

VERS UNE DIFFÉRENCIATION DES ÉPREUVES CANTONALES DE RÉFÉRENCE DE DIXIÈME ANNÉE

Ladislav Ntamakiro

José Ticon

Marc Ferrer

164 / Mars 2015



Unité de recherche pour le pilotage
des systèmes pédagogiques



REMERCIEMENTS

Nous remercions sincèrement les enseignantes et les enseignants qui ont bien voulu répondre à notre questionnaire ainsi que les directrices et les directeurs d'établissement scolaire qui nous ont aidés dans cette démarche. Nous remercions également les chefs de file de français qui nous ont fourni des renseignements et des documents concernant les épreuves communes d'établissement.

Cette étude n'aurait pas été possible sans l'appui de la Direction générale de l'enseignement obligatoire (DGEO) qui nous a fourni les documents et les données relatifs aux épreuves cantonales de référence (ECR) de huitième année. Nous lui en sommes reconnaissants.

AVERTISSEMENT

Au moment où nous avons entrepris cette recherche, l'école obligatoire vaudoise qui comprend 11 degrés depuis 2013, était encore organisée en 9 degrés, de la 1^{ère} à la 9^e année. Par commodité, nous utiliserons l'ancienne numérotation des degrés dans ce rapport, même si le 10^e degré est mentionné dans le titre, en référence à la nouvelle organisation.

Dans cette étude, seul le genre masculin de certains termes est utilisé, dans le seul but de faciliter la lecture du document.

AUTEURS

Ladislav Ntamakiliro

Unité de recherche pour le pilotage des systèmes pédagogiques (URSP)

José Ticon

Haute École pédagogique du Canton de Vaud (HEP Vaud)

Marc Ferrer

Unité de recherche pour le pilotage des systèmes pédagogiques (URSP)

*Dans le cadre des missions de l'URSP,
ses travaux sont publiés sous l'égide
du Département de la formation, de la jeunesse et de la culture.
Les publications expriment l'avis de leurs auteurs
et n'engagent pas les institutions dont ils dépendent.*

SOMMAIRE

1	INTRODUCTION	7
1.1	LA VALIDITÉ DES ÉPREUVES CANTONALES DE RÉFÉRENCE DE HUITIÈME ANNÉE EN QUESTION	8
1.2	QUAND LES NOTES POSENT PROBLÈME.....	9
1.3	L'ALIGNEMENT DES ÉPREUVES EXTERNES SUR LE CURRICULUM ET L'INSTRUCTION	10
1.4	LA DIFFÉRENCIATION DES TÂCHES PAR NIVEAU DE COMPÉTENCES DANS LES TESTS STANDARDISÉS.....	12
1.5	OBJECTIFS, QUESTIONS ET HYPOTHÈSE DE RECHERCHE.....	14
1.6	MÉTHODE	14
2	APPROCHES DE L'ÉVALUATION EXTERNE AU SECONDAIRE I EN SUISSE ROMANDE.....	17
2.1	LES ÉPREUVES COMMUNES AU CYCLE D'ORIENTATION DANS LE CANTON DE GENÈVE	17
2.2	LES ÉPREUVES CANTONALES DE HUITIÈME ANNÉE DANS LE CANTON DE NEUCHÂTEL	19
2.3	LES EXAMENS CANTONAUX AU CYCLE D'ORIENTATION DANS LE CANTON DU VALAIS	20
2.4	SYNTHÈSE	22
3	ALIGNEMENT CURRICULAIRE DES ECR DE HUITIÈME ANNÉE	23
3.1	L'ECR DE FRANÇAIS	23
3.1.1	<i>Objectifs sous la loupe.....</i>	<i>23</i>
3.1.2	<i>Activités sous la loupe.....</i>	<i>26</i>
3.2	L'ECR DE MATHÉMATIQUES	30
3.2.1	<i>Objectifs sous la loupe.....</i>	<i>30</i>
3.2.2	<i>Activités sous la loupe.....</i>	<i>32</i>
3.3	SYNTHÈSE	34
4	LES RÉSULTATS DES ÉLÈVES SELON LES VOIES	37
4.1	RÉSULTATS À L'ECR DE FRANÇAIS.....	38
4.2	RÉSULTATS À L'ECR DE MATHÉMATIQUES	40
4.3	SYNTHÈSE.....	42
5	VALIDITÉ ET UTILITÉ PERÇUES PAR LES ENSEIGNANTS.....	43
5.1	LE QUESTIONNAIRE.....	43
5.2	LES ENSEIGNANTS INTERROGÉS	46
5.3	ANALYSE DES RÉPONSES	48
5.4	VALIDITÉ ET UTILITÉ PERÇUES DE L'ECR DE FRANÇAIS	48
5.4.1	<i>Degré d'alignement curriculaire.....</i>	<i>48</i>
5.4.2	<i>Degré d'alignement pédagogique.....</i>	<i>50</i>
5.4.3	<i>Degré de difficulté.....</i>	<i>51</i>
5.4.4	<i>Prise en compte de l'ECR dans la planification de l'enseignement.....</i>	<i>52</i>
5.4.5	<i>Degré d'adhésion aux modalités d'évaluation caractéristiques de l'ECR.....</i>	<i>53</i>
5.4.6	<i>Degré de cohérence entre l'ECR et les travaux significatifs</i>	<i>55</i>
5.4.7	<i>Degré de cohérence entre l'ECR et les épreuves communes d'établissement</i>	<i>56</i>
5.4.8	<i>Utilisation de l'ECR dans le cadre de l'évaluation certificative</i>	<i>57</i>
5.4.9	<i>Rôle de l'ECR dans la régulation des pratiques évaluatives des enseignants.....</i>	<i>57</i>
5.4.10	<i>Modalités de différenciation de l'ECR en fonction des voies.....</i>	<i>59</i>
5.4.11	<i>Synthèse</i>	<i>60</i>
5.5	VALIDITÉ ET UTILITÉ PERÇUES DE L'ECR DE MATHÉMATIQUES	61
5.5.1	<i>Degré d'alignement curriculaire.....</i>	<i>61</i>

5.5.2	Degré d'alignement pédagogique.....	62
5.5.3	Degré de difficulté.....	63
5.5.4	Prise en compte de l'ECR dans la planification de l'enseignement.....	64
5.5.5	Degré d'adhésion aux modalités d'évaluation caractéristiques de l'ECR.....	65
5.5.6	Degré de cohérence entre l'ECR et les travaux significatifs.....	66
5.5.7	Degré de cohérence entre l'ECR et les épreuves communes d'établissement.....	66
5.5.8	Utilisation de l'ECR dans le cadre de l'évaluation certificative.....	67
5.5.9	Rôle de l'ECR dans la régulation des pratiques évaluatives des enseignants.....	68
5.5.10	Modalités de différenciation de l'ECR en fonction des voies.....	68
5.5.11	Synthèse.....	69
6	LES ÉPREUVES COMMUNES D'ÉTABLISSEMENT.....	71
6.1	UN CADRE LÉGISLATIF ET ADMINISTRATIF RESTREINT POUR UNE PRATIQUE COURANTE	71
6.2	FRÉQUENCE DES ÉPREUVES COMMUNES D'ÉTABLISSEMENT	71
6.3	PRATIQUES D'ÉVALUATION COMMUNE EN FRANÇAIS.....	72
6.3.1	D'après les chefs de file interrogés.....	72
6.3.2	D'après quelques épreuves communes collectées	75
6.4	SYNTHÈSE	77
7	CONCLUSION	79
	BIBLIOGRAPHIE.....	83

PLAN DU RAPPORT

Ce rapport est composé de sept chapitres. Après un premier chapitre traitant de la problématique et des questions de recherche, une analyse comparative des pratiques cantonales en matière d'évaluation externe des élèves du secondaire obligatoire est présentée au deuxième chapitre.

Au troisième chapitre, la cohérence entre les différents éléments permettant d'évaluer le degré d'alignement des épreuves cantonales de référence (ECR) sur le curriculum et l'instruction est examinée. La liste d'objectifs pédagogiques annoncée aux enseignants par la Direction générale de l'enseignement obligatoire (DGEO), comme référentiel d'évaluation des ECR, est examinée du point de vue de sa représentativité du plan d'études. Les contenus des ECR, aussi bien les tâches que les consignes de passation, sont passées en revue du point de vue des objectifs et des niveaux de compétences qu'ils sont destinés à évaluer.

Au quatrième chapitre, la validité différentielle des ECR de 8^e année est examinée à la lumière d'une analyse statistique des résultats des élèves aux ECR de français et de mathématiques passées en 2011.

Les résultats d'une enquête effectuée auprès des enseignants sont présentés au cinquième chapitre.

Le sixième chapitre présente les résultats d'une enquête exploratoire menée auprès de quelques chefs de file de français au sujet des épreuves communes d'établissement.

Enfin, une synthèse des principaux résultats de l'étude et quelques recommandations en vue d'une amélioration de la validité différentielle des ECR de 10^e année sont présentés dans le septième chapitre.

1 INTRODUCTION

La mise en place des épreuves cantonales de référence (ECR) de 8^e année en 2007, après celles de 6^e en 2003, et celles de 4^e et de 2^e, en 2006, représente un changement non négligeable apporté au système d'évaluation. Par le passé, depuis les années 1960, les examens cantonaux ont toujours été administrés dans différents degrés de l'enseignement primaire, mais pas au secondaire I où les élèves sont répartis en filières aux niveaux d'exigences assez différents. L'idée d'un examen cantonal permettant d'harmoniser les conditions d'attribution du certificat de fin d'études obligatoires, hypothèse de travail au sein de l'administration scolaire, n'a jamais été mise en œuvre. C'est dire tout le changement de culture de l'évaluation que représente l'administration d'une épreuve standardisée commune à tous les élèves de 8^e du canton.

Toutes les ECR répondent aux mêmes buts définis dans la loi et le règlement scolaires (Direction générale de l'enseignement obligatoire, 2010). Entre autres, elles contribuent à *harmoniser les exigences de l'enseignement dans le canton en vue d'assurer une égalité de traitement entre les élèves*. Une forme de cette harmonisation recherchée consiste à *mettre à la disposition des maîtres des repères extérieurs à la classe permettant de situer la progression des élèves*.

Le but spécifique des épreuves cantonales de référence de 8^e est précisé dans une directive départementale, le *Cadre général de l'évaluation* (Direction générale de l'enseignement obligatoire, 2010). Il est *de fournir des repères en français et en mathématiques et d'harmoniser les exigences et les pratiques dans chacune des voies*. Ces voies sont trois, la voie secondaire à options (VSO), la voie secondaire générale (VSG) et la voie secondaire de baccalauréat (VSB) ; la VSB est la plus exigeante des trois.

Les enjeux des ECR de 8^e comme de 2^e, de 4^e et de 6^e sont faibles pour les élèves, dans la mesure où leurs résultats ne sont en principe pas pris en compte dans les notes de certification, et partiellement dans les décisions de promotion, ou d'orientation. C'est vrai même si certains enseignants ont pu intégrer les notes obtenues aux ECR dans la moyenne de leurs élèves. Les ECR jouent par contre un rôle important dans la régulation des pratiques d'enseignement et d'évaluation, dans la communication entre l'école et les parents et dans le monitoring de l'enseignement, grâce aux indicateurs de résultats du système scolaire mis à disposition annuellement.

La situation est en train de changer. Une des réformes de la nouvelle *Loi sur l'enseignement obligatoire* (LEO), adoptée par référendum le 7 juin 2011 et progressivement mise en œuvre à partir de la rentrée 2013, consiste à faire passer les épreuves cantonales de référence, du statut d'épreuves à enjeux faibles (*low stakes tests*) à celui d'épreuves à enjeux élevés (*high stakes tests*) pour les élèves. Dans cette perspective, les résultats des élèves aux ECR de français, mathématiques et allemand en fin de 8^e année HarmoS (harmonisation de la scolarité obligatoire), compteront pour un tiers de la note annuelle déterminante pour l'orientation dans les voies et les niveaux, alors que les résultats aux ECR passées en fin de 6^e et de 10^e HarmoS devront être pris en compte par les enseignants, au même titre qu'un travail significatif, dans le calcul de la moyenne annuelle.

1.1 LA VALIDITÉ DES ÉPREUVES CANTONALES DE RÉFÉRENCE DE HUITIÈME ANNÉE EN QUESTION

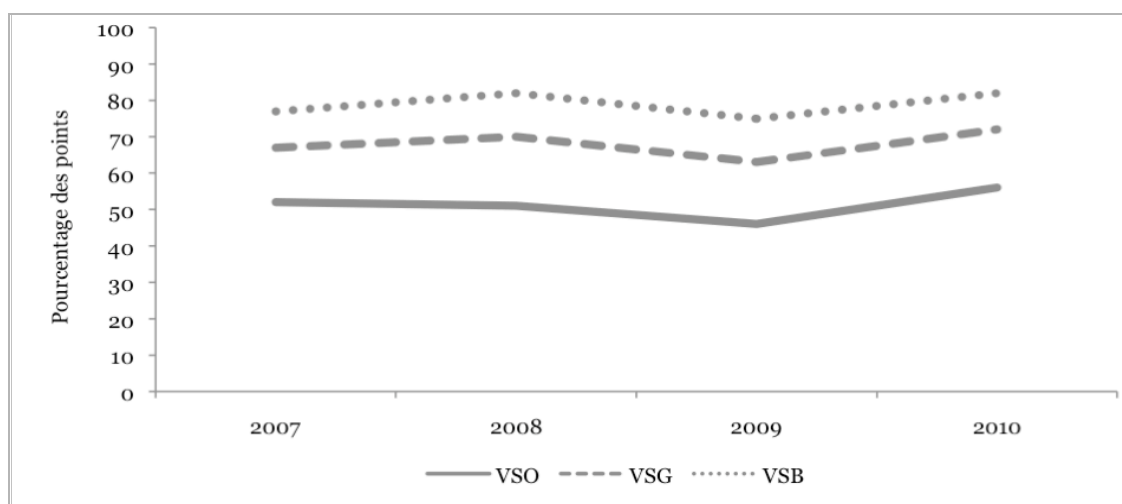
Depuis leur mise en place, les ECR de 8^e suscitent des interrogations concernant leur adéquation avec le niveau d'exigences de l'enseignement assuré dans chaque voie. Les ECR de 8^e ont-elles les qualités requises pour renseigner sur le degré d'efficacité et d'équité du système d'enseignement à l'échelle du canton et des établissements scolaires, permettre aux enseignants de situer la progression de leurs élèves, fournir des repères à leur travail d'enseignement et d'évaluation ? Les résultats aux ECR sont-ils assez fiables pour que les enseignants, qui le souhaitent, en tiennent compte au même titre qu'un travail significatif ?

Responsable du dispositif des ECR, la Direction générale de l'enseignement obligatoire (DGEO) a fait le choix d'une épreuve unique, portant sur les objectifs fondamentaux communs, pour évaluer tous les élèves de 8^e, quelles que soient les différences de niveaux de compétences, bien connues, entre les trois voies.

L'option d'une épreuve unique pour les trois voies est cohérente avec le but principal des ECR, défini dans la loi et le règlement scolaires, celui de contribuer à l'évaluation du système scolaire. Dans quelle mesure convient-elle aux autres buts de l'ECR également définis dans la loi et le règlement scolaires ?

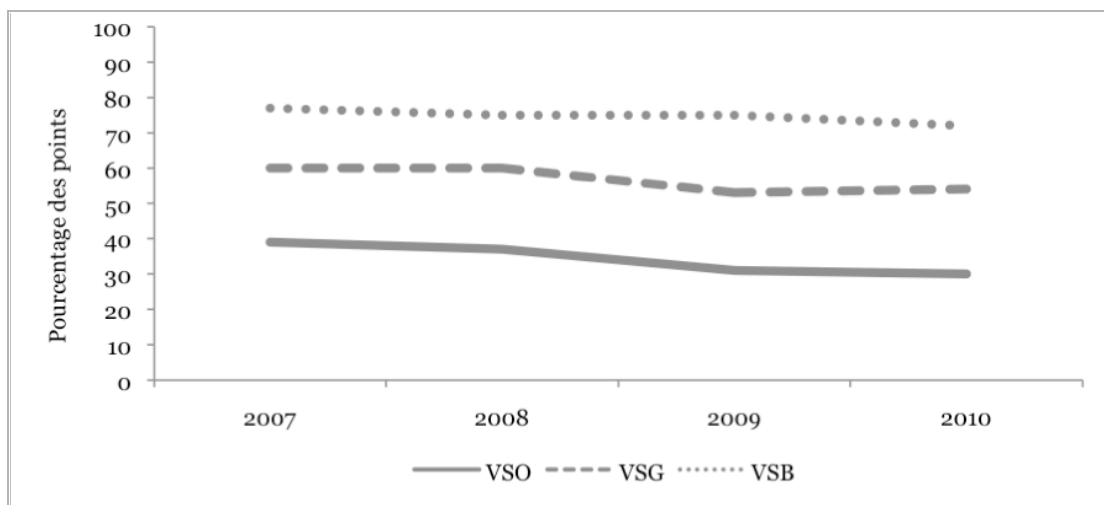
Les statistiques descriptives des résultats aux premières ECR de 8^e, en particulier la moyenne (figures 1.1 et 1.2), donnent une bonne mesure des écarts de niveaux entre les voies. Les résultats moyens des élèves sont systématiquement plus élevés en VSB qu'en VSG, et en VSG qu'en VSO.

Figure 1.1 : Moyenne des résultats selon les voies aux ECR 8^e de français de 2007 à 2010



L'existence d'un large recouvrement des distributions des résultats dans les trois voies, comme aux tests du *Programme international pour le suivi des acquis des élèves PISA* (Moreau, 2004), plaide également en faveur de la bonne qualité des ECR de 8^e, comme instruments de monitoring du système d'enseignement. A l'ECR de français de 2010 par exemple, 10% des élèves de VSO ont obtenu de meilleurs résultats que 20% des élèves de VSB et 50% des élèves de VSG.

Figure 1.2 : Moyenne des résultats selon les voies aux ECR 8^e de mathématiques de 2007 à 2010



Mais la qualité des ECR de 8^e dépend aussi de ce qu'elles fournissent aux enseignants comme repères externes permettant de situer le niveau de leurs élèves et de leur classe en comparaison avec l'ensemble des élèves de la même voie. C'est à ce titre que les ECR de 8^e pourraient contribuer à l'harmonisation des exigences et des pratiques d'évaluation des enseignants dans chaque voie. Pour cela, encore faut-il qu'elles soient adaptées au niveau moyen des élèves de chaque voie en particulier et pas seulement de toutes les voies.

A ce point de vue, l'intérêt des ECR de 8^e semble limité. Si on en juge ne serait-ce que par les moyennes des résultats aux ECR de 2007 à 2010 (figures 1.1 et 1.2), on constate qu'elles ne sont pas assez adaptées au niveau de la voie à exigences élémentaires (la VSO) ni à celui de la voie à exigences élevées (la VSB). Plus précisément, les ECR de français sont plutôt adaptées au niveau des élèves de VSG qui réussissent environ 63 à 72% des items en moyenne qu'à celui des élèves de VSO qui réussissent environ 46 à 56% des items en moyenne. Elles sont très faciles pour les élèves de VSB qui réussissent environ 75 à 82% des items en moyenne.

Les ECR de mathématiques semblent également conçues en fonction du niveau des élèves de VSG qui réussissent environ 53 à 60% des items en moyenne. Elles sont très difficiles pour les élèves de VSO qui ne réussissent qu'environ 30 à 39% des items en moyenne alors qu'elles sont faciles pour les élèves de VSB qui réussissent environ 72 à 77% des items en moyenne.

1.2 QUAND LES NOTES POSENT PROBLÈME

Les résultats aux ECR sont communiqués individuellement aux élèves et à leurs parents. La fiche ad hoc comporte pour chaque épreuve : le nombre total de points, le nombre de points obtenus par l'élève, la moyenne cantonale de la voie, la moyenne cantonale de toutes les voies, ainsi que le barème des notes pour la voie fréquentée par l'élève. Chaque élève connaît ainsi sa note à chaque épreuve. L'échelle des notes, la même que dans les évaluations assurées par les enseignants conformément à la directive départementale (Direction générale de l'enseignement obligatoire, 2010), comprend 11 niveaux, de 1 à 6 avec demi-points.

Le seuil de réussite, la limite inférieure de l'intervalle des points correspondant à la note 4, est déterminé a posteriori, au vu de la distribution des résultats de l'ensemble des élèves

du canton fréquentant le même degré ainsi que la même voie. Il est fixé au nombre de points correspondant à la moyenne moins 1/5 de l'écart type. Les niveaux inférieurs et supérieurs sont fixés à des intervalles réguliers au-dessous et au-dessus du seuil de réussite.

Les notes attribuées selon ces barèmes ne reflètent pas le niveau de réussite à l'épreuve. A l'ECR de français passée en mai 2010, par exemple, le seuil de réussite a été fixé à 53% du total des points en VSO, 70% en VSG et 81% en VSB. L'élève de VSB ayant obtenu 78% des points se trouve ainsi en échec, obtenant une note de 3.5 sur 6, alors que, avec un score identique, un camarade de VSO s'est vu attribuer la note de 5 sur 6.

La contradiction entre les notes et les résultats bruts est également forte à l'épreuve de mathématiques passée en 2010. Le seuil de réussite à cette épreuve a été fixé à 27% du total des points en VSO, 49% en VSG et 69% en VSB. L'élève de VSB ayant obtenu 67% des points se trouve ainsi en échec, avec 3.5 sur 6, alors qu'avec un score identique, celui de VSO se voit attribuer la note de 5 sur 6.

L'attribution de l'échec à des élèves de VSB ayant pourtant bien réussi l'épreuve ou à l'inverse, de la réussite à des élèves de VSO aux résultats plutôt médiocres touche une proportion d'élèves non négligeable, surtout aux ECR de mathématiques. A l'ECR de français passée en mai 2010, 34% des élèves de VSB ayant obtenu plus de 67% des points, le fameux 2/3 des points considéré comme significatif de l'atteinte des objectifs d'apprentissage (Roegiers, 2004), se sont vu attribuer la note de 3.5 sur 6. A l'inverse, 29% des élèves de VSO ayant obtenu entre 27 et 41% des points ont reçu la note de 4 sur 6, à l'ECR de mathématiques. Globalement, des élèves de VSB se voient attribuer une note insuffisante en ayant bien réussi l'épreuve, alors que des élèves de VSO bénéficient d'une note suffisante en ayant échoué la même épreuve.

La notation des élèves aux ECR de 8^e fait l'objet de critiques de nombreux enseignants. Le journal *Le Temps* du 31 août 2011 s'en est fait l'écho en ces termes : *Le calcul de la moyenne, après coup et "corrigée" en tenant compte de la dispersion des résultats, fausserait l'appréciation exacte du profil des élèves et des filières. En gros, 60% de l'effectif serait "fictivement" placé au-dessus de la suffisance. Enfin, les moyennes communiquées aux parents sont pondérées spécifiquement pour chaque filière. Si bien qu'un 4 en VSB n'est pas comparable à la même note obtenue en VSG ou en VSO.* (Danesi, 2011).

