



Note d'

Numéro 36  
Septembre 2009

# information

du SRED

Service de la recherche en éducation  
Département de l'instruction publique - Canton de Genève

Résultats aux épreuves cantonales ou communes de français  
et aux tests diagnostiques des difficultés de lecture

Anne Soussi et Christian Nidegger

Dans le cadre du plan lecture initié à la suite des résultats plutôt médiocres des élèves à l'enquête PISA 2000 en lecture (littératie) et de la priorité du Département de l'instruction publique donnée à la maîtrise de la langue française, les trois directions de l'enseignement ont confié à une équipe de chercheurs de l'Université de Rennes 2, sous la responsabilité d'E. Bonjour, la tâche d'élaborer un outil de diagnostic des difficultés de lecture chez les élèves à trois moments du cursus (fin de la 2P ou début de la 3P; fin de la 6P ou début de la 7<sup>e</sup>; fin de la 9<sup>e</sup> ou début du post-obligatoire).

**A**u départ, ces tests étaient destinés à cerner plus précisément le type de difficulté qu'ont les élèves présentant des lacunes, en vue d'organiser et de proposer aux élèves des activités de remédiation ciblées. Il s'agissait également de voir si les élèves qui n'avaient pas réussi les épreuves cantonales ou communes de français, et de lecture en particulier, auraient de bonnes performances au test diagnostique, ou si à l'inverse ceux qui avaient atteint le seuil de réussite de ces épreuves basées sur le programme genevois pourraient avoir des difficultés qui n'auraient pas été détectées. En effet, la lecture étant présente tout au long du cursus scolaire et dans toutes les disciplines, des lacunes dans ce domaine peuvent de ce fait avoir des répercussions sur la suite des études de ces élèves. C'est pourquoi il est important de savoir précisément où elles se situent afin de pouvoir y remédier.

Au printemps 2007, les instruments ont été prétestés dans des classes aux trois moments-charnières du cursus (fin de 2P, fin de 6P et fin de 9<sup>e</sup>) auprès d'une centaine d'élèves pour chaque degré concerné. A la rentrée 2007, ces tests ont été passés de manière expérimentale mais à un plus grand échantillon d'élèves en 3P, en 7<sup>e</sup> et au secondaire II chez les élèves des dispositifs d'insertion scolaire et professionnelle<sup>1</sup>.

Le SRED a accompagné le processus, analysé les résultats de ces tests diagnostiques et les a mis en relation avec ceux des épreuves cantonales et com-

munes. La présente note porte sur cette phase expérimentale et présentera les résultats des élèves aux tests diagnostiques en les comparant également à ceux obtenus aux épreuves "institutionnelles" que passent tous les élèves d'une volée.

Tous les élèves sont soumis chaque année à des épreuves cantonales ou communes en français notamment en 2P, 4P, 6P ainsi qu'en 7<sup>e</sup>, 8<sup>e</sup> et 9<sup>e</sup>. Ces épreuves externes ont pour objectifs d'attester les acquis des élèves en fin d'année ou de cycle et de réguler l'enseignement. Elles sont basées sur les plans d'études des deux niveaux d'enseignement. Pour les dispositifs d'insertion scolaire et professionnelle ("10<sup>e</sup> degré"), il n'y avait pas de plan d'études à proprement parler au moment de nos analyses, étant donné qu'il s'agit d'un degré supplémentaire destiné à des élèves qui n'ont pas atteint les objectifs prévus à la fin du cycle d'orientation.

La batterie de tests élaborée par l'équipe française de l'Université de Rennes 2 ne porte pas sur les objectifs d'un curriculum scolaire mais sur des compétences ou sous-compétences de lecture. Elle a pour but de repérer de manière plus spécifique les difficultés des élèves dans ce domaine. Elle reprend le modèle utilisé auprès des jeunes de 18 ans en France dans le cadre des Journées d'appel à la défense où l'on cherche à repérer les jeunes ayant de grandes difficultés en lecture parmi lesquels certains sont proches de l'illettrisme. Tous les jeunes Français de 18 ans sont d'abord testés au moyen d'une évaluation nationale por-

tant sur les compétences supposées acquises en fin de scolarité. Seuls certains élèves dont les compétences sont estimées faibles, voire insuffisantes dans le domaine de la lecture seront soumis à une batterie de tests plus spécifiques portant sur des compétences ou micro-compétences liées à la lecture. A Genève, les directions générales des trois ordres d'enseignement (primaire, secondaire I et secondaire II) ont donc confié à l'équipe de Rennes le développement d'une batterie de tests du même type (mais adaptée à l'âge des élèves testés) pour détecter les élèves genevois qui rencontrent des difficultés de lecture. Le diagnostic établi grâce à cette batterie de tests permettra ensuite d'orienter la remédiation nécessaire pour améliorer les compétences en lecture des élèves faibles.

Les deux types d'évaluation externe dont nous parlerons ici se situent davantage du côté de la logique de l'apprenant selon la catégorisation de Weiss (2002) même si l'évaluation institutionnelle peut parfois également servir à renseigner le système et contribuer à son pilotage. Toutefois, ces deux évaluations poursuivent des buts différents. Dans un cas, les évaluations cantonales ou communes institutionnelles servent à vérifier l'atteinte des objectifs définis dans le plan d'études et s'adressent à tous les élèves (fonction plutôt certificative) tandis que dans le second cas, il s'agit d'une évaluation destinée à un petit groupe d'élèves présentant des lacunes en lecture et visant le diagnostic de ces difficultés. La présente note a pour objectif de montrer les liens entre les résultats à ces deux types d'évaluation (institutionnelle et tests diagnostiques).

## Description des épreuves

Le domaine investigué se rapporte à la lecture et à certains sous-domaines du français. Il faut souligner que la compétence de lecture est complexe à construire car reposant sur la maîtrise de nombreuses composantes dont la description peut varier selon les conceptions théoriques. Ainsi, pour Giasson (1992), auteur très connu dans le monde francophone de l'éducation, la lecture d'un texte met en relation trois éléments: le *lecteur* avec ses structures et ses processus, le *texte* composé de l'intention de l'auteur, la forme et le contenu, ainsi que le *contexte* psychologique, social et physique. Depuis de nombreuses années, il est admis par les spécialistes que dès le début de l'apprentissage de la lecture, il est nécessaire de combiner les activités sur le code et sur le sens, c'est-à-dire à

### Encadré 1. Les épreuves cantonales ou communes de français et les tests diagnostiques

Au primaire, les épreuves cantonales de français sont composées de quatre parties se référant aux sous-domaines du français: compréhension de l'écrit, compréhension de l'oral, production écrite et structuration. Les trois premiers sont eux-mêmes subdivisés en quatre composantes: contexte, contenu, textualisation, planification, la structuration se composant de l'orthographe, du vocabulaire, de la conjugaison et de la grammaire.

Les épreuves de français du secondaire I, quant à elles, sont organisées tout à fait différemment: elles comportent trois parties, la première portant sur la compréhension d'un texte informatif et la production d'une courte lettre, la deuxième sur la compréhension d'un texte littéraire et la troisième sur la production d'un texte argumentatif. Dans chacune de ces parties, les questions peuvent relever de trois composantes différentes: le contenu, les moyens langagiers et la langue. L'aspect technique est donc complètement intégré à la compréhension ou à la production.

La batterie de tests élaborée par l'équipe de Rennes (Bonjour et al., 2007) comporte entre 12 et 15 exercices, selon le degré concerné (respectivement 13 exercices pour le test de 3P, 15 pour celui de 7<sup>e</sup> et 12 pour celui du "10<sup>e</sup> degré"), qui relèvent globalement des différents domaines proposés par Gombert. On peut en différencier quatre: les processus dits de bas niveau (reconnaissance des mots), les connaissances lexicales, les connaissances syntaxiques (compréhension des phrases et cohérence textuelle) et la compréhension (et les stratégies de lecture). Ces quatre dimensions composent la compétence générale de lecture. C'est en fonction de ces quatre dimensions ou composantes que seront présentés les résultats.

Les deux instruments sont donc complémentaires, l'un se référant au programme scolaire genevois, l'autre à des compétences plus globales et plus éloignées d'un programme scolaire.

la fois la capacité à déchiffrer à partir de mots et d'éléments iconographiques, de phrases et de connexions entre phrases (décodage), et également la capacité à construire le sens ou à comprendre des textes de genres variés. Pour Goigoux (2004), quatre éléments doivent être développés dans les premiers apprentissages: l'acculturation ou familiarisation avec la culture écrite, la production de texte, la compréhension de texte ainsi que l'identification et la production de mots (connaissances graphiques et phonologiques). Pour les concepteurs de la batterie de tests diagnostiques et notamment Gombert, la lecture suppose à la fois la reconnaissance de mots et le traitement des phrases et des textes. De nombreuses capacités doivent être mises en œuvre en parallèle: la perception, la mémoire, les catégorisations, les traitements linguistiques et la compréhension. Sur la base des différentes capacités nécessaires, trois grandes catégories peuvent être dégagées: ce qui relève des connaissances lexicales, ce qui concerne l'automatisme de la lecture et enfin, les traitements complexes.

