration of the intelligence factor into our model improved the prediction of reading abilities in the group of the more experienced readers (5<sup>th</sup> / 6<sup>th</sup> graders). The improvement though was essentially situated at the level of school classes as unit of analyses. On the individual level only minor improvements were observed. Therefore we would reject the idea of eliminating the listening comprehension factor in favor of an integration of the IQ factor that has been suggested by Tiu, Thompson and Lewis (2003).

Keywords : « simple view of reading », decoding, listening comprehension, reading comprehension, intelligence, multilevel analyses

LE VALET. - Dieu vous donne le bonsoir !... Dites-moi, monsieur savez-vous lire ?

ROMÉO. - Oui, ma propre fortune dans ma misère.

LE VALET. - Peut-être avez-vous appris ça sans livre; mais, dites-moi, savez-vous lire le premier écrit venu ?

ROMÉO. - *Oui, si j'en connais les lettres et la langue.*

LE VALET. - Vous parlez congrûment. Le ciel vous tienne en joie.

(Shakespeare, 1597)

## Table des matières

Propos liminaires .....	1
Introduction .....	6
1 Partie théorique .....	9
1.1 Les modèles de la lecture.....	9
1.2 Revue historique des paradigmes prédominants .....	10
1.2.1 Les débuts de la recherche sur le processus de lecture.....	10
1.2.2 La lecture comme un processus d'apprentissage conditionné .....	11
1.2.3 L'apprentissage naturel : la lecture comme capacité inhérente.....	12
1.2.4 La lecture un traitement d'information.....	15
1.2.5 Modèles développementaux de la lecture .....	24
1.3 Le modèle de la « <i>Simple View of Reading Theory</i> » : $L = D \times CO$ .....	28
1.3.1 Les composantes de la « <i>Simple View of Reading Theory</i> » : $L = D \times CO$ .....	30
1.3.2 Analyse critique des prémisses de la « <i>SVRT</i> ».....	37
1.3.3 Données empiriques concernant la « <i>Simple View of Reading Theory</i> » .....	38
2 Questions de recherche.....	48
2.1 La validité d'un modèle additif de la « <i>Simple View of Reading Theory</i> » .....	48
2.2 Le rôle des opérationnalisations de la compréhension de texte.....	48
2.3 Lecture en L2 .....	50
2.4 Le QI un troisième facteur potentiel?.....	51
3 Méthode.....	56
3.1 Description de l'échantillon.....	56
3.2 Sélection de l'échantillon pour le présent travail de thèse .....	58
3.3 Déroulement du recueil des données.....	65
3.4 Traitement des données .....	67
3.5 Caractéristiques et utilisation du « <i>modèle de Rasch</i> » .....	69
3.5.1 Estimation de paramètres avec le modèle de Rasch.....	70
3.5.2 Vérification de la validité du modèle de Rasch.....	70
3.5.3 Les avantages de l'utilisation du « <i>modèle de Rasch</i> » .....	71
3.5.4 Les analyses selon le « <i>modèle de Rasch</i> » dans la présente étude.....	72
3.6 Structure des données et méthode d'analyse multi-niveaux.....	73
3.6.1 Les modèles d'analyse multi-niveaux.....	73
3.6.2. Modèle conditionnel de croissance.....	75
3.6.3 La centration des variables.....	80
3.6.4 Les inférences statistiques dans le cadre des modèles multi-niveaux .....	80
3.7 Opérationnalisation des variables indépendantes.....	83
3.8. Opérationnalisation des variables dépendantes.....	99
4 Résultats .....	108

---

4.1 Validation de l'importance des deux composantes D et CO .....	108
4.1.1 Analyse des coefficients de régression .....	108
4.1.2 Analyse des coefficients de détermination des composantes D et CO....	112
4.2 Résultats des analyses auprès des élèves en CM2/6 <sup>ème</sup> .....	120
4.3- L'influence du plurilinguisme .....	130
4.4- L'impact du QI en tant que troisième facteur possible de la SVRT .....	140
4.4.1 Analyse des coefficients de régression .....	140
4.4.2 Analyse des coefficients de détermination.....	142
5 Discussion générale et conclusion .....	149
5.1 Résumé des résultats et interprétations.....	149
5.1.1 Impact et force explicative des facteurs D et CO .....	150
5.1.2 L'impact de l'opérationnalisation des compétences en lecture.....	152
5.1.3 Elèves plurilingues et monolingues .....	154
5.1.4 L'apport du QI comme 3 <sup>ème</sup> facteur.....	155
5.2 Limites des études et recherches futures .....	157
Bibliographie .....	160
Annexes.....	173

## Table des figures

Figure 1 Courants théoriques au cours du 20 <sup>ème</sup> siècle .....	11
Figure 2 Caractère à multiples facettes de la lecture (Scarborough, 2001).....	30
Figure 3: Répartition des langues d'origines dans l'échantillon sélectionné.....	60
Figure 4a Préférences de langue en CE1.....	61
Figure 4b Préférences de langue en CM2 .....	61
Figure 5 Distribution des catégories socioprofessionnelles .....	62
Figure 6 Fiche standardisée servant à numériser les données recueillies .....	68
Figure 7 Fonction caractéristique de l'item .....	70
Figure 8 Illustration de l'hiérarchisation de l'échantillon.....	75
Figure 9 Extrait du ELFE partie 1 .....	83
Figure 10 Exemple de question de Compréhension orale (HarmoS_L1) .....	86
Figure 11 Exemple « Repérer des relations spatiales » .....	93
Figure 12 Extrait du questionnaire pour les élèves.....	95
Figure 13 Extrait du questionnaire concernant le statut socioéconomique.....	96
Figure 14 Extrait du questionnaire concernant le temps de lecture extra-scolaire.....	98
Figure 15 Exemple de question du test SLS 1-4 .....	99
Figure 16 Exemple de question du test ELFE partie 3 (compréhension de texte). .....	101
Figure 17 Extrait du test HarmoS-L1 pour la compréhension de texte (CE1/CE2).....	103
Figure 18 Effets d'interactions entre les facteurs D et CO pour le SLS .....	111
Figure 19 Interaction entre D et les compétences de lecture selon le test SLS.....	111
Figure 20 Effets d'interactions entre les facteurs D et CO et les résultats du ELFE....	111
Figure 21 Effets d'interactions entre D et CO et les résultats du test HarmoS.....	112
Figure 22 Droite de régression test SLS .....	122
Figure 23 Droites de régression test ELFE .....	122
Figure 24 Droites de régression test HarmoS.....	122
Figure 25 Evolution de la Compréhension Orale entre le CE1 et le CE2.....	132

## Tableaux

Tableau 1 : Echantillon global lors des deux vagues de recueil de données.....	56
Tableau 2 : Indicateurs sociodémographiques et Index Social .....	57
Tableau 3 : Caractéristiques de l'échantillon sélectionné .....	59
Tableau 4 : Catégorie socioprofessionnelle la plus élevée des parents .....	59
Tableau 5: Proportions des langues d'origine dans l'échantillon sélectionné.....	60
Tableau 6 Comparaisons des groupes monolingues versus plurilingues.....	62
Tableau 7 Catégories socioprofessionnelles des parents d'élèves.....	63
Tableau 8 Comparaison du nombre de livres à la maison .....	63
Tableau 9 Temps consacré à la lecture en allemand en dehors de l'école.....	64
Tableau 10 Résumé du recueil des données .....	65
Tableau 11 Indices de qualité du test ELFE partie 1 (Test de Décodage).....	84
Tableau 12 Distribution de ZWortRoh_2bis3 et ZWortRoh_5bis6.....	85
Tableau 13 Extrait du modèle de HarmoS-L1 .....	87
Tableau 14 Les exercices d'compréhension orale pour les CE1/CE2 .....	88
Tableau 15 Les exercices d'compréhension orale pour les CM2/6ème.....	88
Tableau 16 Indices de qualité du test HarmoS (Compréhension orale).....	89
Tableau 17 Analyses de composantes principales (Harmos Compréhension orale) .....	90
Tableau 18 Indicateurs de « <i>fit</i> » et de difficulté .....	91
Tableau 19 Indicateurs de « <i>fit</i> » et de difficulté (CM2/6 <sup>ème</sup> ).....	92
Tableau 20 Distribution de ZLogit_HörenGesamt_23 et ZLogit_Hören-Gesamt56.....	92
Tableau 21 Indices de qualité du Test CFT 20R .....	94
Tableau 22 Distribution de la variable CFTIQTOT .....	94
Tableau 23 Distribution de la variable MEHSPRAS01 .....	95
Tableau 24 Distribution de l'échelle HOEPOSELk .....	97
Tableau 25 Distribution de la variable LESDE .....	98
Tableau 26 Indices de qualité issus du manuel du SLS 1-4 .....	100
Tableau 27 Distribution des variables ZSLS_pkt23 et Z_SLSPkt_5bis6 .....	100
Tableau 28 Indices de qualité issus du manuel de test ELFE 1-6 .....	102
Tableau 29 Distribution des variables ZTextRoh_2bis3 et ZTextRoh_5bis6.....	102
Tableau 30 Les exercices d'compréhension orale pour les CM2/6ème.....	104
Tableau 31 Les exercices d'compréhension orale pour les CE1/CE2 .....	104
Tableau 32 Indices de qualité extrait 2 du tableau WINSTEPS 3.1 .....	105
Tableau 33 Analyses à composantes principales (HarmonS Compréhension de texte). 105	
Tableau 34 Indicateurs de « <i>fit</i> » et de difficulté (CM2/6 <sup>ème</sup> ) .....	106
Tableau 35 Indicateurs de « <i>fit</i> » et de difficulté (CE1/CE2).....	106
Tableau 36 Distribution des variables Z_Lesen_2bis3_01 et Z_Lesen_5bis9_01 .....	107
Tableau 37 Résultats des analyses multi-niveaux du modèle en CE1 / CE 2 .....	109
Tableau 38.1 Evolution des modèles multi niveaux hiérarchisés relatifs au SLS.....	114
Tableau 38.2 Evolution des taux de variances expliquées relatif au SLS.....	114

---

Tableau 39.1 Evolution des modèles multi-niveaux hiérarchisés de l'ELFE .....	115
Tableau 39.2 Evolution des taux de variances expliquées relatifs à l'ELFE .....	115
Tableau 40.1 Evolution des modèles concernant le test de HarmoS.....	116
Tableau 40.2 Evolution des taux de variances expliquées relatifs au test de HarmoS .	116
Tableau 41 Résultats des analyses multi-niveaux du modèle en CM2 / 6ème.....	120
Tableau 42.1 Evolution des modèles concernant le SLS en CM2 / 6ème .....	125
Tableau 42.2 Les taux de variances expliquées relatif au SLS en CM2/ 6ème.....	125
Tableau 43.1 Evolution des modèles concernant le ELFE en CM2 / 6ème.....	126
Tableau 43.2 Les taux de variances expliquées relatif au ELFE en CM2/ 6ème .....	126
Tableau 44.1 Evolution des modèles concernant le HarmoS en CM2 / 6ème .....	127
Tableau 44.2 Les taux de variances expliquées relatif au HarmoS en CM2/ 6ème .....	127
Tableau 44 Résultats du modèle 4 chez les élèves en CE1/CE2.....	131
Tableau 44.2 Résultats du modèle 1.2 .....	133
Tableau 45 Evolution des modèles du SLS chez les monolingues et plurilingues.....	135
Tableau 46 Evolution des modèles du ELFE chez les monolingues et plurilingues ....	136
Tableau 47 Evolution des modèles du HarmoS chez les monolingues et plurilingues	137
Tableau 48 L'impact du QI .....	141
Tableau 49 Evolution du modèle 1.2 au modèle 2 .....	143
Tableau 50 Evolution du modèle 1.2 au modèle 2 pour l'ELFE.....	144
Tableau 51 Evolution du modèle 1.2 au modèle 2 pour le test de HarmoS.....	145
Tableau 52 Analyse des coefficients de régression .....	146

## Propos liminaires

Le simple fait que le présent travail de thèse ait pour sujet les compétences en lecture en allemand, qu'il se base sur un échantillon d'élèves en Suisse et des données recueillies à l'aide de tests germanophones, justifie la présence de propos liminaires. Leur objectif est d'apporter les clarifications et explications nécessaires à la compréhension du cadre dans lequel ce travail a été réalisé.

La présente thèse fait partie intégrante d'un projet de recherche<sup>1</sup> réalisé au sein de la Haute Ecole de Pédagogie<sup>2</sup> (HEP) de Berne sous la direction de Romano Müller, professeur de psychologie. L'objectif de ce projet de recherche était double : 1) analyser le développement des compétences en langue scolaire (écriture, compréhension orale, compréhension de texte) chez des élèves plurilingues et monolingues en tenant compte de l'impact de caractéristiques spécifiques des contextes scolaires et familiaux ; 2) soutenir de jeunes chercheurs en leur permettant de réaliser un travail de thèse dans le cadre de ce projet.

Quatre thèses ont ainsi été réalisées dans le cadre de ce projet de recherche, qui à son tour a bénéficié du travail coordonné d'une équipe de 4 à 5 personnes. Cela a plusieurs implications dans le présent travail. D'un côté, la réalisation de cette thèse a pu profiter du soutien financier et des moyens en termes de ressources humaines dont bénéficiait le projet. Evaluer les compétences linguistiques de plus de 600 élèves à l'aide de différents tests et à deux reprises, aurait été impossible sans l'appui de toute une équipe de recherche. D'un autre côté, cette thèse a aussi été conditionnée par le projet de recherche global, notamment sur le plan méthodologique. Ainsi, le choix des instruments et des tests de compétences linguistiques a été principalement déterminé par les besoins du projet global et seulement en deuxième lieu par les besoins de la présente thèse. Le temps mis à disposition par les enseignants et par les élèves devait également servir à évaluer les compétences en écriture et à recueillir de multiples données sur les variables des différents contextes scolaires et familiaux qui ne sont pas au centre du présent travail. Cer-

---

<sup>1</sup>[www.sprachlichekompetenzen.ch](http://www.sprachlichekompetenzen.ch)

<sup>2</sup>Équivalent Suisse de l'Institut Universitaire de Formation des Maîtres (IUFM).

tains résultats du projet global seront cependant cités, afin d'expliquer l'approche adoptée au cours du présent travail, chaque fois que cela paraîtra nécessaire.

Les instruments utilisés pour évaluer les compétences en compréhension de texte et en compréhension orale en allemand seront présentés de façon plus détaillée que s'ils avaient été des instruments bien connus en langue française. La littérature citée dans la présente thèse est majoritairement anglophone ou germanophone. Lorsqu'une citation germanophone est utilisée, elle sera traduite en français le plus fidèlement possible par l'auteur et restituée en bas de page dans sa version originale.

Ce qui distingue, entre autres, la langue allemande de la langue française est le fait que l'allemand est une langue transparente, tandis que le français est une langue opaque. La relation grapho-phonémique est plus régulière et moins complexe en allemand qu'en français ou en anglais. Cela peut avoir des impacts sur l'apprentissage de la lecture et de l'écriture. Des études ont ainsi démontré que les lecteurs dyslexiques allemands ne souffrent pas -contrairement aux jeunes lecteurs dyslexiques anglophones- d'un problème de conscience phonologique ni de segmentation phonologique (Landerl & Wimmer, 2000; Wimmer et al., 2000). Les résultats obtenus auprès d'un échantillon germanophone suisse ne peuvent donc pas prétendre avoir une validité dans d'autres systèmes langagiers.

Pour les enfants suisses alémaniques -notre échantillon a été recruté dans le canton de Zürich (cf. annexe 1)- s'ajoute encore une spécificité supplémentaire due à la présence prépondérante du dialecte dans les communications verbales quotidiennes. Pour un enfant suisse alémanique c'est le dialecte local<sup>3</sup> qui est souvent considéré comme la langue première, c'est-à-dire la langue dans laquelle il a appris à parler et dans laquelle il se sent « chez lui ». Il s'agit là d'un autre exemple de différence entre le français qui est une langue plutôt normée par rapport à l'allemand qui est une langue pluricentriste avec une large variété de dialectes que ce soit en Suisse ou en Autriche, mais également à l'intérieur même des frontières de la République Fédérale d'Allemagne (Berthele, 2001).

Le haut allemand<sup>4</sup> en Suisse est souvent associé à l'enseignement scolaire et parfois même considéré comme une deuxième langue, distincte de la première langue: le dia-

---

<sup>3</sup> Les dialectes suisses alémaniques diffèrent considérablement d'une ville à l'autre (ex. dialecte zurichois, dialecte bernois ou le dialecte bâlois) et les différentes régions (ex. le dialecte valaisan). Ils sont fortement liés au sentiment d'identité régionale.

<sup>4</sup> « Hoch Deutsch » qui correspond à l'allemand standard

lecte. La langue de scolarisation est uniquement le haut allemand, même s'il peut arriver en primaire que les enseignants utilisent parfois le dialecte. Dans les écoles qui souhaitent apporter un soutien spécial aux élèves plurilingues, l'utilisation du dialecte et des langues d'origine n'est souvent pas tolérée, afin de travailler au maximum sur les compétences en haut allemand.

Le dialecte local est utilisé dans toutes les communications quotidiennes et son utilisation est indépendante de l'appartenance à un certain milieu social (Haas, 1998). De ce point de vue, on pourrait considérer que le présent travail ne montre pas une comparaison entre des enfants monolingues et plurilingues, puisque les monolingues parlent eux aussi deux langues. Malgré cette diglossie très présente en Suisse Alémanique, nous avons décidé de ne pas considérer le dialecte comme une langue à part entière, mais comme une variation de la langue standard.

Ainsi définissons-nous « enfants monolingues » dans le présent travail, les enfants ayant appris dans leur milieu familial comme langue première (L1) uniquement l'allemand ou le dialecte allemand local. Cela inclut donc les enfants immigrés en Suisse provenant d'Allemagne ou d'Autriche. Notre échantillon des monolingues ne se compose donc pas uniquement d'enfants monolingues de nationalité suisse, mais aussi d'enfants monolingues de nationalité allemande ou autrichienne qui vivent en Suisse. Les enfants plurilingues de notre échantillon ont appris dans leur milieu familial une ou plusieurs autres langues que l'allemand et éventuellement le dialecte locale (cas lorsque les parents forment un couple mixte par exemple). Cela inclut aussi bien les enfants qui ont développé un plurilinguisme simultané que les enfants avec un plurilinguisme consécutif. Au moins un de leurs parents ou de leurs grands-parents a immigré en Suisse. Les élèves plurilingues dont nous parlons ont passé la plus grande partie de leur vie scolaire en Suisse. Souvent ils maîtrisent bien le dialecte local, ce qui peut conduire les enseignants à se méprendre sur leurs véritables compétences en haut allemand. Nous avons exclu de notre échantillon des élèves plurilingues immigrés récemment.

Afin de donner une impression de l'ampleur des phénomènes de plurilinguisme et de migration en Suisse, nous allons terminer ces propos liminaires avec des informations sur des flux migratoires en direction de la Suisse. En 2009, l'office fédéral de migration suisse a recensé sur le territoire suisse 21,7% (1,68 millions) d'étrangers (Office fédéral des migrations, 2010). Et ce sont même, en 2009, un quart des personnes qui travaillent en Suisse qui possède un passeport étranger. La Suisse a donc au sein de l'Europe l'un

des taux d'étrangers le plus élevé. Ainsi en Suisse, l'immigration contribue plus à la croissance démographique que dans les pays d'immigration traditionnels comme les Etats-Unis, le Canada et l'Australie. Cela peut surprendre lorsqu'on pense au fait que la Suisse n'a, contrairement à la France, pas d'afflux migratoires venant d'anciennes colonies et n'a pas la réputation d'un pays qui ouvre grand ses portes aux immigrés. En fait, la plupart de ces 21,7% d'étrangers sont soit des réfugiés politiques<sup>5</sup> soit des immigrés spécifiquement choisis pour leurs qualifications professionnelles en adéquation avec les manques existants sur le marché de l'emploi en Suisse.

Par ailleurs, une politique de naturalisation très restrictive, fait que les immigrés et leurs familles doivent attendre longtemps avant de pouvoir demander la nationalité suisse qui n'est accordée qu'une fois que le demandeur a réussi un examen et a payé des frais de naturalisation. Dans certaines communes, c'est encore une assemblée de citoyens qui décide si oui ou non la nationalité suisse est accordée.

C'est à la fin du 19<sup>ème</sup> siècle que la Suisse est devenue un pays d'immigration. En 1915, le taux de personnes étrangères en Suisse avait déjà atteint 15% de la population totale, ce qui éveilla de plus en plus de ressentiments parmi la population autochtone. Une restriction massive de l'immigration fit par la suite décroître le taux des étrangers à 5% pendant la 2<sup>ème</sup> guerre mondiale. Le développement économique rapide après la guerre s'accompagna d'une forte demande en mains-d'œuvre étrangères dans les domaines de l'agriculture, du bâtiment et de l'industrie. Cette main-d'œuvre est essentiellement d'origine italienne. Malgré des restrictions et un manque de mesures d'intégration, le taux d'étrangers (provenant essentiellement du Portugal, de Yougoslavie et de la Turquie) augmenta et atteint un million en 1970. En 1994, le taux d'étrangers dépasse pour la première fois les 20%. Cette évolution éveille des craintes -auxquelles l'Etat répond en essayant de plafonner les entrées des immigrés- et profite aux partis d'extrême droite. Mais l'économie suisse a besoin de main-d'œuvre et d'experts étrangers.

En 2000, la Suisse signe les accords « Schengen » et autorise désormais la libre circulation et la recherche de travail sur le territoire suisse aux ressortissants de l'Union Européenne.

---

<sup>5</sup> En 2009, au total 17326 demandes d'asile ont été déposées en Suisse, dont 16,3% ont reçu une réponse favorable.

Cette évolution se reflète dans les chiffres actuels concernant l'origine des étrangers en Suisse. En 2009, les immigrés d'origine italienne représentent encore la majorité des étrangers avec 17,2 %. Ils sont suivis des immigrés d'origine allemande (14,9 %), des immigrés d'origine portugaise (12,2 %) et des immigrés venant de Serbie (8,9 %). Au total 63,6 % des étrangers sont issus des l'UE-27/AELE<sup>6</sup>.

Rapportés au taux dans la population totale, les enfants immigrés représentent environ un cinquième des élèves scolarisés en Suisse. Ils restent cependant, comme nous allons le voir, sous-représentés dans les filières d'enseignement sélectif et surreprésentés dans les filières scolaires aux exigences moindres. Une question importante à clarifier est de savoir si cela est à mettre en lien avec leur plurilinguisme –comme cela est souvent supposé.

---

<sup>6</sup> La AELE (l'Association européenne de libre-échange) inclut outre la Suisse, l'Islande, la Norvège et le Liechtenstein.

## Introduction

Le bilinguisme et le plurilinguisme ne relèvent pas d'un phénomène rare puisqu'ils correspondent à la situation de plusieurs centaines de millions de personnes à travers le monde. Les conséquences d'une globalisation croissante accompagnée de flux migratoires en forte augmentation obligent les systèmes éducatifs des pays occidentaux à s'intéresser à la question du plurilinguisme à l'école. La Suisse avec quatre langues nationales semble être le parfait modèle d'un pays plurilingue. Pourtant, le système d'éducation suisse n'intègre pas sans problèmes les milliers d'enfants plurilingues issus de familles immigrées. En Suisse, comme en France ou en Allemagne, l'acquisition de la langue de scolarisation se fait dans le cadre d'un modèle d'assimilation : les langues d'origine des enfants sont ignorées en faveur d'une volonté d'assimilation la plus rapide possible de la langue de scolarisation. Cet « habitus monolingue » des écoles comme il est décrit par Gogolin (1994) est critiqué, car il contribue, entre autres, à un « gâchis » de capital humain.

Ce n'est que depuis peu que le coût de ce « gâchis » est évalué. Andreas Schleicher, responsable coordinateur des études Program of International Student Assessment (dorénavant PISA) commissionnées par l'OCDE, constate que les élèves du 21<sup>ème</sup> siècle apprennent des enseignants et des parents du 20<sup>ème</sup> siècle dans un système éducatif qui date du 19<sup>ème</sup> siècle (Schleicher, 2010). Il présente une étude<sup>7</sup> selon laquelle la Suisse obtiendrait un gain économique qui équivaut à trois fois son produit intérieur brut actuel, si on arrivait à amener les élèves les plus faibles au moins au niveau 2<sup>8</sup> des échelles de compétences scolaires de PISA (Hanushek & Woessmann, 2010). De tels calculs peuvent certes inspirer le scepticisme, mais ils nous indiquent toutefois que l'efficacité de l'enseignement scolaire est de plus en plus comprise en termes économiques. Le fait que selon les études PISA 7 % des élèves suisses à la fin de la scolarité obligatoire, éprouvent de sérieux problèmes dans la compréhension du contenu d'un texte n'est pas acceptable (Moser, 2001)<sup>9</sup>.

---

<sup>7</sup> Ce gain économique sera réalisé au cours des 70 ans de vie d'une future génération qui aura au moins le niveau 2 en compétences de lecture.

<sup>8</sup> PISA a développé une échelle composé de 5 niveaux de compétences, le 5<sup>ème</sup> étant le plus élevé.

<sup>9</sup> Selon le rapport de PISA 2000, 7 % des élèves suisses, 4,2 % des élèves français, 5,4 % des élèves italiens et 9,9% des élèves allemands ne sont pas capables d'extraire les informations clés d'un

Toujours selon PISA, la Suisse serait, avec l'Allemagne, le pays où les compétences en lecture dépendent le plus de l'origine sociale des élèves. Ce sont donc particulièrement les élèves issus de couches défavorisées et/ou issus de familles immigrées qui sont le moins bien préparés pour s'intégrer dans la vie professionnelle. Les résultats d'études menées en Suisse confirment cette tendance. Ainsi seulement 10 % des élèves plurilingues accèdent aux filières générales du baccalauréat contre 21 % des élèves monolingues. A l'inverse, les enfants plurilingues sont de manière relative, quatre fois plus représentés que leurs homologues suisses dans les filières spécialisées pour élèves en difficulté (Kronig et al., 2000; Müller, 1996, 2001). L'échec scolaire des élèves immigrés plurilingues est hautement corrélé avec leurs compétences insuffisantes en langue de scolarisation lorsque l'on contrôle l'influence des compétences dans les autres disciplines qui sont importantes pour la sélection<sup>10</sup>: les mathématiques et le français (Kronig, 2003; Kronig et al., 2000; Müller, 1997, 2000).

Ce n'est pourtant pas un manque d'engagement de la part des enseignants qui est responsable de ce constat, mais plutôt le manque de ressources (personnelles et matérielles) et de connaissances sur les besoins et les possibilités de soutien dans le développement des compétences scolaires.

En effet, il existe un manque de données et de comparaisons systématiques concernant le développement en langue scolaire des élèves plurilingues par rapport aux élèves monolingues, de l'école primaire jusqu'à l'enseignement secondaire I en Suisse.

Un des objectifs de ce travail de thèse consiste à pallier ce manque en examinant de manière quasi-longitudinale le développement des compétences en compréhension de texte et en compréhension orale en langue de scolarisation chez les élèves plurilingues par rapport aux élèves monolingues. Afin d'apporter une réponse empirique sûre, nous utilisons des analyses à niveaux multiples, permettant non seulement de connaître le niveau des compétences à un moment précis de l'année scolaire, mais aussi de modéliser le gain de compétence réalisé au cours d'une année.

---

texte qu'ils devaient lire. PISA examine dans un intervalle de 3 ans les compétences scolaires en lecture, en mathématiques et en sciences naturelles dans un échantillon d'élèves de 15 ans.

<sup>10</sup> L'accès à une des 3 filières du secondaire I dans le système éducatif du Canton de Zurich est conditionné par les résultats en allemand, en mathématiques et en langues étrangères des élèves à la fin du primaire c'est-à-dire lorsque les élèves ont environ 12 ans. Cette sélection précoce est soupçonnée contribuer aux difficultés des élèves immigrés plurilingues à accéder aux filières scolaires à exigences étendues, car il paraît difficile d'avoir compensé déjà à cet âge un éventuel retard en langue de scolarisation.

L'objectif principal de la présente étude consiste à examiner la validité du modèle de la « *Simple View of Reading Theory* » (désormais *SVRT*) très populaire dans la recherche anglophone relative à la lecture. Malgré cette popularité et des preuves empiriques fiables dont témoignent de multiples études anglophones, ce modèle a reçu à ce jour peu d'attention dans la recherche germanophone ou française. Ainsi la validité de plusieurs points clés de la *SVRT* sera examinée dans un échantillon de lecteurs monolingues lisant en langue allemande (L1) et auprès de lecteurs plurilingues lisant dans leur L2 (l'allemand).

La *SVRT* étant explicitement un modèle simple des processus complexes de la lecture, certains auteurs ont tenté d'y ajouter un troisième facteur, afin d'augmenter de façon significative la force explicative de ce modèle. Dans le cadre du présent travail notre troisième objectif consistera à examiner l'impact des capacités intellectuelles en tant que facteur susceptible d'enrichir le modèle de la *SVRT*.

Afin de présenter le travail réalisé dans le cadre de cette thèse de la façon la plus claire possible, nous avons opté pour un découpage simple en trois parties:

- La partie théorique qui exposera en trois chapitres les modèles de compréhension de texte et notamment le modèle de la « *Simple View of Reading Theory* », le cours du développement des compétences en lecture et les résultats de recherche récents sur le lien entre les compétences en lecture et les capacités intellectuelles opérationnalisées par le QI.
- La deuxième partie décrira la méthodologie adoptée au cours de cette étude empirique.
- Et la troisième partie consistera en une présentation et une discussion des résultats de recherche en les mettant en relation avec les résultats les plus récents d'autres travaux de recherche internationaux, en particulier anglo-saxons.

# 1 Partie théorique

## 1.1 Les modèles de la lecture

Les exigences des sociétés modernes en général et du monde du travail en particulier ont conféré à la lecture un statut de compétence de base, indispensable pour pouvoir partager la vie sociétale et culturelle. Ne pas pouvoir lire équivaut à un handicap sérieux dans la vie professionnelle et privée. La lecture est devenue une clé indispensable pour pouvoir apprendre, communiquer et partager, pour pouvoir se construire, s'épanouir et se divertir. Par voie de conséquence, l'apprentissage de la lecture est l'un des objectifs majeurs de l'enseignement scolaire. Les méthodes qui devront être utilisées pour que chaque enfant sans exception arrive au stade où il peut lire de façon rapide et efficace un texte long et y extraire les informations nécessaires, sont régulièrement au centre d'intenses discussions. Ainsi se sont affrontés plusieurs fois au cours des décennies passées les défenseurs de la méthode dite « globale » et les défenseurs d'une approche dite « syllabique ». En 2006, par exemple une circulaire du Ministre de l'Education Nationale Français -Gilles de Robien – proposant la méthode « syllabique » comme étant la seule adaptée à l'enseignement de la lecture pour les jeunes écoliers, a provoqué de vives protestations de la part du syndicat des enseignants du primaire (NOUVELOBS.COM, 05.01.2006). En Grande Bretagne la publication du rapport de l'inspecteur nationale Sir Jim Rose, conseillant une orientation « syllabique » aux futurs programmes d'enseignement, sans nier les mérites d'une approche « globale » (Rose, 2006), a également ranimé un vieux débat parfois mené de façon très émotionnelle. Ces deux exemples montrent l'actualité inaltérée de la question « Comment apprendre au mieux à lire ? ». Bien que la période des débats houleux et envenimés, résumée aux Etats-Unis sous le terme de « *reading wars* », semble révolue, le nombre de recherches et de publications traitant des questions de l'apprentissage de la lecture n'a cessé d'augmenter.

Dans les pages qui suivent, notre objectif est de retracer l'évolution des modèles et théories de lecture qui ont mené au développement des différentes approches d'enseignement de la lecture<sup>11</sup>. Nous allons également présenter des modèles de lecture

---

<sup>11</sup> Pour cela nous nous appuyons en particulier sur les articles de M.J. Fresch (2008), de P.A. Alexander et E. Fox (2004) ainsi que de K. Holle (2009)

qui essaient délibérément d'intégrer les deux approches comme la *Simple View of Reading Theory*, qui sera la base théorique du présent travail de recherche.

## 1.2 Revue historique des paradigmes prédominants

Depuis le début du 20<sup>ème</sup> siècle nous pouvons distinguer une suite de plusieurs courants de recherche en matière de lecture.

Le passage d'un courant à l'autre et d'un paradigme prédominant à l'autre n'est bien entendu pas aussi net que la figure 1 ci-dessus le suggère. La plupart de ces courants persistent jusqu'à nos jours, mais on peut attribuer à chacun d'eux une période pendant laquelle son paradigme a été prédominant dans la recherche psychologique en général et la recherche en lecture en particulier.

### 1.2.1 Les débuts de la recherche sur le processus de lecture

Au début du 20<sup>ème</sup> siècle, les premiers chercheurs s'intéressent au processus de lecture, qu'ils analysent et définissent comme un pur processus de perception. Javal (1879) mais aussi Judd et Buswell (1922) par exemple étudient les mouvements oculaires lors de la lecture. Judd et Buswell trouvent des relations dynamiques entre ces mouvements oculaires, la difficulté du texte et l'intention du lecteur. Ils en concluent que les mouvements oculaires lors de la lecture sont l'expression des processus cognitifs servant à la compréhension du langage. Cette compréhension du langage faisant pour eux partie des processus cognitifs plus généraux de la pensée, la lecture se limitait au niveau de la perception. L'analyse méticuleuse du système visuel grâce aux mesures des temps de fixation, des saccades de progression et de régression lors de la lecture, a continué être un champ de recherche fructueux. On a par exemple pu montrer que les « lecteurs-apprentis » sont beaucoup plus fortement liés à la surface graphique du texte que les bons « lecteurs experts » (F. Smith, 1994). Le nombre de régressions oculaires (lorsque les yeux retournent à un passage antérieur de la phrase) est plus élevé chez les lecteurs apprentis, ce qui peut être interprété comme expression de la recherche active de cohérence au niveau du texte (Rayner, 1998). Cela montre clairement que la lecture est plus qu'un simple processus de perception. Déjà Thorndike (1917) rejette l'idée de la lecture comme simple phénomène de perception en déclarant que lire est penser (« *reading as reasoning* »). Petit à petit, la conception de la lecture fortement liée à la perception est

enrichie de facteurs linguistiques et cognitifs qui dans leur ensemble rendent mieux compte du caractère multifactoriel du processus de lecture.

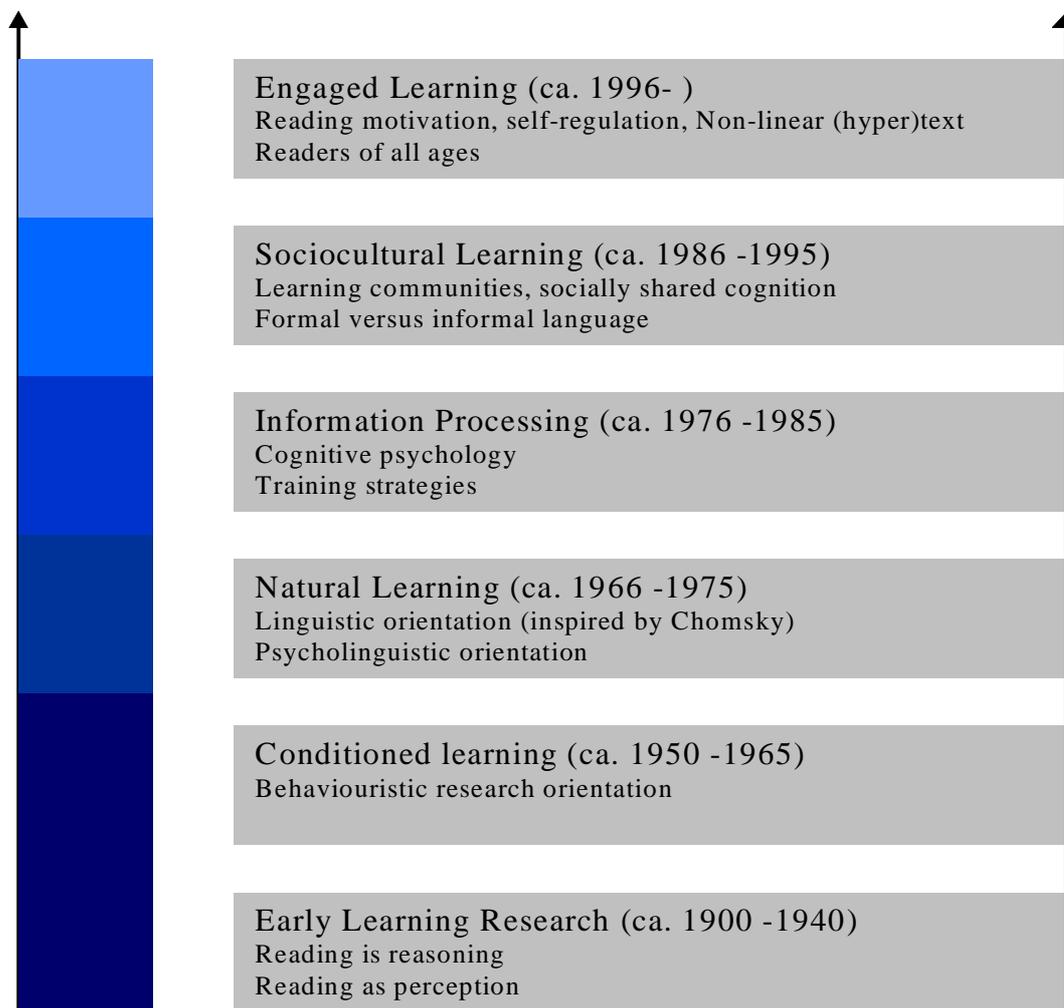


Figure 1 Courants théoriques au cours du 20<sup>ème</sup> siècle

### 1.2.2 La lecture comme un processus d'apprentissage conditionné

Après la deuxième guerre mondiale, on note un intérêt grandissant pour la question de la remédiation des difficultés de lecture. Le « behaviorisme » en tant que paradigme dominant de l'époque promettait de fournir des réponses scientifiques à cette question. La compréhension de texte était conçue comme une forme de comportement conditionné. Elle était dans cette vision traditionnelle le produit qui résulte du fonctionnement automatique de la somme de divers processus de base (« *basic skills* »). On est donc face à l'idée d'une addition de composantes cognitives et non-cognitives. Le behaviorisme se focalise dans sa conceptualisation des processus de lecture uniquement sur le comportement extérieur des personnes, les processus mentaux au sein de la « boîte noire » sont laissés de côté. La lecture de phrases longues et de textes reçoit très peu

d'attention, le mot étant l'unité principale de recherches menées sous le paradigme du « behaviorisme ». Le « behaviorisme » admet que de nombreux processus cognitifs font partie du processus de lecture, mais une séparation stricte est maintenue entre la phase de perception visuelle et la phase suivante, à savoir le traitement cognitif qui mène à la compréhension du langage. Cette vision traditionnelle de la lecture a eu des implications qui sont encore perceptibles dans les idées de la psychologie naïve sur la lecture. Au niveau de la didactique de l'enseignement de la lecture, cette approche traditionnelle additive se traduit par une concentration lors de l'enseignement sur les capacités discriminatoires du système visuel, et sur l'articulation correcte des lettres et graphèmes. De nombreux programmes d'entraînement à la lecture sont conçus en se focalisant sur les capacités de base et leurs sous-parties supposées déficitaires, en faisant fi des processus de compréhension plus complexes.

Les critiques du courant behavioriste faisaient remarquer à quel point l'exclusion des processus mentaux à l'intérieur de la « boîte noire », rendait la recherche sur la lecture (comme sur d'autres phénomènes) stérile. L'idée de base des « gestaltistes » que le tout est toujours plus que la somme des parties, gagnait également de plus en plus de terrain. La « substrata-factor theory » de Holmes (1954) fut l'un des premiers exemples d'une nouvelle orientation dans la recherche sur la lecture. De nouvelles découvertes en neurologie et dans le champ de l'intelligence artificielle accéléraient au milieu des années soixante la fin du « behaviorisme » en tant que paradigme de recherche prédominant. Le point focal de la recherche repassait de l'environnement à l'Homme. Avec le début du « cognitivisme » et une augmentation des recherches interdisciplinaires, commençait une période très fructueuse pour la recherche sur la lecture.

### **1.2.3 L'apprentissage naturel : la lecture comme capacité inhérente**

Deux courants principaux dominaient par la suite la recherche des processus de lecture : d'un côté les linguistes, dans la tradition chomskyenne, de l'autre, les psycholinguistes. Leur conviction commune était que l'Homme est génétiquement prédisposé à acquérir le langage et cela sans besoin d'entraînement dédié. Chomsky était très critique à l'égard du behaviorisme. Il développa la théorie de « grammaire générative », qui supposait des structures langagières universelles et innées qui vont s'exprimer et devenir visibles tout naturellement au cours du développement infantile. Il lui était important que l'on distingue, dans la recherche des phénomènes langagiers, la performance de la compétence d'un individu. La dispute entre les partisans d'une analyse langagière holis-

tique ancrée dans les sciences humaines et les partisans behavioristes d'une recherche strictement objective et apparentée aux sciences naturelles s'atténue au fil du temps. Il en résulte la naissance de la psycholinguistique à orientation cognitiviste qui s'intéresse à la fois aux structures langagières et cognitives mais aussi à la façon dont les phénomènes langagiers se construisent. C'est également l'époque d'importantes publications dans le domaine de la recherche sociolinguistique. Labov et Shuy par exemple, s'intéressent à la fin des années soixante à la relation entre les rôles sociaux et les diverses variations du langage. Ils étaient ainsi les premiers à thématiser la différence entre les langages quotidiens d'enfants issus de différents milieux sociaux et la langue scolaire.

Une théorie très importante issue des travaux psycholinguistes est celle de la devinette psycholinguiste (« *psycholinguistic guessing game* ») de Goodman. Il suppose que les résultats de la lecture ne sont que des significations probables : « *Readers predict, select, confirm and self-correct as they seek to make sense of print. [...] they guess or make hypothesis about what will occur in the text. Then they monitor their own reading to see whether they guessed right or need to correct themselves to keep making sense* » (p.40 Goodman, 2005). Basé sur l'analyse d'erreurs, il distingue les « oral reading miscues » - erreurs de lecture qui se situent au niveau de la compréhension globale de la phrase - des « oral reading errors » - véritables erreurs de lecture. Ainsi s'oppose-t-il à l'idée d'une correspondance directe entre la lecture correcte des mots et la compréhension globale d'un texte : « *accuracy, correctly naming or identifying each word or word part in a graphic sequence, is not necessary for effective reading since the reader can get the meaning without accurate word identification.* » (p. 826 Goodman, 1974). La conception du processus de lecture chez Goodman est donc fortement concentrée sur la construction de sens : « *reading is making sense of print* ». Toutefois, il semble important de distinguer les lecteurs débutants des lecteurs experts, car l'influence du contexte décroît au fur et à mesure que le niveau de maîtrise de la lecture augmente (Perfetti & Roth, 1978; West & Stanovich, 1978). Deviner pour lire ne serait donc pas la caractéristique d'une lecture bien maîtrisée, mais une stratégie de résolution face aux problèmes rencontrés par des lecteurs débutants lors de l'identification des mots.

Frank Smith est un autre représentant important de la recherche psycholinguistique de la lecture. Pour lui, la lecture n'a rien de spécial, elle n'est qu'une autre manière d'acquérir de l'information - à côté de la compréhension orale - et d'en extraire le sens. La seule

différence entre la lecture et la compréhension orale réside pour Smith dans la façon dont l'information est codée (Frank Smith, 1973, 1986). Ainsi l'apprentissage de la lecture ne repose pas sur des capacités prérequis spécifiques. L'enfant est selon lui programmé à apprendre à lire, car l'apprentissage de la lecture est la conséquence logique de son appartenance à une société tournée vers l'écrit. Cette vision n'est pas acceptée par tous comme l'illustre parfaitement l'article « Learning to read : an unnatural act » de Gough et Hillinger (1980).

D'un point de vue didactique, Smith redéfinit le rôle de l'enseignant. Son rôle ne sera pas d'apprendre aux élèves la lecture, mais d'aider les élèves à apprendre par eux-mêmes la lecture. En adéquation avec Goodman, Smith met l'accent sur les connaissances disponibles en amont chez le lecteur. Elles aident à minimiser la dépendance de la base textuelle et permet de faire des « informed predictions » lors de la construction de la compréhension du texte.

Les conséquences didactiques de ces modèles psycholinguistiques de lecture sous la forme de l'enseignement « globale » de la lecture, ont eu une grande influence. Des activités centrées sur l'expérience de la lecture, sur le contenu de textes souvent authentiques, s'accompagne d'une nouvelle appréciation des erreurs lors de la lecture et l'écriture. Au lieu d'être appréhendées en tant que déficit, les erreurs étaient de plus en plus comprises comme une fenêtre privilégiée sur les processus conscients et inconscients de l'apprentissage, qui pouvaient être utilisés pour mieux cibler l'enseignement sur les besoins de chaque enfant.

Le parti-pris d'une large proportion des enseignants en faveur de la méthode globale peut être expliqué (du moins aux Etats-Unis) par le fait qu'elle correspond mieux aux intentions des enseignants et pédagogues en mettant l'accent sur la motivation et la découverte du plaisir de la lecture (Stahl, 1999).

Une approche perçue comme contraire à la méthode « globale » est l'approche « syllabique » ou « synthétique ». Cette dernière met l'accent sur l'enseignement direct des correspondances grapho-phonémiques. Une fois maîtrisé, l'enfant apprend à composer ces éléments de base en syllabes et en mots. Entre les défenseurs de l'approche « globale » et ceux de l'approche « syllabique » des débats très passionnés ont eu lieu, comme nous l'avons déjà mentionné. Pourtant aucune des multiples études menées sur

la question d'un avantage éventuel d'une méthode sur l'autre, n'est parvenue à une conclusion univoque.<sup>12</sup> Partant de ce constat, de nombreux spécialistes recommandent une mixité des approches. En Allemagne et en Suisse, cette intégration des méthodes est une réalité dans les classes de lecture depuis fort longtemps. Il semble en être de même en France où selon un sondage IFOP en 2006, 47% des enseignants aux CP déclarent utiliser des méthodes mixtes alors que 18% disent utiliser la méthode globale et 35 % la méthode synthétique (NOUVELOBS.COM, 2006). La directive ministérielle de G. de Robien en faveur de la méthode syllabique semble néanmoins avoir eu un impact, notamment chez les plus jeunes enseignants. En effet, parmi ceux ayant moins de 5 ans d'ancienneté, 46 % déclarent mettre l'accent sur la maîtrise du code lors de l'enseignement de la lecture.

#### **1.2.4 La lecture un traitement d'information**

Le cognitivisme et le développement rapide d'outils informatiques ont ouvert la voie à la naissance de nouvelles méthodes et théories concernant les processus mentaux. C'est sur ce fond que se sont développés les modèles cognitivistes de la lecture. Les modèles psycholinguistiques étaient toujours d'actualité mais cette approche perdait sa place prédominante au sein de la recherche sur la lecture. L'approche cognitiviste s'intéressait davantage aux processus de construction et de reproduction de connaissances au cours de la lecture de texte qu'aux dispositions innées qui mènent à l'apprentissage de la lecture. Au lieu d'analyser la lecture dans le même cadre conceptuel que la compréhension orale, l'écriture et la parole, l'approche cognitiviste faisait à nouveau ressortir la spécificité des processus de lecture.

L'influence des caractéristiques du texte éveillait l'intérêt des chercheurs cognitivistes et se manifestait dans de nombreuses publications concernant l'impact du genre, de la cohésion et de la structure du texte sur la compréhension de texte. Un autre courant de recherche analysait l'impact des connaissances préalables du lecteur sur la direction de

---

<sup>12</sup> En Allemagne trois études empiriques (Ferdinand 1972, Müller, 1964 et Schmalohr 1961 cités dans (Kirschhock, 2003p. 84) ont montré qu'à la fin de la deuxième année scolaire, toute différence significative dans les performances de lecture a disparu entre des groupes d'élèves ayant reçu un enseignement "syllabique" et ceux ayant reçu un enseignement global. Seuls les élèves les plus faibles semblaient tirer plus de profit d'un enseignement « syllabique », mais dans la quatrième année scolaire les compétences en lecture était équivalentes entre les groupes (Müller,1964). Diverses études nord-américaines comme celle de Bond et Dykstra en 1967 (cité dans Kirschhock, 2003p. 84) ont produit des résultats semblables.

l'attention lors de la lecture, sur l'interprétation et sur la mémorisation du contenu du texte.

L'influence de la métaphore de l'ordinateur comme modèle de pensée se faisait également remarquer comme en témoignent deux modèles concernant la lecture à voix haute : le «Dual Route Cascade Model» ou les modèles parallèles distribués.

La lecture pose aux chercheurs la question fondamentale du mécanisme qui mène à l'activation de la représentation phonologique, sémantique et syntactique d'un mot à partir de sa forme orthographique. Deux mécanismes responsables de ce travail d'assemblage ont été imaginés, dont un qui se base sur un ensemble de règles de conversion entre graphèmes et phonèmes (désormais CGP) que le lecteur débutant devra apprendre. La « théorie des deux voies »(Coltheart, 1978) combine les forces des deux mécanismes : elle postule deux voies d'accès possibles au lexique mental, qui est l'unité de stockage de toutes les caractéristiques sémantiques, orthographiques et phonologiques des mots. La voie directe (lecture visuelle) permet une identification rapide à partir des caractéristiques visuelles. Au sein du lexique mental, ces dernières sont complétés par les caractéristiques phonologiques correspondantes et le mot peut ainsi être lu (à voix haute). La deuxième voie (lecture phonologique) est appelée indirecte car le mot est d'abord segmenté en graphèmes et traduit en une séquence de phonèmes en s'appuyant sur les règles CGP. Ainsi le mot peut être lu à voix haute sans que pour autant le lecteur ait accès à sa signification. Aucun mécanisme à lui seul n'est satisfaisant. Les règles CGP aboutiront pour des mots à prononciation irrégulière à une fausse prononciation, tandis que la voie directe doit déclarer forfait dans le cas de lecture de pseudo-mots, qui ne figurent pas dans le stock lexical. La théorie des deux voies combine les deux mécanismes et suppose que les faiblesses de chacun soient compensées par les forces de l'autre. L'effet de la régularité<sup>13</sup> était un argument en faveur de la théorie de Coltheart. Cette théorie a ensuite été étayée par une étude en neuropsychologie (Marshall & Newcombe, 1973) sur les erreurs de lecture de sujets cérébro-lésés qui suggérait une perte spécifique de l'un ou de l'autre mécanisme. La théorie des deux voies a constitué jusqu'en 1990 le principal cadre de recherche sur la prononciation et l'accès à la signification.

---

<sup>13</sup>Effet de régularité: les mots réguliers sont lus plus rapide que les mots irréguliers.

La théorie a aussi été l'objet de diverses critiques, entre autres sur la validité de la distinction entre mots réguliers et irréguliers. Elle était remise en question notamment par la démonstration de Glushko (1979) qui montrait que même lorsqu'un mot est régulier d'un point de vue des règles CGP, on observe que la performance de prononciation est affectée lorsqu'il existe pour un mot une prononciation alternative pour la même rime orthographique ("ille" dans ville et fille). Le modèle classique de Coltheart ne pouvait pas expliquer ces résultats. De multiples variantes de la théorie des deux voies ont été développées comme par exemple l'introduction d'une mémoire phonologique temporaire qui combinerait les informations générées par les deux mécanismes (Coltheart et al., 1993) attribuant l'effet de consistance à la contribution de la voie orthographique directe. Face à la montée connexionniste proposant des « modèles parallèles distribués » qui présentent l'avantage d'une simulation des données empiriques sous forme de programme informatique (Seidenberg & McClelland, 1989). Coltheart et Rastles (1994) ont développé une implémentation sur ordinateur du modèle à deux voies de la théorie des deux voies : le « *Dual Route Cascade Model* » (« double voie en cascade » désormais DRC). Il conçoit le lexique mental comme un ensemble de sous-systèmes (systèmes orthographique, sémantique et phonologique) qui se suivent en cascade et qui incluent des relations bidirectionnelles entre les sous-systèmes.

Les « modèles parallèles distribués » issus du courant connexionniste peuvent être considérés comme une proposition alternative à la « théorie des deux voies ». Ils sont d'emblée conçus comme des modèles de simulation sur ordinateur. Leur thèse principale est que le processus de lecture peut être compris sans avoir recours à deux sources de connaissance distinctes et sans avoir recours à un système de règles CGP appris en amont. Dans les modèles parallèles distribués, c'est un ensemble d'associations entre des patrons orthographiques et des patrons phonologiques qui lie la forme orthographique à la forme phonologique. L'apprentissage de certaines relations stimulus/réaction est ainsi considéré comme suffisant pour produire le code phonologique correspondant à un certain code orthographique.

Le « triangle framework model » de Seidenberg et McClelland (1989) est le plus connu des « modèles parallèles distribués ». Il rejette l'idée d'une représentation spécifique locale pour chaque mot en faveur d'une représentation distribuée sous la forme d'un ensemble d'activations neuronales, représentant le codage d'une certaine suite de lettres. Ayant entré un corpus constitué de 3000 mots anglais monosyllabiques dans un système

informatique, Seidenberg et McClelland ont entraîné le réseau à convertir les formes orthographiques en formes phonologiques correspondantes. Cet entraînement montrait des résultats assez proches des performances observées chez un lecteur habile (par exemple, les mots fréquents sont mieux « lus » que les mots moins fréquents.). Une extension de ce modèle (Plaut et al., 1996) aboutissait à des résultats encore plus proches des données empiriques, en modifiant la forme des représentations orthographiques et phonologiques utilisées par le système.

Les postulats des « modèles parallèles distribués » sont purement hypothétiques et ne reposent pas sur des données empiriques obtenues lors de l'enregistrement de l'activité neuronale d'un lecteur. Mais contrairement au modèle DRC de Coltheart et Rastles (1994), le « triangle framework » essaie d'inclure une certaine dynamique de développement. Tandis que le modèle DRC ne modélise que le statut quo d'un lecteur confirmé et ne se prononce pas sur la question de l'apprentissage des règles grapho-phonémiques, le modèle « triangle framework » essaie d'expliquer la maîtrise de ces règles par la fortification des relations neuronales les plus fréquentes (McKague et al., 2001). Malgré cela, les deux approches sont loin de la réalité d'un enfant qui apprend à lire. Elles partent du principe que les éléments phonologiques et orthographiques sont « appris » en même temps, or en réalité un enfant qui commence à apprendre le code écrit dispose déjà d'immenses connaissances dans le domaine phonologique. Cette incohérence a été critiquée à plusieurs reprises (Hulme et al., 1991; McKague et al., 2001) et a été considérée lors des modifications du « triangle framework model » faites par (Harm & Seidenberg, 1999).

Si l'architecture du système et la nature des mécanismes qui permettent d'obtenir la forme phonologique des mots restent controversées, on observe également une convergence progressive entre les théories, dont la plupart ont recours au formalisme des réseaux (Content & Zesiger, 2000). Les différences essentielles portent sur la nature des représentations des unités lexicales (locales ou distribués), sur la connectivité entre les différents niveaux de représentation, sur la distinction entre connaissances globales (associations spécifiques des mots) et analytiques (correspondances graphème-phonème) et sur les mécanismes qui assurent la correspondance (règles de conversion versus associations graduelles).

Les théories et modèles de lecture développés au sein du courant cognitiviste ne se sont pas uniquement intéressés aux processus fins de conversion grapho-phonémiques au

niveau du mot. L'un des concepts les plus importants issu de la recherche cognitive est celui du « schéma ». Le psychologue britannique Sir F. Bartlett (1932) est le premier à utiliser le terme de « schéma » pour expliquer son observation du fait que les sujets de son étude remplaçaient fréquemment des informations non familières par des éléments beaucoup plus familiers lors de la reproduction du contenu d'une narration de peuples indigènes. Le terme de schéma était réintroduit dans le cadre de la « *frame theory* » (Minsky, 1975) avant que Rumelhart (1980) ne formule une théorie psychologique explicite sur la représentation mentale des connaissances complexes sur le monde : la « théorie du schéma ». Anderson (1977, 1978) introduisait la théorie du schéma dans le domaine de l'éducation et expliquait comment les connaissances antérieures d'une personne peuvent avoir un impact sur l'acquisition de nouvelles informations. Très vite la théorie du schéma était utilisée pour expliquer le processus de la compréhension de texte.