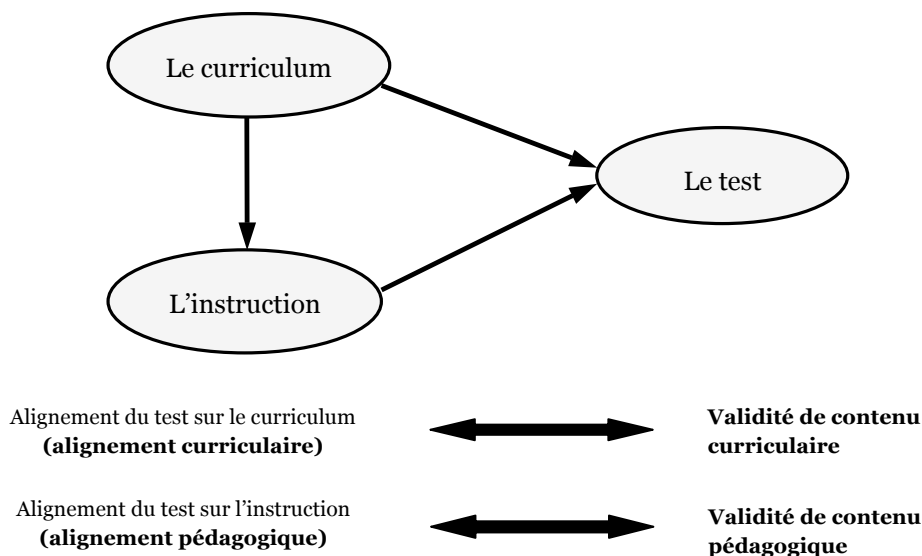
On imagine, dans ce contexte, la difficulté des enseignants à prendre en considération les notes des élèves aux ECR de 8^e, comme éléments indicatifs complémentaires dans leurs procédures d'évaluation certificative.

1.3 L'ALIGNEMENT DES ÉPREUVES EXTERNES SUR LE CURRICULUM ET L'INSTRUCTION

Suivant les théories de la mesure en éducation, les évaluations externes, doivent s'aligner à la fois sur le curriculum et l'instruction (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2004 ; Martone & Sireci, 2009 ; Resnick, Rothman, Slattery & Vranck, 2003 ; Shiel, Kellaghan & Moran, 2010). Il en va de leur validité de contenu curriculaire (curricular validity) et pédagogique (instructional validity) sans laquelle les résultats des élèves ne seraient pas fiables (Figure 1.3). Les ECR auraient une validité de contenu curriculaire dans la mesure où elles seraient alignées sur les objectifs fondamentaux définis dans le

plan d'études vaudois. Elles auraient une validité de contenu pédagogique dans la mesure où elles seraient alignées sur l'enseignement effectivement assuré dans les 3 voies.

Figure 1.3 : Schéma simplifié des relations entre le curriculum, l'instruction et le test



L'alignement des ECR sur le plan d'études ne va pas de soi. En l'absence de standards de formation décrivant les connaissances et compétences attendues de façon suffisamment opérationnelle pour être plus facilement traduites en tâches ou en situation problèmes (Klieme, 2004), l'arrimage des tâches aux objectifs pédagogiques n'est jamais sûr à cent pour cent. L'analyse des items à partir des résultats du prétest, une étape indispensable du travail d'élaboration des épreuves, permet d'ajuster les tâches mais ne garantit pas l'alignement total de celles-ci sur les objectifs pédagogiques correspondants.

Plus problématique est cependant l'alignement des ECR sur l'enseignement effectivement assuré dans les trois voies. En début d'année, la DGEO communique aux enseignants les objectifs pédagogiques visés par l'ECR qui sera passée à la fin de l'année. Au risque de favoriser le *teaching to the test*¹, cette information s'avère indispensable si on veut que tous les élèves aient reçu un enseignement correspondant aux objectifs visés dans le test. Le problème est que, pour les mêmes objectifs, l'enseignement assuré aux élèves diffère considérablement d'une voie à l'autre, tant en qualité qu'en quantité. Les filières à exigences élevées, moyennes ou faibles ne se distinguent pas seulement par la poursuite de quelques objectifs pédagogiques en plus ou en moins. Le niveau de connaissances et de compétences, visé, s'agissant des objectifs communs, n'est pas le même pour la VSO que pour la VSG ou la VSB.

Dans ces conditions, l'épreuve unique fait courir le risque d'une sous-représentation de niveaux de compétences valorisées dans l'une ou l'autre filière, comme le relèvent les associations américaines de spécialistes de la mesure et de l'évaluation en éducation :

Of particular concern is the extent to which construct Under-representation or construct-irrelevant components may give an unfair advantage or disadvantage to one or more subgroups of examinees (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2004, p. 12).

¹ L'évaluation standardisée des élèves comporte le risque que « l'on enseigne surtout en fonction de ce qu'on sait des tâches que l'on retrouve dans les tests » (Blais, 2005, p. 137), rétrécissant ainsi le programme de formation.

On peut penser que les ECR de 8^e ne présentent pas le même degré d'alignement à l'instruction pour les trois voies, certains niveaux de connaissances et de compétences davantage valorisés dans l'une ou l'autre voie étant sous-représentés dans ces épreuves uniques.

1.4 LA DIFFÉRENCIATION DES TÂCHES PAR NIVEAU DE COMPÉTENCES DANS LES TESTS STANDARDISÉS

Une autre question à se poser, au regard des distributions des résultats par voie, est celle de savoir dans quelle mesure les tâches proposées aux ECR de 8^e année sont représentatives des différents niveaux d'acquisition des compétences fondamentales. La différenciation des tâches en fonction des niveaux de compétences constitue en effet une qualité essentielle des tests standardisés. Les tests d'intelligence, modèles de l'évaluation standardisée, sont basés sur ce principe. Le tout premier d'entre eux, l'échelle métrique de l'intelligence mise au point en 1904 par Alfred Binet et Théodore Simon, comporte des questions de difficulté progressive, regroupées par niveau d'âge. Le test permet ainsi de déterminer l'âge mental, soit l'âge le plus élevé auquel l'enfant testé est capable de réussir 100% des questions. L'âge mental peut être inférieur, égal ou supérieur à l'âge réel.

La procédure d'étalonnage permet de calibrer les questions auprès d'une population de référence, sur laquelle le test est prétesté avant d'être validé. Les tâches caractéristiques d'un âge mental sont celles que plus de 75% des enfants de l'âge réel correspondant de la population de référence ont réussi.

L'enfant testé commence par l'épreuve du plus jeune âge et poursuit jusqu'au moment où il est incapable de réussir les exercices demandés. On détermine ainsi son âge mental.

La différenciation des tâches en fonction du niveau de développement caractérise également les tests d'intelligence plus complexes, mis au point à la suite du Binet-Simon. L'échelle d'intelligence pour enfants de David Wechsler (*Wechsler Intelligence Scale for Children*, WISC), aujourd'hui l'un des plus utilisés dans le monde, permet de déterminer le quotient intellectuel (QI) des enfants de 4 ans à 16 ans dans six domaines d'intelligence verbale (Information, Compréhension, Similitudes, Arithmétique, Vocabulaire, Mémoire des chiffres, Séquence lettres-chiffres) et 7 domaines de performance (Complètement d'images, Cubes, Matrices, Codes, Symboles, Arrangement d'images, Assemblage d'objets).

Le niveau d'intelligence est évalué par domaine à l'aide d'une échelle composée d'items de difficulté progressive, groupés par âge. Chaque enfant examiné effectue les tâches correspondant à son âge, à un âge inférieur s'il échoue les premières tâches, généralement les plus faciles à son âge, et à un âge supérieur s'il réussit toutes les tâches prévues pour son âge. L'examen concernant un âge donné dans un domaine d'intelligence est interrompu dès lors qu'un enfant échoue toute une série de questions.

Les premiers tests pédagogiques standardisés ont adopté le principe de différenciation des tâches par niveau de compétences, déjà éprouvé dans le testing psychologique. Le *Stanford Achievement Test* dont la première version date de 1922, se décline en 13 niveaux correspondant aux degrés de l'enseignement primaire et secondaire. Pour chaque niveau ou degré scolaire, plusieurs sous-tests permettent de déterminer le niveau de l'élève dans différents domaines tels que la compréhension orale, la compréhension écrite, l'expression écrite, les mathématiques, les sciences, etc. L'étalonnage des différents sous-tests permet d'attribuer un grade ou un âge équivalent au niveau de performances.

De nombreux tests pédagogiques standardisés, étalonnés comme les tests d'intelligence, sont encore utilisés aux Etats-Unis d'Amérique et ailleurs dans le monde aux fins de l'évaluation diagnostique, certificative ou pronostique des élèves, ou du monitoring du système d'enseignement. Toutefois, la majorité des tests pédagogiques standardisés, utilisés aujourd'hui, emprunte d'autres approches que celle des tests d'intelligence. Une grande partie des épreuves externes est conçue sur le modèle des tests à référence critériée (*criterion-referenced tests*), qui se sont développés à partir des années 1950, dans le courant des travaux de Ralph Tyler (1902-1994), de Benjamin Bloom (1913-1999) et d'autres spécialistes de la mesure en éducation (Madaus & Stufflebeam, 1988 ; Madaus & Kellaghan, 2000). A la différence des tests normatifs (*norm-referenced tests*) dont les standards de contenu et de performance sont définis à la faveur d'un étalonnage des items auprès d'une population de référence, les tests critériés reposent sur un alignement des tâches sur le curriculum et l'instruction. Dans un test standardisé à référence critériée, la différenciation des niveaux de compétences est assurée systématiquement en prévoyant des tâches représentatives de différents niveaux d'apprentissage dans chaque domaine de contenu. Les taxonomies des objectifs pédagogiques (Bloom, Englehart, Furst, Hill & Krathwohl, 1956) ont été élaborées dans ce but. La table de spécification des items assure au test la double représentation des domaines de contenu d'enseignement et des niveaux d'apprentissage dans la discipline évaluée.

L'avènement de la théorie des réponses aux items (TRI) et ses applications dans le *testing* éducatif à large échelle a permis un degré de différenciation des tests pédagogiques standardisés aux différences interindividuelles jamais atteint auparavant. Le Programme international pour le suivi des acquis des élèves (PISA) en est la manifestation la plus connue mais de plus en plus de *surveys* pédagogiques nationaux recourent également aux applications de la théorie des réponses aux items (Bertrand & Blais, 2004), afin d'optimiser la validité et la fiabilité des tests standardisés.

Réalisés tous les trois ans depuis 2000 auprès de jeunes de 15 ans, les tests PISA se répartissent en trois échelles distinctes permettant d'évaluer respectivement la compréhension de l'écrit, la culture mathématique et la culture scientifique (OCDE 2004). Chaque échelle comporte différents niveaux de compétences caractérisés par des tâches spécifiques. Les items correspondant à chaque niveau présentent de nombreuses caractéristiques communes alors qu'ils diffèrent systématiquement des items des niveaux supérieurs ou inférieurs. Le niveau de compétences de l'élève est établi selon des règles bien précises, en fonction de ses réponses à une série d'items caractéristiques des différents niveaux de l'échelle de compétences. A chaque élève est attribué le niveau dont il a réussi au moins 50% des tâches tout en ayant réussi toutes les tâches des niveaux inférieurs. Les limites de chaque niveau ont été en partie définies de manière à ce que les élèves situés au bas d'un niveau aient un taux attendu de réussite de 50% à tout test virtuel qui serait composé d'une sélection aléatoire d'items de ce niveau.

La différenciation de l'évaluation standardisée en fonction du niveau des élèves est poussée jusqu'à son individualisation, dans le cas du *testing* adaptatif sur ordinateur (TAO), en anglais *computerized adaptive testing* (CAT), une autre application de la TRI. Dans ce modèle d'évaluation, aussi bien le niveau de difficulté que le nombre et l'ordre de présentation des questions peuvent varier d'une personne à une autre, en fonction de son profil de connaissances et de compétences, estimé avant le début du test, ainsi que des performances observées et analysées au fur et à mesure qu'elle passe le test.

1.5 OBJECTIFS, QUESTIONS ET HYPOTHÈSE DE RECHERCHE

Au début des ECR de 8^e en 2007, l'option d'une épreuve externe unique pour évaluer les élèves des trois voies semblait relever du bon sens. Les objectifs fondamentaux de l'enseignement assuré dans les trois voies, référentiel officiel de l'évaluation certificative assurée par les enseignants, sont en effet les mêmes.

Le doute sur la validité des ECR de 8^e est né de la prise de connaissance des résultats de chaque voie aux différentes ECR successives ; il a été encore renforcé par les retours d'enseignants de VSO et ceux de VSB disant que le niveau d'exigences des ECR est très élevé pour les premiers, très bas pour les seconds, par rapport à celui de leur enseignement et de leurs travaux significatifs. Cette situation est d'autant plus inconfortable pour la Direction générale de l'enseignement obligatoire que ces mêmes enseignants disent ne pas tenir compte de ces ECR comme éléments indicatifs du niveau de leurs élèves ni, encore moins, comme régulateur du niveau d'exigences de leurs propres évaluations.

Les doutes persistants sur la validité et l'utilité des ECR de 8^e sont à l'origine de cette étude conduite par deux chercheurs de l'URSP et un chercheur de la Haute Ecole pédagogique vaudoise. L'hypothèse principale de cette recherche est que les ECR de 8^e n'ont pas une validité de contenu satisfaisante, permettant d'assurer une mesure fiable du niveau des élèves de chaque voie. Leur principal défaut serait une sous-représentation à la fois de niveaux inférieurs et de niveaux supérieurs de compétences, associés aux objectifs fondamentaux évalués. Une sous-représentation des niveaux inférieurs de compétences expliquerait en partie les contreperformances des élèves de VSO aux ECR de mathématiques et dans une moindre à celles de français. A l'inverse, une sous-représentation des niveaux de compétences supérieurs expliquerait en partie la réussite facile des élèves de VSB aux ECR de français et dans une moindre mesure aux ECR de mathématiques.

1.6 MÉTHODE

Notre étude repose en partie sur une démarche comparative. Nous avons enquêté sur les pratiques d'évaluation externe des élèves du secondaire obligatoire dans trois cantons romands. Les responsables des épreuves externes dans ces cantons nous ont renseignés sur la conception et les conditions d'administration de ces épreuves et nous ont fourni des exemples d'épreuves déjà passées que nous avons examinés.

L'essentiel de notre travail de recherche consiste cependant à évaluer le degré de validité des ECR de 8^e passées en mai 2011. Nous avons suivi trois démarches de recherche à cet effet. La première consiste à confronter les contenus des ECR, aussi bien les tâches que les consignes de passation, au référentiel d'évaluation, à savoir la liste d'objectifs pédagogiques évalués que la DGEO a communiquée aux enseignants et aux parents, au début de l'année scolaire. La deuxième démarche consiste à analyser les résultats des élèves. A cet effet, nous nous sommes basés sur le fichier des résultats de tous les élèves du canton, aux ECR de 8^e de 2011, mis à notre disposition par la DGEO.

La troisième piste est une enquête auprès d'enseignants de français et de mathématiques ayant fait passer les ECR en question et ayant corrigé les copies de leurs élèves. Au moyen d'un questionnaire papier-crayon également mis en ligne, nous avons interrogé les enseignants de français et de mathématiques dans les classes de 8^e de 25 établissements scolaires secondaires, sur leur perception de la validité des différentes parties des ECR,

ainsi que sur l'utilité et la conception de celles-ci. Les avis des enseignants représentent un des plus sûrs moyens d'évaluer la validité du contenu pédagogique des épreuves standardisées, même en admettant que certains d'entre eux hésitent à reconnaître qu'ils n'ont pas enseigné les contenus évalués (Hardy, 1984).

La dernière piste de recherche consiste à comparer les ECR avec les épreuves communes d'établissements. Dans ce but, nous avons mené un entretien avec les chefs de file de français dans 5 établissements connus pour organiser des épreuves communes. Les chefs de file interrogés nous ont également fourni des exemples d'épreuves communes déjà passées dans leur établissement.

2 APPROCHES DE L'ÉVALUATION EXTERNE AU SECONDAIRE I EN SUISSE ROMANDE

Quatre cantons romands sur sept recourent à l'évaluation externe des élèves du secondaire obligatoire ; mais un seul d'entre eux, le Canton de Vaud, y consacre une épreuve unique par discipline, pour tous les élèves, quelles que soient les filières. Dans les cantons de Genève, Neuchâtel et Valais, les épreuves externes destinées aux élèves du secondaire I diffèrent totalement ou partiellement selon la filière ou le niveau d'études.

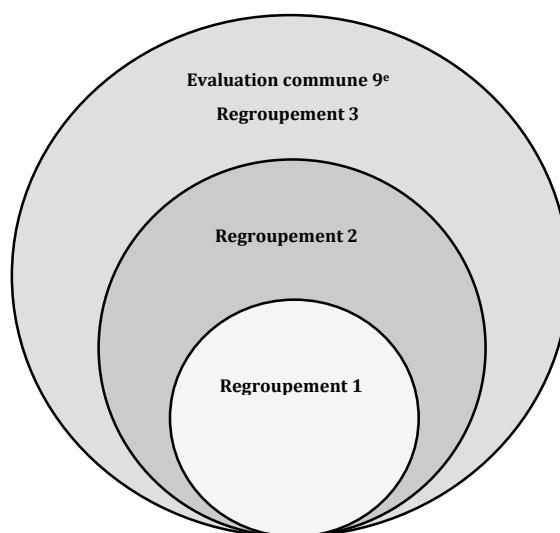
Comment les épreuves externes sont-elles différenciées selon les filières ou les niveaux dans ces cantons ? Quel est le degré de différenciation ? Nous avons interrogé et récolté de la documentation auprès des responsables des épreuves externes au secondaire I dans les cantons de Genève, Neuchâtel et Valais.

Dans ce chapitre, nous allons présenter les modalités de différenciation des épreuves communes en fonction des filières, respectivement dans les cantons de Genève, Neuchâtel et Valais.

2.1 LES ÉPREUVES COMMUNES AU CYCLE D'ORIENTATION DANS LE CANTON DE GENÈVE

Dans le Canton de Genève, des épreuves communes sont organisées dans les trois degrés du secondaire I (9^e, 10^e et 11^e) et dans les trois filières, à savoir, par ordre croissant d'exigences, le regroupement 1 (R1), le regroupement 2 (R2) et le regroupement 3 (R3).

Figure 2.1 : Modalités de différenciation des épreuves communes de 9^e année dans le Canton de Genève



Plusieurs disciplines font l'objet d'une épreuve commune passée le plus souvent vers la fin du deuxième semestre, au mois de mai : le français, l'allemand et les mathématiques dans les trois degrés ; la physique, le latin et l'anglais en plus, en 10^e et en 11^e. Les résultats obtenus à ces épreuves comptent pour 20 à 30% de la moyenne semestrielle, selon les disciplines.

Les épreuves sont partiellement différenciées en fonction des regroupements. Chacune comporte trois parties hiérarchiquement incluses l'une dans l'autre, sur le modèle des

poupées russes : une partie commune aux trois regroupements, une partie commune aux regroupements 2 et 3 et une partie spécifique au regroupement 3.

Un aperçu de la structure des épreuves de français et de mathématiques passées en 9^e année, en mai 2012, permet d'illustrer ce modèle de différenciation de l'évaluation externe en fonction des filières.

L'épreuve commune de français est conçue en trois parties : une partie commune aux trois regroupements, une partie spécifique aux regroupements 2 et 3 et une partie spécifique au seul regroupement 3. Chaque partie comprend un nombre variable d'items correspondant aux trois domaines de compétences évaluées : la compréhension de l'écrit, la production de l'écrit et la connaissance de la langue.

Tableau 2.1 : Nombre d'items (et de points) de chaque domaine de compétences évaluées à l'épreuve commune de français des 9^e selon les regroupements R1, R2 et R3

	Compréhension de l'écrit	Production de l'écrit	Connaissance de la langue	Total
1 ^{ère} partie (R1 + R2 + R3)	22 (41)	4 (9)	16 (34)	42 (84)
2 ^{ème} partie (R2 + R3)	4 (8)	2 (2)	2 (2)	8 (12)
3 ^{ème} partie (R3)	8 (11)	1 (1)		9 (12)
Total	34 (60)	7 (12)	18 (36)	59 (108)

Dans la première partie basée sur un texte narratif, la compréhension de l'écrit est évaluée à l'aide de 22 items comptant pour 41 points ; la production de l'écrit, 4 items pour 9 points ; la connaissance de la langue, 16 items pour 34 points. A raison de 42 items comptant pour 84 points sur 108, la première partie, commune aux trois regroupements, représente environ les quatre cinquièmes de l'épreuve.

La deuxième partie est basée sur un texte informatif. La compréhension de l'écrit y est évaluée à raison de 4 items comptant pour 8 points ; la production de l'écrit à raison de 2 items comptant pour 2 points ; la connaissance de la langue à raison de 2 items comptant pour 2 points.

Dans la troisième partie, sur la base d'un extrait d'un poème, la compréhension de l'écrit est évaluée à raison de 8 items comptant pour 11 points ; la production de l'écrit à raison de 1 item comptant pour 1 point. La connaissance de la langue n'y est pas évaluée.

A noter que le temps imparti à l'épreuve est le même pour les trois filières, soit 95 minutes. La quantité de tâches à effectuer dans le même temps varie ainsi selon les filières. En compréhension de l'écrit, par exemple, les élèves du regroupement I avaient pour tâche un texte à lire et un questionnaire de 41 items, ceux du regroupement II, 2 textes à lire et un questionnaire de 45 items et ceux du regroupement III, trois textes à lire et un questionnaire de 53 items.

La même structure à trois parties emboîtées caractérise l'épreuve de mathématiques passée en mai 2012. Les élèves de 9^e année du regroupement devaient résoudre 17 exercices comptant pour 62 points. Les élèves du regroupement 2 avaient 3 questions de plus que ceux du regroupement 1, soit en tout 21 questions comptant pour 78 points. Enfin, les élèves du regroupement 3 avaient 2 questions de plus que ceux du regroupement 2, soit en tout 23 questions comptant pour 95 points.

La différenciation des épreuves communes en fonction des filières ne se limite pas au contenu. En mathématiques comme en français, si les mêmes critères de correction et de cotation des items sont appliqués à tous les élèves, les barèmes de notation diffèrent. Pour chaque regroupement, un barème normatif est établi a posteriori en regard de la distribution des points.

2.2 LES ÉPREUVES CANTONALES DE HUITIÈME ANNÉE DANS LE CANTON DE NEUCHÂTEL

Dans le Canton de Neuchâtel, des épreuves standardisées à enjeux élevés sont organisées en fin de 6^e année (8^e HarmoS) et en fin de 8^e (10^e HarmoS).