Au niveau des plans d'études genevois des deux niveaux d'enseignement considérés (primaire et secondaire I), il faut relever des différences de centration, liées notamment au fait qu'ils ont été rédigés par des personnes provenant de courants différents. Le plan d'études genevois pour le primaire (*Les objectifs d'apprentissage à l'école primaire*) repose quant à lui sur d'autres modèles (que ceux mentionnés précédemment),

basés d'une part sur les liens entre lecture et écriture (compréhension et production) et d'autre part sur les textes (plus particulièrement sur la théorie des genres et des actions langagières). Les aspects de codes et de structuration sont au service de la compréhension et de la production. Au début de l'apprentissage, il y a, bien sûr, une centration sur le code écrit mais il est également précisé qu'il doit y avoir un va-et-vient quotidien entre activité spécifique de décodage et compréhension. Au secondaire I, le modèle est un peu différent: on parle de savoir lire et interpréter des textes. Cette compétence complexe suppose la mise en œuvre de capacités variées, socioculturelles, cognitives, linguistiques et sémiotiques ainsi qu'une capacité à s'adapter aux enjeux multiples des lectures. La référence est davantage aux types de texte et aux genres littéraires. Ces conceptions variées ont forcément un effet sur la construction des différents tests ou épreuves externes.

### Échantillon

Au cours de l'année 2007-2008, il a été décidé de passer le test diagnostique à plus large échelle, avec des démarches un peu différentes au niveau de la sélection des populations concernées selon les ordres d'enseignement:

- à l'école primaire dans les classes du REP en 3P (ce qui représente 284 élèves en tout). Il faut rappeler que les écoles du REP<sup>2</sup> comportent une proportion plus élevée que les autres d'élèves dont les pa-

rents sont ouvriers ou "divers/sans indication" et d'élèves allophones, ce qui ne signifie pas pour autant que tous ces élèves présentent des difficultés en lecture;

- au CO, trois établissements contrastés (Aubépine, Coudriers et Florence) ont été sélectionnés et la passation a eu lieu dans toutes les classes de 7<sup>e</sup> de ces collèges (c'est-à-dire auprès de 610 élèves répartis dans 32 classes), étant donné que l'on a constaté que si les élèves n'ayant pas atteint le seuil de réussite à l'épreuve cantonale de français en fin de 6P se trouvaient pour une proportion importante en B, on en trouvait également un certain nombre en A. Par ailleurs, l'enquête PISA avait également mis en évidence la présence d'élèves en difficulté en littérature dans toutes les filières. C'est pourquoi la direction du CO a décidé de passer le test non seulement en B, comme cela avait été initialement décidé, mais également en A et en H;

- au PO, dans toutes les classes des dispositifs d'insertion scolaire et professionnelle, c'est-à-dire des classes de complé-

ments de formation des écoles de culture générale et de commerce ainsi que des classes du SCAI, notamment<sup>3</sup>. Ces classes regroupent des élèves qui n'ont pas obtenu les résultats pour l'admission dans les filières certifiantes du PO et donc a priori des élèves présentant certaines difficultés scolaires, mais pas forcément dans le domaine de la lecture (tableaux 1-2-3). Nous allons maintenant présenter les résultats par degré.

### Résultats aux tests en 3P

Ce sont donc un peu moins de 300 élèves provenant des sept écoles entrées dans le REP en 2006-2007 ou en 2007-2008 qui ont été soumis au test diagnostique des difficultés de lecture. Nous présenterons les résultats généraux dans ces quatre dimensions et essayerons de mettre en évidence des profils de compétence puis montrerons les liens entre les deux types de tests (l'épreuve cantonale de 2P et le test diagnostique de 3P). Enfin, nous tenterons de dégager ce qui prédit le mieux les résultats à l'épreuve cantonale entre

les caractéristiques individuelles des élèves (genre, origine sociale, âge, première langue parlée) et les résultats aux quatre dimensions (graphique 1).

De manière générale, on observe des différences de résultats significatives statistiquement entre écoles pour deux des quatre composantes: le lexique et la syntaxe. La réussite est variable d'un domaine à l'autre. C'est en compréhension que les rendements sont les moins élevés, ce qui n'est pas étonnant étant donné que cette composante fait appel à des compétences plus complexes que les trois autres et en particulier, la reconnaissance des mots.

Si l'on regarde les profils de réussite, on constate que les deux tiers des élèves ont atteint le "seuil de réussite" fixé sur la base de la moyenne et d'un écart-type<sup>4</sup>. Quelques rares élèves n'ont atteint le seuil de réussite pour aucune composante. Près de 20% ont échoué à l'une ou l'autre. Il peut s'agir aussi bien d'un domaine que d'un autre, variable selon les élèves ou les écoles (graphique 2).

Tableau 1. Échantillon d'élèves de 3P

Écoles	Châtelaine		Cité-Jonction		Libellules			Meyrin-Bellavista		Onex-Tattes		Palettes				Pâquis-Centre		Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Nombre d'élèves	10	10	20	18	20	19	7	17	15	17	16	20	20	21	20	18	16	284

Tableau 2. Échantillon d'élèves de 7<sup>e</sup>

Collèges	Regroupement			Total
	Regroupement A	Regroupement B	Classe hétérogène	
Aubépine*	141 70.9%	58 29.1%	-	199 100%
Coudriers**	-	-	206 100%	206 100%
Florence*	181 88.3%	24 11.7%	-	205 100%
<b>Total</b>	<b>322</b> 52.8%	<b>82</b> 13.4%	<b>206</b> 33.8%	<b>610</b> 100%

\* Établissements à regroupements différenciés (A et B) - \*\* Établissement à niveaux et options (classes hétérogènes).

Tableau 3. Répartition des élèves par filière de formation au "10<sup>e</sup> degré"

Type d'école et de filière	N	%
CF - Écoles de commerce (18 classes)	241	40.6
CF - ECG (6 classes)	91	15.3
SCAI CEFP (8 classes)	79	13.3
SCAI CIPA (4 classes)	30	5.1
SCAI - Préapprentissage (13 classes)	153	25.8
<b>Total</b>	<b>594</b>	<b>100%</b>

**Tableau 4. Liens entre les différentes composantes et l'épreuve cantonale de français\***

	Composantes du test diagnostique				EC
	Reconnaissance des mots	Lexique	Syntaxe	Compréhension	Français 1
Reconnaissance des mots	1	.443**	.576**	.484**	.571**
Lexique	.443**	1	.516**	.414**	.450**
Syntaxe	.576**	.516**	1	.482**	.615**
Compréhension	.484**	.414**	.482**	1	.472**

\* Il s'agit de l'épreuve de français I de 2007 qui comportait essentiellement de la compréhension de l'écrit. Précisons que les résultats portent sur 94% environ des élèves ayant passé le test diagnostique en 3P. Parmi les 6% restants, il peut y avoir des élèves arrivés récemment qui n'avaient pas passé l'épreuve en 2P ne maîtrisant pas suffisamment le français, ou d'élèves venant d'arriver ou encore d'élèves ayant redoublé la 3P.

\*\* La corrélation est significative à un seuil de 1%.

Avant de présenter les liens entre ces résultats et ceux obtenus à l'épreuve cantonale de 2P qui porte sur la compréhension d'un texte, nous nous intéressons aux corrélations entre les différents domaines proposés (**tableau 4**).

On peut constater que toutes les corrélations<sup>5</sup> sont significatives. Toutefois, certains domaines sont plus liés entre eux que d'autres: les corrélations sont plus élevées entre reconnaissance des mots et syntaxe, entre syntaxe et lexique, entre français I et reconnaissance des mots ou encore entre français I et syntaxe. Il est intéressant de relever que la corrélation entre la dimension compréhension et français I qui comprend une part importante de compréhension de l'écrit n'est

pas la plus élevée (0.47). En d'autres termes, la réussite (ou l'échec) à la compréhension de l'écrit mesurée dans le test diagnostique n'irait pas forcément de pair avec la réussite (ou l'échec) en français I.

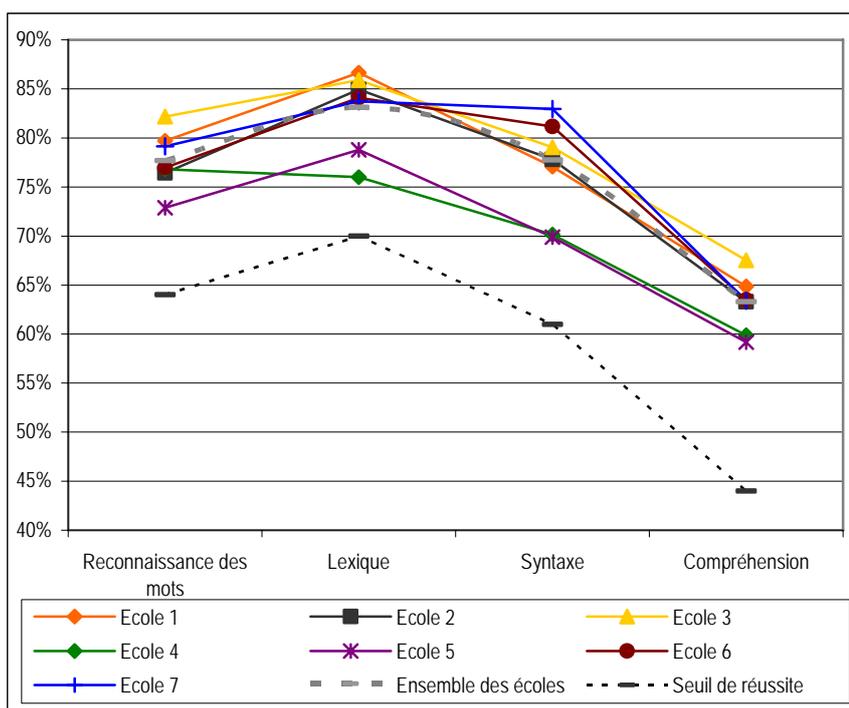
Si l'on observe les résultats des élèves au test diagnostique en fonction de la réussite à l'épreuve cantonale, on s'aperçoit que globalement, les élèves qui avaient atteint le seuil de réussite ont également bien réussi dans les quatre composantes et qu'à l'autre extrême, ceux qui avaient échoué ne réussissent pas non plus le test diagnostique, comme l'illustre le **graphique 3**. Ainsi, les élèves n'ayant pas du tout réussi le test diagnostique avaient également échoué à l'épreuve de 2P. La correspondance est

un peu moins bonne chez ceux qui ont réussi la totalité du test: 90% avaient atteint le seuil de réussite à l'épreuve cantonale mais 10% ne l'avaient pas atteint.