Selon Rumelhart, les schémas peuvent être décrits en tant que « *building blocks of cognition* » (1980). Les connaissances d'un lecteur sont imaginées comme des blocs façonnés par les apprentissages antérieurs et ils sont considérés nécessaires à la compréhension d'un texte. Du point de vue de la théorie du schéma, un texte ne peut être véritablement compris que lorsque le lecteur dispose de schémas qui correspondent aux informations clés du texte. Anderson (1984) attirait l'attention sur la nécessité d'aider les enfants apprenant à lire, à faire ce lien entre ce qu'ils lisent et ce qu'ils savent. Les schémas aident à enregistrer les informations en adéquation avec les connaissances antérieures, ils orientent l'attention du lecteur et lui permettent de tirer des conclusions, y compris sur des éléments évoqués seulement implicitement. La théorie a permis d'étudier le rôle joué par les structures de connaissances antérieures dans les processus de traitement de texte. La théorie du schéma a ainsi rappelé l'importance des processus descendants (« *top-down* »)<sup>14</sup> et contrebalancé l'importance souvent mise en avant des processus ascendants (« *bottom-up* »)<sup>15</sup>. De nombreux programmes d'aide ont désor-

---

<sup>14</sup> Les processus „descendants“ désignent les processus cognitifs lors du traitement d'informations qui sont liés aux connaissances et hypothèses préalables. Dans les modèles "*top-down*" le flux des informations est ainsi descendant et le traitement des informations est prioritairement dirigé par les connaissances préalables (*concept-driven*) (Zagar, 1992).

<sup>15</sup> Les processus « ascendants » désignent ici les processus cognitifs lors du traitement de texte fondés sur la base textuelle. Les modèles "*bottom-up*" décrivent ainsi prioritairement le flux d'information ascendant, dirigé par les données (*data-driven*) avec un traitement d'information séquentielle de bas en haut (Zagar, 1992).

mais intégré ce point de vue en proposant une aide à l'utilisation de stratégies métacognitives lors du processus de lecture.

La théorie du schéma et les approches théoriques qui s'y réfèrent, ne prétendent cependant pas, que seuls les processus « top down » seraient importants pour la compréhension d'un texte. Suivant que l'on adopte une vision plutôt large ou étroite des schémas, on peut supposer que les schémas ne se réfèrent pas uniquement au domaine de l'interprétation mais qu'ils existent aussi au niveau syntactique, morphologique et grapho-phonémique. Ainsi différents schémas sont activés à différents niveaux lors du processus de lecture. Ils interagissent grâce à une combinaison de processus « bottom-up » et « top down ». Ainsi aurait lieu une interaction ininterrompue entre les processus « bottom-up » (décodage des unités linguistiques individuelles comme les graphèmes, mots et syntagmes) et « top-down » (élaboration de prédictions sur la signification du texte à partir des connaissances antérieures activées).

La théorie du schéma fait partie au sein du courant cognitiviste d'une approche centrée sur le « produit » (connaissances, représentations) résultant du processus de la lecture (Marin & Legros, 2008). On peut distinguer une deuxième approche qui est plus centrée sur les processus cognitifs au cours de la lecture (Kintsch & Van Dijk, 1978; 1983).

Le premier modèle de Kintsch et Van Dijk (1978) s'intéresse à la construction d'une base de texte composée d'une micro - et macrostructure cohérentes par rapport au contenu du texte. Les auteurs suggèrent qu'une liste structurée de propositions incluant un prédicat et un ou plusieurs arguments représente la signification d'un texte. La microstructure d'un texte correspond donc à un réseau de propositions. La macrostructure sémantique d'un texte correspond au résumé d'un texte construit par le lecteur à partir d'un traitement de texte. Ce dernier implique à la fois des processus de condensation sémantique (sélection, suppression et généralisation de propositions) et des processus inférentiels basés sur les connaissances antérieures du lecteur, afin de combler des trous sémantiques et d'augmenter la cohérence.

Le deuxième modèle de Van Dijk et Kintsch (1983) reste l'un des modèles de compréhension de texte le plus largement accepté. Il introduit la notion du « modèle de situation » et propose une modélisation de la façon dont les connaissances du lecteur jouent un rôle indispensable à la construction de la signification d'un texte. Selon ce modèle le lecteur doit construire activement sa compréhension du texte en fonction de ses connaissances, de ses compétences mais aussi de ses objectifs de lecture. Le modèle distingue

trois niveaux de représentation : « la surface de texte » composé des mots et de la syntaxe du texte ; la « base de texte » composé des propositions contenues dans le texte et représenté à un niveau locale (microstructure) et globale (macrostructure) et finalement le « modèle de situation » représentant le compte rendu que le lecteur a extrait du texte et intégré à ses connaissances.

Le passage de la base de texte au modèle de situation fait intervenir les différents types de connaissances du lecteur, basées sur ses expériences et apprentissages au sein de son milieu familial, scolaire et culturel afin de compléter, interpréter et rendre cohérentes les informations contenues dans la base de texte.

Le modèle « *constuction-integration* » de Kintsch (1988) est conçu dans le prolongement des deux modèles précédents. Il s'intéresse à un niveau plus fin aux processus de compréhension de texte c'est-à-dire comment ils sont intégrés aux connaissances du lecteur. Il tente de décrire de manière intégrative comment l'activation se distribue tout au long de la lecture, en s'intéressant particulièrement au processus de sélection des informations pertinentes et congruentes, lors de la construction d'un modèle de situation. Selon le modèle, ce processus est composé de deux étapes principales. La phase de construction aboutit à la construction d'un réseau, composé de toutes les propositions construites et des nœuds lexicaux activés. La phase d'intégration, est le moment où les connaissances individuelles du lecteur jouent un rôle décisif car elles déterminent les éléments non-pertinents et incohérents qui seront supprimés tandis que la valeur d'activation des éléments pertinents augmente.

Ce troisième modèle de Kintsch intègre des principes à la fois du cognitivisme et du connexionnisme. Il illustre à quel point les modèles de la compréhension du texte issus de l'approche cognitive, soulignent le rôle actif du lecteur et l'importance de ses connaissances et ses activités mnésiques lors de la lecture.

Les théories et modèles cognitivistes ont considérablement enrichi et modifié nos connaissances sur les processus de lecture et la compréhension de texte. Certains aspects restaient cependant peu thématés, comme le font remarquer par exemple Rosenblatt (1994) ou Sadoski et Pavio (2004). Rosenblatt défendait une vision holistique de la lecture et mettait en avant les aspects esthétiques de la lecture par rapport aux aspects rationnels. Elle distinguait une attitude de lecture non-esthétique (chez Rosenblatt appelée « *effereant lecture* ») qui se concentre avant tout sur l'information retenue en « résidu » après la lecture d'une attitude esthétique qui se concentre sur le vécu personnel à

travers la lecture. « ...*In aesthetic reading, the reader's attention is centered directly on what he is living through his relationship with that particular text* » (Rosenblatt 1978/1994, pp. 24-25). En admettant que les aspects émotionnels et l'intérêt liés au texte jouent aussi un rôle dans la compréhension, Kintsch (1980) constate que les modèles cognitivistes de la compréhension de texte n'ont, pour l'instant, rien à dire sur le vécu et l'esthétique de la lecture.

Pavio et Sadoski (2004) peuvent également être considérés comme des représentants d'une vision plus globale de la lecture. Ils font remarquer qu'un lecteur ne construit pas nécessairement uniquement des formes de représentations langagières au cours de la lecture. Dans leur « dual coding theory », ils proposent qu'au cours de la lecture se construisent en plus des représentations langagières (qu'ils appellent « logogenes ») également des représentations non-langagières (des « imagenes »). Leur théorie est intéressante car elle implique l'ensemble des modalités des sens (2004). Elle aborde aussi des formes de lecture peu thématiques tel que le Braille. Selon Sadoski et Paivio, le Braille serait le « logogène » pour le sens haptique tandis que le ressenti tactile sera l'« imagène » de cette modalité.

Au milieu des années quatre-vingt, la communauté de recherche en lecture semblait s'orienter dans une nouvelle direction. Les résultats des programmes d'apprentissage conçus selon les principes cognitivistes étaient moins bons que prévu. Nombre d'élèves amélioreraient moins que prévu leurs compétences en lecture, suite à un enseignement explicite des stratégies de lecture et de compréhension (Paris, Wasik, Turner, 1991 cité dans Alexander & Fox, 2004)

Les stratégies enseignées aux élèves étaient multiples et pouvaient être appliquées avant, pendant ou après la lecture. Elles étaient fondées sur des techniques qui aident à la planification et au contrôle de sa lecture, à la construction et au contrôle du sens extrait du texte. La métacognition et la prise de conscience de son comportement en tant que lecteur étaient considérées comme les clés indispensables d'une amélioration des compétences en lecture. Or, les auteurs ont constaté que la connaissance des stratégies de lecture n'est pas une garantie du fait qu'elles seront utilisées de manière spontanée et efficace. Comme le font remarquer, Paris et al. (1991) ces stratégies n'ont pas d'effet si elles ne sont pas fondées sur la motivation et une attitude positive. Ils recommandent ainsi un enseignement qui enracine les stratégies de lecture dans des situations sociales en classe (par ex. « *peer tutoring* », « *cooperative and scaffolded learning* »). Cette

prise de conscience de l'importance des orientations en groupe et de la connaissance partagée était au cœur de cette nouvelle approche socio-culturelle dans la recherche sur la lecture.

Ainsi l'objectif primordial de l'apprentissage n'était plus la construction d'une base de connaissance individuelle mais plutôt une compréhension mutuelle qui émerge des interactions sociales d'individus spécifiques dans une situation spécifique. L'élève qui apprend est considéré en tant que membre d'une communauté d'apprentissage immergé dans une expérience collaborative, partagée et socioculturelle. Ceci est illustré par un nombre accru de travaux de recherche à cette époque qui s'intéresse à l'amélioration des échanges sociaux au sein de la classe en vue d'une amélioration de l'enseignement. Entre autres sous l'influence de la redécouverte des travaux de Vygotski (1986), ces travaux redéfinissaient le rôle de l'enseignant et soulignaient son importance en tant que guide ou facilitateur.

Le modèle de lecture développé par Irwin peut être cité en tant qu'exemple de cette époque. Il illustre comment un modèle cognitiviste fondé sur l'idée d'un réseau propositionnel peut être mis en relation avec les résultats de recherche sur les processus lors de l'enseignement en classe (Irwin, 1986).

Dans les années quatre-vingt dix du vingtième siècle, la prolifération et la présence accrue d'hypermedias et d'hypertextes ont amené de nombreux chercheurs à s'intéresser à un nouveau format de texte : les textes non-linéaires. Cette révolution technique a à la fois produit un nombre inédit de sources d'information et de genres de textes et elle a augmenté et compliqué les demandes auxquelles le lecteur moderne doit faire face (Alexander & Fox, 2004). Ainsi la recherche sur le potentiel et les difficultés spécifiques de ces textes, sur la façon dont les lecteurs se repèrent et naviguent dans ces textes, mais aussi sur les stratégies pédagogiques adéquates par rapport à ces nouveaux médias, reste aujourd'hui un champ de recherche très actif. Les travaux sur les compétences de jeunes lecteurs à évaluer la crédibilité d'un texte, à détecter des incohérences ou à identifier les sources sur lesquels un texte électronique se base sont un exemple des tendances actuelles dans la recherche, sur la compréhension de texte (Britt & Rouet, 2009). Les conditions motivationnelles et les processus d'autorégulation, dont font preuve les bons lecteurs relèvent d'un autre domaine de recherche qui émerge au milieu des années quatre-vingt-dix. En même temps que l'augmentation de l'espérance de vie dans les sociétés modernes, la disparition des travaux purement manuels et la nécessité

d'une bonne maîtrise des techniques de lecture et écriture à tous les âges, ont amené la recherche à ne plus se focaliser uniquement sur les processus d'apprentissage de lecture chez les enfants. Il est devenu impossible d'ignorer que la lecture est un domaine qui concerne des lecteurs de niveaux et d'âges très hétérogènes : « *It has become increasingly difficult to ignore that reading is a domain that relates not only to the strong or struggling readers, but also to readers of all abilities and ages. [...] Thus, earlier dichotomization of reading into "learning to read" and "reading to learn" stages is shifting to a more integrated and developmental perspective.* » (Alexander & Fox, 2004 : p 51).

### 1.2.5 Modèles développementaux de la lecture

Plusieurs chercheurs se sont efforcés de non seulement décrire les processus sous-jacents de la lecture mais de proposer une modélisation du développement de ces processus chez l'enfant qui apprend à lire et à écrire. Ces modèles développementaux de la lecture sont souvent conçus comme une suite de plusieurs stades. Ils postulent que le développement des compétences en lecture a lieu selon un certain ordre. Les changements observés au cours de ce développement sont plutôt qualitatifs que quantitatifs. Souvent ces modèles supposent des liens de complémentarité entre le développement de la lecture et de l'écriture. Un modèle très connu est celui de Frith (1985). Il distingue trois<sup>16</sup> stades qui se succèdent de façon séquentielle. Ces stades se caractérisent par la domination d'une stratégie d'identification de mots à un moment donné dans le développement de la lecture et de l'écriture.

Au stade logographique les mots sont identifiés comme des images. Il s'agit donc d'un stade de pré-lecture où l'enfant reconnaît des mots grâce à des indices visuels et à l'utilisation du contexte. Il dispose d'un vocabulaire visuel, qui lui permet de reconnaître des mots familiers, mais il est incapable de déchiffrer les composantes de ces mots et de les utiliser pour décoder des mots inconnus. La méthode globale s'appuie sur le processus d'association entre l'image / la forme du mot et sa signification qui est typique de ce stade logographique.

Au stade alphabétique, les enfants prennent connaissance du principe alphabétique de la langue écrite. Ils identifient les graphèmes comme les éléments constitutifs des mots et

---

<sup>16</sup> Il existe également une version du modèle de Frith avec six au lieu de trois stades. Nous nous référons ici uniquement au modèle classique composé de trois stades.

commencent ainsi à pouvoir décoder les mots grâce aux règles de correspondances grapho-phonémiques (CGP) qu'ils sont en train d'acquérir. Les facteurs phonologiques jouent à ce stade un rôle très important. Le jeune lecteur devient désormais plus autonome dans sa lecture, car il peut de mieux en mieux décoder des mots inconnus. On peut observer à ce stade des erreurs typiques, qui sont dues au fait que les enfants se fient encore exclusivement aux règles CGP. Suivant la méthode d'enseignement cette tendance est parfois même renforcée lorsqu'on encourage les enfants à bien écouter les mots et à les écrire comme ils les entendent, sans encore leur expliquer que certains graphèmes ne sont pas entendus à l'oral comme c'est le cas pour les terminaisons en « en » en allemand. De nombreuses études ont pu confirmer l'existence de cette phase alphabétique ((Doctor & Coltheart, 1980); Coltheart & Leahy, 1992 ; Richter ; 1992 cité dans Kirschhock 2003)).

Au stade suivant, le jeune lecteur prend connaissance des règles orthographiques. Ainsi il découvre le morphème en tant qu'unité de base du traitement des mots. Il apprend à se détacher dans l'écriture des mots de ce qu'il entend, pour y intégrer par exemple des « h » aspirés et d'autres graphèmes qu'on ne peut distinguer à l'oral. La connaissance des règles orthographiques et les connaissances sur la structure morphologique des mots deviennent alors la référence de base. Une sur-généralisation de ces règles produit des erreurs typiques qui sont des indicateurs fiables de ce stade. L'enfant est désormais capable de lire des mots irréguliers, grâce à sa pratique de la lecture, l'accès au sens est accéléré et il lit de manière plus fluide et plus rapide.

Le modèle de Frith a exercé une grande influence sur la façon dont le développement des compétences en lecture et écriture a été conceptualisé. La nécessité de maîtriser d'abord le code alphabétique et les règles CGP avant de passer au traitement orthographique fait notamment aujourd'hui l'unanimité. En revanche, l'importance du stade logographique a été relativisée grâce à des études dans des langues autres que l'anglais, ce qui laisse penser que cette phase a un caractère facultatif dans les langues ayant une plus grande cohérence grapho-phonémique telle que l'allemand (Wimmer, 1993; Wimmer et al., 1990). La nécessité d'un enseignement explicite du principe alphabétique a suscité des débats. Pour certains comme Beech (1987) l'enfant est amené automatiquement à analyser plus spécifiquement les unités composantes des mots lorsqu'il est confronté à un nombre grandissant de mots visuellement proches. Cela va dans le sens de la méthode globale, qui dans sa forme stricte ne prévoit pas d'enseignement des cor-

respondances CGP. Pour d'autres, comme les défenseurs d'un enseignement de la lecture selon une stratégie phonémique, les enfants ne découvrent pas spontanément le principe alphabétique. Sans enseignement explicite, certains enfants ne trouvent donc pas l'accès au stade alphabétique (Byrne & Fielding-Barnsley, 1989). Le caractère strictement séquentiel des stades est également remis en question. L'idée d'une coexistence des stratégies d'identification des mots est notamment défendue par Seymour (1994) dans son modèle à « double fondation » qui suppose que les processus logographiques et alphabétiques contribuent au développement du traitement orthographique du mot.

Parmi les nombreux autres modèles développementaux celui d'Ehri mérite également d'être cité. Il existe en plusieurs versions depuis les années quatre-vingt (Ehri, 1984, 1992, 1997). Il se constitue autour du concept du « *sight word* », désignant ainsi « *that sight of the word, activates that word in memory, including information about its spelling, pronunciation, typical role in sentences and meaning.* » (Ehri 1997, S. 168). Cette définition évoque un traitement selon « la voix directe », mais contrairement au modèle de Coltheart, pour Ehri l'identification immédiate du « *sight word* », inclut un traitement phonologique. Le « *sight word* » est donc un mot qui est identifié immédiatement (« à la vue ») lors de la lecture sans que le lecteur ait procédé à un déchiffrage et à une traduction explicite et lente de chaque graphème en phonème (Scheerer-Neumann, 1995). Il faut une bonne connaissance du niveau des compétences d'un lecteur pour pouvoir distinguer si un mot fait partie du stock des « *sight words* » car sa lecture est automatisée ou si son identification repose sur un apprentissage par cœur des images de mots (stade logographique). Ehri met l'accent sur l'importance de la maîtrise du recodage phonologique pour accéder à une lecture automatisée. Elle s'inspire des modèles connexionnistes pour expliquer la construction du stock des « *sight words* » à travers un processus de création de connexions (« *connection forming-process* »). La primauté du principe alphabétique dans le modèle d'Ehri (1997) est reflétée par la dénomination des quatre phases qui se succèdent dans le développement de la lecture avant d'arriver à une lecture qui repose sur l'identification immédiate d'unités consolidées qui peuvent être des morphèmes, des syllabes et même des mots entiers. L'utilisation de telles unités lors de la lecture automatisée de mots contrairement à une identification lettre par lettre a pu être confirmée (Bowey & Hansen, 1994; Treiman et al., 1990).

Les modèles développementaux, dont nous n'avons présenté que deux exemples, sont importants surtout en tant que cadre d'orientation pour le diagnostic, la pédagogie et la

remédiation dans le domaine de l'apprentissage de la lecture. Ils permettent une appréciation des compétences individuelles d'un lecteur, sans avoir recours à son âge ou son groupe sociale en tant que référence de base. Bien que la suite séquentielle des stades soit aujourd'hui contestée en faveur d'une utilisation flexible de diverses stratégies et que la validité universelle de ces modèles soit remise en question, ils ont amélioré notre compréhension des prérequis essentiels au développement des compétences en lecture.

Nous avons passé en revue dans cette partie théorique diverses approches pour modéliser la complexité de l'ensemble des processus impliqués dans la lecture ainsi que le contexte historique dans lequel elles sont apparues. Nous allons maintenant nous focaliser sur l'approche qui fournit le cadre théorique du présent travail.

*Résumé:*

*Au sein du chapitre précédent nous avons donné un aperçu des principaux courants qui ont marqué l'évolution de la recherche dans le domaine de la lecture. Les théories que nous avons évoquées se situent au sein des grands courants dominants de la psychologie. Ainsi nous avons thématiqué les approches encore centrées majoritairement sur les processus de perception du début du 20ème siècle. Nous voyons là des traces de la psychologie physiologique inspirée par Wilhelm Wundt.*

*L'idée que la lecture est plus un acte de raisonnement qu'un acte de perception est défendue par exemple par Thorndike. Sous l'ère du behaviorisme la recherche se limitait souvent à la lecture des mots, aux capacités de discrimination et d'articulation des lettres et des graphèmes. Par la suite ce sont d'un côté les linguistes d'orientation chomskyenne et de l'autre côté les psycholinguistes qui dominent la recherche des processus de lecture. Les psycholinguistes comme par exemple Goodman ou Smith prônent une approche globale centrée sur l'extraction du sens lors de la lecture.*

*Par la suite le paradigme du cognitivisme prédomina la recherche dans le champs de la lecture. La « théorie des deux voies » de Coltheart, les « modèles parallèles distribués » de Seidenberg et McClelland et les modèles de Kintsch et VanDijk sont des exemples issus de ce courant, qui influence encore la recherche contemporaine en tant que modèles de référence.*

*Puisque ce présent travail de thèse s'intéresse aux compétences de lecture chez les enfants, nous avons terminé le chapitre en présentant les modèles développementaux de Frith (modèles des trois stades « logographique, alphabétique et orthographique ») et Ehri (« sight word reading »).*

*Ainsi nous pensons avoir exposé les modèles fondamentaux dont la Simple View of Reading Theory s'est nourrie et qu'elle essaie d'intégrer tout en simplifiant la complexité des nombreux processus impliqués dans le processus de la lecture.*

### 1.3 Le modèle de la «*Simple View of Reading Theory*» : $L = D \times CO$ <sup>17</sup>

La *SVRT* conçoit la lecture (L) et plus précisément la compréhension de texte comme étant le produit de deux composantes :

- 1) la capacité de décoder de façon correcte et efficace la suite de graphèmes constituant les mots (le décodage D) ;
- 2) et la capacité de compréhension à l'orale (CO).

La particularité de l'approche de la *SVRT* est qu'elle essaie de fournir un modèle à la fois simple et efficace et en même temps intégratif. De nombreux modèles concernant le processus de lecture ont été proposés par des chercheurs et éducateurs spécialisés dans la question. Comme ces modèles sont basés sur une théorie spécifique, ils ne concernent souvent qu'un aspect particulier des processus de lecture tel que la reconnaissance des mots (par ex. la théorie de Goodman, le modèle de Coltheart, les modèles connexionnistes de Seidenberg et McClelland). D'autres modèles sont de nature très globale et décrivent des aspects de la cognition en général (par ex. les « *schema theory models* »). Les questions concernant les processus de compréhension de phrases et de textes sont traitées souvent séparément dans le cadre de modèles de la compréhension de textes (par ex. modèle interactif de Rumelhart (1985), ou les modèles de Kintsch (Kintsch, 2004; Kintsch & Van Dijk, 1978).

Le modèle de la *SVRT* essaie de rendre compte à la fois de la lecture des mots et de la compréhension de texte, car selon les auteurs de la *SVRT* ce sont les deux composantes essentielles de la lecture. Les auteurs de la *SVRT* proposent un modèle pragmatique qui essaie de lier des aspects de diagnostic et d'instruction pour être utile à l'identification des lecteurs faibles et à la remédiation de leurs difficultés. Créé au moment des « *reading wars* » aux Etats-Unis ce modèle veut réconcilier l'approche « globale » et l'approche « synthétique » en montrant que chacune des deux approches a quelque chose à apporter. La composante CO (compréhension orale) symbolise les complexités de la compréhension du langage qui sont impliquées dans la compréhension de la lecture. La composante D (décodage) représente l'importance du principe de l'association phonème

---

<sup>17</sup> Dans la publication originale de la « *The Simple View of Reading* » (Hoover & Gough, 1990) la formule est  $R(eading) = D(ecoding) \times L(inguistic\ comprehension)$ . Dans le présent travail nous utilisons une version francisée de cette formule :  $L(ecture) = D(écodage) \times CO(compréhension\ orale)$ .

/ graphème dans l'écriture alphabétique. La *SVRT* constituait donc à l'époque un premier exemple de ce qui est désormais appelé « *balanced literacy* », une méthode d'instruction qui intègre des aspects de l'approche syllabique et de la méthode globale.

Le modèle de la *SVRT* essaie également de rendre compte des aspects développementaux dans l'apprentissage de la lecture. Les auteurs (Hoover & Gough, 1990) postulant une asymétrie entre une influence plus forte des capacités de décodage sur l'explication de la variance en compréhension de texte chez les lecteurs apprentis et à l'inverse une influence plus forte des capacités de compréhension de langue chez les lecteurs affirmés.

Le modèle de la *SVRT* peut paraître réducteur. Il n'a toutefois pas été proposé comme une théorie élaborée et complète sur les processus détaillés de la lecture. Les auteurs souhaitent proposer un cadre théorique général qui permet d'étudier de nombreux aspects d'un phénomène complexe. Il est le résultat d'une simplification scientifique et d'une réduction nécessaire de la réalité, comme tout autre modèle de recherche. A cet égard, Kirby et Savage (2008) comparent la *SVRT* à une carte postale envoyée d'un grand voyage, et non à la présentation détaillée d'analyses et d'expériences résultant de ce voyage. La vraie question selon Kirby et Savage (2008) est de savoir si la réduction s'avère utile ou si elle a tellement appauvri le fond du phénomène à étudier, que le modèle devient inutile ou même nuisible pour un travail de recherche.

Si le modèle de la *SVRT* paraît simple, ses composantes (D et CO) ne le sont pas, car chacune englobe une multitude de processus complexes (Kirby et Savage, 2008). Le décodage paraît simple à un lecteur confirmé, mais il ne l'est pas pour un lecteur apprenti. En ce qui concerne la compréhension du langage (CO), nous continuons à apprendre tout au long de la vie. La *SVRT* reconnaît qu'au cours du développement de ces deux composantes, une multitude de facteurs « distaux » influencent l'émergence de ces capacités de compréhension de langue et de décodage. Mais le modèle prend explicitement le parti pris de se restreindre à la description et à l'étude des facteurs « proximaux » de la lecture. Les facteurs « distaux » ne sont pas considérés comme étant sans importance, mais le modèle de la *SVRT* postule qu'ils sont suffisamment pris en compte à travers les deux facteurs « proximaux ».

Une illustration de Scarborough (2001 voir ci-dessous) se prête particulièrement bien à l'illustration à la fois de l'interdépendance des deux facteurs « proximaux »: décodage

(*word recognition*) et compréhension de langue (*language comprehension*), ainsi qu'à l'illustration de la multiplicité des éléments «distaux » qu'ils impliquent.

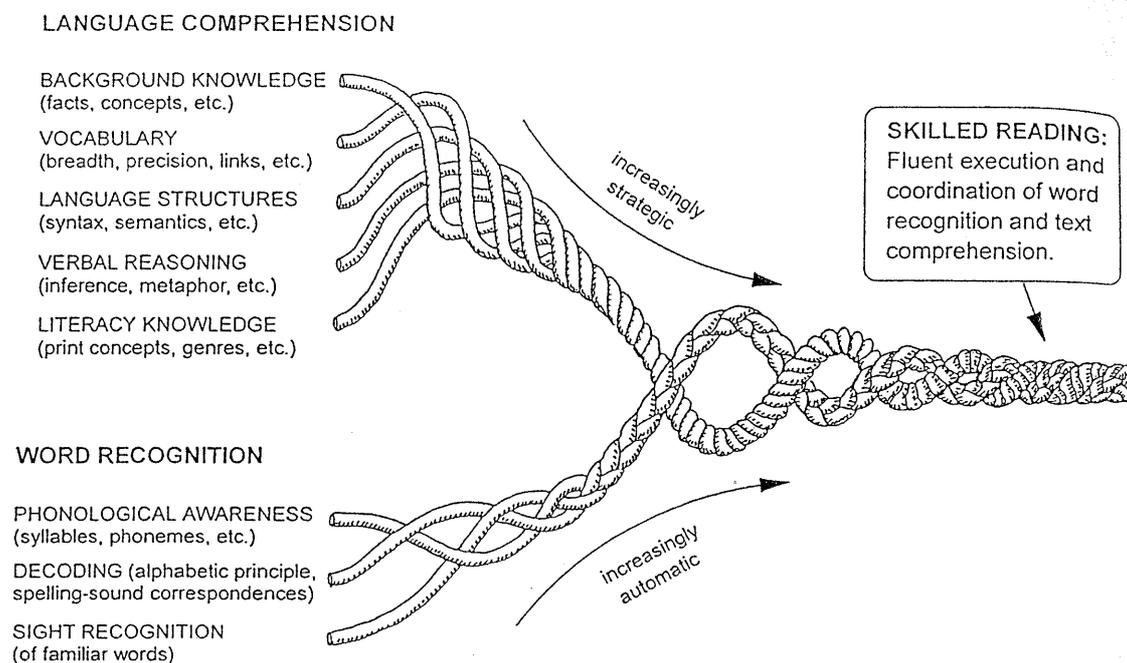


Figure 2 Caractère à multiples facettes de la lecture (Scarborough, 2001)

### 1.3.1 Les composantes de la «Simple View of Reading Theory» : $L = D \times CO$

#### 1) *L*(ecture)

Dans le cadre de la *SVRT*, la lecture repose – tout comme la compréhension orale – sur la compréhension de langue à l'oral (*CO*) à savoir sur la capacité d'extraire ou de déduire le sens d'une information d'ordre lexical. Le seul aspect spécifique de la compréhension de lecture est selon la *SVRT* le fait que cette information lexicale est disponible sous forme graphique et nécessite un travail de décodage (*D*). La *SVRT* se positionne donc comme une approche moniste en considérant que la compréhension à l'oral et à l'écrit relève de processus linguistiques réceptifs très similaires. Plusieurs arguments en faveur d'une telle vision moniste ont été cités (voir revue dans Vandergrift, 2006). Ces deux modes de compréhension dépendent d'une part des mêmes sources de connaissance (connaissances du monde, connaissance de la langue, compétences lexicales, grammaticales, sémantiques et syntaxiques) et d'autre part des processus ascendants et descendants sont nécessaires dans les deux cas pour intégrer ces connaissances et pour construire un modèle mental du contenu de ce qui a été lu ou entendu. On peut ajouter que les deux modes de compréhension sont flexibles, s'adaptant à différents objectifs

(Danks & End, 1987), en étant en même temps hautement automatisés, inconscients et irrépressibles. Maîtrisant une langue, on ne peut pas s'empêcher de comprendre du moins littéralement des mots écrits ou entendus comme en témoigne « l'effet Stroop »<sup>18</sup>(Stroop, 1935).

Cette approche moniste peut être critiquée, car malgré de nombreuses similarités, des différences existent entre la compréhension à l'oral et à l'écrit. Parmi les différences les plus importantes nous souhaitons mentionner le fait que la compréhension orale est un processus sériel qui s'inscrit dans le temps. Elle repose sur un matériel éphémère, contrairement à la base textuelle durable, qui permet au lecteur de faire des allers et retours au sein du texte et de récupérer une information lorsqu'elle a été oubliée. D'un autre côté, l'écoute dispose de sources d'informations paraverbales supplémentaires (prosodie, intonation, gestes, expressions mimiques) qui peuvent faciliter la compréhension en offrant plus d'informations utiles à la compréhension du message mais qui peuvent aussi rendre le processus de la compréhension orale plus exigeante en termes de charge cognitive (Buck, 2001). D'autres auteurs (Aust, 2006) voient dans les caractéristiques linguistiques du « matériel » une différence significative entre la compréhension à l'oral (souvent associée à une forme de langue plus spontanée et « naturelle ») et la compréhension basée sur un matériel écrit (considéré comme une forme de langue plus formelle, plus complexe en termes de structure mais aussi en termes de syntaxe et de lexique). L'association entre langue orale et « langue naturelle » d'un côté et langue écrite et « langue formelle » d'un autre côté (Olson, 1977 cité dans Hoover et Gough, 1990) a été remise en cause ces vingt dernières années. A titre d'exemple, on peut citer les travaux de Koch et Oesterreicher (1994) qui distinguaient et caractérisaient clairement un aspect « médial » et un aspect « conceptionnel » dans la définition de l'oralité et de la « scripturalité » (Koch, 1995). Tandis que l'aspect « médial » se réfère à la réalisation matérielle d'un énoncé et donc à la dichotomie entre graphique et phonique, l'aspect « conceptionnel » concerne les différences dans les situations communicatives et les variétés linguistiques qu'y correspondent. Ainsi l'oralité et la « scripturalité conceptionnelle » ne dépendent pas du médium, mais de la diversité des conditions situationnelles et des buts communicatifs dans l'activité langagière humaine. Koch et Oesterrei-

---

<sup>18</sup>Lorsque le nom d'une couleur est écrit dans une autre couleur on observe un ralentissement dans l'identification de cette couleur d'écriture du à l'interférence causée par l'identification du mot; par ex.: l'identification de la couleur noir du mot « rouge ». L'effet « Stroop » doit son nom au psychologue Ridley J. Stroop qui l'a découvert.

cher caractérisent « l'oralité conceptionnelle » par l'immédiat communicatif et la « scripturalité conceptionnelle » par la distance communicative. Selon eux l'immédiat extrême correspond à une conversation privée, spontanée, émotive avec un recours direct à la situation. La distance extrême correspond au monologue bien préparé et réfléchi qui s'adresse à des étrangers, sans recours à la situation et dépourvu d'émotions. Outre les prototypes de « l'oralité conceptionnelle » réalisés par la voie phonique et de la « scripturalité conceptionnelle » réalisés par la voie graphique, on trouve aussi des constellations mixtes, c'est à dire de « l'oralité conceptionnelle » à la réalisation graphique comme par exemple une lettre très privée ou les très populaires « sms ». A l'inverse un discours officiel ou un cours magistral universitaire sont des exemples d'une « scripturalité conceptionnelle » réalisée par la voie phonique.

Hoover et Gough n'abordent la distinction entre l'aspect « médial » et l'aspect « conceptionnel » que très brièvement. Ils y voient un argument en faveur de leur approche qui consiste à utiliser la compréhension orale comme facteur prédictif de la compréhension écrite à condition de rester dans le même genre de texte (par exemple narratif ou expositif) pour les tâches de lecture et d'écoute.

Habituellement les compétences de lecture (L) sont opérationnalisées au moyen de tests de lecture standardisés, consistant dans la lecture de textes plus ou moins courts (souvent sous contrainte du temps) et une deuxième partie où le lecteur doit répondre à des questions concernant le texte. Bien que valide, cette pratique ne semble pouvoir capter qu'une image restreinte des compétences en compréhension de texte (Kirby, 2007; S. G. Paris, 2007; Sweet, 2005). Ainsi on doit se demander si la compréhension de texte ne devrait pas être testée à l'aide de textes plus longs, des processus de construction de sens plus complexes, plus de questions ouvertes et ambiguës auxquelles on ne peut pas répondre grâce à des connaissances générales sur le monde. Cela rendrait les mesures de la compréhension de la lecture plus valides, mais aussi plus complexes et moins faciles à administrer.

Une question qui attire de plus en plus d'attention concerne l'impact des différentes mesures utilisées pour les compétences en lecture (Lyster, 2010; Rupp et al., 2006). En effet on peut supposer que différentes approches pour évaluer la compréhension de textes impliquant différents processus pourraient, par voie de conséquence, mener à des résultats différents lorsqu'on teste la validité d'un modèle de lecture. Les résultats de

Rupp et al. (2006) indiquent, par exemple, que des tests de lecture utilisant des réponses à choix multiple suscitent des processus spécifiques à ce format de test.

## 2) Compréhension de langue à l'orale (CO)

Dans la version originale anglophone de la *SVRT* une des deux composantes du modèle est dénommé *linguistic comprehension* et symbolisé par l'abréviation L. Nous avons traduit cette composante par la compréhension de langue à l'orale (CO). Le terme CO représente la capacité de compréhension de langue et englobe une multitude de processus allant de l'analyse syntactique, sémantique et morphologique jusqu'à la construction d'un modèle mental du contenu du texte. La compréhension d'un texte (à l'oral comme à l'écrit) nécessite la construction d'une représentation mentale qui intègre successivement les informations du texte. Tout en assurant la cohérence de la représentation élaborée sur la base d'informations déjà traitées, il faut intégrer de nouveaux éléments qui peuvent amener à des modifications et réajustements du modèle mental (Fayol & de Morais, 2004). Dans la compréhension textuelle, deux sources de difficultés particulières ont été identifiées, à savoir le traitement d'anaphores et le traitement d'inférences, les deux étant sensibles aux entraînements spécifiques. De façon plus générale, la difficulté de la compréhension d'un texte est souvent déterminée par : le genre textuel (par ex. texte narratif ou informatif), le contenu du texte (connu ou pas connu, facile ou difficile d'accès) et l'objectif de la lecture (par ex. se distraire, s'informer, apprendre).

Certains des modèles de recherche que nous avons évoqués au chapitre précédent essaient de modéliser des processus spécifiques contribuant à la compréhension, comme par exemple le modèle de construction et d'intégration (Kintsch, 2004), les travaux de Gathercole et Baddley (1993) sur le lien entre les capacités verbales et la mémoire de travail, ceux d'Oakhill & Cain (1999) sur la capacité de traiter des inférences ainsi que ceux de Maryanne Wolf sur l'influence du vocabulaire et de la morphologie (Wolf, 2001).

Les auteurs de la *SVRT* ont choisi pour des raisons de simplicité de réunir la grande complexité de ces processus multiples sous le seul terme de *linguistic comprehension* (compréhension de langue à l'oral). Mettre la grande complexité des capacités verbales réceptives sous un seul terme peut paraître comme une solution non seulement pragmatique mais aussi confortable. Il convient toutefois de rappeler que les auteurs de la *SVRT* ne prétendent pas présenter une théorie sur les capacités de compréhension verbale, qui sont par ailleurs bien étudiées et conceptualisées dans le cadre de nombreux travaux de

recherche, comme ceux que nous venons d'évoquer. Le terme de *linguistic comprehension* représente dans le cadre de l'approche moniste défendue par la *SVRT*, les processus communs de la compréhension à oral et à l'écrit. Par voie de conséquence, les auteurs de la *SVRT* utilisent la compréhension à l'oral afin de pouvoir opérationnaliser la capacité de compréhension de langue et d'expliquer la variance dans les capacités de compréhension de texte ( $L=D \times CO$ ).

De nombreuses études ont pu confirmer que les compétences en compréhension de texte sont corrélées de manière significative avec les compétences en compréhension orale chez les enfants (A. H. Paris & Paris, 2003; van den Broek et al., 2005; van Kraayenoord & Paris, 1996). Ces résultats justifient le choix des auteurs de la *SVRT* d'utiliser la compréhension orale comme opérationnalisation des processus de compréhension de langue qui sous-tendent la compréhension à l'oral et à l'écrit. .

### c) *D (écodage)*

Dans le cadre de la *SVRT*, la capacité de décodage se traduit par une reconnaissance et une identification rapide et correcte des mots qui sont présentés à la voie visuelle. Il s'agit donc ici du processus qui distingue la lecture de la compréhension orale, à savoir la traduction de la forme graphique des mots en représentations mentales sémantiques. Pour les auteurs de la *SVRT* c'est le niveau de maîtrise du décodage qui décide si oui ou non la compréhension en lecture atteint le niveau de la compréhension du même texte à l'oral.

Ce qui a été dit auparavant concernant la composante de la compréhension verbale est également valable pour le décodage : la *SVRT* ne prétend pas qu'il s'agit d'une composante simple, mais elle voit dans l'évaluation des capacités de reconnaissance des mots un moyen pratique et efficace pour expliquer en partie les difficultés en compétences de lecture. Le décodage s'appuie sur une multitude de processus lexicaux et sub-lexicaux. Nous avons évoqué dans le chapitre précédent des modèles qui ont tenté de décrire les processus menant à l'identification des mots lors de la lecture chez des lecteurs apprentis (voir les modèles développementaux par ex. de Frith ou de Ehri présentés au chapitre 1.2.5) ou de lecteurs experts (voir par exemples, les modèles parallèles distribués ou le modèle DRC de Coltheart et Rastles présentés au chapitre 1.2.4). Le modèle de double accès de Coltheart suggère par exemple deux possibilités de décodage des mots : par la voie phonique ou (plus directe) par la voie orthographique.

L'analyse phonique dépend de la conscience phonologique que l'enfant a pu acquérir au cours de son développement linguistique. La notion de conscience phonologique « renvoie à la connaissance de la structure sonore des mots » (Fayol & de Morais, 2004 p. 28). Elle est intrinsèquement liée aux expériences faites avec le système alphabétique. Ainsi Morais et ses collègues ont pu montrer que les personnes ne connaissant pas le système alphabétique ne savent pas réaliser la manipulation intentionnelle des phonèmes (Morais et al., 2004). L'importance de la conscience phonologique en tant que prérequis du développement des compétences en lecture est à ce jour bien étudiée. Ehri et ses collègues ont par exemple procédé à une méta-analyse de 54 études expérimentales et concluent qu'un entraînement de la conscience phonologique a un effet positif sur la reconnaissance des mots jusqu'à un à deux ans plus tard (Ehri et al., 2001).

L'analyse orthographique dépend des expériences de lecture et du contact avec le monde écrit dont l'enfant a bénéficié dès sa prime enfance. Des influences d'ordre sémantique, syntactique et contextuel jouent un rôle important dans la reconnaissance des mots. C'est à cause de l'aspect sémantique que les deux composantes de la *SVRT*- le décodage (D) et la compréhension de langue (CO)- interfèrent et ne peuvent de ce fait pas être considérées comme complètement indépendantes.

Les capacités de décodage peuvent être opérationnalisées de deux manières: par un test de reconnaissance de mots réels ou par la lecture de pseudo-mots. La dernière s'appuie exclusivement sur la connaissance des correspondances entre graphèmes et phonèmes et demande de lier lors de la prononciation les phonèmes, sans l'appui de la connaissance du mot. Hoover et Gough (1990) suggéraient d'utiliser la lecture/prononciation de pseudo - mots chez les lecteurs apprenants et la lecture de mots réels chez les lecteurs plus avancés afin d'éviter un effet de seuil chez ces derniers. Une étude utilisant la lecture de mots réels et la lecture de pseudo – mots ne montrait pas de différences significatives entre ces deux possibilités d'opérationnalisation du facteur D (Chen & Vellutino, 1997).

Dans des langues plutôt transparentes, ayant des correspondances grapho-phonémiques régulières telles que l'allemand, l'italien ou le finlandais, un décodage par analyse phonique peut être utile lors de l'apprentissage de la lecture. Toutefois, il ne suffit pas pour le décodage des mots dans une langue opaque comme le français ou l'anglais. Certains chercheurs proposent donc que les lecteurs d'une langue opaque utilisent des unités plus larges qu'une seule lettre (Ziegler & Goswami, 2005) et des informations spécifiques

sur les mots qui ne dépendent pas directement des règles de correspondance graphophoniques (CGP). Avec une lecture de plus en plus fluide et maîtrisée, c'est la voie orthographique qui est le plus souvent empruntée. Johnston et Kriby considèrent donc qu'un modèle de la *SVRT* qui opérationnalise D uniquement à travers la prononciation de pseudo - mots (analyse phonique) sera moins pertinent que celui qui utilise la reconnaissance de mots réels (Johnston & Kirby, 2006).

#### *d) Produit D x CO*

Des deux composantes de la *SVRT*, c'est le décodage (D) qui représente les processus de base fonctionnant plutôt sur un mode de « bas en haut » tandis que la compréhension de langue à l'oral (CO) fait intervenir de nombreux processus d'ordre supérieur. D'autres chercheurs (par exemple Rumelhart, 1985 dans le cadre du « interactive model of reading ») ont insisté avant la *SVRT* sur la nécessité d'une perpétuelle interaction entre les processus de base et les processus d'ordre supérieur grâce à un mélange de processus de bas en haut (bottom-up) et de haut en bas (top down)

Suivant cette logique, Hoover et Gough (1990) ont considéré que l'interaction entre D et CO était de nature multiplicative. Cela n'est pas sans importance car cela signifie que le niveau de compréhension de lecture dépend de l'ensemble des deux composantes. Un niveau zéro en compréhension de langue ou en décodage ne peut pas être compensé par l'autre composante. Il faut un niveau minimal dans les deux, pour un minimum de compréhension de texte. Cela paraît surtout important dans les cas extrêmes, par exemple lorsqu'un enfant maîtrise une langue à l'oral, mais ne connaît pas encore le principe du décodage et des règles CGP. A l'inverse, un étranger peut parfaitement maîtriser le décodage, mais il sera privé de toute compréhension tant qu'il ne comprend pas la langue. Une illustration anecdotique de ce sujet est citée dans différents travaux (Davis, 2006; Philipp B. Gough et al., 1996; Snowling & Hulme, 2005). Il s'agit de l'exemple du poète anglais John Milton qui, avec l'âge, fût privé de la capacité de lire ses textes préférés écrits en grec ancien. Il enseigna alors à ses filles le décodage du grec et leur demanda d'oraliser ces textes qu'il restait le seul à comprendre. Ainsi la compréhension de texte a pu être menée à bien grâce aux compétences de décodage des filles et à la compréhension du grec ancien par le père.

Toutefois, ces cas extrêmes (absence totale de la compréhension de la langue ou des compétences de décodage) ne font souvent pas partie de l'échantillon typique d'une

recherche sur les capacités en lecture. On peut donc s'interroger sur la nécessité de l'utilisation du terme multiplicatif dans le modèle de la *SVRT*.

### 1.3.2 Analyse critique des prémisses de la « *SVRT* »

Après avoir été proposé en 1986 (par Gough et Tunmer), le modèle de la *SVRT* a été vérifié au cours d'une étude longitudinale auprès d'un échantillon d'enfants bilingues (anglais - espagnol) dont les résultats ont été publiés par Hoover et Gough en 1990. Des spécifications supplémentaires concernant la *SVRT* ont été publiées en 1996 par Gough, Hoover et Peterson. A la base de ces trois articles fondateurs, on peut identifier quatre idées clés qui caractérisent ce modèle de la lecture :

- 1) Les compétences en lecture reposent essentiellement sur deux composantes distinctes : les capacités spécifiques au traitement du langage écrit (le décodage D) et les compétences linguistiques générales de l'individu (la compréhension orale CO), cette dernière étant opérationnalisée par la compréhension à l'oral.
- 2) Les deux composantes D et CO sont nécessaires, aucune des deux à elle seule ne suffit pour la réussite du processus de lecture. C'est pourquoi la relation entre D et CO s'exprime par une interaction multiplicative. D et CO contribuent chacune de façon significative, à l'explication des variances dans les compétences de compréhension écrite, mais le produit des deux ( $D \times CO$ ) améliore encore significativement l'explication des variances.
- 3) Par voie de conséquence, chez les enfants ayant des difficultés dans l'apprentissage de la lecture, on peut identifier trois profils possibles en fonction de l'origine pré-supposée des difficultés. Ces dernières peuvent résider principalement dans les processus de décodage. Pour un autre groupe de lecteurs en difficulté, elles se situent principalement au niveau de la compréhension et enfin un troisième groupe cumule les deux faiblesses.
- 4) La façon dont les deux composantes D et CO contribuent à l'explication des variances interindividuelles en compréhension écrite change au cours du développement du jeune lecteur apprenti. En effet, le modèle de développement de la *SVRT* (Gough, Hoover et Peterson, 1996) suggère qu'au début de l'apprentissage ce sont en premier lieu les capacités de décodage qui déterminent les compétences de compréhension écrite. C'est dans ce domaine qu'il y a de très grandes différences entre les enfants tandis que les différences interindividuelles en compréhension de langue ne pèseront

pas encore beaucoup sur la compréhension des textes -d'un contenu encore assez simple- auxquels ils sont confrontés. Au cours de l'enseignement institutionnalisé de la lecture, le niveau des capacités de décodage est moins hétérogène, tandis que des textes plus complexes font de plus en plus peser les différences interindividuelles en compréhension de langue sur la compréhension de texte. Ainsi le poids des deux prédicateurs D et CO change de façon asymétrique.

La *SVRT* a attiré depuis ses débuts l'attention de nombreux chercheurs, surtout dans le monde anglophone. Les points clés de la *SVRT* ont été vérifiés par exemple dans le cadre d'études de « validation croisée » qui tentent de reproduire dans un nouvel échantillon les résultats obtenus ou prédits par les auteurs de la *SVRT* (Chen & Vellutino, 1997; Dreyer & Katz, 1992). De nombreuses études ont vérifié des points spécifiques de la *SVRT*. Ainsi certains travaux ont comparé l'efficacité<sup>19</sup> de la *SVRT* avec d'autres modèles de lecture (Verhoeven & Leeuwe, 2008), d'autres études ont testé la validité de la *SVRT* auprès d'un groupe particulier de personnes (par ex. enfants pré-lecteurs (Aouad & Savage, 2009), étudiants d'université (Savage & Wolforth, 2007), enfants canadiens indigènes (Georgiou et al., 2009) ou en utilisant une méthode d'analyse spécifique comme l'analyse en composantes principales, (Kendeou et al., 2009; Protopapasa et al., 2012). Nous allons par la suite passer en revue les résultats principaux de ces études afin de pouvoir identifier les points forts et faibles de la *SVRT* avant d'aborder les questions relatives à la *SVRT* qui restent encore en suspens.

### **1.3.3 Données empiriques concernant la « Simple View of Reading Theory »**

#### *1.3.3.1 L'importance des deux facteurs : Compréhension Orale et Décodage*

Considérant le nombre d'études ayant confirmé par le biais de différentes techniques que le décodage (D) et la compréhension orale (CO) sont deux composantes principales de la lecture, la première idée clé de la *SVRT* est généralement admise. Grâce à des études de corrélation, de régressions multiples, des analyses factorielles ou des modèles d'équations structurelles, il a été confirmé que D et CO sont deux facteurs distincts (Kendeou, Savage, van den Broek, 2009 ; Aouad & Savage, 2009) qui contribuent indépendamment l'un de l'autre à l'explication d'une grande partie de la variance dans les compétences en compréhension de texte (Catts et al., 2005; Chen & Vellutino, 1997;

---

<sup>19</sup> L'efficacité du modèle de la *svrt* est estimée par nombreux auteurs en termes de pourcentage de variance en compréhension de texte expliqué par les facteurs (D x CO) de la *svrt*.

Dreyer & Katz, 1992; Philipp B. Gough et al., 1996; Johnston & Kirby, 2006; Joshi & Aaron, 2000; Savage, 2001; Janet Tilstra et al., 2009). La proportion de variance expliquée se situe dans ces études entre 40 et 80 % pour des lecteurs âgés de 7 à 15 ans.

Le modèle de la *SVRT* a également reçu le soutien d'études dans le domaine de la génétique comportementale. Après avoir procédé à une étude comparative entre des jumeaux monozygotes et des jumeaux dizygotes Keenan et ses collègues ont conclu que les relations entre le décodage (D) et la compréhension de lecture (L) ainsi qu'entre la compréhension orale (CO) et la compréhension de lecture (L) sont largement dues aux gènes et moins aux environnements partagés (Keenan et al., 2006). Prises ensemble, les relations génétiques entre D et L et CO et L constituent la totalité de l'influence génétique sur les compétences en compréhension de lecture (L). Il convient de noter que l'échantillon de cette étude impliquait une grande variabilité des âges et des compétences en lecture.

Marx et Jungmann (2000) ont confirmé le pouvoir prédictif du modèle de la *SVRT* auprès d'un échantillon de jeunes lecteurs germanophones de l'école élémentaire. Megherbi, Seigneuric et Ehrlich (2006) ont confirmé l'influence des facteurs D et CO sur la compréhension de texte auprès d'un groupe de lecteurs apprentis francophones, âgés de 6 à 8 ans. Aouad et Savage (2009) mettent en évidence à l'aide d'une analyse factorielle l'existence des composantes de pré-décodage et de compréhension chez les pré-lecteurs canadiens anglophones.

Il semble donc que le modèle de la *SVRT* soit un bon prédicteur des compétences de lecture surtout au cours des quatre premières années d'apprentissage. Tilstra et ses collègues (2009) ont examiné la validité du modèle de la *SVRT* auprès de trois groupes d'âge. Leurs résultats montrent que la *SVRT* a dans les trois groupes un pouvoir explicatif significatif. Néanmoins, elle explique une partie de variance bien plus importante chez les élèves âgés de 10 ans (61%) que chez les élèves âgés de 13 ans (48%) et de 15 ans (38 %). Cela suggère, comme les auteurs le font remarquer, que la compréhension de texte implique au cours du cursus scolaire d'autres facteurs supplémentaires, qui vont au delà du décodage et de la compréhension orale, comme par exemple les capacités de la mémoire de travail, la capacité de faire des inférences, la connaissance des structures des textes (grammaire de récit) spécifiques au genre (narratif, expositif, informatif) ou encore la capacité d'autorégulation des processus de compréhension (Perfetti et al., 2005). De plus, il est probable que les textes proposés aux élèves sont tant pour le vocabulaire que pour la structure de plus en plus complexes et différents du langage oral

(Diakidoy et al., 2005). L'influence de la compréhension orale sur la compréhension de ces textes sera donc moindre, ce qui expliquerait l'efficacité décroissante de la *SVRT*.

Toutefois même décroissant, le modèle de la *SVRT* explique encore une partie importante de la variance en compétences de lecture chez les lecteurs de 15 ans selon Tilstra et ses collègues (2009). De plus, les études de Savage (2001) et de Savage et Wolforth (2007), laissent supposer que la combinaison de D et CO garde un pouvoir prédictif jusqu'à l'âge des études supérieures. Il existe aujourd'hui un consensus dans la communauté scientifique qui s'occupe des questions de lecture, selon lequel le modèle de la *SVRT* capte les processus importants de l'apprentissage de la lecture et ce malgré son apparente simplicité.

### 1.3.3.2 L'interaction entre D et CO

Chen et Vellutino (1997) ainsi que Dreyer et Katz (1992) ont procédé à une validation croisée de la « *Simple View of Reading Theory* ». Chen et Vellutino ont tenté de reproduire les résultats publiés par Hoover et Gough (1990) auprès d'un groupe d'élèves monolingues (anglais). Leurs résultats confirment le premier point que nous venons d'aborder ci-dessus, c'est-à-dire l'idée d'un découpage général entre les capacités de décodage (mesurées par des tests de « *phonetic ability* » ou de « *real word naming* ») et de compréhension linguistique (mesuré par des tests de compréhension orale). Ils sont toutefois plus réservés en ce qui concerne la nécessité d'une combinaison multiplicative de ces deux composantes. Chen et Vellutino s'accordent avec Gough et Tunmer (1986) sur le fait qu'en temps normal, la combinaison multiplicative entre D et CO explique une part substantielle de la variance de la compréhension de lecture. Si la compréhension de langue (CO) est nulle (dans le cas d'une lecture en langue étrangère) ou les compétences de décodage (D) nulles (dans le cas d'un lecteur apprenti), la compréhension de lecture sera également nulle. Il n'y aurait pas d'effet de compensation entre les deux composantes en dessous d'un niveau critique de maîtrise minimale. Ainsi pour Gough et Tunmer, la combinaison multiplicative est une nécessité pour assurer la validité écologique de la *SVRT*.

Mais contrairement à Hoover & Gough (1990), qui trouvent une augmentation de la variance expliquée lorsqu'on inclut le produit de  $D \times CO$  après avoir pris en compte la combinaison linéaire ( $L = b_0 + b_1D + b_2CO + b_3D \times CO$ ), Chen et Vellutino (1997) ne trouvent pas de variance supplémentaire expliquée. Ils observent cependant une aug-

mentation de la variance expliquée lorsqu'on introduit la somme après la combinaison multiplicative.

Chen et Vellutino (1997) proposent deux raisons possibles au fait qu'ils n'ont pas pu répliquer les résultats de Hoover et Gough (1990). D'une part, les effets d'interactions sont difficiles à détecter avec des analyses à régressions multiples car le terme multiplicatif est toujours moins fiable que ses composantes prises séparément à cause de la multiplication des erreurs de mesure. McClelland & Judd (1993) concluent que les analyses de régressions multiples n'ont qu'une efficacité de 20% par rapport aux méthodes expérimentales en ce qui concerne la détection d'un effet d'interaction.

La deuxième explication possible proposée par Chen et Vellutino (1997) concerne les caractéristiques de l'échantillon. L'étude de Hoover et Gough (1990) se référait à un échantillon bilingue avec une plus grande variation dans les capacités de décodage, de compréhension orale et de compréhension de texte que dans l'échantillon de Chen et Vellutino. Cette variation a également un effet sur la fiabilité des composantes et du terme multiplicatif.

Chen et Vellutino (1997) concluent à la vue de leurs résultats que la meilleure prédiction des compétences en lecture est réalisée avec une version « atténuée » mais plus compliquée de la *SVRT* ( $L = D + CO + D \times CO$ ) qui implique l'utilisation d'éléments additifs et multiplicatifs.