Les épreuves standardisées passées en fin de 6^e année, appelées épreuves cantonales d'orientation, jouent un rôle important dans la procédure d'orientation des élèves dans les trois filières du secondaire I : la section de maturités, la section moderne et la section préprofessionnelle. Ces épreuves portant sur trois disciplines (le français, les mathématiques et l'allemand) sont passées en deux sessions : en novembre et en mai. Les résultats obtenus aux deux sessions comptent pour un tiers de la note déterminante de la décision d'orientation, les évaluations internes assurées par les enseignants constituant les deux tiers restant.

Les épreuves standardisées passées en fin de 8^e année sont très récentes. Les premières ont eu lieu en avril 2009. Elles ont pour visée principale de contribuer à l'évaluation certificative des élèves en français et en mathématiques. Les résultats à ces épreuves sont intégrés dans la moyenne annuelle de la discipline.

Les épreuves cantonales neuchâteloises de 8^e année sont différenciées en fonction des sections. La structure des épreuves est commune aux trois sections dans la mesure où il s'agit des mêmes domaines de compétences, mais le contenu est en principe différent d'une section à l'autre, seules quelques tâches pouvant être accessoirement les mêmes. Il en résulte trois épreuves de français et trois épreuves de mathématiques.

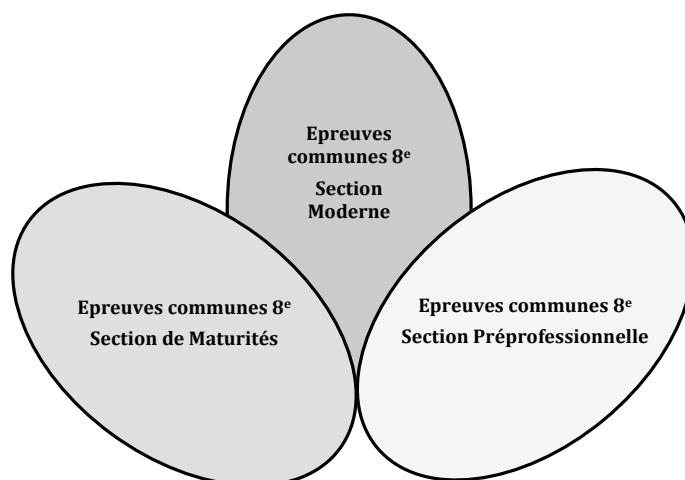
A titre d'exemple, prenons les épreuves de français et de mathématiques passées en mai 2011. Chaque épreuve de français comporte trois parties : la compréhension de l'écrit, la production de l'écrit et le fonctionnement de la langue. Chaque partie dure 45 minutes.

La compréhension de l'écrit est évaluée à l'aide de deux textes et d'un questionnaire de compréhension. Un des deux textes est commun aux trois sections. Le deuxième texte est propre à chaque section. Les questions de compréhension, y compris celles qui se rapportent au texte commun, sont en revanche différentes d'une section à l'autre.

La tâche de production de l'écrit en 45 minutes est identique aux trois sections : même situation de communication, mêmes consignes d'écriture et mêmes critères d'évaluation.

Les exercices d'évaluation des quatre composantes du fonctionnement de la langue (grammaire, orthographe, conjugaison et vocabulaire) sont tous différents d'une section à l'autre.

Figure 2.2 : Modalités de différenciation des épreuves communes de 8^e année dans le Canton de Neuchâtel



Les épreuves de mathématiques, aussi, présentent la même structure pour les trois sections. Elles sont conçues en trois parties couvrant cinq domaines prévus dans le plan d'études neuchâtelois : nombres et opérations, et géométrie, dans la première partie ; modélisation, dans la deuxième partie ; calcul littéral et grandeurs et mesures, dans la troisième partie. Un sixième domaine, fonctions, est cependant testé, dans la troisième, uniquement dans la section de maturités.

Les problèmes à résoudre par domaine diffèrent d'une section à l'autre. Peu de problèmes sont communs aux trois sections. En principe, leur niveau de difficulté correspond aux niveaux d'exigences de chaque section. C'est ainsi que quelques problèmes prévus dans la section de maturités se retrouvent en version plus simplifiée dans la section générale ou dans la section préprofessionnelle.

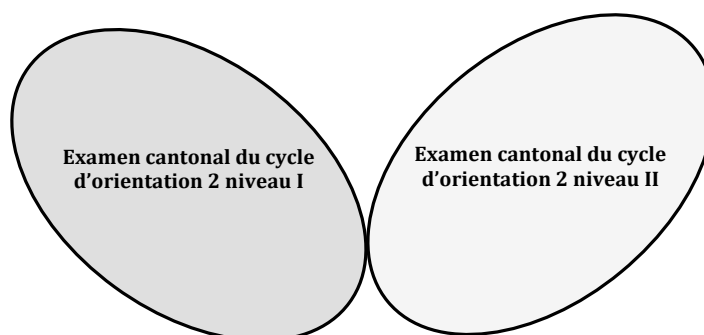
Au final, ce sont trois épreuves cantonales de français ou de mathématiques ayant quelques éléments en commun.

2.3 LES EXAMENS CANTONAUX AU CYCLE D'ORIENTATION DANS LE CANTON DU VALAIS

Le Canton du Valais organise des épreuves externes appelées examens cantonaux à l'école primaire, en fin de 4^e (6^e HarmoS) et de 6^e (8^e HarmoS), en français et en mathématiques, ainsi qu'au cycle d'orientation (secondaire I), en fin de 8^e (10^e HarmoS) et de 9^e (11^e HarmoS), en français, en allemand et en mathématiques.

Les examens cantonaux valaisans participent à l'évaluation certificative. En fin de 4^e année, le résultat à l'examen cantonal compte pour un cinquième de la note du 2^e semestre. En fin de 6^e année, il compte pour un cinquième de la moyenne annuelle de la discipline. En fin de 8^e et de 9^e, le résultat à l'examen cantonal compte pour un quart de la moyenne annuelle de la discipline. Autant dire que les enjeux des examens cantonaux valaisans sont élevés pour les élèves. Ils n'en constituent pas moins un outil de monitoring du système d'enseignement. Les statistiques des résultats aux examens cantonaux sont systématiquement établies et communiquées aux autorités scolaires et aux enseignants, et en partie au public.

Figure 2.3 : Modalités de différenciation des examens cantonaux du cycle d'orientation dans le Canton du Valais



La construction des examens cantonaux des 8^e et 9^e années est différenciée dans le respect de l'organisation de l'enseignement en classes de niveau.

L'enseignement secondaire inférieur en Valais est un tronc commun. L'enseignement est dispensé dans des classes hétérogènes pour les matières secondaires, et dans des classes de niveau pour les matières principales que sont le français, l'allemand et les mathématiques, auxquelles s'ajoute l'anglais en 8^e et en 9^e.

Deux niveaux sont distingués. Le niveau I pour les élèves forts (moyenne annuelle de 5 sur 6, ou plus dans la discipline) et le niveau II pour les élèves moyens ou faibles (moyenne annuelle inférieure à 5 sur 6).

Les standards de contenu et de performance des examens cantonaux de 8^e et 9^e varient d'un niveau à l'autre. Afin d'illustrer cette différence, prenons, à titre d'exemple, les examens cantonaux de français et de mathématiques passés en fin de 8^e, en 2011.

En français, les élèves de niveaux I et II ont été évalués à l'aide de deux épreuves distinctes. L'ossature de l'épreuve est la même. Elle est faite de quatre parties comptant pour 20 points chacune : l'expression orale (20 points), l'expression écrite (20 points), la compréhension de l'écrit (20 points) et la structuration (20 points).

L'expression orale est commune aux deux niveaux. En passation individuelle, l'élève reçoit une fiche d'une page contenant un texte court à lire et une image à analyser et commenter. Une grille d'autoévaluation également utilisée comme grille d'évaluation par l'enseignant indique à l'élève les critères d'évaluation. L'élève dispose d'environ 20 minutes de préparation et 5 minutes de lecture et de communication orale de son analyse de l'image. Les enseignants disposent d'une vingtaine de fiches qu'ils distribuent aléatoirement à chaque élève. Toutes se rapportent au thème de la science-fiction que les enseignants doivent avoir préalablement traité en classe, en guise de préparation à l'évaluation cantonale.

Les tâches d'évaluation de l'expression écrite, de la compréhension de l'écrit et de la structuration, diffèrent d'un niveau à l'autre.

L'épreuve de mathématiques est composée de deux parties passées en 75 minutes chacune pour le niveau I comme pour le niveau II. La première partie comprend 5 exercices pour le niveau I, 4 pour le niveau II, à effectuer sans calculatrice ni aide-mémoire. La deuxième partie, 5 exercices pour le niveau I et 5 pour le niveau II, à effectuer avec une calculatrice et un aide-mémoire. Aucun exercice n'est commun aux deux niveaux.

Tableau 2.2 : Le contenu de l'examen cantonal de français passé en 8^e année, en 2011, selon les niveaux I et II de l'enseignement secondaire inférieur valaisan

	Niveau I	Niveau II
<i>Expression orale</i>	Analyse et commentaire oral d'une fiche de science-fiction (une image + un texte court)	Analyse et commentaire oral d'une fiche de science-fiction (une image + un texte court)
<i>Expression écrite</i>	Rédaction d'un texte de 25-30 lignes décrivant une scène de roman policier : la découverte d'une victime d'agression. La structure et le contenu du texte ainsi qu'un guide de relecture sont donnés.	Rédaction d'un texte de 200 à 250 mots qui serait la suite d'une lettre narrative en respectant le genre narratif et le temps des verbes utilisé et en référence au guide de relecture donné.
<i>Compréhension de l'écrit</i>	Lecture d'un texte extrait d'un récit de voyage. Questionnaire de compréhension de 15 items	Lecture d'un texte extrait d'un roman policier. Questionnaire de compréhension de 16 items.
<i>Structuration</i>	Neuf exercices d'orthographe, vocabulaire, grammaire, conjugaison.	Douze exercices d'orthographe, vocabulaire, grammaire et conjugaison.

2.4 SYNTHÈSE

La construction des épreuves externes repose sur un compromis entre deux logiques d'évaluation complémentaires : celle de l'élève dont il s'agit d'estimer le degré d'apprentissage dans un but formatif ou certificatif et celle du système d'enseignement dont on voudrait assurer un monitoring aussi fiable que possible en vue de son pilotage. Les trois modalités différentes de l'évaluation externe au secondaire obligatoire illustrent bien les tensions existant entre les deux logiques.

L'approche vaudoise, l'épreuve unique sans éléments spécifiques à l'une ou l'autre voie, privilégie l'évaluation du système, au détriment de l'évaluation des élèves. Les approches neuchâteloise et valaisanne, la différenciation totale des épreuves selon les filières ou les niveaux, privilégient l'évaluation des élèves, au détriment de l'évaluation du système.

L'approche genevoise, la différenciation partielle des épreuves selon les filières, représente un compromis entre les deux logiques, mais son caractère inclusif des compétences évaluées, dans une voie à exigences moins élevées dans celles de la voie à exigences plus élevées, limite la validité de l'évaluation. Les filières à exigences moins élevées se définissent aussi par des apprentissages spécifiques et pas seulement par ce qui est commun à toutes les filières. En plus de la partie commune à toutes les filières, l'épreuve externe devrait alors justifier d'une partie spécifique à chaque filière, à l'instar du programme d'enseignement, en partie identique, en partie différent.

3 ALIGNEMENT CURRICULAIRE DES ECR DE HUITIÈME ANNÉE

Dans ce chapitre, le contenu des ECR passées en 8^e année, en mai 2011, est regardé à la loupe, afin d'évaluer son degré d'alignement au curriculum. Les ECR sont composées d'activités ou tâches conçues pour évaluer le niveau d'atteinte des objectifs fondamentaux, définis dans le *Plan d'études vaudois* (PEV)² en termes de compétences visées et de compétences associées. En référence aux modèles de définition des objectifs pédagogiques (De Landsheere et De Landsheere, 1980), les compétences visées s'apparentent aux objectifs généraux, les compétences associées pouvant être assimilées aux objectifs spécifiques. Les objectifs spécifiques définis dans le *Plan d'études vaudois* représentent les différents niveaux de complexité mis en évidence par les taxonomies des objectifs pédagogiques comme celles de Bloom et al. (1969).

Comme dans toute évaluation sommative - l'évaluation standardisée en particulier - les ECR couvrent une partie seulement des objectifs fondamentaux de l'enseignement assuré en 8^e année. En début d'année scolaire, la DGEO communique aux enseignants, via leurs directeurs d'établissement scolaire, les objectifs généraux et les objectifs spécifiques, visés par les ECR passées à la fin de la même année scolaire.

L'analyse du degré d'alignement curriculaire des ECR de français d'abord, de mathématiques, ensuite, est basée sur deux questions principales. La première question concerne le degré de représentativité des objectifs généraux (les compétences visées) et spécifiques (les compétences associées aux compétences visées). Les objectifs généraux et les objectifs spécifiques couverts par l'ECR représentent-ils une part significative du référentiel d'évaluation ? La deuxième question concerne la représentativité des différents niveaux de connaissances et de compétences. Les tâches prévues dans les ECR permettent-elles d'évaluer les différents niveaux de connaissances et de compétences prévues dans le plan d'études ? Plus précisément, dans quelle mesure l'équilibre établi dans le plan d'études entre les niveaux inférieurs (connaissance, compréhension et application) et les niveaux supérieurs (analyse, synthèse, évaluation) de la taxonomie de Bloom et al. (1969) est-il respecté à travers le contenu des ECR ?

3.1 L'ECR DE FRANÇAIS

3.1.1 OBJECTIFS SOUS LA LOUPE

Le *Plan d'études vaudois* prévoit en VSB, VSG et VSO respectivement vingt-cinq, vingt-trois et vingt-deux objectifs fondamentaux répartis en 4 domaines : *Lire et écouter*, *Ecrire et dire*, *Structurer*, *Préciser et vérifier* (tableau 3.1).

L'ECR de 2011 vérifie 20 objectifs fondamentaux, regroupés en trois parties dans la liste communiquée aux enseignants, au début de l'année scolaire : (1) *Lire, comprendre et analyser un texte poétique et narratif* ; (2) *Produire un message à visée argumentative, ici une lettre pour convaincre* ; (3) *Raisonner sur le fonctionnement de la langue : orthographe, vocabulaire, conjugaison, grammaire*.

La première partie comprend 5 objectifs fondamentaux du domaine *Lire et écouter*. La deuxième partie comprend un objectif du domaine *Ecrire et dire* et trois objectifs spécifiques à l'activité d'évaluation, mais non prévus dans le PEV. La troisième partie

2 Le *Plan d'études vaudois* (PEV) est resté en vigueur jusqu'à la mise en place du *Plan d'études romand* (PER) en 2012.

comprend quatre objectifs fondamentaux du domaine *Structurer*, trois objectifs fondamentaux du domaine *Préciser et vérifier* ainsi que trois objectifs spécifiques à la situation d'évaluation, mais non prévus dans le PEV.

Tableau 3.1 : Nombre d'objectifs fondamentaux d'enseignement du français par domaine et par voie

	Nombre d'objectifs fondamentaux prévus dans le PEV		
	En 8 ^e VSB	En 8 ^e VSG	En 8 ^e VSO
Lire et écouter	7	9	7
Ecrire et dire	7	4	3
Structurer	9	8	10
Préciser et vérifier	2	2	2
Total	25	23	22

D'après la liste des objectifs évalués que la DGEO a communiquée aux enseignants au début de l'année scolaire, l'alignement de l'ECR de français sur le *Plan d'études vaudois* est partiel. D'une part, quelques objectifs évalués ne sont pas considérés comme objectifs fondamentaux dans le plan d'études. D'autre part, plusieurs objectifs fondamentaux évalués ne sont pas communs aux trois voies.

Dans la partie *Lire, comprendre et analyser un texte poétique et narratif*, les objectifs *Définir le rôle des personnages, identifier les thèmes d'un texte et repérer les caractéristiques d'un texte poétique* sont prévus en 7^e, 8^e et 9^e, dans les trois voies. Leur évaluation est donc tout à fait légitime. L'objectif *Analyser l'organisation d'un texte*, prévu en 9^e VSB et 9^e VSG, n'est pas un objectif fondamental en VSO. En VSO, il est prévu en 8^e et 9^e, mais n'a pas le statut d'objectif fondamental. Quant à l'objectif fondamental *Repérer et analyser des contraintes poétiques*, il concerne la 8^e VSO et la 8^e VSG, alors qu'en VSB, il est prévu en 9^e année.

Dans la partie *Produire un message à visée argumentative*, l'objectif *Respecter l'univers propre au genre* fait partie des objectifs de 8^e VSO, VSG et VSB, mais il n'est pas considéré comme un objectif fondamental. De plus, il est apparenté au genre narratif, pas au genre argumentatif. L'objectif *Respecter un point de vue*, également apparenté au genre narratif et pas au genre argumentatif, est prévu en 8^e VSB. En VSG, il est prévu en 9^e année. Il est bien prévu en 8^e VSO, mais il n'est pas considéré comme un objectif fondamental. L'objectif *Respecter un temps de base donné est un objectif fondamental* en 7^e et 8^e VSB, en 7^e et 9^e VSG, ainsi qu'en 7^e VSO. Son évaluation est légitime bien qu'il soit apparenté au genre narratif plutôt qu'au genre argumentatif. Il relèverait a priori du domaine *Structurer* plutôt que du domaine *Ecrire et dire*, mais le plan d'études le veut ainsi. Enfin, le double objectif *Dans une situation polémique et tenant compte du destinataire, émettre une opinion étayée par quelques arguments ; reformuler un argument et contre-argumenter* est prévu en 7^e, 8^e et 9^e en VSB, VSG et VSO, sauf qu'il n'est pas considéré comme un objectif fondamental devant faire l'objet d'une évaluation certificative.

Dans la partie *Raisonnement sur le fonctionnement de la langue*, six objectifs figurent dans les objectifs fondamentaux communs aux trois voies : *Appliquer des règles d'accord, Appliquer les règles d'orthographe d'usage, Utiliser les signes de ponctuation, Choisir dans le dictionnaire l'information qui correspond au contexte, Recourir efficacement aux ouvrages de référence, Gérer la conjugaison et l'emploi des temps dans toute production*

écrite. Les cinq autres ne sont pas reconnus dans les trois voies comme objectifs fondamentaux d'enseignement en 8^e année.

Tableau 3.2 : Conformité des objectifs pédagogiques évalués à l'ECR de français de 2011 selon le plan d'études de chaque voie en 8^e année

	VSB	VSG	VSO
Lire, comprendre et analyser un texte poétique			
Définir le rôle des personnages			
Analyser l'organisation d'un texte			
Identifier les thèmes d'un texte			
Repérer les caractéristiques d'un texte poétique			
Repérer et analyser des contraintes poétiques			
Produire un message (une lettre) à visée argumentative			
Respecter l'univers propre au genre			
Respecter un point de vue			
Respecter un temps de base donné			
Émettre une opinion étayée par quelques arguments			
Reformuler un argument et contre-argumenter			
Raisonnement sur le fonctionnement de la langue			
Appliquer des règles d'accord			
Appliquer les règles d'orthographe d'usage			
Utiliser les signes de ponctuation			
Choisir dans le dictionnaire l'information qui correspond au contexte			
Gérer la conjugaison et l'emploi des temps dans toute production écrite			
Recourir efficacement aux ouvrages de référence			
Repérer un champ lexical autour d'un thème			
Repérer les diverses possibilités de désigner un personnage ou un objet			
Adapter son registre de langue à la situation de communication			
Repérer et analyser des mots qui participent à l'organisation d'un texte			
Mobiliser les connaissances textuelles, syntaxiques et orthographiques			

Légende :

Objectif fondamental prévu dans le plan d'études de la voie en 8^e année

	Oui
	Non

L'objectif *Repérer un champ lexical autour d'un thème* est prévu en 8^e en VSB, VSG et VSO, mais il n'est pas considéré comme un objectif fondamental. L'objectif *Adapter son registre de langue à la situation de communication* figure parmi les objectifs fondamentaux en VSB et en VSG, mais il doit être évalué en 9^e, pas en 8^e. Il est également prévu en 9^e VSO où il n'est toutefois pas considéré comme un objectif fondamental. L'objectif *Repérer les diverses possibilités de désigner un personnage ou un objet* est prévu en 7^e VSO mais pas comme objectif fondamental ; il n'est prévu ni en VSB ni en VSG. L'objectif *Repérer et analyser des mots ou expressions qui participent à*

l'organisation d'un texte est prévu dans les trois voies, mais en 9^e et non en 8^e. Enfin, l'objectif *Mobiliser les connaissances textuelles, syntaxiques et orthographiques acquises* ne figure pas dans le plan d'études comme un objectif fondamental.

En résumé, près de la moitié des objectifs évalués ne sont pas considérés comme objectifs fondamentaux de 8^e année, dans le plan d'études de chaque voie. Il s'agit plutôt soit d'objectifs fondamentaux de 7^e ou de 9^e année, soit d'objectifs non fondamentaux de 8^e année. Le choix du genre poétique comme trame des activités d'évaluation explique en partie pourquoi l'ECR déborde tant de son référentiel d'évaluation. Le plan d'études prévoit, en effet, peu d'objectifs fondamentaux concernant la compréhension de textes poétiques.

3.1.2 ACTIVITÉS SOUS LA LOUPE

L'ECR de français de 2011 est passée en deux sessions de 90 minutes chacune. La première session commence par une dictée d'un texte court de 96 mots en 30 minutes et se poursuit par la lecture des textes des chansons *Lily* de Pierre Perret et *Djamel* de Michel Bühler. Au terme de la dictée, les élèves écoutent une fois les deux chansons pendant qu'ils en lisent les textes écrits. Ils répondent ensuite à dix-huit questions de difficulté variable, conçues pour évaluer leur compréhension des deux textes lus, ainsi que leur connaissance du fonctionnement de la langue. Tout au long de la première partie, les élèves ont le droit d'utiliser le dictionnaire ainsi que d'autres ouvrages de référence usuels.

La deuxième session est entièrement consacrée à l'écriture d'un texte argumentatif suivant les consignes données.

En somme, l'ECR évalue le niveau d'atteinte des objectifs fondamentaux à l'aide de trois activités principales : la dictée, la compréhension de deux textes poétiques et la production d'un texte argumentatif. Le tableau de correspondance entre les trois activités d'évaluation d'une part, les objectifs généraux et spécifiques évalués d'autre part, constitue une pièce maîtresse du guide de correction fourni aux enseignants. Nous nous y référons afin de savoir dans quelle mesure les niveaux de compétence élevés aussi bien que les niveaux élémentaires, prévus dans le PEV, sont bien représentés dans ces activités.

La dictée

La dictée est destinée à évaluer cinq objectifs fondamentaux du domaine *Structurer*, couramment appelé *Structuration* ou *Fonctionnement de la langue*. La tâche, relativement simple, est conçue pour évaluer le niveau d'acquisition élémentaire de ces objectifs. Le texte est bref, de registre standard, et ne comporte pas de difficultés particulières dans les accords ou dans les homophones, éléments les plus complexes pour les élèves du secondaire 1. Il est dicté une première fois en entier, une deuxième fois lentement avec la ponctuation, en laissant le temps d'écrire, et une troisième et dernière fois avant que les élèves ne puissent corriger leurs copies, en recourant au dictionnaire ou autres ouvrages de référence à leur disposition. A la correction, les points, de 0 à 10, sont attribués en fonction du nombre d'erreurs de ponctuation et d'orthographe lexicale ou grammaticale.