Si l'on considère les élèves qui avaient réussi l'épreuve cantonale, il ressort que près de 80% d'entre eux réussissent parfaitement le test diagnostique et 14% la quasi-totalité (trois composantes sur quatre). Parmi ceux qui n'avaient pas réussi l'épreuve, la situation est plus hétérogène puisque près de 30% ont réussi le test diagnostique et 31% trois composantes sur quatre. En d'autres termes, les deux instruments sont complémentaires et mettent en évidence d'autres éléments.

Pour finir avec l'analyse des résultats en 3P, nous avons effectué une analyse multivariée (de régression multiple)<sup>6</sup> prenant en compte simultanément les caractéristiques des élèves ainsi que les résultats aux différentes composantes du test diagnostique. Il faut préciser que les classes analysées font partie du REP et comportent donc des similitudes de compositions (même si les écarts peuvent être conséquents par rapport à la proportion d'enfants de milieux défavorisés [ouvriers et divers/sans indication] et d'allophones). Les différentes variables (caractéristiques individuelles et rendement aux quatre composantes du test diagnostique) décrivent 46% de variance, c'est-à-dire une part relativement importante des différences de résultats entre élèves. Parmi les diverses caractéristiques qui expliquent le score obtenu à l'épreuve de français et en particulier en compréhension de l'écrit, ce sont surtout les résultats obtenus aux quatre composantes qui ont des effets sur les résultats davantage que les caractéristiques individuelles. D'autres études ont montré que les scores initiaux avaient un plus grand effet sur les scores finaux que les caractéristiques individuelles.

**Graphique 1. Résultats pour les quatre composantes dans les sept écoles**



téristiques individuelles des élèves (voir notamment Bressoux, 1994, 1995; Duru-Bellat et al., 2004). On peut faire l'hypothèse que les caractéristiques individuelles des élèves (et en particulier la proportion d'enfants d'ouvriers et divers/sans indication et celle d'allophones<sup>7</sup>) sont ici plus homogènes que dans d'autres contextes scolaires. Une analyse complémentaire, l'analyse de régression logistique<sup>8</sup> qui permet d'observer ce qui détermine la réussite ou l'échec à l'épreuve cantonale, permet de confirmer ces résultats: la réussite à trois composantes (reconnaissance des mots, syntaxe et compréhension) semble être liée à celle de l'épreuve cantonale (et dans une moindre mesure au lexique).

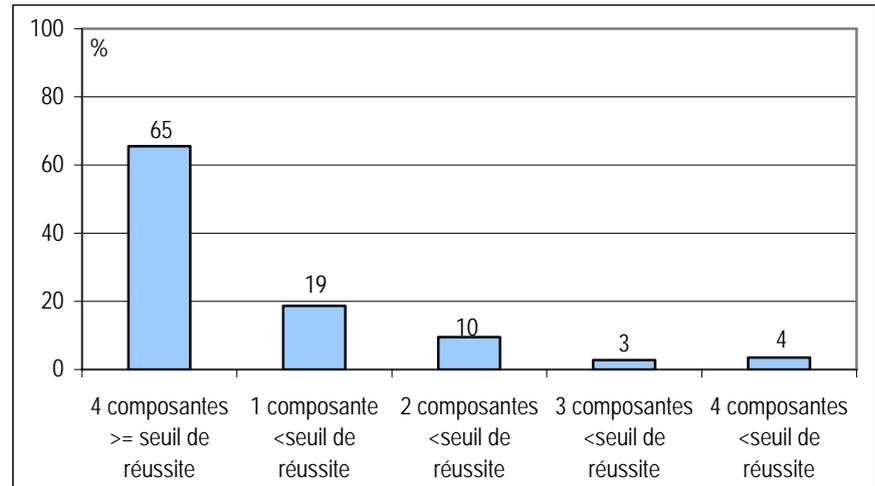
Quand on regarde l'effet des caractéristiques individuelles sur chacune des quatre dimensions, on s'aperçoit d'une part que la part de variance expliquée est nettement plus faible (c'est-à-dire qu'il faut chercher dans d'autres dimensions l'explication des différences de résultats entre élèves) et que ces différentes caractéristiques (le fait d'être une fille ou un garçon, d'avoir comme première langue parlée le français ou non, de provenir d'un milieu plutôt défavorisé ou favorisé, etc.) semblent avoir assez peu d'effets sur les résultats dans ce contexte d'élèves issus d'écoles du REP (et qui présentent donc des caractéristiques relativement semblables).

### Résultats aux tests en 7<sup>e</sup>

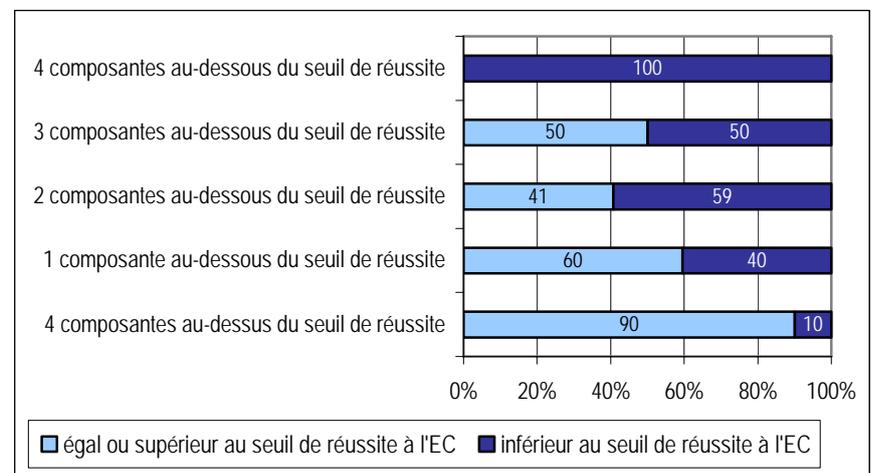
Une population d'environ 600 élèves ont passé le test en 7<sup>e</sup> année parmi lesquels 53% proviennent du regroupement A, 13% de B et 34% de classes hétérogènes (**graphique 4**). Relevons qu'en 7<sup>e</sup> le test a été soumis à une population d'élèves non sélectionnés (par opposition à celui de 3P passé dans des classes d'écoles du REP dont les élèves sont plus nombreux à provenir de milieux défavorisés et à avoir comme première langue parlée une autre langue que le français, et à celui de 10<sup>e</sup> année soumis à des élèves des dispositifs d'insertion scolaire et professionnelle, considérés comme n'ayant pas acquis les objectifs de 9<sup>e</sup>).

Comme on pouvait le supposer, les résultats suivent une certaine progression: les élèves de A obtiennent de meilleurs résultats dans les quatre composantes (reconnaissance de mots, lexique, syntaxe et compréhension), ceux des classes hétérogènes obtiennent des résultats légèrement en-dessous tout en restant proches et ceux de B obtiennent des résultats inférieurs aux deux autres grou-

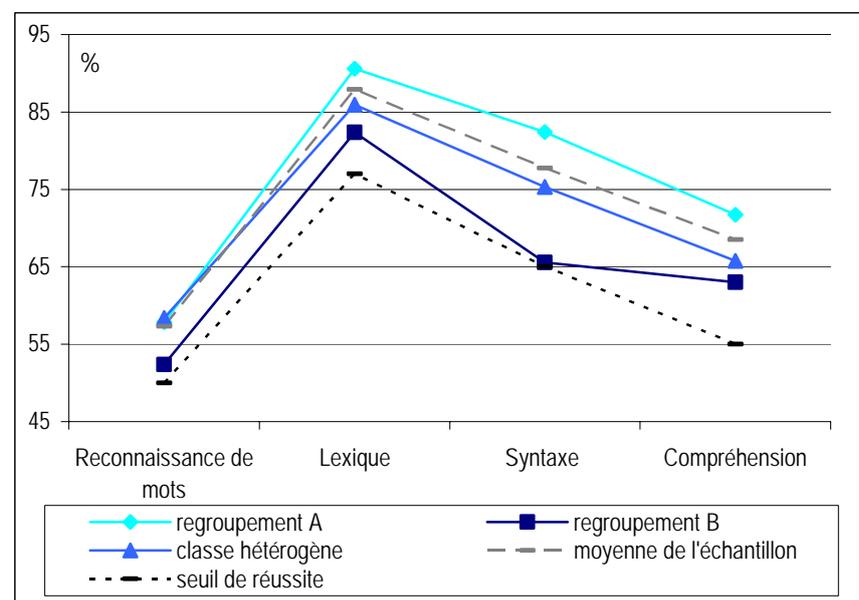
**Graphique 2. Profils de compétences (en %)**