A la base d'une autre étude de validation croisée, Dreyer et Katz (1992) confirment comme Chen et Vellutino (1997) le pouvoir prédictif du modèle de la *SVRT* qu'ils ont évalué doublement. D'abord en testant la validité du modèle auprès d'un groupe d'élèves monolingues anglais âgés d'environ 8 ans (validité concurrente) et ensuite en testant le pouvoir prédictif de  $D \times CO$  mesuré à 8 ans sur les compétences en lecture ( $L$ ) deux ans plus tard, lorsque les élèves ont environ 10 ans (validité prédictive). Ils obtiennent dans les deux cas une explication de variance de la compréhension de texte entre 37 et 48 %. Dreyer et Katz font toutefois remarquer que l'utilisation du terme additif ( $D + CO$ ) a un pouvoir prédictif supérieur à celui du terme multiplicatif ( $D \times CO$ ), le terme additif augmentant la partie de variance expliquée de 7.4% (8 ans) et de 8.1% (10 ans) par rapport à la seule utilisation du terme multiplicatif.

Parmi les études récentes, Joshi et Aaron (2000), Savage et Wolforth (2007), Georgiou, Das et Hayward (2009) n'ont trouvé aucun pouvoir explicatif supérieur de l'interaction

multiplicative par rapport à l'interaction additive. Lyster (2010) obtenait une explication de variance plus élevée pour la version additive. La nécessité d'une interaction multiplicative continue ainsi à être mise en doute.

### 1.3.3.3 Dissociation entre *D* et *L* visible dans différents profils de lecteurs

L'idée d'une dissociation et (quasi)indépendance entre les composantes de décodage et de compréhension de langue a reçu - comme nous l'avons vu au chapitre 1.3.3.1 - un fort soutien empirique. Des travaux utilisant des analyses factorielles ont confirmé *D* et *CO* comme deux facteurs distincts, (Kendeou, Savage & van den Broek, 2009 ; Aouad & Savage, 2009) tandis qu'une étude de génétique comportementale auprès de jumeaux indique même que différents facteurs génétiques et environnementaux influencent les deux facteurs *D* et *CO* (Keenan et al., 2006). Des études neuropsychologiques confirment également l'idée d'une dissociation en montrant par exemple, par le biais de l'imagerie neuronale, que la rétention d'information phonologique (nécessaire au décodage) ou sémantique (nécessaire à la compréhension de langue) dans la mémoire du travail, implique différentes structures du cortex cérébral (Poldrack & Wagner, 2004).

Cette dissociation entre les composantes *D* et *CO* va de pair avec la troisième idée clé de la *SVRT* concernant les trois facteurs supposés générer de faibles compétences en compréhension de texte :

- 1) de faibles compétences en compréhension de langue (*CO*) ;
- 2) de faibles compétences en décodage (*D*) ;
- 3) ou un troisième groupe qui cumule les faibles compétences en décodage et en compréhension de langue.

En effet, plusieurs études ont pu identifier des lecteurs faibles ayant des faiblesses en termes de décodage mais une bonne compréhension de langue. Le cas extrême de ce type de profil correspond à la dyslexie. Le profil inverse paraît moins répandu, même s'il est bien documenté depuis quelques années grâce à des recherches surtout menées en Grande Bretagne. Ainsi Stothard et Hulme (1992) et Nation et Snowling (1998, 1999) ont étudié les différentes compétences en lecture dans des groupes lecteurs qui disposaient du même niveau de capacités de décodage mais dont le niveau de compréhension de texte se différençait très clairement. Ces études indiquent qu'environ 10 à 15 % des élèves âgés de 7 à 9 ans montrent de faibles niveaux de compréhension de texte malgré une bonne maîtrise du décodage. Dans une étude auprès d'enfants francophones âgés de

6 à 8 ans, Megherbi, Seigneuric et Ehrlich (2006) mettent en évidence que parmi les enfants ayant une faible compréhension de texte, environ 50 % disposent de bonnes compétences de décodage.

Ce groupe de lecteurs faibles est appelé chez différents auteurs « technical readers » ou « word callers ». Oakhill et ses collègues (1988, 1999, 2003) ont pu montrer les faiblesses suivantes dans la compréhension et production de langue dans des groupes d'enfants de 7 à 8 ans possédant ce même profil:

- des déficiences dans le domaine de la construction et du maintien de cohérence (Yuill & Oakhill, 1988);
- des déficiences dans l'intégration d'information et la capacité à faire des inférences (Cain & Oakhill, 1999; J. V. Oakhill et al., 2003).

Lerkkanen et ses collègues identifient dans une étude longitudinale auprès d'un échantillon de 90 élèves finlandais de 7 à 8 ans trois groupes de lecteurs (Lerkkanen et al., 2004):

1. des lecteurs compétents avec un bon niveau sur le plan de l'identification de mots et de la compréhension de texte (« competent readers ») ;
2. un deuxième groupe avec une bonne maîtrise de l'identification des mots mais une faible compréhension de texte (« technical readers ») ;
3. et un troisième groupe avec des difficultés dans les deux domaines (« poor readers »).

L'appartenance à ces groupes était assez stable au cours d'une année, surtout dans le group des « competent readers » et des « technical readers ». Cependant, comme les compétences en compréhension orale ne faisaient pas partie de cette étude, il n'est pas surprenant que Lerkkanen et al. n'aient pas identifié un quatrième groupe avec de faibles compétences en identification de mots mais une bonne compréhension de la langue. Il est probable que ce profil est noyé dans le groupe des « poor readers ».

Les résultats de Braibant présentent à nos yeux un défi à l'idée d'une nécessité absolue des deux composantes D et CO pour une compréhension de texte satisfaisante (Braibant, 1994). Braibant démontre dans une étude auprès de 157 enfants belges (2<sup>ème</sup> année de l'école primaire), l'existence de quatre groupes de lecteurs, dont deux ayant un

profil « homogène » et deux ayant un profil « atypique » (Braibant, 1994 p. 191). Le profil « homogène » se caractérise par une compréhension écrite en adéquation avec les compétences en décodage (soit les deux sont faibles, soit les deux sont bons), tandis que le profil atypique regroupe le profil des « bons décodeurs - mauvais compreneurs » (p.191) qui correspond au « technical readers ». Ces trois profils s'insèrent parfaitement dans les prédictions de la *SVRT* et dans la lignée des résultats d'études citées auparavant. Toutefois, l'existence d'un groupe de bons lecteurs avec une bonne compréhension de texte malgré un faible niveau de décodage semble déconcertante. La *SVRT* suppose qu'une bonne compréhension de langue puisse être accompagnée d'une faible maîtrise du décodage (puisque ce sont deux facteurs distincts), mais il en résulte une faible compréhension de texte. Or chez Braibant, 14% des enfants testés sont au dessus de la moyenne en compréhension de lecture (et donc des bons lecteurs) tout en étant des « mauvais décodeurs », (Braibant, 1994, p 191). L'auteur cite des résultats comparables obtenus par Aaron (1989) et Yuill et Oakhill, (1991) et fait appel à l'hypothèse de compensation (Keith E. Stanovich, 1980) pour interpréter ces résultats. Braibant fait remarquer que son épreuve de compréhension écrite<sup>20</sup> favorise particulièrement la compensation de difficultés en décodage par l'utilisation d'informations contextuelles car elle présente simultanément du texte et des images et se réalise sans contrainte de temps. Des nombreuses relectures ainsi que l'utilisation des informations fournies par les images permettraient selon Braibant à ces « mauvais décodeurs » mais « bon compreneurs » d'atteindre des performances en compréhension de texte comparables à celles des « bons décodeurs ».

Nous touchons là un aspect qui sera abordé en détail au chapitre suivant, à savoir la dépendance à l'égard des instruments de mesure des compétences en lecture pour évaluer la validité d'un modèle de lecture comme la « *Simple View of Reading Theory* ».

#### 1.3.3.4 Développement asymétrique

Le quatrième point clé de la *SVRT* aborde la question du développement des compétences en lecture et formule l'idée d'une asymétrie dans les contributions des deux facteurs décodage (D) et compréhension orale (CO).

---

<sup>20</sup> Braibant utilisait « L'épreuve d'évaluation de la compétences en lecture : Lecture de mots et compréhension » de A. Khomsi (1990) Paris : Editions du Centre de Psychologie Appliquée.

Dans une meta-analyse de 10 études différentes (17 échantillons), Gough, Hoover et Peterson (1996) ont examiné les corrélations obtenues entre les compétences de lecture (L) et de décodage (D) ainsi que les compétences de lecture (L) et de compréhension orale (CO). Pour être incluses dans cette meta-analyse, les études devaient satisfaire à deux conditions :

- a) les méthodes d'évaluation des compétences en compréhension orale n'étaient pas « polluées » par les compétences en décodage ;
- b) les compétences en compréhension orale et écrite étaient mesurées avec des méthodes qui correspondaient l'une à l'autre.

Les 17 échantillons des 10 études en question, ont été catégorisés en quatre groupes (CP/CE1 ; CE2/CM1 ; CM 2/6<sup>ème</sup> ; > 6<sup>ème</sup>). Les analyses effectuées par Gough, Hoover et Peterson (1996) montraient pour les deux types de corrélations (L & D, L & CO) un effet entre-groupes significatif, indiquant ainsi que les corrélations dépendent du niveau scolaire des enfants testés. Les corrélations moyennes pour L & D diminuent de 0.61 (CP/CE1) à 0.53 (CE2/CM1) à 0.48 (CM2/ 6<sup>ème</sup>) à 0.39 (> 6<sup>ème</sup>). A l'inverse les corrélations moyennes entre L et CO montaient de 0.41 (CP/CE1) à 0.50 (CE2/CM1) puis à 0.72 (CM2/ 6<sup>ème</sup>) et à 0.68<sup>21</sup> (> 6<sup>ème</sup>). Les auteurs conclurent à la vue de ces résultats de meta-analyse que le modèle développementale de la *SVRT* est largement confirmé, du moins en ce qui concerne les tendances générales. Lorsqu'on regarde les 17 études séparément, on constate que 4 des 17 études analysées ne montrent pas l'asymétrie supposée par la *SVRT* (Cunningham et al., 1990; Palmer et al., 1985; Keith E. Stanovich et al., 1984; Keith E. Stanovich et al., 1986).

Parmi les études plus récentes, on trouve également des travaux aux résultats contradictoires. Catts Hogan et Adloff (2005) ont confirmé que la contribution relative de ces deux composantes varie avec l'âge des lecteurs. Utilisant une approche de régressions multiples dans un échantillon d'élèves anglophones en 2<sup>ème</sup>, 4<sup>ème</sup> et 8<sup>ème</sup> classe, les auteurs ont obtenu une augmentation du taux de variance expliquée par le facteur CO (9% en 2<sup>ème</sup> classe, soit l'équivalent du CE1, 21 % en 4<sup>ème</sup> soit l'équivalent du CM1 et 36 % en 8<sup>ème</sup> classe, soit l'équivalent de la 4<sup>ème</sup>), et une diminution du taux de variance expliquée dû au facteur D (27 % en 2<sup>ème</sup>, 13 % en 4<sup>ème</sup> et moins de 1 % en 8<sup>ème</sup> classe).

---

<sup>21</sup> Gough, Hoover et Peterson expliquent le déclin de la corrélation moyenne entre L et CO dans le 4<sup>ème</sup> groupe par le fait qu'une des trois études dans ce groupe employait une méthode d'évaluation de la compréhension orale peu valide.

Tilstra et al. (2009) observent une diminution de la force explicative de la *SVRT* chez les lecteurs de treize à quinze ans et confirment également l'augmentation de l'importance du facteur CO en même temps que l'influence du facteur D diminue au cours du développement des compétences en lecture. Vellutino et ses collègues utilisent des analyses de modèles d'équations structurelles et observent également le développement asymétrique des deux facteurs tel que les auteurs de la *SVRT* l'ont supposé (Vellutino et al., 2007).

A l'inverse Carver (1998) conclut à la vue des ses résultats que D reste un facteur de haute importance même auprès des lecteurs en 6<sup>ème</sup>. Megherbi, Seigneuric et Ehrlich (2006), au contraire, constatent dans leur étude auprès de jeunes lecteurs francophones, que la compréhension orale est déjà en CP et CE1 plus importante que le décodage dans l'explication de la variance de la compréhension de texte. De Jong et Van der Leij (2002) obtiennent des résultats similaires dans un échantillon de lecteurs néerlandais. Les résultats de Verhoeven et Leeuwe (2008), auprès de lecteurs néerlandais, confirment une influence incohérente du facteur décodage dans l'explication des compétences en lecture de textes.

Ces divergences dans les résultats nous semblent être en partie liées à des divergences entre les méthodes d'évaluation mais aussi à des différences d'ordre linguistique. Nous allons approfondir ces deux points au chapitre suivant.

Afin de conclure sur la présentation de la « *Simple View of Reading Theory* », nous pouvons constater que ce modèle, dans ses grandes lignes du moins, a reçu un soutien empirique considérable en provenance de différents champs de la recherche et utilisant des méthodes variées. Le décodage et la compréhension orale sont bien établis en tant que prédicteurs fiables des compétences en compréhension de texte. La différenciation entre des groupes de lecteurs en difficulté qui se distinguent suivant que leurs faiblesses se trouvent dans le domaine du décodage, de la compréhension ou dans les deux, fournit un cadre théorique important pour le développement de méthodes d'enseignement et de remédiation. Il est important de prendre en compte les variances développementales dans l'influence de ces deux facteurs lorsqu'il s'agit de concevoir des méthodes d'enseignement de la lecture adaptées aux groupes ciblés.

Actuellement, de nombreuses recherches, notamment anglophones s'appuient plus ou moins explicitement sur le modèle de la *SVRT* en tant que cadre théorique. En Angleterre la *SVRT* a récemment remplacé le « *searchlight model* » comme base théorique au

sein de la « *National Literacy Strategy (NLS)* » (Rose, 2006). Néanmoins la *SVRT* demande comme tout autre modèle de recherche à être mis et remis à l'épreuve, voire à être amélioré. Dans ce sens, la *SVRT* reste un modèle «in progress».

Dans le chapitre suivant nous allons brièvement passer en revue les manques qui ont été identifiés dans le modèle de la *SVRT* ainsi que les modifications qui ont été proposées pour y remédier. Nous allons conclure cette partie en développant les questions clé qui, de notre point de vue, méritent encore d'être traitées sur le thème de la « *Simple View of Reading Theory* ». Elles constituent le fondement sur lequel reposent les analyses statistiques que nous présenterons dans la troisième partie du présent travail de doctorat.

*Résumé :*

*La Simple View of Reading Theory résulte de la volonté de fournir un modèle utile à la recherche et la remédiation qui simplifie la complexité des multiples processus impliqués dans la lecture. Ce modèle veut également concilier les positions en intégrant les points clés d'une approche globale de l'enseignement de la lecture et d'une approche synthétique.*

*La SVRT a suscité dans la communauté des chercheurs anglophones beaucoup d'intérêt et ses postulats principaux ont reçu du soutien empirique. Le pouvoir prédictif des deux facteurs décodage et compréhension orale a été confirmé par de nombreuses études.*

*Ce modèle semble bien pouvoir capter, malgré son apparente simplicité, les processus importants dans l'apprentissage de lecture en anglais.*

## 2 Questions de recherche

### 2.1 La validité d'un modèle additif de la « *Simple View of Reading Theory* »

La première question qui guide nos analyses, concerne la validité de la « *Simple View of Reading Theory* » auprès d'un échantillon d'élèves germanophones de différents âges. Etant donné que la nécessité d'une combinaison multiplicative des deux facteurs décodage (D) et compréhension orale (CO) est contestée comme nous l'avons vu au chapitre 1.3.2, le modèle choisi pour nos analyses se focalise sur les deux composantes D et CO, mais n'introduit pas le produit  $D \times CO$ . Il nous permet de tester si le décodage et la compréhension orale contribuent de manière significative à l'explication de variances résiduelles de nos trois variables dépendantes. En nous appuyons sur les travaux des auteurs de la *SVRT* (Philipp B. Gough et al., 1996; Philipp B. Gough & Tunmer, 1986; Hoover & Gough, 1990) nous formulons deux hypothèses :

- H1a : Nous supposons que le décodage et la compréhension orale contribuent de manière significative à l'explication de variances des compétences en lecture.
- H1b : Nous supposons que l'influence du facteur Compréhension Orale est plus grande dans le groupe des élèves en CM2 / 6<sup>ème</sup>, tandis que le poids du facteur Décodage est plus grand dans le groupe des élèves en CE1 / CE2.

### 2.2 Le rôle des opérationnalisations de la compréhension de texte

Une question à laquelle la *SVRT* n'a pas encore répondu de manière satisfaisante selon Kirby et Savage (2008) concerne le rôle de l'opérationnalisation concrète des compétences en lecture dans les différents tests utilisés dans les travaux de recherche autour de la *SVRT*.

En effet, des stratégies spécifiques servent à localiser une information au sein du texte, à déterminer la structure du texte, ou à utiliser par exemple des indices typographiques apportant à la compréhension de texte des éléments qui vont bien au delà de la compréhension orale et du décodage. L'impact de ces stratégies spécifiques sur les résultats de la *SVRT*, dépend fort probablement de la façon dont la compréhension de texte sera évaluée. Ainsi on peut s'imaginer que la force explicative de la *SVRT* et le poids de ses deux facteurs D et CO varient en fonction de la méthode et de l'instrument choisis pour évaluer les compétences en lecture de textes. En effet Nation et Snowling (1997) ont

constaté lors de la comparaison de deux tests britanniques de la compréhension de texte (« *Neale Analysis of Reading Ability* » et « *Suffolk Reading Scale* ») que les deux étaient influencés par le facteur D mais que le facteur CO avait une influence significative uniquement dans le test de *Neale*. Le test de *Suffolk* consistant à compléter des phrases avec le mot adéquat (*cloze test*), ne reçoit ainsi pas d'influence du paramètre compréhension orale (Kate Nation & Snowling, 1997). Depuis, la comparabilité et la correspondance des tests de compréhension textuelle – jusque là souvent présumée – est remise en cause par certains chercheurs. Cutting et Scarborough (2006), ainsi que Keenan, Betjemann et Olson (2008), ont comparé systématiquement plusieurs tests très répandus dans la recherche en lecture aux Etats-Unis. Leurs résultats indiquent une contribution très variable du décodage et de la compréhension orale à l'explication de variance dans la compréhension de texte selon les tests utilisés. Keenan et al. (2008) ont aussi testé l'idée d'une asymétrie développementale dans les contributions des deux facteurs. Ils constatent que seulement deux des cinq tests analysés montrent une augmentation de l'influence de la compréhension orale et une diminution de l'influence du Décodage au cours du développement, tel que la *SVRT* le postule. Ces deux tests sont le « *Woodcock Johnson Passage Comprehension Subtest* » (*WJPC*) et le « *Reading Comprehension Subtest from the Peabody Individual Achievement Test* » (*PIAT*). Contrairement aux autres trois tests qui demandent une lecture à voix haute de passages plus longs, ils se distinguent par une lecture silencieuse de phrases ou de passages très courts. Le *PIAT* demande l'identification de l'image qui correspond au sens de phrases lues, tandis que le *WJPC* est un *cloze test* et demande l'identification du mot manquant au sein d'une phrase. Dans les deux cas, l'identification de la bonne réponse dépend très fortement de la lecture correcte de certains mots qui servent d'indices et qui à eux seuls suffisent pour trouver la bonne réponse même si le sens global de la phrase n'a pas été compris. Les auteurs illustrent cela grâce à l'exemple suivant : La phrase test est « *The patients were amazed by the giraffe in the lobby* » (Keenan et al., 2008 p.296). Les images parmi lesquelles il faut choisir sa réponse, montrent des scènes qui se réfèrent à des mots distracteurs comme par exemple « *parents* » au lieu de « *patients* » et « *graffiti* » au lieu de « *giraffe* ». Ayant correctement identifié les mots « *patients* » ou « *giraffe* » les enfants peuvent sans problème choisir la bonne image. La situation est très similaire avec le *WJPC*, car dans les phrases à compléter il y a toujours un ou deux mots qui servent d'indice pour trouver la bonne réponse. Par exemple (p.296) « *I thought that the painting was enormous. I did not however feel like arguing about the -----* ». La ré-

ponse correcte est le mot « *size* » le mot « *enormous* » dans la phrase précédente est le seul indice.

Parmi les caractéristiques spécifiques des tests de compréhension textuelle ayant une influence sur la validité de la *SVRT*, il faut aussi penser au genre textuel. En effet Diakidoy et al. (2005) ont constaté une relation plus forte entre la compréhension orale (CO) et la lecture de texte (L) pour des textes narratifs que pour les textes expositifs. Des différences dans les méthodes d'évaluation de la variable dépendante pourront ainsi être à l'origine de certains résultats hétérogènes concernant la *SVRT*.

Notre deuxième question de recherche concerne donc l'impact de l'opérationnalisation de la compréhension de texte sur la validité de notre modèle additif de la *SVRT*. Les hypothèses que nous souhaitons vérifier sont les suivantes :

- H2a : Nous supposons que le pouvoir de prédiction et d'explication de notre modèle de la *SVRT* varie suivant l'opérationnalisation des compétences en lecture.
- H2b : Nous supposons que le modèle de la *SVRT* explique mieux les compétences en lecture opérationnalisées par le test Salzburger Lesescreening ou ELFE demandant une lecture rapide de textes très courts, que les compétences évalués par des textes complexes proposés par le test de HarmoS-L1.

### 2.3 Lecture en L2

Comme nous l'avons vu au chapitre précédent, la « *Simple View of Reading Theory* » utilise la compréhension orale comme opérationnalisation des capacités de compréhension de langue. Elle soutient aussi que le niveau de compréhension orale d'un lecteur définit le niveau maximal qu'il peut atteindre en compréhension de texte lorsque ses compétences en décodage sont parfaites. C'est ainsi qu'il faut lire la formule de la *SVRT* ( $L=D \times CO$ ). Lorsque 0 représente une incompétence totale et 1 représente la compétence maximale on obtient pour  $D=1$  ;  $L=CO$ .

L'idée d'une asymétrie dans le développement de l'impact des deux facteurs D et CO (voir chapitre précédent) est fondée sur l'observation d'un niveau déjà assez élevé en compréhension orale par rapport à des compétences en décodage plus faibles et plus hétérogènes en début de scolarisation. Cela paraît juste pour le groupe des enfants dont la langue de scolarisation est identique à leur première langue (L1). Dans le cas des en-

fants plurilingues, apprenant la lecture dans une langue qui n'est pas leur première langue (L1)<sup>22</sup>, la situation peut être différente (Kirby & Savage, 2008). Notre deuxième question de recherche concerne la validité de notre modèle additif de la *SVRT* dans un échantillon d'élèves plurilingues par rapport à un échantillon d'élèves monolingues.

Lorsque des enfants de langue première autre que l'allemand, ont eu peu de contact avec cette langue qui sera celle de leur scolarisation, ils risquent de commencer l'apprentissage de la lecture avec des compétences en compréhension orale aussi faibles et hétérogènes qu'en décodage. Le développement de l'influence des deux composantes D et CO peuvent ainsi différer de la trajectoire asymétrique supposée. Nous formulons donc deux hypothèses à l'égard de cette question :

- H3a : Nous supposons qu'aussi bien dans un échantillon d'élèves plurilingues que monolingues le décodage (D) et la compréhension orale (CO) ont une force explicative significative pour les compétences de lecture (L).
- H3b : Nous supposons cependant, que la compréhension orale (CO) ait plus de force explicative pour les compétences de lecture (L) chez les élèves plurilingues que monolingues.

Bien que Hoover et Gough aient publié en 1990 leur étude de validation du modèle de la *SVRT* basée sur un échantillon de bilingues anglais - espagnol, la question du développement asymétrique ne faisait pas partie de leur travail. La question de l'impact du plurilinguisme sur la validité de la *SVRT* n'ayant pas été étudiée de façon systématique à notre connaissance, nous souhaitons pallier ce manque par le présent travail.

## 2.4 Le QI un troisième facteur potentiel?

Comme nous l'avons vu au chapitre précédent, de nombreuses études ont confirmé que le décodage et la compréhension orale ont une influence significative sur les compétences en compréhension de texte et expliquent une grande partie de sa variance. Les quotas de variance expliquée par ces deux facteurs varient selon les études entre 59% à 80%. Toutefois ils n'expliquent pas toute la variance, indiquant ainsi que d'autres facteurs jouent un rôle dans l'explication et la prédiction des compétences en lecture, (Catts et al., 2005; Dreyer & Katz, 1992; Hoover & Gough, 1990). Divers facteurs ont ainsi été

---

<sup>22</sup>Nous comprenons ici par le terme « première langue » (L1) la langue qu'un enfant a appris en premier, qu'il utilise dans son environnement familial et qui est souvent aussi désignée par le terme « langue maternelle ». Un enfant peut avoir plusieurs langues premières.

proposés afin d'améliorer la *SVRT* et augmenter la part de variance expliquée. Parmi ces facteurs dont l'apport potentiel a été testé, on peut retenir :

- 1) la rapidité d'exécution des processus cognitif (*processing speed*) opérationnalisée par la rapidité de dire des mots (*naming speed*) ou des lettres (*naming letters*) (Georgiou et al., 2009; Johnston & Kirby, 2006; Joshi & Aaron, 2000) ou encore la rapidité dans la perception (*perceptual speed*) (Tiu, Thompson, & Lewis, 2003),
- 2) la mémoire de travail; (J. Oakhill & Yuill, 1996; Seigneuric & Ehrlich, 2005),
- 3) les processus de contrôle de l'attention (Conners, 2008),
- 4) la fluidité de la lecture (Adlof et al., 2006; Janet Tilstra et al., 2006). S'appuyant sur l'idée d'une capacité de traitement d'information limitée, de nombreux chercheurs voient dans la fluidité de la lecture un facteur important qui manque à la *SVRT*. En effet la plupart des études concernant la *SVRT* opérationnalisent le décodage uniquement comme l'identification correcte des mots ou des pseudo-mots. Or selon la théorie très influent de LaBerge et Samuels un décodage correct mais trop lent, n'est pas en mesure de soutenir le processus de la lecture (LaBerge & Samuels, 1974). Le décodage doit être automatique, c'est à dire correct et rapide / fluide, afin que les mots décodés restent pendant un certain temps ensemble dans la mémoire de travail et que le lecteur puisse identifier leurs relations syntaxiques et sémantiques.
- 5) l'intelligence (QI). La nature de la relation entre les compétences en lecture et les performances intellectuelles a provoqué au cours des 20 dernières années de vifs débats. Par exemple Jensen postule en 1980 que «la compréhension de texte dépend avant tout des capacités intellectuelles générales (« ...*reading comprehension is largely a matter of g* » Jensen 1980, p. 325). Stanovich, Cunningham et Feeman (1984) réfutent cette idée en montrant que l'apport du QI dans l'explication des variances en compréhension de texte est moindre que celui de la rapidité du Décodage, de la Compréhension Orale et de la conscience phonologique. La critique de Stanovich, Cunningham et Feeman ainsi que d'autres auteurs (Aaron et al., 2008; Savage, 2001) à l'égard de l'utilisation du QI n'est pas seulement fondée sur le pouvoir prédictif insatisfaisant du QI.

Ces auteurs soutiennent qu'une notion globale comme l'intelligence offre peu de pistes précises pour identifier les origines des déficiences en lecture et par voie de conséquence pour trouver des méthodes de remédiation adéquates. Malgré ces critiques, le

niveau des compétences en lecture et le niveau des capacités intellectuelles globales sont encore souvent perçus comme étant fortement liés. Ainsi l'identification des lecteurs en difficulté, supposés être atteints de dyslexie, se fait souvent à partir de l'observation d'une divergence entre le niveau intellectuel, voire les compétences en lecture attendues à la base du QI, et les compétences en lecture réelles. Cette approche se base sur l'idée traditionnelle selon laquelle le QI serait la meilleure opérationnalisation des processus de niveau supérieur sous-jacents à la lecture. Etant toujours d'actualité, cette approche de la divergence a été confrontée à des critiques sévères, dont nous allons mentionner seulement quelques points clés.

Ayant fait des progrès considérables dans l'identification des processus spécifiques impliqués dans la compréhension de texte, de nombreux chercheurs estiment que la compréhension orale est un indicateur plus probant que le QI dont les définitions sont multiples et qui véhicule une vision déterministe et peu encourageante (Savage, 2001).

Sur le plan empirique, les corrélations entre les compétences en lecture et le QI ne sont que de taille modeste chez les jeunes lecteurs débutants. Elles augmentent cependant avec l'âge des lecteurs. Stanovich et al. (1984) ont passé en revue plusieurs dizaines d'études réalisées entre 1955 et 1982. Ils concluent que chez des jeunes lecteurs (6 à 9 ans) la corrélation est d'environ 0.3-0.5, chez les lecteurs au milieu de leur scolarité (9 à 14 ans) elle est de .45-.65, et pour les lecteurs adultes (> 14 ans) elle peut aller de 0.55 - 0.82. Il semble que la relation entre les compétences en lecture et le QI est plus faible dans le groupe des lecteurs jeunes pour lesquels la nécessité d'un diagnostic correct et précoce d'éventuelles déficiences en lecture est pourtant plus grande.

Un autre moyen de tester la fiabilité du QI comme critère de référence pour un diagnostic de déficiences en lecture consiste à distinguer deux groupes de lecteurs faibles : dont un sans divergence entre le QI et les compétences en lecture (faible en lecture, faible QI) et un deuxième groupe avec une divergence (QI normal, faible en lecture). Comparant ces deux groupes, il s'avère qu'il n'y a pas de différence majeure entre les deux groupes : ils ont le même niveau de faiblesse en décodage, ils font les mêmes types d'erreurs (Frankel-Tal & Siegal, 1996 cité dans Savage 2001) et les deux groupes bénéficient des mêmes mesures d'intervention et de remédiation.

L'utilisation du QI semble alors problématique à deux égards. Premièrement, la présence ou non d'une divergence entre le faible niveau en lecture et le QI ne joue aucun rôle, ce qui rend inutile une telle distinction. Deuxièmement, le groupe des lecteurs fai-

bles sans divergence n'est pas reconnu comme ayant besoin d'une remédiation adaptée suivant la logique « ils sont mauvais en lecture c'est normal puisqu'ils ne sont pas très intelligents » (Savage 2001).

Des analyses de régressions multiples ont également été employées pour vérifier la valeur prédictive du QI pour les compétences en lecture. Les résultats obtenus sont hétérogènes. Stanovich et al (1984) rapporte une explication de la variance de la compréhension de texte par le QI (au delà du Décodage, de la Compréhension Orale et de la conscience phonologique) qui est faible pour les groupes de 6 ans et de 11 ans (9.1% / 0.5%). Pour le groupe des 8 / 9 ans l'explication de variance supplémentaire est de 36.9%. Pour les auteurs une explication à cette exception observée dans ce dernier groupe «n'est pas immédiatement disponible » (p. 296, traduit par N. Dittmann-Domenichini).

D'autres travaux (Siegel, 1989) concluent que l'apport supplémentaire du QI est faible lorsqu'il est ajouté après le Décodage. Savage (2001) rapporte que la Compréhension Orale est un meilleur prédicteur que le QI dans son étude auprès d'un échantillon d'adolescents très restreint (N=14).

A l'inverse Tiu, Thompson et Lewis (2003) demandent que le QI soit intégré en tant que troisième facteur dans la formule de la « *Simple View of Reading Theory* ». Dans leur échantillon de lecteurs de 11 ans, ils obtenaient une explication de la variance significative grâce au QI après que le décodage, la compréhension orale et la vitesse de traitement aient été pris en compte.

Kershaw et Schatschneider (2010) obtiennent également une contribution significative du QI, après la prise en compte des deux facteurs de la *SVRT*. Il est intéressant que le résultat inexplicé d'un apport très significatif du QI chez Stanovich et al. (1984) concerne également les lecteurs de 8 à 9ans de cette étude. Kershaw et Schatschneider essaient d'expliquer ces résultats par le fait que c'est à cet âge, que les enfants doivent commencer à "lire pour apprendre" plutôt que « d'apprendre à lire », ce qui placerait les capacités intellectuelles de raisonnement sur le devant de la scène.

A la vue des ces travaux, il paraît important de repenser l'approche d'un diagnostic de déficiences en lecture basé sur la divergence entre le QI et les compétences en lecture. Toutefois la question du pouvoir explicatif du QI dans la compréhension de texte reste ouverte. Dans le présent travail, qui permet une analyse de différents groupes d'âge et

de différentes méthodes d'évaluation de la compréhension de texte nous souhaitons clarifier la question de l'apport du QI dans un modèle additif de la *SVRT*.

- H4a : Nous supposons que c'est dans le groupe des lecteurs plus âgés (CM2 / 6<sup>ème</sup>) que le QI peut avoir un impact significatif, l'effet sur le groupe des plus jeunes lecteurs sera moindre.
- H4b : Nous supposons également que l'apport du QI varie fortement en fonction du test utilisé pour évaluer les compétences en lecture.
- H4c : Finalement nous supposons qu'introduit après le décodage, le QI n'est pas un meilleur prédicteur que la compréhension orale.

## 3 Méthode

### 3.1 Description de l'échantillon

Le présent travail de thèse repose sur l'analyse d'une partie spécifique des données collectées au sein du projet de recherche de la HEP de Berne sous la direction du Professeur Romano Müller<sup>23</sup>. Ce recueil de données a été réalisé en deux vagues afin de pouvoir effectuer une comparaison quasi-longitudinale. Le premier recueil de données a eu lieu entre octobre 2007 et mars 2008, la deuxième entre janvier et mars 2009. Le tableau 1 indique le nombre d'élèves et de classes dont nous avons pu collecter les données. Ne faisant pas partie de l'échantillon final de la présente thèse, l'échantillon des élèves de 4<sup>ème</sup>/3<sup>ème</sup> est signalé en gris.

Tableau 1 : Echantillon global lors des deux vagues de récolte de données

	CE1 / CE2		CM2 / 6ème		4ème / 3ème		Total		Total %	
	1ère vague	2ème vague	1ère vague	2ème vague	1ère vague	2ème vague	1ère vague	2ème vague	1ère vague	2ème vague
monolingues	80	77	87	85	180	155	347	317	30.7%	30.5%
plurilingues	295	276	207	194	283	252	785	722	69.3%	69.5%
<b>Total</b>	<b>375</b>	<b>353</b>	<b>294</b>	<b>279</b>	<b>463</b>	<b>407</b>	<b>1132</b>	<b>1039</b>	<b>100%</b>	<b>100%</b>
<b>Classes</b>	<b>18</b>		<b>15</b>		<b>23</b>		<b>56</b>		<b>56</b>	

Au total 56 classes se sont portées volontaires pour participer à l'étude de la HEP de Berne. Ces classes se situaient dans différents endroits de l'agglomération de Zürich (voir annexe 1). Dans toutes ces classes, le taux des élèves immigrés plurilingues était supérieur à 30 %. Le tableau 2 résume les indicateurs qui servent à une description sociodémographique de la population dans les 22 écoles et 12 communes dont les élèves sont issus. Parmi les indicateurs cités se trouvent : le taux des personnes sans emploi, le taux des étrangers, le taux des propriétaires de maison ainsi que l'Indice Social. Ce dernier est calculé chaque année pour toutes les communes sur la base d'une multitude d'indicateurs. Plus l'Indice Social d'une commune est élevé, plus elle est « défavorisée ». Pour mieux comprendre la signification des données, elles sont rapportées à la moyenne du Canton de Zürich (ZH).

<sup>23</sup> Voir site web « [www.sprachlichekompetenzen.phbern.ch](http://www.sprachlichekompetenzen.phbern.ch) »

Tableau 2 : Indicateurs sociodémographiques et Index Social

Commune	Ecole	Taux de chômage	Chômage rapporté à la moyenne du Canton de ZH	Taux d'étrangers	% Etrangers rapporté à la Ø du Canton de ZH	Taux des propriétaires de maisons	% propriétaires rapporté à la Ø du Canton de ZH	Indice Social 2008	Indice Social '08 rapporté à la Ø du Canton de ZH
Adliswil	Kronenwiese	2.72	0.09	25.88	3.36	11.07	-6.61	113.89	2.28
Bülach	Böswisli	3.07	0.44	22.27	-0.25	14.42	-3.25	112.28	0.67
Bülach	Lindenhof	3.07	0.44	22.27	-0.25	14.42	-3.25	112.28	0.67
Dietikon	Steinmürli	4.20	1.57	39.74	17.23	8.62	-9.05	118.08	6.47
Dietikon	Wolfsmatt	4.20	1.57	39.74	17.23	8.62	-9.05	118.08	6.47
Dietikon	Zentral A	4.20	1.57	39.74	17.23	8.62	-9.05	118.08	6.47
Dietikon	Zentral C	4.20	1.57	39.74	17.23	8.62	-9.05	118.08	6.47
Horgen	Waldegg	2.66	0.03	26.84	4.32	15.18	-2.49	113.43	1.82
Kloten	Nägelimoss Oberstufe	3.12	0.49	26.18	3.67	8.60	-9.07	114.19	2.58
Regensdorf	Chritzächer	1.48	-1.15	16.17	-6.34	15.69	-1.99	106.47	-5.14
Regensdorf	Ruggenacher I	3.39	0.76	30.96	8.45	14.52	-3.16	115.56	3.95
Regensdorf	Ruggenacher II	3.30	0.67	27.53	5.02	15.79	-1.89	114.59	2.98
Regensdorf	Ruggenacher III	3.39	0.76	30.96	8.45	14.52	-3.16	115.56	3.95
Schlieren	Kalktarren	4.23	1.60	42.02	19.50	6.53	-11.15	118.39	6.78
Schlieren	Zelgli	4.23	1.60	42.02	19.50	6.53	-11.15	118.00	6.39
Urdorf	Moosmatt	1.92	-0.71	18.57	-3.95	12.52	-5.16	111.63	0.02
Volketswil	Lindenbüel	3.30	0.67	21.96	-0.56	21.44	3.77	112.73	1.12
Wallisellen	Bürgli Nord	3.21	0.58	25.29	2.77	16.66	-1.01	113.42	1.81
Wetzikon-Seegräben	Walenbach Oberstufe	2.84	0.21	21.81	-0.70	21.81	4.13	112.02	0.41
Winterthur-Mattenbach	Mattenbach Oberstufe	3.30	0.67	24.71	2.19	12.12	-5.55	113.57	1.96
Winterthur-Töss	Eichliacker	3.95	1.32	32.70	10.18	16.45	-1.22	116.00	4.39
Winterthur-Töss	Rebwiesen	3.95	1.32	32.70	10.18	16.45	-1.22	116.00	4.39

Ainsi pour le taux de chômage, les écoles Chrüzacher et Moosmatt se trouvent à des endroits où le taux de chômage est légèrement inférieur à la moyenne du Canton de Zürich. Pour les autres écoles, il correspond à peu près à la moyenne du Canton de Zürich. Les divergences sont plus grandes en ce qui concerne le taux d'étrangers. Dans les communes de Dietikon, Schlieren et Winterhur Töss, le taux d'étrangers est de 10 à 20 % plus élevé que la moyenne du Canton qui est de 22.5 % (chiffre 2008). Quant au taux des propriétaires de maisons, il est dans toutes ces communes inférieur à la moyenne (17.6 % de la population possède une maison). L'Indice Social résume la situation des communes. Pour 4 des 12 communes l'Indice Social indique une situation similaire (Bülach, Urdorf, Wetzikon) voire meilleure (Regensdorf / Chrüzacher) que pour la moyenne du Canton de Zürich. Dans les 8 communes restantes, la situation globale est un peu moins favorable, sans pour autant s'écarter beaucoup de la moyenne. L'échantillon semble donc assez fidèlement représenter les caractéristiques sociodémographiques du Canton de Zürich.

### 3.2 Sélection de l'échantillon pour le présent travail de thèse

Afin d'apporter aux questions de recherche des réponses reposant sur des analyses empiriques fiables, nous avons sélectionné pour la présente étude, le groupe des élèves en classe de CE1/CE2 et de CM2/6<sup>ème</sup>. Nous allons maintenant décrire de façon plus détaillée cet échantillon sélectionné.

Le tableau 3 montre que la proportion de garçons et de filles est équilibrée dans les deux groupes d'âges. Le taux des enfants se déclarant bilingue ou trilingue<sup>24</sup> est à peu près 10% plus grand dans le groupe des CE1/CE2. Le QI global des deux groupes se trouve dans la moyenne. Le nombre de livres à la maison nous a servi d'indicateur du milieu socioculturel dont l'enfant est issu<sup>25</sup>. Les données présentées dans le tableau 4 reposent sur les réponses données par les parents. Pour chaque famille, c'est la catégorie la plus élevée entre le père et la mère, qui a été prise en compte.

---

<sup>24</sup> Les élèves étaient priés de cocher dans un questionnaire les langues qu'ils maîtrisaient du moins à l'oral et qu'ils avaient apprises dans leur milieu familial et non par le biais de l'école (langues étrangères comme l'anglais ou le français).

<sup>25</sup> Le groupe des enfants en CE1/CE2 devait pour cela compter le nombre de livres à la maison ; pour le groupe des enfants plus grands nous avons montré une rangée de livres type dans une étagère contenant environ 50 livres. Les élèves devaient ensuite estimer le nombre de livres qui se trouvaient chez eux.

Tableau 3 Caractéristiques de l'échantillon sélectionné

Sexe	CE1/CE2		CM2/6ème		
	Garçons	N=186	48.90%	N=157	48.50%
Filles	N=194	51.10%	N=157	51.50%	
Age en années	Moyenne (M)	7.88		11.09	
	Ecart-type (SD)	0.56		0.59	
	Amplitude	4		3	
Groupes de (Pluri)lingues	Monolingues	N=81	21.30%	N=94	30.80%
	Bilingues	N=260	68.40%	N=173	56.70%
	Trilingues	N=39	10.30%	N=38	12.50%
Nombre de livres à la maison	< 50 livres	N=123	37.30%	N=66	23.80%
	50 à 150 livres	N=110	33.30%	N=111	40.10%
	> 150 livres	N=97	29.40%	N=100	36.10%
Catégorie socio-professionnelle	Moyenne (M)	4.59		4.95	
	Ecart-type (SD)	1.97		2.2	
	Amplitude	9		9	
QI	Moyenne (M)	102.06		105	
	Ecart-type (SD)	14.4		13.89	
	Amplitude	82		75	

Tableau 4 : Catégorie socioprofessionnelle la plus élevée des parents

	CE1/CE2		CM2/6ème	
	N	%	N	%
Homme / Femme au foyer	9	2.6	6	2.1
Employé(e) sans formation	37	10.9	29	10.4
Employé(e) avec ou sans formation mais avec de bonnes qualifications	71	20.9	53	18.9
Chef d'équipe, Chef de petite entreprise avec 0 à 5 mains-d'oeuvres	62	18.2	39	13.9
Employé(e) bien qualifié responsable d'un domaine spécifique. Chef / Propriétaire d'une PME (6 à 20 employés)	58	17.1	52	18.6
Cadre (de base) dans différents domaines du service (par ex. bureau, hôpital, police)	42	12.4	28	10.0
Propriétaire d'une PME (250 employés); Cadre moyen; Employés avec tâches de direction dans une grande entreprise (> 250 employés); Enseignants niveau primaire ou secondaire	29	8.5	39	13.9
Cadre supérieur / Enseignants de lycée ou dans les formations supérieures (lycée, université, IUFM), Etudiants d'université	19	5.6	12	4.3
Professions libérales avec formations supérieures (par ex. Médecins, Avocats, Architectes)	11	3.2	11	3.9
Cadre de niveau le plus élevé. Très hautes tâches executives (grandes entreprises ou l'administration); membres du gouvernement	2	.6	11	3.9
Total	340	100.0	280	100.0

Les élèves plurilingues constituent environ deux tiers de notre échantillon. Ils sont pour la plupart nés en Suisse et issus de la 2ème génération d'immigration (CE1/CE2 68.2% nés en Suisse, CM2/6ème 74,8% nés en Suisse). Ainsi la grande majorité des élèves plurilingues a suivi toute sa scolarité en Suisse. Les élèves n'ayant rejoint l'enseignement en Suisse alémanique que récemment, ont été exclus des analyses statistiques ultérieures. Ceci concernait cinq élèves en CM2 qui étaient arrivés depuis seulement un ou

deux ans en Suisse alémanique, et trois élèves en CE1 qui étaient arrivés depuis un an. Pour la présente étude, ce n'est pas la nationalité mais la situation linguistique (monolingue versus plurilingue) qui est une variable indépendante importante dans les analyses statistiques présentées au chapitre 4. Toutefois, on rencontre une multitude de constellations, résumées sous le terme de plurilinguisme ou de bilinguisme. Le tableau 5 et la figure 3 montrent les différentes constellations de bilinguisme dans notre échantillon.

Tableau 5: Proportions des langues d'origine dans l'échantillon sélectionné

	CE1/CE2		CM2/6ème		
	N	%	N	%	
Monolingue allemand (suisse-allemand)	70	20.6	81	29.1	
Bilingue italien-allemand	40	11.8	26	9.4	
Bilingue espagnol-allemand	8	2.4	13	4.7	
Bilingue portugais-allemand	25	7.4	15	5.4	
Bilingue français-allemand	2	.6	1	.4	
Bilingue turc-allemand	37	10.9	23	8.3	
Bilingue langue slave des Balkans-allemand	50	14.7	35	12.6	
Bilingue albanais-allemand	64	18.8	46	16.5	
Bilingue anglais/ ou lunge scandinave / ou hollandais-allemand	6	1.8	5	1.8	
Bilingue langue slave est-européen-allemand	6	1.8	8	2.9	
Bilingue langue asiatique-allemand	25	7.4	23	8.3	
Bilingue grec-allemand	5	1.5	0	.0	
Bilingue langue africaine-allemand	2	.6	2	.7	
	N	340	100.0	278	100.0

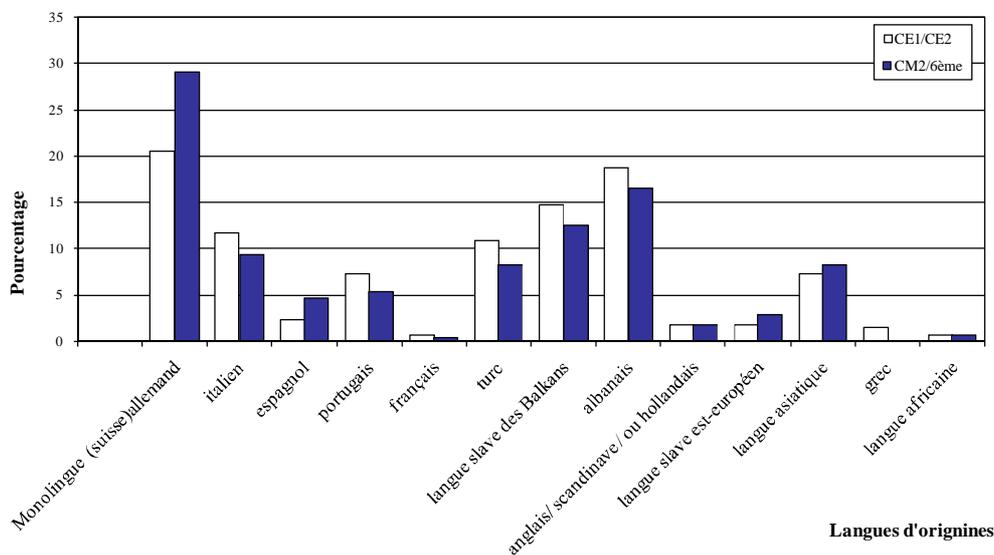


Figure 3: Répartition des langues d'origines dans l'échantillon sélectionné

Afin d'inclure les constellations relatives au bilinguisme et au trilinguisme (10 à 12 % des élèves sont trilingues) nous parlons en général d'élèves plurilingues.

A défaut d'avoir pu tester la maîtrise et les compétences en Langue d'origine, nous avons demandé aux élèves plurilingues de nous renseigner sur leurs préférences de lan-

gue dans différentes situations de communication (voir figures 4a et 4b). On retrouve dans les préférences de langues de notre échantillon le phénomène connu d'utilisation privilégiée de la langue d'origine dans la communication avec les parents. A l'inverse, la langue du pays d'immigration est privilégiée lorsqu'il s'agit de parler avec les camarades et amis ayant la même Langue d'Origine, mais aussi avec les frères et sœurs.

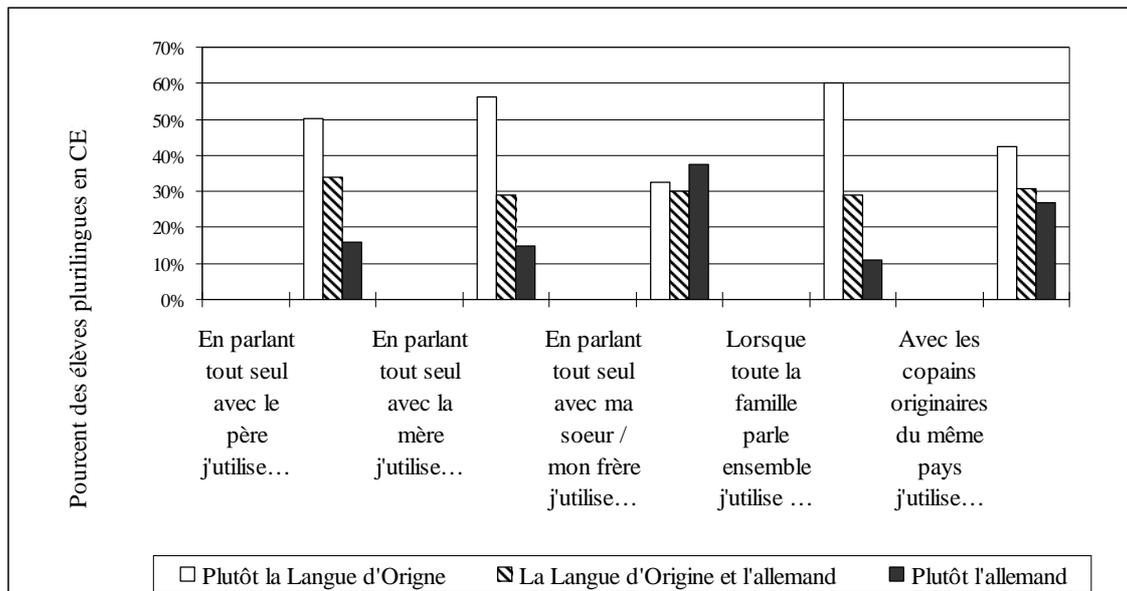


Figure 4a Préférences de langue en CE1

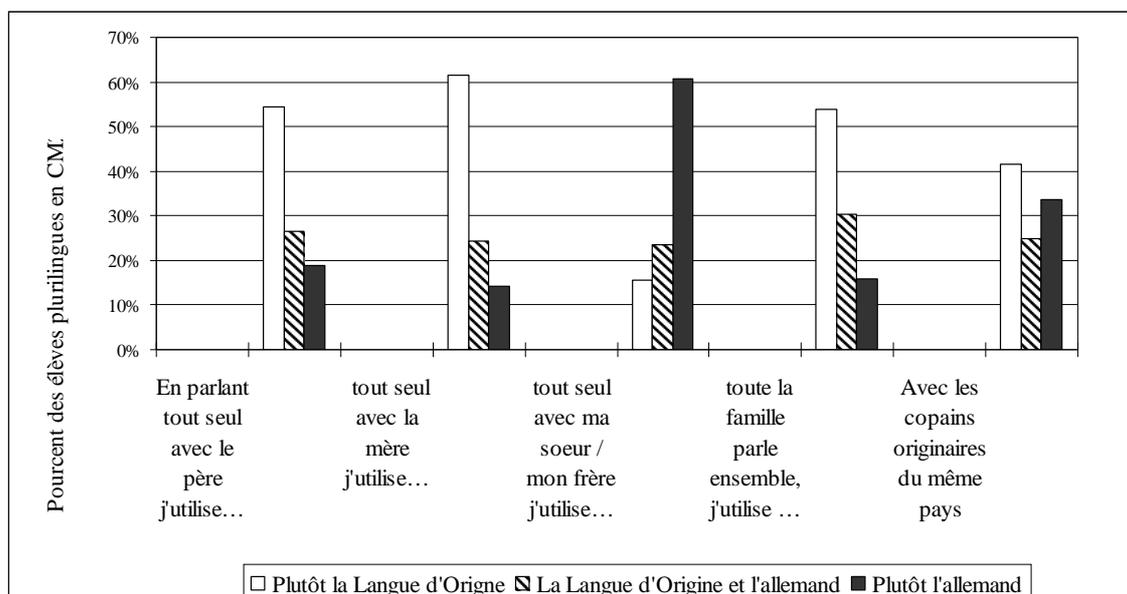


Figure 4b Préférences de langue en CM2

Notre échantillon d'élèves plurilingues peut donc être décrit de façon globale comme a) étant issu de la deuxième génération d'immigration, b) ayant avec la Langue d'Origine un contact surtout dans le milieu familial et c) ayant un contact régulier avec l'allemand,

au moins par le biais de l'école suisse, qu'ils fréquentent pour la plupart depuis le début de leur scolarisation.

Afin de vérifier la validité de la « *Simple View of Reading Theory* » au sein de notre échantillon, nous avons séparé l'échantillon des élèves de CE1/CE2 en deux groupes : les monolingues et les plurilingues. Nous avons ensuite procédé à des analyses de régressions multiples multi-niveaux avec ces deux sous-échantillons. Avant de présenter les résultats de ces analyses, nous souhaitons mettre en évidence les différences éventuelles entre les deux groupes.

Tableau 6 Comparaisons des groupes monolingues versus plurilingues

Variable	Groupe	N	Moyenne (M)	Ecart-type	Comparaison	Somme des carrés	df	Moyenne des carrés	F	Significativité
Nombre de livres à la maison	monolingue	64	3.8	1.5	Inter-groupes	140.8	1.0	140.8	69.1	0.00 ***
	plurilingue	232	2.2	1.4	Intra-groupes	598.9	294.0	2.0		
Temps de lecture extra-scolaire	monolingue	64	2.8	1.2	Inter-groupes	0.9	1.0	0.9	0.7	0.39 n.s.
	plurilingue	232	2.6	1.1	Intra-groupes	373.6	294.0	1.3		
QI	monolingue	64	106.2	14.5	Inter-groupes	1151.9	1.0	1151.9	5.5	0.02 *
	plurilingue	232	101.4	14.4	Intra-groupes	61043.7	294.0	207.6		
Catégorie socio-professionnelle	monolingue	64	6.0	1.8	Inter-groupes	152.1	1.0	152.1	45.5	0.00 ***
	plurilingue	232	4.3	1.8	Intra-groupes	982.5	294.0	3.3		

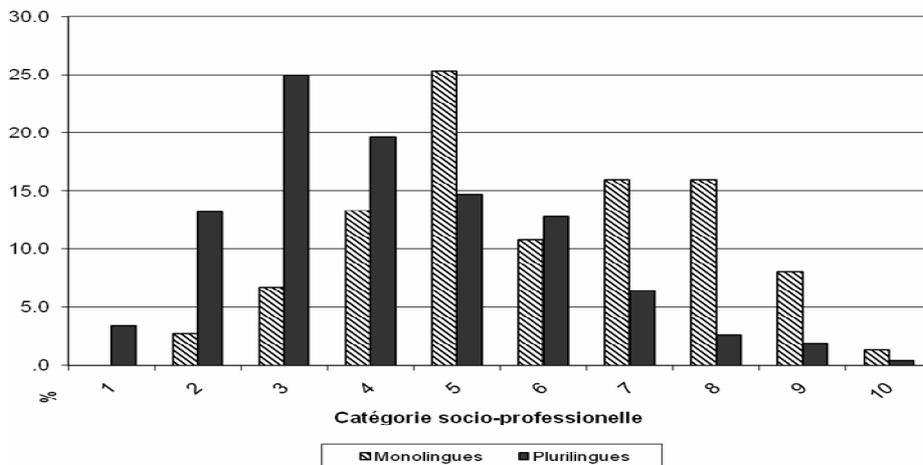


Figure 5 Distribution des catégories socioprofessionnelles

Le tableau 6 ci-dessus nous montre que nos deux sous-échantillons se distinguent de manière significative en ce qui concerne le nombre de livres à la maison et la catégorie socio-professionnelle des parents. La moyenne du QI atteint dans les deux groupes marque un avantage d'environ 5 points en faveur des monolingues. Cette différence est encore significative avec une probabilité d'erreur acceptée de 5 %. Toutefois le QI moyen des plurilingues étant légèrement au dessus de 100 est normal.

La distribution des catégories socioprofessionnelles (figure 5) montre que les parents d'élèves plurilingues occupent plus souvent des positions de femme au foyer,

d'employé(e)s avec ou sans formation ou de chef d'équipe de manutention. Les parents d'élèves monolingues occupent plus souvent, et ce de manière significative, des positions de dirigeant, voire de chef d'entreprise de PME, de cadre ou de membre de professions libérales.