La compréhension de deux textes poétiques

Durant environ 60 minutes, les élèves lisent les deux textes, écoutent l'enregistrement des deux chansons pendant leur lecture, puis complètent un questionnaire d'évaluation des compétences dans les domaines *Lire, comprendre et analyser un texte* (douze questions) et *Structurer* (six questions).

L'objectif *Analyser l'organisation d'un texte*, est évalué par une question consistant à demander aux élèves, à partir des textes des deux chansons, de retrouver les étapes de la vie des héros selon leur ordre d'apparition dans le texte. Cette question est relativement simple puisqu'elle ne fait pas appel à la connaissance des marqueurs temporels et des retours en arrière que ces derniers pourraient signaler. L'exercice ne demande que le repérage de moments de l'histoire à replacer dans l'ordre que le texte propose. Comme cet objectif ne concernait pas les VSO (voir tableau 3.2), le degré de difficulté de la question a pu être ajusté.

- Quatre questions correspondent à l'objectif *Repérer les caractéristiques d'un texte poétique*. Relativement simples, elles consistent à repérer et copier les vers répétés dans les chansons (une question), compter les syllabes (deux questions), identifier et souligner des paires de rimes dans un extrait de sept vers de la chanson *Lily* (une question).
- L'objectif *Repérer et analyser les contraintes poétiques* fait l'objet d'une question de niveau élevé, consistant à demander à l'élève d'expliquer le procédé d'éliision (suppression de la dernière voyelle, du dernier son vocalique d'un mot), utilisé dans deux vers du texte de la chanson *Djamel*, afin de conserver le même nombre de syllabes ainsi que le même rythme.
- L'objectif *Définir le rôle des personnages* est évalué à l'aide de deux questions relativement simples : une question consistant à choisir parmi trois couples de qualificatifs celui qui convient pour décrire le comportement de Djamel illustré par un extrait de quatre vers et une question ouverte à réponse courte consistant à repérer dans les deux textes trois exemples de *boulots* proposés aux immigrés. L'objectif *Identifier les thèmes d'un texte* est évalué à l'aide d'une question simple du type *vrai ou faux* comportant six énoncés.

Huit questions relativement complexes sont destinées à évaluer les objectifs du domaine *Raisonnement sur le fonctionnement de la langue*, à savoir *appliquer les règles d'accord, choisir dans le dictionnaire l'information qui correspond au contexte, adapter son registre de langue à la situation de communication*. Elles consistent à retrouver dans les textes et copier correctement certaines expressions familières correspondant au sens donné (une question), expliquer en peu de mots le sens de certains termes ou expressions utilisés dans les deux textes (quatre questions), compléter des phrases à trou, à l'aide d'une liste de mots de même famille tout en assurant les accords (une question), expliquer le choix du mot *gugusse* pour désigner des manifestants qui luttent pour leurs droits, dans le texte de la chanson *Lily* (une question). Une question nécessitant un niveau élevé de compétences ne se réfère explicitement à aucun des objectifs annoncés aux enseignants mais s'apparente au domaine du fonctionnement de la langue. Elle consiste à identifier parmi trois phrases celle qui contient une intention ironique. Elle correspond à l'objectif *Repérer quelques procédés rhétoriques (illustration, hyperbole, répétition, ironie, litote...)*, prévu en 9^e VSB et en 9^e VSG mais non prévu en VSO.

La lettre argumentative

Le sujet de rédaction est inspiré d'un passage issu d'un des textes lus dans la première partie, celui de la chanson *Lily*, dans lequel les parents d'un jeune homme blanc s'opposent au mariage de leur fils avec la jeune fille noire.

Elle aima un beau blond frisé

Qu'était tout prêt à l'épouser

Mais la belle-famille lui dit nous

Ne sommes pas racistes pour deux sous

Mais on ne veut pas de ça chez nous.

Les élèves doivent écrire une lettre argumentative destinée à convaincre les parents de changer d'attitude. Afin d'aider l'élève à s'engager dans le thème et le genre du texte, le début et la fin de la lettre sont donnés. Comme dans la première partie, les élèves disposent du dictionnaire ainsi que d'autres ouvrages de référence usuels.

Cette lettre doit permettre d'évaluer cinq objectifs du domaine *Ecrire et dire* : *respecter l'univers propre au genre ; respecter un point de vue ; dans une situation polémique et tenant compte du destinataire, émettre une opinion étayée par quelques arguments ; reformuler un argument et contre-argumenter ; respecter un temps de base donné.*

Les critères d'évaluation de ces objectifs sont explicités aux élèves afin de leur faciliter la révision de leur texte :

- La lettre doit comporter au moins trois difficultés du parcours de Lily, et au moins trois qualités de Lily, ainsi que des sentiments de l'auteur de la lettre à l'égard de Lily et de ses parents ;
- Le contenu de la lettre doit être cohérent avec le contexte de la chanson, ainsi qu'avec le début et la fin qui sont imposés.
- Nonobstant le fait que la planification, le début et la fin du texte soient proposés aux élèves, l'activité évalue des niveaux élevés de compétences en lien avec les objectifs visés.

La même activité permet également d'évaluer six objectifs du domaine *Structurer* : *respecter un temps de base donné, appliquer les règles d'accord, appliquer les règles d'orthographe d'usage, mobiliser les connaissances textuelles, syntaxiques et orthographiques acquises, utiliser les signes de ponctuation, adapter son registre de langue à la situation de communication, repérer et analyser des mots ou expressions qui participent à l'organisation d'un texte, repérer un champ lexical autour d'un thème ; conjointement avec trois objectifs du domaine *Préciser et vérifier* : *recourir efficacement aux ouvrages de référence, choisir dans le dictionnaire l'information qui correspond au contexte, gérer l'orthographe dans toute production écrite.* Comme pour le domaine *Ecrire et dire*, l'élève est formellement incité à vérifier que son travail satisfasse aux critères d'évaluation qui lui sont décrits : le présent et le passé composé sont les temps principaux, le texte est organisé par des paragraphes et/ou des organisateurs ; l'orthographe est correcte et les phrases sont construites de manière variée et ponctuées correctement, le vocabulaire est précis, varié et riche ; le texte comporte au moins 200 mots.*

Contrairement à la dictée faisant appel aux compétences de base, le fait de mobiliser en action, dans un processus d'élaboration de ses idées, les connaissances du fonctionnement de la langue correspond à un niveau de compétences élevé.

L'alignement curriculaire de l'ECR se vérifie également à travers la répartition des points entre les quatre domaines d'enseignement du français : *Lire et écouter*, *Ecrire et dire*, *Structurer*, *Préciser et vérifier*. Le premier domaine se voit attribuer 19 points (27% du total de 71 points), le deuxième, 16 points (23%), le troisième et le quatrième domaines réunis, 36 points (50%).

Tableau 3.3 : Spécification des activités d'évaluation en français par objectif et niveau de compétences visés

	Dictée	Compréhension de textes écrits	Production d'un texte écrit
Raisonnement sur le fonctionnement de la langue			
Appliquer des règles d'accord	élémentaire		
Appliquer les règles d'orthographe d'usage	élémentaire		
Utiliser les signes de ponctuation	élémentaire		
Mobiliser les connaissances textuelles, syntaxiques et orthographiques acquises	élémentaire		
Recourir efficacement aux ouvrages de référence	élémentaire		
Lire, comprendre et analyser un texte poétique et narratif			
Définir le rôle des personnages		élémentaire	
Analyser l'organisation d'un texte		élémentaire	
Identifier les thèmes d'un texte		élémentaire	
Repérer les caractéristiques d'un texte poétique		élémentaire	
Repérer et analyser des contraintes poétiques		élémentaire	
Raisonnement sur le fonctionnement de la langue			
Appliquer des règles d'accord		élémentaire	
Adapter son registre de langue à la situation de communication		élémentaire	
Repérer quelques procédés rhétoriques (hyperbole, répétition, ironie, ...)		élémentaire	
Choisir dans le dictionnaire l'information qui correspond au contexte		élémentaire	
Produire un message (une lettre) à visée argumentative			
Émettre une opinion étayée par quelques arguments			élémentaire
Respecter l'univers propre au genre			élémentaire
Respecter un point de vue			élémentaire
Raisonnement sur le fonctionnement de la langue			
Appliquer des règles d'accord			élémentaire
Appliquer les règles d'orthographe d'usage			élémentaire
Utiliser les signes de ponctuation			élémentaire
Mobiliser les connaissances textuelles, syntaxiques et orthographiques acquises			élémentaire
Gérer la conjugaison et l'emploi des temps dans toute production écrite			élémentaire
Respecter un temps de base donné			élémentaire
Adapter son registre de langue à la situation de communication			élémentaire
Recourir efficacement aux ouvrages de référence			élémentaire
Légende :			
	élémentaire		
Niveau de compétences visé par l'activité	élévée		

Les points accordés dans l'épreuve pour chacune de ces trois compétences visées s'élèvent à 19 pour la compréhension (27% de l'attribution totale des points), 16 pour la production écrite (23%) et 36 pour *Raisonnement sur le fonctionnement de la langue* (50%).

Aucune directive sur l'évaluation n'établit de règles précises pour la répartition des points entre les compétences en compréhension, rédaction et fonctionnement de la langue. Avant la réforme Ecole vaudoise en mutation (EVM) en 1996, le français était classé en deux évaluations d'égale valeur : l'expression (y compris la compréhension) et la structuration. L'ECR 2011 s'inscrit dans cette perspective, en accordant la moitié des points à des éléments en lien avec le fonctionnement de la langue, un quart environ pour la compréhension de textes et un dernier quart pour l'expression écrite. Il faut relever que la grammaire est absente de cette évaluation.

En résumé, les activités d'évaluation correspondent aux objectifs annoncés aux enseignants. On note, toutefois, une certaine confusion entre les questions du domaine de la compréhension de l'écrit et celles du domaine du fonctionnement de la langue. Les questions de compréhension du sens de termes ou expressions utilisés dans le texte, par exemple, sont souvent à cheval entre les deux domaines.

L'équilibre entre les niveaux élémentaires et les niveaux élevés de connaissances et de compétences ne semble pas assuré, surtout si on en juge domaine par domaine. Seul le domaine *Structurer* respecte un certain équilibre entre les deux niveaux d'acquisition : le niveau élémentaire évalué par la dictée et les niveaux élevés par les questions de compréhension du vocabulaire utilisé dans les textes écrits et, surtout, la mobilisation de connaissances syntaxiques et orthographiques dans l'activité de production textuelle.

Le domaine *Lire et écouter*, parent pauvre de cette ECR, est évalué principalement par des questions relativement faciles. Les capacités d'interprétation et d'analyse du contenu des textes, prévues dans le plan d'études, sont peu investiguées. Du domaine *Ecrire et dire*, enfin, seules les compétences de niveau élevé sont évaluées à travers l'activité d'écriture d'une lettre argumentative malgré toute l'aide apportée à l'élève notamment au niveau de la planification et de l'organisation du travail d'écriture.

3.2 L'ECR DE MATHÉMATIQUES

3.2.1 OBJECTIFS SOUS LA LOUPE

Les objectifs fondamentaux d'enseignement des mathématiques en 8^e comme en 7^e et en 9^e sont répartis en 6 domaines :

- *Nombres et opérations : résoudre des problèmes numériques*
- *Grandeurs et mesures : mobiliser la mesure pour comparer des grandeurs*
- *Analyse de données : représenter des situations de la vie courante*
- *Calcul littéral : résoudre des problèmes par voie algébrique*
- *Fonctions : modéliser des situations de la vie courante*
- *Géométrie : modéliser le plan et l'espace.*

L'ECR de mathématiques passée en 8^e, en mai 2011, touche à tous les domaines sauf un : *L'Analyse des données*. Elle couvre ainsi un maximum de domaines. Reste à savoir dans

quelle mesure, pour chaque voie, les objectifs évalués sont représentatifs des cinq domaines touchés.

Le domaine *Résoudre des problèmes numériques* se décline en 8 objectifs spécifiques répartis entre deux objectifs généraux : utiliser les nombres et leurs représentations et calculer en situation et porter un regard critique sur le résultat. La tâche *Fractions* évalue un objectif spécifique du deuxième objectif général et non du premier, comme il a été annoncé aux enseignants : additionner, soustraire, multiplier et diviser dans \mathbb{N} , \mathbb{Z} , \mathbb{Q} , \mathbb{R} . Cet objectif spécifique est prévu au programme de 8^e en VSG et de VSO. En VSB en revanche, il est prévu au programme de 7^e. L'évaluation de cet objectif peut être considéré comme légitime dans les trois voies.

Le domaine *Calcul littéral* n'est normalement pas abordé avant la 9^e année en VSO. En VSB et VSG, trois objectifs sont prévus en 8^e (résoudre une équation du premier degré, traduire un langage courant en langage algébrique et viceversa et vérifier les solutions), deux autres en 9^e (résoudre une équation du deuxième degré et résoudre un système d'équations). L'évaluation de l'objectif traduire un langage courant en langage algébrique et viceversa est donc cohérent avec le programme de VSB et de VSG mais pas de VSO ; il n'est du reste pas représentatif de tous les apprentissages réalisés en VSB et VSG dans le domaine du calcul littéral.

Le domaine *Fonctions* (modéliser des situations de la vie courante) recouvre deux objectifs fondamentaux prévus en 8^e dans les trois voies : *Reconnaître et utiliser les propriétés des fonctions affines et représenter les fonctions affines*. L'ECR comporte des tâches d'évaluation des deux objectifs.

Le domaine *Grandeurs et mesures (mobiliser la mesure pour comparer des grandeurs)* recouvre en 8^e année deux objectifs fondamentaux communs aux trois voies : *Mesurer et/ou calculer des volumes de cylindres et utiliser le théorème de Pythagore pour calculer des longueurs, des aires*. En VSO, toutefois, la teneur du deuxième objectif est différente, l'application du théorème de Pythagore se limitant au calcul des aires et des longueurs dans le plan. Un autre objectif est prévu dans le plan d'études mais pas dans la listes des objectifs fondamentaux devant faire l'objet de l'évaluation certificative : *Effectuer des changements d'unités dans les cas usuels (en 8^e : capacité, volume, masse, temps)*. Enfin, un objectif fondamental interdomaines prévu dans le plan d'études est mentionné ici comme faisant partie du domaine *Grandeurs et mesures* : *communiquer ses résultats*.

Le domaine *Géométrie (modéliser le plan et l'espace)* se décline en deux objectifs généraux. Le premier, *Construire des figures géométriques* comporte trois objectifs fondamentaux communs aux trois voies : *Construire des figures planes*, prévu uniquement en 7^e ; *Représenter des solides en perspective et par leur développement*, en 7^e, 8^e et 9^e ; *Reconnaître et utiliser des transformations géométriques (isométries et homothétie)*, prévu en 7^e et en 8^e. Le deuxième objectif général *Recourir au raisonnement déductif pour justifier des propriétés des figures* comprend deux objectifs fondamentaux réservés à la 9^e VSB : *Définir des figures par certaines de leurs propriétés* et *Justifier des propriétés des figures*.

L'ECR s'est affranchi du plan d'études dans le domaine de la géométrie, en évaluant l'objectif *Construire des figures planes* prévu en 7^e dans les trois voies, ainsi que l'objectif *Recourir au raisonnement déductif pour calculer les angles d'une figure*, prévu en 9^e VSB, non prévu en VSO ni en VSG. Un seul objectif du domaine *Géométrie*, commun aux trois voies, à savoir *Représenter des solides en perspective*, est évalué conformément au plan d'études.

Tableau 3.4 : Conformité des objectifs pédagogiques évalués à l'ECR de mathématiques de 2011 selon le plan d'études de chaque voie en 8^e année

	VSB	VSG	VSO
Résoudre des problèmes numériques			
Additionner, soustraire, multiplier et diviser dans N, Z ;Q, R			
Résoudre des problèmes par voie algébrique			
Calculer la valeur numérique d'une expression littérale			
Modéliser des situations de la vie courante			
Représenter une fonction par une représentation graphique			
Reconnaître et utiliser les propriétés des fonctions linéaires et affines			
Mobiliser la mesure pour comparer des grandeurs			
Utiliser les théorèmes métriques pour calculer des longueurs			
Calculer des longueurs, des aires et des volumes			
Effectuer des changements d'unités dans les cas usuels			
Communiquer ses résultats			
Modéliser le plan et l'espace			
Construire des figures géométriques planes			
Représenter des solides en perspective			
Recourir au raisonnement déductif pour calculer les angles d'une figure			

Légende :		Oui
Objectif fondamental prévu dans le plan d'études de la voie en 8 ^e année		Non

En bref, l'échantillon d'objectifs évalués par l'ECR de mathématiques n'a pas la même cohérence avec le plan d'études de chaque voie. En VSB, sur les 11 objectifs évalués, huit sont conformes au programme d'enseignement de 8^e année, deux objectifs sont inscrits au programme de 7^e année et n'apparaissent plus en 8^e, et un objectif est prévu en 9^e année. En VSG, neuf objectifs sont prévus en 8^e, un en 7^e, le onzième en 9^e. En VSO, seulement sept objectifs sur 11 sont prévus en 8^e, un en 7^e, deux en 9^e, le onzième ne figure pas au programme de la voie.

3.2.2 ACTIVITÉS SOUS LA LOUPE

L'ECR 8^e de mathématiques est passée en une session unique de 90 minutes. Comme d'autres ECR de mathématiques passées en 4^e et en 6^e, elle comporte un nombre restreint de problèmes relativement complexes, chacun nécessitant la mobilisation de compétences variées de différents domaines. Cinq problèmes plus ou moins complexes les uns que les autres sont destinés à évaluer les objectifs fondamentaux de l'enseignement des mathématiques en 8^e année. Ces problèmes sont présentés aux élèves dans un cahier de 12 pages comportant également pour chaque problème, de l'espace pour les calculs et les réponses, dans l'ordre : *fractions, périmètre, angles, verre d'eau et chocolat*.

Le problème *Fractions* vise un objectif fondamental du domaine *Nombres et opérations* : *calculer en situation et porter un regard critique sur le résultat*.

Quatre exercices de calcul de nombres fractionnaires, un par opération (addition, soustraction, multiplication et division) sont présentés avec une solution. L'élève doit effectuer les calculs et, en fonction de son résultat, dire si la solution proposée est correcte ou incorrecte. Les points lui sont attribués en fonction à la fois de la solution choisie (*correct* ou *incorrect*) et des calculs effectués. Selon l'opération en jeu, les 4 exercices sont de difficulté variable, mais ils ciblent un même objectif. La conception du problème *Fractions* semble être en décalage avec l'option d'évaluer les compétences visées à l'aide de tâches complexes. Les compétences en résolution de problèmes complexes ne sont donc pas évaluées.

Tableau 3.5 : Spécification des activités de l'ECR de mathématiques par objectif et niveau de compétences visés

	FRACTIONS	PERIMETRE	ANGLES	VERRE D'EAU	CHOCOLAT
Résoudre des problèmes numériques					
Utiliser les nombres et leur représentation					
Résoudre des problèmes par voie algébrique					
Calculer la valeur numérique d'une expression littérale					
Modéliser des situations de la vie courante					
Représenter une fonction par une représentation graphique					
Reconnaître et utiliser les propriétés des fonctions linéaires et affines					
Mobiliser la mesure pour comparer des grandeurs					
Utiliser les théorèmes métriques pour calculer des longueurs					
Calculer des longueurs, des aires et des volumes					
Effectuer des changements d'unités dans les cas usuels					
Communiquer ses résultats					
Modéliser le plan et l'espace					
Construire des figures géométriques planes					
Représenter des solides en perspective					
Recourir au raisonnement déductif pour calculer les angles d'une figure					

Légende :

Niveau de compétences visées par l'activité

	élémentaire
	élevé

Au problème *Périmètre*, cinq croquis (a, b, c, d et e) de formes différentes (un carré, un rectangle, un triangle équilatéral, un polygone dont cinq angles sont droits et un losange) sont présentés à l'élève. Ce dernier doit effectuer quatre tâches différentes sur ces figures. La première tâche consiste à déterminer le périmètre des figures en empruntant la bonne formule de calcul. La deuxième question consiste à calculer les variations du périmètre du carré et du rectangle en fonction de la variation de la mesure d'un côté. La troisième tâche consiste à représenter graphiquement les variations des périmètres du carré et du rectangle calculés dans la deuxième tâche. La quatrième tâche consiste à déterminer, à

partir d'une représentation graphique des variations du périmètre de deux figures quelconques, en fonction de la longueur d'un de leurs côtés, pour quelle valeur du côté le périmètre des deux figures est identique. Il suffit à l'élève de calculer la longueur de l'axe des abscisses correspondant au point de croisement des droites représentant les deux périmètres. La dernière question enfin consiste à calculer le périmètre identique des deux figures représentées dans la quatrième tâche, en projetant le point de croisement sur l'axe des ordonnées.

Problème à tiroirs et relativement complexe, le *Périmètre* est conçu pour évaluer des niveaux de compétences élevés dans trois domaines distincts : *Calcul littéral*, *Grandeurs et mesure* et *Fonctions*.

Le problème *Les Angles* consiste à calculer les angles des triangles représentés sur un croquis. La résolution de ce problème relativement complexe passe par une suite de calculs mettant en jeu une bonne connaissance des propriétés des triangles et des angles, ainsi qu'un bon raisonnement déductif.

Le *Verre d'eau* est un problème classique d'application du théorème de Pythagore aux mesures de longueur. Le problème est complexe pour les élèves de VSB, théoriquement hors de portée pour la majorité des élèves de VSG et de VSO qui aborde un peu ce genre de problèmes, seulement en 9^e.

Le *Chocolat* est un problème également complexe, portant sur deux domaines : *Grandeurs et mesure* d'un côté, *Géométrie* de l'autre. La situation de départ est celle d'un croquis en perspective d'un chocolat ayant la forme d'un prisme ainsi que les mesures de la base du triangle équilatéral et de la longueur du rectangle. La première tâche consiste à calculer la surface totale de ce prisme ; la deuxième, à construire un développement de l'emballage du chocolat à l'échelle 1:2 et la troisième, à dessiner un croquis en perspective d'une boîte d'emballage du chocolat en forme de parallépipède rectangle.

En résumé, les connaissances et compétences de base sont sous-évaluées par l'ECR de mathématiques. Un seul problème est réservé à celles-ci. Les autres problèmes s'adressent aux compétences en résolution de problèmes complexes. Le problème du *Verre d'eau* constitue un cas particulier du non-alignement de certaines parties de l'épreuve au curriculum, dans la mesure où il porte sur des objectifs n'ayant pas fait l'objet de l'enseignement normalement assuré en 8^e VSO.