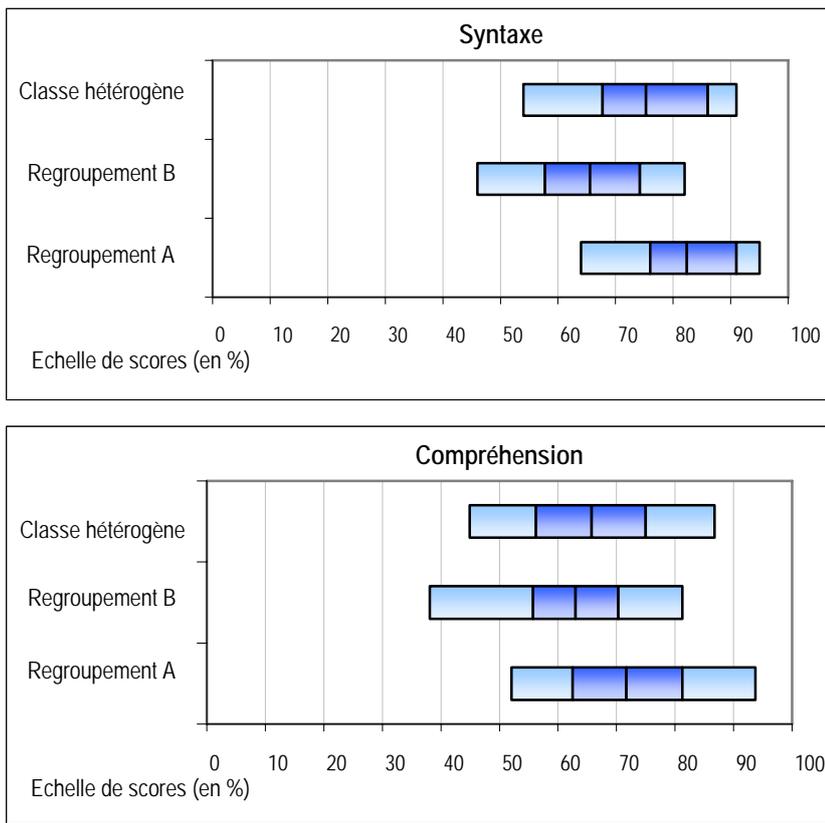
**Graphique 3. Profils de compétences en fonction de la réussite à l'EC de 2P**



**Graphique 4. Résultats moyens des élèves dans les quatre dimensions en fonction du type de regroupement ou de classe (en %)**



**Graphique 5. Distribution des résultats à l'intérieur de chaque regroupement ou type de classe pour la syntaxe et la compréhension**



ves (15%) n'a réussi que dans une seule composante, voire dans aucune. C'est pour la syntaxe que les écarts sont les plus importants puisque 52% seulement des élèves ont atteint le seuil de réussite contre 94% en A et 81% en H.

**Liens entre les résultats aux deux types d'évaluation**

Tout d'abord, nous nous intéresserons aux corrélations entre les différentes composantes du test diagnostique et les parties de l'épreuve cantonale de 6P en français. Il est intéressant de constater que les corrélations entre les quatre composantes du test, bien que significatives, ne sont pas très élevées (la plus importante étant de .42 entre le lexique et la syntaxe). Elles semblent donc bien mesurer des compétences différentes qui composent la compétence en lecture. Elles sont plus élevées entre parties de l'épreuve cantonale ou encore entre ces dernières et les composantes du test diagnostique. Ainsi, c'est avec la composante syntaxe que les liens avec l'épreuve cantonale sont les plus élevés (plus de .50, atteignant .60 avec le français I) et dans une moindre mesure entre la compréhension et le français I (.51).

Nous montrerons à titre d'exemple (la situation avec la partie structuration de l'épreuve de 6P est très comparable) les rendements moyens des élèves de 7<sup>e</sup> dans les quatre composantes en fonction de leur réussite à l'épreuve de 6P en compréhension de l'écrit. Comme on peut le constater, quelle que soit la composante, les élèves qui avaient réussi la partie compréhension de l'écrit de l'épreuve de français 6P obtiennent de

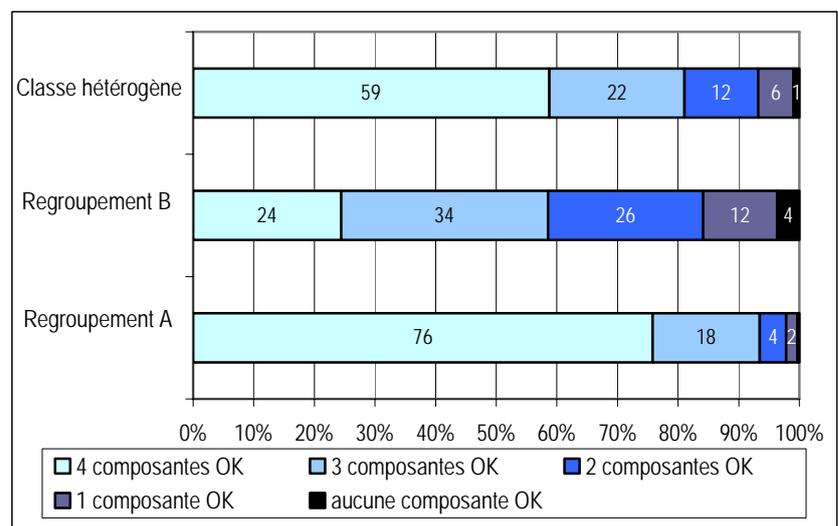
pes (pour la syntaxe et la reconnaissance de mots, ils se trouvent en moyenne au niveau du seuil de réussite)<sup>9</sup>. Quelle que soit la composante, les différences de résultats sont significatives statistiquement. On observe des différences de réussite entre les différents domaines dus sans doute au type d'exercice les composant<sup>10</sup>. Le lexique et la syntaxe donnent lieu à des résultats très élevés en moyenne, en particulier en A. Par ailleurs, c'est pour la composante compréhension que les écarts de réussite entre les élèves de A et de B sont les plus importants (de 94 à 52%), ce qui n'est pas étonnant: la compréhension est supposée plus complexe par exemple que la reconnaissance de mots qui fait appel à des processus de bas niveaux, en principe acquise par la majorité des élèves de cet âge (**graphique 5**).

Comme on peut le constater aussi bien en classe hétérogène que dans les deux regroupements A et B, on observe des écarts entre élèves. En d'autres termes, un certain nombre d'élèves ont obtenu des résultats caractéristiques aussi bien d'élèves de A, de B ou de H.

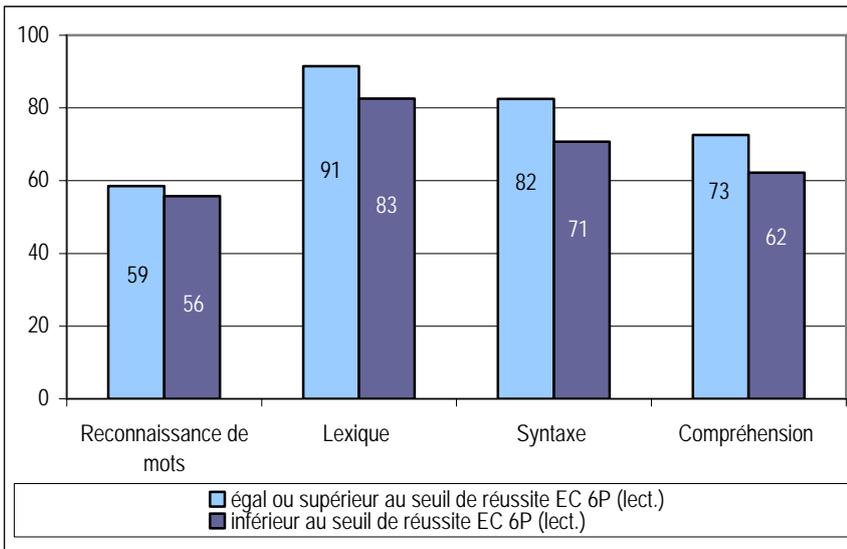
Le **graphique 6** confirme les observations précédentes. En A, plus des trois quarts des élèves atteignent ou dépassent le seuil de réussite pour les quatre

composantes et 18% réussissent trois composantes (une très petite proportion d'élèves ne réussissent qu'une, voire aucune composante). Dans les classes hétérogènes, la proportion d'élèves ayant réussi dans les quatre composantes est de 59% et de près d'un élève sur quatre dans les classes de B. Chez ces derniers, soulignons qu'une minorité d'élè-

**Graphique 6. Profils de résultats en fonction du type de regroupement ou de classes**



**Graphique 7. Rendements\* dans les quatre composantes en fonction de la réussite à l'EC de 6P (compréhension de l'écrit)**



\*Le rendement est le rapport entre le nombre de points obtenus et le nombre de points maximum.

des mots, syntaxe et dans une moindre mesure compréhension) semblent déterminées aussi bien par les caractéristiques individuelles (appartenance au regroupement A ou B ou le fait d'être "en retard") que par les autres composantes; le lexique semble étonnement n'être influencé que par la syntaxe.

### Résultats aux tests au "10<sup>e</sup> degré"

Près de 600 élèves provenant des compléments de formation (CF) des écoles de commerce et de l'ECG et du SCAI (CEFP, CIPA et préapprentissage)<sup>11</sup> ont été soumis au test diagnostique.

Ces filières de formation sont assez différentes au niveau de leur composition. Si les deux types de compléments de formation se ressemblent beaucoup (les élèves peuvent être attribués à l'un ou l'autre des deux types d'écoles indépendamment de leur projet de formation future), les trois filières du SCAI sont assez différentes entre elles et par rapport aux deux types de CF. Ainsi, la proportion d'élèves provenant du milieu le plus défavorisé varie de 54-56% dans les CF à près des deux tiers des élèves pour les élèves de CEFP ou de préapprentissage, voire à trois quarts pour ceux des CIPA. On peut faire le même type d'observation pour la proportion d'élèves dont la première langue parlée n'est pas le français (un peu plus de la moitié dans les classes de CF ou de préapprentissage versus près des trois quarts dans les classes de CIPA et de CEFP). Cette composition a inévitable-

meilleurs résultats que leurs camarades (graphique 7).