Tableau 7 Catégories socioprofessionnelles des parents d'élèves

Catégorie socio-professionnelle		Monolingues CE1		Plurilingues CE1	
		N	%	N	%
1	Homme / Femme au foyer	0	.0	9	3.4
2	Employé(e) sans formation	2	2.7	35	13.2
3	Employé(e)c avec ou sans formation mais avec des bonnes qualifications	5	6.7	66	24.9
4	Chef d'équipe, Chef de petite entreprise avec 0 à 5 mains d'oeuvres	10	13.3	52	19.6
5	Employé(e) bien qualifié responsable d'un domaine spécifique. Chef / Propriétaire d'une PME (6 à 20 employés)	19	25.3	39	14.7
6	Cadre (de base) dans différents domaines du service (par ex. bureau, hôpital, police)	8	10.7	34	12.8
7	Propriétaire d'une PME (250 employés); Cadre moyen; Employés avec tâches dirigeantes dans une grande entreprise (> 250 employés); Enseignants	12	16.0	17	6.4
8	Cadre supérieur/ Enseignants de lycée ou dans les formations supérieures (lycée, université, IUFM) Etudiants d'université	12	16.0	7	2.6
9	Professions libérales avec formations supérieures (par ex. Médecins, Avocats, Architecte)	6	8.0	5	1.9
10	Cadre de niveau le plus élevé. Très hautes tâches executives (grandes entreprises ou l'administration); membres du gouvernement	1	1.3	1	.4
	Total	75	100.0	265	100.0

Tableau 8 Comparaison du nombre de livres à la maison

Nombre de livres à la maison	% Monolingues		% Plurilingues	
	N	%	N	%
< 50 livres	10	10.1	44	44.4
50 à 100 livres	14	14.5	27	27.2
100 à 150 livres	7	7.2	9	9.2
150 à 200 livres	10	10.1	5	5.7
> 200 livres	58	58	13	13.4

Tableau 9 Temps consacré à la lecture en allemand en dehors de l'école.

Temps de lecture extra-scolaire en allemand	% Monolingues		% Plurilingues	
	jama is	10.3	10.9	
< 30 minutes	44.1	51.3		
entre 30 min et 1h	23.5	18.5		
entre 1h et 2h	5.9	9.4		
> 2h	16.2	9.8		

Le nombre de livres à la maison est un indicateur du milieu socio-économique et éducatif des élèves. Comme dans la plupart des enquêtes internationales sur les résultats scolaires, nous l'utilisant pour avoir une idée des ressources culturelles, socio-économiques et éducatifs dans les différentes familles.

Malgré un nombre de livres en moyenne bien plus élevé à la maison, les élèves monolingues ne consacrent pas plus de temps à la lecture en allemand à la maison que les élèves plurilingues (voir tableau 8 et 9). Lorsque nous regardons d'éventuelles différences dans le niveau de compétence en lecture entre des élèves plurilingues et les élèves monolingues nous devons tenir compte des différences présentées ci-dessus entre ces deux groupes d'élèves.

*Résumé:*

*Nous avons procédé entre octobre 2007 et mars 2009 à deux recueils de données dans le canton de Zürich, recrutant ainsi 18 classes d'élèves en CE1 et 15 classes d'élèves en CM2.*

*En CE1 la moyenne d'âge des élèves était de 8 ans, en CM2 elle était de 11 ans. Environ 20 % à 30% des élèves étaient monolingues tandis que 70 à 80 % étaient plurilingues. Les langues d'origines des enfants plurilingues étaient majoritairement l'albanais, les langues slaves, le turc ou l'italien. Le taux des filles et des garçons était équilibré dans les deux groupes d'âges.*

*Du point de vue socio-économique notre échantillon semble également être représentatif. Toutes les catégories socioprofessionnelles étaient représentées, mais le gros de l'échantillon était composé d'enfants venant du milieu des catégories socioprofessionnelles moyennes.*

*Entre les élèves plurilingues et monolingues en CE, il existe cependant des différences significatives en faveur des monolingues en ce qui concerne le nombre de livres à la maison, la catégorie socioprofessionnelle des parents et le QI. La différence du QI était significative mais faible avec un écart de 5 points entre les deux groupes. Le QI moyen dans notre groupe d'élèves plurilingues était légèrement au-dessus de 100 mais cependant normal. L'impact du plurilinguisme et du QI sur les compétences en lecture est présenté de façon détaillée, aux chapitres 4.4 et 4.5.*

### 3.3 Déroulement du recueil des données

La majorité des données concernant les variables indépendantes a été recueillie entre octobre 2007 et mars 2008 par les assistants du projet de recherche de la HEP de Berne, dont l'auteur de la présente thèse faisait partie.

Tableau 10 Résumé du recueil des données

Instruments	Recueilli par	Quand	Versions	Temps requis
Questionnaire pour les élèves	assistants de recherche	janvier-mars 2008	2 versions: 1) CE1 / CE2 2) CM2 / 6ème	50 minutes
Questionnaire pour les parents	parents	juillet-août 2008	1 version allemande traduit en 10 langues	60 minutes
CFT 20 (test de QI)	assistants de recherche	janvier-mars 2008	versions A/B	60 minutes
Tests de lecture (décodage et compréhension de texte)	enseignants	1) janvier-mars 2008 2) janvier-mars 2009	1) HarmoS test (A / B) 2) ELFE (A / B) 3) Salzburger Lesescreening (SLS) (A / B)	1) 50 minutes 2) 20 minutes 3) 10 minutes
Tests de compréhension orale (HarmoS)	enseignants	1) janvier-mars 2008 2) janvier-mars 2009	A / B	50 minutes

Deux rendez-vous ont été fixés avec les enseignants ; le premier pour effectuer le test de QI en classe pendant environ 60 minutes et en présence de l'enseignant ; le deuxième pour présenter aux élèves le questionnaire qu'ils devaient remplir par la suite. Ce questionnaire portait sur les données sociodémographiques, les attitudes et la motivation envers l'école, le concept de soi ainsi que, pour les plurilingues, sur l'utilisation des langues d'origines. Afin d'exclure des difficultés de compréhension, chaque phrase a été lue à haute voix et tous les élèves remplissaient le questionnaire au même rythme.

En vue d'éviter que la vie scolaire ne soit trop perturbée par la réalisation de notre étude, les tests de lecture et de compréhension orale ont été réalisés par les enseignants. Afin de s'assurer que les enseignants étaient en mesure de recueillir des données valides, nous avons créé un document d'accompagnement très détaillé. De plus, tous les enseignants avaient participé à une réunion avec les membres de l'équipe de recherche lors de laquelle ils recevaient les consignes importantes. Les enseignants étaient par exemple informés qu'il fallait respecter les limites de temps et suivre les consignes des

tests à la lettre ; qu'il ne fallait pas préparer les élèves aux tests, qu'il ne fallait pas faire passer plus qu'un seul test par jour.

Les tests ont ensuite été renvoyés à l'équipe de recherche qui procédait au nettoyage et à l'analyse des données. Pour remercier les enseignants de leur aide, nous leur avons transmis une description des performances de leur classe concernant les tests de langue. En revanche, aucune information relative aux résultats du test de QI et aux questionnaires des élèves n'a été donnée.

Les informations concernant le statut social avaient été recueillies par le biais d'un questionnaire destiné aux parents. Les enseignants ont distribué ce questionnaire aux élèves afin qu'ils le fassent remplir par leurs parents. Après l'avoir rempli, les parents scellaient le questionnaire dans une enveloppe, et le rendaient ainsi aux enseignants qui par la suite les envoyaient à l'équipe de recherche. Ainsi avons-nous pu collecter les données d'environ 90% des parents d'élèves.

### 3.4 Traitement des données

L'incommensurable quantité de données recueillies à l'aide des instruments indiqués ci-dessus au sein du tableau 10, nous a obligé à trouver un procédé efficace pour numériser les données et calculer les scores obtenus par chaque élève dans chaque test. Pour nous aider, nous avons eu recours à des classes d'élèves volontaires (en 3<sup>ème</sup>) qui ne faisaient pas partie de notre étude. Ainsi en 2008 nous sommes allés pendant 4 demi-journées dans un collège à Täuffelen (CH)<sup>26</sup>. Nous y avons présenté les grandes lignes de notre projet avant de familiariser les élèves avec les tests de compréhension orale et de lecture ainsi que leurs grilles d'évaluation. Une matinée a été dédiée à entraîner les élèves à l'évaluation correcte des tests. Chaque classe était spécialisée dans un seul test, afin de rendre les évaluations aussi fiables que possible. Comme les tests de lecture et de compréhension orale comportaient un grand nombre de questions fermées (questions vrai/faux ou à choix multiple) cela posait peu de problèmes aux élèves. Pour l'évaluation des quelques questions ouvertes, qui se trouvaient dans certains tests, nous avons préparé des documents qui listaient les idées clés qui devaient être contenues dans une réponse correcte. De plus au sein des classes, une assistante du projet se tenait continuellement à disposition pour toutes les questions concernant l'évaluation. La figure 6 montre un exemple des fiches standardisées sur lesquelles les résultats de chaque élève étaient transcrits. Les réponses données aux questions à choix multiples étaient seulement copiées, les scores obtenus aux questions vrai / faux (cases blanches sur la fiche) et aux questions ouvertes (cases grises) étaient marqués à l'aide d'un crayon noir. Nous avons ainsi obtenu par la suite une fiche par test et par élève. Cette fiche comportait outre le nom de l'individu, un numéro d'identification pour chaque élève de notre échantillon.

Selon un principe aléatoire, l'équipe de projet sélectionnait des tests pour juger de la qualité des évaluations. Elle était de manière générale bonne, peu de corrections ont été nécessaires. En guise de récompense, les classes recevaient une petite somme d'argent pour leur caisse commune. La plupart des élèves étaient motivés et ont perçu ce travail comme une distraction bienvenue dans leur quotidien scolaire. Aucun élève n'a refusé de participer à cette tâche.

---

<sup>26</sup>Le même procédé a été reproduit en 2009 avec des élèves en 3<sup>ème</sup> d'un collège à Burgdorf (CH) pour les données de la deuxième vague.

Par la suite, les fiches, créées à l'aide d'un outil informatique (*Evasys*), ont été scannées et numérisées au centre informatique de l'université de Berne.

Code du n°d'identification de l'élève      Nom de l'élève

PHBern

1 1 1 2 2 1 4 1 0 5  
 Name  
 Daw

Rater-ID  
 Vorname  
 Gabriel

Klasse  
 14

HAR\_225 / HarmoS-Hören 5. Klasse      B

**Hinweise zum Ausfüllen:**  
 Markieren Sie so:      
 Korrektur: erfolgt über Ausradieren

Dieses Auswertungsblatt daher sehr wichtig, dass i  
 sauber schreibt und die Kästchen markiert. Benutzt dafür  
 nur den von uns ausgegebenen Bleistift.

Nom et version du test ist ehr

**Pribjat**

**Aufgabe 1**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 2**  
 Richtig ist:  
 niemand oder keiner  
 oder keine Menschen  
 wenn Antwort richtig= 1 Punkt  
 sonst gib 0 Punkte  
   
 0 1

**Aufgabe 3**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 4**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 5**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 6**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 7**  
 1   
 2   
 3   
 4

**Aufgabe 8**  
 1   
 2   
 3   
 4

**Aufgabe 9**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 10**  
 Je nach Qualität der  
 Antwort, gib 1 oder 0  
 Punkte (siehe Blatt zur  
 Auswertungshilfe).  
   
 0 1

**Ich fand den Hörtest**  
 1 2 3 4 5

**Schuluniform**

**Aufgabe 1**  
 Je nach Qualität der Antwort,  
 gib 1 oder 0 Punkte (siehe  
 Blatt zur Auswertungshilfe).  
   
 0 1

**Aufgabe 2**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 3**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 4**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 5**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 6**  
 1   
 2   
 3

**Aufgabe 7**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 8**  
 1   
 2   
 3   
 4   
 5

**Aufgabe 9**  
 Je nach Qualität der Antwort,  
 gib 1 oder 0 Punkte (siehe  
 Blatt zur Auswertungshilfe).  
   
 0 1

**Aufgabe 10**  
 1   
 2   
 3

**Ich fand den Hörtest**  
 1 2 3 4 5

Numéro de l'item au sein du test

Figure 6 Fiche standardisée servant à numériser les données recueillies

Après la numérisation, nous avons procédé au nettoyage des données. Nous avons exclu des procédures statistiques les élèves avec des données manquantes, les élèves présentant des défauts de l'ouïe, souffrant d'un retard ou d'une pathologie concernant les compétences en langues (par ex. dyslexie). Nous avons dû écarter une classe de CM2/6<sup>ème</sup> dont le test de lecture fluide (*Salzburger Lesescreening*) avait été réalisé sans respecter les limites de temps.

Contrairement aux tests de lecture « ELFE » et « Salzburger Lesescreening », les épreuves construites par le groupe de HarmoS-L1 pour la compréhension de texte et la compréhension orale ne sont pas standardisées. Elles ne proposent pas une standardisation du score obtenu par rapport à la norme d'un échantillon normatif. Pour vérifier la qualité des données obtenues par les épreuves de HarmoS-L1 et pour transformer ces données dans une unité de mesure qui satisfait les conditions d'une échelle à intervalle, nous avons utilisé le « modèle de Rasch ».

### 3.5 Caractéristiques et utilisation du « modèle de Rasch »

Le modèle dit « de Rasch » est l'approche mathématique la plus simple qui est utilisée dans le cadre de la théorie des réponses aux items (TRI ou IRT dans la littérature anglophone). Il a été développé dans la deuxième moitié du 20<sup>e</sup> siècle par le mathématicien Georg Rasch. Son objectif est de modéliser la relation fondamentale, postulée par cette théorie, entre le trait latent (compétence, capacité, aptitude, etc.) de l'individu et sa probabilité de réussir un item. Les modèles probabilistes comme le modèle de Rasch reposent sur le postulat que la probabilité d'un individu de fournir une réponse correcte à un item est déterminée essentiellement par deux facteurs:

- (a) d'une part, la compétence de l'individu. Elle est généralement qualifiée de trait latent, car elle n'est pas directement accessible à l'observation et l'évaluation.
- (b) d'autre part, les caractéristiques de l'item lui-même, en particulier sa difficulté ou son pouvoir de discrimination.

Ainsi lorsque la compétence d'un individu est égale à la difficulté de l'item, la probabilité de l'individu de réussir un item est de 0.5. La probabilité de réussite est de  $> 0.5$  lorsque la compétence est supérieure à la difficulté. Lorsque la compétence est inférieure à la difficulté, la probabilité de réussite est de  $< 0.5$ .

Cette relation est exprimée par une fonction mathématique (appelée fonction caractéristique de l'item) et elle est représentée par une courbe en forme de S (voir figure 7).

Probabilité de réussite

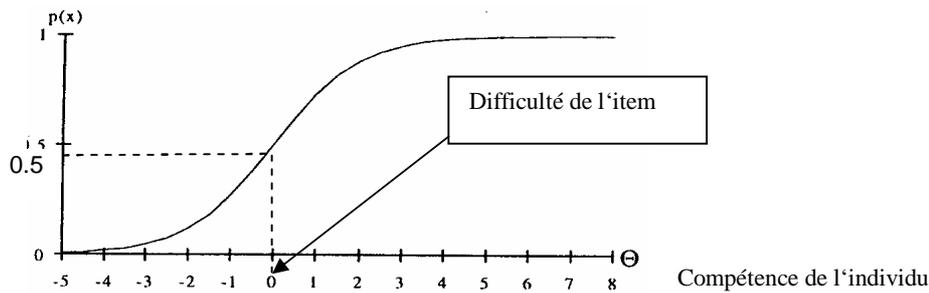


Figure 7 Fonction caractéristique de l'item

### 3.5.1 Estimation de paramètres avec le modèle de Rasch

Le « logit » est le logarithme naturel du rapport issu de la probabilité conditionnelle qu'un sujet réussisse un item, et de la probabilité conditionnelle que ce même sujet ne réussisse pas l'item.

$$\ln \frac{P(Y_i = 1 | p_U = u)}{P(Y_i = 0 | p_U = u)} = \ln P(Y_i = 1 | p_U = u) - \ln P(Y_i = 0 | p_U = u) \quad 27$$

On utilise le logarithme naturel de ce ratio afin d'éviter la restriction des valeurs dans l'intervalle  $]0 ; +\infty[$ . Ce logarithme, appelé le « logit » du sujet répondant relatif à l'item  $Y_i$ , peut maintenant varier dans l'intervalle  $]-\infty ; +\infty[$ .

Pour estimer les paramètres individuels avec la méthode du maximum de vraisemblance, le score individuel sera associé à la variable latente  $x$ , valeur pour laquelle le comportement de réussite du sujet sera le plus probable.

La différence entre deux paramètres d'items peut être calculée à partir des probabilités de réussite (inconditionnelles) estimables de manière empirique. Lorsqu'on fixe un paramètre d'item, tous les autres paramètres d'item peuvent être calculés sur la base de la différence avec le paramètre fixé.

### 3.5.2 Vérification de la validité du modèle de Rasch

Comme dans les modèles de la théorie classique des tests, on peut déduire du modèle de Rasch des propositions qui peuvent être vérifiées empiriquement. Une contrainte parti-

<sup>27</sup> Nous nous appuyons sur les formules décrites dans le manuel du programme d'analyse WINSTEPS (Linacre, 1991-2006) et le chapitre sur la théorie des réponses aux items (TRI) dans le livre «Messen und Testen » (Steyer & Eid, 2001).

culièrement exigeante du modèle de Rasch, suppose que tous les items d'un test ont le même pouvoir discriminatif (égal à 1). On postule par ailleurs que tous les items appartenant à un instrument de test utilisé, permettent d'appréhender une même caractéristique sous-jacente. Lorsque cette condition est satisfaite, on peut dire que les items et les variables sont « Rasch-homogènes ». Une épreuve étant construite exclusivement d'items « Rasch-homogènes » présente l'avantage que toutes les informations importantes relatives à la compétence ou à la caractéristique que nous voulons mesurer sont incluses dans un seul indice: la somme des points récoltés, puisque tous les items mesurent la même chose, ils ont le même pouvoir discriminatif. Une échelle classique n'étant pas « Rasch-homogène » nécessite de prendre aussi en considération les caractéristiques de la distribution des réponses.

Lorsqu'une de ces conséquences testables ne répond pas à une application empirique, il faut alors rejeter l'hypothèse de la validité du « modèle de Rasch ». Si l'on ne veut pas revenir à un autre modèle de mesure, on peut aussi changer les instruments de mesure de telle manière qu'ils soient compatibles avec le « modèle de Rasch ». On peut par exemple exclure les items du test qui ne sont pas compatibles avec le modèle.

### **3.5.3 Les avantages de l'utilisation du « modèle de Rasch »**

Les modèles probabilistes comme le « modèle de Rasch » permettent notamment de faire face à des problèmes auxquels la théorie classique de la mesure n'a pas toujours su apporter de réponses satisfaisantes. Par exemple, l'évaluation des propriétés techniques d'un item (comme le calcul des indices de difficulté ou de discrimination) fournit des résultats qui sont toujours relatifs à l'échantillon particulier auquel l'item a été soumis. De ce fait, un item jugé facile ou difficile au sein d'un échantillon d'individus, peut ne plus l'être (ou plus autant) s'il était soumis à un échantillon différent. Par rapport à ce genre de situation, la théorie des réponses aux items (TRI) s'efforce de donner une évaluation des propriétés de l'item qui soit indépendante d'un groupe particulier. En d'autres termes, il s'agit de parvenir à l'élaboration d'instruments de mesure dont les caractéristiques soient aussi peu influencées que possible par le groupe de référence.

Dans le cadre général qui vient d'être esquissé, l'objectif de l'utilisation du modèle de Rasch est double et simultané. Il s'agit, d'une part, d'estimer les caractéristiques des items (calcul des paramètres dits de difficulté, de discrimination et de « Rasch homogé-

néité ») et, d'autre part, d'estimer le niveau de l'individu par rapport au trait latent considéré.

### 3.5.4 Les analyses selon le « modèle de Rasch » dans la présente étude

Nous avons procédé à des analyses selon le modèle de Rasch en utilisant l'application WINSTEPS de J.M. Linacre (1991 / 2006). Cette application nous permet d'obtenir des mesures de Rasch sur la base des réponses de notre échantillon par rapport aux items des épreuves de HarmoS-L1. Ces réponses ont été d'abord enregistrées dans une fiche de SPSS (répertoire des données) et ensuite importées dans le programme WINSTEPS. Dans une fiche de contrôle nous avons spécifié le format des réponses et les analyses auxquelles nous avons procédé par la suite. Le format des items comprenait des réponses dichotomiques (juste/faux) et des réponses dites de type « *partial credit* » (faux = 0 points, à peu près juste = 1 point, juste = 2 points). A la base de la fiche de contrôle et le répertoire des données, le logiciel WINSTEPS nous fournit suite aux analyses :

- une estimation du *logit* de chaque individu ;
- une estimation de la difficulté de chaque item ;
- et une estimation de l'homogénéité des items.

Les « *logits* » des individus sont fondés sur le logarithme naturel du rapport issu de la probabilité conditionnelle qu'un individu réussisse les items de l'épreuve de lecture de HarmoS, et de la probabilité conditionnelle qu'il ne les réussisse pas. Ainsi les « *logits* » des individus correspondent à l'estimation de leur compétence en compréhension orale et en compréhension de texte. Afin de les rendre plus compréhensibles, nous avons par la suite standardisé les « *logits* » sur une échelle « z » avant de les utiliser pour procéder à des analyses multi-niveaux.

### 3.6 Structure des données et méthode d'analyse multi-niveaux

La présente étude s'intéresse au développement des compétences en lecture chez les élèves plurilingues et monolingues en Suisse. Ces élèves sont des individus ayant une multitude de caractéristiques individuelles. Mais ils sont également les membres de différents contextes (famille, pairs, club de sport, classe, etc.). En tant que membres d'une classe, ils partagent un certain nombre de caractéristiques comme par exemple le même enseignant, la même approche didactique, le même lieu d'enseignement. Différentes classes d'élèves peuvent appartenir à la même école et par voie de conséquence partager de nombreuses caractéristiques liées à cette école. Plusieurs écoles peuvent se trouver dans une même commune et par voie de conséquence avoir des similarités, par exemple en termes de ressources financières ou de contextes sociaux et politiques. Effectuer une analyse auprès d'un groupe d'élèves nous confronte donc à un échantillon fortement hiérarchisé.

#### 3.6.1 Les modèles d'analyse multi-niveaux

Afin de prendre en compte les dimensions contextuelles hiérarchisées, comme l'appartenance à une classe ou une école, dans l'analyse individuelle, il faut utiliser des modèles multi-niveaux, car ils prennent en compte « des effets de grappes » (Dinaucourt). Ces modèles effectuent simultanément des régressions multiples à plusieurs niveaux. Développés en particulier par J.J. Hox et H. Goldstein pour les sciences de l'éducation, on les utilise pour rechercher simultanément des corrélations entre, d'un côté les indicateurs individuels et d'autre part les variables à des niveaux supra-individuels. Les méthodes classiques d'analyse à un seul niveau reposent sur l'idée indispensable d'une indépendance des observations. Lorsque cette condition est violée (ex. les performances des élèves ne sont pas indépendantes de leur appartenance à une même classe avec un même enseignant), les méthodes classiques conduisent à une sous-estimation des écarts-types et à des interprétations biaisées. L'étude d'Aitkin et Longford démontre clairement l'importance de la prise en compte de « l'effet de classe » (Aitkin & Longford, 1981). Reproduisant une étude sur les différences des méthodes d'enseignement de la lecture, Aitkin et Longford montrent, que lorsque l'on tient compte des classes dans lesquelles les élèves sont regroupés, les différences significatives entre les élèves suivants des méthodes d'enseignement différentes disparaissent.

A l'inverse, dans le domaine des études sociologiques, on s'est rendu compte qu'on ne peut pas utiliser une corrélation écologique qui a été mesurée à un niveau agrégé, comme substitut d'une corrélation individuelle. Cette inaptitude à déceler des comportements individuels à partir de mesures agrégées est à l'origine de ce que Courgeau et Baccaïni appellent « l'erreur écologique » (Daniel Courgeau & Baccaïni, 1997). Cependant il fallait attendre quelques années jusqu'à ce que nous disposions d'outils adaptés, comme MLwiN, pour la prise en compte simultanée de multiples niveaux.

Toutefois, les modèles multi-niveaux ne sont pas parfaits. Leur qualité dépend, comme pour les méthodes classiques de la qualité de la théorie qui les sous-tend et de la qualité des données recueillies. Ainsi selon Courgeau et Baccaïni (1997), les modèles multi-niveaux présentent le risque de conclure à l'existence d'effets d'agrégation alors qu'il n'y en a pas. Ceci peut arriver lorsque le modèle spécifié ne prend pas en compte une variable au niveau individuel qui influence à la fois la variable dépendante et les variables au niveau contextuel. Un manque de fiabilité dans les mesures des variables au niveau individuel peut également résulter dans une surestimation au niveau contextuel lorsque ces variables sont agrégées et introduites au niveau contextuel (Bellin, 2009).

Dans les analyses réalisées pour le présent travail de thèse, nous avons utilisé un modèle à trois niveaux. Afin de mesurer l'amélioration du modèle grâce à l'entrée de nouvelles variables indépendantes, nous avons construit au total quatre modèles. Chaque modèle comportait les trois mêmes niveaux, mais se distinguait des autres modèles par la quantité de variables indépendantes prises en compte.

- 1) Le premier niveau était destiné à prendre en compte le développement dans le temps des compétences en décodage et compréhension orale et son effet sur le niveau de compétences en lecture.
- 2) Le deuxième niveau était celui des caractéristiques individuelles. Il comportait les caractéristiques individuelles comme le niveau de compétences en décodage (modèle 1.1) et de compréhension orale (modèle 1.2), ensuite celui du QI (modèle 2), de l'arrière-plan linguistique (plurilingue versus monolingue), du statut socio-économique et du temps de lecture extra-scolaire (modèle 3).
- 3) Le troisième niveau était celui du contexte, à savoir celui des différentes classes scolaires.

### 3.6.2. Modèle conditionnel de croissance

Nous disposons grâce au caractère quasi-longitudinal de notre étude de deux mesures répétées; cela nous permet de prendre en considération la dimension temporelle des acquisitions de compétences en lecture.<sup>28</sup> Contrairement aux analyses des modèles de régression utilisant la méthode des moindres carrés ordinaires (MCO), l'utilisation de modèles multi-niveaux s'est vite relevée très appropriée pour l'analyse de données longitudinales. En effet, dans le cas de mesures répétées « ...les erreurs tendent à être auto-corrélées parce que l'effet d'une caractéristique non contrôlée des individus va avoir tendance à se répéter dans le temps » (Bressoux, 2010 p. 361). Les modèles multi-niveaux de croissance ont une composante stochastique dont la structure complexe des erreurs autorise que ces derniers soient autocorrélées et hétéroscédastiques au niveau intra-individuel. La non-indépendance des erreurs et leurs hétéroscédasticité peut ainsi être modélisée, même si en règle générale on suppose que la structure des erreurs se répète à l'identique d'un individu à l'autre (Bressoux, 2010).

Dans le cadre des modèles de croissance multi-niveaux, les mesures répétées sont considérées comme étant groupées au sein d'une unité supérieure qui est dans notre cas l'individu. Dans le modèle que nous utilisons ce n'est donc pas la personne mais les mesures répétées dans le temps qui se situent au niveau le plus bas de la hiérarchie des données. Les élèves se trouvent au niveau intermédiaire et les classes scolaires au niveau supérieur.

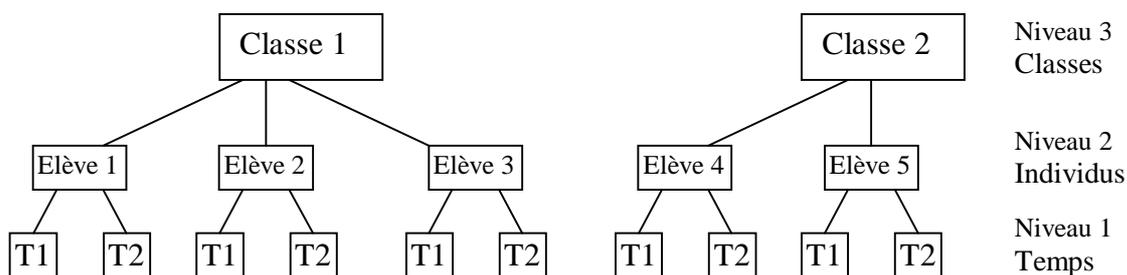


Figure 8 Illustration de l'hiérarchisation de l'échantillon

<sup>28</sup>Toutefois, nous avons conscience de deux limites majeures de notre étude. A cause de ces seulement deux et non trois points de mesure nous ne pouvons pas : a) déterminer la forme que prend le changement dans le temps et b) nous ne pouvons pas distinguer entre les „vrais“ changements et les erreurs de mesure. Pour des raisons de faisabilité, il nous était impossible d'ajouter un 3<sup>ème</sup> point de mesure. C'est pourquoi notre travail ne se focalise pas sur des questions concernant cet aspect temporel.

Dans un exemple proche de notre échantillon avec des élèves regroupés dans diverses écoles, une régression linéaire simple peut être exprimée de la façon suivante :

$$Y_i = \alpha + \beta X_i + e_i \quad (\text{I})$$

avec  $i$  ayant des valeurs de 1 à  $n_j$ , et  $n_j$  étant le nombre d'élèves dans la  $j$ ème école. Pour le  $i$ ème élève  $y_i$  est la „réponse“ (variable dépendante = variable réponse) comme par exemple le niveau de compétence en lecture.  $x_i$  est une variable explicative (variable indépendante), comme par exemple dans notre cas, la compétence en décodage ou en compréhension orale.

Nous pouvons simplement adapter l'équation I au style de notation utilisé par MLwiN

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (\text{II})$$

$\beta_0$  et  $\beta_1$  étant les coefficients de régressions, dont  $\beta_0$  indique où la droite de régression rencontre l'axe vertical<sup>29</sup> et  $\beta_1$  est l'indice de la pente de la droite de régression ;  $u_i$  signifie la variance résiduelle (variance d'erreur) due à la différence entre la compétence en lecture prédite et la compétence en lecture réelle de l'élève  $i$ .

Lorsque nous travaillons avec deux écoles, nous pourrions estimer deux régressions linéaires du type suivant, où  $u_{ij}$  est censé suivre une loi normale  $N(0, \sigma_j)$ :

$$\begin{array}{l} Y_{i1} = \beta_{01} + \beta_1 X_{i1} + u_{i1} \\ Y_{i1} = \beta_{02} + \beta_1 X_{i2} + u_{i2} \\ \text{-----} \downarrow \\ Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + u_{ij} \end{array} \quad (\text{III})$$

Comme nous avons dans notre échantillon un grand nombre de classes, avec parfois peu d'élèves, ces estimations pourraient donc nous mener à des résultats peu significatifs et biaisés par un bruit de fond. Ce dernier risque de masquer ce qui nous intéresse, à savoir les relations entre les caractéristiques individuelles qui sont modulées par le groupe sco-

---

<sup>29</sup> Notez que dans le logiciel MLwiN, le point de rencontre ou d'*interception* de la droite de régression et l'axe vertical  $\beta_0$  est nommé „*intercept*“. Un nom équivalent pour ce point d'interception est la « constante », car  $\beta_0$  est le coefficient de régression d'une variable fictive ayant la valeur constante 1 (voir équation V). En plus de correspondre à ce point d'*interception*, elle indique en même temps la valeur attendue de  $Y$  pour un individu ayant un score de 0 pour les différentes variables explicatives.

laire dans lequel se trouve l'individu. Il est donc souhaitable d'utiliser un modèle unique incluant les différentes classes (voir équation III).

L'indice  $j$  peut prendre des valeurs entre 1 à 17 une pour chaque classe dans notre échantillon des CM1/CM2. L'indice  $i$  se réfère maintenant à l'élève  $i$  dans la classe  $j$ . En général lorsqu'un paramètre contient les indices  $ij$ , il varie entre les classes et entre les élèves. Lorsqu'un paramètre n'a que l'indice  $j$  il varie seulement entre les classes mais il est constant pour les élèves au sein d'une classe. Lorsqu'un paramètre ne contient pas d'indice, sa valeur est constante pour tous les élèves et toutes les classes.

Tout comme pour les paramètres au niveau des individus, il est probable qu'une variance résiduelle existe au niveau des classes, due à une divergence des valeurs de paramètres estimées et des valeurs prédites. Il faut donc intégrer ce résidu du niveau des classes ( $v_j$ ) dans notre équation, ainsi nous obtenons l'équation suivante :

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + v_j + u_{ij} \quad (\text{IV})$$

Etant situées à des niveaux différents (individus versus classes), nous supposons que les deux variables résiduelles sont indépendantes l'une de l'autre. Etant donné qu'il s'agit de paramètres au caractère aléatoire, nous supposons également que leur distribution correspond à une distribution normale avec une moyenne autour de 0. C'est la variance ( $\sigma_u^2$  et  $\sigma_v^2$ ) de ces variables résiduelles qui nous intéresse plus spécifiquement.

Comme il s'agit d'un modèle multi-niveaux, notre exemple (équation IV) comporte des paramètres fixés ( $\beta_0$  et  $\beta_1$  qu'il faut estimer) et des paramètres aléatoires (les variances  $\sigma_u^2$  et  $\sigma_v^2$ ). Les paramètres aléatoires peuvent être appelés des résidus comme pour les modèles ayant un seul niveau. Cependant le fait d'en avoir deux à des niveaux d'agrégation différents, fait la particularité des modèles multi-niveaux.

Une version généralisée de l'équation (IV) est donc :

$$Y_{ij} = \beta_{0ij} X_{0ij} + \beta_1 X_{1ij}$$

$$\beta_{0ij} = \beta_0 + v_j + u_{0ij}$$

$$[v_j = v_j, u_{0ij} = u_{ij}, X_{0ij} = 1 = \text{constante}] \quad (\text{V})$$

Ce type de modèle constitué de paramètres aléatoires à plusieurs niveaux d'agrégation demande l'utilisation de méthodes d'estimation plus complexes que celles des régressions linéaires.

Ces méthodes ont été décrites par exemple par Goldstein (2003). En nous appuyant sur Courgeau (2004), nous allons citer seulement quelques éléments clés.

Pour estimer la relation entre une variable réponse et un ensemble de variables explicatives à des niveaux variés, MLwiN utilise la méthode de « maximum de vraisemblance » des données d'un échantillon, sur la base d'une matrice de variances – covariances. Le principe statistique utilisé pour l'estimation de notre modèle spécifique est celui du maximum de vraisemblance « complet ». En maximisant la vraisemblance des données de l'échantillon, la moyenne et la variance de la variable réponse sont estimées à l'aide d'un algorithme des moindres carrés itératifs généralisés (en anglais *iterative generalized least squares IGLS*)<sup>30</sup>.

Comme nous avons déjà indiqué, dans le cadre de nos analyses nous avons construit un modèle à trois niveaux. Dû au caractère quasi-longitudinal de notre étude, nous avons introduit en tant que troisième niveau le niveau du temps, c'est-à-dire le changement des compétences entre le moment de la première et de la deuxième collecte des données.

Nous avons également incorporé dans notre modèle une mesure du temps en tant que variable explicative, en supposant que les compétences de lecture (variable réponse) sont, entre autres, influencées par le niveau scolaire des élèves. Nous avons ainsi exprimé le temps passé entre les deux moments de collecte des données par le changement d'appartenance de la classe de CE1 à la classe de CM2 (respectivement de la classe de CM2 à la classe de 6ème). Outre l'introduction du temps comme variable explicative, nous avons introduit un terme d'interaction entre le temps et chacune des autres variables explicatives.

Chaque modèle multi-niveaux comporte des paramètres fixés et des paramètres aléatoires. Dans notre modèle, toutes les variables explicatives, y compris la constante, sont des paramètres fixés. Les paramètres aléatoires sont les variances (résiduelles) et les covariances aux niveaux des individus et des classes. Nous avons spécifié notre modèle de base (appelé modèle vide, car il ne contient pas encore des variables explicatives) de façon à ce que  $\beta_0$  et  $\beta_1$  puissent varier entre les individus et les classes. C'est-à-dire que la constante ( $=\beta_0$  ou l'*intercept*, voir note de bas de page, page 76), l'endroit où la droite de régression rencontre l'abscisse peut être différent suivant les individus et les classes.

---

<sup>30</sup> Cette méthode n'est pas sans problème (voir Bressoux, 2010 p. 317).

De même, l'influence du facteur temps ( $\beta_1$ , qui équivaut à la pente de la droite de la régression) peut varier entre les individus et les classes.

Pour les variables explicatives qui s'ajoutent au fur et à mesure que nous spécifions d'avantage notre modèle (voir modèles 1.1 à modèles 2 dans l'annexe 2), nous partons du principe que leur influence sur la variable réponse ( $\beta_2$  à  $\beta_5$ ) est constante pour tous les individus et toutes les classes. C'est pourquoi les coefficients de  $\beta_2$  à  $\beta_5$  ne contiennent pas d'indices. Cependant, les variables explicatives elles mêmes ont une valeur particulière suivant le moment de la collecte des données (CE1 ou CE 2, signalé par l'indice  $zeit01$ ), suivant l'individu (signalé par l'indice  $IDSk$ ) et suivant la classe (signalé par l'indice  $KLANRS$ ).

Nous présentons les modèles multi-niveaux hiérarchisés sur lesquels se base nos analyses dans l'annexe 2.

### 3.6.3 La centration des variables

Dans un modèle multi-niveaux l'estimation de la variance est étroitement liée aux variables explicatives. La variance est estimée à l'origine des abscisses c'est-à-dire pour une valeur de  $X = 0$ . Il peut arriver que cette valeur nulle, n'ait pas de signification particulière, puisqu'elle peut se situer en dehors de l'intervalle des valeurs observées de  $X$  et même en dehors de ses valeurs théoriques (Bressoux, 2010).

L'exemple du test d'intelligence CFT 20 illustre la nécessité d'une centration des variables. Les valeurs de QI résultant du test CFT 20 peuvent varier autour de la moyenne de 100, mais jamais un individu n'obtiendra un QI de 0. Même en n'ayant eu aucune réponse correcte, cela donnera après les transformations un  $QI > 0$ . La constante (= l'intercept =  $\beta_0$ ) représente le point de rencontre de la droite de régression avec l'abscisse et elle indique ainsi en même temps la valeur attendue de  $Y$  pour un individu ayant un score de 0 sur la variable explicative  $X$  ( $\beta_0 = E(Y_{ij}/X_{ij}=0)$ ). Il est préférable que la valeur 0 représente un cas dont la signification est bien déterminée (Bressoux, 2010, p. 327), ce qui n'est pas vrai pour un QI de 0. Une solution possible consiste à centrer les variables explicatives sur leurs moyennes générales ( $X_{ij} - \bar{X}$ ) ce que nous avons fait pour le QI. Ainsi la constante représente la valeur attendue de  $Y$  pour un individu ayant un score de 0 (= la moyenne) sur la variable explicative  $X$ . Il en est le même pour nos autres variables explicatives telles que les compétences de décodage ou de compréhension orale, puisque nous les avons transformées sur une échelle de  $z$ , ce qui amène également la moyenne à être représentée par la valeur de 0. Dans le cas des variables explicatives telles que l'origine linguistique, la valeur 0 codifie le fait d'être monolingue. En ce qui concerne le statut socio-économique la valeur 0 signifie la catégorie socioprofessionnelle moyenne.

### 3.6.4 Les inférences statistiques dans le cadre des modèles multi-niveaux

#### *La significativité testée à l'aide de la déviance*

La significativité des modèles construits et l'amélioration de ces modèles au fur et à mesure que des variables explicatives supplémentaires y sont intégrées nous est signalée par la décroissance de la déviance par rapport aux modèles moins complets. Lorsque les modèles multi-niveaux utilisés sont construits de manière à ce que les modèles moins complets soient toujours inclus dans le modèle suivant plus complet (*nested models*), on

peut tester la significativité du modèle en observant la décroissance de la déviance par rapport aux modèles précédents. Suivant une loi du Chi<sup>2</sup> avec  $m$  degrés de liberté on peut ainsi facilement estimer si la décroissance de la déviance (et donc l'augmentation de la vraisemblance) est significative en passant d'un modèle au suivant.

#### *La significativité de chaque paramètre*

Comme pour les régressions linéaires, on peut estimer la significativité de chaque paramètre en rapportant la valeur des coefficients de régression à leurs erreurs-types. Lorsque le rapport entre ces deux valeurs est au moins égale à 1.96 et sous la condition d'une distribution normale, l'hypothèse nulle (de non significativité) peut être rejetée avec une certitude de 95%. Cependant, il ne s'agit là que d'approximation, car l'estimation des erreurs-types est sujette à caution dans les analyses multi-niveaux (Bressoux, 2010 p. 316). Ainsi le logiciel MLwiN ne propose aucun test classique de significativité. On obtient toutefois une indication de la tendance, raison pour laquelle cette méthode classique simple est toujours assez utilisée dans le cadre d'analyses multi-niveaux.

#### *Le pouvoir explicatif du modèle*

Étant donné que le but de chaque modélisation est la construction d'une représentation schématique du réel, la qualité d'un modèle s'explique par son pouvoir explicatif. Dans le cadre des modélisations utilisant les moindres carrés ordinaires, le coefficient de détermination  $R^2$  nous indique ce pouvoir explicatif en se basant sur la proportion de la variance expliquée par rapport à la variance résiduelle. Dans le cadre des modèles multi-niveaux, nous avons des composantes de variance à chacun des niveaux, il faut donc calculer un coefficient de détermination pour chaque niveau, nommé « pseudo  $R^2$  » selon Singer et Willet (2003 pp.103-104).

Pour cela nous regardons dans le cadre de notre modèle, au niveau des individus et des classes, la réduction de la variance résiduelle en la comparant aux composantes de variances indiquées par le modèle vide. Ce dernier nous fournit les « valeurs de base » et lors de l'interprétation toutes les spécifications subséquentes par introduction de variables explicatives lui sont rapportées (Bressoux, 2010). Le modèle « vide » est inconditionnel car aucune variable explicative n'y est intégrée, il nous renseigne cependant sur la décomposition de la variable réponse (compétences en compréhension de texte) entre le niveau des individus et des classes.

Pour déterminer le pouvoir explicatif de notre modèle, nous calculons l'écart entre la variance indiquée par le modèle vide et la variance résiduelle par le modèle plus complet dont le pouvoir explicatif est à déterminer. Cet écart est ensuite divisé par la variance indiquée par le modèle vide.

Il faut toutefois ajouter que le débat entre experts concernant les méthodes de calcul du pouvoir explicatif des modèles multi-niveaux n'est pas clos. L'indice « pseudo  $R^2$  » présente l'inconvénient de ne pas pouvoir empêcher le problème d'une possible augmentation de la variance résiduelle. Il peut en effet arriver lorsque la répartition de la variance résiduelle entre les niveaux est très disproportionnée que l'introduction d'une nouvelle variable explicative réduise la partie de la variance résiduelle élevée alors qu'elle peut en même temps augmenter la variance résiduelle au niveau où elle était la plus faible. Cela conduit par voie de conséquence à un coefficient « pseudo  $R^2$  » négatif aberrant et contraire à l'idée qu'on se fait du pouvoir explicatif d'un modèle (Bressoux, 2010). L'interprétation des coefficients de détermination des modèles multi-niveaux exige donc un minimum de prudence.

### 3.7 Opérationnalisation des variables indépendantes

Nous allons par la suite présenter l'opérationnalisation des variables indépendantes et dépendantes, dans l'ordre par lequel elles rentrent dans les analyses à niveaux multiples. Cette description porte sur le fondement des tests utilisés ainsi que sur leurs caractéristiques et leurs qualités métriques.

#### a) *Décodage Test ELFE 1-6 partie 1 (ZWortRoh\_2bis3 et ZWortRoh\_5bis6<sup>31</sup>)*

Auteurs : Lenhard, Wolfgang et Schneider, Wolfgang (2006)

Items: 72

Versions : A et B (différence dans l'ordre des items)

Public ciblé : jeunes lecteurs de la 1<sup>ère</sup> classe (CP) à la 6<sup>ème</sup> classe (6<sup>ème</sup>)

Méthode: Cocher le mot correspondant au dessin parmi 3 mots semblables du point de vue graphique ou phonologique (voir image ci-dessous).

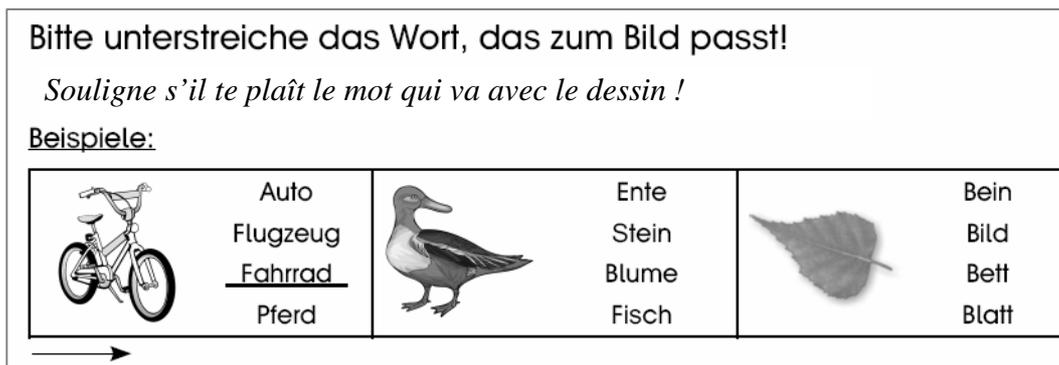


Figure 9 Extrait du ELFE partie 1

Les différentes études concernant la théorie de la *Simple View of Reading* évaluent les capacités de décodage à travers la lecture de pseudo-mots ou à travers la lecture de mots réels isolés, c'est dire sans qu'ils soient intégrés au contexte d'une phrase. La lecture de pseudo-mots nécessite une lecture à haute voix pour pouvoir juger de la prononciation correcte de ces mots qui indiquera en retour s'il y a une correspondance correcte entre le code graphique et le code phonologique. La lecture des mots réels n'implique pas nécessairement une lecture à voix haute et se destine ainsi mieux à une évaluation en classe.

<sup>31</sup> Derrière chaque variable opérationnalisée au sein du chapitre 3.7 vous trouvez indiqué entre parenthèses la dénomination de ces variables comme on les utilise dans les modèles statistiques présentés au sein de l'annexe 2.

Par ailleurs, les auteurs de la *SVRT* (Hoover et Gough, 1990) préconisent l'utilisation des pseudo-mots pour les jeunes lecteurs débutants, tandis que pour les lecteurs avancés, ils préconisent la méthode des mots réels pour éviter un effet de plafond. Pour la présente étude nous avons choisi d'utiliser la méthode des mots réels comme le test ELFE 1-6 de Lenhard & Schneider (2006) le propose dans sa première partie.

Le test ELFE est un instrument pour l'évaluation des capacités de lecture à trois niveaux et sous contrainte de temps: le mot, la phrase et le texte. Il a été développé pour des enfants entre la 1<sup>ère</sup> classe (CP) et la 6<sup>ème</sup> classe (6<sup>ème</sup>). Dans la description du fondement théorique du ELFE, les auteurs se réfèrent explicitement à *SVRT*. La première partie du ELFE 1-6 évalue la capacité à lire des mots de façon correcte et rapide, en identifiant parmi 4 mots visuellement semblables le mot qui correspond à l'objet montré sur un dessin. Au total 72 mots de longueurs et complexités variables, doivent être ainsi identifiés en 3 minutes<sup>32</sup>. Sur les 72 items, 17 contiennent des mots monosyllabiques, 19 items contiennent des mots à deux syllabes, 18 items proposent des mots à trois syllabes et 18 items utilisent des mots à quatre syllabes. Le nombre de mots correctement identifiés est le score obtenu dans cette première partie du test ELFE. Il a été par la suite transformé sur une échelle de z pour les analyses statistiques de niveaux multiples. L'utilisation de deux versions du test au contenu identique mais avec un ordre de mots différent (A et B) avait pour but d'empêcher que des élèves assis côte à côte copient leurs réponses. Cette précaution et l'utilisation d'un chronomètre devaient assurer la fiabilité des résultats.

Tableau 11 Indices de qualité du test ELFE partie 1 (Test de Décodage)

Objectivité	Le manuel du test prévoit des instructions pré formulées et des indications précises pour la passation et l'évaluation du test.
Homogénéité Fidélité	Cronbach $\alpha = 0.97$
	Corrélation entre deux résultats de tests réalisés à 14 jours d'intervalle (retest) $r = 0.91$
Validité convergente <sup>33</sup>	Corrélation de $r = 0.64$ avec le test de lecture « Würzburger Leise Leseprobe » (WLLP).
	Corrélation de $r = 0.52$ avec l'évaluation des compétences en lecture de la part de l'enseignant.
Validité discriminative	Corrélation de $r = 0.16$ avec l'évaluation de l'enseignant des compétences en mathématiques et de $r = 0.27$ pour les compétences en écriture

<sup>32</sup> Le manuel du test ELFE prévoit une limite de temps de 3 minutes pour les élèves du CP au CM1 et une limite du temps 2 minutes pour les élèves qui sont en CM2 et en 6<sup>ème</sup>. Pour rendre les résultats comparables entre les groupes d'âges, nous avons cependant appliqué une limite de temps de 3 minutes à l'ensemble des élèves.

<sup>33</sup> Les corrélations indiquées dans ce tableau concernent des classes de CM1.

Tableau 12 Distribution des résultats du test de décodage du ELFE

Variable		N	Minimum	Maximum	Moyenne	Déviati on Standard	Extremité	Curtosis
ELFE	CE 1 t 1	312	2	63	24.08	9.253	.653	1.480
	CE 2 t 2	334	5	72	36.21	11.751	.451	.520
Décodage	CM 2 t 1	268	22	72	52.10	12.071	-.092	-.569
	6ème t 2	261	26	72	62.77	9.697	-1.275	1.139

Nous avons, pour les analyses à niveaux multiples, procédé à une transformation des scores obtenus par les élèves (WortRoh) sur une échelle de z. Ainsi nous avons créé les variables ZWortRoh\_2bis3 (CE1/CE2) et ZWortRoh\_5bis6 (CM2 / 6<sup>ème</sup>).

Le tableau 12 ci-dessus nous montre la distribution des deux variables qui résument les résultats pour le test de décodage dans les deux groupes d'âges. Nous remarquons une nette augmentation des moyennes, mais aussi une taille d'échantillon qui varie entre les deux moments de recueil des données (t1-t2). Cette variation s'explique d'une part par des mouvements démographiques naturels au sein des classes dus à des déménagements ou des redoublements. D'autre part, certains élèves ont pu être absents de l'école le jour où un test de lecture, de décodage ou de compréhension orale avait été réalisé. En ce qui concerne le questionnaire et le test d'intelligence nous avons essayé de collecter toutes les données manquantes. Pour les tests de langue, cela n'était pas possible. Notre échantillon comporte ainsi toujours un petit nombre d'élèves, pour lesquels nous n'avons les données qu'au moment t1 ou t2.

b) *HarmoS\_L1 Test de Compréhension Orale (ZLogit\_HörenGesamt\_23 et ZLogit\_HörenGesamt56)*

Auteurs : HarmoS-L1 Consortium (HarmoS-L1, 2007)

Versions : A et B (différence dans l'ordre des items)

Public ciblé : enfants de la 2<sup>ème</sup> classe (CE2) et de la 6<sup>ème</sup> classe (6<sup>ème</sup>)

Format : questions ouvertes et QCM

Méthode : après l'écoute du texte, cocher les réponses correctes.

**Strassenkinder in Indien** *Les enfants de rue en Inde*

**Beantworte die folgenden Fragen. Lies die Fragen genau durch. Es gibt immer nur eine richtige Antwort, kreuze sie an.**

*Réponds aux questions suivantes après les avoir lu attentivement. Il n'y a toujours qu'une seule réponse correcte.*

Was ist ein Strassenkind? *Qu'est ce qu'est un enfant de rue?*

1.  Ein Kind, das Strassen putzt. *Un enfant qui nettoie les rues.*
2.  Ein Kind, das auf der Strasse und von der Strasse lebt. *Un enfant vivant dans la rue et de la rue.*
3.  Ein Kind, das sich in den Strassen gut auskennt. *Un enfant connaissant bien les rues.*
4.  Ein Kind, das gerne auf der Strasse spaziert. *Un enfant qui aime se balader dans la rue.*
5.  Ein Kind, das auf der Strasse geboren wird. *Un enfant né dans la rue.*

Figure 10 Exemple de question de Compréhension orale (HarmoS\_L1)

L'évaluation des compétences en compréhension orale se heurte souvent à un choix très limité en matière de tests standardisés par comparaison avec la multitude de tests disponibles pour l'évaluation de la compréhension de texte.

Dans le cas du projet de recherche de la HEP de Berne, nous avons pu profiter du matériel de tests développés par le projet HarmoS pour la compréhension orale et la compréhension de texte. En effet le projet HarmoS vise une harmonisation de l'éducation scolaire entre les différents cantons suisses. Un des points clés du projet HarmoS est la description des standards de compétence en mathématiques, en allemand et en sciences naturelles qu'un élève doit avoir acquis à des moments précis de sa scolarité. Un groupe de travail (HarmoS-L1 Consortium) constitué de nombreux chercheurs et enseignants venant entre autres des HEP de Zürich, de Berne, de Lucerne et de Genève s'est appliqué à formuler ces standards de compétence pour la compréhension orale, la compréhension de texte et l'écriture en langue de scolarisation.

Avant de formuler de tels standards, il était nécessaire de connaître le niveau des compétences actuelles des élèves dans ces trois domaines de langue. Basé sur un modèle théorique fédérateur, le groupe HarmoS-L1 Consortium entreprit la construction de tests

pour la compréhension orale, la compréhension de texte et l'écriture en langue de scolarisation. Le modèle de HarmoS L1 est un modèle « pragmatique ». Il concerne les aspects nécessaires à la réalisation « d'actions langagières » dans le domaine de la lecture, de l'écriture, de la compréhension ou de la production orale. Le groupe HarmoS L1 distingue cinq aspects de ces « actions langagières ». De plus différents 6 niveaux de difficulté ou de complexité ont été définis pour chaque aspect. Le tableau 13 ci-dessous décrit les 5 aspects tandis que les tableaux 14 et 15 indiquent pour les 4 tâches de compréhension orale que nous avons utilisée, l'aspect et le niveau de complexité visé par chacune des questions.

Tableau 13 Extrait du modèle de HarmoS-L1

Aspects	Description
Situer une action langagière	Identifier la fonction et l'intention d'un texte
Planifier une action langagière	Définir son objectif et des stratégies de compréhension de texte
Réaliser une action langagière	Trouver des informations explicites et implicites, faire des déductions, construire un modèle mental cohérent
Evaluer une action langagière	Evaluer le contenu et les effets recherchés d'un texte, exposer son point de vue par rapport au texte
Réparer une action langagière	Poser des questions pour préciser ou modifier son modèle mental basé sur le texte

Le groupe HarmoS L1 a ensuite essayé de construire des « tâches » comprenant des textes authentiques et des contextes de différents genres textuels pour la lecture et la compréhension orale ainsi que des questions qui ciblent au mieux les cinq aspects définis ci-dessus. Cependant il est rare qu'une « tâche » réussisse à cibler ces cinq aspects. Plusieurs « tâches » ont été construites pour différents groupes d'âges. La participation du Prof. Romano Müller (chef du projet de recherche de la HEP de Berne) au sein du groupe HarmoS L1 nous a permis de choisir certaines de ces tâches pour l'évaluation des compétences en compréhension orale, compréhension de texte et la rédaction de textes.

Les « tâches » de la compréhension orale présentaient pour nous l'avantage d'avoir été construites avec beaucoup de soin pour un échantillon actuel d'élèves suisses en utilisant un matériel authentique auquel les jeunes en Suisse pourraient être confrontés en écoutant la radio ou la télévision. Nous avons choisi pour l'évaluation des compétences en compréhension orale dans le cadre de la présente thèse quatre tâches différentes dont:

- deux pour le niveau CE1/CE2 (la narration « Gâteau au chocolat » et le reportage « Les enfants de la rue en Inde »)

- - et deux pour le CM2/ 6<sup>ème</sup> (le reportage « Pribjat » et une argumentation sur le « pour et le contre de l'uniforme scolaire »).