Dans tous les domaines évalués, on observe une sous-représentation, soit des niveaux élémentaires, soit des niveaux élevés de connaissances et de compétences. Dans le domaine des *Nombres et opérations (résoudre des problèmes numériques)*, seules les connaissances et compétences de base sont évaluées, alors que dans les quatre autres domaines, seules les capacités de résolution de problèmes complexes sont mises à l'épreuve. Ce déséquilibre est évidemment préjudiciable à la validité de l'épreuve. En particulier, les élèves en difficulté, s'agissant de résoudre des problèmes complexes, n'auraient pas l'occasion de montrer leur degré de maîtrise des connaissances et compétences de base.

3.3 SYNTHÈSE

Les épreuves cantonales de référence de 8^e ne sont que partiellement alignées au curriculum. A l'ECR de français, par exemple, près de la moitié des objectifs évalués ne sont pas considérés comme objectifs fondamentaux de 8^e année dans le plan d'études de chaque voie. Une partie d'entre eux sont considérés comme objectifs fondamentaux de

l'enseignement en 7^e ou 9^e mais non en 8^e année, les autres ne sont reconnus comme objectifs fondamentaux dans aucune voie.

Les épreuves cantonales de référence de 8^e sont en plus inégalement alignées au plan d'études de chaque voie. A l'ECR de mathématiques, plus particulièrement, sur 11 objectifs évalués, respectivement huit, neuf et sept sont reconnus comme objectifs fondamentaux en VSB, VSG et VSO.

Les activités d'évaluation qui composent les ECR de 8^e sont presque toutes partiellement alignées dans la mesure où elles représentent en partie des objectifs fondamentaux et en partie des objectifs non considérés comme fondamentaux. Certaines sont toutefois non alignées au référentiel d'évaluation que constitue l'ensemble des objectifs fondamentaux de chaque voie. L'activité de rédaction d'une lettre argumentative, par exemple, est entièrement non alignée au plan d'études dans la mesure où aucun objectif correspondant n'est considéré comme fondamental dans aucune des trois voies d'enseignement pour la 8^e année. Dans l'ECR de mathématiques, le problème *Fractions* n'est pas aligné au plan d'études de VSB, celui des *Angles*, aux plans d'études des trois voies.

Une transgression aussi importante du référentiel conventionnel de l'évaluation certificative que constitue l'ensemble des objectifs fondamentaux de l'enseignement en 8^e année, est sans doute préjudiciable à la validité de contenu de l'épreuve. Elle est par ailleurs une source d'inégalité de traitement entre les voies dans la mesure où certains objectifs évalués ont le statut d'objectifs fondamentaux dans une voie et pas dans une autre. Le plus souvent, à l'ECR de mathématiques surtout, la VSO est la plus nettement désavantagée.

Le fait que les enseignants soient informés, en début d'année, des objectifs évalués à l'ECR ne règle pas le problème. Au contraire, en demandant aux enseignants de préparer les élèves à passer l'ECR en avançant d'une année une partie du programme, ou en renforçant les apprentissages consacrés aux objectifs non fondamentaux, on encourage le *teaching to the test* aux inconvénients si souvent dénoncés. Il n'est du reste pas sûr que les enseignants soient en mesure ou acceptent de changer significativement leur planification de l'enseignement des deux dernières années du secondaire I. L'enquête menée auprès d'une partie de ces enseignants nous éclairera à ce sujet au chapitre 5.

La diversité d'activités composant chaque épreuve permet d'évaluer les compétences de niveau élémentaire aussi bien que celles de niveau élevé. L'équilibre entre les deux ordres hiérarchiques de compétences n'est toutefois pas assuré si on en juge domaine par domaine. A l'ECR de français, l'équilibre entre les compétences de niveau élémentaire et celles de niveau élevé est assuré dans le domaine du fonctionnement de la langue, mais pas dans celui de la compréhension de l'écrit où les compétences d'ordre supérieur ne sont guère examinées ni dans celui de la production écrite où les élèves n'ont pas l'occasion de démontrer leur maîtrise de simples tâches d'écriture. A l'ECR de mathématiques, quatre problèmes sur cinq ne visent que les compétences en résolution de problèmes complexes. Quelques problèmes simples en plus seraient nécessaires afin d'équilibrer les deux ordres de compétences. Le temps de passation de l'épreuve serait en conséquence plus long que les deux périodes de 45 minutes actuelles.

4 LES RÉSULTATS DES ÉLÈVES SELON LES VOIES

Dans ce chapitre, après avoir examiné quelques-unes des qualités éducatives des ECR du point de vue de leur alignement curriculaire, nous tentons d'approcher leurs qualités psychométriques au moyen d'une analyse statistique des résultats des élèves.

De nombreux spécialistes de la mesure et de l'évaluation scolaires (Bain, 2010 ; Carver, 1974 ; De Ketele & Gerard, 2005 ; Laveault & Grégoire, 2002 ; Trahan & Dassa, 1978) s'accordent sur la double dimensionnalité des épreuves externes standardisées : éducatif et psychométrique. Les qualités éducatives sont assurées grâce à un échantillon de tâches suffisamment représentatives de l'univers de référence de l'évaluation en question, ainsi qu'un bon alignement curriculaire et pédagogique des épreuves. Dans le cas de l'évaluation continue des progrès des élèves, les enseignants se contenteraient d'assurer les qualités éducatives requises à leurs outils d'évaluation sans se soucier des qualités psychométriques. Il n'en va pas de même s'agissant de l'évaluation standardisée. Une épreuve externe ne serait pas standardisée sans les qualités psychométriques lui permettant de mesurer les différences existant entre élèves.

Tout en couvrant aussi largement que possible les principaux domaines de l'enseignement assuré, les épreuves d'évaluation scolaire standardisées sont donc conçues de telle sorte que la distribution des résultats s'apparente à la courbe normale (De Ketele & Gerard, 2005 ; Laveault & Grégoire, 2002 ; Trahan & Dassa, 1978).

La double contrainte, éducatif et psychométrique, provoque de nombreuses tensions tout au long du processus d'élaboration des épreuves standardisées. Bain (2010) évoque à ce sujet un malentendu docimologique entre d'une part les enseignants chargés d'élaborer les épreuves, et d'autre part les spécialistes de l'évaluation (docimologues, statisticiens) qui supervisent leurs travaux. Les enseignants ne comprennent pas lorsque les docimologues éliminent de l'épreuve les questions réussies par une grande majorité d'élèves, au motif qu'elles sont peu discriminatives ou n'ont pas les caractéristiques statistiques requises. Pour eux, ce genre de questions qui attestent notamment de l'efficacité de l'enseignement ont leur place dans une épreuve standardisée. Pour les spécialistes de l'évaluation, au contraire, il faut éviter les questions trop faciles qui limitent la validité et la fidélité de l'épreuve standardisée.

Les résultats de tous les élèves de 8^e année ayant passé les ECR de 2011, ont été mis à notre disposition par la Direction générale de l'enseignement obligatoire. Il s'agit, pour chaque élève, du nombre de points obtenus à chaque domaine de compétence évalué, soit le score partiel, ainsi que du score total équivalent à la somme des scores partiels. A l'ECR de français, six scores partiels sont attribués respectivement à la dictée, la compréhension 1, la compréhension 2, le fonctionnement de la langue (FDL) en lecture, le fonctionnement de la langue (FDL) en production textuelle, et la production textuelle, sont attribués. A l'ECR de mathématiques, quatre scores partiels sont attribués respectivement aux domaines *Résolution de problèmes*, *Fonctions*, *Grandeurs et mesure*, et *Géométrie*.

Le nombre d'élèves ayant passé chaque ECR est variable du fait des absences le jour de la passation. Il s'élève à 2183 en VSO, 2455 en VSG et 2707 en VSB, soit 7345 au total, concernant l'ECR de français ; à 2189 en VSO, 2464 en VSG et 2712 en VSB, soit 7365 au total, s'agissant de l'ECR de mathématiques.

L'analyse de ces données a pour but non pas d'estimer le niveau d'acquisition des élèves de chaque voie, mais de questionner le degré de validité et de fiabilité des résultats aux épreuves. Une analyse comparée de la distribution des résultats dans chaque voie nous

permettra de savoir si la validité de l'ECR est équivalente ou différente pour les trois voies. L'hypothèse principale de notre étude est en effet celle d'un biais de validité et de fidélité des mesures assurées par les ECR de 8^e année dans la mesure où elles ne sont pas différenciées, comme l'enseignement, en fonction des voies.

Dans cette analyse, l'ampleur des écarts entre les moyennes des résultats des trois voies constitue un préalable, mais comme le précisent Laveault & Grégoire (2002, p. 193), *l'existence d'une différence entre les moyennes des scores de deux groupes de la population n'est pas en soi la preuve de l'existence d'un biais*. Il est normal que les élèves de la voie à exigences plus élevées réussissent mieux le même test que ceux de la voie à exigences moins élevées. Nous vérifierons l'hypothèse d'un biais de validité des ECR, en comparant les résultats des élèves des trois voies à l'aide des indicateurs ci-après.

Deux indicateurs de la dispersion des résultats :

- L'étendue de la distribution, soit la différence entre le score minimum et le score maximum obtenu par les élèves de la voie ;
- Le coefficient de variation, soit le rapport de l'écart-type à la moyenne des résultats, qui mesure le degré de variabilité des résultats autour de la moyenne.

Deux indicateurs de la forme de la distribution :

- Le coefficient de symétrie (*skewness*) qui renseigne sur le degré d'asymétrie à gauche ou à droite de la distribution ;
- Le coefficient d'aplatissement (*kurtosis*) qui renseigne sur le degré de concentration des résultats aux extrémités de la distribution.

Deux indicateurs de la cohérence entre les différentes parties du test :

- Les coefficients de corrélation entre les scores partiels et le score total, une mesure du degré de cohérence entre chaque partie et tout le test ;
- Le coefficient alpha de Cronbach qui est une mesure de la cohérence interne, de la consistance interne en d'autres termes ; variant entre 0 et 1 ; considéré comme satisfaisant à partir de 0.7.

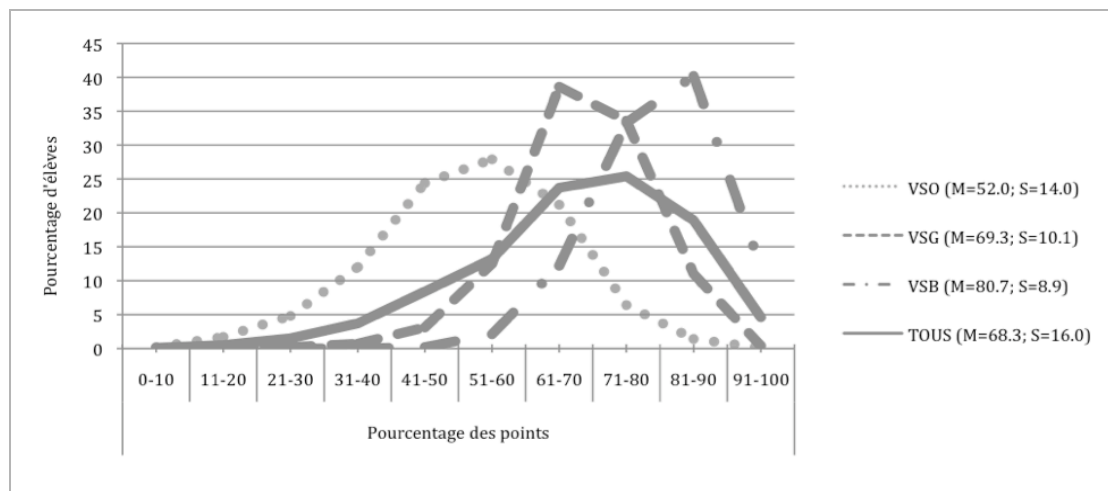
Afin de faciliter les comparaisons des statistiques descriptives ci-dessus des résultats obtenus dans les trois voies, les scores individuels sont convertis en pourcentage du score maximum possible de l'épreuve. A cet effet, le ratio du nombre de points obtenus sur le nombre maximum possible de points a été multiplié par 100.

4.1 RÉSULTATS À L'ECR DE FRANÇAIS

L'analyse de la variance, $F(2, 7342) = 4087.14$, $P = .00$, $\eta^2 = .527$, met en évidence une différence significative entre les moyennes des résultats des trois voies. L'effet de la voie sur la variance des résultats, mesuré à l'aide d'êta carré (η^2) est considérable, puisqu'il explique plus de 50% de la variance des résultats entre les élèves.

Les variances de la distribution des résultats dans les trois voies ne sont pas du tout homogènes, ainsi qu'en attestent les résultats du test de Levene : $F(2, 7342) = 235.843$; $p = .00$. La dispersion des résultats est beaucoup plus faible en VSB et en VSG qu'en VSO. Le coefficient de variation varie du simple au double de la VSB ou la VSG à la VSO.

Figure 4.1 : Distribution des résultats à l'ECR de français selon les voies



Au niveau des trois voies ainsi que de toute la population, les courbes des résultats à l'ECR de français sont asymétriques à droite, mais l'asymétrie est moins prononcée en VSO que dans les deux autres voies. La VSO, avec une moyenne plus près du centre et de la médiane, une légère asymétrie à droite et une moindre concentration des résultats aux extrémités, présente une distribution des résultats plus proche de la distribution normale.

Tableau 4.1 : Statistiques descriptives de la distribution des résultats à l'ECR de français

	VSO (N=2183)	VSG (N=2455)	VSB (N=2707)	TOUS (N=7345)
Etendue	92.96	70.42	54.93	100.00
Médiane	52.11	70.42	81.69	70.42
Moyenne	51.99	69.25	80.66	68.32
Ecart-type	14.02	10.10	8.93	16.05
Coefficient de variation	0.27	0.15	0.11	0.23
Coefficient d'asymétrie (skewness)	-.24	-.52	-.54	-.68
Erreur standard d'asymétrie	.05	.05	.05	.03
Coefficient d'aplatissement (kurtosis)	.05	.61	.20	.17
Erreur standard d'aplatissement	.11	.10	.09	.06

En VSG et en VSB, l'étendue de la distribution est limitée respectivement aux deux tiers et à la moitié de l'échelle des points. La moitié des élèves ont obtenu plus de 70% des points en VSG (le score médian équivalant à 70.42% des points), plus de 81% des points en VSB (le score médian étant de 81.69% des points).

Le faisceau d'indices de la cohérence interne présenté au tableau 4.2 renforce l'impression que les qualités éducatives de l'ECR varient d'une voie à l'autre.

Tableau 4.2 : Coefficients de corrélation entre les scores partiels et le score total et coefficient alpha de Cronbach à l'ECR de français

		VSO	VSG	VS	TOUS
Corrélation r de Bravais-Pearson*	Score en dictée - score total	.70	.62	.54	.81
	Score de compréhension 1 - score total	.61	.55	.57	.73
	Score de compréhension 2 - score total	.55	.44	.44	.66
	Score de FDL en lecture - score total	.60	.56	.61	.72
	Score de FDL en production textuelle - score total	.67	.59	.68	.71
	Score de production textuelle - score total	.78	.72	.73	.84
Coefficient de consistance interne (Alpha Cronbach) entre les six scores partiels		.72	.60	.64	.84

* Toutes les corrélations entre les scores partiels et le score total sont significatives à .01.

Au niveau de tous les élèves des trois voies, les corrélations entre les scores partiels et le score total sont élevées. Elles sont supérieures à 0.7, à l'exception de la corrélation entre le score total et le score en compréhension 2 qui, pour rappel, porte sur les caractéristiques du genre poétique. Le coefficient alpha de Cronbach qui s'élève à 0.84 confirme cette cohérence interne de l'épreuve au niveau de tous les élèves.

La consistance interne de l'ECR de français reste satisfaisante au niveau de la VSO, alors qu'elle se fissure au niveau de la VSG et de la VS. En VSO, à l'exception de la compréhension 2, toutes les parties sont bien corrélées avec le tout et le coefficient alpha de Cronbach est satisfaisant. En VSG et en VS, trois parties sur 6 ne sont que moyennement corrélées avec le tout et le coefficient alpha de Cronbach n'est guère satisfaisant.

L'ECR de français, relativement bien adaptée au niveau des élèves de la VSG, s'avère plutôt difficile pour les élèves de la VSO et assez facile pour ceux de la VS.

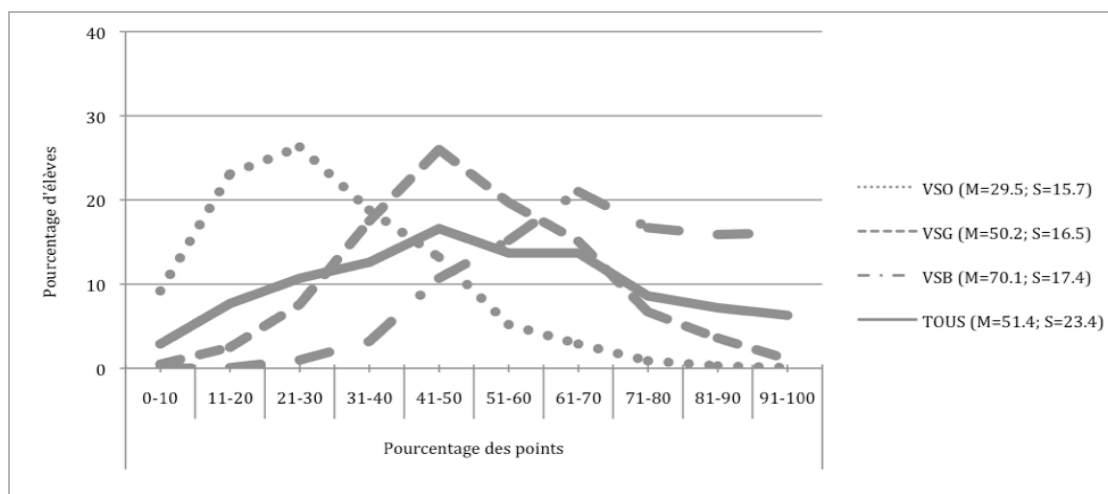
4.2 RÉSULTATS À L'ECR DE MATHÉMATIQUES

Les écarts de moyenne entre les voies sont très importants à l'ECR de mathématiques : 21% des points entre la VSG et la VSO, 20% entre la VS et la VSG, 41% entre la VS et la VSO. L'analyse de la variance confirme ces différences : $F(2, 7362) = 3633.52$, $P = .00$, $\eta^2 = .497$. L'effet de la voie sur la variance des résultats, mesuré à l'aide d'éta carré (η^2) explique 50% de la variance des résultats.

Les variances des résultats ne sont pas du tout homogènes, ainsi qu'en attestent les résultats du test de Levene : $F(2, 7362) = 31.884$; $p = .00$. La forme de la distribution des résultats est tout à fait différente d'une voie à l'autre.

Le contraste entre les résultats des élèves de VSO et ceux de VS domine le tableau des indices de tendance centrale et de dispersion (tableau 4.3). Asymétrie à gauche d'un côté, à droite de l'autre ; concentration des résultats à l'extrémité de droite d'un côté, de gauche de l'autre. La majorité des élèves de VSO obtient de faibles résultats alors que la majorité de VS a réalisé des scores proches du maximum. L'ECR de mathématiques n'est pas assez sensible aux différences individuelles, ni en VSO ni en VS. Les compétences des élèves sont sous-estimées dans un cas, alors qu'elles sont surestimées dans l'autre cas.

Figure 4.2 : Distribution des résultats à l'ECR de mathématiques selon les voies



La courbe des résultats de l'ensemble des élèves, anormalement aplatie, avec presque autant d'élèves au centre qu'aux extrémités de la distribution s'explique par le déséquilibre de la distribution des résultats en VSO et en VSB. La distribution des résultats est plus proche de la normale en VSG qu'en VSO et en VSB. L'épreuve semble avoir été conçue en fonction du niveau des élèves de cette voie.

Tableau 4.3 : Statistiques descriptives de la distribution des résultats à l'ECR de mathématiques

	VSO (N=2189)	VSG (N=2464)	VS (N=2712)	TOUS (N=7365)
Etendue	100.00	100.00	79.63	100.00
Médiane	27.78	50.00	70.37	50.00
Moyenne	29.48	50.22	70.13	51.39
Ecart-type	15.75	16.47	17.40	23.42
Coefficient de variation	0.53	0.33	0.25	0.46
Coefficient d'asymétrie (skewness)	.71	.22	-.25	.08
Erreur standard d'asymétrie	.052	.049	.047	.029
Coefficient d'aplatissement (kurtosis)	.70	-.21	-.69	-0.78
Erreur standard d'aplatissement	.105	.099	.094	.057

Tableau 4.4 : Coefficients de corrélation item-test et coefficient alpha de Cronbach à l'ECR de mathématiques

		VSO	VSG	VS	TOUS
Corrélation r de Bravais-Pearson*	Résolution de problèmes	.77	.66	.57	.78
	Fonctions	.68	.62	.68	.78
	Grandeurs et mesures	.80	.79	.89	.88
	Géométrie	.78	.72	.68	.84
Coefficient de consistance interne (Alpha Cronbach) entre les quatre scores		.75	.55	.64	.82

* Toutes les corrélations entre le score partiel et le score total sont significatives à .01.

Du point de vue de l'ensemble des élèves, avec une corrélation partie-tout supérieure à 0.78 et un coefficient alpha de 0.82, l'épreuve affiche une consistance interne satisfaisante. Entre les voies, en revanche, la consistance interne est inégale : satisfaisante en VSO, insuffisante en VSG et en VSB.

4.3 SYNTHÈSE

Différents indices statistiques de la distribution des résultats dans les trois voies, le coefficient de variation et le coefficient alpha de Cronbach, entre autres, témoignent de la validité différentielle limitée des ECR de 8^e année.

Ces épreuves qui semblent avoir été conçues en fonction du niveau des élèves de la voie secondaire générale (VSG), à exigences moyennes, donnent une mesure biaisée des compétences des élèves de la voie secondaire à options (VSO) et de ceux de la voie secondaire de baccalauréat (VSB) qu'elle tendent à sous-estimer dans le premier cas, à surestimer dans le deuxième.

5 VALIDITÉ ET UTILITÉ PERCUES PAR LES ENSEIGNANTS

5.1 LE QUESTIONNAIRE

Afin de renseigner sur la validité de contenu curriculaire et pédagogique des épreuves standardisées, une démarche courante consiste à interroger les enseignants qui non seulement organisent et gèrent les apprentissages en question dans ces évaluations mais aussi, la plupart des fois, sont chargés d'administrer et de corriger celles-ci (Antonietti & al., 2005 ; Ntamakiliro & Tessaro, 2002 ; Hardy, 1984). C'est à cet effet que nous avons réalisé une enquête par questionnaire auprès des enseignants de français et de mathématiques en 8^e année, en VSO, VSG ou VSB.

Le questionnaire a été envoyé par la poste le 16 mai 2011, peu après que les enseignants avaient terminé la correction des copies d'élèves. L'ECR de français a été passée les 3 et 10 mai, celle de mathématiques, le 5 mai 2011.