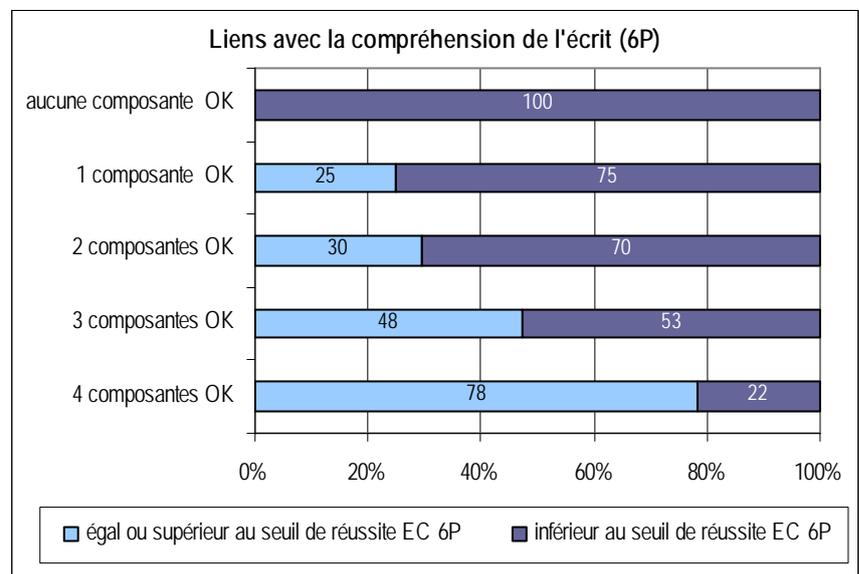
Si l'on observe les profils, on constate une correspondance un peu moins bonne qu'en 3P, bien que tout à fait satisfaisante. Si tous les élèves ayant échoué à l'épreuve de 6P en compréhension de l'écrit n'ont réussi dans aucune des quatre composantes du test diagnostique, à l'autre extrême, la correspondance est un peu moins bonne pour ceux qui ont réussi dans les quatre composantes et surtout dans trois. Ainsi, 80% des élèves ayant réussi dans les quatre composantes avaient également atteint le seuil de réussite de l'épreuve de 6P en compréhension de l'écrit mais un élève sur cinq ne l'avait pas réussi. Parmi ceux qui réussissent dans trois composantes sur quatre, la situation est plus contrastée puisque moins de la moitié des élèves seulement avait atteint le seuil de réussite en compréhension de l'écrit (graphique 8).

Enfin, nous allons essayer de mettre en évidence au moyen d'analyses de régressions multiple et logistique ce qui, parmi différentes caractéristiques (individuelles et composantes du test diagnostique), explique le mieux les résultats ou la réussite à l'épreuve de français de 6P. La réussite à l'épreuve en compréhension de l'écrit (analysée par une régression logistique) semble être assez peu expliquée par les différentes variables prises en compte à l'exception du fait de ne pas être "en retard", de se retrouver dans le regroupement A l'année suivante ou encore d'avoir réussi également en français I. L'analyse de régression multiple montre que les résultats à cette

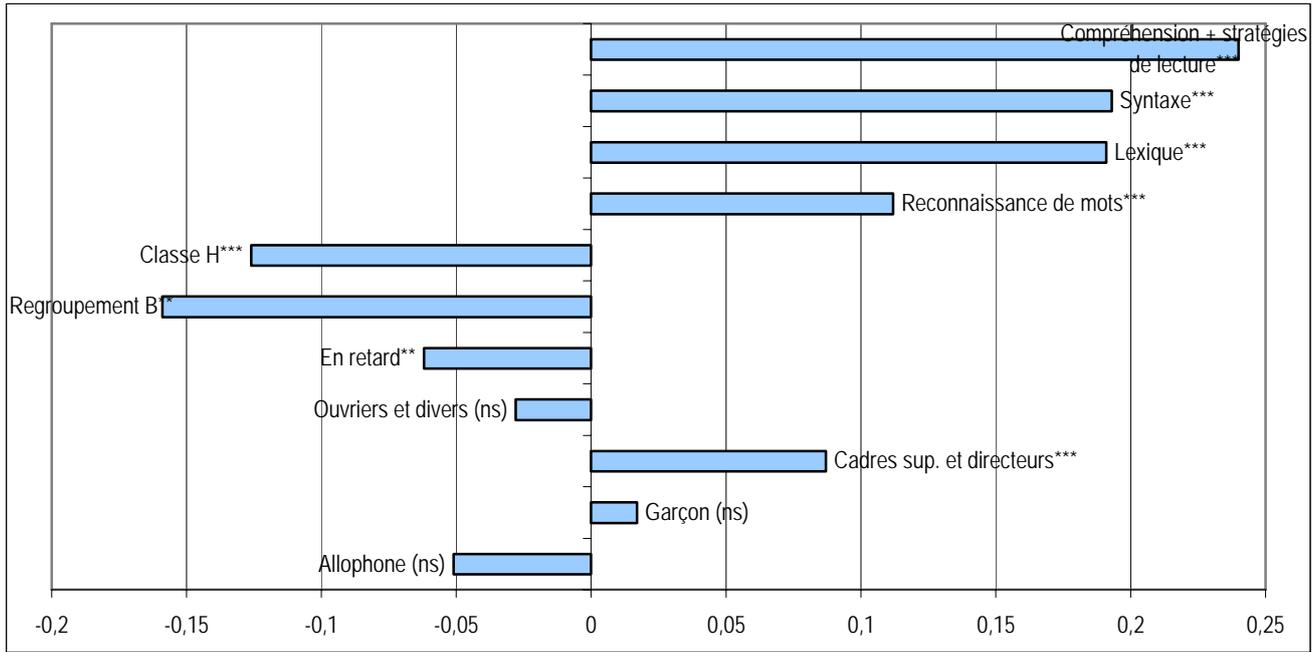
épreuve (et en particulier en compréhension de l'écrit) sont liés à plusieurs facteurs. Ainsi, le fait d'avoir du "retard", de se retrouver dans le regroupement B semblent avoir des effets négatifs sur les résultats tandis que le fait d'être enfants de cadres supérieurs a un effet positif. Par ailleurs, les résultats dans les différentes composantes paraissent également avoir des liens avec ceux de l'épreuve cantonale (graphique 9).

Enfin, si l'on regarde de plus près les effets des caractéristiques individuelles et des composantes sur l'une ou l'autre d'entre elles, il ressort comme en 3P que certaines composantes (reconnaissance

**Graphique 8. Profils de compétences en fonction de la réussite à l'EC (compréhension de l'écrit) de 6P**



**Graphique 9. Résultats de l'analyse de régression multiple pour la 7<sup>e</sup> - Score de compréhension de l'écrit (6P)**



\*\*\* Significatif à un seuil de .01 - \*\* Significatif à un seuil de .05 - (ns) Non significatif - Part de variance expliquée: 56%

ment des effets sur les résultats, comme on pourra le constater (graphique 10).

Globalement, les résultats suivent la même progression dans les différents types de formation: le rendement est meilleur pour le lexique que pour les trois autres composantes, malgré une proportion relativement importante d'élèves allophones. Comme on pouvait le supposer, les résultats diffèrent selon le type de formation. Dans l'ensemble, les élèves des CF obtiennent de meilleurs résultats que ceux du SCAI (et en particulier ceux des classes de CEFP). Selon les composantes, les résultats sont un peu différents: pour la reconnaissance de mots, les élèves des CF-ECG ont des résultats qui se différencient statistiquement de ceux des CF-EC, ces derniers ayant des résultats comparables aux élèves de préapprentissage. Pour les autres composantes, les élèves des CF des deux types d'écoles ne se différencient pas. Pour la composante compréhension, le clivage au niveau des résultats se fait clairement entre les CF et le SCAI (graphique 11).

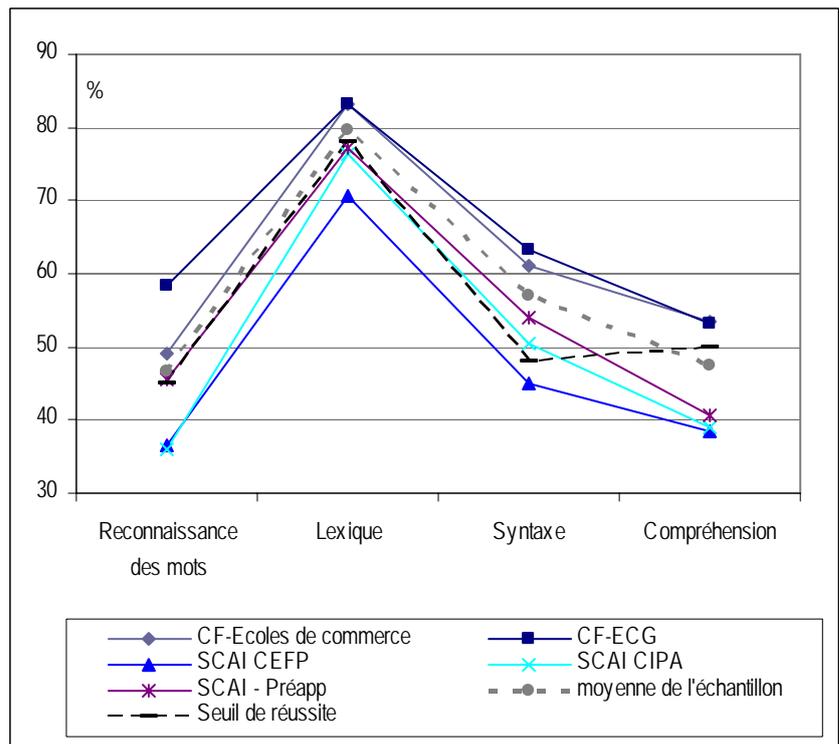
Comme on peut le constater au travers de ces deux graphiques, les écarts entre élèves sont relativement importants à l'intérieur de chaque type de formation. On observe un certain recouvrement de résultats même si les résultats des élèves des CF sont plus élevés que pour

ceux des autres types de formation (graphique 12).