Tableau 14 Les exercices d'compréhension orale pour les CE1/CE2

Tâche	Genre		Question	Aspect	Description de l'aspect
Le gâteau au chocolat	narratif	monologue	1	rea6	trouver de l'information explicite
			2	rea6	trouver de l'information explicite
			3	rea2	activer des connaissances linguistiques
			4	sit2	identifier la fonction du texte
			5	rea6	trouver de l'information explicite
			6	rea2	activer des connaissances linguistiques
			7	rea5	trouver de l'information implicite / deduction
			8	rea6	trouver de l'information explicite
			9	sit2	identifier la fonction du texte
Les enfants des rues en Inde	expositif	dialogue	1	rea2	activer des connaissances linguistiques
			2	rea6	trouver de l'information explicite
			3	rea5	trouver de l'information implicite / deduction
			4	rea5	trouver de l'information implicite / deduction
			5	sit2	identifier la fonction du texte
			6	rea6	trouver de l'information explicite
			7	rea6	trouver de l'information explicite
			8	rea5	trouver de l'information implicite / deduction
			9	eva1	résumer l'essentiel du contenu
			10	eva3	reflexion sur le contenu et la forme du texte
			11	sit2	identifier la fonction du texte
			12	rea5	trouver de l'information implicite / deduction

Tableau 15 Les exercices d'compréhension orale pour les CM2/6ème

Tâche	Genre		Question	Aspect	Description de l'aspect
Pribjat	expositif	monologue	1	rea6	trouver de l'information explicite
			2	rea5	trouver de l'information implicite / deduction
			3	rea6	trouver de l'information explicite
			4	rea6	trouver de l'information explicite
			5	sit2	identifier la fonction du texte
			6	rea6	trouver de l'information explicite
			7	rea6	trouver de l'information explicite
			8	rea2	activer des connaissances linguistiques
			9	rea2	activer des connaissances linguistiques
			10	rea5	trouver de l'information implicite / deduction
Uniforme scolaire	argumentatif	dialogue	1	sit2	identifier la fonction du texte
			2	eva3	reflexion sur le contenu et la forme du texte
			3	eva3	reflexion sur le contenu et la forme du texte
			4	rea2	activer des connaissances linguistiques
			5	rea6	trouver de l'information explicite
			6	rea5	trouver de l'information implicite / deduction
			7	rea6	trouver de l'information explicite
			8	rea6	trouver de l'information explicite
			9	eva5	évaluer le contenu par rapport à sa propre point de vue
			10	rea6	trouver de l'information explicite

Les critères qui ont guidé notre choix concernaient la validité des tâches, la sélectivité le caractère discriminatif des items, le genre textuel (qui devait varier), la difficulté des tâches (qui devait être moyenne avec des questions allant du très simple au très complexe), les aspects ciblés et le temps nécessaire pour écouter les textes et répondre aux questions. Afin de réduire le risque que les élèves copient les réponses sur leur voisin, nous avons créé deux versions en changeant l'ordre des questions. Après une explication du déroulement du test, les enseignants faisaient écouter une seule fois le texte avant de lire à haute voix, ensemble avec les élèves, les questions et de les laisser y répondre. Nous avons examiné la qualité des tâches non-standardisées de HarmoS-L1 à l'aide des analyses de Rasch réalisées avec le logiciel WINSTEPS (Linacre 1991-2006).

Tableau 16 Indices de qualité du test HarmoS (Compréhension orale)

HarmoS Compréhension orale	CE1/CE2		CM2/6ème	
	Real	Modèle	Real	Modèle
<i>Root Mean Square Standard Error (RMSE)<sup>34</sup></i>	0.55	0.53	0.55	0.54
<i>Adjusted Standard Deviation (ADJSD)<sup>35</sup></i>	0.91	0.92	0.7	0.71
<i>Separation<sup>36</sup></i>	1.65	1.72	1.27	1.33
<i>Person Reliability</i>	0.73	0.75	0.62	0.64

Sous la colonne MODEL sont réunis les indicateurs calculés en supposant que les données de notre échantillon sont en adéquation avec le modèle de Rasch. Cette colonne rapporte donc les estimations les plus optimistes, tandis que la colonne REAL rapporte les estimations les plus pessimistes, basées sur la supposition que les spécifications du modèle de Rasch ne sont pas parfaitement remplies dans les données de notre échantillon. Nous nous intéressons ici en particulier à l'indicateur « Person Reliability ». Dans ces calculs de WINSTEPS, ce dernier est l'équivalent de la fiabilité calculée par « Crohnbachs  $\alpha$  ». Nous voyons dans le tableau une estimation de fiabilité qui est meilleure pour les tâches du CE1/CE2 (0.73-0.75) que pour celles des CM2/6<sup>ème</sup> (0.62-0.64). Nous la jugeons cependant satisfaisante pour les deux groupes de notre échantillon.

Le tableau 17 ci-dessous rapporte les résultats d'une analyse de composantes principales, réalisée dans le cadre des analyses de Rasch à l'aide du logiciel WINSTEPS. Cette

<sup>34</sup> RMSE est la racine carrée de la moyenne de la variance d'erreur dans l'échantillon.

<sup>35</sup> ADJ. S.D. est l'écart-type „ajusté“ c'est à dire que c'est l'écart-type de l'échantillon  $(ADJ.S.D.)^2 = (S.D. \text{ of MEASURE})^2 - (RMSE)^2$

<sup>36</sup> La SEPARATION rapporte la proportion entre ADJ.S.D la déviation standard ajustée et RMSE, la déviation standard des erreurs. SEPARATION<sup>2</sup> est la proportion « signal-to-noise » c'est à dire la proportion entre la vraie variance et la variance d'erreur.

analyse des composantes principales ne doit pas être confondue avec une analyse factorielle. Cette dernière s'intéresse à la structure des corrélations entre les items et les facteurs latents. L'analyse des composantes principales s'attache cependant à falsifier l'hypothèse que tous les résidus de variance (donc la variance d'erreur) sont dû au hasard en essayant de trouver des composantes qui expliquent un maximum de la variance des résidus.

Tableau 17 Résultats des analyses de composantes principales (Harmos Compréhension Orale)

<b>HarmoS Compréhension orale</b>	<b>CE1/CE2</b>		<b>CM2/6ème</b>	
<i>Standardized Residual Variance (Eigenvalue units)</i>				
<i>Total variance in observations</i>	38.3	100%	30	100%
<i>Variance explained by measures</i>	17.3	45.10%	10	33.30%
<i>Unexplained variance (total)</i>	21	54.90%	20	66.70%
<i>Unexplained variance in 1st contrast</i>	1.7	4.50%	1.6	5.20%
<i>Unexplained variance in 2nd contrast</i>	1.6	4.10%	1.4	4.60%
<i>Unexplained variance in 3rd contrast</i>	1.4	3.70%		
<i>Unexplained variance in 4th contrast</i>	1.4	3.70%		
<i>Unexplained variance in 5th contrast</i>	1.3	3.40%		

Dans le tableau 17, ces composantes sont indiquées par le terme « contraste ». Leur force d'explication est exprimée par les unités appelées « Eigenvalue ». Lorsqu'un « contraste » a un « Eigenvalue »  $< 2$ , et donc une force d'explication de moins de deux items, la variance des résidus est jugée être due au hasard. Dans notre échantillon, toutes les unités « Eigenvalue » étant  $< 2$  il n'existe donc pas de dimension cachée dans nos tâches de compréhension orale. Tous les items visent une même compétence latente.

Tableau 18 Indicateurs de « *fit* » et de difficulté

HarmoS Compréhension orale CE1/CE2		<i>Infit</i>		<i>Outfit</i>
Numéro d'Item	Difficulté	MNSQ	MNSQ	MNSQ
Gâteau chocolat 4	-0.72	1.29	1.77	
Enfants de Rue 5	0.73	1.15	1.29	
Gâteau chocolat 9	1.02	1.18	1.27	
Gâteau chocolat 7	1.53	1.04	1.09	
Gâteau chocolat 8	1.49	0.99	1.09	
Enfants de Rue 10	0.67	1.06	1.04	
Gâteau chocolat 5	-0.49	1.02	1.06	
Gâteau chocolat 3	0.23	1.03	1.05	
Enfants de Rue 1	-0.95	0.98	1.04	
Enfants de Rue 9	-1.43	1.03	1.04	
Enfants de Rue 7	-0.03	0.96	1.01	
Gâteau chocolat 1	-0.6	0.98	0.89	
Enfants de Rue 6	1.07	0.91	0.98	
Enfants de Rue 12	0.08	0.96	0.96	
Enfants de Rue 11	1.39	0.96	0.96	
Enfants de Rue 3	-1.34	0.9	0.79	
Enfants de Rue 2	-0.39	0.9	0.82	
Gâteau chocolat 2	-1.57	0.9	0.83	
Enfants de Rue 4	-0.36	0.88	0.8	
Enfants de Rue 8	-0.89	0.88	0.8	
Gâteau chocolat 6	0.58	0.81	0.81	

A l'inverse, un individu a résolu un item alors qu'il devait être trop difficile pour lui. Le « *infit* » est sensible aux déviations pour des items dont le niveau de difficulté (par ex. facile) correspond au niveau de compétences de l'individu (par ex. faible). Ainsi un individu de compétence faible n'aurait pas résolu, par exemple, un item qui était pourtant facile et donc à sa portée. Le « *outfit* » comme nous l'avons décrit est sensible aux déviations qui concernent des items qui sont loin au dessus ou en dessous du niveau de compétence de l'individu. Dans le cadre du modèle de Rasch des estimations de « *infit* » et « *outfit* » entre 0.5 et 2 sont acceptables. Les estimations <.05 apportent moins d'informations sans pour autant dégrader

trop les analyses de Rasch mais elles produisent une estimation de fiabilité trop positive. Les estimations > 2 cependant indiquent la présence de biais qui dégradent les analyses de Rasch.

Tableau 19 Indicateurs de « fit » et de difficulté (CM2/6<sup>ème</sup>)

Compréhension orale CM2/6 <sup>ème</sup>		Infit	Outfit
Numéro d'Item	Difficulté	MNSQ	MNSQ
Pribjat 10	2.01	1.04	1.29
Pribjat 5	0.44	1.03	1.17
Pribjat 4	0.54	1.11	1.15
Uniforme scolaire 3	-0.22	1.08	1.09
Uniforme scolaire10	-1.63	0.98	1.08
Uniforme scolaire 5	0.22	1.08	1.05
Uniforme scolaire 9	1.4	1.06	1.02
Pribjat 8	0.52	0.95	1.04
Uniforme scolaire 8	-0.7	1.03	1.03
Uniforme scolaire 2	-0.09	1.02	1.01
Uniforme scolaire 7	0.44	1.01	1.02
Uniforme scolaire 4	0.44	1.01	0.99
Pribjat 7	-0.79	0.97	1.01
Pribjat 1	-0.39	1.01	0.99
Uniforme scolaire 1	-0.98	0.96	0.94
Pribjat 3	-0.52	0.95	0.92
Uniforme scolaire 6	0.45	0.94	0.95
Pribjat 2	-0.54	0.94	0.93
Pribjat 9	-0.96	0.92	0.85
Pribjat 6	0.48	0.9	0.85

Comme pour le groupe des CE1/CE2, les données issues du groupe des CM2/6<sup>ème</sup> montrent des estimations de « fit » satisfaisantes. Les deux derniers tableaux contiennent dans une troisième colonne une estimation de la difficulté de chaque item. Cette estimation est basée sur la probabilité de réussite pour chaque item au sein de notre échantillon. La difficulté moyenne se situe à 0, en dessous se trouvent les items plus faciles en dessous les items plus difficiles. Nous voyons dans ces deux tableaux que la difficulté des items varie de manière satisfaisante,

condition de base pour pouvoir évaluer de façon suffisamment différenciée les compétences en compréhension orale.

Le tableau 20 ci-dessous nous montre la distribution des compétences ainsi évaluées au sein de nos deux échantillons.

Tableau 20 Distribution des résultats du test de HarmoS pour la Compréhension Orale

Variable			N	Minimum	Maximum	Moyenne	Déviati on Standard	Extremité	Curtosis
Har moS Com pré hension orale	CE 1	t 1	334	-3	3	.28	1.07	0.02	0.26
	CE 2	t 2	347	-3	5	.80	1.13	0.01	1.02
	CM 2	t 1	283	-4	2	-.37	0.90	-0.23	1.44
	6 <sup>ème</sup>	t 2	274	-3	5	.09	0.98	0.80	1.59

Les estimations de compétence en compréhension orale obtenues par les analyses de Rasch pour chaque élève ont été transformées sur une échelle de z. Ainsi nous avons obtenu les variables ZLogit\_HörenGesamt\_23 et ZLogit\_HörenGesamt56.

c) *Culture Fair Test CFT 20R (CFTIQTOT)*

Auteurs : Weiss, Reinhard H. (Weiss, 2006)

Items : Continuer des suites logiques, parties 1 et 2 (14 items)

Faire des classifications, parties 1 et 2 (15 items)

Résoudre des matrices, parties 1 et 2 (14 items)

Repérer des relations spatiales, parties 1 et 2 (10 items)

Versions : A et B (différence dans l'ordre des items)

Public ciblé : Enfants et adultes âgés de 8,7 ans à 70 ans

Méthode: Cocher la bonne réponse dont l'identification implique, selon les sous-tests, des capacités cognitives différentes (voir exemple ci-dessous).

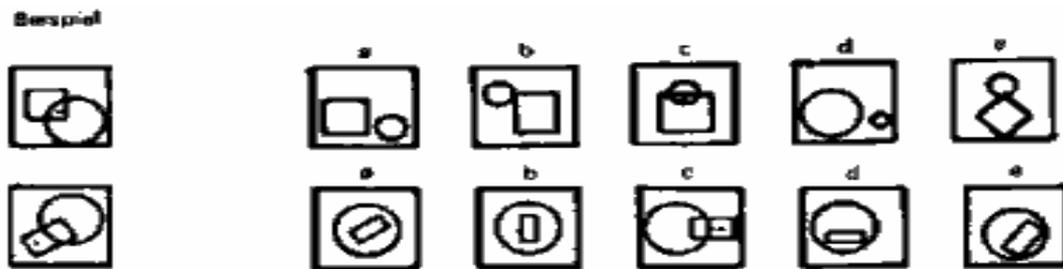


Figure 11 Exemple « Repérer des relations spatiales »

Nous avons utilisé le CFT 20R de Weiss (2006), qui est une version modifiée et renouvelée du test d'intelligence « Cultur Fair Test » CFT développé par Cattell et Weiss dans les années soixante-dix pour l'évaluation de l'intelligence fluide. Basée sur la théorie de l'intelligence cristallisée et de l'intelligence fluide (Cattell, 1963), cette dernière est définie par Cattell comme étant la capacité à découvrir des relations figurales et à résoudre rapidement des problèmes de logique d'une complexité croissante. Selon Cattell cette capacité cognitive serait peu conditionnée par des influences du contexte socioculturel environnant. Souhaitant une évaluation des capacités intellectuelles, entre autres d'enfants plurilingues et issus de l'immigration qui soit indépendante des influences d'ordre linguistique, culturel ou sociale, le CFT 20R a été le parfait instrument de notre choix. Il consiste en deux parties dont chacune est composée de 4 sous-tests qui, sous contrainte de temps, évaluent grâce à un matériel très pictural et donc assez peu verbal, la capacité d'une pensée déductive et logique. Le CFT 20R a été passé en groupe au sein des classes pendant environ 60 minutes.

Une assistante de l'équipe de recherche (dont l'auteur de la présente thèse) a dirigé la passation du test après avoir distribué les cahiers du test et après avoir expliqué la struc-

ture du test, les tâches et la manière d'y répondre. Les limites de temps prévues pour chacune des 8 sous-parties du test ont été respectées à l'aide d'un chronomètre. Les temps de récréation habituelle ont été ignorés afin de pouvoir faire une pause après la moitié du test comme les auteurs du CFT 20R le préconisent. En profitant au maximum de la place disponible dans les classes pour créer de l'espace entre les élèves et en utilisant les versions A et B du CFT 20R pour les élèves voisins, nous avons essayé de décourager toute tentative de copie qui aurait contribué à fausser les résultats. Comme les élèves savaient que le résultat du test ne serait pas communiqué aux enseignants ou aux parents, la grande majorité des élèves prenait le test pour un défi bienvenu, permettant d'évaluer ses propres capacités de pensée logique. Bien évidemment la participation n'était pas obligatoire, mais tous les élèves étaient volontaires. Les scores obtenus au CFT 20R ont ensuite été transformés en valeurs de QI en ajoutant l'âge exact de chaque élève à la formule.

Tableau 21 Indices de qualité du Test CFT 20R

Objectivité	Le manuel du test prévoit des instructions pré-formulées et des indications très précises pour la passation du test ; une grille d'évaluation à superposer facilite le comptage des points obtenus.
Homogénéité	Corrélation entre deux parties du test (split half) $r=0.95$
Validité convergente	Avec d'autres tests d'intelligence (HAWIK, CPM, WIP) entre $r=0.57$ et $r=.73$
Validité factorielle	Tous les sous-tests ont une valence forte sur le facteur d'intelligence générale « g ».
Validité pronostique	Corrélation satisfaisante avec le parcours scolaire 6 à 10 ans plus tard.

Ces indices de qualité sont rapportés dans le manuel du test CFT 20R. Nous les avons jugés satisfaisants.

Tableau 22 Distribution de la variable CFTIQTOT

CFTIQTOT Quotient d'intelligence		CE1/CE2	CM2/6ème
		Moyenne (M)	102.06
	Ecart-type (SD)	14.4	13.89
	Amplitude	82	75

d) *Plurilinguisme (MEHRSPRAS01 / LMUT 1, 2 et 3)*

Auteurs: L'équipe du projet de recherche de la PH Berne (2007)

Instrument: Questionnaire pour les élèves

Nombre d'items: 1

Format de réponse: (0) une langue, (1) deux langues, (2) trois langues

		1 Sprache 1 langues	2 Sprachen 2 langues	3 Sprachen 3 langues
MEHRSPRAS	<b>Sprichst Du 1, 2 oder 3 Sprachen?</b> <i>Est-ce que tu parles 1, 2 ou 3 langues</i>			

Figure 12 Extrait du questionnaire pour les élèves

Grâce à la question « Combien de langues parles-tu ? Laquelle/ Lesquelles ? » que nous posions aux élèves dans le questionnaire, nous pouvions par la suite identifier les élèves qui ne parlaient que l'allemand (ou suisse allemand) et ceux qui parlaient en plus de l'allemand une ou deux autres langues. Nous avons expliqué aux élèves que « parler une langue » signifie : avoir appris une langue dans le contexte familial et comprendre et s'exprimer au moins oralement dans une autre langue que l'allemand (/suisse allemand). Cela excluait, de facto, les langues étrangères apprises à l'école. Par un recodage nous avons ensuite créé la variable dichotomique MEHRSPRAS01 qui identifie les monolingues par le code 0 et les plurilingues (parlant au moins deux langues, selon la définition donnée ci-dessus) par le code 1.

MEHRSPRAS01

(0) une langue —→ (0) monolingues

(1) deux langues —→ (1) plurilingues

(2) trois langues —→ (1) plurilingues

Cette variable MEHRSPRAS01 qui a été utilisée pour des analyses à niveaux multiples

Tableau 23 Distribution de la variable MEHRSPRAS01

Groupes de (Plu-ri)lingues	Monolingues	CE1/CE2		CM2/6ème	
		N=62	21.4%	N=78	30.7%
	Plurilingues	N=228	78.6%	N=176	69.3%

Les données rassemblées sous la variable MEHRSPRAS01 ont été mises en relation avec les informations recueillies concernant les langues d'origine.

e) Statut socioéconomique (POSMU, POSVA)

Source: Müller, Romano (Müller, 1998)

Instrument: Questionnaire pour les élèves

Nombre d'items: 10

Format de réponse: oui / non

Veuillez lire attentivement ce tableau et répondre par oui ou non à chaque ligne, même si actuellement votre mère ou père sont sans emploi ou à la retraite.		Mère		Père	
		Oui	Non	Oui	Non
	<b><i>J'occupe la position suivante (plusieurs réponses possibles)</i></b>				
POSMU1 POSVA1	Homme / Femme au foyer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
POSMU2 POSVA2	Employé(e) sans formation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
POSMU3 POSVA3	Employé(e) avec formation ou sans formation mais avec de bonnes qualifications.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
POSMU4 POSVA4	Chef d'équipe, Chef de petite entreprise avec 0 à 5 employés	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
POSMU5 POSVA5	Employé(e) qualifié responsable d'un domaine spécifique. Chef/Propriétaire d'une PME (6 à 20 employés)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
POSMU6 POSVA6	Cadre dans différents domaines du service (par ex. bureau, hôpital, police)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
POSMU7 POSVA7	Propriétaire d'une PME (250 employés); Cadre moyen; Employés avec des tâches de direction dans une grande entreprise (> 250 employés); Enseignants	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
POSMU8 POSVA8	Cadre supérieur/ Enseignants de lycée ou dans les formations supérieures (université, IUFM) Etudiants d'université.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
POSMU9 POSVA9	Professions libérales avec formations supérieures (par ex. Médecins, Avocats, Architectes.).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
POSMU10 POSVA10	Cadres de très haut niveau. Très hautes tâches exécutives dans des grandes entreprises ou dans l'administration, Membres du gouvernement.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 13 Extrait du questionnaire concernant le statut socioéconomique<sup>37</sup>

En s'appuyant sur les variables POSMU et POSVA nous avons créé la variable HOE-POSELk avec une distribution d'échelle de 1 à 10 (voir tableaux ci-dessous). C'est la variable HOEPOSELk qui a été utilisée pour des analyses à niveaux multiples. La variable HOEPOSELk désigne la position socioéconomique la plus élevée (en allemand:

<sup>37</sup>Traduction française, faite par l'auteur de la présente thèse

HOEchste POSition der ELtern) des parents. Basée sur le système de classification internationale ISCO (International Standard Classification of Occupation) développé par l'Organisation Internationale du Travail (International Labour Organization, 2009), cette échelle contient dix catégories socioprofessionnelles. Dans notre questionnaire, les deux parents doivent répondre pour chaque catégorie si oui ou non, ils ont déjà occupé cette position au cours de leurs parcours professionnels (variable POSMU 1 à 10 et POSVA 1 à 10). Nous avons ensuite créé la variable HOEPOSELk qui résume les réponses des parents en retenant seulement la position la plus élevée qui est (a été) occupée par le père ou la mère. Ainsi une mère ayant occupé une position en tant qu'enseignante (catégorie 7) et le père étant chef d'équipe (catégorie 4), la position socioprofessionnelle la plus élevée qui sera retenue pour la variable HOEPOSELk correspond à la position 7.

Tableau 24 Distribution de l'échelle HOEPOSELk

	CE1/CE2		CM2/6ème	
	N	%	N	%
1=Homme / Femme au foyer	9	2.6	6	2.1
2=Employé(e) sans formation	37	10.9	29	10.4
3=Employé(e) avec formation ou sans formation mais avec de bonnes qualifications.	71	20.9	53	18.9
4=Chef d'équipe, Chef de petite entreprise avec 0 à 5 employés	62	18.2	39	13.9
5=Employé(e) qualifié responsable d'un domaine spécifique. Chef / Propriétaire d'une PME (6 à 20 employés)	58	17.1	52	18.6
6=Cadre dans différents domaines des service (par ex. bureau, hôpital, police)	42	12.4	28	10.0
7=Propriétaire d'une PME (250 employés); Cadre moyen; Employés avec tâches dirigeantes dans une grande entreprise (> 250 employés); Enseignants	29	8.5	39	13.9
8=Cadre supérieur/ Enseignants de lycée ou dans les formations supérieures (université, IUFM) Etudiants	19	5.6	12	4.3
9=Professions libérales avec formations supérieures (par ex. Medecins, Avocats, Architectes.).	11	3.2	11	3.9
10=Cadres de très haut niveau. Très hautes tâches exécutives dans des grandes entreprises ou dans l'administration, Membres du gouvernement.	2	.6	11	3.9
Total	340	100.0	280	100.0
Moyenne (M)	4.59		4.95	
Ecart-type (SD)	1.97		2.2	
Amplitude	9		9	

f) *Temps de lecture extra-scolaire (LESDE)*

Auteurs: L'équipe du projet de recherche de la PH Berne (2007)

Instrument: Questionnaire pour les élèves

Nombre d'items: 1

Format de réponse: (0) Jamais, (1) 30 minutes, (2) Entre 30 minutes et 1 heure (3) Entre 1 et 2 heures ; (4) Plus que 2 heures

	Nie <i>jamais</i>	30 min. <i>30 minutes</i>	30 min. bis 1 h <i>Entre 30 minutes et 1 heure</i>	1 bis 2 h <i>Entre 1 et 2 heures</i>	Mehr als 2 h <i>Plus que 2 heures</i>
Wie lange liest Du pro Tag zu Hause auf Deutsch? <i>Combien de temps par jour lis-tu en allemand à (la maison)<sup>38</sup>?</i>					

Figure 14 Extrait du questionnaire concernant le temps de lecture extra-scolaire

Afin de pouvoir prendre en considération lors d'une comparaison entre les élèves le temps passé à l'exercice de la lecture nous l'avons introduit dans la liste des variables indépendantes. La question posée dans le questionnaire aux élèves sur le temps passé à la lecture chaque jour concerne la lecture de tout genre effectuée à la maison, hors du contexte scolaire.

Tableau 25 Distribution de la variable LESDE

LESDE =Temps de lecture en dehors de l'école		CE1/CE2		CM2/6ème	
	jamais	N=36	10.80%	N=13	4.70%
< 30 minutes	N=166	49.80%	N=99	35.70%	
entre 30 min et 1h	N=65	19.50%	N=88	31.80%	
entre 1h et 2h	N=29	8.70%	N=49	17.70%	
> 2h	N=37	11.10%	N=28	10.10%	

Pour les analyses à niveau multiples, nous avons procédé à une transformation de la variable sur une échelle de z. Ainsi avons-nous créé la variable ZLESDE.

<sup>38</sup> Lors de la réalisation du questionnaire, il a été précisé aux élèves que nous voulions savoir combien de temps ils lisent en moyenne par jour lorsqu'ils ne sont pas à l'école.

### 3.8. Opérationnalisation des variables dépendantes

Nous allons maintenant présenter les trois opérationnalisations différentes des compétences de lecture que nous avons choisies pour évaluer la validité de la SVRT.

a) *Salzburger Lese-Screening SLS 1-4 (ZSLS\_pkt23) et Salzburger Lesescreening SLS 5-8 (Z\_SLSpkt\_5bis6)*

- Auteurs SLS 1-4 : (Mayringer & Wimmer, 2003)  
 Auteurs SLS 5-8 : (Auer et al., 2005)  
 Items : 70  
 Versions : A1, A2, B1, B2 (Différentes versions de phrases entre A et B et différence dans l'ordre de phrases entre 1 et 2)  
 Public ciblé: SLS 1-4 : jeunes lecteurs de la 1<sup>ère</sup> (CP) à la 4<sup>ème</sup> classe (CM1)  
 SLS 5-8 : jeunes lecteurs de la 5<sup>ème</sup> (CM2) à la 8<sup>ème</sup> classe (4<sup>ème</sup>)  
 Méthode : Lire autant de phrases possibles en 3 minutes et juger si le contenu des phrases lues est correct (✓) ou incorrect (x) en entourant les symboles (voir image ci-dessous).

1. Zuckerwatte ist klebrig und schmeckt süß. <input type="radio"/>	<i>La barbe à papa colle et a un gout sucré.</i>
2. Jemand, der Bücher schreibt, ist ein Sänger. <input checked="" type="checkbox"/>	<i>Quand on écrit des livres on est un chanteur.</i>
3. Ein Löwenzahn ist eine fleischfressende Pflanze. <input type="checkbox"/>	<i>Le pissenlit est une plante carnivore</i>
4. Eine Woche hat sieben Tage. <input type="radio"/> x	<i>Une semaine a sept jours.</i>
5. Auf der Tafel schreibt man mit Kreide. <input type="radio"/> x	<i>Sur le tableau noir on écrit avec la craie.</i>

Figure 15 Exemple de question du test SLS 1-4

Le test SLS vise à évaluer un aspect „technique“ précis des compétences en lecture, la rapidité de lecture de phrases simples et décontextualisées. Les travaux des chercheurs autour de Prof. Heinz Wimmer (Landerl & Wimmer, 2000; Wimmer, 1993; Wimmer et al., 1990; Wimmer et al., 1998, 2000) ont montré que la rapidité ou fluidité de la lecture serait un indicateur puissant des compétences en lecture lorsqu'il s'agit de langues peu opaques comme l'espagnol, le néerlandais et l'allemand. La capacité d'une lecture rapide ou fluide repose sur trois pré-requis : (1) un décodage rapide et exact, (2) une automatisation des processus cognitifs impliqués dans la compréhension de la lecture et (3) une intonation expressive lors de la lecture à haute voix (Kuhn et al., 2006; Kuhn &

Stahl, 2003). La troisième composante étant difficile à tester lors d'une évaluation en classe, le SLS permet une évaluation des deux premières composantes en demandant une lecture silencieuse, rapide et exacte du plus grand nombre de phrases isolées possible en trois minutes. Pour assurer que la compréhension des phrases était correcte, les élèves doivent cocher si oui ou non le contenu de la phrase présente un fait correct. Afin de ne pas biaiser l'évaluation de la rapidité et l'automatisation des processus de base par le niveau de connaissances du lecteur et son contexte socioculturel, les phrases du SLS présentent des faits très simples et familiers à tous les enfants comme par exemple : « les vaches sont bleues ». La différence entre le SLS 1-4 (pour les élèves du CP au CM1) et SLS 5-8 (élèves de CM2 au 4<sup>ème</sup>) réside seulement dans la complexité des phrases à lire. La méthode et la limite du temps de 3 minutes ne changent pas. Cela permet une comparaison des résultats du SLS sur plusieurs années scolaires. Il existe au total quatre versions du test SLS. Pour la présente étude nous n'avons utilisé que deux versions, avec des phrases identiques mais d'un ordre différent. Les instructions du manuel du test SLS sont précises en ce qui concerne la passation du test et l'évaluation des résultats.

Tableau 26 Indices de qualité issus du manuel du SLS 1-4

Objectivité	Le manuel du test prévoit des instructions pré-formulées et des indications très précises pour la passation du test ; une grille d'évaluation à superposer facilite l'évaluation du test.
Homogénéité	Cronbach $\alpha = 0.97$
	Corrélation entre deux versions parallèles du test $r = 0.89$
Validité convergente	La corrélation avec la lecture à voix haute de textes est de $r = 0.78$ .

Tableau 27 Distribution des variables ZSLS\_pkt23 et Z\_SLSPkt\_5bis6

Variable	N	Minimum	Maximum	Moyenne	Déviat Standard	Extremité	Curtosis	
SLS	CE 1 t 1	318	2	65	26.35	9.69	0.24	0.89
	CE 2 t 2	338	3	70	37.08	10.18	0.06	0.50
	CM 2 t 1	281	3	59	33.06	8.36	0.37	0.69
	6 <sup>ème</sup> t 2	273	18	70	38.95	8.78	0.65	0.72

Pour les analyses à niveau multiples, nous avons procédé à une transformation des scores obtenus par les élèves (SLS\_Pkt) sur une échelle z. Ainsi nous avons créé les variables ZSLS\_pkt23 (CE1/CE2) et Z\_SLSpkt\_5bis6 (CM2 / 6<sup>ème</sup>).

*b) Lecture de Textes Courts: Test ELFE 1-6 partie 3 (ZTextRoh\_2bis3 et ZTextRoh\_5bis6)*

Auteurs : Lenhard, Wolfgang et Schneider, Wolfgang (2006)

Items: 20

Versions : A et B (différence dans l'ordre des items)

Public ciblé : jeunes lecteurs de la 1<sup>ère</sup> classe (CP) à la 6<sup>ème</sup> classe (6<sup>ème</sup>)

Méthode: Questions à choix multiples

Ein Pferd, das ist ein großes Tier. Es hat auch Beine und zwar vier.	
<i>Un cheval est un animal grand. Il a des jambes au nombre de quatre</i>	
<u>Ein Pferd ...</u>	<i>Un cheval...</i>
<input type="radio"/> ist ganz klein. <i>est tout petit.</i>	<input type="radio"/> hat vier Beine. <i>a quatre pattes..</i>
<input type="radio"/> hat braune Haare. <i>a une crinière brune</i>	<input type="radio"/> frisst Gras. <i>mange du foin.</i>

Figure 16 Exemple de question du test ELFE partie 3 (compréhension de texte).

Nous avons utilisé la partie 1 du test ELFE pour l'évaluation des capacités de décodage. La partie 3 du test ELFE nous a permis d'évaluer les compétences de compréhension de texte très courts. Ces textes n'étaient composés que de 4 à 5 phrases sur des sujets très divers (par ex. les mammoths, les baleines, à la plage).

Après la lecture de chaque petit texte, les élèves devaient répondre à une ou deux questions à choix multiple. Ces questions ne visaient pas une information explicite mais implicite du texte. Afin de pouvoir y répondre, il fallait faire des déductions pour tirer les bonnes conclusions. Comme pour le SLS il s'agissait aussi, pour la partie 3 du test ELFE, d'une lecture sous contrainte de temps. Au-delà de l'évaluation de la capacité à déchiffrer et à comprendre un texte en reliant l'information de plusieurs phrases afin d'en tirer les bonnes conclusions, le ELFE évalue aussi la rapidité de ce processus.

Tableau 28 Indices de qualité issus du manuel de test ELFE 1-6

Objectivité	Le manuel du test prévoit des instructions pré formulées et des indications précises pour la passation et l'évaluation du test ;
Homogénéité Fidélité	Cronbach $\alpha = 0.91$
	Corrélation entre deux résultats de tests à 14 jours d'intervalle (rétest) $r = 0.8$
Validité convergente <sup>39</sup>	avec le test de lecture « Würzburger Leise Leseprobe » (WLLP) $r = 0.48$
	Corrélation avec l'évaluation de la part de l'enseignant des compétences en lecture $r = 0.79$
Validité discriminative	Corrélation avec l'évaluation de l'enseignant des compétences en mathématiques $r = 0.46$ <sup>40</sup> et de $r = 0.49$ pour les compétences en écriture

Tableau 29 Distribution des variables ZTextRoh\_2bis3 et ZTextRoh\_5bis6

Variable			N	Minimum	Maximum	Moyenne	Dévi- ation Standard	Extrenité	Curtosis
ELFE partie 3	CE 1	t 1	313	0	20	6.24	3.59	0.84	1.20
	CE 2	t 2	332	0	20	10.02	4.49	0.20	-0.66
	CM2	t 1	268	4	20	14.89	4.01	-0.54	-0.70
	6ème	t 2	261	6	20	16.55	3.30	-1.16	0.71

Pour les analyses à niveaux multiples, nous avons procédé à une transformation des scores obtenus par les élèves (TextRoh) sur une échelle z. Ainsi nous avons créé les variables ZTextRoh\_2bis3 (CE1/CE2) et ZTextRoh\_5bis6 (CM2 / 6<sup>ème</sup>).

<sup>39</sup> Les corrélations indiquées dans ce tableau concernent des classes de CM1.

<sup>40</sup> Les indicateurs de la validité convergente et la validité discriminative sont très semblables, ce qui n'est pas souhaité. Les auteurs du ELFE ne commentent pas cet élément. Dans le cadre de la présente thèse nous le voyons comme une mise en garde ; la façon dont le ELFE partie 3 évalue la compréhension de texte est inhabituel et se distingue d'autres tests de la compréhension de texte. Il utilise très peu de phrases et met l'accent sur les capacités de déduction. On peut supposer que ces dernières jouent également un rôle important dans les compétences en mathématique.

c) HarmoS Lecture Textes Complexes (Z\_Lesen\_2bis3\_01 et Z\_Lesen\_5bis9\_01)

Auteurs : HarmoS\_L1 Consortium (2007)

Versions : A et B (différence dans l'ordre des items)

Public ciblé : enfants de la 2<sup>ème</sup> classe (CE1) et de la 6<sup>ème</sup> classe (6<sup>ème</sup>)

Format : questions ouvertes, questions vrai / faux et QCM

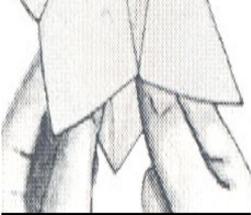
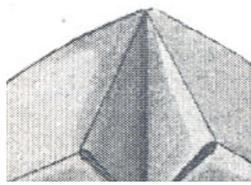
	<p>Frage 4 / Questions 4          Wofür sind die Bilder auf dem Blatt? Kreuz die richtige Antwort an.  <i>Pourquoi y a-t-il des dessins sur la feuille de texte? Coche la bonne réponse.</i></p> <p><input type="checkbox"/> Die Bilder sollen das Blatt einfach schöner machen.  <i>Les dessins rendent la feuille plus jolie.</i></p> <p><input type="checkbox"/> Die Bilder helfen, den Text besser zu verstehen.  <i>Les dessins aident à mieux comprendre le texte.</i></p> <p><input type="checkbox"/> Die Bilder zeigen, wie der Frosch lebt.  <i>Les dessins montrent comment vie la grenouille.</i></p> <p><input type="checkbox"/> Die Bilder sollen Kinder zum Lachen bringen.  <i>Les dessins doivent faire rire les enfants.</i></p>
	

Figure17 Extrait du test HarmoS-L1 pour la compréhension de texte (CE1/CE2)

Nous avons déjà décrit la démarche du groupe de HarmoS-L1 pour la construction de tâches afin d'évaluer les compétences en langue scolaire des élèves en Suisse. Outre les tâches pour la compréhension orale nous avons utilisé les tâches pour la compréhension de texte. Ces tâches comprenaient des textes authentiques de sujets et de genre divers. Comme pour les tâches de compréhension orale, ce sont les critères de la validité des tâches, le caractère discriminatoire des items, la variété des genres textuels, la difficulté des tâches, les aspects ciblés (planifier, situer, réaliser, évaluer, réparer) et le temps nécessaire qui conditionnaient le choix des tâches pour la compréhension de texte.

Nous avons ainsi choisi quatre tâches différentes dont une pour le niveau CE1/CE2 (les instructions de bricolage pour une « Grenouille » en carton) et trois tâches pour le niveau CM2/ 6<sup>ème</sup>, dont une recette de cuisine pour une compote de pomme (recette 1), une narration fantaisiste d'un dialogue entre légumes (recette 2) et les conseils d'utilisation de la notice accompagnant l'huile de millepertuis.

Comme pour les tâches de compréhension orale, nous avons créé deux versions de chaque tâche en changeant l'ordre des questions. Pour prévenir un effet de répétition de tâches, nous avons créé des tâches parallèles pour la deuxième vague du recueil de données en 2008. Ces tâches parallèles contenaient exactement les mêmes questions mais des textes différents sur exactement les mêmes sujets. Ainsi la deuxième année, les élè-

ves de CE1/CE2 ont été invités à lire les instructions de bricolage pour un « Lapin » en carton. Au niveau des CM2/ 6<sup>ème</sup>, la recette de cuisine concernait une compote de prunes et les instructions de la notice concernaient une autre huile pharmaceutique (huile de PoHo). Les tableaux 30 et 31 résument les caractéristiques des différentes tâches pour la compréhension de texte.

Tableau 30 Les exercices d'compréhension orale pour les CM2/6ème

Tâche / Parallèle	Genre	Question	Aspect	Description
Huile pharmaceutique " „millepertuis“, „PoHo“	e x p o s i t i f	1	rea6	trouver de l'information explicite
		2	sit2	identifier la fonction du texte
		3	sit3	identifier la situation communicative (contexte et destinataire du message)
		4	rea6	trouver de l'information explicite
		5	pla3	choisir une stratégie et un but de lecture
		6	rea6	trouver de l'information explicite
		7	rea6	trouver de l'information explicite
		8	rea6	trouver de l'information explicite
		9	rea6	trouver de l'information explicite
		10	rea6	trouver de l'information explicite
		11	sit2	identifier la fonction du texte
Recette 1 "Compote de pommes" / "C. de prunes"	f	1	rea5	trouver de l'information implicite / deduction
		2	rea5	trouver de l'information implicite / deduction
		3	rea5	trouver de l'information implicite / deduction
		4	pla1	organiser sa lecture
		5	sit2	identifier la fonction du texte
		6	rea6	trouver de l'information explicite
Recette 2 "Dialogue de légumes"	narratif	1	rea6	trouver de l'information explicite
		2	rea2	activer des connaissances linguistiques
		3	rea6	trouver de l'information explicite
		4	pla1	organiser sa lecture
		5	rea5	trouver de l'information implicite / deduction
		6	eva3	identifier l'objectif de l'auteur, se mettre à distance, réfléchir sur le contenu
		7	sit2	identifier la fonction du texte
		8	eva3	identifier l'objectif de l'auteur, se mettre à distance, réfléchir sur le contenu
		9	rea5	trouver de l'information implicite / deduction
		10	rea2	activer des connaissances linguistiques
		11	rea6	trouver de l'information explicite
		12	sit2	identifier la fonction du texte

Tableau 31 Les exercices d'compréhension orale pour les CE1/CE2

Instructions de bricolage "Grenouille" / "Lapin"	e x p o s i t i f	1	sit2	identifier la fonction du texte
		2	sit3	identifier la situation communicative (contexte et destinataire du message)
		3	pla1	organiser sa lecture
		4	pla2	identifier les caractéristiques formelles du textes
		5	pla2	identifier les caractéristiques formelles du textes
		6	rea6	trouver de l'information explicite
		7	rea5	trouver de l'information implicite / deduction
		8	rea5	trouver de l'information implicite / deduction
		9	rea4	mettre en lien le contenu du texte avec ses connaissances préalables

Après une explication du déroulement du test, les enseignants distribuèrent les cahiers de tâches contenant les textes et les questions. Les élèves lisaient chacun en silence les textes avant de répondre sous contrainte de temps aux questions, tout en gardant les textes sous les yeux.

Comme pour les tâches de compréhension orale, nous avons examiné la qualité des tâches non-standardisées de HarmoS-L1 pour la compréhension de textes à l'aide des analyses de Rasch réalisées avec le logiciel WINSTEPS (Linacre 2006).

Tableau 32 Indices de qualité

HarmoS Compréhension texte	CE1/CE2		CM2/6ème	
	Real	Model	Real	Model
<i>Root Mean Square Standard Error (RMSE)</i> <sup>41</sup>	0.77	0.72	0.4	0.38
<i>Adjusted Standard Deviation (ADJSD)</i> <sup>42</sup>	1.13	1.17	0.72	0.73
<i>Separation</i> <sup>43</sup>	1.46	1.63	1.8	1.92
<i>Person Reliability</i>	0.68	0.73	0.76	0.79

L'indicateur « Person Reliability » est, dans ces calculs de WINSTEPS, équivalent au « Crohnbachs  $\alpha$  ». Il nous rapporte une fiabilité satisfaisante des tâches pour l'évaluation de la compréhension de textes dans les deux groupes CE1/CE2 (0.68-0.73) et CM2/6<sup>ème</sup> (0.76 -0.79).

Le tableau 32 ci-dessous nous montre les résultats des analyses de composantes principales. La partie de la variance expliquée (*variance explained by measures*) est très grande par rapport à la variance non expliquée (*unexplained variance*). Aucun contraste ne dépasse un « *Eigenvalue* » de 2, nous pouvons donc en conclure que les items du test évaluent tous la même variable latente : la compréhension de texte.

Tableau 33 Analyses à composantes principales (HarmoS Compréhension de texte)

HarmoS Compréhension Texte	CE1/CE2		CM2/6ème	
<i>Standardized Residual Variance (Eigenvalue units)</i>				
<i>Total variance in observations</i>	48.4	100%	71.6	100%
<i>Variance explained by measures</i>	38.4	79.30%	39.6	55.30%
<i>Unexplained variance (total)</i>	10	20.70%	32	44.70%
<i>Unexplained variance in 1st contrast</i>	1.4	2.90%	2	2.80%
<i>Unexplained variance in 2nd contrast</i>	1.3	2.70%	1.7	2.30%
<i>Unexplained variance in 3rd contrast</i>	1.3	2.70%	1.6	2.30%
<i>Unexplained variance in 4th contrast</i>	1.2	2.40%	1.5	2.10%
<i>Unexplained variance in 5th contrast</i>	1.1	2.30%	1.4	2.00%

<sup>41</sup> RMSE est la racine carrée de la moyenne de la variance d'erreur dans l'échantillon.

<sup>42</sup> ADJ. S.D. est l'écart-type „ajusté“ c'est à dire que c'est l'écart-type de l'échantillon  $(ADJ.S.D.)^2 = (S.D. of MEASURE)^2 - (RMSE)^2$

<sup>43</sup> La SEPARATION rapporte la proportion entre ADJ.S.D l'écart-type ajusté et RMSE, l'écart-type des erreurs. SEPARATION<sup>2</sup> est la proportion « signal-to-noise » c'est à dire la proportion entre la vraie variance et la variance d'erreur.

Tableau 34 Indicateurs de « *fit* » et de difficulté (CM2/6<sup>ème</sup>)

Compréhension de texte CM2/6ème		<i>Infit</i>		<i>Outfit</i>
Numéro d'item	<i>measue</i>	MNSQ	MNSQ	MNSQ
Recette 9	0.11	1.57		1.56
Huile de millepertuis 3	0.86	1.11		1.43
Compote de pomme 1	-3.4	1.05		1.33
Huile de millepertuis 7	0.72	1.32		1.29
Recette 11	-0.43	1.15		1.18
Huile de millepertuis 10	-64	1.05		1.12
Compote de pomme 2	-0.99	1.06		1.1
Huile de millepertuis 11	1.35	0.92		1.08
Huile de millepertuis 9	0.4	1.02		1.06
Huile de millepertuis 8	0.31	1.03		1.05
Recette 10	-0.26	1.04		0.98
Huile de millepertuis 5	-0.46	1.03		1.03
Recette 1	0.39	0.99		1.03
Recette 12	2.03	0.96		1.02
Compote de pomme 3	-2.27	0.91		1.02
Huile de millepertuis 2	0.78	1.01		1
Compote de pomme 4	0.04	0.87		1.01
Compote de pomme 6	-0.56	1		1.01
Recette 6	-1.11	1		0.95
Compote de pomme 5	-1.22	0.99		0.97
Recette 3	2.2	0.99		0.78
Huile de millepertuis 4	1.38	0.97		0.96
Recette 4	0.37	0.95		0.92
Recette 8	-0.35	0.93		0.91
Huile de millepertuis 1	-0.22	0.93		0.91
Recette 7	-2.39	0.91		0.73
Huile de millepertuis 6	1.77	0.87		0.78
Recette 5	-0.82	0.8		0.77
Recette 2	0.08	0.68		0.77

Les tableaux 33 et 34 ci-contre nous rapportent le « *infit* » et « *outfit* » de chaque item des tâches visant la compréhension de texte, conçu par le groupe de HarmoS-L1. Nous voyons qu'à l'exception de l'item 9 du texte Recette, toutes les estimations du « *infit* » et « *outfit* » varient entre 0.5 et 1.5. Ils nous indiquent ainsi que les données obtenues rentrent dans le modèle de Rasch. Les estimations de la difficulté des items varient suffisamment, nous indiquant ainsi que nos tâches de lecture de HarmoS-L1 nous permettent de différencier les niveaux de compétence en compréhension de texte des élèves de notre échantillon.

Tableau 35 Indicateurs de « *fit* » et de difficulté (CE1/CE2)

Compréhension de texte CE1/CE2		<i>Infit</i>		<i>Outfit</i>
Item number	Difficulté	MNSQ	MNSQ	MNSQ
Grenouille 9	1.88	1.17		1.34
Grenouille 8	2.08	1.28		1.12
Grenouille 3	0.76	1.03		0.95
Grenouille 7	-1.23	0.88		1.01
Grenouille 2	-0.09	0.99		1
Grenouille 5	1.24	0.93		0.96
Grenouille 4	0.28	0.86		0.82
Grenouille 6	-1.84	0.8		0.74
Grenouille 1	-0.17	0.71		0.6

Le tableau 36 montre la distribution des données obtenues dans les deux groupes à l'aide du test HarmoS-L1 compréhension de texte.

Tableau 36 Distribution des variables Z\_Lesen\_2bis3\_01 et Z\_Lesen\_5bis9\_01

Variable			N	Minimum	Maximum	Moyenne	Déviati on Standard	Extremité	Curtosis
HarmoS Compréhension de textes	CE 1	t 1	329	-4	3	.06	1.40	-0.74	0.69
	CE 2	t 2	343	-2	5	1.45	1.37	-0.23	-0.09
	CM 2	t 1	282	-3	2	-.35	0.83	-0.22	0.02
	6ème	t 2	269	-3	3	.56	0.98	-0.16	0.89

Ces données obtenues pour chaque élève sous la forme de « logits » ont été transformées sur une échelle z. Ainsi avons-nous obtenu les variables Z\_Lesen\_2bis3\_01 et Z\_Lesen\_5bis9\_01.

#### *Résumé des variables opérationnalisées*

*Nous pouvons distinguer parmi les variables présentées au sein des chapitres 3.7 et 3.8 :*

- *Les variables qui servent à opérationnaliser les deux composantes de l'équation  $L=D \times CO$  de la SVRT. Le test ELFE 1-6 partie 1 mesure la capacité de lecture de mots. Il est dans notre modèle de la SVRT l'indicateur pour les compétences en décodage (D). Les différents exercices créés par le groupe de travail HarmoS\_L1 pour mesurer la compréhension orale, nous servent d'indicateur pour la deuxième composante CO de la SVRT.*
- *Le Culture Fair Test (CFT 20-R) de Weiss, basé sur les travaux de Cattell, nous fournit une mesure du QI qui essaie de faire abstraction des compétences en langue. Il nous permet de tester la force prédictive d'un modèle modifié de la SVRT qui intègre le QI.*
- *Le questionnaire pour les élèves, développé par l'équipe de projet de la HEP dont l'auteur faisait partie, nous fournit des données concernant le statut socio-économique. De plus, il nous a permis de calculer le temps de lecture extra-scolaire de chaque élève et de savoir si l'élève était monolingue ou plurilingue. Les trois variables indépendantes qui en résultent, nous ont permis de distinguer des éventuels groupes et d'éventuels biais dans notre échantillon et de tester séparément la validité de la SVRT dans le groupe des monolingues et des plurilingues.*
- *Les tests de SLS, ELFE 1-6 partie 3 et HarmoS pour la compréhension de texte, nous ont permis de créer trois variables dépendantes, opérationnalisant chacune un aspect spécifique des compétences en lecture.*

## 4 Résultats

Conformément à l'ordre de nos questions de recherche, nous allons d'abord présenter les résultats des analyses multi-niveaux qui nous permettent de juger si les prédictions clés de la « *Simple View of Reading Theory* » (SVRT) concernant l'impact des deux facteurs décodage (D) et compréhension orale (CO) sont validées au sein de notre échantillon.

La présentation des résultats issus des analyses multi-niveaux avec le logiciel WINSTEPS, suit l'ordre des hypothèses formulées au chapitre 2. Pour chacune de nos hypothèses nous présenterons d'abord une analyse des coefficients de régression et ensuite une analyse des coefficients de détermination.

### 4.1 Validation de l'importance des deux composantes D et CO

#### 4.1.1 Analyse des coefficients de régression

Il s'agit dans un premier temps de vérifier l'hypothèse H1a qui postule que les facteurs décodage (D) et compréhension orale (CO) contribuent de manière significative à l'explication de la variance des compétences en lecture. Pour cela, nous allons présenter les résultats obtenus avec le logiciel WINSTEPS sous forme de tableau Excel (voir tableau 37). Ces tableaux contiennent une première colonne à gauche indiquant entre autres le nom des variables explicatives qui sont les composantes fixes de notre modèle multi-niveaux, les variances résiduelles et les covariances faisant partie des composantes aléatoires.

Les colonnes qui suivent présentent les résultats concernant nos trois variables réponses qui sont les compétences de lecture mesurées selon les trois tests SLS, ELFE et Har-moS-L1. Elles indiquent les coefficients de régression (de  $\beta_0$  à  $\beta_5$ ) de chaque variable explicative et l'erreur standard (standard error SE), indispensable pour calculer la significativité des coefficients de régression. En bas de ces trois colonnes nous trouvons les variances résiduelles ainsi qu'une information concernant la qualité du modèle ( $2 \cdot \log\text{likelihood}$ ). A la fin de chaque colonne est indiqué le nombre d'unités qui ont été incluses dans les analyses.

Tableau 37 Analyses multi-niveaux du modèle de la SVRT en CE1 / CE 2

Variables réponses	SLS		ELFE		HarmoS-L1	
	ZSLS_Pkt23_01		ZTextRoh_2bis3		Z_Lesen_2bis3	
Variables explicatives (Parties fixes)	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.091	0.065	-0.112*	0.053	-0.214	0.115
Temps (compétences t1-t2)	0.208***	0.06	0.21***	0.059	0.377**	0.115
Décodage (D)	0.766***	0.041	0.535***	0.044	0.255***	0.059
Interaction D & Temps	-0.162***	0.044	-0.032	0.051	0.082	0.069
Compréhension Orale (CO)	0.071*	0.032	0.187***	0.037	0.339***	0.047
Interaction CO & Temps	0.013	0.039	0.034	0.05	-0.03	0.06
Variances (Parties aléatoires)						
Niveau III: Classes						
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.049	0.022	0.022	0.014	0.176	0.071
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.026	0.015	0.027	0.01	-0.114	0.059
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.028	0.016	0.008	0.013	0.147	0.066
Niveau II: Individus						
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.218	0.019	0.31	0.027	0.395	0.034
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.105	0.016	-0.157	0.026	-0.304	0.037
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.251	0.023	0.492	0.044	0.628	0.057
-2*loglikelihood:	774.828		1072.657		1147.904	
N° d'Unités Niv. III (Classes)	17		17		17	
N° d'Unités Niv. II (Individus)	318		318		317	
N° d'Unités Niv. I (Temps)	571		580		570	

\* le resultat est significatif avec un niveau d'erreur de  $\alpha = 5\%$ ; \*\*  $\alpha = 1\%$ ; \*\*\*  $\alpha=0.1\%$

Nous voyons dans le tableau 37 par exemple que pour l'analyse concernant le test SLS les analyses se basent sur 318 individus en 17 classes de CE1, avec deux moments de recueil de données en CE1 et CE2 ; nous avons au total 571 valeurs intégrées dans les analyses<sup>44</sup>. Comme nous avons transformé les valeurs de nos variables réponses et explicatives sur une échelle de z, tous les  $\beta$  et SE sont exprimés dans cette unité de mesure z (avec une moyenne de 0 et un écart type de 1).

La première ligne du tableau présente le coefficient de la constante ( $\beta_0$ ). Nous avons vu au chapitre précédent (note en bas de page 76) que la constante représente la valeur attendue de la variable réponse Y pour un individu ayant un score de 0 (ou de la moyenne) sur les variables explicatives X intégrées dans le modèle. La constante nous indique donc le niveau de compétence en lecture mesuré par le test de SLS, le test ELFE et le test de HarmoS des élèves en CE1 ayant des compétences de décodage et de compréhension orale exactement dans la moyenne (d'une valeur de  $z = 0$ .) Ces élèves ont

<sup>44</sup>Idéalement nous aurions dû avoir deux valeurs pour chaque élève (une du CE1 et une du CE2). Ainsi nos unités au niveau du temps devons être de  $318 \times 2 = 636$ . Dans la réalité les deux échantillons ne sont pas parfaitement identiques à cause de l'absence de certains élèves le jour du test, des élèves ayant quitté l'établissement scolaire mais aussi des élèves nouveaux qui n'étaient pas dans la classe en CE1.

une compétence de lecture de phrases (test SLS) de  $\beta = -0.091$ , leur compétence en lecture de textes courts (test ELFE) est de  $\beta = -0.112$  et leur compétence en lecture de textes authentiques (test HarmoS-L1) est de  $\beta = -0.214$ .

La variable temps indique le développement des compétences en lecture en affichant en valeurs  $z$  de combien la compétence en lecture (SLS ; ELFE ou HarmoS-L1) a en moyenne changée entre la première (CE1) et la deuxième vague (CE2) dans le processus de recueil des données. Selon le tableau 37, c'est dans la lecture des textes authentiques (HarmoS-L1) que les élèves font en moyenne le plus de progrès ( $\beta = 0.377$ ). Ce progrès est significatif, tout comme celui que l'on constate dans les deux autres tests. Regardons ensuite les deux variables explicatives qui concernent spécifiquement notre hypothèse sur la *Simple View of Reading Theory (SVRT)*. Le facteur du « décodage » (D) a un impact très significatif sur les trois variables réponses. Il est particulièrement élevé pour le test de SLS avec un coefficient de régression de  $\beta = 0.766$ , ce qui est équivalent à environ trois quart d'un écart type. Cela signifie que pour un élève qui n'a pas une capacité de décodage de 0 (valeur utilisée pour le calcul de la constante) mais de 1 (en valeur  $z$  1 = 1 écart type) sa compétence en lecture de phrases n'est pas de  $-0.091$  mais de  $-0.091 + 0.766 = 0.675$  points. L'impact sur la lecture de textes courts (ELFE) est un peu moins élevé ( $\beta = 0.535$ ), il est équivalent à un demi écart type. Pour la lecture de textes complexes et authentiques (test de HarmoS-L1), l'impact du décodage est toujours significatif mais plus faible avec seulement  $\beta = 0.255$ .