Le questionnaire aux enseignants est conçu en deux versions destinées aux enseignants de français et de mathématiques respectivement. Les deux versions contiennent les mêmes questions, dans le même ordre, la seule différence étant la référence à l'ECR de français pour les uns, à l'ECR de mathématiques pour les autres.

En tout 18 questions sont posées. Les cinq premières sont destinées à l'identification des caractéristiques personnelles de l'enseignant (genre, âge, nombre d'années d'expérience professionnelle, type et niveau de formation, voie dans laquelle il enseigne). Les 13 questions suivantes servent à évaluer sa perception de l'ECR de français ou de mathématiques selon plusieurs points de vue, détaillés ci-après.

Le degré d'alignement curriculaire de l'ECR

Une question destinée à évaluer le degré d'alignement perçu de l'ECR au plan d'études est posée en ces termes :

Estimez-vous que les éléments suivants de l'ECR 8^e de français / mathématiques passée en mai 2011 correspondent aux objectifs fondamentaux d'enseignement définis dans le plan d'études vaudois pour la voie concernée ?

Pour l'ECR de français, les éléments en question sont les cinq parties de l'épreuve : *la dictée, les questions de compréhension des deux chansons (compréhension 1), les questions relatives aux caractéristiques du texte de chanson (compréhension 2), les exercices de vocabulaire (vocabulaire) et la rédaction d'une lettre pour convaincre (rédaction)*. Pour l'ECR de mathématiques, les éléments en question sont les cinq problèmes à résoudre qui constituent l'épreuve : *fractions, périmètres, angles, verre d'eau et chocolat*. L'échelle d'évaluation de type Likert comporte quatre niveaux (*tout à fait, un peu, pas tellement et pas du tout*) ainsi qu'une case *sans opinion* pour les enseignants qui n'ont pas ou ne veulent pas donner d'avis.

Le degré d'alignement pédagogique de l'ECR

Une question destinée à évaluer le degré d'alignement perçu de l'ECR à l'enseignement effectivement assuré est posée en ces termes :

Estimez-vous que les éléments suivants de l'ECR 8^e de français / mathématiques passée en mai correspondent à l'enseignement dont ont effectivement bénéficié vos élèves ?

La même liste des éléments et la même échelle de Likert qu'en ce qui concerne le degré d'alignement curriculaire sont appliquées à cette question.

Le degré de difficulté de l'ECR

Une question relative au degré de difficulté perçu de l'ECR est posée en ces termes :

Quel est selon vous, le degré de difficulté des diverses parties de l'ECR 8e de français / mathématiques de mai 2011 pour la majorité de vos élèves ?

La liste des parties de l'ECR est la même qu'à la question précédente. L'échelle de Likert appliquée à cette question directe comprend quatre niveaux (*très facile, plutôt facile, plutôt difficile, très difficile*) ainsi qu'une case *sans opinion*.

La prise en compte de l'ECR dans la planification de l'enseignement

Une question porte sur la communication, au début de l'année scolaire, par la DGEO, des objectifs d'apprentissage évalués par l'ECR. Le but de cette communication est de permettre aux enseignants de traiter au préalable les contenus évalués à l'ECR. L'annonce des objectifs évalués serait préjudiciable à la qualité de l'enseignement si les objectifs évalués aux ECR étaient privilégiés au détriment d'autres objectifs importants du plan d'études, entraînant ce qu'il est convenu d'appeler un rétrécissement du curriculum enseigné (Blais, 2005 ; Ntamakiliro & Tessaro, 2010). Il n'en reste pas moins vrai que l'annonce en début d'année des objectifs évalués aux épreuves externes s'impose dans la mesure où la planification de l'enseignement varie d'un enseignant à l'autre et d'une voie à l'autre.

La question à laquelle les enseignants pouvaient répondre par *oui* ou *non*, est la suivante :

Les objectifs fondamentaux et les genres textuels évalués à l'ECR 8e de français de mai 2011 / les objectifs fondamentaux évalués à l'ECR 8e de mathématiques de mai 2011, ont été annoncés aux enseignants en début d'année. Ces informations ont-elles influencé votre planification de l'enseignement ?

Le degré d'adhésion aux modalités d'évaluation caractéristiques de l'ECR

Une modalité d'évaluation caractéristique de l'ECR de français consiste à évaluer le niveau de connaissance du fonctionnement de la langue (orthographe, vocabulaire, grammaire, conjugaison) à travers les activités de compréhension de textes écrits d'une part, la production de textes écrits d'autre part. Deux questions apparentées ont été posées aux enseignants.

- *Dans cette ECR, les exercices de structuration (vocabulaire) sont liés à la compréhension des textes lus. Quel est votre degré d'adhésion à cette modalité d'évaluation des connaissances et compétences en structuration ?*
- *Dans cette ECR, les connaissances et compétences en structuration sont en partie évaluées par le biais de l'activité de production textuelle. Quel est votre degré d'adhésion à cette modalité d'évaluation des connaissances et compétences en structuration ?*

En mathématiques, une modalité d'évaluation propre à l'ECR consiste à soumettre aux élèves des problèmes relativement complexes, permettant d'évaluer plusieurs compétences à la fois. Cette approche diffère de l'approche traditionnelle selon laquelle à chaque problème correspond une compétence spécifique. Une autre modalité consiste à

attribuer plus ou moins de points, selon les cas à la solution trouvée, qu'aux procédures et démarches de résolution du problème. Les deux questions ont été posées en ces termes :

- *Dans cette ECR, un même problème peut évaluer plusieurs compétences visées. Quel est votre degré d'adhésion à cette modalité d'évaluation des connaissances et compétences en mathématiques de vos élèves ?*
- *Dans cette ECR, les points sont attribués autant à la solution trouvée qu'aux procédures et démarches de résolution du problème. Quel est votre degré d'adhésion à cette modalité d'évaluation ?*

Les enseignants ont pu exprimer leur position sur une échelle de Likert à 4 niveaux (*tout à fait / un peu / pas tellement / pas du tout d'accord*).

Le degré de cohérence entre l'ECR et les travaux significatifs

Une question destinée à sonder le degré de cohérence perçu par les enseignants entre l'ECR et les travaux dits significatifs, c'est-à-dire leurs propres épreuves d'évaluation certificative des élèves est posée en ces termes :

- *Les activités d'évaluation proposées dans l'ECR 8^e de français / mathématiques de mai 2011 correspondent-elles aux travaux significatifs proposés à vos élèves tout au long de l'année ?*

Le degré de cohérence entre l'ECR et les épreuves communes d'établissement

Le Règlement d'application de la loi scolaire de 1984 prévoit l'organisation, au sein des établissements scolaires, d'épreuves communes aux différentes classes et différentes voies. Les chefs de file de disciplines sont notamment chargés de mettre en œuvre cette disposition réglementaire. La pratique de ces épreuves communes varie, cependant, très largement d'un établissement à l'autre. C'est pourquoi il fallait se renseigner sur leurs pratiques d'épreuves communes avant d'interroger les enseignants sur la cohérence perçue entre l'ECR et ces épreuves. Trois séries de questions leur ont été posées en ce sens.

- *Avez-vous cette année scolaire mis en place une ou plusieurs épreuves communes à plusieurs classes (oui / non) ?*
- *(Si oui), ces épreuves communes ont-elles été passées dans des classes d'une même voie ou de voies différentes (classes d'une même voie ; classes de 2 voies différentes, VSB + VSG, VSG + VSO, VSB + VSO ; classes de 3 voies différentes) ?*
- *(Si oui), trouvez-vous que, en comparaison, l'ECR 8^e de français / mathématiques est plus facile ou plus difficile que ces épreuves communes ?*

Le degré d'utilité perçue de l'ECR

Le degré d'utilité perçue de l'ECR est examiné à l'aide de deux questions :

- *L'ECR 8^e de français / mathématiques contribue-t-elle à l'harmonisation des pratiques évaluatives des enseignants ? (réponse oui ou non).*
- *Est-ce que vous prenez en compte, d'une façon ou d'une autre, les résultats à l'ECR 8^e de français / mathématiques dans le cadre de l'évaluation de vos élèves ? (échelle : tout à fait / un peu / pas tellement / pas du tout).*

Modalités de différenciation de l'ECR en fonction des voies

Enfin, les enseignants ont été invités à donner leur avis sur les trois conceptions différentes des épreuves externes tenant compte des différences de niveau entre les trois voies.

- *L'ECR 8^e de français / mathématiques doit être la même pour les trois voies comme c'est le cas actuellement.*
- *Il devrait y avoir trois ECR 8^e de français / mathématiques différentes, respectivement pour la VSO, la VSG et la VSB.*
- *L'ECR 8^e de français / mathématiques devrait comprendre une partie commune aux trois voies et une partie spécifique à chaque voie.*

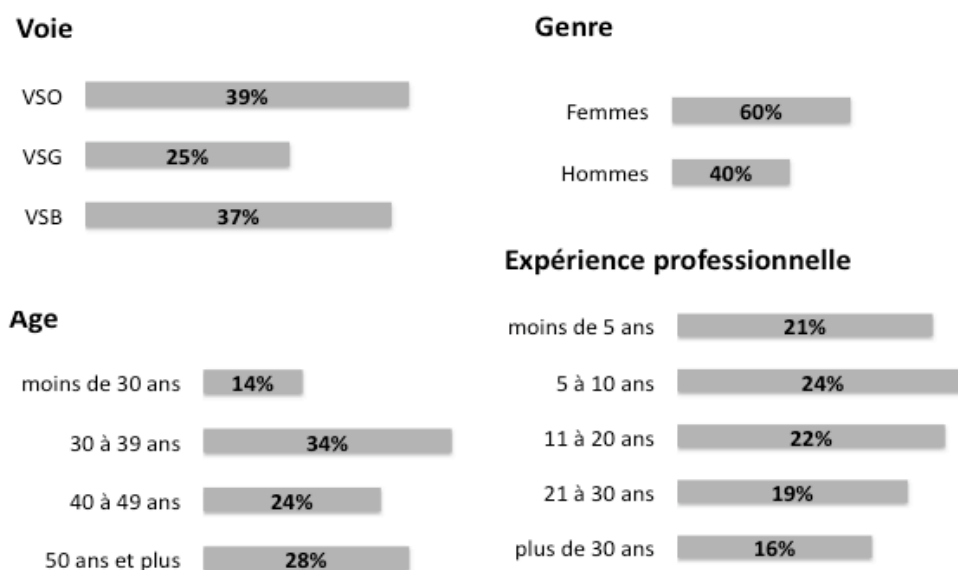
5.2 LES ENSEIGNANTS INTERROGÉS

Les enseignants de français et de mathématiques dans les classes de 8^e année de 23 sur les 24 établissements secondaires ont tous reçu un questionnaire en format papier, par l'intermédiaire de leurs directeurs d'établissement. Nous n'avons pas envoyé de questionnaire aux enseignants d'un établissement secondaire que nous avons sollicité au même moment en vue d'avoir accès aux copies d'élèves.

La lettre d'accompagnement du questionnaire précise les objectifs de l'étude ainsi que les modalités de réponse au questionnaire. Le questionnaire, totalement anonyme, pouvait être complété sur papier ou en ligne via Internet, suivant le lien indiqué.

Le questionnaire de français a été envoyé à 199 enseignants : 74 en VSO, 49 en VSG, 69 en VSB, et 7 enseignants dans des classes relevant de deux voies différentes. En retour, le questionnaire papier a été complété par 82 enseignants, le questionnaire en ligne par 21 enseignants. Au total, nous avons récolté les réponses de 103 enseignants, soit un taux de retour de 52%.

Figure 5.1 : Caractéristiques de l'échantillon d'enseignants de français (N=98)

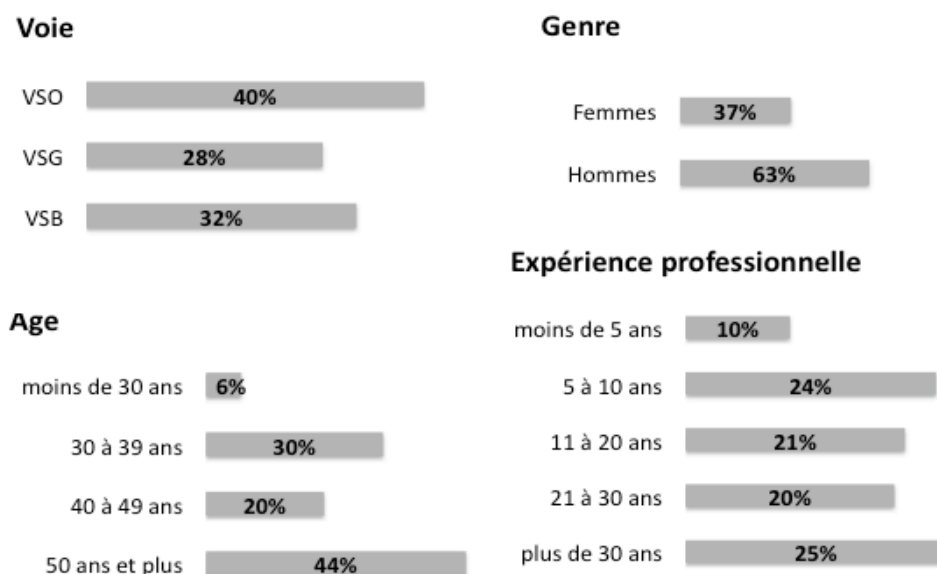


Le questionnaire de mathématiques a été envoyé à 182 enseignants : 72 enseignants en VSO, 44 en VSG, 55 en VSB, et 11 enseignants dans des classes relevant de deux voies différentes. En retour, le questionnaire papier a été complété par 61 enseignants, le questionnaire en ligne par 30 enseignants. Au total, nous avons récolté les réponses de 91 enseignants, soit un taux de retour satisfaisant de 50%.

Un des objectifs principaux de cette enquête étant de mettre en évidence les différences de perception de l'ECR selon la voie, nous avons demandé à ceux qui enseignent dans deux voies distinctes d'en choisir une qui servirait de référence dans leurs réponses. Les répondants concernés, cinq enseignants de français et deux enseignants de mathématiques, qui n'ont pas donné cette indication, n'ont pas été pris en compte dans l'analyse des données. Au final, l'analyse des données a porté sur 98 enseignants de français et 89 enseignants de mathématiques, dont les caractéristiques sont résumées dans les figures 5.1 et 5.2.

Le taux de retour des questionnaires aux enseignants de français dans chaque voie est satisfaisant : 38 enseignants sur 74 (51%) en VSO, 24 sur 49 (49%) en VSG et 36 sur 69 (52%) en VSB.

Figure 5.2 : Caractéristiques de l'échantillon d'enseignants de mathématiques (N=89)



L'échantillon des 98 enseignants de français est composé en majorité de femmes (60%), à l'instar de la population de référence (les enseignants de français en 8^e année dans les établissements secondaires) qui en compte 62%. 84% d'entre eux ont plus de 30 ans ; 79% ont plus de 5 ans d'expérience professionnelle ; 39% enseignent en VSO, 24% en VSG, 37% en VSB.

Le taux de retour des questionnaires aux enseignants de mathématiques dans chaque voie est également satisfaisant : 36 enseignants sur 72 (50%) en VSO, 25 sur 44 (57%) en VSG, 28 sur 55 (51%) en VSB.

L'échantillon des 89 enseignants de mathématiques comprend une majorité d'hommes, à peu près dans les mêmes proportions que la population de référence, celle des enseignants de mathématiques en 8^e année dans les établissements secondaires, soit 63% et 65%,

respectivement. 94% d'entre eux ont plus de 30 ans ; 90% ont plus de 5 ans d'expérience professionnelle ; 40% enseignent en VSO, 28% en VSG, 32% en VSB.

5.3 ANALYSE DES RÉPONSES

Les réponses aux différentes questions ont fait l'objet d'une analyse statistique qui a permis de mettre en évidence le point de vue majoritaire de tous les enseignants interrogés d'abord, les différences dans leur perception des ECR selon la voie dans laquelle ils enseignent, ensuite. L'analyse des différences entre les enseignants des trois voies contribuera à la vérification de l'hypothèse principale de cette étude, celle d'une validité différentielle insatisfaisante des ECR. Autrement dit, le degré d'alignement curriculaire et pédagogique perçu de l'ECR, son degré de difficulté perçu, son degré de cohérence avec les pratiques d'évaluation en classe, son degré d'utilité perçu varient systématiquement selon la voie dans laquelle on enseigne.

Nous avons emprunté les procédures d'analyse de variance multiple (MANOVA – *multivariate analysis of variance*) afin de comparer les enseignants des trois voies au niveau du degré d'alignement curriculaire, du degré d'alignement pédagogique et du degré de difficulté perçus. Au niveau des autres questions, nous nous sommes basé sur le test du Chi² pour analyser les différences de perceptions entre les enseignants des trois voies.

5.4 VALIDITÉ ET UTILITÉ PERÇUES DE L'ECR DE FRANÇAIS

5.4.1 DEGRÉ D'ALIGNEMENT CURRICULAIRE

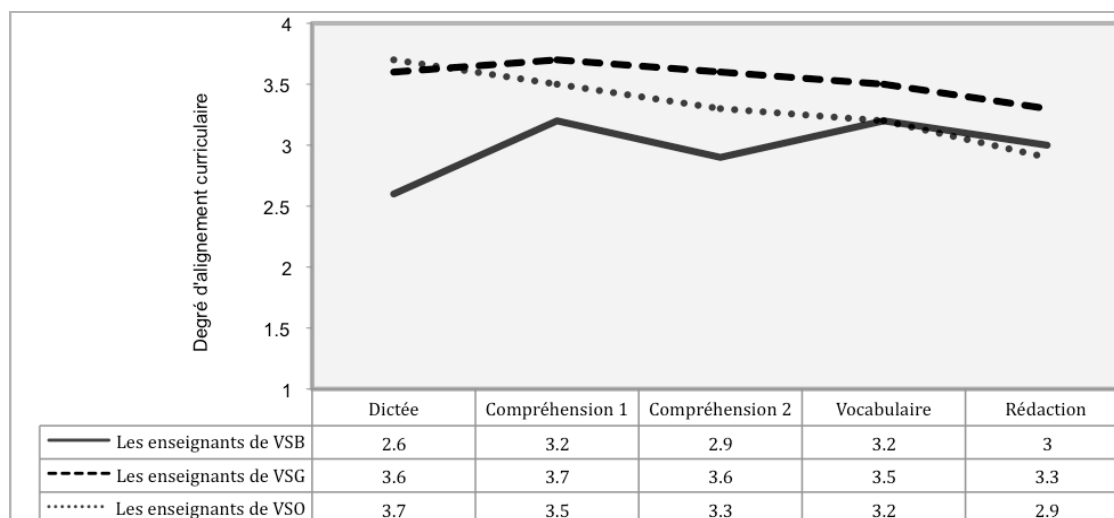
Pour la majorité des enseignants des trois voies, l'alignement curriculaire de l'ECR de français ne fait aucun doute. Sur une échelle de 1 à 4, le degré moyen d'alignement curriculaire perçu des différentes parties de l'épreuve varie entre 3.1 et 3.4. Toutefois, le degré d'alignement curriculaire est apprécié différemment selon la voie dans laquelle on enseigne. Comme il apparaît sur la figure 5.3, pour les enseignants de VSG, toutes les parties de l'ECR sont créditées d'un degré élevé d'alignement curriculaire, entre 3.3 et 3.7 sur 4. A l'opposé, pour les enseignants de VSB, toutes les parties de l'ECR sont créditées d'un degré d'alignement curriculaire plutôt moyen, entre 2.6 et 3.2 sur 4. Quant aux enseignants de VSO, leur perception du degré d'alignement curriculaire varie d'une partie à l'autre de l'épreuve. La dictée et la compréhension 1 sont créditées d'un degré élevé d'alignement curriculaire alors que la compréhension 2, le vocabulaire et la rédaction sont perçus comme moyennement alignés sur le plan d'études.

Les différences de perception du degré d'alignement curriculaire des différentes parties de l'ECR de français selon la voie dans laquelle on enseigne sont statistiquement significatives, d'après les résultats de l'analyse de variance multivariée : lambda de Wilks = .68, $F(10,162) = 3.41$, $p < .001$. C'est au niveau de la dictée que les différences d'appréciation sont les plus fortes. Cette partie de l'épreuve ne serait pas assez en phase avec le plan d'études pour les enseignants de VSB alors qu'elle est créditée d'un degré élevé d'alignement curriculaire par les enseignants de VSG et ceux de VSO.

Les différences de perception du degré de validité curriculaire de la compréhension 1 et de la compréhension 2 sont statistiquement significatives entre la VSB et la VSG, mais pas entre la VSB et la VSO, ni entre la VSG et la VSO. Au niveau du vocabulaire ainsi que de la rédaction, les différences entre les trois catégories d'enseignants ne sont pas significatives.

Dans leurs commentaires, quelques enseignants disent que la dictée est *trop facile*. Elle est *indigne de mes élèves*, dit un d'entre eux.

Figure 5.3 : Degré moyen d'alignement curriculaire perçu des différentes parties de l'ECR de français par les enseignants de VSO, VSG et VSB



Au sujet de la première partie des questions de compréhension (compréhension 1), quelques remarques traitent de la difficulté des textes des chansons et de celle des consignes et de leur formulation. Un répondant s'étonne que l'on n'exige pas plus de connaissances en versification, il juge le programme flou et trop faible dans ses contenus.

Au sujet de la rédaction, plusieurs enseignants, dans les trois voies, mettent en évidence le fait que l'argumentation se travaille en 9^e, en raison notamment d'une meilleure maturité des élèves pour aborder ce genre de textes. Un répondant souligne que le texte informatif serait plus judicieux pour des élèves de 8^e. Une personne s'inquiète des critères de corrections qui ne permettent pas de valoriser des idées intéressantes qui deviennent de fait hors critères. Deux autres répondants trouvent que l'évaluation de la production écrite est trop peu exigeante.