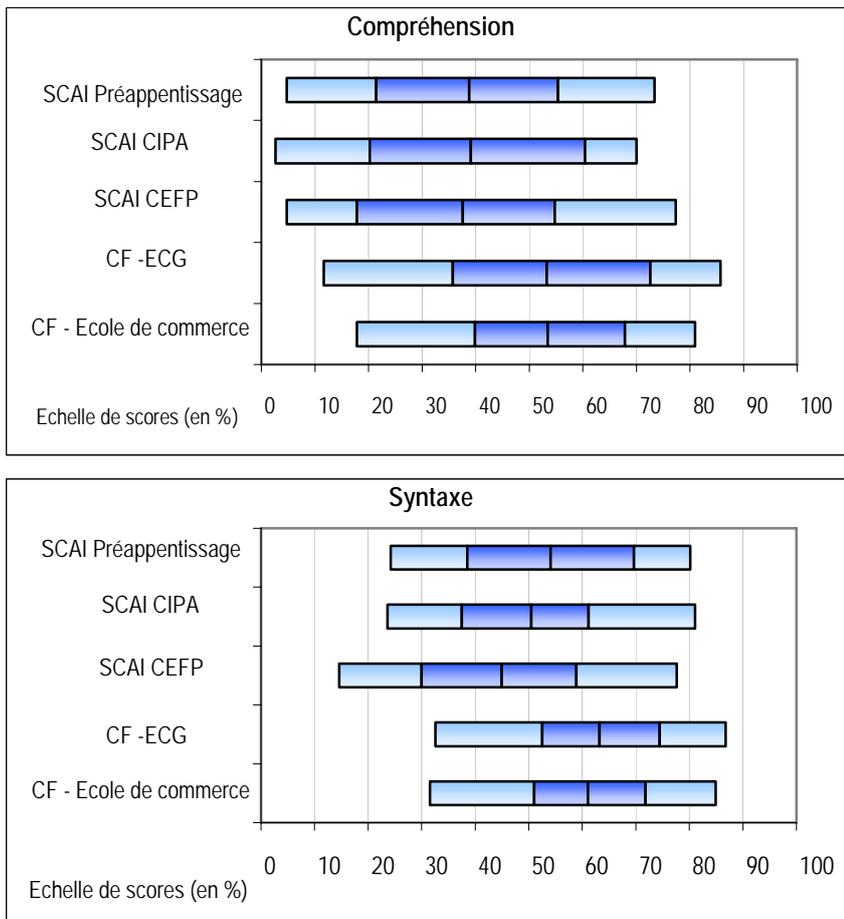
Le profil de compétences confirme les résultats précédents: dans les CF, près d'un tiers des élèves ont atteint le seuil de réussite pour les quatre composantes (un peu moins chez les élèves de l'ECG);

la proportion est nettement inférieure pour les élèves du SCAI et en particulier des classes de CIPA ou de CEFP. A l'autre extrême, très rares sont ceux qui n'ont réussi dans aucune composante chez les élèves des CF (aucun à l'ECG). Au SCAI, la proportion varie de 17 à 32%.

**Graphique 10. Résultats dans les différentes composantes en fonction du type de formation**



**Graphique 11. Distribution des résultats à l'intérieur de chaque type de formation pour la syntaxe et la compréhension**



maîtrisant pas suffisamment la langue ou pas encore arrivés à Genève)<sup>12</sup>.

Si l'on observe les corrélations entre les quatre composantes, on constate que ce sont entre le lexique et la syntaxe qu'elles sont le plus élevées (.53). La reconnaissance des mots qui fait appel à des processus de "bas niveau" présente de manière générale des liens plus faibles avec les trois autres composantes (moins de .40). Ce sont pour les mêmes composantes (lexique et syntaxe) que l'on trouve des liens avec la partie tronc commun de l'épreuve commune de français (autour de .60). Il y aurait donc plus de lien entre les composantes faisant plutôt appel à des aspects formels et l'épreuve commune de français qu'avec la reconnaissance de mots et la compréhension.

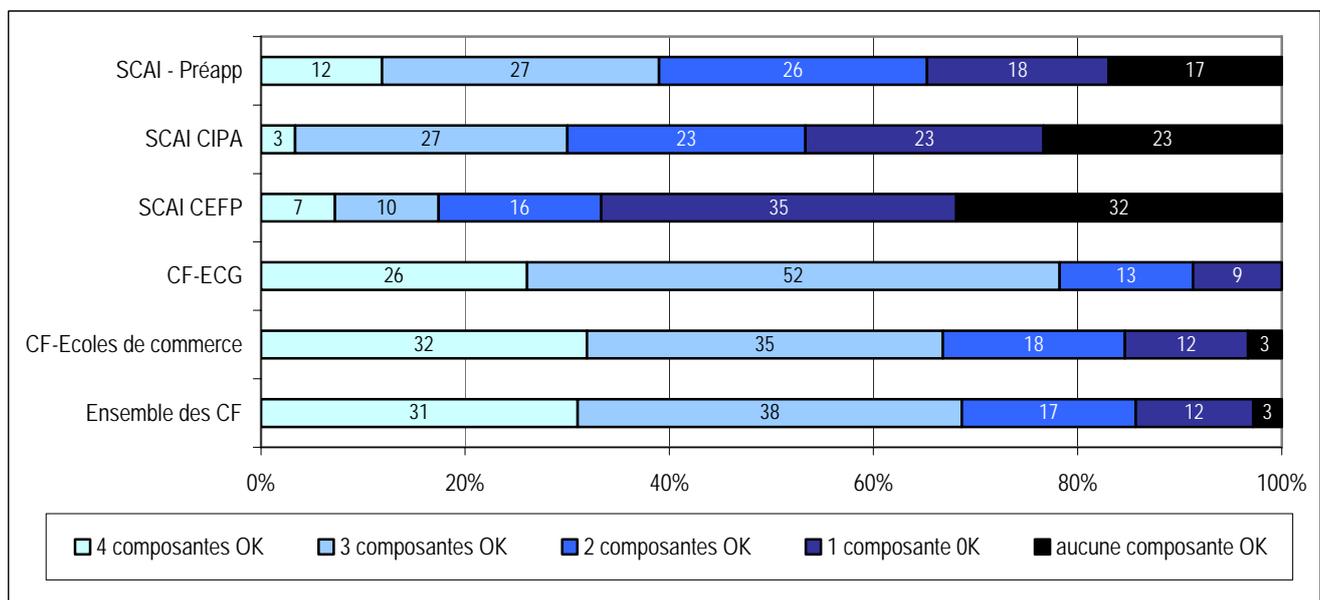
On observe également une assez bonne correspondance entre les résultats obtenus à l'épreuve commune de français de 9<sup>e</sup> et les résultats au test diagnostique: trois quarts des élèves qui avaient réussi l'épreuve ont atteint le seuil de réussite aux quatre composantes. A l'autre extrême, 86% des élèves qui n'avaient pas réussi l'épreuve ont échoué dans les quatre composantes. On peut s'interroger sur les résultats des 14% qui avaient réussi l'épreuve et échouent à toutes les composantes (**graphique 13**). Ce résultat est difficile à expliquer si ce n'est par le fait que les deux types d'évaluation ne mesurent pas exactement les mêmes compétences et sont donc complémentaires. De plus, certains exercices du test diagnostique peuvent s'avérer assez éloignés des tâches scolaires habituelles.

**Liens entre les résultats aux deux types d'évaluation**

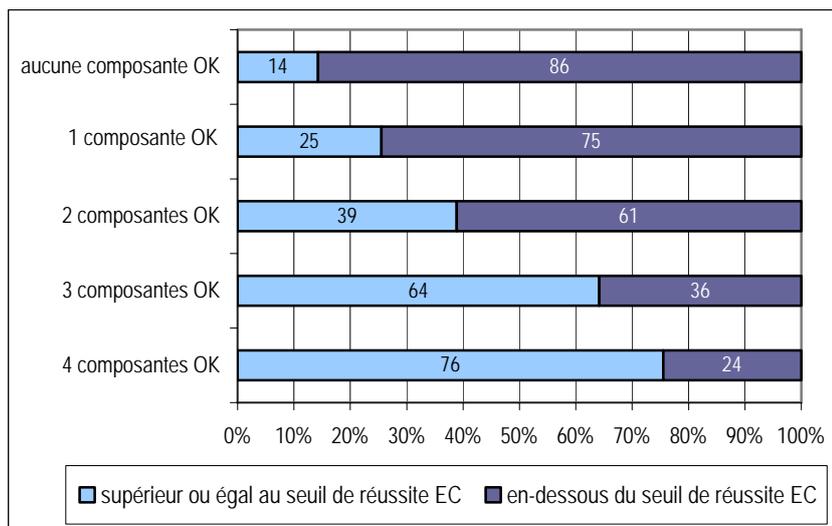
Pour cette partie, l'échantillon d'élèves est nettement réduit (69% des élèves qui ont passé le test diagnostique), étant

donné qu'une partie non négligeable d'élèves n'ont pas passé l'épreuve commune de français de 9<sup>e</sup> en 2007 (en particulier au SCAI) pour des raisons diverses (déjà en formation au PO, ne