Lorsque nous regardons l'impact du facteur de la compréhension orale (CO), nous observons le phénomène inverse. C'est dans la lecture de textes authentiques (test de HarmoS-L1) que le coefficient de régression de la compréhension orale est le plus élevé  $\beta = 0.339$  tandis qu'il est le plus faible pour le test de lecture de phrases (SLS) avec  $\beta = 0.071$ .

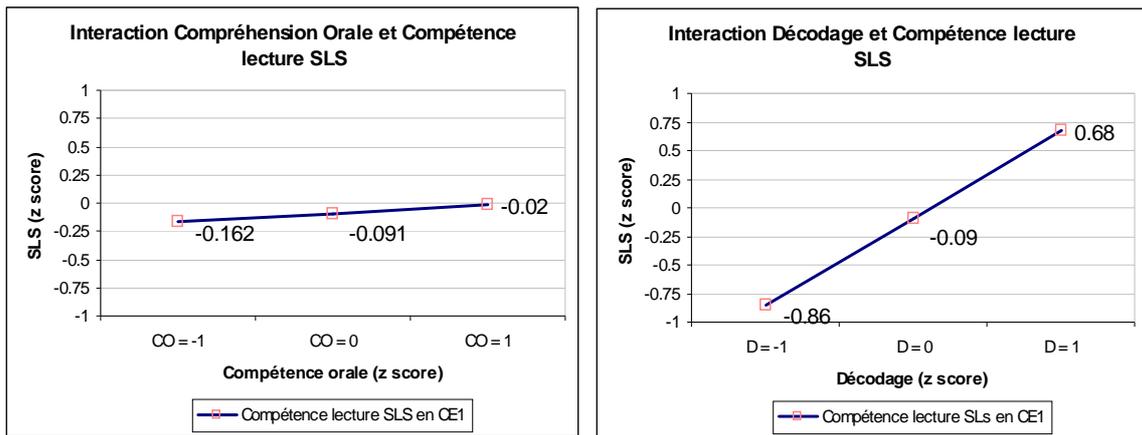


Figure 18 Effets d'interactions entre les facteurs D et CO pour le SLS

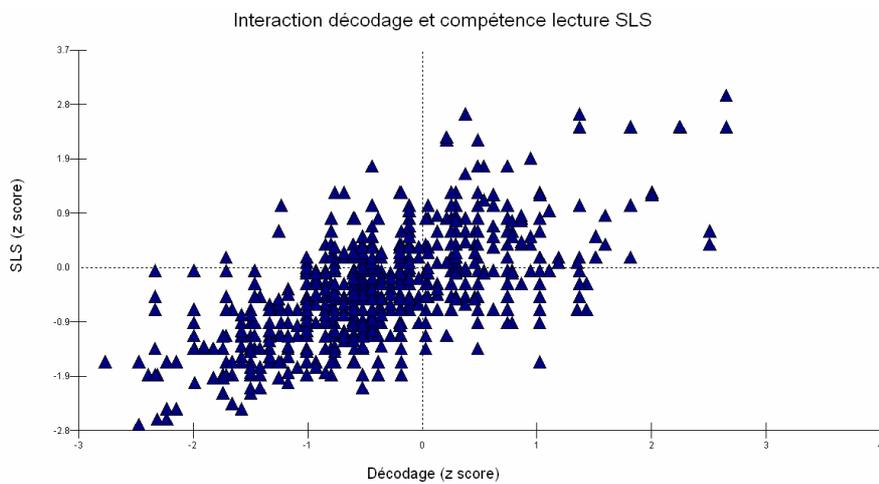


Figure 19 Interaction entre D et les compétences de lecture selon le test SLS

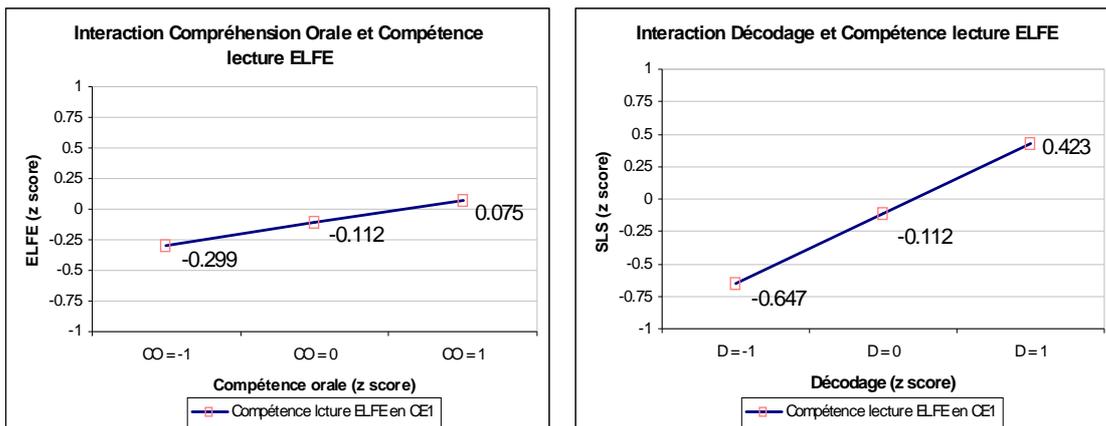


Figure 20 Effets d'interactions entre les facteurs D et CO et les résultats du ELFE

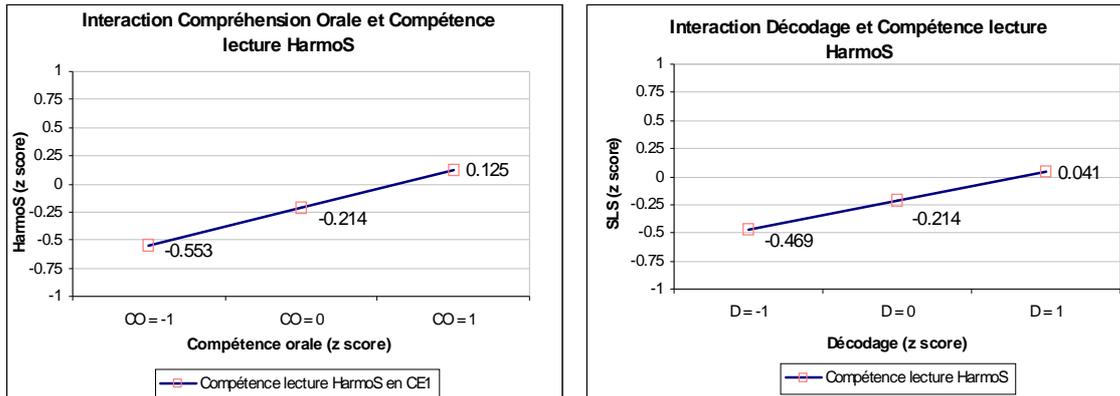


Figure 21 Effets d'interactions entre D et CO et les résultats du test HarmoS

Les Figures 18 à 21 ci-dessus présentent une illustration des droites de régression résultant des relations entre les deux facteurs décodage et compréhension orale pour chacun de nos trois tests de lecture. La figure 19 illustre la relation qui existe entre le décodage et les compétences de lecture selon le test SLS lorsque l'on affiche chaque individu sur le graphique au lieu de résumer la relation par une droite de régression.

Les coefficients de régression ne sont qu'un indicateur parmi d'autres qui peuvent nous renseigner sur la question de l'impact des facteurs Décodage et Compréhension Orale sur les compétences en lecture dans notre modèle de *SVRT*. Ainsi, avant de tirer des conclusions, nous allons maintenant regarder les coefficients de détermination.

#### 4.1.2 Analyse des coefficients de détermination des composantes D et CO

Le rapport entre le modèle abstrait et théorique et la réalité empirique est la question essentielle lorsque nous souhaitons mettre à l'épreuve le modèle de la *Simple View of Reading Theory*. Le coefficient de détermination  $R^2$  nous renseigne sur ce rapport en nous indiquant le rapport entre la variabilité des résidus (qu'on souhaite faible) sur la variabilité de notre variable-reponse (variable dépendante = Y). Lorsque le modèle théorique est une bonne représentation des données empiriques nous obtenons des points de mesure très rapprochés à la droite de régression et une variabilité des résidus faible par rapport à la variabilité totale de Y.  $R^2$  est ainsi un indicateur du pouvoir explicatif de notre modèle construit sur la base de la *SVRT* (Bressoux, p. 82). Dans le cadre des analyses à niveau multiple nous calculons un « pseudo »  $R^2$  qui correspond à l'écart entre la variance estimée par le modèle 0 (modèle « vide ») ne contenant aucun facteur explicatif et le modèle plus complet, divisée ensuite par la variance estimée par le modèle «vide » (cf. tableau 38.1, cadres en pointillés gris).

En effet, lorsque nous regardons les tableaux 38.1 à 40.2 sur les trois pages suivantes, nous pouvons observer pour chacune de nos trois variables réponses, l'évolution des variances et des valeurs de la déviance ( $2 \times \text{loglikelihood}$ ) depuis le modèle 0 (modèle « vide ») au modèle 1.2 qui contient les facteurs D et CO.

En ce qui concerne les compétences en lecture rapide de phrases (test SLS), nous constatons dans le tableau 38.1 au niveau des classes, une baisse considérable dans les variances résiduelles de la constante et du temps. Le tableau 38.2 montre que le pouvoir d'explication augmente de 76.7% entre le modèle 0 (« vide ») et le modèle contenant le facteur du décodage. L'introduction du facteur de la compréhension orale l'augmente encore une fois de 9.22 %. Au total, entre le modèle « vide » et le modèle 1.2, le pouvoir d'explication augmente de 78.88%. La différence entre les classes dans les compétences de lecture rapide de phrases peut donc être expliquée en très grande partie par les différences en compétences de décodage. La prise en compte des différences de compétences en compréhension orale augmente encore la partie de la variance expliquée.

En ce qui concerne les différences entre les classes dans le développement des compétences en lecture de phrases au cours d'une année, nous observons également une forte baisse des variances résiduelles et une augmentation de la variance expliquée de 51.72% lorsque les facteurs D et CO sont introduits (voir tableau 38.2 variance résiduelle de  $\sigma_1$ ). Au niveau des individus, les résultats sont semblables en ce qui concerne la variance résiduelle de la constante ; au total, le pouvoir d'explication augmente de 60.79% suite à l'introduction des facteurs D et CO. Cependant ici l'apport spécifique de deux facteurs est encore plus important car le « pseudo R2 » n'augmente que de 1.8% suite à l'introduction de la compréhension orale au sein du modèle 1.2.

Tableau 38.1 Evolution des modèles multi niveaux hiérarchisés relatifs au SLS

Variables réponses	SLS :		SLS :		SLS :	
	Modèle 0 (vide)		Modèle 1.1		Modèle 1.2	
Variables explicatives (Parties fixes)	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.541	0.125	-0.096	0.067	-0.091	0.065
Temps (compétences t1-t2)	0.942	0.067	0.221	0.062	0.208	0.06
Décodage (D)			0.79	0.039	0.766	0.041
Interaction D & Temps			-0.166	0.042	-0.162	0.044
Compréhension Orale (CO)					0.071	0.032
Interaction CO & Temps					0.013	0.039
Variances (Parties aléatoires)						
Niveau III: Classes						
	(0.232-0.054)/0.0232=76.72%					
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.232	0.091	0.054	0.023	0.049	0.022
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.026	0.035	-0.026	0.016	-0.026	0.015
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.058	0.025	0.031	0.017	0.028	0.016
Niveau II: Individus						
	(0.556-0.222)/0.556=60.07%					
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.556	0.047	0.222	0.019	0.218	0.019
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.073	0.022	-0.106	0.016	-0.105	0.016
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.203	0.019	0.251	0.023	0.251	0.023
-2*loglikelihood:	1085.499	301.902	783.597	8.769	774.828	
		df=2		df=2		
Test de Chi2		0.000***		0.012*		
N° d'Unités Niv. III (Classes)	17		17		17	
N° d'Unités Niv. II (Individus)	318		318		318	
N° d'Unités Niv. I (Temps)	571		571		571	

Tableau 38.2 Evolution des taux de variances expliquées relatif au SLS<sup>45</sup>

Niveau III: Classes	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	76.72%	78.88%	9.26%
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	46.55%	51.72%	9.68%
Niveau II: Individus			
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	60.07%	60.79%	1.80%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	-23.65%	-23.65%	0.00%

<sup>45</sup>Pour vérifier l'asymétrie dans l'apport spécifique de chacun des facteurs D et CO nous avons procédé à des analyses de modèle multi-niveaux avec des modèles constitués d'un ordre inversé des variables. Au modèle 1.1Inversé c'est donc la compréhension orale qui est introduite et le décodage est introduit ensuite. Les résultats de ces analyses sont présentés au sein de l'annexe 3.

Tableau 39.1 Evolution des modèles multi-niveaux hiérarchisés de l'ELFE

Variables réponses	ELFE:		ELFE:		ELFE:	
	Modèle 0 (vide)		Modèle 1.1		Modèle 1.2	
Variables explicatives (Parties fixes)	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.461	0.09	-0.12	0.061	-0.112	0.053
Temps (compétences t1-t2)	0.841	0.062	0.242	0.066	0.21	0.059
Décodage (D)			0.601	0.044	0.535	0.044
Interaction D & Temps			-0.04	0.049	-0.032	0.051
Compréhension Orale (CO)					0.187	0.037
Interaction CO & Temps					0.034	0.05
Variations (Parties aléatoires)						
Niveau III: Classes						
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.105	0.047	0.034	0.019	0.022	0.014
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	0.051	0.025	0.031	0.013	0.027	0.01
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.038	0.022	0.019	0.017	0.008	0.013
Niveau II: Individus						
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.546	0.046	0.336	0.029	0.31	0.027
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.088	0.031	-0.155	0.027	-0.157	0.026
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.438	0.039	0.489	0.043	0.492	0.044
-2*loglikelihood:	1277.362	168.124	1109.238	36.581	1072.657	
		df=2		df=2		
Test de Chi <sup>2</sup>		0.000***		0.000***		
N° d'Unités Niv. III (Classes)	17		17		17	
N° d'Unités Niv. II (Individus)	318		318		318	
N° d'Unités Niv. I (Temps)	580		580		580	

Tableau 39.2 Evolution des taux de variances expliquées relatifs à l'ELFE

Niveau III: Classes	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	67.62%	79.05%	35.29%
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	50.00%	78.95%	57.89%
Niveau II: Individus			
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	38.46%	43.22%	7.74%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	-11.64%	-12.33%	-0.61%

Tableau 40.1 Evolution des modèles concernant le test de HarmoS

Variables réponses	HarmoS:Z_Lesen_2b		HarmoS:Z_Lesen_2b		HarmoS:Z_Lesen_2b	
	Modèle 0		Modèle 1.1		Modèle 1.2	
Variables explicatives (Parties fixes)	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.451	0.157	-0.231	0.13	-0.214	0.115
Temps (compétences t1-t2)	0.843	0.113	0.431	0.118	0.377	0.115
Décodage (D)			0.379	0.06	0.255	0.059
Interaction D & Temps			0.039	0.067	0.082	0.069
Compréhension Orale (CO)					0.339	0.047
Interaction CO & Temps					-0.03	0.06
Variances (Parties aléatoires)						
Niveau III: Classes						
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.374	0.142	0.229	0.091	0.176	0.071
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.133	0.081	-0.138	0.068	-0.114	0.059
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.166	0.072	0.153	0.068	0.147	0.066
Niveau II: Individus						
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.532	0.046	0.469	0.041	0.395	0.034
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.244	0.039	-0.313	0.04	-0.304	0.037
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.59	0.054	0.647	0.059	0.628	0.057
-2*loglikelihood:	1306.967	78.767	1228.2	80.296	1147.904	
		df=2		df=2		
Test de Chi <sup>2</sup>		0.000***		0.000***		
N° d'Unités Niv. III (Classes)	17		17		17	
N° d'Unités Niv. II (Individus)	317		317		317	
N° d'Unités Niv. I (Temps)	570		570		570	

Tableau 40.2 Evolution des taux de variances expliquées relatifs au test de HarmoS

Niveau III: Classes	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	38.77%	52.94%	23.14%
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	7.83%	11.45%	3.92%
Niveau II: Individus			
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	11.84%	25.75%	15.78%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	-9.66%	-6.44%	2.94%

Les résultats changent lorsque nous regardons l'explication de la variance résiduelle du temps au niveau des individus. En effet, on observe des gains de variance expliquée négatifs dus à une augmentation de la variance résiduelle. Cela se reproduit dans le cas des analyses du test ELFE (tableau 39.2) et du test HarmoS (tableau 40.2). Nous avons discuté de la possibilité d'une telle incohérence au chapitre 3.6. Selon Singer et Willet (2003), le problème d'une variance résiduelle qui augmente alors qu'on introduit des variables explicatives au sein du modèle peut se présenter lorsque la répartition de la variance résiduelle entre les niveaux est très importante (Bressoux, 2010).

Ceci est effectivement le cas dans nos trois exemples d'analyses présentés ci-dessus : à chaque fois la variance résiduelle du facteur « temps » est bien plus grande au niveau des individus qu'au niveau des classes. Cela nous indique que la modélisation de l'évolution des compétences en lecture au cours d'une année est plus difficile à réaliser que celle du niveau initial des compétences en lecture. Cela nous rappelle aussi que l'interprétation des coefficients de détermination « pseudo-R » des modèles multi-niveaux requiert de la prudence. Nous allons donc devoir nous restreindre dans l'interprétation des coefficients de détermination obtenus, à l'explication des compétences en lecture au moment du premier recueil des données c'est-à-dire en CE1 et laisser de côté l'évolution de ces compétences entre le CE1 et le CE2.

Concernant notre deuxième variable réponse - la compétence de lecture rapide de textes très courts selon le test ELFE - le tableau 39.2 montre que la partie de variance expliquée au niveau des classes est de 78.95 % (Temps) et de 79.05 % (Constante) lorsque les deux facteurs D et CO sont introduits. Au niveau des individus le pouvoir explicatif du modèle 1.2 est moins grand, avec un taux de 43.22% de variance expliquée, à cause d'une baisse de la variance résiduelle de la constante de 0.546 (modèle vide) à 0.31 (modèle 1.2).

Nous remarquons que la compréhension orale a un apport spécifique plus important dans les analyses de la variable réponse du test ELFE par rapport à des analyses du test SLS. Tandis que la baisse de la variance résiduelle grâce à l'introduction du facteur CO augmente le pouvoir explicatif entre le modèle 1.1 et 1.2 d'environ 9 % au niveau des classes et de 1.8 % au niveau des individus, le facteur CO apporte un pouvoir explicatif compris entre 35 à 57 % au niveau des classes et de 7 % au niveau des individus.

En ce qui concerne la compétence de lecture de textes authentiques sans contrainte de temps (test de HarmoS, cf tableau 40.1 et 40.2), le pouvoir explicatif du modèle 1.2 avec les facteurs D et CO intégrés est moins élevé que dans les analyses du test ELFE ou du test SLS. Il est respectivement de 52.94% (Constante) et 11.45 % (Temps) au niveau des classes et de 25,75% au niveau des individus. Les facteurs décodage et compréhension orale semblent ainsi être importants pour la maîtrise de la lecture de textes authentiques, mais la part de variance résiduelle non expliquée qui persiste dans le modèle 1.2 suggère que d'autres facteurs doivent être pris en compte pour expliquer de manière satisfaisante les différences entre les individus et entre les classes. Nous pouvons constater également, que comme pour le test ELFE, l'apport de pouvoir explicatif du facteur CO est proportionnellement plus grand dans les analyses du test HarmoS que dans celles du test SLS.

Cette spécificité des compétences de lecture rapide de phrases évaluée par le test de SLS est confirmée lorsque nous regardons l'évolution des valeurs de la déviance ( $2 \cdot \log\text{likelihood}$ ). Nous avons en effet procédé à une évaluation de la différence (par simple soustraction) des déviations entre le modèle vide et le modèle 1.1 ainsi qu'entre le modèle 1.1 et 1.2. Ces différences sont inscrites en gris au sein des tableaux 38.1, 39.1 et 40.1. En dessous également en gris se trouvent les résultats des tests de  $\chi^2$  qui ont été effectués. Ils sont significatifs pour l'ensemble des analyses. Cependant dans le cadre du SLS (tableau 39.1), la différence entre les déviations du modèle vide et du modèle 1.1 est de 301.9. Cette différence est extrêmement significative à un niveau d'erreur de 0.01%. L'introduction du facteur décodage améliore par ailleurs de beaucoup notre modèle d'analyse. L'introduction de la compréhension orale améliore également le modèle, mais la différence de seulement 8.769 entre le modèle 1.1 et 1.2 n'est significative qu'à un niveau d'erreur de 5%.

En ce qui concerne le test ELFE, les résultats sont moins nets, tout en allant dans le même sens : la différence des déviations entre le modèle vide et le modèle 1.1 est de 164.12 ; celle entre le modèle 1.1 et le modèle 1.2 est de 36.58. Dans les deux cas, ces différences sont très significatives à un niveau d'erreur de 0.01%. A l'inverse, les analyses concernant la variable réponse du HarmoS nous montrent des différences de déviations semblables (78.76 et 80.29) entre les modèles. Cela nous indique que l'introduction de la compréhension orale en tant que deuxième facteur réduit davantage la déviance du modèle que l'introduction du premier facteur décodage.

**Résumé :**

*Les coefficients de détermination et les analyses de déviations confirment ainsi ce que les coefficients de régression nous ont indiqué : les deux facteurs Décodage (D) et Compréhension Orale (CO) ont des impacts significatifs sur l'ensemble des compétences en lecture que nous avons mesurées dans le groupe des élèves en CE1. L'apport spécifique de chacun des deux facteurs et la qualité finale du modèle d'analyse dépendent de la façon dont les compétences en lecture ont été opérationnalisées à travers nos trois tests SLS, ELFE et HarmoS.*

*Ainsi les compétences en lecture rapide de phrases (test SLS) et de textes courts (test ELFE) peuvent être bien expliquées par les deux composantes de la SVRT. Les résultats sont significatifs pour chacune des composantes, toutefois l'impact du Décodage est nettement plus grand.*

*Les compétences en lecture de textes authentiques, sans contrainte de temps (test HarmoS), sont moins bien modélisées par notre modèle de la SVRT. La variance expliquée n'est que d'environ 50 % comparé à 79 % pour le SLS et ELFE. Les deux composantes ont un impact significatif. L'impact du facteur Compréhension Orale est plus grand que sur les compétences de lecture sollicitée par les tests SLS et ELFE*

## 4.2 Résultats des analyses auprès des élèves en CM2/6<sup>ème</sup>

Nous allons maintenant présenter les résultats concernant le groupe des élèves en CM2/6<sup>ème</sup>, afin de vérifier la validité de notre hypothèse H1 sur l'influence des facteurs de décodage et de compréhension orale auprès d'un échantillon de lecteurs plus avancés.

Tableau 41 Résultats des analyses multi-niveaux en CM2 / 6<sup>ème</sup>

Variables réponses	SLS		ELFE		HarmoS-L1	
	Z_SLSPkt_5bis6		ZTextRoh_5bis6		Z_Lesen_5bis9	
<b>Variables explicatives (Parties fixes)</b>	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.078	0.08	0.013	0.082	-0.7***	0.078
Temps (compétences t1-t2)	0.142*	0.06	-0.042	0.111	0.456***	0.082
Décodage (D)	0.396***	0.04	0.369***	0.051	0.208***	0.038
Interaction D & Temps	0.131*	0.061	0.022	0.066	0.09	0.058
Compréhension Orale (CO)	0.229***	0.043	0.414***	0.055	0.312***	0.041
Interaction CO & Temps	-0.042	0.051	-0.153**	0.059	0.042	0.05
<b>Variances (Parties aléatoires)</b>						
Niveau III: Classes						
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.066	0.033	0.058	0.034	0.069	0.032
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.017	0.017	-0.053	0.038	-0.035	0.026
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.008	0.013	0.128	0.06	0.06	0.032
Niveau II: Individus						
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.414	0.037	0.612	0.055	0.307	0.028
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.181	0.031	-0.393	0.046	-0.233	0.029
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.449	0.043	0.576	0.054	0.442	0.042
-2*loglikelihood:	975.594		1063.101		834.688	
N° d'Unités Niv. III (Classes)	15		15		15	
N° d'Unités Niv. II (Individus)	267		267		266	
N° d'Unités Niv. I (Temps)	500		502		499	

\* le resultat est significatif avec un niveau d'erreur de  $\alpha = 5\%$ ; \*\*  $\alpha = 1\%$ ; \*\*\*  $\alpha = 0.1\%$

Conformément au tableau 37, dans le tableau 41 la constante nous indique le niveau de compétence en lecture mesuré par les trois différents tests de lecture (SLS ou ELFE ou le test de HarMoS) des élèves en CM2 ayant des compétences de décodage et de compréhension orale exactement dans la moyenne (d'une valeur de  $z=0$ ). En ce qui concerne le SLS, le niveau de compétence de ces élèves est de -0.078, leur compétence en lecture de textes courts (ELFE) est de 0.013 et leur compétence en lecture de textes authentiques (test de HarMoS) est de - 0.7.

Regardons maintenant la variable Temps, qui nous indique le développement des compétences en lecture en affichant, en notes z, de combien la compétence en lecture a, en moyenne, changé entre le premier et le deuxième moment de recueil de données. Tandis

que les élèves en CE1/CE2 font des progrès significatifs entre les deux moments de la collecte des données, c'est seulement dans le test de HarmoS (avec des textes longs et authentiques) que les élèves de CM2/6ème manifestent une nette progression, de presque la moitié d'un écart type ( $\beta=0.456$ ). En effet avec le test ELFE, rien de significatif n'est observé et en ce qui concerne la lecture rapide de phrases (test SLS), la progression existe mais demeure assez faible avec une amélioration de seulement 0.142 points (sur une échelle de z)<sup>46</sup>.

Concernant les coefficients de régression du Décodage (D), nous observons des valeurs importantes qui varient entre un tiers (SLS :  $\beta=0.396$  ; ELFE  $\beta=0.369$ ) et un quart d'écart type (HarmonS  $\beta=0.208$ ). Comme pour le groupe des jeunes lecteurs en classe de CE1/CE2, c'est dans la lecture rapide de phrases (SLS) que le décodage a l'impact le plus grand pour les élèves de CM2/6ème. Cependant cet impact reste bien plus faible que celui constaté dans le groupe des CE1 /CE2 (SLS :  $\beta=0.766$  voir tableau 37 et figure 22). Cette perte de poids est également visible pour la lecture rapide de texte courts (ELFE) où le coefficient de régression en CE1/CE1 était de  $\beta=0.535$  ( $\beta=0.369$  en CM2, voir figure 23).

En ce qui concerne le test de HarmoS, nous ne pouvons parler que d'une légère tendance puisque les coefficients en CE1 ( $\beta=0.255$ ) et en CM2 ( $\beta=0.208$ ) sont assez proches.

Cette perte d'influence du facteur décodage pour les résultats du SLS et de l'ELFE s'accompagne d'une augmentation de l'influence de la compréhension orale. Le coefficient de régression du facteur CO est pour le SLS de  $\beta=0.229$  en CM2 (par rapport à  $\beta=0.071$  en CE1) ; pour le ELFE il est de  $\beta=0.414$  en CM2 (par rapport à  $\beta=0.187$  en CE1). Encore une fois, pour le test de HarmoS nous n'observons pas un tel changement, les coefficients de régression de la compréhension orale restant assez proches entre le groupe en CE1 ( $\beta=0.339$ ) et le groupe en CM2 ( $\beta=0.312$ ). Les figures 22 à 24 illustrent les propos formulés ci-dessus.

---

<sup>46</sup> Nous rappelons que toutes les données dans nos tableaux se réfèrent à une échelle de z avec une moyenne de 0 et un écart type de 1.

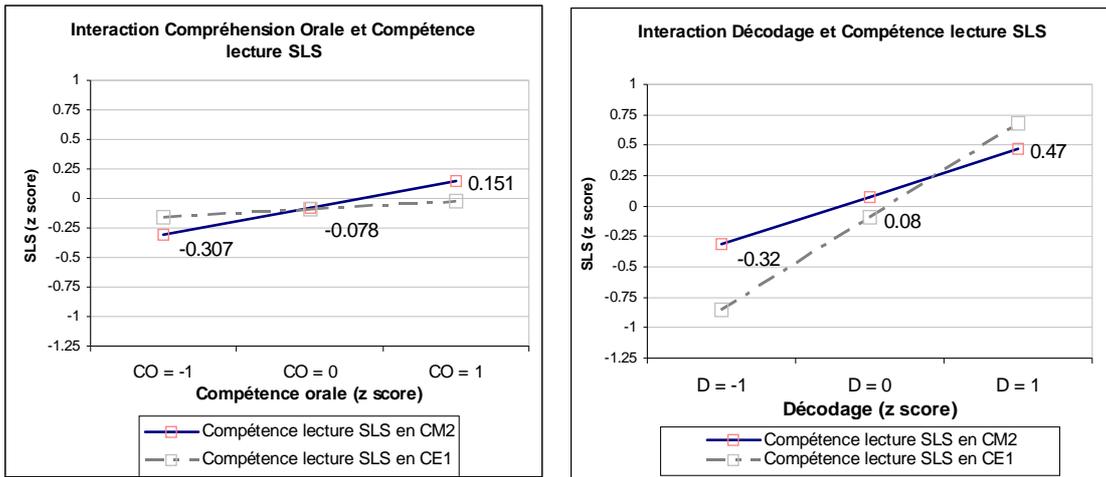


Figure 22 Droite de régression test SLS

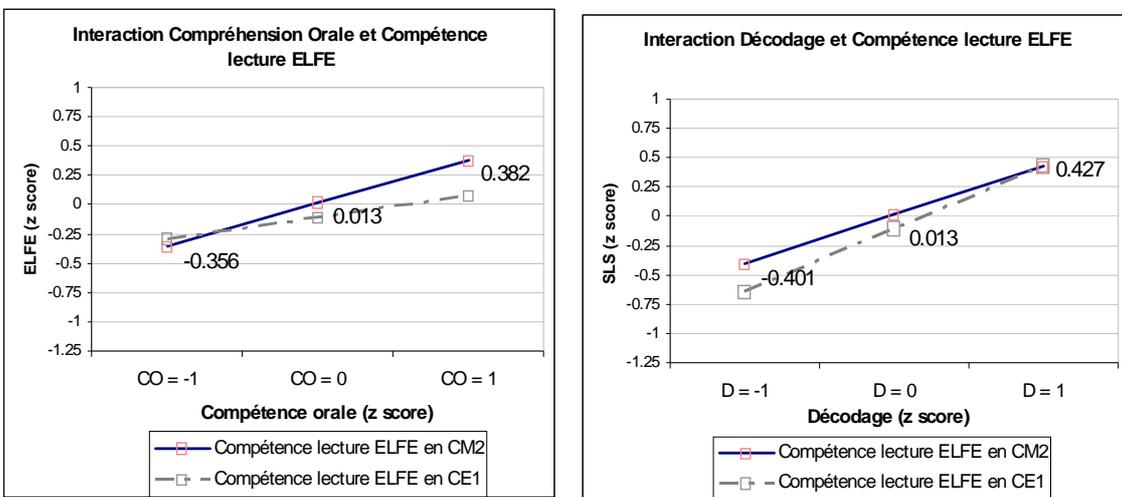


Figure 23 Droites de régression test ELFE

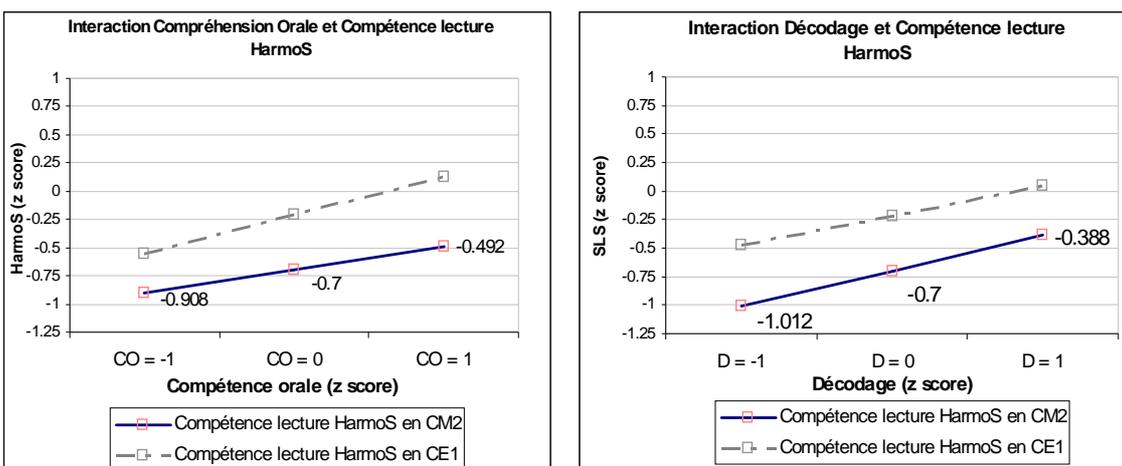


Figure 24 Droites de régression test HarmoS

Les interactions entre les facteurs décodage / compréhension orale et la variable temps se manifestent dans des coefficients de régression peu cohérents et non significatifs (voir tableau 41). Deux exceptions tout de même : le test SLS révèle un coefficient significativement positif (0.131) entre le décodage et le développement des compétences de lecture tandis que le test ELFE révèle un coefficient négatif important (-0.153) entre la compréhension orale et le développement des compétences de lecture. Pour le SLS une meilleure maîtrise du décodage semble favoriser la progression des compétences en lecture entre le CM2 et la 6<sup>ème</sup>, tandis que pour le test ELFE ce sont les élèves les plus faibles en compréhension orale qui font le plus de progrès entre le CM2 et la 6ème. Au vu de ces résultats, les deux facteurs D et CO de la *SVRT* ne semblent pas pouvoir expliquer de manière satisfaisante l'évolution des compétences de lecture dans le temps.

Ces résultats sont à nouveau confirmés par l'évolution des déviations au fil des modèles et les coefficients de détermination « pseudo R2 » indiqués sur les pages suivantes dans les tableaux 42.1 à 44.2 (pages 126-128).

La réduction de la déviance ( $2 \times \log\text{likelihood}$ , en gris dans les tableaux 42.1 à 44.1) est dans les trois analyses du SLS, ELFE et du HarMoS hautement significative à un niveau d'erreur de seulement 0.01 %.

Nous pouvons par ailleurs constater que l'introduction de la compréhension orale au modèle 1.2 conduit à une réduction de la déviance proportionnellement plus grande que dans les analyses du groupe des CE1/CE2. Ainsi, par exemple, la réduction de la déviance dans le modèle 1.2 du SLS était de seulement 8.76 pour les élèves de CE1 (voir tableau 38.1), alors qu'elle est de 32.9 2 pour les élèves de CM2 (voir tableau 42.1).

Dans les analyses du test ELFE en CE1, la réduction de la déviance est de 168.12 (voir tableau 39.1) lors de l'introduction du facteur Décodage et de 36.58, lors de l'introduction du facteur compréhension orale. Selon le tableau 43.1 en CM2 les deux facteurs apportent une réduction de la déviance presque similaire de 81.03 (au modèle 1.1 après l'introduction de du facteur D) voire de 70.17 (au modèle 1.2 après l'introduction du facteur CO).

Pour le test de HarmoS réalisé en CM2, c'est le facteur compréhension orale qui améliore sensiblement le modèle d'analyse en diminuant la déviance de 108.37 (versus 39.99 pour le facteur Décodage) alors qu'il n'arrive que dans un deuxième temps dans le modèle. En CE1, la baisse de la déviance due aux deux facteurs est très équilibrée (entre 78 et 80) comme nous le montre le tableau 40.1., sur la page 117.

L'analyse des déviances confirme ainsi le constat d'un impact significatif des deux facteurs avec une tendance à l'augmentation de l'influence du facteur compréhension orale au fil du temps. Avant de pouvoir tirer des conclusions de ces résultats, il convient de prendre en considération deux aspects supplémentaires : le taux de variance expliquée et les modèles inversés, c'est-à-dire les analyses de modèles qui introduisent d'abord la compréhension orale et ensuite le décodage.

L'analyse des variances résiduelles doit se restreindre aux taux de variance de la constante, c'est à dire à la variance qui existe entre les compétences de lecture des individus et des classes des CM2. L'analyse de la variance qui existe entre les taux de progression des individus et des classes (variable temps) doit être laissée de côté, car nous sommes encore une fois confrontés au phénomène des taux de variance expliquée négatifs. Ceci nous indique que la spécification de notre modèle avec les deux facteurs D et CO n'arrive pas à rendre suffisamment compte des divergences dans le développement des compétences de lecture lorsqu'on est confronté à un échantillon qui se caractérise par un taux de variance très élevé au niveau des individus avec, à l'inverse, très peu de variance entre les classes. L'interprétation de nos coefficients de détermination requiert ainsi la prudence.

Tableau 42.1 Evolution des modèles concernant le SLS en CM2 / 6ème

Variables réponses	SLS :		SLS :		SLS :	
	Modèle 0 (vide)		Modèle 1.1		Modèle 1.2	
Variables explicatives (Parties fixes)	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.303	0.095	-0.125	0.082	-0.078	0.08
Temps (compétences t1-t2)	0.642	0.045	0.228	0.055	0.142	0.06
Décodage (D)			0.413	0.039	0.396	0.04
Interaction D & Temps			0.122	0.056	0.131	0.061
Compréhension Orale (CO)					0.229	0.043
Interaction CO & Temps					-0.042	0.051
Variations (Parties aléatoires)						
Niveau III: Classes						
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.089	0.049	0.068	0.035	0.066	0.033
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.011	0.017	-0.011	0.012	-0.017	0.017
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.008	0.011	0	0	0.008	0.013
Niveau II: Individus						
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.764	0.069	0.471	0.043	0.414	0.037
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.118	0.035	-0.165	0.032	-0.181	0.031
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.344	0.033	0.432	0.04	0.449	0.043
-2*loglikelihood:	1107.693	99.176	1008.517	32.923	975.594	
Test de Chi2		df=2 0.000***		df=2 0.000***		
N° d'Unités Niv. III (Classes)	15		15		15	
N° d'Unités Niv. II (Individus)	267		267		267	
N° d'Unités Niv. I (Temps)	500		500		500	

Tableau 42.2 Les taux de variances expliquées relatif au SLS en CM2/ 6ème

Niveau III: Classes	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	23.60%	25.84%	2.94%
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	100.00%	0.00%	
Niveau II: Individus			
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	38.35%	45.81%	12.10%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	-25.58%	-30.52%	-3.94%

Tableau 43.1 Evolution des modèles concernant le ELFE en CM2 / 6ème

Variables réponses	ELFE:		ELFE:		ELFE:	
	Modèle 0 (vide)		Modèle 1.1		Modèle 1.2	
Variables explicatives (Parties fixes)	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.236	0.102	-0.061	0.098	0.013	0.082
Temps (compétences t1-t2)	0.442	0.095	0.075	0.11	-0.042	0.111
			0	0	0	0
Décodage (D)			0.431	0.051	0.369	0.051
Interaction D & Temps			0.006	0.062	0.022	0.066
Compréhension Orale (CO)					0.414	0.055
Interaction CO & Temps					-0.153	0.059
Variances (Parties aléatoires)						
Niveau III: Classes						
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.096	0.057	0.093	0.05	0.058	0.034
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.059	0.044	-0.063	0.045	-0.053	0.038
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.102	0.05	0.123	0.058	0.128	0.06
Niveau II: Individus						
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	1.037	0.093	0.765	0.069	0.612	0.055
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.446	0.057	-0.423	0.051	-0.393	0.046
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.534	0.051	0.57	0.054	0.576	0.054
-2*loglikelihood:	1214.312	81.033	1133.279	70.178	1063.101	
		df=2		df=2		
Test de Chi2		0.000***		0.000***		
N° d'Unités Niv. III (Classes)	15		15		15	
N° d'Unités Niv. II (Individus)	267		267		267	
N° d'Unités Niv. I (Temps)	502		502		502	

Tableau 43.2 Les taux de variances du ELFE en CM2 / 6ème

Niveau III: Classes	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	3.13%	39.58%	37.63%
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	-20.59%	-25.49%	-4.07%
Niveau II: Individus			
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	26.23%	40.98%	20.00%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	-6.74%	-7.87%	-1.05%

Tableau 44.1 Evolution des modèles concernant le HarmoS en CM2 / 6ème

Variables réponses	HarmoS: Modèle 0		HarmoS: Modèle 1.1		HarmoS: Modèle 1.2	
	$\beta$	SE	$\beta$	SE	$\beta$	SE
<b>Variables explicatives (Parties fixes)</b>						
Constante	-0.861	0.095	-0.77	0.092	-0.7	0.078
Temps (compétences t1-t2)	0.832	0.07	0.595	0.083	0.456	0.082
	0	0	0	0	0	0
Décodage (D)			0.226	0.038	0.208	0.038
Interaction D & Temps			0.11	0.055	0.09	0.058
Compréhension Orale (CO)					0.312	0.041
Interaction CO & Temps					0.042	0.05
<b>Variances (Parties aléatoires)</b>						
Niveau III: Classes						
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.105	0.049	0.099	0.045	0.069	0.032
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.014	0.026	-0.027	0.029	-0.035	0.026
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.05	0.027	0.062	0.032	0.06	0.032
Niveau II: Individus						
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.48	0.043	0.388	0.035	0.307	0.028
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.14	0.029	-0.189	0.029	-0.233	0.029
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.34	0.033	0.41	0.039	0.442	0.042
-2*loglikelihood:	983.065	39.999	943.066	108.378	834.688	
		df=2		df=2		
Test de Chi2		0.000***		0.000***		
N° d'Unités Niv. III (Classes)	15		15		15	
N° d'Unités Niv. II (Individus)	266		266		266	
N° d'Unités Niv. I (Temps)	499		499		499	

Tableau 44.2 Les taux de variances expliquées relatif au HarmoS en CM2/ 6ème

Niveau III: Classes	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	5.71%	34.29%	30.30%
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	-24.00%	-20.00%	3.23%
Niveau II: Individus			
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	19.17%	36.04%	20.88%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	-20.59%	-30.00%	-7.80%

Ce que nous pouvons observer en comparant les résultats des tableaux 38.2 - 40.2 (CE1/CE2) et 42.2 –44.2 (CM2/6<sup>ème</sup>) est qu'au total le taux de variance expliquée entre le modèle vide et le modèle 1.2 a tendance à être plus faible dans les modèles qui concernent les élèves plus âgés. Concrètement, chez les élèves en CM2/6<sup>ème</sup> pour le modèle complet contenant les deux facteurs D et CO (modèle 1.2) le taux de variance expliquée de 25 à 45% pour le test SLS. Il est de 39 à 40% pour le test ELFE et de 34 à 36% pour le test de HarmoS (voir tableaux 42.2, 43.2. et 44.2). Dans le groupe des élèves en CE1/CE2 (voir tableau 38.2), ce taux de variance expliquée est de 78 à 60 % pour le SLS, de 79 à 43% pour l'ELFE et de 52 à 25 % pour le test de HarmoS.

En ce qui concerne l'apport de chaque composante, nous constatons que dans le groupe des élèves en CM2 pour les test ELFE et HarmoS le facteur compréhension orale a un impact plus important dans l'explication de la variance résiduelle que dans le groupe de élèves plus jeunes (voir tableaux évolution entre le modèle 1.1 et 1.2 en comparant les tableaux 39.2 / 40.2 et 42.2 / 43.2).

*Résumé des résultats en CM2/6<sup>ème</sup> :*

*Nos analyses dans le groupe des élèves en CM2/6<sup>ème</sup> donnent une image assez différente de celles en CE1/CE2. C'est surtout en ce qui concerne la lecture rapide de phrases (test SLS) et de textes courts (test ELFE) que les différences sont marquées. Ainsi dans ces deux aspects de la lecture, les compétences n'évoluent quasiment pas entre l'année en CM2 et l'année en 6<sup>ème</sup>. A l'inverse les compétences dans la lecture de texte authentiques (test HarmoS) s'améliore significativement d'en moyenne un demi écart-type.*

*La force prédictive du modèle qui était très bonne en CE1/CE2 pour le SLS a considérablement baissée chez les élèves en CM2/6<sup>ème</sup>. Cela va de pair avec une perte d'influence importante du facteur décodage pour le test de SLS et ELFE.*

*En CM2/6<sup>ème</sup> la force prédictive concernant le test ELFE baisse surtout au niveau des classes, mais reste stable au niveau des individus.*

*Dans ce groupe d'élèves plus âgés la compréhension orale est globalement un facteur plus important qu'il a été dans le groupe des élèves de CE1/CE2.*

*Résumé des résultats à l'égard des hypothèses 1 et 2*

Au regard de notre première hypothèse et au vu de nos résultats, nous pouvons confirmer que les deux facteurs décodage et compréhension orale ont une influence significative dans l'explication des compétences en lecture, comme le postule la « *Simple View of Reading Theory* ». L'importance des deux facteurs dépend cependant de la façon dont les compétences de lecture sont opérationnalisées au sein de nos trois tests spécifiques. L'importance des deux facteurs dépend également de l'âge voire du niveau des lecteurs dont les compétences sont analysées. Les deux prédicteurs de la *SVRT* ont un impact plus important dans le groupe des jeunes lecteurs en CE1/CE2 et pour les tests qui mesurent une lecture rapide et efficace de textes très courts comme le SLS ou le test ELFE. Lorsqu'on regarde l'apport de chaque facteur pris individuellement, on s'aperçoit d'une augmentation de l'impact de la compréhension orale dans le groupe des lecteurs plus âgés (CM2/6ème), en ce qui concerne la lecture de textes courts comme dans le test ELFE et la lecture de textes plus longues et authentiques, sans contrainte du temps comme dans le test HarmoS.

### 4.3- L'influence du plurilinguisme

La question de recherche à laquelle nous souhaitons répondre sur les pages suivantes concerne la question de la validité de la *SVRT* au sein d'un échantillon uniquement plurilingue ou monolingue. Pour cela nous avons séparé notre échantillon d'élèves de CE1/CE2 en deux groupes, en fonction de l'origine linguistique. Nous avons présenté au sein du chapitre 3.1 (pages 56-64) les deux groupes et leurs éventuelles différences en ce qui concerne le statut socioprofessionnel des parents, le temps consacré à la lecture en allemand en dehors de l'école, le nombre de livres à la maison et les résultats obtenus dans un test d'intelligence.

Avant de vérifier la validité de la *SVRT* dans les deux groupes, il nous semblait utile de vérifier d'éventuelles différences concernant leurs compétences en lecture, en compréhension orale et en décodage.

Le tableau 45 montre que le fait d'être plurilingue a un impact négatif significatif uniquement sur le niveau des compétences en compréhension orale. Dans ce domaine, le niveau de compétence des élèves plurilingues est en moyenne un demi-écart type plus faible que celui des élèves monolingues ( $\beta = -0.527$ ). Dans le domaine du décodage ( $\beta = -0.224$ ) et des compétences de lecture mesurés par le test ELFE ( $\beta = -0.227$ ) l'écart moyen entre les élèves plurilingues et monolingues en CE1 est d'environ un écart type. Pour les tests de HarMoS et le SLS les élèves plurilingues ne se distinguent pas significativement des élèves monolingues.

Le temps consacré à la lecture, en dehors de l'école, a une influence positive significative sur les résultats des tests de lecture impliquant une lecture rapide et correcte comme le SLS et l'ELFE. Il est intéressant de noter que les enfants qui déclarent de lire beaucoup à la maison (entre 1 et 2 heures par jour) ont des résultats significativement meilleurs dans la lecture de textes authentiques (HarMoS), mais pas dans la lecture rapide de phrases et de textes courts.

Le statut professionnel des parents a - tout comme le fait d'être plurilingue - un impact significatif sur les compétences en compréhension orale. En effet, ces compétences augmentent en moyenne d'une catégorie professionnelle à l'autre de 0.069. De la catégorie professionnelle la plus faible à la plus élevée l'écart est de  $10 \times 0.069 = 0.69$ , ce qui est légèrement plus grand même que l'impact du facteur plurilinguisme (-0.527).

Tableau 45 Résultats du modèle 4 chez les élèves en CE1/CE2

Variables explicatives (Parties fixes)		Compréhension Orale		Décodage	
		$\beta$	SE	$\beta$	SE
Constante		-0.181	0.22	-0.534**	0.192
Temps (compétences t1-t2)		0.736***	0.22	1.047***	0.19
Intelligence (QI)		0.015***	0.003	0.017***	0.003
Interaction QI & Temps		0.006*	0.003	0.005	0.003
Plurilinguisme		-0.527***	0.123	-0.224*	0.1
Plurilinguisme & Temps		-0.034	0.12	-0.022	0.101
Statut Socioprofessionnelle (SES)		0.069**	0.025	0.009	0.02
Interaction SES & Temps		-0.018	0.024	-0.018	0.021
Lecture (allemand) hors-scolaire 30 min		0.058	0.164	0.092	0.133
Lecture (allemand) hors-scolaire 30' - 1h		0.284	0.185	0.187	0.15
Lecture (allemand) hors-scolaire 1h - 2h		0.238	0.216	0.272	0.175
Lecture (allemand) hors-scolaire > 2h		0.182	0.202	0.154	0.164
Lecture (allm.) hs-scol. 30 min & Temps		-0.132	0.171	0.024	0.146
Lecture (allm.) hs-scol. 30'-1h & Temps		-0.3	0.191	0.133	0.163
Lecture (allm.) hs-scol. 1h - 2h & Temps		-0.056	0.217	0.111	0.184
Lecture (allm.) hs-scol. > 2h & Temps		-0.317	0.208	0.153	0.176
<b>Variances (Parties aléatoires)</b>					
Niveau III: Classes					
$\sigma^2_{v_0}$ = Variance résiduelle $\beta_0$ (Constante)		0.052	0.029	0.116	0.047
$\sigma_{v_{01}}$ = Covariance de $\beta_0$ et $\beta_1$		-0.015	0.018	0.024	0.022
$\sigma^2_{v_1}$ = Variance résiduelle de $\beta_1$ (Temps)		0.025	0.02	0.031	0.019
Niveau II: Individus					
$\sigma^2_{u_0}$ = Variance résiduelle $\beta_0$ (Constante)		0.563	0.048	0.365	0.031
$\sigma_{u_{01}}$ = Covariance de $\beta_0$ et $\beta_1$		-0.225	0.036	-0.053	0.023
$\sigma^2_{u_1}$ = Variance résiduelle de $\beta_1$ (Temps)		0.509	0.046	0.358	0.032
-2*loglikelihood		1264.88		1113.44	
N° d'Unités Niv. III (Classes)		17		17	
N° d'Unités Niv. II (Individus)		318		318	
N° d'Unités Niv. I (Temps)		580		580	

Variables explicatives (Parties fixes)		SLS		ELFE		HarmoS	
		ZSLS_pk123_01	SE	ZTextRoh_2bis3_01	SE	Z_Lesen_2bis3_01	SE
Constante		-0.59**	0.214	-0.474*	0.187	-0.477*	0.233
Temps (compétences t1-t2)		0.9***	0.158	0.89***	0.205	0.52*	0.256
Intelligence (QI)		0.019***	0.003	0.021***	0.003	0.019***	0.003
Interaction QI & Temps		0	0.002	0.006*	0.003	0.005	0.003
Plurilinguisme		-0.211	0.11	-0.276**	0.104	-0.153	0.113
Plurilinguisme & Temps		-0.047	0.08	-0.003	0.109	-0.049	0.13
Statut Socioprofessionnelle (SES)		0.012	0.022	0.028	0.021	0.018	0.022
Interaction SES & Temps		0.005	0.016	-0.021	0.022	0.021	0.026
Lecture (allemand) hors-scolaire 30 min		0.106	0.148	0.038	0.139	0.029	0.149
Lecture (allemand) hors-scolaire 30' - 1h		0.292	0.167	0.299	0.156	0.213	0.169
Lecture (allemand) hors-scolaire 1h - 2h		0.433*	0.195	0.437*	0.183	0.086	0.198
Lecture (allemand) hors-scolaire > 2h		0.34	0.182	0.265	0.172	0.335	0.187
Lecture (allm.) hs-scol. 30 min & Temps		0.072	0.116	0.07	0.157	0.327	0.184
Lecture (allm.) hs-scol. 30'-1h & Temps		0.068	0.13	-0.019	0.175	0.3	0.206
Lecture (allm.) hs-scol. 1h - 2h & Temps		0.11	0.146	-0.145	0.198	0.503*	0.234
Lecture (allm.) hs-scol. > 2h & Temps		-0.007	0.138	0.074	0.19	0.037	0.225
<b>Variances (Parties aléatoires)</b>							
Niveau III: Classes							
$\sigma^2_{v_0}$ = Variance résiduelle $\beta_0$ (Constante)		0.141	0.058	0.043	0.023	0.269	0.104
$\sigma_{v_{01}}$ = Covariance de $\beta_0$ et $\beta_1$		-0.025	0.028	0.034	0.018	-0.116	0.068
$\sigma^2_{v_1}$ = Variance résiduelle de $\beta_1$ (Temps)		0.059	0.026	0.046	0.025	0.161	0.069
Niveau II: Individus							
$\sigma^2_{u_0}$ = Variance résiduelle $\beta_0$ (Constante)		0.454	0.038	0.413	0.035	0.439	0.038
$\sigma_{u_{01}}$ = Covariance de $\beta_0$ et $\beta_1$		-0.075	0.02	-0.101	0.027	-0.262	0.035
$\sigma^2_{u_1}$ = Variance résiduelle de $\beta_1$ (Temps)		0.201	0.018	0.419	0.037	0.562	0.051
-2*loglikelihood		1010.9		1168.3		1200.2	
N° d'Unités Niv. III (Classes)		17		17		17	
N° d'Unités Niv. II (Individus)		318		318		317	
N° d'Unités Niv. I (Temps)		571		580		570	

\* le resultat est significatif avec un niveau d'erreur de  $\alpha = 5\%$ ; \*\*  $\alpha = 1\%$ ; \*\*\*  $\alpha=0.1\%$

Un élève monolingue issu d'une famille d'un niveau socioprofessionnel peu élevé aura, potentiellement, comme un élève plurilingue, un niveau de compréhension orale plus faible. Pour une majorité des élèves plurilingues, les deux effets risquent de s'additionner, car notre échantillon d'élèves plurilingues est plus souvent issu d'un milieu socioprofessionnel moins élevé que celui des élèves monolingues. La figure 25 ci-dessous montre la relation entre les facteurs plurilinguisme et statut socioprofessionnel et leur impact sur l'évolution des compétences en compréhension orale.

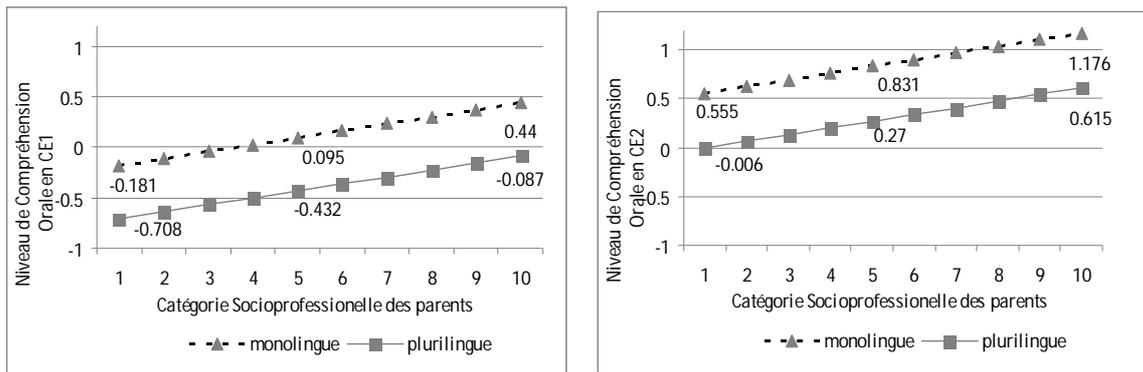


Figure 25 Evolution des Compétences en Compréhension Orale entre le CE1 et le CE2

Nous voyons qu'en CE1 lorsque toutes les autres caractéristiques (QI, Lecture extrascolaire,) sont égales, un élève plurilingue de la catégorie socioprofessionnelle la plus haute a à peu près le même niveau de compréhension orale qu'un élève monolingue de la catégorie socioprofessionnelle 3. Entre les deux droites, nous avons indiqué le niveau de compréhension orale pour la catégorie socioprofessionnelle moyenne.

Les écarts entre les élèves plurilingues et monolingues observés en ce qui concerne les compétences de décodage et surtout de compréhension orale nous amène à nous poser la question de savoir si ces différences auront des conséquences sur la validité de la *SVRT* dans le groupe des plurilingues par rapport aux monolingues.

Le premier constat que nous pouvons faire à la vue des données présentées dans le tableau 45 est que la faiblesse des élèves plurilingues en compréhension orale, ne semble guère affecter les compétences en lecture mesurées par le test SLS (lecture rapide de phrases courtes) ou le test de HarmoS (lecture de textes authentiques). Dans ces deux tests aucune différence significative n'existe entre le groupe des monolingues et celui des plurilingues.