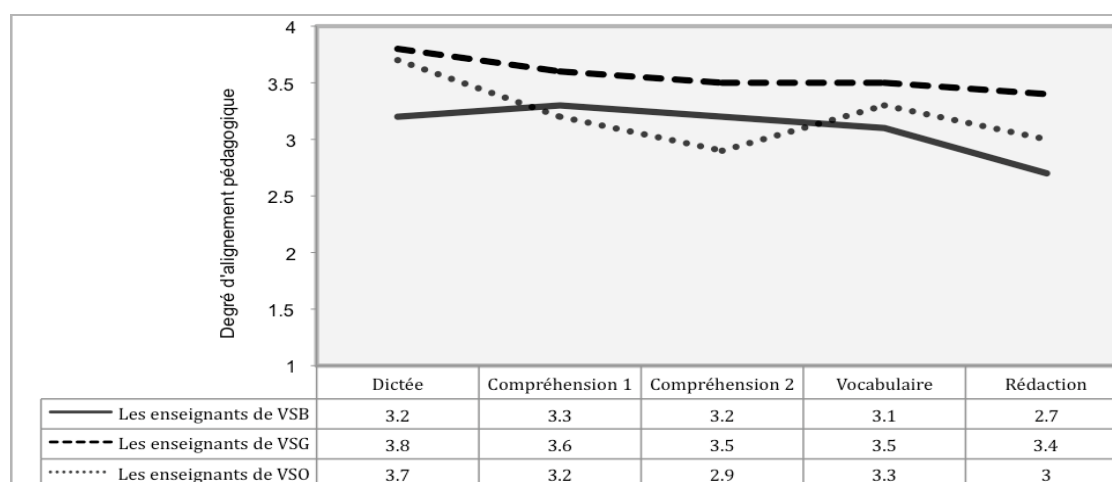
La question du niveau d'exigence et des critères d'évaluation conduit les enseignants, d'après leurs commentaires, à mal évaluer la correspondance entre les objectifs fondamentaux et l'ECR. Il est intéressant de mettre ces résultats en lien avec l'analyse de l'alignement de l'ECR sur les objectifs fondamentaux du PEV (voir chapitre 3). Ainsi, la production d'une lettre argumentative en VSO en particulier s'est révélée hors plan d'études. Malgré ce fait, il n'y a que 17% des enseignants de cette voie qui considèrent que l'activité d'écriture ne correspond *pas tellement* ou *pas du tout* aux objectifs fondamentaux du PEV. Le même non-alignement en compréhension des textes de chanson avait été constaté avec les objectifs fondamentaux de la VSO. Pourtant, ce ne sont que 9% des enseignants de cette voie qui signalent une non-correspondance avec les objectifs fondamentaux. L'alignement plus précis de l'ECR sur le plan d'études des élèves de VSB n'entraîne pourtant pas un meilleur alignement selon les enseignants de VSB puisque, pour chaque partie de l'ECR, ce sont eux qui considèrent, dans la proportion la plus importante, que les activités de l'ECR ne correspondent *pas tellement* ou *pas du tout* au plan d'études. Ces quelques constats conduisent à émettre l'hypothèse que les objectifs fondamentaux issus du PEV sont peu connus des enseignants et que ces derniers fondent leur appréciation (et leur enseignement) sur d'autres balises, par exemple un savoir d'expérience qui établit certains critères dans les exigences et dans la planification des

activités des élèves de la 7^e à la 9^e. Le questionnaire parle d'objectifs fondamentaux, les enseignants répondent en mettant en évidence le niveau d'exigence des activités.

5.4.2 DEGRÉ D'ALIGNEMENT PÉDAGOGIQUE

Les enseignants de toutes les voies confondues ont perçu un degré relativement élevé d'alignement pédagogique de toutes les parties de l'ECR de français, entre 3.0 et 3.5 sur 4. Mais, comme en ce qui concerne la validité curriculaire, les différences d'appréciation selon la voie dans laquelle on enseigne, sont assez importantes. La figure 5.4 met en évidence un degré élevé d'alignement pédagogique de toutes les parties pour les enseignants de VSG et un degré moyen pour les enseignants de VSB. Quant aux enseignants de VSO, un degré élevé pour certaines parties (la dictée et le vocabulaire) et moyen pour les autres (compréhension 1 et 2, et rédaction. Les différences d'appréciation entre les enseignants des trois voies sont statistiquement significatives comme en témoignent les résultats de l'analyse de variance multivariée : lambda de Wilks = .78, $F(10,162) = 2.17$, $p < .05$.

Figure 5.4 : Degré moyen d'alignement pédagogique perçu des différentes parties de l'ECR de français par les enseignants de VSO, VSG et VSB



Les principales divergences de vue se révèlent au niveau de la dictée et de la rédaction. Au niveau de la dictée, les comparaisons a posteriori à l'aide du test de Tukey font apparaître une différence significative entre les enseignants de VSB et ceux de VSO ($p < .05$), d'une part, ceux de VSG ($p < .05$), d'autre part. Au niveau de la rédaction, les différences d'appréciation du degré d'alignement pédagogique, qui sont apparues entre les enseignants de VSB et ceux de VSG, s'avèrent statistiquement significatives ($p < .05$). Au niveau des autres parties de l'épreuve, les différences de perception du degré de validité pédagogique, selon la voie dans laquelle on enseigne, ne sont pas statistiquement significatives.

Dans leurs commentaires, les enseignants de VSG, davantage que leurs collègues de VSO et de VSB, voient une très bonne adéquation entre les exercices de l'ECR et l'enseignement effectivement dispensé. Les enseignants de VSO sont également nombreux à reconnaître que les différentes parties de l'ECR correspondent tout à fait à l'enseignement assuré. L'un d'entre eux l'exprime ainsi dans ses commentaires : *l'ECR correspondait bien aux objectifs des classes de VSO et les textes choisis étaient intéressants ; ils touchaient les élèves, leur parlaient de manière concrète*. Aucune des activités demandées aux élèves ne lui paraît déconnectée des activités menées en classe.

Les enseignants de VSB sont plus réservés quant à la validité pédagogique des différentes parties de l'ECR. Dans leurs commentaires, plusieurs d'entre eux relèvent qu'il y a un écart entre ce qui est demandé à l'ECR et leurs exigences beaucoup plus élevées en particulier en dictée. Ils protestent contre des activités trop simples demandées aux élèves. Un enseignant de VSB évoque *un manifeste nivellement par le bas*.

Au sujet des questions sur le texte poétique (compréhension 2), quelques enseignants de VSB déplorent qu'on n'ait pas vérifié les connaissances des élèves en versification. La rédaction n'est pas entièrement du goût de certains maîtres de VSB qui regrettent qu'elle ne permette pas d'apprécier les capacités créatives des élèves, ou que des points soient trop facilement accordés aux élèves.

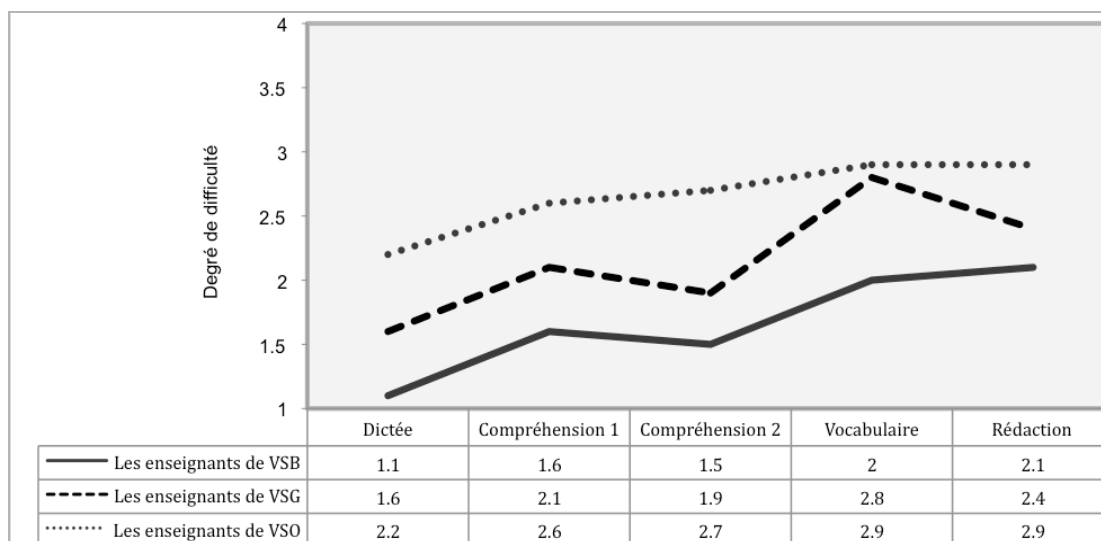
5.4.3 DEGRÉ DE DIFFICULTÉ

La question du degré de difficulté perçue divise encore plus les enseignants des trois voies que celle relative à l'alignement curriculaire ou celle du degré d'alignement pédagogique. La grande majorité des enseignants de VSB trouve que toutes les parties de l'ECR sont très faciles ou plutôt faciles pour ses élèves. Dans une moindre mesure que ceux de VSB, les enseignants de VSG estiment également que quatre parties de l'ECR (la dictée, la compréhension 1, la compréhension 2 et la rédaction) sont faciles voire très faciles, seul le vocabulaire étant perçu comme relativement difficile.

A l'opposé de ceux de VSB et de VSG, les enseignants de VSO trouvent trois parties de l'ECR (la compréhension 2, le vocabulaire et la rédaction) relativement difficiles pour leurs élèves, seules la dictée et la compréhension 1 leur apparaissant relativement faciles.

L'analyse de variance multivariée montre que ces différences d'appréciation du degré de difficulté de l'ECR entre les enseignants des trois voies sont statistiquement significatives : lambda de Wilks = .24, $F(10,160) = 16.76$, $p < .001$. Les analyses a posteriori à l'aide du test de Tukey montrent que les divergences de vue entre les trois catégories d'enseignants sont statistiquement significatives à toutes les parties de l'ECR, à l'exception de la compréhension 2, entre la VSB et la VSG, du vocabulaire, entre la VSG et la VSO, et de la rédaction entre la VSG et la VSB.

Figure 5.5 : Degré moyen de difficulté perçue des différentes parties de l'ECR de français par les enseignants de VSO, VSG et VSB



Les commentaires portent également la trace d'un fort clivage entre les enseignants de VSO et ceux de VSB. Un maître de VSO constate : *J'ai trouvé que pour la première fois les ECR étaient adaptées au niveau VSO. Mes élèves ont pu faire l'épreuve entièrement sans se décourager.* Un répondant de VSB s'insurge : *Forcément, vouloir prouver que les 3 voies peuvent rivaliser amène à des aberrations.*

Plusieurs enseignants de VSB s'inquiètent du barème trop généreux à leurs yeux, dans la correction de la dictée et dans l'attribution des points d'orthographe dans la production écrite (cinq erreurs permettent, en effet, d'obtenir le maximum de points). D'autres avis plus nuancés de la part de maîtres de VSB apparaissent également : *Cela me semblait facile, mais les résultats des élèves n'étaient pas si bons.*

Les enseignants de VSO s'inquiètent en revanche des critères stricts de correction qui font perdre des points aux élèves : *Les élèves ne sont pas assez précis, donc ils perdent des points.* Certains relèvent qu'ils corrigent de façon plus souple en cours d'année en particulier pour la compréhension de texte. Les questions de vocabulaire sont pointées comme problématiques, en particulier par les enseignants de VSO : *Pour les allophones, les nuances sont peu, voire pas perceptibles ; ils ont souvent un vocabulaire de base peu fourni.* Quelques enseignants de VSG leur emboîtent le pas : *Il n'était pas normal de pénaliser autant les élèves qui avaient trouvé le mot exact mais qui n'avaient pas effectué les accords.* L'accès à la culture pose également problème selon cet autre maître de VSG : *Les exercices de vocabulaire sont trop difficiles pour des élèves allophones ; aucun de mes élèves n'a compris "nous ne sommes pas racistes pour deux sous". Cela fait appel à une culture qu'ils n'ont pas.* Ces exercices de vocabulaire jugés les plus difficiles par les enseignants de VSO et VSG ne suscitent pas de commentaires de leurs collègues de VSB.

Manifestement, un clivage dans les perceptions de ces activités intervient. Ces questions - parce qu'elles reposent sur une grande précision dans la maîtrise de mots et expressions et n'ont pas fait l'objet d'un enseignement explicite - fâchent les enseignants s'occupant d'élèves scolairement moins performants. Elles ne posent, en revanche, aucun problème à leurs collègues oeuvrant avec des élèves plus performants scolairement.

Un enseignant de VSB met en cause le nombre de critères, selon lui trop précis et orientés, portant sur la lettre. Il considère que le processus d'écriture est entravé. Un autre questionne le corrigé : *Que faire des arguments qui vont au-delà de ce qui est demandé ?* Ces quelques commentaires démontrent que les pratiques évaluatives reposant sur des objectifs fondamentaux décrits en critères et indicateurs (et ayant fait l'objet d'un enseignement explicite) font débat et que les pratiques en évaluation de la production écrite sont très hétérogènes.

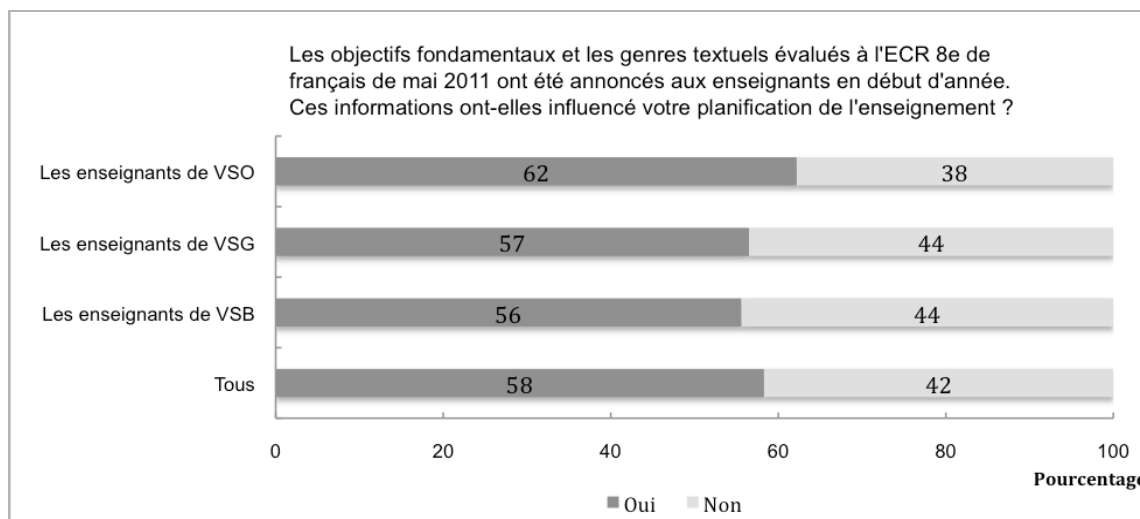
5.4.4 PRISE EN COMPTE DE L'ECR DANS LA PLANIFICATION DE L'ENSEIGNEMENT

Une majorité des répondants déclare avoir inclus dans sa planification un travail spécifique orienté vers l'ECR. Les écarts entre les voies sont minimes : 62% de oui en VSO ; 57% en VSG et 56% en VSB. La proportion apparaît faible dans un contexte où les parents, la direction et les enseignants reçoivent en début d'année scolaire les objectifs qui seront évalués lors de l'ECR. Malgré l'annonce par une directive officielle, 43% des enseignants disent ne pas s'inspirer de ces informations.

Parmi les réponses positives, 39 répondants signalent l'impact précis de l'annonce des objectifs sur leur planification. 25 enseignants mentionnent ainsi qu'ils ont anticipé le travail sur l'argumentation et 23 annoncent qu'ils ont ciblé un travail sur la poésie. Quelques répondants de VSB s'étonnent de l'écart entre le travail qu'ils ont mené -

notamment en versification - et les activités effectivement demandées aux élèves. Un professeur conclut que son travail de préparation n'a pas été utile aux élèves tant ses exigences étaient plus élevées que celles de l'ECR : *J'ai fait plus vite de la versification, mais en fait rien n'était utile pour cette ECR... je me suis assuré que mes élèves connaissent les principes de base de l'argumentation, mais là encore quel écart avec ce qui a été finalement demandé ; j'aurais tout aussi bien pu ne rien préparer....*

Figure 5.6 : Pourcentage d'enseignants de français ayant tenu compte de l'ECR dans la planification de l'enseignement



On constate donc un impact de l'ECR sur la planification, en contraignant un certain nombre d'enseignants à travailler l'argumentation et la poésie à un moment qui ne leur est pas coutumier ; et aussi à y consacrer plus de temps qu'habituellement, notamment pour la poésie.

Deux répondants relèvent d'ailleurs que les manuels d'enseignement utilisés ne prévoient aucun travail sur la poésie (ce qui est effectivement le cas pour *L'Atelier du langage* alors que *Le Livre unique* propose des séquences sur le genre poétique). La forte minorité (42%) d'enseignants ne tenant pas compte des objectifs de l'ECR questionne sur leur posture face à des démarches d'évaluation standardisée. Le constat par plusieurs enseignants de l'écart en exigence du travail effectué en prévision de l'ECR et les activités demandées dans l'épreuve a-t-il conduit une part significative de professionnels expérimentés à négliger un travail ciblé sur l'ECR ? Ou s'agit-il d'une mesure de défiance face à l'imposition de cette épreuve cantonale ?

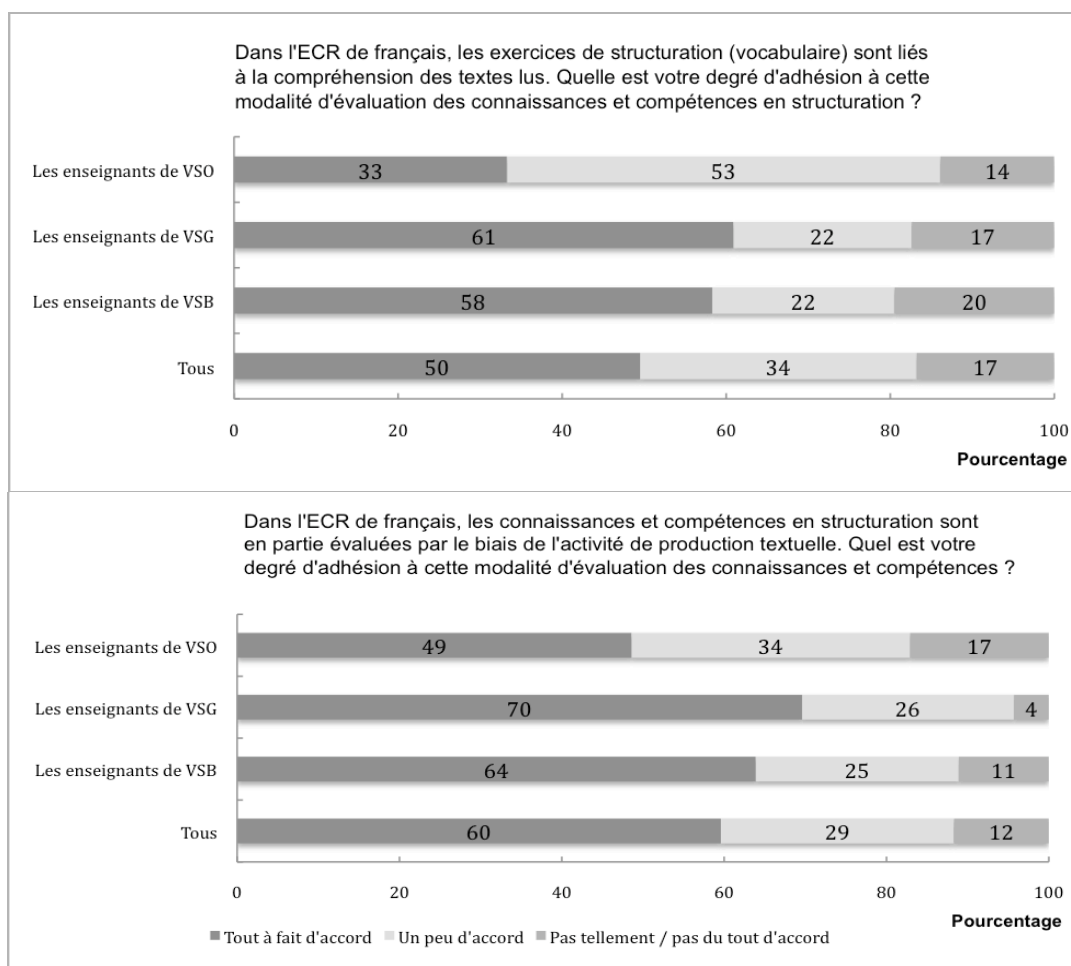
5.4.5 DEGRÉ D'ADHÉSION AUX MODALITÉS D'ÉVALUATION CARACTÉRISTIQUES DE L'ECR

Le fait d'évaluer la connaissance du fonctionnement de la langue, principalement le vocabulaire et l'orthographe, par le biais des questions de compréhension du texte est admis par la majorité des enseignants, sans différence significative entre les voies.

Plusieurs enseignants dévoilent les raisons de leur soutien à cette modalité d'évaluation : *Il faut étudier la langue en contexte, donc je suis tout à fait d'accord avec cette démarche. La structuration du français ne peut être négligée dans l'expression.* Les quelques reproches face à cette modalité reposent sur deux aspects : le niveau des textes proposés dans l'ECR (chansons) qui ne permettrait pas vraiment de créer des activités d'un bon

niveau en structuration et l'inadéquation de l'épreuve pour les élèves en difficulté. Les reproches contre le niveau des textes émanent des enseignants de VSB et de VSG. L'un d'entre eux s'exprime ainsi : *Oui, mais à condition de trouver des textes qui permettent des exercices du niveau des élèves, ce qui n'est pas évident ; trop facile pour les 8B ; la pauvreté littéraire des textes choisis est discutable.*

Figure 5.7 : Pourcentage d'enseignants favorables aux modalités d'évaluation du fonctionnement de la langue dans l'ECR de français



Le 2^e aspect problématique concerne les élèves en difficulté. Ils se trouvent face à un double problème : effectuer des activités de structuration et les mettre en lien avec une compréhension parfois partielle du texte de départ. Ce sont des enseignants de VSO et parfois de VSG qui mettent en évidence cet écueil :

- *Une mauvaise compréhension générale fragilise les élèves qui sont sûrs de rater parce qu'ils n'ont pas compris.*
- *Pour les élèves qui auraient mal compris le texte, les exercices de structuration deviennent très compliqués.*
- *L'activité de vocabulaire sur l'émigration est très difficile si la problématique n'est pas comprise.*

L'évaluation de la maîtrise du fonctionnement de la langue, principalement le vocabulaire et l'orthographe, à travers l'activité de production écrite est également validée par les

enseignants. 84% d'entre eux se disent en effet *un peu* ou *totalemment d'accord* avec cette modalité d'évaluation.

L'adhésion majoritaire à cette modalité d'évaluation ne manque pourtant pas de divergences. Seuls 33% des enseignants en VSO contre 61% et 58% en VSG et en VSB sont *tout à fait d'accord* avec le fait d'évaluer le fonctionnement de la langue au travers d'une production textuelle. Cette différence de point de vue entre les deux catégories d'enseignants est statistiquement significative ($\chi^2(4) = 9.68; p < 0.05$).

Quelques commentaires renseignent sur les arguments de ceux qui n'adhèrent *pas tellement* ou *pas du tout* à cette modalité d'évaluation des compétences en structuration. Il y a d'abord le temps de travail nécessaire pour la correction de productions écrites et la non-adéquation de cette modalité d'évaluation avec les élèves de VSO. Il y a ensuite le faible degré de fiabilité car les élèves *contournent les difficultés, rédigent des phrases trop peu complexes* ou font preuve d'un vocabulaire pauvre. La trop forte hétérogénéité des productions péjore la valeur des points attribués et la syntaxe est sous-évaluée.