**Graphique 12. Profils de résultats en fonction du type de formation**



**Graphique 13. Profils de compétences en fonction de la réussite à l'EC de français (tronc commun) de 9<sup>e</sup>**



téristiques individuelles, seul le fait de se trouver en classe de préapprentissage a un effet (négatif). Il est intéressant de constater que le type de formation occulte, par la composition des élèves les fréquentant, les différentes caractéristiques individuelles (le fait d'être un garçon ou une fille, d'être "en retard", de provenir d'un milieu défavorisé ou favorisé, d'être suisse ou non). L'analyse de régression logistique qui cherche à prédire la réussite à l'épreuve commune 2007 confirme certains résultats: un effet négatif de l'appartenance à ces différents types de formation (par opposition aux CF), un lien avec la réussite dans la composante syntaxe mais également le fait d'être une fille qui a un effet positif. Certaines caractéristiques telles que le fait d'être issu d'un milieu défavorisé et provenir d'une classe de CEFP ont une influence négative sur la réussite au tronc commun de l'épreuve

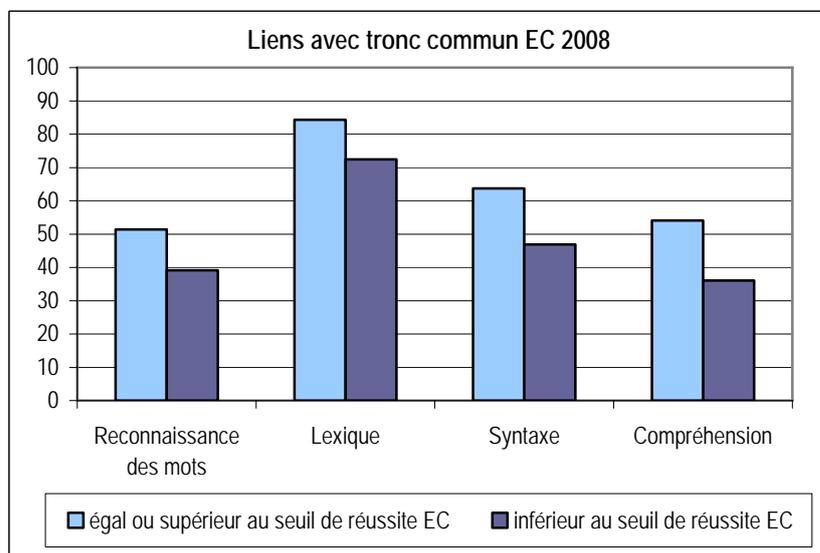
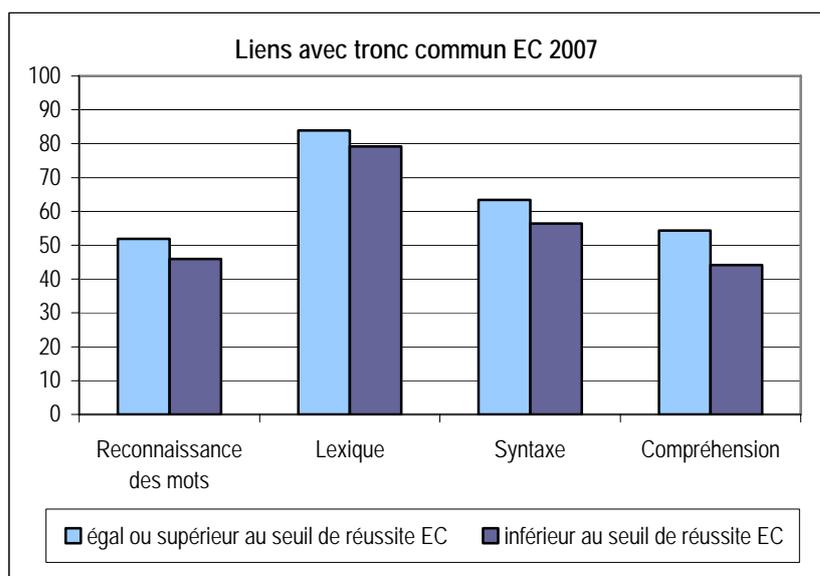
Les élèves qui avaient réussi le tronc commun de l'épreuve commune de français en 2007 ont toujours des résultats meilleurs aux différentes composantes du test diagnostique (graphique 14).

Si l'on peut observer une gradation de résultats aux épreuves communes selon le type de formation (meilleurs résultats chez les élèves des CF que dans les trois types de formation du SCAI), les écarts s'atténuent pour le tronc commun de l'épreuve commune 2008. En d'autres termes, les élèves des trois types de formation du SCAI semblent avoir davantage progressé que ceux des autres formations (graphique 15).

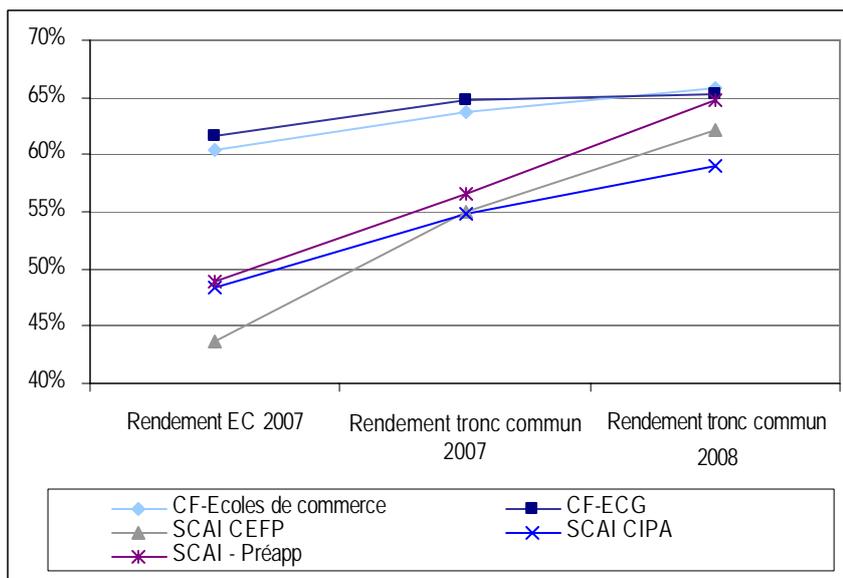
Enfin, les analyses multivariées incluant les variables individuelles, y compris le type de formation, ainsi que les résultats aux différentes épreuves communes ou parties d'épreuves communes de français permettent de dégager les éléments suivants: les résultats à l'épreuve commune de français sont liés négativement au fait de se trouver l'année d'après dans les classes de CEFP, de CIPA ou de préapprentissage mais également (positivement) aux différentes composantes du test diagnostique (reconnaissance de mots, lexique, syntaxe ou encore compréhension). En d'autres termes, les élèves fréquentant les classes de CEFP, de CIPA ou de préapprentissage ont tendance à avoir des résultats plus faibles que les autres, tandis qu'avoir de bons résultats dans les quatre composantes a un effet positif sur les résultats à l'épreuve commune de français.

Si l'on considère seulement la partie tronc commun de la même épreuve, les différentes composantes jouent également un rôle tandis que parmi les caractéristiques individuelles, seul le fait de se trouver en classe de préapprentissage a un effet (négatif).

**Graphique 14. Rendement aux quatre composantes en fonction de la réussite à l'EC de français de 9<sup>e</sup> (2007)**



**Graphique 15. Résultats aux EC de français 2007 et 2008 en fonction des types de formation**



commune 2008, tandis que la réussite aux quatre composantes a un effet positif sur les résultats au tronc commun en 2008.

Si l'on regarde les effets des caractéristiques individuelles et du type de formation suivie pour chacune des composantes, on peut faire les constats suivants: dans la plupart des cas, le fait d'appartenir à une classe de CEFP, de CIPA et le plus souvent de préapprentissage (sauf pour la reconnaissance des mots) a un effet négatif sur les résultats. Sinon, selon les cas, le fait de ne pas être suisse ou d'être allophone ou encore d'être "en retard" semble également jouer un rôle négatif au-delà de l'appartenance à un type de formation.

## Discussion

Comme attendu, les deux types d'épreuves donnent lieu à des résultats complémentaires et ce, quel que soit le degré considéré. Ces deux épreuves externes poursuivent des buts différents par définition: l'une (épreuve cantonale ou commune), conçue de manière institutionnelle, sert à vérifier les acquis des élèves à la fin d'un cycle et participe à leur orientation; l'autre cherche à diagnostiquer précisément où se situent les lacunes du point de vue de la lecture. S'il y a des liens entre les résultats obtenus aux deux types d'épreuves – les élèves ayant atteint le seuil de réussite pour l'épreuve institutionnelle ont également mieux réussi le test diagnostique –, les dimensions mesurées dans les deux types d'épreuves ne semblent pas être de même nature: par exemple, ce qui relève de la compréhension dans l'une n'est pas for-

cément corrélé avec la dimension équivalente dans l'autre test. Toutefois, aux trois degrés, on a pu observer une plus grande influence des composantes du test diagnostique sur les résultats aux épreuves cantonales ou communes que des caractéristiques individuelles. Il y a donc un lien indéniable entre les deux types d'instruments.

Il est intéressant de relever que malgré des logiques et des concepteurs différents concernant les épreuves cantonales ou communes dans les trois degrés considérés (3<sup>e</sup>, 7<sup>e</sup> et "10<sup>e</sup>" degrés) (cf. Soussi, Guilley, Guignard et Nidegger, 2009), on peut faire les mêmes constats lorsqu'on compare les résultats des élèves à ces épreuves et au test diagnostique. On observe donc à la fois une certaine cohérence mais également une complémentarité entre ces deux évaluations.