Tableau 46 Résultats du modèle 1.2

Variables réponses	CE 1 / CE 2 modèle 1.2 Elèves plurilingues						CE 1 / CE 2 modèle 1.2 Elèves monolingues					
	SLS		ELFE		HarmoS-L1		SLS		ELFE		HarmoS-L1	
	ZSLS_Pkt23_01	SE	ZTextRoh_2bis3	SE	Z_Lesen_2bis3	SE	Z_SLS_Pkt_5bis6	SE	ZTextRoh_5bis6	SE	Z_Lesen_5bis6	SE
<b>Variables explicatives (Parties fixes)</b>												
Constante	-0.096*	0.042	-0.192***	0.045	-0.118	0.067	0.049	0.086	0.05	0.098	-0.082	0.088
Temps (compétences t1-t2)	0.189***	0.048	0.283***	0.062	0.22**	0.08	0.196	0.125	0.179	0.157	0.343*	0.15
Décodage (D)	0.736***	0.044	0.471***	0.049	0.393***	0.074	0.865***	0.099	0.679***	0.111	0.283**	0.101
Interaction D & Temps	-0.093*	0.047	-0.006	0.06	0.043	0.083	-0.286**	0.109	-0.206	0.119	-0.022	0.128
Compréhension Orale (CO)	0.123**	0.039	0.185***	0.043	0.395***	0.066	0.014	0.087	0.232*	0.092	0.351***	0.091
Interaction CO & Temps	-0.055	0.045	0.025	0.06	-0.069	0.08	0.103	0.114	0.091	0.129	-0.121	0.135
<b>Variances (Parties aléatoires)</b>												
Niveau II: Individus												
$\sigma^2_{uo}$ = Variance résiduelle $\beta_0$ (Constante)	0.244	0.023	0.284	0.027	0.585	0.056	0.312	0.056	0.448	0.081	0.316	0.058
$\sigma_{uo1}$ = Covariance de $\beta_0$ et $\beta_1$	-0.111	0.019	-0.13	0.028	-0.472	0.059	-0.205	0.052	-0.129	0.061	-0.239	0.063
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.255	0.026	0.503	0.049	0.854	0.085	0.375	0.07	0.447	0.083	0.56	0.104
-2*loglikelihood:	597.191		813.242		964.408		190.422		249.461		219.371	
N° d'Unités Niv. II (Individus)	246		246		245		67		67		67	
N° d'Unités Niv. I (Temps)	440		448		439		123		124		123	

\* le resultat est significatif avec un niveau d'erreur de  $\alpha = 5\%$ ; \*\*  $\alpha = 1\%$ ; \*\*\*  $\alpha = 0.1\%$

Cela semble discréditer l'importance que la SVRT accorde à la compréhension orale en tant que prédicteur de la compétence de lecture. Le test ELFE montre l'exception, car il met en exergue une différence significative de  $\beta = -0.276$  en faveur des monolingues (voir tableau 45). Cela semble confirmer, ce que les tableaux 41.2, 42.2 et 43.2 nous ont déjà suggéré : dans notre échantillon, la force explicative de la SVRT est plus grande pour le test ELFE que pour des deux autres tests de lecture utilisés.

Le tableau 46 ci-contre nous permet de comparer les coefficients de régression obtenus lors des analyses multi-niveaux séparément pour le groupe des élèves plurilingues et monolingues. Nous devons ajouter que le nombre relativement faible d'élèves monolingues, nous a contraint à enlever le niveau des classes de nos modèles d'analyses ; hormis cet aspect, les modèles sont strictement restés identiques aux modèles de l'échantillon général.

Les coefficients significatifs de la variable temps nous indiquent que

le groupe des élèves monolingues en CE2 fait des progrès significatifs dans les trois tests entre la première et la deuxième collecte de données. Les progrès du groupe d'élèves monolingues ne sont significatifs que dans le test HarmoS.

En ce qui concerne l'impact du facteur décodage, on voit une décroissance similaire dans les deux groupes avec une influence particulièrement forte sur les résultats du test SLS ( $\beta=0.736$  et  $\beta=0.865$ ). L'impact sur les résultats du test ELFE est encore assez fort ( $\beta=0.471$  et  $\beta=0.679$ ), mais plus faible sur les résultats du test HarmoS ( $\beta=0.393$  et  $\beta=0.291$ ). Nous constatons par ailleurs que, chez les monolingues, l'importance du décodage a tendance à être plus grande que chez les plurilingues, à l'exception du test de HarmoS où chez les plurilingues la compréhension orale a autant d'influence que le décodage.

Lorsque nous regardons l'interaction entre les facteurs décodage et temps, nous avons dans les deux groupes une interaction négative peu élevée ( $\beta=-0.093$  et  $\beta=0.286$ ). Elle nous indique que les enfants qui sont d'abord plus faibles en Décodage font, en moyenne, de plus grands progrès dans la lecture rapide de phrases (SLS). Cet effet est plus faible chez les plurilingues que chez les monolingues. En ce qui concerne le facteur Compréhension Orale nous n'observons qu'une seule différence entre les deux groupes : l'impact significatif de la compréhension orale sur les résultats du SLS chez les plurilingues est absent chez les monolingues. Le niveau de lecture rapide de phrases est chez les monolingues très largement déterminé par le niveau de décodage des élèves. Chez les plurilingues par contre, la faculté de compréhension du contenu des phrases semble aussi jouer un rôle.

Nous allons dans les pages suivantes présenter des tableaux spécifiques pour chacun de nos trois tests utilisés, afin de montrer de façon comparative pour les plurilingues et les monolingues l'évolution de la déviance et du coefficient de détermination « pseudo  $R^2$  ».

Tableau 47 Evolution des modèles du SLS chez les monolingues et plurilingues

Variables réponses	Elèves plurilingues en CE1/CE2						Elèves monolingues en CE1/CE2					
	SLS : Modèle 0 (vide)		SLS : Modèle 1.1		SLS : Modèle 1.2		SLS : Modèle 0 (vide)		SLS : Modèle 1.1		SLS : Modèle 1.2	
	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE
<b>Variables explicatives (Parties fixes)</b>												
Constante	-0.597	0.054	-0.111	0.042	-0.096	0.042	-0.139	0.107	0.054	0.074	0.049	0.086
Temps (compétences t1-t2)	0.969	0.035	0.202	0.049	0.189	0.048	0.955	0.073	0.269	0.104	0.196	0.125
Décodage (D)			0.787	0.042		0.736			0.873	0.099		0.865
Interaction D & Temps			-0.122	0.043		-0.093			-0.264	0.109		-0.286
Compréhension Orale (CO)						0.123						0.014
Interaction CO & Temps						-0.055						0.103
<b>Variances (Parties aléatoires)</b>												
Niveau II: Individus												
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.694	0.065	0.255	0.024	0.244	0.023	0.74	0.132	0.312	0.057	0.312	0.056
$\sigma_{u0,1}$ = Covariance de $\beta_0$ et $\beta_1$	-0.062	0.03	-0.117	0.02	-0.111	0.019	-0.199	0.068	-0.213	0.053	-0.205	0.052
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.249	0.025	0.259	0.026	0.255	0.026	0.309	0.058	0.393	0.073	0.375	0.07
-2*loglikelihood:	889.73	281.88	607.85	10.661	597.19		251.82	59.71	192.11	1.687	190.42	
Test de Chi2				df=2				df=2		df=2		
N° d'Unités Niv. II (Individus)	246		246		246		67		67		67	
N° d'Unités Niv. I (Temps)	440		440		440		123		123		123	
Niveau II: Individus	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2	Modèle 1.1 → 1.2	Modèle 1.1 → 1.2	
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	63.26%	64.84%	4.31%	57.84%	57.84%	4.31%	57.84%	57.84%	57.84%	0.00%	0.00%	
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	-4.02%	-2.41%	1.54%	-27.18%	-21.36%	1.54%	-27.18%	-21.36%	-21.36%	4.58%	4.58%	

Tableau 48 Evolution des modèles du ELFE chez les monolingues et plurilingues

Variables réponses	Elèves plurilingues en CE1/CE2						Elèves monolingues en CE1/CE2					
	ELFE: Modèle 0 (vide)		ELFE: Modèle 1.1		ELFE: Modèle 1.2		ELFE: Modèle 0 (vide)		ELFE: Modèle 1.1		ELFE: Modèle 1.2	
	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE
<b>Variables explicatives (Parties fixes)</b>												
Constante	-0.549	0.047	-0.217	0.047	-0.192	0.045	-0.01	0.113	0.15	0.093	0.05	0.098
Temps (compétences t1-t2)	0.856	0.048	0.307	0.063	0.283	0.062	0.918	0.084	0.323	0.133	0.179	0.157
Décodage (D)			0.541	0.047	0.471	0.049			0.689	0.117	0.679	0.111
Interaction D & Temps			-0.019	0.055	-0.006	0.06			-0.166	0.119	-0.206	0.119
Compréhension Orale (CO)					0.185	0.043					0.232	0.092
Interaction CO & Temps					0.025	0.06					0.091	0.129
<b>Variances (Parties aléatoires)</b>												
Niveau II: Individus												
$\sigma^2_{u_0}$ = Variance résiduelle $\beta_0$ (Constante)	0.517	0.049	0.311	0.03	0.284	0.027	0.831	0.148	0.506	0.091	0.448	0.081
$\sigma_{u_0,1}$ = Covariance de $\beta_0$ et $\beta_1$	-0.05	0.034	-0.123	0.029	-0.13	0.028	-0.082	0.077	-0.139	0.066	-0.129	0.061
$\sigma^2_{u_1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.481	0.047	0.5	0.049	0.503	0.049	0.409	0.076	0.465	0.087	0.447	0.083
-2*loglikelihood:	973.23	134.27	838.96	25.713	813.24		288.96	29.348	259.61	10.153	249.46	
Test de Chi2				df=2				df= 2		df=2		
N° d'Unités Niv. II (Individus)	246		246		246		67		67		67	
N° d'Unités Niv. I (Temps)	448		448		448		124		124		124	
Niveau II: Individus	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2	Modèle 1.1 → 1.2	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2			
$\sigma^2_{u_0}$ = Variance résiduelle $\beta_0$ (Constante)	39.85%	45.07%	8.68%	8.68%	39.11%	46.09%	39.11%	46.09%	11.46%	11.46%	11.46%	11.46%
$\sigma^2_{u_1}$ = Variance résiduelle de $\beta_1$ (Temps)	-3.95%	-4.57%	0.00%	0.00%	-13.69%	-9.29%	-13.69%	-9.29%	3.87%	3.87%	3.87%	3.87%

Tableau 49 Evolution des modèles du HarmoS chez les monolingues et plurilingues

Variables réponses	Elèves plurilingues en CE1/CE2						Elèves monolingues en CE1/CE2					
	HarmoS		HarmoS		HarmoS		HarmoS		HarmoS		HarmoS	
	Modèle 0 (vide)	Modèle 1.1	Modèle 1.1	Modèle 1.2	Modèle 0 (vide)	Modèle 1.1	Modèle 1.1	Modèle 1.2	Modèle 0 (vide)	Modèle 1.1	Modèle 1.1	Modèle 1.2
<b>Variables explicatives (Parties fixes)</b>												
Constante	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$
SE	SE	SE	SE	SE	SE	SE	SE	SE	SE	SE	SE	SE
Temps (compétences t1-t2)	-0.522	-0.17	0.071	-0.118	0.067	0.022	0.091	0.084	-0.082	0.088	0.082	0.088
	0.843	0.271	0.084	0.22	0.08	0.646	0.338	0.128	0.343	0.15	0.343	0.15
Décodage (D)		0.554	0.072	0.393	0.074		0.328	0.111	0.283	0.101	0.283	0.101
Interaction D & Temps		-0.024	0.079	0.043	0.083		-0.023	0.132	-0.022	0.128	-0.022	0.128
Compréhension Orale (CO)				0.395	0.066				0.351	0.091		0.091
Interaction CO & Temps				-0.069	0.08				-0.121	0.135		0.135
<b>Variances (Parties aléatoires)</b>												
Niveau II: Individus												
$\sigma^2_{u_0}$ = Variance résiduelle $\beta_0$ (Constante)	0.881	0.084	0.066	0.585	0.056	0.454	0.393	0.072	0.316	0.058	0.316	0.058
$\sigma^2_{u_{01}}$ = Covariance de $\beta_0$ et $\beta_1$	-0.387	0.065	0.064	-0.472	0.059	-0.254	-0.281	0.073	-0.239	0.063	-0.239	0.063
$\sigma^2_{u_1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.8	0.081	0.886	0.854	0.085	0.571	0.6	0.112	0.56	0.104	0.56	0.104
-2*loglikelihood:	1126	102.97	1023	58.598	964.41	249.65	13.961	16.317	219.37		219.37	
	df=2	df=2	df=2	df=2	df=2	df=2	df=2	df=2	df=2		df=2	
Test de Chi2	0.000***	0.000***	0.000***	0.000***	0.001***	0.001***	0.000***	0.000***	0.000***		0.000***	
N° d'Unités Niv. II (Individus)	245	245	245	245	245	67	67	67	67		67	
N° d'Unités Niv. I (Temps)	439	439	439	439	439	123	123	123	123		123	
Niveau II: Individus	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2	Modèle 1.1 → 1.2	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 0 → 1.1	Modèle 0 → 1.2	Modèle 1.1 → 1.2			
$\sigma^2_{u_0}$ = Variance résiduelle $\beta_0$ (Constante)	22.47%	33.60%	14.35%	14.35%	13.44%	30.40%	13.44%	30.40%	19.59%	19.59%	19.59%	19.59%
$\sigma^2_{u_1}$ = Variance résiduelle de $\beta_1$ (Temps)	-10.75%	-6.75%	3.61%	3.61%	-5.08%	1.93%	-5.08%	1.93%	6.67%	6.67%	6.67%	6.67%

*Le test SLS*

Le niveau de compétence en lecture rapide de phrases (voir tableau 45) est, comme nous l'indiquaient les coefficients de régressions, avant tout déterminé par le niveau de décodage. Chez les monolingues, l'introduction du facteur CO n'améliore pas le modèle (la déviance ne diminue pas de façon significative) et n'augmente pas la part de variance expliquée. Chez les plurilingues, l'introduction du facteur CO augmente la part de variance expliquée de seulement 4,31%. Puisque nous sommes, dans le cas du SLS, comme dans celui des deux autres tests, à nouveau confronté aux problèmes des variances négatives dues à une augmentation de variance résiduelle du facteur temps, nous écartons ce facteur prudemment de nos interprétations. Notons toutefois que ces augmentations sont minimales. Étant donné que les analyses multi-niveaux se font par itérations successives, qui nous fournissent des estimations finales à trois décimales, ces augmentations minimales de la variance résiduelle peuvent être liées à la marge d'erreur des estimations. Ce que nous pouvons dire avec certitude, c'est que nos modèles qui intègrent les facteurs de la *SVRT*, arrivent à expliquer une bonne partie des variances résiduelles de la constante dans les deux groupes d'élèves plurilingues et monolingues. Cependant ils n'arrivent pas à expliquer la grande variabilité du développement des compétences de lecture dans le temps.

*Le test ELFE*

Le tableau 46 montre, ce qui concerne le test ELFE, qu'il existe entre les deux groupes une grande ressemblance des parts de variance expliquée lors de la transition du modèle vide au modèle 1.1., puis au modèle 1.2. L'introduction d'abord du décodage et ensuite de la compréhension orale réduit dans la même mesure la variance résiduelle de la constante. L'évolution de la déviance montre une amélioration significative de la qualité du modèle aussi bien dans le groupe des plurilingues que des monolingues.

*Le test HarmoS*

Lorsque nous regardons les modèles spécifiques au test de HarmoS (tableau 47), nous constatons que l'explication de la variance résiduelle de la constante atteint dans les deux groupes un taux d'environ 30% dans le modèle complet (modèle 1.2) avec les deux facteurs D et CO. Comparé aux résultats du SLS et de l'ELFE, c'est le « pseudo R<sup>2</sup> » le plus faible des trois tests. La contribution relative de chacun des deux facteurs varie cependant entre les deux groupes. Chez les monolingues l'apport de la compré-

hension orale (19.45%) surpasse celui du décodage (13.44%)- bien qu'il ne soit introduit que dans un deuxième temps.

Chez les plurilingues, l'introduction du facteur CO contribue à une augmentation de la variance expliquée de la constante de 14.34%, ce qui demeure inférieur à l'apport dû à l'introduction du facteur D (22.47%). Ceci confirme l'information des coefficients de régression : le Décodage est aussi important que la compréhension orale dans l'explication des compétences en lecture mesurée par le test de HarmoS, tandis que chez les monolingues c'est la compréhension orale qui est plus déterminante.

*Résumé :*

*Le plurilinguisme peut avoir – au même titre qu'un statut socio-économique peu élevé - un impact négatif sur les compétences en langue opérationnalisées par la compréhension orale. Il n'a pas d'impact généralement négatif sur les compétences en lecture, à part pour une lecture rapide de textes courts, qui demandent en particulier de savoir faire des inférences entre différentes phrases en utilisant par exemple des pronoms comme indicateurs.*

*A la vue des résultats obtenus dans notre échantillon en CE1/CE2, nous pouvons dire qu'en général le modèle de la SVRT tel que nous l'avons testé ne s'applique pas beaucoup mieux à notre groupe d'élèves monolingues qu'à notre groupe d'élèves plurilingues.*

*Les différences observées entre les deux groupes concernent l'importance relative des deux facteurs dans les différents tests de compétence en lecture. Ainsi dans le groupe des élèves plurilingues la compréhension orale est un facteur prédictif important pour chacun des trois tests de lecture, y compris le SLS. De même dans ce groupe la force prédictive du modèle s'améliore un peu plus (environ 5%) que chez les monolingues lorsque la compréhension orale est introduite dans le modèle. Cela mène global total à une explication de variance un peu plus élevée chez les plurilingues.*

*Un impact clairement plus important de la compréhension orale dans le groupe des élèves plurilingues n'a cependant pas pu être confirmé.*

#### 4.4- L'impact du QI en tant que troisième facteur possible de la SVRT

La dernière de nos questions de recherche concerne la nécessité d'introduire un facteur complémentaire comme le QI dans le modèle de la SVRT. Nous avons exposé dans la partie théorique (chapitre 2.4) le point de vue d'auteurs comme Tiu et al. (2003) défendant l'utilité du QI dans l'identification d'enfants ayant des troubles d'apprentissage de la lecture. Les origines du débat autour du rôle du QI dans la lecture remontent à la citation de Jensen (1980, cité dans Stanovich, Cunningham, Feeman, 1984) « ...reading comprehension is largely a matter of g ».<sup>47</sup>

Les tableaux sur les pages suivantes présentent les résultats de nos analyses visant à analyser l'impact du QI en tant que prédicteur et possible facteur complémentaire à intégrer à la SVRT. Le tableau 50 sur la page suivante montre la force explicative du QI en tant que facteur indépendant unique.

##### 4.4.1 Analyse des coefficients de régression

###### *Les élèves de CE1/CE2*

Les coefficients de régression dans le groupe d'élèves de CE1/CE2 sont, pour les trois tests de lecture, d'une remarquable similitude ( $\beta = 0.021$  à  $0.024$ ) chaque point de QI « gagné » s'accompagne donc d'une augmentation de  $0.021 / 0.024$  des compétences en lecture. La moyenne de QI de notre échantillon est de 102. Ainsi Un élève de CE1 ayant un QI de 103, sa compétence de lecture, mesurée par les trois tests, augmente de  $0.021$  à  $0.024$ , selon le test, par rapport à la constante, donnée par la moyenne sur l'ensemble des élèves en CE1. Ramené à l'échelle du QI, cela veut dire qu'un élève, ayant un QI de 115, aura une note en compétence de lecture sur l'échelle de z qui s'écartera de presque un tiers de déviation standard ( $0.021 \times 13 = 0.273$ ) au dessus de la constante.

L'impact du QI sur le développement des compétences en lecture au cours d'une année n'est pas significatif comme nous le montre le coefficient de l'interaction entre le QI et la variable temps (voir tableau 50).

---

<sup>47</sup> „La compréhension de lecture repose essentiellement sur l'intelligence générale –g–.“

Tableau 50 L'impact du QI sur les compétences en lecture dans les deux groupes d'âges

Variables responses	CE1 / CE2						CE1 / CE2						CE1 / CE2						CE2 / 6ème						
	SLS		ELFE		HarmoS		SLS		ELFE		HarmoS		SLS		ELFE		HarmoS		SLS		ELFE		HarmoS		
	Modèle 2.1	β	SE	Modèle 2.1	β	SE	Modèle 2.1	β	SE	Modèle 2.1	β	SE	Modèle 2.1	β	SE	Modèle 2.1	β	SE	Modèle 2.1	β	SE	Modèle 2.1	β	SE	
<b>Variables explicatives (Parties fixes)</b>																									
Constante	-0.53	0.111	0.066	-0.448	0.073	0.142	-0.319	0.087	-0.253	0.067	-0.874	0.069	0.943	0.066	0.841	0.063	0.841	0.112	0.642	0.045	0.45	0.092	0.831	0.068	
Temps (compétences t1-t2)	0.021***	0.003	0.024***	0.003	0.003	0.003	0.024***	0.004	0.036***	0.004	0.026***	0.003	0	0.002	0.005	0.003	0.005	0.003	0	0.003	-0.012	0.003	0.002	0.003	
Intelligence (QI)																									
Interaction QI & Temps	0	0.002	0.005	0.005	0.003	0.003	0	0.003	0.003	0.003	0.003	0.003	0	0.002	0.005	0.003	0.005	0.003	0	0.003	-0.012	0.003	0.002	0.003	
<b>Variables (Parties aléatoires)</b>																									
Niveau III: Classes																									
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.179	0.072	0.065	0.031	0.304	0.116	0.073	0.041	0.02	0.024	0.05	0.026	-0.027	0.031	0.046	0.021	-0.125	0.073	-0.015	0.016	-0.035	0.028	-0.008	0.019	
$\sigma^2_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	0.057	0.025	0.04	0.023	0.164	0.071	0.009	0.011	0.093	0.046	0.025	0.025	0.057	0.025	0.04	0.023	0.164	0.071	0.009	0.011	0.093	0.046	0.047	0.025	
Niveau II: Individus																									
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.472	0.04	0.448	0.038	0.454	0.039	0.663	0.06	0.811	0.073	0.367	0.033	-0.072	0.02	-0.114	0.028	-0.267	0.037	-0.116	0.033	-0.38	0.049	-0.153	0.026	
$\sigma^2_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	0.204	0.019	0.432	0.039	0.587	0.053	0.344	0.033	0.517	0.049	0.34	0.033	0.204	0.019	0.432	0.039	0.587	0.053	0.344	0.033	0.517	0.049	0.34	0.033	
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)																									
-2*Loglikelihood:	1030.96																								
Test de Chi <sup>2</sup>																									
N° d'Unités Niv. III (Classes)	17																								
N° d'Unités Niv. II (Individus)	318																								
N° d'Unités Niv. I (Temps)	571																								
* le resultat est significatif avec un niveau d'erreur de $\alpha = 5\%$ ; ** $\alpha = 1\%$ ; *** $\alpha = 0.1\%$																									
Niveau III: Classes																									
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	22.84%	(78%)	38.10%	(79%)	18.72%	(52%)	17.98%	(24%)	79.17%	(39%)	52.38%	(34%)	1.72%	(60%)	-5.26%	Modèle 0 $\rightarrow$ 2.1	1.20%	Modèle 0 $\rightarrow$ 2.1	-12.50%	Modèle 0 $\rightarrow$ 2.1	8.82%	6.00%	6.00%	6.00%	
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	1.72%	(60%)	-5.26%	Modèle 0 $\rightarrow$ 2.1	1.20%	Modèle 0 $\rightarrow$ 2.1	-12.50%	Modèle 0 $\rightarrow$ 2.1	8.82%	6.00%	6.00%	6.00%	1.72%	(60%)	-5.26%	Modèle 0 $\rightarrow$ 2.1	1.20%	Modèle 0 $\rightarrow$ 2.1	-12.50%	Modèle 0 $\rightarrow$ 2.1	8.82%	6.00%	6.00%	6.00%	
Niveau II: Individus																									
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	15.11%	(60%)	17.95%	(43%)	14.66%	(25%)	13.22%	(38%)	21.79%	(40%)	23.54%	(36%)	15.11%	(60%)	17.95%	(43%)	14.66%	(25%)	13.22%	(38%)	21.79%	(40%)	23.54%	(36%)	23.54%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	-0.49%	1.37%	0.51%	0.00%	0.00%	0.00%	0.00%	0.00%	3.18%	0.00%	0.00%	0.00%	-0.49%	1.37%	0.51%	0.00%	0.00%	0.00%	0.00%	3.18%	0.00%	0.00%	0.00%	0.00%	

### *Les élèves en CM2/6<sup>ème</sup>*

Lorsque nous regardons dans le tableau 50 les résultats du groupe des élèves en classe de CM2 / 6<sup>ème</sup>, le poids de l'interaction entre le QI et la variable temps n'est guère plus grand et toujours pas significatif. Cependant les coefficients de régression du QI montrent un impact significatif sur les compétences en lecture en CM2 qui est dans le test ELFE supérieur à celui observé en CE1/CE2 ( $\beta = 0.039$  par rapport à  $\beta = 0.024$ ). Dans les deux autres tests (SLS et HarmoS) ces coefficients ont seulement une légère tendance à augmenter dans le groupe des lecteurs du CM2.

#### **4.4.2 Analyse des coefficients de détermination**

Les coefficients de déterminations que nous indiquons en bas du tableau 50 nous montrent pour les deux groupes le gain de variance expliquée grâce à l'introduction du QI dans le modèle (modèle 2.1) par rapport au modèle vide. Encore une fois nous nous concentrons sur l'explication de la variance résiduelle de la constante qui est marquée en noir. A droite de ces coefficients de détermination noirs, sont indiqués en gris les coefficients correspondants du modèle 1.2 (issus des tableaux 38.2 à 44.2) contenant les deux facteurs D et CO de la *SVRT*. La comparaison montre que pour le test ELFE et HarmoS l'apport du QI, dans l'explication de variance résiduelle, augmente dans le groupe des lecteurs plus âgés. Nous voyons également que c'est surtout au niveau des classes, que le QI apporte le plus d'explication de la variance résiduelle de la constante (ELFE = 79.17%, HarmoS = 52.38%). Au niveau des individus l'effet est moins important.

Ceci nous conduit à la conclusion que le QI en tant que seul prédicteur, est d'une valeur explicative supérieure au décodage et à la compréhension orale en ce qui concerne le niveau de compétence en lecture (mesuré par l'ELFE et HarmoS) des différentes classes de CE2 / 6<sup>ème</sup>. Les deux facteurs de la *SVRT* ont cependant un pouvoir explicatif supérieur en ce qui concerne les variances en compétences en lecture entre les individus en CE2 / 6<sup>ème</sup> et l'échantillon des élèves en classe de CM1 / CM2.

Les tableaux 51 à 53 sur les pages suivantes nous permettent d'analyser l'éventuelle amélioration du modèle de la *SVRT*, lorsqu'on y ajoute le facteur de l'intelligence. Ils montrent les coefficients de régression du QI en tant que troisième facteur explicatif. Ces coefficients sont significatifs pour les trois tests, dans les deux groupes d'âge de notre échantillon. Lorsque nous regardons de plus près le groupe des élèves en CM2 /

6<sup>ème</sup>, les coefficients du QI sont particulièrement élevés pour le test de l'ELFE et de HarmoS.

Tableau 51 Evolution du modèle 1.2 au modèle 2

Variables réponses	CE1 / CE 2				CM2 / 6ème			
	SLS :		SLS		SLS :		SLS :	
	Modèle 1.2		Modèle 2		Modèle 1.2		Modèle 2	
	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE
<b>Variables explicatives (Parties fixes)</b>								
Constante	-0.091	0.065	-0.119	0.066	-0.078	0.08	-0.111	0.078
Temps (compétences t1-t2)	0.208	0.06	0.256***	0.062	0.142	0.06	0.196***	0.059
Décodage (D)	0.766	0.041	0.721***	0.042	0.396	0.04	0.361***	0.04
Interaction D & Temps	-0.162	0.044	-0.147**	0.045	0.131	0.061	0.125*	0.061
Compréhension Orale (CO)	0.071	0.032	0.047	0.032	0.229	0.043	0.183***	0.045
Interaction CO & Temps	0.013	0.039	0.017	0.04	-0.042	0.051	-0.042	0.054
Intelligence (QI)			0.007***	0.002			0.011***	0.003
Interaction QI & Temps			-0.001	0.002			0	0.004
<b>Variances (Parties aléatoires)</b>								
Niveau III: Classes								
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.049	0.022	0.051	0.022	0.066	0.033	0.061	0.031
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.026	0.015	-0.027	0.016	-0.017	0.017	-0.016	0.015
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.028	0.016	0.03	0.016	0.008	0.013	0.006	0.012
Niveau II: Individus								
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.218	0.019	0.213	0.018	0.414	0.037	0.41	0.037
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.105	0.016	-0.098	0.016	-0.181	0.031	-0.165	0.03
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.251	0.023	0.239	0.022	0.449	0.043	0.423	0.04
-2*loglikelihood:	774.828	13.25	761.578		975.594	12.448	963.146	
Test de Chi2		df=2				df=2		
		0.001***				0.002**		
N° d'Unités Niv. III (Classes)	17		17		15		15	
N° d'Unités Niv. II (Individus)	318		318		267		267	
N° d'Unités Niv. I (Temps)	571		571		500		500	

\* le resultat est significatif avec un niveau d'erreur de  $\alpha = 5\%$ ; \*\*  $\alpha = 1\%$ ; \*\*\*  $\alpha=0.1\%$

Niveau III: Classes	Modèle 1.2 → 2	Modèle 1.2 → 2
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	4.00%	7.58%
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	-7.14%	25.00%
Niveau II: Individus		
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	2.29%	0.97%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	4.78%	5.79%

Selon les résultats présentés dans le tableau 51 ci-dessus, le QI augmente le pouvoir explicatif du modèle à l'égard de la variance résiduelle de la constante au niveau des classes de seulement 7.58%, au maximum. A l'échelle des individus, l'apport est presque négligeable.

Tableau 52 Evolution du modèle 1.2 au modèle 2 pour l'ELFE

Variables réponses	CE1 / CE 2				CM2 / 6ème			
	ELFE: Modèle 1.2		ELFE: Modèle 2		ELFE: Modèle 1.2		ELFE: Modèle 2	
	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE
<b>Variables explicatives (Parties fixes)</b>								
Constante	-0.112	0.053	-0.153***	0.052	0.013	0.082	-0.048	0.067
Temps (compétences t1-t2)	0.21	0.059	0.29***	0.063	-0.042	0.111	0.042	0.107
					0	0	0	0
Décodage (D)	0.535	0.044	0.469***	0.045	0.369	0.051	0.303***	0.049
Interaction D & Temps	-0.032	0.051	-0.015	0.052	0.022	0.066	-0.047	0.066
Compréhension Orale (CO)	0.187	0.037	0.153***	0.037	0.414	0.055	0.319***	0.055
Interaction CO & Temps	0.034	0.05	0.015	0.051	-0.153	0.059	-0.103	0.061
Intelligence (QI)			0.012***	0.002			0.023***	0.004
Interaction QI & Temps			0.002	0.003			-0.01*	0.004
<b>Variations (Parties aléatoires)</b>								
Niveau III: Classes								
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.022	0.014	0.02	0.013	0.058	0.034	0.026	0.022
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	0.027	0.01	0.029	0.011	-0.053	0.038	-0.038	0.029
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.008	0.013	0.015	0.015	0.128	0.06	0.116	0.055
Niveau II: Individus								
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.31	0.027	0.296	0.025	0.612	0.055	0.567	0.051
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.157	0.026	-0.153	0.025	-0.393	0.046	-0.357	0.043
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.492	0.044	0.469	0.042	0.576	0.054	0.537	0.051
-2*loglikelihood:	1072.66	28.325	1044.33		1063.1	34.369	1028.73	
		df=2				df=2		
Test de Chi <sup>2</sup>		0.000	***			0.000	***	
N° d'Unités Niv. III (Classes)	17		17		15		15	
N° d'Unités Niv. II (Individus)	318		318		267		267	
N° d'Unités Niv. I (Temps)	580		580		502		502	

\* le resultat est significatif avec un niveau d'erreur de  $\alpha = 5\%$ ; \*\*  $\alpha = 1\%$ ; \*\*\*  $\alpha=0.1\%$

Niveau III: Classes	Modèle 1.2 → 2	Modèle 1.2 → 2
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	9.09%	55.17%
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	-87.50%	9.38%
Niveau II: Individus		
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	4.52%	7.35%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	4.67%	6.77%

Comme les résultats du tableau 50 l'ont déjà suggéré, l'introduction du QI augmente le pouvoir explicatif de la SVRT de 4 à 7 % au niveau des individus. Au niveau des classes de CM2, cette augmentation est de 55 %. Le pseudo R<sup>2</sup> de - 87% pour la variance résiduelle relative au temps, nous rappelle que notre modèle n'est pas adéquat pour la prise en compte de la dimension développementale des compétences en lecture, mesurées au cours de deux recueils de données.

Tableau 53 Evolution du modèle 1.2 au modèle 2 pour le test de HarmoS

Variables réponses	CE1 / CE 2				CM2 / 6ème			
	HarmoS: Modèle 1.2		HarmoS: Modèle 2		HarmoS: Modèle 1.2		HarmoS: Modèle 2	
	$\beta$	SE	$\beta$	SE	$\beta$	SE	$\beta$	SE
<b>Variables explicatives (Parties fixes)</b>								
Constante	-0.214	0.115	-0.259*	0.116	-0.7	0.078	-0.746***	0.066
Temps (compétences t1-t2)	0.377	0.115	0.466***	0.114	0.456	0.082	0.527***	0.078
Décodage (D)	0.255	0.059	0.182**	0.061	0.208	0.038	0.162***	0.037
Interaction D & Temps	0.082	0.069	0.102	0.071	0.09	0.058	0.094	0.057
Compréhension Orale (CO)	0.339	0.047	0.301***	0.047	0.312	0.041	0.237***	0.041
Interaction CO & Temps	-0.03	0.06	-0.048	0.061	0.042	0.05	0.052	0.052
Intelligence (QI)			0.012***	0.003			0.016***	0.003
Interaction QI & Temps			0.002	0.004			-0.002	0.003
<b>Variances (Parties aléatoires)</b>								
Niveau III: Classes								
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.178	0.071	0.178	0.07	0.069	0.032	0.045	0.022
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.114	0.059	-0.112	0.058	-0.035	0.026	-0.022	0.02
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.147	0.066	0.145	0.064	0.06	0.032	0.051	0.028
Niveau II: Individus								
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	0.395	0.034	0.374	0.033	0.307	0.028	0.278	0.025
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.304	0.037	-0.297	0.035	-0.233	0.029	-0.213	0.026
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	0.628	0.057	0.609	0.055	0.442	0.042	0.409	0.039
-2*loglikelihood:	1147.9	28.933	1118.97		834.688	47.631	787.057	
		df=2				df=2		
Test de Chi <sup>2</sup>		0.000***				0.000***		
N° d'Unités Niv. III (Classes)	17		17		15		15	
N° d'Unités Niv. II (Individus)	317		317		266		266	
N° d'Unités Niv. I (Temps)	570		570		499		499	

\* le resultat est significatif avec un niveau d'erreur de  $\alpha = 5\%$ ; \*\*  $\alpha = 1\%$ ; \*\*\*  $\alpha=0.1\%$

Niveau III: Classes	Modèle 1.2 → 2	Modèle 1.2 → 2
$\sigma^2_{v0}$ = Variance résiduelle $\beta_0$ (Constante)	0.00%	34.78%
$\sigma^2_{v1}$ = Variance résiduelle de $\beta_1$ (Temps)	1.36%	15.00%
Niveau II: Individus		
$\sigma^2_{u0}$ = Variance résiduelle $\beta_0$ (Constante)	5.32%	9.45%
$\sigma^2_{u1}$ = Variance résiduelle de $\beta_1$ (Temps)	3.03%	7.47%

Comme dans le cas des résultats au test ELFE, l'apport du QI pour expliquer la variance résiduelle des résultats du test HarmoS est très inégale. Un apport remarquable au niveau des classes en CM2 /6<sup>ème</sup> va de pair avec un pseudo R<sup>2</sup> de zéro au niveau des classes en CE1 / CE2. A l'échelle des individus, l'apport complémentaire du QI est d'environ 5% chez les plus jeunes (CE1/CE2), et de 9% dans le groupe des plus grands (CM2/6<sup>ème</sup>).

Tableau 54 Analyse des coefficients de régression

CE1 / CE 2

Niveau III: Classes	SLS	SLS	ELFE	ELFE	HarmoS	HarmoS
$\sigma^2_{v_0}$ = Variance résiduelle $\beta_0$ (Constante)	78.88%	76.72%	79.05%	72.38%	52.94%	40.64%
Niveau II: Individus	Modèle 0 → 1.2	Modèle 0 → 2.2	Modèle 0 → 1.2	Modèle 0 → 2.2	Modèle 0 → 1.2	Modèle 0 → 2.2
$\sigma^2_{u_0}$ = Variance résiduelle $\beta_0$ (Constante)	60.79%	61.15%	43.22%	42.86%	25.75%	19.17%

CM2 / 6ème

Niveau III: Classes	SLS	SLS	ELFE	ELFE	HarmoS	HarmoS
$\sigma^2_{v_0}$ = Variance résiduelle $\beta_0$ (Constante)	25.84%	33.71%	39.58%	65.63%	34.29%	49.52%
Niveau II: Individus	Modèle 0 → 1.2	Modèle 0 → 2.2	Modèle 0 → 1.2	Modèle 0 → 2.2	Modèle 0 → 1.2	Modèle 0 → 2.2
$\sigma^2_{u_0}$ = Variance résiduelle $\beta_0$ (Constante)	45.81%	41.62%	40.98%	36.84%	36.04%	33.13%

Tiu et al. (2003) avaient trouvé dans les analyses de leur échantillon un pouvoir de détermination du QI qui égalait, voire surpassait celui de la compréhension orale, tandis que le décodage restait le premier facteur du modèle. Afin de pouvoir répondre de façon exhaustive à la question de la valeur ajoutée du QI dans un modèle de la *SVRT*, nous avons construit un modèle 2.2, qui introduit le QI en tant que deuxième prédicteur, après le facteur décodage. La compréhension orale n'était pas intégrée dans ce modèle. Le tableau 52, ci-contre, compare le pouvoir explicatif de ce nouveau modèle 2.2. (D et QI) avec le pouvoir explicatif du modèle 1.2 de la *SVRT* (contenant les facteurs D et CO) sur la variance résiduelle de la constante.

Lorsqu'on regarde les résultats dans le groupe des CE1 / CE2 en haut du tableau 54, les parts de variance expliquées des deux différents modèles sont assez proches pour les tests SLS et ELFE. Pour le test de HarmoS, le modèle 1.2 de la *SVRT* arrive à expliquer une plus grande partie de variance résiduelle (surtout au niveau des classes) que le nouveau modèle qui remplace la Compréhension Orale par le QI.

Dans le groupe des élèves de CE1 / CE2, les résultats se distinguent selon le niveau de l'analyse. Au niveau des classes, c'est le modèle avec le QI (modèle 2.2) qui explique une plus grande partie de variance résiduelle de la constante ; au niveau des individus, c'est le modèle de la *SVRT* qui a les meilleurs résultats.

*Résumé:*

*Lorsque nous introduisons le QI en tant que troisième facteur indépendant, après le décodage et la compréhension orale, dans notre modèle, nous observons qu'il est pour chacun de nos trois tests de lecture un prédicteur significatif. Son apport en terme d'explication de variance résiduelle se vérifie particulièrement dans l'échantillon CM2/6<sup>ème</sup> et au niveau des classes, moins au niveau des individus.*

*Lorsque nous remplaçons le facteur de la compréhension orale par le QI en créant ainsi un nouveau modèle, nous remarquons que le modèle « classique » de la *SVRT* a une meilleure force explicative dans le groupe des élèves de CE1/CE2 et en particulier lorsqu'il s'agit de la lecture de texte authentiques (test HarmoS).*

*Dans le groupe des élèves de CM2/6<sup>ème</sup>, le nouveau modèle contenant le QI à la place de la compréhension orale explique plus de variance résiduelle au niveau des classes. Au niveau des individus, c'est le modèle de la *SVRT* qui est plus performant.*

*Résumé des résultats à l'égard des hypothèses 3 et 4*

En nous appuyant sur les nombreuses analyses effectuées, nous pouvons conclure que le QI n'améliore pas suffisamment la qualité du modèle pour justifier son introduction dans le modèle la *SVRT*. Surtout dans le groupe des plus jeunes lecteurs, l'effet du QI est limité, tandis que les compétences en lecture des enfants plus âgés de notre échantillon semblent plus influencées par les capacités intellectuelles telles que nous les avons mesurées.

L'impact du QI dans nos modèles dépend aussi, outre l'âge des enfants, du test de lecture qui était utilisé. Le test de lecture standardisé ELFE mesure, à l'aide de questions à choix multiples, la capacité à lire et comprendre des textes très courts sous contrainte de temps. Cela implique de comprendre des anaphores, de relier et de déduire des informations de plusieurs phrases. Ce format de test requiert probablement plus de capacités intellectuelles et linguistiques que la lecture et les réponses aux phrases simples du test SLS. Les deux tests sont standardisés, ils ont avec la contrainte du temps et les réponses à choix multiple un format « typique » de test qu'on retrouve dans le test d'intelligence CFT 20R.

Le test de HarmoS est très probablement plus proche d'une lecture « naturelle » que le test ELFE ou SLS. Le test de HarmoS ne s'effectue en revanche pas sous contrainte de temps. Il utilise des textes narratifs ou expositifs authentiques. Il offre la possibilité d'une relecture et propose un format de réponse mêlant à la fois des questions ouvertes et des questions à choix multiples. L'explication et la prédiction de ses résultats peuvent être améliorées dans une certaine mesure au niveau des classes (en CM2), mais il apporte peu au niveau des individus. Or, lorsqu'il s'agit d'identifier des enfants ayant des troubles d'apprentissage de la lecture, on travaille essentiellement au niveau de l'individu. Ce sont alors les facteurs de la *SVRT* qui offrent à la fois une meilleure explication, mais aussi des pistes plus concrètes pour la remédiation que des facteurs relatifs aux compétences de raisonnement intellectuelles générales, opérationnalisés par le QI.

## 5 Discussion générale et conclusion

Avant de discuter en détail les résultats présentés dans la partie précédente, nous allons brièvement rappeler les objectifs de notre travail de recherche.

Le principal objectif que nous nous étions assigné était de vérifier la validité d'un modèle additif de la « *Simple View of Reading Theory* » (*SVRT*), uniquement constitué des facteurs décodage (D) et compréhension orale (CO). Pour ce faire, nous avons construit des modèles multi-niveaux en introduisant les variables D et CO comme facteurs prédictifs. Trois tests de lecture différents (SLS, ELFE et HarmoS) ont constitué les variables dépendantes de nos modèles d'analyses. Nous avons alors analysé la qualité d'un modèle additif de la « *Simple View of Reading Theory* » :

- dans deux échantillons d'élèves d'âges différents (CE1 / CE2 et CM2 / 6<sup>ème</sup>) ;
- dans des groupes d'élèves plurilingues par rapport aux monolingues ;

et ce tout en variant systématiquement la façon dont les compétences de lecture étaient opérationnalisées à travers les trois différentes variables dépendantes.

Le deuxième objectif consistait, quant à lui, à analyser la valeur ajoutée apportée par l'introduction du facteur « intelligence » pour le modèle de la *SVRT*, reprenant ainsi une proposition élaborée par Tiu, Thompson et Lewis (2003).

### 5.1 Résumé des résultats et interprétations

Nos deux premières hypothèses concernaient la qualité et le pouvoir d'explication d'un modèle constitué des facteurs décodage et compréhension orale ainsi que l'apport de chacun de ces deux facteurs dans le groupe des élèves plus jeunes (CE1/CE2) par rapport à celui des élèves plus âgés (CM2/6<sup>ème</sup>). Notre modèle multi-niveaux ne comportait pas le produit des facteurs D x CO, en raison des résultats très divergents sur la nécessité d'une telle multiplication. Ce modèle nous a permis d'évaluer le postulat de la *svrt* selon lequel les facteurs D et CO à eux deux expliquent suffisamment les compétences en lecture.

### 5.1.1 Impact et force explicative des facteurs D et CO

Les résultats que nous avons obtenu au regard de notre première hypothèse H1a confirment que les deux facteurs décodage et compréhension orale ont des apports significatifs et déterminants pour l'explication des compétences en lecture. Dans chacun de nos modèles multi-niveaux, les facteurs D et CO ont des coefficients de régression très significatifs. Selon le test de lecture utilisé, ces deux facteurs suffisent pour expliquer 50% à 79% des variances en compétences de lecture entre les classes de CE1/CE2 et 25% à 60 % des variances entre les élèves.

Conformément à notre hypothèse H1b, l'apport des deux facteurs dépend de l'âge des lecteurs. Bien qu'ayant toujours des coefficients de régression significatifs, le pouvoir d'explication est généralement plus faible dans le groupe des lecteurs en CM2/6<sup>ème</sup> que dans le groupe des CE1/CE2. L'explication de variance résiduelle qu'on atteint dans ce groupe plus âgé, varie entre 25% et 39% au niveau des classes et entre 36% et 45 % au niveau des individus selon le test de lecture utilisé. Nos résultats montrent bien que les deux facteurs D et CO mis en avant par la *SVRT* se prêtent bien à la modélisation des compétences des jeunes lecteurs débutants. Pour les lecteurs de 10 à 12 ans qui sont au milieu de leur parcours scolaire, D et CO semblent à eux seuls être des facteurs explicatifs insuffisants.

Lorsqu'on regarde l'apport de chaque facteur individuellement, on perçoit un gain d'importance de la compréhension orale dans le groupe des lecteurs plus âgés (CM2/6<sup>ème</sup>) ainsi qu'une importante diminution de l'apport du décodage, du moins dans les tests SLS et ELFE pour lesquels le décodage avait une forte influence dans le groupe des plus jeunes lecteurs. .

Nos résultats confirment ainsi les travaux de Tilstra, McMaster, Van den Broek, Kendeou et Rapp (2009) qui avaient également constaté une perte de pouvoir explicatif de leur modèle de la *SVRT* dans un échantillon composé de trois groupes d'âges (9 / 10 ans, 12 / 13 ans et 14 / 15 ans). Pour l'expliquer, les auteurs avancent qu'avec l'âge des élèves, les textes et les tâches de lecture auxquels ils sont confrontés, sont de plus en plus élaborés et font intervenir des processus cognitifs de plus en plus complexes. Ils supposent aussi que les élèves plus âgés peuvent davantage palier les difficultés de lecture par leurs préacquis.

Savage et Wolforth (2007) ont conclu, grâce à leurs résultats de recherche, que la « *Simple View of Reading Theory* » gardait un pouvoir explicatif important au delà des années de scolarisation obligatoire. Cette contradiction des résultats peut cependant s'expliquer par le fait que l'étude de Savage et Wolforth portait sur des étudiants d'université (N=60) souffrant d'importants retards dans l'acquis des compétences en lecture.

Concernant les résultats que nous avons obtenus, les pistes de réflexion proposées par Tilstra et al. (2009) nous paraissent intéressantes. Cependant nous devons ajouter qu'au sein de notre échantillon et de nos modèles statistiques, cette perte d'influence concerne aussi les résultats pour le même test de lecture. Au delà de la nature des textes qui change, il semble que le modèle même de la *SVRT* ne soit plus en mesure de modéliser les composantes du processus de lecture. Ainsi le test ELFE est identique pour nos deux groupes d'âge et il parvient à expliquer environ 79 % de la variance résiduelle entre les classes et 43% entre les individus dans le groupe des CE1 / CE2 alors qu'il explique seulement 39 % (entre classes) et 40 % (entre individus) dans le groupe des CM2 / 6<sup>ème</sup>.

Ces résultats sont à nos yeux essentiellement liés à la perte d'influence du facteur décodage qui n'est que partiellement compensé par une augmentation de l'influence du deuxième facteur compréhension orale. Une analyse descriptive des résultats de décodage montre qu'en CM2 / 6<sup>ème</sup>, les capacités de décodage de la grande majorité des élèves sont très développés. En moyenne, les élèves de notre échantillon plus âgés atteignent 52 points sur 72 points possibles (en CE1 la moyenne est de 23 sur 72 points). La valeur maximale de 72 points est même atteinte par 7 % des élèves en CM2. Pour ces élèves, il existe donc un effet de plafond. Une grande partie des élèves de ce groupe ne se distingue plus suffisamment dans sa capacité à déchiffrer rapidement des mots, pour faire du décodage un prédicteur de haute importance. Le modèle de la *SVRT* n'étant composé que de deux facteurs, la perte d'influence de l'un entre eux, remet en question la qualité du modèle dans son ensemble.

### 5.1.2 L'impact de l'opérationnalisation des compétences en lecture

Les résultats que nous venons de discuter, doivent cependant être nuancés. En effet, nous constatons, conformément à nos hypothèses H2a et H2b, une forte influence de la façon dont les compétences en lecture sont opérationnalisées.

Le pouvoir de prédiction et d'explication de notre modèle de la *SVRT* varie en effet beaucoup suivant les trois tests SLS, ELFE et HarmoS. Comme nous l'avons supposé à la vue des recherches menés autour de la *SVRT*, ce modèle explique mieux les compétences en lecture rapide, de textes plutôt courts comme le propose le test ELFE. Dans nos deux groupes d'âge, c'est pour les le test ELFE que notre modèle de la *SVRT*, atteint des parts de variance expliquée les plus élevées au niveau des classes (79 % en CE1/CE2 et 39 % en CM2/6<sup>ème</sup>). Le taux de variance expliquée au niveau des individus est de 43 % en CE1/CE2 et de 40 % en CM2/6<sup>ème</sup>.

Nous observons que le pouvoir explicatif de notre modèle reste stable au niveau des individus, mais diminue beaucoup au niveau des classes. La lecture rapide de phrases proposée par le test SLS montre des résultats similaires. En CE1/CE2 les variances au niveau des classes (78%) et des individus (60%) sont également bien expliquées par notre modèle avec un impact très important du facteur décodage. Mais le pouvoir explicatif du modèle pour le SLS chute à seulement 25% (classes) voire 45 % (individus) dans le groupe des CM2/6<sup>ème</sup>. En ce qui concerne les variances entre les classes, les compétences requises par le SLS en matière de compréhension de langue ne sont probablement pas assez complexes pour faire de la compréhension orale un facteur explicatif qui compense la perte d'influence du décodage.

La comparaison des résultats du HarmoS entre les deux groupes d'âge requiert plus de prudence que dans le cas du SLS ou ELFE, puisque pour le test de HarmoS, la nature des textes change en termes de vocabulaire, de complexité et de longueur selon le groupe d'âge de notre échantillon. Cependant une lecture de textes authentiques, sans contrainte de temps, offrant la possibilité de retourner au texte en cas de doute et proposant des questions ouvertes, correspond à une tâche proche de la lecture en situation réelle.

Le pouvoir explicatif de notre modèle de la *SVRT* pour les résultats du test HarmoS est moindre que pour les tests ELFE et SLS. Une exception existe cependant au niveau des classes de CM2/ 6<sup>ème</sup>; la variance résiduelle expliquée des résultats du test HarmoS est

de 34%. Pour les résultats du SLS, elle est de 25%. Les résultats du test HarmoS montrent deux autres exceptions par rapport au SLS et ELFE :

- 1) les coefficients de régression du facteur décodage restent stables entre les deux groupes d'âge et ;
- 2) au niveau des individus, le pouvoir d'explication du modèle passe de 25% en CE1/CE2 à 36 % en CM2/6<sup>ème</sup>.

Si les prémisses de la « *Simple View of Reading Theory* » correspondent bien aux observations faites sur les résultats des tests ELFE et SLS, c'est moins le cas avec les résultats du test de HarmoS.

Nos résultats confirment ainsi les analyses de Cutting et Scarborough (2006), ainsi que ceux de Keenan, Betjeman et Olson (2008). Les différences entre divers tests de lecture peuvent être très importantes et mènent par voie de conséquence à des résultats divergents. Notre modèle de la *SVRT* s'applique assez bien à un test standardisé comme ELFE qui est relativement facile à mettre en œuvre, et qui est couramment utilisé par les chercheurs et les enseignants. Toutefois, lorsqu'on observe les résultats obtenus pour le test de HarmoS non standardisé avec des textes authentiques on peut émettre des doutes quant à la validité de notre modèle de la *SVRT* et quant au fait que seulement deux facteurs D et CO suffisent à expliquer les compétences en lecture de manière satisfaisante.

La majorité des études opérationnalise les compétences de lecture de textes comme le résultat de tests standardisés, consistant dans la lecture de passages de textes courts, suite auxquels le lecteur doit répondre à des questions dans un temps limité. Cette pratique ne semble pas pouvoir capter tout ce que le terme « compréhension de lecture » induit. Il paraît en effet indispensable que des textes plus longs soient utilisés, que des processus de construction de sens plus complexes soient mesurés, que des questions très proches du texte, ouvertes et même ambiguës soient employées. Cela rendrait cependant plus complexe la réalisation de tests sur de grands échantillons. L'interprétation des résultats, en particulier lorsqu'on utilise des questions ouvertes, et l'objectivité dans l'évaluation étant plus difficile à garantir.

A l'inverse, les divergences entre les résultats des différents tests alimentent facilement le désaccord dans la communauté scientifique lorsqu'il s'agit de mettre à l'épreuve une nouvelle théorie. Des tests de lecture comme ELFE et HarmoS semblent viser des processus cognitifs sous-jacents différents, ce qui rend difficile un diagnostic valable des

besoins spécifiques d'un jeune lecteur en difficulté. Il nous semble ainsi indispensable que les différentes opérationnalisations des compétences de lecture proposées par des tests différents soient d'avantage analysées afin de pouvoir identifier clairement les processus sous-jacents ciblés par chaque test.

### 5.1.3 Elèves plurilingues et monolingues

Gough et Tunmer avaient utilisé dans leur étude fondatrice de la *svrt* un échantillon composé de bilingues majoritairement anglais-espagnol. Les études concernant la *svrt* qui suivirent, privilégiaient, en général, des échantillons monolingues anglais ou des échantillons mixtes plurilingues et monolingues, sans comparer de façon systématique ces deux groupes. A nos yeux, l'origine linguistique est un facteur qui nécessite d'être pris en compte. Kirby et Savage (2008), considèrent même que la question de la validité de la *SVRT*, selon qu'il s'agit d'une lecture en L1 ou L2 n'est pas résolue.

Nous pensons cependant avoir pu apporter avec nos résultats une réponse partielle à cette question. Elle est partielle, car elle repose sur un échantillon de plurilingues qui exclue les enfants immigrés récemment et qu'elle se focalise uniquement sur les élèves de CE1/CE2. Notre groupe d'élèves plurilingues a passé la plupart de sa vie dans l'environnement germanophone suisse, et correspond ainsi à la situation d'une majorité des enfants plurilingues. Par voie de conséquence, tous nos élèves plurilingues étaient familiarisés avec l'allemand, même si une différence significative en faveur des monolingues persistait dans les compétences en compréhension orale. Cette différence n'a cependant pas de conséquence sur la validité de notre modèle de la *SVRT* pour le groupe des plurilingues.

En effet, d'après nos analyses, le modèle de la *SVRT* tel que nous l'avons testé s'applique de la même manière au groupe d'élèves monolingues qu'à celui d'élèves plurilingues. Les légères différences qui existent concernent l'importance relative des deux facteurs dans les trois tests de compétence de lecture (Harmos, ELFE et SLS). Le modèle de la *SVRT* capture donc des composantes qui jouent un rôle important dans le processus de lecture en L1 aussi bien qu'en L2.

Ayant ainsi pu confirmer notre hypothèse H3a, nous devons cependant rejeter notre hypothèse H3b qui postulait un impact plus important de la compréhension orale dans le groupe des élèves plurilingues. Nous avons vu que les compétences en compréhension

orale (pour l'allemand) sont plus faibles chez les élèves plurilingues. Contrairement à notre hypothèse cela ne se reflète pas dans les compétences en lecture.

L'apport de notre étude se situe aussi au niveau d'une validation « inter-linguistique » de la *SVRT*. Peu d'études concernant la *SVRT* ont été réalisées auprès d'un échantillon germanophone (Marx & Jungmann, 2000; Rost & Hartmann, 1992).