Si ces évaluations se situent en premier lieu davantage du côté de l'apprenant que du système, elles visent toutefois des objectifs différents (l'une plutôt certificative, l'autre diagnostique) et s'adressent à d'autres publics (l'une à tous les élèves, l'autre à un petit échantillon d'élèves présentant des lacunes dans le domaine de la lecture). Les évaluations externes institutionnelles peuvent jouer plusieurs rôles, comme le soulignent Ntamakiro et Tessaro (2002): avoir un rôle certificatif, mais aussi servir à réguler les pratiques et informer les familles. Le test diagnostique devrait quant à lui être déclencheur de remédiation et être en cela formatif, à condition de fournir une aide aux enseignants au niveau de l'interprétation des résultats et de la remédiation (formations, activités). Il faudrait

laisser à ce test ce rôle et le destiner aux élèves présentant des difficultés car il n'est pas conçu au départ pour vérifier les compétences de l'ensemble des élèves et ne devrait pas avoir des répercussions sur le programme scolaire.

Dans le cas des épreuves genevoises, elles sont clairement construites, comme nous l'avons souligné, dans la droite ligne de ce qui est enseigné, tandis que le test diagnostique porte sur des compétences plus éloignées des programmes scolaires mais entrant dans la construction de la compétence de lecteur.

Comme on a pu le constater dans les analyses des résultats, une majorité d'élèves réussissent dans les quatre composantes du test diagnostique. Parmi ceux qui échouent il est important de regarder de plus près où se situent les lacunes et c'est à cela que ce test doit servir: diagnostiquer plus finement les difficultés que ne pourrait le faire une épreuve cantonale ou commune ou encore les épreuves des enseignants, dont ce n'est pas le but. Une fois le diagnostique opéré, il s'agit d'apporter des remédiations. Le test devrait permettre de confirmer les difficultés mises en évidence par d'autres instruments davantage liés au programme scolaire et l'atteinte des objectifs.

Actuellement, le test diagnostique fait l'objet de modifications et d'adaptations par l'intermédiaire de chercheurs de la FPSE (Prof. P. Zesiger, S. Cèbe et Th. Thévenaz) et une réflexion concernant les situations de remédiations à proposer aux enseignants est en cours dans les trois niveaux d'enseignement. Une fois le test modifié, un manuel d'utilisation et de correction élaboré et le dispositif de remédiation mis en place, il sera nécessaire de décider précisément à qui s'adresse l'outil. Chaque direction d'enseignement devrait définir la manière d'utiliser ces tests: de manière généralisée auprès d'une population d'élèves considérée "à risque" ou mis à disposition des enseignants. ■

## Notes

<sup>1</sup> A partir de maintenant, nous appellerons les dispositifs d'insertion scolaire et professionnelle le "10<sup>e</sup> degré" pour des raisons de lisibilité bien qu'il ne s'agisse pas véritablement d'un 10<sup>e</sup> degré mais d'un degré de transition entre le secondaire I et le secondaire II.

<sup>2</sup> La principale condition pour entrer dans le Réseau d'enseignement prioritaire (REP) est que l'école doit être composée d'une proportion au moins égale à 55% d'enfants de milieux défavorisés (ouvriers et divers/sans indication).

<sup>3</sup> A titre expérimental, le test a également été passé à un échantillon d'élèves provenant des formations initiales en 2 ans dont nous ne parlerons pas ici.

<sup>4</sup> Nous avons choisi de fixer ainsi le seuil de réussite en regardant également les résultats obtenus aux prétests de l'année précédente passés dans des classes contrastées du canton. Il n'était pas possible de se baser sur une norme attendue étant donné que certains exercices limités en temps ne pouvaient être terminés. Il ne s'agit pas véritablement d'un seuil de réussite mais plutôt d'un seuil au-dessous duquel un élève peut être considéré comme ayant des difficultés importantes. Toutefois, pour des raisons de lisibilité, nous utiliserons le terme de seuil de réussite.

<sup>5</sup> Une corrélation mesure le lien entre deux variables. Elle varie de -1 à +1. Plus elle s'approche de 1, plus le lien est fort entre les deux variables. Quand elle est proche de 0, le lien est faible voire inexistant. Une valeur négative indique un lien, mais inverse.

<sup>6</sup> L'analyse de régression multiple permet de décrire la relation entre une variable dépendante quantitative (ici le score de l'EC) et des variables indépendantes (ici les différentes caractéristiques individuelles telles que le genre, l'âge, le milieu social, la première langue parlée ainsi que les résultats aux quatre composantes du test diagnostique). Elle permet de contrôler l'effet individuel des différentes variables par la méthode des moindres carrés ordinaires (MCO). En d'autres termes, on essaie de savoir laquelle ou lesquelles de ces variables prédit le mieux le score en question.

<sup>7</sup> Précisons que si la proportion d'allophones des classes de l'échantillon se situe autour de 64%, elle varie toutefois d'une classe à l'autre, allant de 44 à 80%. On retrouve les mêmes écarts pour la proportion d'enfants d'ouvriers et divers/sans indication (67% en moyenne, avec une variation de 41 à 80%).

<sup>8</sup> La régression logistique est une technique statistique qui a également pour objectif de prédire une variable dépendante mais cette fois catégorielle, le plus souvent binaire (ici la réussite ou l'échec) à partir d'une série de variables explicatives continues ou binaires.

<sup>9</sup> Le seuil de réussite a été calculé de la même manière que celui du test de 3P (moyenne - 1 écart-type) mais on s'est également appuyé sur les résultats du prétest passé en mai auprès de 100 élèves de 6P provenant d'écoles contrastées sur le plan de la composition socio-économique et de la proportion d'élèves allophones.

<sup>10</sup> Il peut s'agir d'exercices faisant appel à des compétences plus complexes ou bien d'autres en temps limité qui ne peuvent donc être terminés (par exemple l'exercice de segmentation dans la composante reconnaissance de mots).

<sup>11</sup> En fait, en plus de ces élèves, d'autres provenant de diverses formations initiales en 2 ans ont également été soumis aux tests diagnostiques à titre expérimental. Il faut toutefois préciser que nous avons choisi d'analyser les résultats de ces 600 élèves à qui était destiné en premier lieu ce test et dont une bonne partie avait passé l'épreuve de français de 9<sup>e</sup>.

<sup>12</sup> Près de 85% des élèves se trouvaient au CO, y compris en classe d'accueil ou atelier, les autres fréquentaient le plus souvent une filière du PO (apprentissage, SCAI, notamment).

## Références

- Bonjour, E., De La Haye, F., Gombert, J.-E., Marec-Breton, N. & Ménager, N. (2007). *Outils d'évaluation-diagnostique des difficultés de lecture, destinés aux élèves genevois du primaire, du cycle d'orientation et de l'enseignement secondaire post-obligatoire* (convention D.I.P. - C.R.P.C.C. : épreuves de lecture non publiées). Centre de Recherches en Psychologie, Cognition et Communication, Université de Rennes 2 - Haute Bretagne.
- Bressoux, P. (1994). Estimer et expliquer le cas des acquisitions en lecture. In *Mesure et évaluation en éducation*, vol. 17, no 1, 75-94.
- Bressoux, P. (1995). Les effets du contexte scolaire sur les acquisitions des élèves: effet-école et effets-classes en lecture. In *Revue française de sociologie*, XXXVI, 273-294.

- De la Haye, F., Gombert, J.-E., Rivière, J.-Ph., Rocher, Th. (2006). Les évaluations en lecture dans le cadre de la journée d'appel de préparation à la défense. Année 2005. *Note évaluation 06.03 octobre*. Ministère éducation nationale enseignement supérieur recherche.

- Duru-Bellat, M., Le Bastard-Landrier, S., Piqué, C., Suchaut, B. (2004). Tonalité sociale du contexte et expérience scolaire des élèves au lycée et à l'école primaire. In *Revue française de sociologie*, 45-3, 441-468.

- Giasson, J. (1992). *La compréhension en lecture*. Bruxelles: De Boeck.

- Goigoux, R. (2004). Méthodes et pratiques d'enseignement de la lecture. In *L'apprentissage de la lecture*, Revue des HEP de Suisse romande et du Tessin, 7, 37-56.

- Gombert, J.-E., Colé, P., Valdois, S. Mousty, P., Fayol, M et Goigoux, R. (coord.) (2000). *Enseigner la lecture au cycle 2*. Paris: Nathan pédagogie.

- Gombert, J.E. (2003). Compétences et processus mobilisés par l'apprentissage de la lecture. *Document envoyé au PI-REF en vue de la conférence de consensus sur l'enseignement de la lecture à l'école primaire les 4 et 5 décembre 2003*.

- Ntamakiliro, L., Tessaro, W. (2002). La perception de l'évaluation externe des élèves par les enseignants primaire genevois. In Symposium *Epreuves externes: évaluation de la qualité des apprentissages ou des curricula*. Actes du colloque de l'Admée, Lausanne, 2002.

- Soussi, A., Guilly, E., Guignard et Nidegger, Ch. (2009). *Évaluation des acquis des élèves à l'école obligatoire. L'évaluation cantonale et communale: état de situation du dispositif existant, points forts et points faibles, perspectives d'avenir*. Genève: SRED.

- Weiss, J. (2002). L'évaluation externe dans un concept général et cohérent d'évaluation. In Symposium *Epreuves externes: évaluation de la qualité des apprentissages ou des curricula*. Actes du colloque de l'Admée, Lausanne, 2002.

### Informations complémentaires :

[anne.soussi@etat.ge.ch](mailto:anne.soussi@etat.ge.ch), 022 546 71 39

[christian.nidegger@etat.ge.ch](mailto:christian.nidegger@etat.ge.ch), 022 546 71 19

### Edition :

[narain.jagasia@etat.ge.ch](mailto:narain.jagasia@etat.ge.ch), 022 546 71 14

### Version électronique de cette note :

<http://www.ge.ch/sred/publications/notesinfo>