Les études qui évaluent l'impact de la transparence orthographique montrent une meilleure maîtrise du processus de décodage dans des langues, comme l'allemand, ayant une correspondance grapho-phonique relativement régulière (Wimmer 1993, Wimmer et al. 1990, 1998, 2000). Dans le groupe des lecteurs plus âgés, cela peut être l'une des raisons pour la diminution importante de l'apport du décodage et de la qualité globale du modèle de la *SVRT*. Cela ne rend pas la composante du décodage inutile pour autant, du moins dans le groupe des très jeunes lecteurs et pour une lecture de textes courts qui met l'accent sur la rapidité comme le SLS ou l'ELFE.

#### **5.1.4 L'apport du QI comme 3<sup>ème</sup> facteur**

Les études concernant la *SVRT* ont démontré l'importance des facteurs D et CO pour les compétences en lecture. Cependant, ces deux facteurs ne peuvent prétendre à l'exhaustivité, et plusieurs auteurs se sont intéressés à l'identification de facteurs déterminants supplémentaires. Selon certaines études (Savage, 2006 ; Tiu et al. 2003) le QI pouvait être un facteur explicatif important. Tiu et al. (2003) proposaient même de remplacer le facteur CO par le QI, sachant que dans leur échantillon d'élèves en 5<sup>ème</sup> année de scolarisation (CM2), c'est la combinaison des composantes D et QI qui montre la plus grande force explicative.

Dans la présente étude, nous avons introduit le QI en tant que troisième facteur, après le décodage et la compréhension orale. Les coefficients de régression obtenus pour le facteur QI sont significatifs dans les deux groupes d'âge de notre échantillon. Nous avons cependant observé, dans l'explication des variances résiduelles de la constante, que l'apport du QI est nettement plus important pour le groupe des élèves en CM2/6<sup>ème</sup>. Nous avons également pu montrer que le pouvoir d'explication gagné avec l'introduction du QI varie, en fonction du test de lecture utilisé, entre moins de 10 % (SLS) et 55% (ELFE).

A la vue de nos résultats, nous ne pouvons donc pas rejoindre Tiu et al (2003) dans leur proposition de remplacer le facteur compréhension orale par le QI dans le modèle de la *SVRT*. En particulier chez élèves plus jeunes, le QI n'est pas un meilleur prédicteur que la compréhension orale lorsqu'il est introduit après le décodage. En effet, dans le groupe des élèves en CE1/CE2, les parts de variance expliquée pour les tests SLS et ELFE ne sont pas plus grandes pour un modèle contenant les facteurs décodage et QI que pour notre modèle de la *SVRT* (Décodage et Compréhension Orale).

Pour le test HarmoS c'est le modèle de la *SVRT* qui atteint une meilleure explication de variances résiduelles.

Dans le groupe des élèves de CM2/6<sup>ème</sup>, les résultats diffèrent selon le niveau des analyses. Nous constatons que l'explication des variances entre les classes profite davantage de l'introduction du QI que de l'introduction de la compréhension orale. Les différences entre les individus sont à l'inverse mieux expliquées par le modèle de la *SVRT*, qui propose avec le décodage et la compréhension orale deux facteurs que nous supposons être plus intrinsèquement liés au processus de la lecture que les capacités cognitives liés à la conception d'une intelligence fluide. Compte tenu du fait que dans la pratique les professionnels sont le plus souvent intéressés par les difficultés de certains enfants jeunes lecteurs, nos résultats peuvent être interprétés comme un encouragement à utiliser le modèle de la *SVRT* sans nécessairement y inclure de mesure de l'intelligence.

Savage (2001) présente des arguments qui vont au-delà des résultats empiriques. Nous devons impérativement en tenir compte lorsque nous discutons la relation entre le QI et les mesures des compétences de lecture. En nous rappelant les liens intrinsèques entre le langage et les concepts utilisés Savage déclare « Most pernicious in this vocabulary is the use of 'ability' and 'IQ'. The assumptions often conveyed by the use of these terms, that potential is fixed and immutable, is one that is most likely to get in the way of interventions for those assessed as having 'low ability' <sup>48</sup>» (Savage, 2001 p. 27). Le facteur de la compréhension orale a bien moins une connotation déterministe que les divers tests de capacités cognitives générales. D'ailleurs, et c'est ce que démontrent Stanovich (1986, 1988, 1991, 1993) Siegel (19989, 1993), et Savage (2001), l'utilisation de la

---

<sup>48</sup> « L'utilisation des termes QI et « capacité » est particulièrement difficile. Souvent ces termes évoquent l'idée d'un potentiel stable et immuable, une telle idée va à l'encontre des interventions faites pour ceux qui ont été catégorisé comme ayant des « faible capacité » (traduction de Nora Dittmann-Domenichini).

compréhension orale pour mesurer les processus cognitifs hautement complexes impliqués dans la compréhension de langue nous paraît plus justifiable que celle du QI – du moins pour le groupe des jeunes lecteurs.

Ces résultats illustrent encore une fois que chez les élèves plus âgés, la *SVRT* a plus de mal à modéliser les processus de lecture. Parallèlement à une baisse générale de la force explicative de notre modèle de la *SVRT* dans le groupe des élèves en CM2 / 6<sup>ème</sup>, nous constatons une augmentation de l'impact d'un facteur de mesure de l'intelligence fluide. Etant loin d'atteindre les 90 à 100%, la part de variance expliquée dans nos analyses nécessite la prise en compte d'autres facteurs. Nos résultats peuvent donc être rapprochés de ceux de Kershaw et Schatschneider (2011): au moment où les enfants passent de l'apprentissage de la lecture à l'apprentissage par la lecture, les processus et les facteurs impliqués dans la lecture changent.

## 5.2 Limites des études et recherches futures

Nous allons brièvement résumer l'apport de la présente étude, avant d'aborder plus en détail les limites de notre travail, qui serviront en même temps de point de départ pour nos travaux et recherches futures.

Outre le fait d'avoir utilisé un échantillon germanophone pour vérifier une théorie répandue surtout dans la recherche anglo-américaine et d'avoir procédé à une comparaison systématique entre des élèves monolingues et plurilingues, l'apport de notre travail se situe au niveau méthodologique. Sans une analyse à niveaux multiples, « l'effet de grappe » n'aurait pas été contrôlé, les différences de résultats au niveau des individus et des classes auraient été masquées. Nos résultats témoignent de l'importance de pouvoir distinguer les variances résiduelles au niveau des individus des variances résiduelles au niveau des classes.

L'utilisation d'une nouvelle méthode implique des erreurs, des questionnements et des incertitudes qu'il faut cependant surmonter afin de construire un modèle bien pensé même s'il n'est pas irréprochable. Il s'est par exemple avéré que l'introduction de la variable Temps dans nos analyses a rendu notre modèle extrêmement complexe par rapport à la base de données dont nous disposions. Ainsi l'explication des variances résiduelles au niveau du temps a mené à des résultats parfois négatifs, qui nous ont mis en garde et ont suggéré une structure de modèle moins complexe, avec moins de niveaux. Notre objectif initial de tester aussi le pouvoir explicatif de notre modèle atténué de la

*SVRT* par rapport au développement du niveau de compétence en lecture entre le CE1 et le CE 2 (et entre le CM2 et la 6<sup>ème</sup>) a ainsi du être abandonné. Cependant nous souhaitons poursuivre dans cette direction et construire un modèle statistique qui sera en mesure d'intégrer la variable du développement des compétences.

Nos analyses sont de plus partiellement incomplètes, car la structure de nos données ne nous a pas permis d'analyser séparément le groupe des élèves plurilingues et le groupe des élèves monolingues de CM2/6<sup>ème</sup>. En effet, le nombre d'élèves monolingues dans ce groupe d'âge était trop faible pour procéder à des estimations de paramètres par itérations. Ainsi, nos conclusions sur la validité de notre modèle de la *SVRT* dans le groupe des plurilingues et des monolingues, ne se base que sur les élèves de CE1/CE2. Avec un modèle statistique moins complexe tel que nous souhaitons le construire dans nos futurs travaux, cette comparaison pourrait s'étendre au groupe des élèves de CM2/6<sup>ème</sup>.

Notre travail a également été limité par le fait que nous n'avons que partiellement pu réaliser des mesures convergentes pour nos deux facteurs décodage et compréhension orale. Le test de la compréhension orale de HarmoS consistait en deux tâches distinctes (un reportage et une narration courte) qui étaient la plupart du temps réalisées à quelques jours d'intervalle. Nous pourrions donc supposer que le score global obtenu se compose de deux mesures convergentes, ce qui augmente sa fiabilité. En revanche, pour le décodage nous ne disposons pas de deux mesures. Nous devons nous limiter à un seul test en raison de la charge de travail que nos évaluations représentaient pour les nombreux élèves et enseignants qui ont participé à notre étude.

Les tests utilisés pour évaluer les compétences en lecture posent une autre limite à la portée de nos résultats. Nous avons obtenu les meilleurs résultats pour le modèle de la *SVRT* avec les tests SLS et ELFE. Comme nous l'avons dit auparavant, ces deux tests se ressemblent par une évaluation de textes voire de phrases, courtes, sous limite de temps et en utilisant des questions à choix multiple. Or, ces deux tests sont aussi standardisés et normés contrairement au test de HarmoS. Nous ne pouvons pas exclure que la fiabilité des résultats du SLS et ELFE soit plus grande et que cela améliore l'explication des variances lorsque les deux facteurs D et CO sont introduits au modèle statistique.

Finalement, la portée de nos résultats est limitée par le fait qu'ils reposent exclusivement sur des analyses statistiques inférentielles, qui évaluent le pouvoir de prédiction ou d'explication d'un facteur dépendant par de nombreux facteurs indépendants. De nombreuses études que nous avons citées visant à évaluer la validité d'un modèle théorique comme la *SVRT* appliquent comme nous ce paradigme. Or, il manque à cette démarche la preuve « ultime » qui serait apportée par la pratique des enseignants ou des psychologues scolaires et qui démontrerait clairement qu'un programme de remédiation focalisé sur le décodage ou la compréhension orale améliorerait réellement les compétences en lecture.

De ce dernier point, nous retenons donc la possibilité d'orienter nos futurs travaux de recherche vers la réalisation d'une étude d'intervention qui évaluerait l'adaptation de la *SVRT* dans une démarche de remédiation.

La « *Simple View of Reading Theory* » s'est cependant montré être d'après nos analyses une approche fertile et économe à la fois pour analyser les compétences en lecture surtout lorsqu'il s'agit d'enfants jeunes et d'une lecture qui est évaluée par des instruments standardisés avec des textes très courts et sous limite de temps.

## Bibliographie

- Aaron, P. G. (1989) *Dyslexia and hyperlexia. Diagnosis management of developmental reading disabilities*. Dordrecht, Boston, London: Kluwer Academic Publishers.
- Aaron, P. G., Joshi, R. M., Gooden, R., & Bentum, K. E. (2008) Diagnosis and Treatment of Reading Disabilities Based on the Component Model of Reading: An Alternative to the Discrepancy Model of LD. *Journal Learning Disabilities* 41(1), 67-84.
- Adlof, S., M., Catts, H., W. , & Little, T., D. . (2006) Should the Simple View of Reading Include a Fluency Component? *Reading and Writing* 19(9), 933-958.
- Aitkin, M., & Longford, N. (1981) Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society*.
- Alexander, P., A. , & Fox, E. (2004) A Historical Perspective on Reading Research and Practice. In R. B. Ruddell & N. J. Unrau (Eds.), *A Historical Perspective on Reading Research and Practice* (5 ed., pp. 33-68). Newark: International Reading Association.
- Anderson, R. C. (1977) The Notion of Schemata and the Educational Enterprise: General Discussion of the Conference. In R. C. Anderson, R. J. Spiro & W. E. Montague (Eds.), *Schooling and the acquisition of Knowledge*. Hillsdale N.J.: Erlbaum.
- Anderson, R. C. (1978) Schema-Directed Process in Language Comprehension. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema & R. Glaser (Eds.), *Cognitive Psychology and Instruction*. New York: Plenum.
- Anderson, R. C. (1984) Role of the Reader's Schema in Comprehension. In R. C. Anderson, J. Osborn & R. J. Tierney (Eds.), *Learning to Read in American Schools: Basal Readers and Content Text*. Hillsdale N.J.: Erlbaum.
- Aouad, J., & Savage, R. (2009) The Component Structure of Preliteracy Skills: Further Evidence for the Simple View of Reading. *Canadian Journal of School Psychology* 24(2), 183-200.
- Auer, M., Gruber, G., Mayringer, H., & Wimmer, H. (2005) *SLS 5-8. Salzburger Lesescreening für die Klassenstufen 5-8* Göttingen, Bern, Wien: Hogrefe.
- Aust, H. (2006) Entwicklung des Textlesens. In U. Bredel, H. Günther, P. Klotz, J. Ossner & G. Sieber-Ott (Eds.), *Didaktik der deutschen Sprache. Ein Handbuch* (2. ed., Vol. 1, pp. 525-535). Paderborn, München, Wien, Zürich: Ferdinand Schöningh.
- Bartlett, F. C. (1932) *Remembering*. Cambridge: Cambridge University Press.
- Beech, J. R. (1987) Early reading development. In J. R. Beech & A. Colley (Eds.), *Cognitive approaches to reading* (pp. 187-211). Chichester: Wiley.

- Bellin, N. (2009) *Klassenkomposition, Migrationshintergrund und Leistung. Mehrebenenanalysen zum Sprach- und Leseverständnis von Grundschulern*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Berthele, R. (2001) Die vielsprachige Schweiz - Vorbild oder schlechtes Beispiel? In C. De Boot, S. Kroon, P. H. Nelde & H. Van de Velde (Eds.), *Institutional Status and Use of National Languages in Europe*. Sankt Augustin.
- Bowey, J., & Hansen, J. (1994) The development of orthographic rhymes as units of word recognition. *Journal of Experimental Child Psychology* 58, 465-488.
- Braibant, J.-M. (1994) Le décodage et la compréhension. In J. Grégoire & B. Pierrart (Eds.), *Evaluer les troubles de la lecture* (pp. 173-194). Bruxelles: De Boeck Université.
- Bressoux, P. (2010) *Modélisation statistique appliquée aux sciences sociales* (2 ed.). Bruxelles: De Boeck Université.
- Britt, M. A., & Rouet, J.-F. (2009) The comprehension of multiple sources of information: When do readers build documents models?, *EARLI congress* (pp. 27.08.2009). Amsterdam.
- Buck, G. (2001) *Assessing Listening*. Cambridge: Cambridge University Press.
- Byrne, B., & Fielding-Barnsley, R. (1989) Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology* 81(3), 313-321.
- Cain, K., & Oakhill, J. V. (1999) Inference making ability and its relation to comprehension failure in young children. *Reading and Writing* 11(5), 489-503.
- Carver, R. P. (1998) Predicting reading level in grades 1 to 6 from listening level and decoding level: Testing theory relevant to the simple view of reading. *Reading and Writing: An Interdisciplinary Journal* 10, 121-154.
- Cattell, R. B. (1963) Theory of fluid and crystallized intelligence: A critical experiment *Journal of Educational Psychology* 54, 1-22.
- Catts, H. W., Hogan, T. P., & Adloff, S. M. (2005) Developmental Changes in Reading and Reading Disabilities. In H. W. Catts & A. G. Kamhi (Eds.), *The Connections between Language and Reading Disabilities* (pp. 25-40): Psychology Press.
- Chen, R. S., & Vellutino, F. R. (1997) Prediction of Reading Ability. A Cross-Validation of the simple view of reading. *Journal of Literacy Research* 29(1), 1-24.
- Coltheart, M. (1978) Lexical access in simple reading tasks In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London: Academic Press.
- Coltheart, M., Curtis, A., Atkins, B., & Haller, M. (1993) Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review* 100, 589-608.

- Coltheart, M., & Rastle, K. (1994) Serial processing in reading aloud: evidence for dual route models of reading. *Journal of Experimental Psychology: Human Perception and Performance* 20, 1197-1211.
- Conners, F. A. (2008) Attentional Control and the Simple View of Reading. *Reading and Writing*.
- Content, A., & Zesiger, P. (2000) L'acquisition du langage écrit. In J.-A. Rondal & X. Seron (Eds.), *Troubles du langage: bases théoriques, diagnostiques et rééducation* (pp. 179-209). Hayen: Mardaga.
- Courgeau, D. (2004) *Du groupe à l'individu. Synthèse multiniveau*. Paris: INED.
- Courgeau, D., & Baccaini, B. (1997) Analyse multi-niveau en sciences sociales. *Population* 52(4), 831-863.
- Cunningham, A. E., Stanovich, K. E., & Wilson, M. R. (1990) Cognitive variation in adult college students differing in reading ability. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: component skills approaches* (pp. 129-159). San Diego: Academic Press.
- Cutting, L. E., & Scarborough, H. S. (2006) Prediction of Reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive Skills Can Depend on How Comprehension Is Measured. *Scientific Studies of Reading* 10(3), 277-299.
- Danks, J. H., & End, J. L. (1987) Processing strategies for reading and listening. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending Oral and Written Discourse*. San Diego CA: Academic Press.
- Davis, M. (2006) The Simple View of Reading, *Common Knowledge. The Newsletter of the Core Knowledge Foundation* (Vol. 19).
- de Jong, P. F., & van der Leij, A. (2002) Effects of Phonological Abilities and Linguistic Comprehension on the Development of Reading. *Scientific Studies of Reading* 6(1), 51-77.
- Diakidoy, I.-A. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005) The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology* 26, 55-80.
- Dinaucourt, M. 15.10.2010]
- Doctor, E., & Coltheart, M. (1980) Phonological recoding in childrens reading for meaning. *Memory and Cognition* 80, 195-209.
- Dreyer, L. G., & Katz, L. (1992) An examination of "the simple view of reading". In C. K. Kinzer & D. J. Leu (Eds.), *Literacy Research, Theory and Practice: Views from Many Perspectives* (pp. 169-176). Chicago: National Reading Conference.

- Ehri, L. C. (1984) How orthography alters spoken language competencies in children learning to read and spell. In J. Downing & R. Valtin (Eds.), *Language awareness and learning to read* (pp. 119-147). New York: Springer.
- Ehri, L. C. (1992) Reconceptualizing the development of sight word reading and its relationship to recoding. In P. B. Gough & R. Treiman (Eds.), *Reading acquisition* (pp. 107-143). Hillsdale N.Y.: Erlbaum.
- Ehri, L. C. (1997) Sight word reading in normal readers and dyslexics. In B. Blachman (Ed.), *Foundations and Reading Acquisition and Dyslexia* (pp. 163-190). London: Lawrence Erlbaum Associates.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001) Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analyses. *Reading Research Quarterly* 36, 250-287.
- Fayol, M., & de Morais, J. J. (2004) *La lecture et son apprentissage*: Observatoire Nationale de Lecture.
- Fresch, M. J. (2008) Standing on the Shoulders of Giants. In M. J. Fresch (Ed.), *An Essential History of Current Reading Practices* (pp. 1-11): International Reading Association.
- Frith, U. (1985) Beneath the surface of developmental dyslexia. Are comparisons between developmental and acquired disorders meaningful? In K. E. Patterson, J. C. Marshall & M. Coltheart (Eds.), *Surface Dyslexia* (pp. 301-330). London: Lawrence Erlbaum Associates.
- Gathercole, S., & Baddley, A. D. (1993) *Working Memory And Language: Essays In Cognitive Psychology*. Hove East Sussex, New York: Taylor & Francis.
- Georgiou, G. K., Das, J. P., & Hayward, D. (2009) Revisiting the "Simple View of Reading" in a Group of Children with Poor Reading Comprehension. *Journal of Learning Disabilities* 42(1), 76-84.
- Glushko, R. J. (1979) The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology. Human Perception and Performance* 5(4), 674-691.
- Gogolin, I. (1994) *Der monolinguale Habitus der multilingualen Schule*. Münster / New York: Waxmann.
- Goldstein, H. (2003) *Multilevel statistical models* (3 ed.). London: Arnold.
- Goodman, K. S. (1974) Effective teachers of reading know language and children. *Elementary English* 51, 823-828.
- Goodman, K. S. (2005) *What's whole in whole language?* (20 ed.). Woolsey: RDR Books.
- Gough, P. B., & Hillinger, M. L. (1980) Learning to Read: an Unnatural Act. *Annals of Dyslexia* 30, 179-196.

- Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996) Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 1-13). Mahwah N.J.: Erlbaum.
- Gough, P. B., & Tunmer, W. (1986) Decoding, reading and reading disability. *Remedial and Special Education* 7, 6-10.
- Haas, W. (1998) Diglossie im historischen Wandel oder: Schweizerdeutsch auf dem Weg zur eigenen Sprache? In C. V. J. Russ (Ed.), *Sprache-Kultur-Nation. Papers from a conference held on 19 October 1996 at the Guildhall, York* (pp. 77-101): Hull.
- Hanushek, E. A., & Woessmann, L. (2010) *The High cost of Low Educational Performance. The long-run economic impact of improving PISA outcomes.*: OECD.
- Harm, M. W., & Seidenberg, M. S. (1999) Phonology, Reading Acquisition and Dyslexia: Insights From Connectionist Models. *Psychological Review* 106(3), 491-528.
- HarmoS-L1. (2007) *Schulsprache – Langue de scolarisation – Lingua di scolarizzazione. Kompetenzmodell – modèle de compétence – modello die competenza.* Bern: EDK/CDIP.
- Holmes, J. A. (1954) *The substrata-factor theory of reading.* Berkeley, CA: California Book.
- Hoover, W., A. , & Gough, P. B. (1990) The simple view of reading. *Reading and Writing* 2(2), 127-160.
- Hulme, C., Snowling, M. J., & Quinlan, P. (1991) Connectionism and learning to read: towards a psychologically plausible model. *Reading and Writing* 3, 159-168.
- International Labour Organization. (2009) ISCO. The International Standard Classification of Occupations
- Irwin, J. W. (1986) *Teaching reading comprehension processes.* New Jersey: Prentice-Hall.
- Javal, E. (1879) Essai sur la physiologie de la lecture. *Annales d'Oculistique* 82, 242-253.
- Johnston, T., C., & Kirby, J., R. (2006) The Contribution of Naming Speed to the Simple View of Reading. *Reading and Writing* 19(4), 339-361.
- Joshi, R. M., & Aaron, P. G. (2000) The Component model of Reading: Simple view of Reading made a little more complex *Reading Psychology* 21(2), 85-97.
- Judd, C. H., & Buswell, G. T. (1922) *Silent reading: A Study of the various types.* Chicago, IL: University of Chicago Press.
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008) Reading Comprehension Tests Vary in the Skills they Assess: Differential Dependence on Decoding and Oral Comprehension. *Scientific Studies of Reading* 12(3), 281-300.

- Keenan, J. M., Betjemann, R. S., Wadsworth, S. J., DeFries, J. C., & Olson, R. K. (2006) Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading* 29(1), 75-91.
- Kendeou, P., Savage, R., & van den Broek, P. (2009) Revisiting the simple view of reading. *British Journal of Educational Psychology* 79, 353-370.
- Kershaw, S., & Schatschneider, C. (2010) A latent variable approach to the simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 1-32.
- Kintsch, W. (1980) Learning from text, levels of comprehension, or why anyone would read a story any way. *Poetics*, 9 87-98
- Kintsch, W. (1988) The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review* 95(2), 163-182.
- Kintsch, W. (2004) The Construction-Integration Model of Text Comprehension and Its Implication for Instruction. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical Models and Processes of Reading* (pp. 1270-1328). Newark Del.: International Reading Association.
- Kintsch, W., & Van Dijk, T. A. (1978) Towards a model of text comprehension and production. *Psychological Review* 85, 363-394.
- Kirby, J. R. (2007) Reading Comprehension: Its Nature and Development., *Encyclopedia of Language and Literacy Development* (pp. 1-8). London: Canadian Language and Literacy Research Network.
- Kirby, J. R., & Savage, R. S. (2008) Can the simple view deal with the complexities of reading? *Literacy* 42, 75-82.
- Kirschhock, E. M. (2003) *Die Entwicklung schriftsprachlicher Kompetenzen im ersten Schuljahr*. Unpublished Inaugural-Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg.
- Koch, P. (1995) *Subordination, intégration syntaxique et oralité*. Paper presented at the Congrès "La subordination dans les langues romanes". , Copenhagen.
- Koch, P., & Oesterreicher, W. (1994) Schriftlichkeit und Sprache. In H. Günther & O. Ludwig (Eds.), *Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung. 1. Halbband*. (pp. 587-604). Berlin: de Gruyter.
- Kronig, W. (2003) *Flawed Selection - the Example of Immigrant Children in Swiss Education Systems*. Jyväskylä: University of Jyväskylä. Institute for Educational Research.
- Kronig, W., Haeblerlin, U., & Eckhard, M. (2000) *Immigrantenkinder und schulische Selektion. Pädagogische Visionen, theoretische Erklärungen und empirische Untersuchungen zur Wirkung integrierender und separierender Schulformen in den Grundschuljahren*. Bern / Stuttgart / Wien: Haupt.

- Kuhn, M. R., Schwanenflugel, P. J., Morris, R. D., Mandel Morrow, L., Gee Woo, D., Meisinger, E. B., et al. (2006) Teaching Children to Become Fluent and Automatic Readers. *Journal of Literacy Research* 38(4), 357-387.
- Kuhn, M. R., & Stahl, S. A. (2003) Fluency: A Review of Developmental and Remedial Practices. *Journal of Educational Psychology* 95(1), 3-21.
- LaBerge, D., & Samuels, S. J. (1974) Toward a theory of automatic information processing in reading. *Cognitive Psychology* 6(2), 293-323.
- Landerl, K., & Wimmer, H. (2000) Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics* 21, 243-262.
- Lenhard, W., & Schneider, W. (2006) ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler. Göttingen, Bern, Wien: Hogrefe.
- Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J.-E. (2004) Reading performance and its developmental trajectories during the first and the second grade. *Learning and Instruction* 14(2), 111-130.
- Linacre, J. M. (1991-2006) A Users Guide to Winsteps® (pp. 349).
- Lyster, S.-A. H. (2010) Reading comprehension: Unanswered questions and reading instruction challenges. *Advances in Learning and Behavioral Disabilities Literacy and Learning*(23), 115-153.
- Marin, B., & Legros, D. (2008) *Introduction à la psycholinguistique cognitive : lecture, compréhension et production de textes*. Paris: De Boeck.
- Marshall, J., & Newcombe, F. (1973) Patterns of paralexia: a psycholinguistic approach. *Journal of Psycholinguistic Research* 2, 175-199.
- Marx, H., & Jungmann, T. (2000) Abhängigkeit der Entwicklung des Leseverstehens von Hörverstehen und grundlegenden Lesefertigkeiten im Grundschulalter: Eine Prüfung des Simple View of Reading-Ansatzes. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 32(2), 81-93.
- Mayringer, H., & Wimmer, H. (2003) *SLS 1-4. Salzburger Lese-Screening für die Klassenstufen 1-4* Göttingen, Bern, Wien: Hogrefe.
- McClelland, G. H., & Judd, C. M. (1993) Statistical Difficulties of Detecting Interactions and Moderator Effects. *Psychological Bulletin* 114(2), 376-390.
- McKague, M., Pratt, C., & Johnston, M. B. (2001) The effect of vocabulary on reading visually novel words: a comparison of the dual-route-cascaded and triangle framework. *Cognition* 80, 231-262.
- Megherbi, H., Seigneuric, A., & Ehrlich, M.-F. (2006) Reading comprehension in French 1st and 2nd grade children: Contribution of decoding and language comprehension. *European Journal of Psychology of Education* XXI(2), 135-147.

- Minsky, M. (1975) A Framework for Representing Knowledge. In P. H. Winston (Ed.), *The Psychology of Computer Vision*. New York: McGraw-Hill.
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (2004) Does awareness of speech as a sequence of phones arise spontaneously. *Cognition* 7, 323-331.
- Moser, U. (2001) *Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen - Kurzfassung des nationalen Berichtes PISA 2000*: Bundesamt für Statistik und EDK.
- Müller, R. (1996) Ist das Schulversagen von zweisprachigen Migrantenkindern selbstverständlich? Oder: Mit schulischer Auslese und Zuweisung zu Kleinklassen wird ein Missstand aus den Augen, aber nicht aus der Welt geschafft. *Zeitschrift des Vereins Berner Logopäden VBL-Bulletin* 1, 3 - 18.
- Müller, R. (1997) *Sozialpsychologische Grundlagen des schulischen Zweitspracherwerbs bei MigrantenschülerInnen. Theoretische Grundlagen und empirische Studien bei zweisprachigen und einsprachigen SchülerInnen aus der 6. - 10. Klasse in der Schweiz*. (Vol. 21). Aarau / Frankfurt a. M. / Salzburg: Sauerländer.
- Müller, R. (1998) *Berufliches Wahlverhalten, berufliche Anspruchsniveaus und Bildungskarrieren von zweisprachigen Jugendlichen. Vorauswertung der Ergebnisse aus den Berufswahlraten von SchülerInnen der 8. - 10. Klasse. Unveröffentlichter Forschungsbericht des Autors*. Bern: Staatliches Seminar Hofwil.
- Müller, R. (2000) Über den Monolingualismus der Schweizer Schulen und seine Folgen für die mehrsprachigen SchülerInnen. *Babylonia. A Journal of Language Teaching and Learning* 1(8), 74-80.
- Müller, R. (2001) Die Situation der ausländischen Jugendlichen auf der Sekundarstufe II in der Schweizer Schule - Integration oder Benachteiligung. *Schweizerische Zeitschrift für Bildungswissenschaften* 23(2), 265-298.
- Nation, K., & Snowling, M. J. (1997) Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology* 67, 359-370.
- Nation, K., & Snowling, M. J. (1998) Individual differences in contextual facilitation: evidence from dyslexia and poor reading comprehension. *Child Development* 69, 996-1011.
- Nation, K., & Snowling, M. J. (1999) Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition* 70, 1-13.
- NOUVELOBS.COM. Un adieu définitif à la méthode globale. *NOUVELOBS.COM* (05.01.2006).
- NOUVELOBS.COM. (2006) Un adieu définitif à la méthode globale, *Nouvel Observateur*.

- Oakhill, J., & Yuill, N. (1996) Higher Order Factors in Comprehension Disability: Processes and Remediation. In C. Cornoldi & J. Oakhill (Eds.), *Reading Comprehension Difficulties: Processes and remediation*. Mahwah NJ: Lawrence Erlbaum Associates.
- Oakhill, J. V., Cain, K., & Bryant, P. E. (2003) The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes* 18(4), 443-468.
- Office fédéral des migrations. (2010) *Rapport sur la migration 2009*. Bern-Wabern: Office fédéral des migrations (ODM).
- Palmer, J., MacLeod, C. M., Hunt, E., & Davidson, J. E. (1985) Information processing correlates of reading. *Journal of Memory and Language* 24, 59-88.
- Paris, A. H., & Paris, S. G. (2003) Assessing narrative comprehension in young children. *Reading Research Quarterly* 38(1), 36-76.
- Paris, S. G. (2007) Assessment of Reading Comprehension *Encyclopedia of Language and Literacy Development* (pp. 1-8). London: Canadian Language and Literacy Research Network.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005) The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading. A handbook* (pp. 227-247). Malden MA: Blackwell Publishing.
- Perfetti, C. A., & Roth, S. F. (1978) Some of the interactive processes in reading and their role in reading skill. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 269-297). Hillsdale NJ: Erlbaum.
- Plaut, D. C., McClelland, J. L., Seidenberg, M., & Patterson, K. E. (1996) Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* 103, 56-115.
- Poldrack, R. A., & Wagner, A. D. (2004) What can neuroimaging tell us about mind? *Current Directions in Psychological Science* 13, 177-181.
- Protopapasa, A., Simosb, P. G., Sideridisb, G. D., & Mouzakic, A. (2012) The Components of the Simple View of Reading: A Confirmatory Factor Analysis.
- Rayner, K. (1998) Eye Movements in Reading and Information Processing. 20 Years of Research. *Psychological Bulletin* 124, 372-422.
- Rose, J. (2006) *Independent Review of the Teaching of Early Reading* (No. ISBN: 1-84478-684-6). Nottingham: Department for Education and Skills.
- Rosenblatt, L. M. (1994 ) The Transactional Theory of Reading and Writing. In R. B. Ruddell & e. al. (Eds.), *Theoretical Models and Processes of Reading* (4 ed., pp. 1057-1092). Newark Del.: International Reading Association (original publication 1978).
- Rost, D. H., & Hartmann, A. (1992) Lesen, Hören, Verstehen. *Zeitschrift für Psychologie* 200, 345-361.

- Rumelhart, D. E. (1980) Schemata: The Building Blocks of Cognition. In R. J. Spiro, B. C. Bruce & W. F. Brewer (Eds.), *Theoretical Issues in Reading Comprehension*. Hillsdale N.J.: Erlbaum.
- Rumelhart, D. E. (1985) Toward an interactive model of reading. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and the processes of reading* ( 3rd edition ed., pp. 960 pages). Newark, DE: International Reading Association
- Rupp, A., Ferne, T., & Choi, H. (2006) How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective *Language Testing* 23(4), 441-474.
- Sadoski, M., & Paivio, A. (2004) A Dual Coding Theoretical Model of Reading. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical Models and Processes of Reading* (pp. 1329-1362). Newark Del.: International Reading Association.
- Savage, R. (2001) The Simple View of Reading : some evidence and possible implications *Educational Psychological in Practice* 17(1), 17-33.
- Savage, R., & Wolforth, J. (2007) An Additive Simple View of Reading Describes the Performance of Good and Poor Readers in Higher Education. *Exceptionality Education Canada* 17(2), 243-268
- Scarborough, H. (2001 ) Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 1, pp. 97-110). New York: Guilford Press.
- Scheerer-Neumann, G. (1995) Wortspezifisch: ja - Wortbild: nein. Ein letztes Lebewohl and die Wortbildtheorie. In H. Balhorn & H. Brügelmann (Eds.), *Rätsel des Schriftspracherwerbs. Teil "Rechtschreiben"* (pp. 230-244). Konstanz: Libelle.
- Schleicher, A. (2010) Viel Ideologie, sehr wenig Wissen. *NZZ am Sonntag* 9/24.
- Seidenberg, M. S., & McClelland, J. L. (1989) A distributed developmental model of word recognition and naming. *Psychological Review* 96, 523-586.
- Seigneuric, A., & Ehrlich, M.-F. (2005) Contribution of Working Memory Capacity to Children's Reading Comprehension: A Longitudinal Investigation. *Reading and Writing* 18(7), 617-656.
- Seymour, P. H. K. (1994) Un modèle du développement orthographique à double fondation. In J. P. Jaffré, L. Sprenger-Charolles & M. Fayol (Eds.), *Lecture/Ecriture: Acquisition; Les actes de La Villette* (pp. 57-79). Paris: Nathan.
- Shakespeare, W. (1597) *Roméo et Juliette*. London.
- Siegel, L. S. (1989) IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities* 22(8), 469-478.
- Singer, J. D., & Willett, J. B. (2003) *Applied Longitudinal Data Analysis* New York: Oxford University Press.

- Smith, F. (1973) *Psychology and reading*. New York: Holt, Rinehart & Winston.
- Smith, F. (1986) *Understanding reading*. Hillsdale, NJ: Lawrence Erlbaum.
- Smith, F. (1994) *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read* (5 ed.). Hillsdale NJ: Erlbaum.
- Snowling, M. J., & Hulme, C. (Eds.). (2005) *The Science of Reading. A Handbook*: Blackwell Publishing.
- Stahl, S. A. (1999) Why innovations come and go (and mostly go): the case of whole language. *Educational Researcher* 28, 13-22.
- Stanovich, K. E. (1980) Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly* 16, 32-71.
- Stanovich, K. E., Cunningham, A. E., & Feeman, D. J. (1984) Intelligence, cognitive skills and early reading progress. *Reading Research Quarterly* XIX(3), 278-303.
- Stanovich, K. E., Nathan, R. G., & Vala-Rossi, M. (1986) Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis *Reading Research Quarterly* 21(3), 267-283.
- Steyer, R., & Eid, M. (2001) *Messen und Testen*. Heidelberg: Springer Verlag,
- .
- Stothard, S. E., & Hulme, C. (1992) Reading comprehension difficulties in children. *Reading and Writing* 4(3), 245-256.
- Stroop, R. J. (1935) Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18, 643-662.
- Sweet, A. P. (2005) Assessment of reading comprehension: The RAND reading study group vision. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 3-12). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Thorndike, E. L. (1917) Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology* 8(6), 323-332.
- Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009) Simple but complex: components of the simple view of reading across grade levels. *Journal of Research in Reading* (-).
- Tilstra, J., van den Broek, P., McMaster, K., Kendeou, P., & Rapp, D. (2006) The Contribution of Fluency, Vocabulary and Listening Comprehension to Reading Comprehension in 4th, 7th and 9th Grade Readers, *Annual Meeting of the Society for the Scientific Study of Reading* Vancouver.
- Tiu, R. D., Thompson, L. A., & Lewis, B. A. (2003) The Role of IQ in a Component Model of Reading. *Journal of Learning Disabilities* 36(5), 424-436.

- Treiman, R., Goswamin, U., & Bruck, M. (1990) Not all non-words are alike: Implications for reading development and theory. *Memory and Cognition* 18, 559-567.
- van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M. J., et al. (2005) Assessment of comprehension abilities in young children. In S. G. Paris & S. A. Stahl (Eds.), *Current issues in reading comprehension and assessment*. Mahwah, NJ: Erlbaum.
- Van Dijk, T. A., & Kintsch, W. (1983) *Strategies of discourse comprehension*. New York: Academic Press.
- van Kraayenoord, C. E., & Paris, S. G. (1996) Story construction from a picture book: An assessment activity for young learners. *Early Childhood Research Quarterly* 11, 41-61.
- Vandergrift, L. (2006) Second Language Listening: Listening Ability or Language Proficiency? . *The Modern Language Journal*, 90(1), 6-18(13).
- Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007) Components of Reading Ability: Multivariate Evidence for a Convergent Skills Model of Reading Development. *Scientific Studies of Reading* 11(1), 3-32.
- Verhoeven, L., & Leeuwe, J. v. (2008) Prediction of the development of reading comprehension: a longitudinal study. *Applied Cognitive Psychology* 22(3), 407-423.
- Vygotsky, L. S. (1986 ) *Thought and language (traduit par A. Kozalin)*. Cambridge M.A.: MIT Press (original publication 1934).
- Weiss, R., H. (2006) *CFT 20-R Grundintelligenztest Skala 2 - Revision*. Göttingen: Hogrefe.
- West, R. F., & Stanovich, K. E. (1978) Automatic contextual facilitation in readers of three ages. *Child Development* 49, 717-727.
- Wimmer, H. (1993) L'acquisition de la lecture dans une orthographe plus régulière que celle de l'anglais. In S. Brun-Bourguignon & L. Dercourt (Eds.), *Les actes de la Villette: Lecture-écriture: acquisition* (pp. 97-106). Paris: Editions Nathan.
- Wimmer, H., Hartl, M., & Moser, E. (1990) Passen "englische " Modelle des Schriftsprachenerwerbs auf "deutsche" Kinder? Zweifel an der Bedeutung der der logographischen Stufe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 22, 136-154.
- Wimmer, H., Mayringer, H., & Landerl, K. (1998) Poor reading: a deficit in skill automatization or a phonological deficit? *Scientific Studies of Reading* 2(4), 321-340.
- Wimmer, H., Mayringer, H., & Landerl, K. (2000) The Double-Deficit Hypothesis and Difficulties in Learning to Read a Regular Orthography. *Journal of Educational Psychology* 92(4), 668-680.
- Wolf, M. (2001) *Dyslexia, Fluency, and the Brain*: York Press.

- Yuill, N., & Oakhill, J. (1988) Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology* 2, 33-45.
- Yuill, N., & Oakhill, J. (1991) *Childrens problems in text comprehension*. Cambridge: Cambridge University Press.
- Zagar, D. (1992) L'approche cognitive de la lecture: de l'accès au lexique au calcul syntaxique. In M. Fayol, J. E. Gombert, P. Lecoqu, L. Sprenger-Charolles & D. Zagar (Eds.), *Psychologie cognitive de la lecture*. Paris: puf.
- Ziegler, J. C., & Goswami, U. (2005) Reading Acquisition, Developmental Dyslexia and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin* 131(1), 3-29.



## Annexe 2

Modèle 0 (Modèle vide) sans et avec les valeurs estimées (voir encadré pointillé)

$$Z\text{TextRoh\_2bis3\_01}_{zeit01, IDS\%, KLANRS} \sim N(\mathcal{X}B, \Omega)$$

$$Z\text{TextRoh\_2bis3\_01}_{zeit01, IDS\%, KLANRS} = \beta_{0IDS\%, KLANRS} \text{kons} + \beta_{1IDS\%, KLANRS} \text{LZ von 2. zu 3.Klasse}_{zeit01, IDS\%, KLANRS}$$

$$\beta_{0IDS\%, KLANRS} = \beta_0 + v_{0KLANRS} + u_{0IDS\%, KLANRS}$$

$$\beta_{1IDS\%, KLANRS} = \beta_1 + v_{1KLANRS} + u_{1IDS\%, KLANRS}$$

$$\begin{bmatrix} v_{0KLANRS} \\ v_{1KLANRS} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} \sigma_{v0}^2 & \\ & \sigma_{v1}^2 \end{bmatrix}$$

$$\begin{bmatrix} u_{0IDS\%, KLANRS} \\ u_{1IDS\%, KLANRS} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ & \sigma_{u1}^2 \end{bmatrix}$$

Les coefficients de régression  $\beta_0$  et  $\beta_1$  ainsi que leurs erreurs (variances résiduelles) au niveau des classes (v) et des individus (u)

Variances (résiduelles) et Covariances : au niveau des classes (v) et des individus (u)

$-2 * \text{loglikelihood(IGLS Deviance)} = 1277.362(580 \text{ of } 580 \text{ cases in use})$  Indice de la variance

$$Z\text{TextRoh\_2bis3\_01}_{zeit01, IDS\%, KLANRS} \sim N(\mathcal{X}B, \Omega)$$

$$Z\text{TextRoh\_2bis3\_01}_{zeit01, IDS\%, KLANRS} = \beta_{0IDS\%, KLANRS} \text{kons} + \beta_{1IDS\%, KLANRS} \text{LZ von 2. zu 3.Klasse}_{zeit01, IDS\%, KLANRS}$$

$$\beta_{0IDS\%, KLANRS} = -0.461(0.090) + v_{0KLANRS} + u_{0IDS\%, KLANRS}$$

$$\beta_{1IDS\%, KLANRS} = 0.841(0.062) + v_{1KLANRS} + u_{1IDS\%, KLANRS}$$

$$\begin{bmatrix} v_{0KLANRS} \\ v_{1KLANRS} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.105(0.047) & \\ 0.051(0.025) & 0.038(0.022) \end{bmatrix}$$

$$\begin{bmatrix} u_{0IDS\%, KLANRS} \\ u_{1IDS\%, KLANRS} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.546(0.046) & \\ -0.088(0.031) & 0.438(0.039) \end{bmatrix}$$

$-2 * \text{loglikelihood(IGLS Deviance)} = 1277.362(580 \text{ of } 580 \text{ cases in use})$

## Modèle 1.1

incluant les variables indépendantes: « Temps » (LZ von 2. Zu 3. Klasse) et « Décodage » (ZWortRoh\_2bis3\_01)

$$Z_{\text{TextRoh\_2bis3\_01}}^{\text{zeit01, IDSk, KLANRS}} \sim N(\chi\beta, \Omega)$$

$$Z_{\text{TextRoh\_2bis3\_01}}^{\text{zeit01, IDSk, KLANRS}} = \beta_{0\text{IDSk, KLANRS}}^{\text{kons}} + \beta_{1\text{IDSk, KLANRS}}^{\text{LZ von 2. zu 3. Klasse}} + \beta_2^{\text{ZWortRoh\_2bis3\_01}} + \beta_3^{\text{ZWortRoh\_2bis3\_01.LZ von 2. zu 3. Klasse}}$$

$$\beta_{0\text{IDSk, KLANRS}} = \beta_0 + v_{0\text{KLANRS}} + u_{0\text{IDSk, KLANRS}}$$

$$\beta_{1\text{IDSk, KLANRS}} = \beta_1 + v_{1\text{KLANRS}} + u_{1\text{IDSk, KLANRS}}$$

$$\begin{bmatrix} v_{0\text{KLANRS}} \\ v_{1\text{KLANRS}} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} \sigma_{v0}^2 & \\ & \sigma_{v1}^2 \end{bmatrix}$$

$$\begin{bmatrix} u_{0\text{IDSk, KLANRS}} \\ u_{1\text{IDSk, KLANRS}} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ & \sigma_{u1}^2 \end{bmatrix}$$

Interaction: Décodage et Temps.  
Elle montre l'influence du progrès en décodage entre CM1 et CM2 sur les compétences en compréhension de texte (épreuve ELFE)

Variances et covariances :  
au niveau des classes (v) et des individus (u)

$-2 * \log\text{likelihood (IGLS Deviance)} = 1109.238(580 \text{ of } 580 \text{ cases in use})$

ZTextRoh\_2bis3\_01= Y (la variable réponse = la compréhension de texte selon le test ELFE)

Zeit01= k le niveau du temps / IDSk = i le niveau des individus / KLANRS = j le niveau des classes

$\beta_{0\text{IDSk, KLANRS}}$  = le coefficient de régression indiquant où la droite de régression rencontre l'axe verticale ; il varie entre les individus et des classes = *intercept* (voir note en bas de page 76), = La performance (le nombre de points obtenus dans une tâche de compréhension de lecture spécifique) pour un individu qui aurait la valeur « 0 » ou la moyenne dans chaque variable indépendantes qui est introduite dans cette équation de régression. Dans ce modèle 1.1 le  $\beta_{0\text{IDSk, KLANRS}}$  est la performance en compréhension de texte d'un individu ayant  $z=0$  (compétence dans la moyenne) en tâche de décodage.

$\beta_{1\text{IDSk, KLANRS}}$  = le coefficient de régression de la variable LZ von 2. zu 3. Klasse (évolution des compétences en lecture entre CE1 et CE2), il varie entre les individus et des classes

$\beta_2$  ZWortRoh\_2bis3\_01 = coefficient de régression des compétences en Décodage

$\beta_3$  ZWortRoh\_2bis3\_01.LZ von 2.zu 3. Klasse = coefficient de régression de l'interaction entre les compétences de Décodage et de l'évolution dans le temps, nous indiquant le poids régression de l'évolution des compétences en Décodage

$\beta_4$  Z\_Hoeren\_2bis3\_01 = coefficient de régression des compétences en Compréhension Orale

### Modell 1.2

Avec les variables indépendantes: « Temps » (LZ von 2. Zu 3. Klasse) et « Décodage » (ZWortRoh\_2bis3\_01), et « Compréhension Orale » (Z\_Hoeren\_2bis3\_01)

$$Z\_Lesen\_2bis3\_01_{zeit01, IDSk, KLANRS} \sim N(\chi B, \Omega)$$

$$Z\_Lesen\_2bis3\_01_{zeit01, IDSk, KLANRS} = \beta_{0IDSk, KLANRS} \text{kons} + \beta_{1IDSk, KLANRS} LZ \text{ von 2. zu 3. Klasse}_{zeit01, IDSk, KLANRS} + \beta_2 ZWortRoh\_2bis3\_01_{zeit01, IDSk, KLANRS} + \beta_3 ZWortRoh\_2bis3\_01.LZ \text{ von 2. zu 3. Klasse}_{zeit01, IDSk, KLANRS} + \beta_4 Z\_Hoeren\_2bis3\_01_{zeit01, IDSk, KLANRS} + \beta_5 Z\_Hoeren\_2bis3\_01.LZ \text{ von 2. zu 3. Klasse}_{zeit01, IDSk, KLANRS}$$

$$\beta_{0IDSk, KLANRS} = \beta_0 + v_{0KLANRS} + u_{0IDSk, KLANRS}$$

$$\beta_{1IDSk, KLANRS} = \beta_1 + v_{1KLANRS} + u_{1IDSk, KLANRS}$$

$$\begin{bmatrix} v_{0KLANRS} \\ v_{1KLANRS} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} \sigma_{v0}^2 & \\ & \sigma_{v01} \sigma_{v1}^2 \end{bmatrix}$$

$$\begin{bmatrix} u_{0IDSk, KLANRS} \\ u_{1IDSk, KLANRS} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ & \sigma_{u01} \sigma_{u1}^2 \end{bmatrix}$$

Interaction: Compréhension Orale et Temps. L'interaction montre l'influence du progrès en compréhension orale entre CM1 et CM2 sur la compréhension de texte (ELFE)

Variances et covariances : au niveau des classes (v) et des individus (u)

$-2 * \loglikelihood(IGLS \text{ Deviance}) = 1147.904(570 \text{ of } 580 \text{ cases in use})$

$\beta_5 Z\_Hoeren\_2bis3\_01.LZ \text{ von 2.zu 3. Klasse}$  = coefficient de régression d'interaction, nous indiquant le poids de régression de l'évolution des compétences en compréhension orale de CE1 à CE2

$\sigma_{v0}^2$  = Variance résiduelle de  $\beta_0$  (= intercept = coefficient de régression de la constante) au niveau des classes

$\sigma_{v01}$  = Covariance de  $\beta_0$  et  $\beta_1$  au niveau des classes

$\sigma_{v1}^2$  = Variance résiduelle de  $\beta_1$  (coefficient de régression du temps « LZ von 2.zu 3. Klasse » évolution de CE1 à CE2) au niveau des classes

$\sigma_{u0}^2$  = Variance résiduelle de  $\beta_0$  (= intercept = coefficient de régression de la constante) au niveau des individus

$\sigma_{u01}$  = Covariance de  $\beta_0$  et  $\beta_1$  au niveau des individus

$\sigma_{u1}^2$  = Variance résiduelle de  $\beta_1$  (coefficient de régression du temps « LZ von 2.zu 3. Klasse » l'évolution de CE1 à CE2) au niveau des individus

## Annexe 3

## Modèle statistique d'un ordre inversé pour le groupe des élèves en CE1/CE2 : test SLS

Variables réponses	SLS :		SLS :		SLS :	
	Modèle 0 (vide)		Modèle 1.1 Inversé		Modèle 1.2 Inversé	
Variables explicatives (Parties fixes)	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.541	0.125	-0.502	0.115	-0.091	0.065
Temps (compétences t1-t2)	0.942	0.067	0.87	0.064	0.208	0.06
Compréhension Orale (CO)			0.15	0.036	0.071	0.032
Interaction CO & Temps			0.004	0.036	0.013	0.039
Décodage (D)					0.766	0.041
Interaction D & Temps					-0.162	0.044
Variances (Parties aléatoires)	(0.232-0.19)/0.023218.10%					
Niveau III: Classes						
$\sigma_{v0}^2$ = Variance résiduelle $\beta_0$ (Constante)	0.232	0.091	0.19	0.076	0.049	0.022
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.026	0.035	-0.023	0.03	-0.026	0.015
$\sigma_{v1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	0.058	0.025	0.048	0.022	0.028	0.016
Niveau II: Individus	(0.556-0.507)/0.556=8.8%					
$\sigma_{u0}^2$ = Variance résiduelle $\beta_0$ (Constante)	0.556	0.047	0.507	0.043	0.218	0.019
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.073	0.022	-0.076	0.022	-0.105	0.016
$\sigma_{u1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	0.203	0.019	0.214	0.02	0.251	0.023
-2*loglikelihood:	1085.499		1064.154		774.828	
Test de Chi2						
N° d'Unités Niv. III (Classes)	17		17		17	
N° d'Unités Niv. II (Individus)	318		318		318	
N° d'Unités Niv. I (Temps)	571		571		571	

	Modèle 0 → 1.1 Inversé	Modèle 0 → 1.2 Inversé	Modèle 1.1 → 1.2 Inversé
Niveau III: Classes			
$\sigma_{v0}^2$ = Variance résiduelle $\beta_0$ (Constante)	18.10%	78.88%	74.21%
$\sigma_{v1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	17.24%	51.72%	41.67%
Niveau II: Individus			
$\sigma_{u0}^2$ = Variance résiduelle $\beta_0$ (Constante)	8.81%	60.79%	57.00%
$\sigma_{u1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	-5.42%	-23.65%	-17.29%

*Modèle statistique d'un ordre inversé pour le groupe des élèves en CE1/CE2 : test ELFE*

Variables réponses	ELFE:		ELFE:		ELFE:	
	Modèle 0 (vide)		Modèle 1.1 Inversé		Modèle 1.2 Inversé	
Variables explicatives (Parties fixes)	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.461	0.09	-0.391	0.07	-0.112	0.053
Temps (compétences t1-t2)	0.841	0.062	0.703	0.056	0.21	0.059
Compréhension Orale (CO)			0.28	0.039	0.187	0.037
Interaction CO & Temps			0.03	0.046	0.034	0.05
Décodage (D)					0.535	0.044
Interaction D & Temps					-0.032	0.051
Variances (Parties aléatoires)						
Niveau III: Classes						
$Z_{v0}^2$ = Variance résiduelle $\beta_0$ (Constante)	0.105	0.047	0.054	0.028	0.022	0.014
$Z_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	0.051	0.025	0.038	0.017	0.027	0.01
$Z_{v1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	0.038	0.022	0.02	0.017	0.008	0.013
Niveau II: Individus						
$Z_{u0}^2$ = Variance résiduelle $\beta_0$ (Constante)	0.546	0.046	0.459	0.039	0.31	0.027
$Z_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.088	0.031	-0.11	0.03	-0.157	0.026
$Z_{u1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	0.438	0.039	0.465	0.041	0.492	0.044
-2*loglikelihood:	1277.362		1220.572		1072.657	
Test de Chi <sup>2</sup>						
N° d'Unités Niv. III (Classes)	17		17		17	
N° d'Unités Niv. II (Individus)	318		318		318	
N° d'Unités Niv. I (Temps)	580		580		580	

	Modèle 0 → 1.1 Inversé	Modèle 0 → 1.2 Inversé	Modèle 1.1 → 1.2 Inversé
Niveau III: Classes			
$\sigma_{v0}^2$ = Variance résiduelle $\beta_0$ (Constante)	48.57%	79.05%	59.26%
$\sigma_{v1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	47.37%	78.95%	60.00%
Niveau II: Individus			
$\sigma_{u0}^2$ = Variance résiduelle $\beta_0$ (Constante)	15.93%	43.22%	32.46%
$\sigma_{u1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	-6.16%	-12.33%	-5.81%

*Modèle statistique d'ordre inversé pour le groupe des élèves en CE1/CE2: test HarmoS*

Variables réponses	HarmoS		HarmoS		HarmoS	
	Modèle 0 (vide)		Modèle 1.1 Inversé		Modèle 1.2 Inversé	
<b>Variables explicatives (Parties fixes)</b>	$\beta$	SE	$\beta$	SE	$\beta$	SE
Constante	-0.451	0.157	-0.349	0.129	-0.214	0.115
Temps (compétences t1-t2)	0.843	0.113	0.649	0.111	0.377	0.115
Compréhension Orale (CO)			0.385	0.045	0.339	0.047
Interaction CO & Temps			0.007	0.056	-0.03	0.06
Décodage (D)					0.255	0.059
Interaction D & Temps					0.082	0.069
<b>Variances (Parties aléatoires)</b>						
Niveau III: Classes						
$\sigma_{v0}^2$ = Variance résiduelle $\beta_0$ (Constante)	0.374	0.142	0.245	0.095	0.176	0.071
$\sigma_{v01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.133	0.081	-0.104	0.064	-0.114	0.059
$\sigma_{v1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	0.166	0.072	0.155	0.068	0.147	0.066
Niveau II: Individus						
$\sigma_{u0}^2$ = Variance résiduelle $\beta_0$ (Constante)	0.532	0.046	0.417	0.036	0.395	0.034
$\sigma_{u01}$ = Covariance de $\beta_0$ et $\beta_1$	-0.244	0.039	-0.264	0.036	-0.304	0.037
$\sigma_{u1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	0.59	0.054	0.604	0.055	0.628	0.057
-2*Loglikelihood:	1306.97		1203.48		1147.9	
N° d'Unités Niv. III (Classes)	17		17		17	
N° d'Unités Niv. II (Individus)	317		317		317	
N° d'Unités Niv. I (Temps)	570		570		570	

	Modèle 0 $\rightarrow$ 1.1 Inversé	Modèle 0 $\rightarrow$ 1.2 Inversé	Modèle 1.1 $\rightarrow$ 1.2 Inversé
Niveau III: Classes			
$\sigma_{v0}^2$ = Variance résiduelle $\beta_0$ (Constante)	34.49%	52.94%	28.16%
$\sigma_{v1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	6.63%	11.45%	5.16%
Niveau II: Individus			
$\sigma_{u0}^2$ = Variance résiduelle $\beta_0$ (Constante)	21.62%	25.75%	5.28%
$\sigma_{u1}^2$ = Variance résiduelle de $\beta_1$ (Temps)	-2.37%	-6.44%	-3.97%