Un enseignant de VSO peu favorable à l'approche intégrée de l'évaluation du fonctionnement de la langue met en avant la différence de niveau entre les élèves de VSO et ceux des deux autres voies : *Je pense que ce genre d'exercices correspond très bien à des VSB et VSG, mais pour les VSO, c'est une autre histoire ! Oui, mais pas avec l'évaluation de la syntaxe sous-évaluée, certains textes très bien écrits ont peu de points et inversement.*

Un autre motif pédagogique conduit un enseignant de VSB à rejeter cette modalité : l'aspect démotivant pour un élève de voir son texte annoté à divers niveaux (*On ne peut pas tout évaluer dans une production textuelle car l'élève ne peut pas progresser si on le corrige à tous les niveaux : mots, phrase ou texte*).

Le modèle didactique d'évaluation de savoir-faire en situation complexe (compréhension et production de texte) apparaît largement accepté par les enseignants du secondaire. Il sera intéressant de voir dans quelle mesure les épreuves communes d'établissements confortent cette position (voir chapitre 6).

5.4.6 DEGRÉ DE COHÉRENCE ENTRE L'ECR ET LES TRAVAUX SIGNIFICATIFS

Une majorité d'enseignants de VSB (56%) considère que les ECR ne correspondent *pas du tout* ou *pas tellement* à leurs propres épreuves d'évaluation sommative. A l'inverse, la majorité des enseignants de VSG (91%) et de VSO (73%) estime que les ECR correspondent *un peu* ou *tout à fait* aux travaux significatifs de son cru. Cette différence de point de vue est statistiquement significative ($\chi^2(2) = 14.40; p < 0.01$).

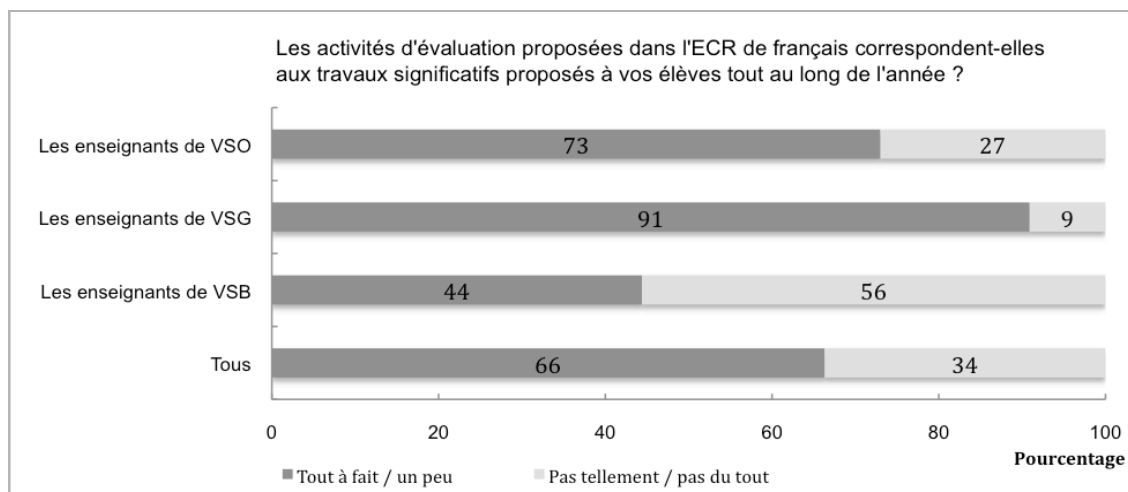
On peut relever, à nouveau, que les enseignants de VSG déclarent pratiquer une évaluation plus proche de l'ECR que leurs collègues des autres voies. En revanche, la correspondance entre l'ECR et les travaux significatifs paraît faible pour les élèves de VSB.

Les commentaires permettent d'éclaircir les raisons des écarts constatés. 11 enseignants, dont 10 de VSB, relèvent le faible niveau d'exigence des questions, la superficialité et la facilité des tâches prévues dans l'ECR.

Plus concrètement, quelques enseignants remarquent que les questions de leurs évaluations demandent plus un *développement* qu'un *remplissage*. Un maître de VSG signale que *la compréhension de textes est évaluée toute l'année par le biais des lectures suivies ; je ne faisais pas une compréhension d'un texte ex nihilo*, dit-il. En revanche, deux

enseignants de VSO trouvent les consignes difficiles d'accès pour leurs élèves et le nombre d'objectifs testés trop important.

Figure 5.8 : Degré de correspondance perçue par les enseignants entre l'ECR de français et les travaux significatifs



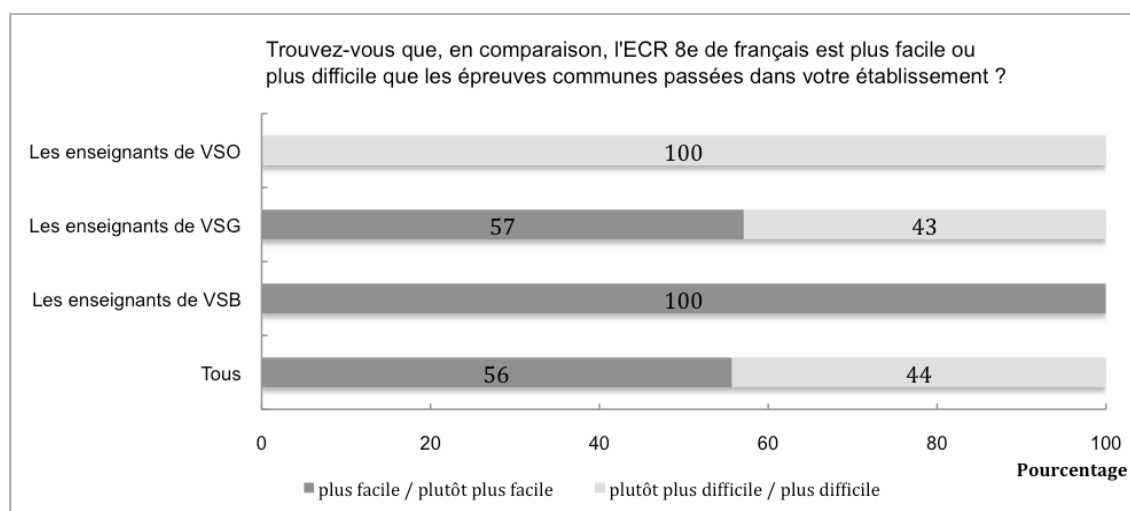
Quelques remarques signalent un écart du fait de la sous-représentation de la grammaire et de la conjugaison dans l'ECR. Le décompte des points, notamment les fautes admises pour obtenir le maximum de points, laisse plusieurs répondants dubitatifs.

5.4.7 DEGRÉ DE COHÉRENCE ENTRE L'ECR ET LES ÉPREUVES COMMUNES D'ÉTABLISSEMENT

Sept enseignants sur dix disent avoir eu recours aux épreuves communes d'établissement. La proportion d'enseignants ayant fait passer à leurs élèves une ou plusieurs épreuves communes d'établissement est à peu près la même dans toutes les voies.

Les épreuves communes d'établissements sont conçues dans 91% des cas pour des classes de même voie et seulement dans 9% des cas pour des classes de 2 ou 3 voies différentes.

Figure 5.9 : Différence de difficulté perçue par les enseignants entre l'ECR de français et les épreuves communes d'établissement



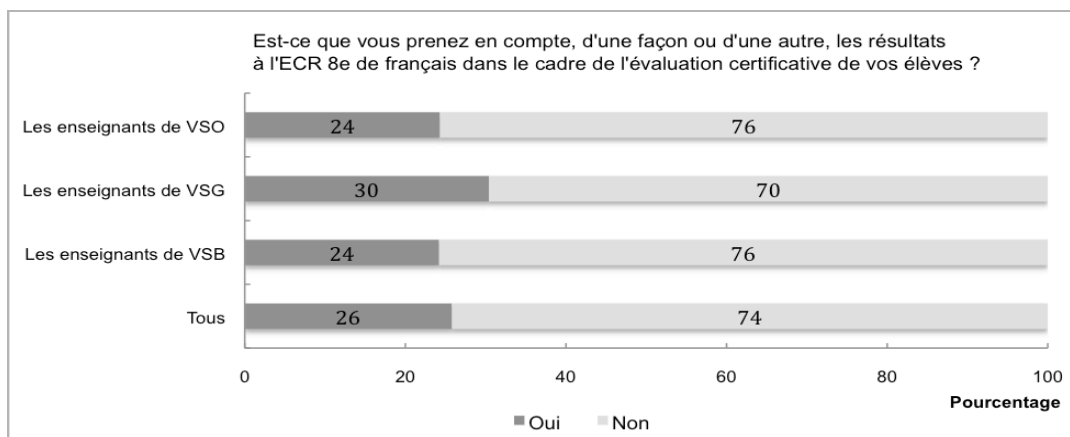
En comparaison, tous les enseignants de VSO estiment que les ECR sont plus difficiles que les épreuves communes d'établissements destinés à leurs élèves. A l'inverse, tous les enseignants de VSB considèrent que les ECR sont plus faciles. Les enseignants de VSG sont plus partagés, 57% estimant que leurs épreuves communes sont plus difficiles, 43% étant d'un avis opposé.

Quelques commentaires d'enseignants de VSB et de VSO indiquent que les épreuves communes d'établissement portent sur moins d'objectifs et ciblent des savoirs plus précis que les ECR. Ces enseignants soulignent également le fait que les épreuves communes d'établissement portent le plus souvent sur la structuration ou sur la littérature et moins sur la production écrite.

5.4.8 UTILISATION DE L'ECR DANS LE CADRE DE L'ÉVALUATION CERTIFICATIVE

Environ un quart des enseignants, toutes voies confondues, disent utiliser les résultats de leurs élèves à l'ECR dans le cadre de l'évaluation certificative.

Figure 5.10 : Pourcentage d'enseignants de français utilisant les résultats de l'ECR dans le cadre de l'évaluation certificative de leurs élèves



La proportion de ces enseignants est un peu plus élevée en VSG (30%) qu'en VSO (24%) et en VSB (24%), mais la différence n'est pas statistiquement significative. Cette proportion paraît importante dans un contexte où les ECR ont, dans la loi scolaire, un statut informatif et non certificatif.

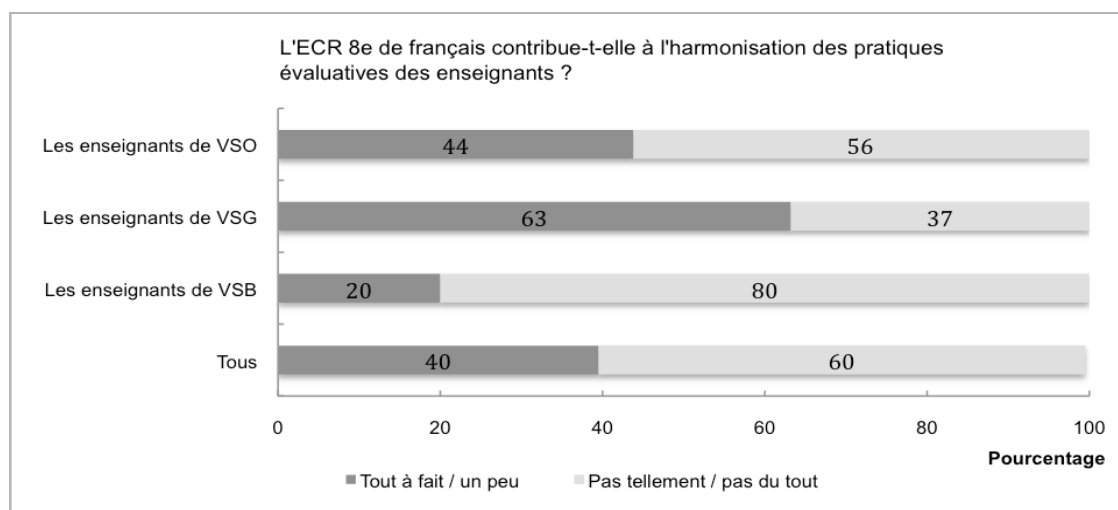
Dans les réponses complémentaires, plusieurs enseignants mentionnent que la démarche n'est pas autorisée, mais que parfois ils prennent en compte les résultats à l'ECR pour argumenter le passage ou l'échec, ou pour une décision de prolongement de scolarité. Trois enseignants proposent de faire de la note de l'ECR une note facultative en rédaction. D'autres relèvent que le résultat est utile pour échanger avec les parents ou à titre d'évaluation formative pour reprendre l'année suivante ce qui n'a pas été compris. Plusieurs regrettent que cela ne compte pas alors que la correction demande beaucoup de temps.

5.4.9 RÔLE DE L'ECR DANS LA RÉGULATION DES PRATIQUES ÉVALUATIVES DES ENSEIGNANTS

La majorité des enseignants de VSB (80%) et de VSO (56%) ne croit pas que l'ECR jouerait un rôle régulateur de sa pratique d'évaluation alors qu'en VSG, la majorité des enseignants

(63%) en est persuadée. Les divergences de vue à ce sujet entre les enseignants de VSO, ceux de VSG et ceux de VSB sont statistiquement significatives ($\chi^2(2) = 9.47$; $p < 0.01$).

Figure 5.11 : Pourcentage d'enseignants pour lesquels l'ECR joue un rôle régulateur de leurs pratiques évaluatives



De nombreux commentaires critiques sur la démarche d'harmonisation des pratiques évaluatives s'expriment. Il y a d'abord plusieurs enseignants qui se prononcent clairement contre toute tentative d'harmoniser les pratiques d'évaluation : *risque de tout harmoniser = dérive, les enseignants sont des êtres humains, tous différents les uns des autres* (un enseignant de VSB) ; *c'est toujours difficile d'entrer dans la pratique évaluative de quelqu'un d'autre et en général frustrant* (un enseignant de VSB) ; *je ne suis pas pour cette harmonisation* (un enseignant de VSG) ; *doux rêve caressé par le Département et la HEP... vive la mondialisation des pratiques absurdes* (un enseignant de VSB) ; *certain font du bachotage ce qui ne correspond pas aux buts visés* (un enseignant de VSO) ; *ces pratiques de piètre qualité ne peuvent devenir un modèle* (un enseignant de VSB) ; *les gens suivent les consignes des ECR, mais ensuite chacun garde son propre système* (un enseignant de VSB).

Bachotage, imposition de pratiques vides de sens, difficulté de réellement évaluer de façon commune, inutilité de toute démarche qui ne reconnaît pas le statut particulier de la relation pédagogique, par définition hors norme, mauvaise qualité des épreuves, autant d'arguments utilisés par ce groupe minoritaire qui s'oppose à des évaluations standardisées.

Quelques pistes de transfert dans les pratiques sont aussi évoquées par des répondants dévoilant ainsi une attitude plus intéressée envers ce type d'épreuve : *l'ECR peut fournir des exemples d'épreuves quant à sa forme, en particulier le centrage sur le texte ou thème, pour l'évaluation de compétences variées, sinon, elle ne correspond pas aux connaissances et compétences exigées* (un enseignant de VSB) ; *elles offrent un autre regard et permettent des discussions intéressantes* (un enseignant de VSO).

La dernière remarque appuie le constat qu'un quart des enseignants prennent en compte l'ECR dans l'évaluation certificative des élèves en particulier en cas d'échec ou de demande de prolongation de scolarité.

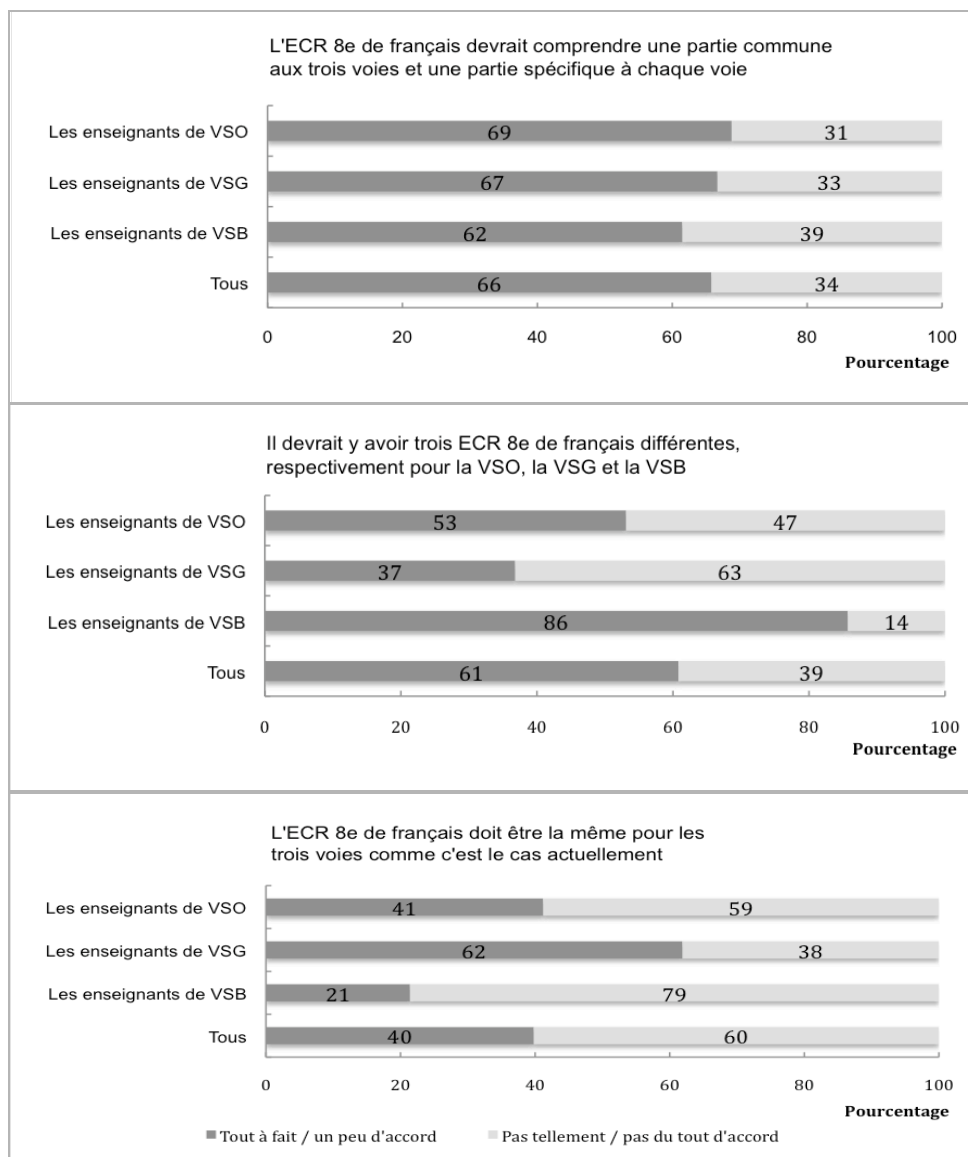
Un dernier commentaire signale la qualité de l'ECR du fait de ses conditions privilégiées de production : *Pas possible de rivaliser, nous ne faisons pas qu'une épreuve par année !*

Manifestement, des résistances importantes s'expriment face à la visée d'harmoniser les pratiques évaluatives. La relative nouveauté des ECR en 8^e année explique peut-être partiellement la méfiance ressentie par une majorité d'enseignants. La question du niveau de l'épreuve et du bien-fondé d'une évaluation commune aux trois voies joue peut-être également un rôle important dans cette méfiance affichée.

5.4.10 MODALITÉS DE DIFFÉRENCIATION DE L'ECR EN FONCTION DES VOIES

La majorité des enseignants (60%) ne veut pas d'une épreuve commune aux trois voies. Ces enseignants préfèrent qu'il y ait trois épreuves différentes, une épreuve par voie à 61%, ou mieux encore une épreuve comprenant une partie commune et une partie spécifique selon les voies à 66%. Les avis n'apparaissent pas si tranchés : on le voit, les ECR communes comptent - malgré tous les commentaires négatifs recensés - une minorité importante (40%) d'adeptes.

Figure 5.12 : Degré d'accord des enseignants de français avec trois modèles de différenciation de l'ECR de 8^e année



Les différences entre les voies, en revanche, sont marquées : seuls les enseignants de VSG se déclarent majoritairement pour un maintien du statu quo (62% contre 41% en VSO et seulement 21% en VSB). Quant à une possible évolution de l'épreuve actuelle vers trois épreuves selon les voies, 86% des répondants de VSB soutiennent cette solution ; ils ne sont que 37% en VSG et 53% en VSO. Le modèle d'une ECR qui comprendrait une partie commune aux trois voies et une partie spécifique retient l'adhésion de la majorité des répondants quelle que soit leur voie d'enseignement (69% des enseignants de VSO, 67% de VSO et 62% de VSB). Sur cette dernière question, les différences entre les voies ne sont pas statistiquement significatives.

La piste qui obtient les faveurs des répondants est précisée en ces termes dans les commentaires : *Des textes communs oui, mais trois épreuves différentes avec des activités mieux ciblées, ainsi les résultats pourraient compter et la démarche aurait plus de sens ; la comparaison partielle est une bonne solution, les mêmes exercices avec des consignes différentes.*

Les autres commentaires à cette question laissent entrevoir deux tendances. Celle d'abord des enseignants qui trouvent ce type d'évaluation standardisée inutile et chronophage. Ces derniers relèvent qu'il n'est pas possible d'harmoniser les corrections ni de disposer de données vraiment scientifiques, en raison des biais induits par une correction menée isolément par chaque enseignant.

Une autre tendance se dégage avec des enseignants qui souhaitent que la note à l'ECR compte et que le temps consacré aux corrections soit ainsi valorisé, également pour que les élèves accordent plus d'attention à cette épreuve.

Un enseignant exprime un avis partagé par de nombreux collègues de VSB : *3 voies = 3 modes didactiques, pédagogiques et sociaux donc 3 épreuves si l'élève veut pouvoir se situer dans ses apprentissages.*

5.4.11 SYNTHÈSE

- 1 L'alignement curriculaire de l'ECR de français ne fait guère de doute pour la majorité des enseignants des trois voies confondues. Toutes les parties de l'ECR sont perçues comme suffisamment alignées au plan d'études. Selon la voie dans laquelle on enseigne, pourtant, le degré d'alignement curriculaire perçu de l'ECR de français est relatif. Ainsi l'exercice de la dictée oppose les enseignants de VSB à ceux de VSO. Les premiers considèrent que cette partie de l'épreuve est au-dessous des attentes minimales, les deuxièmes signalent qu'elle est adaptée aux objectifs fondamentaux.
- 2 L'alignement pédagogique perçu par les enseignants se révèle plus nuancé. La majorité des répondants issus de toutes les voies reconnaît que les activités proposées dans l'ECR correspondent à l'enseignement effectivement assuré et sont de bonne facture. Mais ces répondants perçoivent un grand écart entre le niveau d'exigences de l'ECR et celui de leurs propres épreuves d'évaluation certificative. Les enseignants de VSO avouent que les activités proposées dans l'ECR, excepté la dictée, sont beaucoup plus difficiles que leurs propres épreuves communes d'établissement. A l'opposé, ceux de VSB regrettent le nivellement vers le bas de l'ECR en comparaison avec le niveau d'exigence de leurs travaux significatifs ou de leurs épreuves communes. Entre ceux de VSB et ceux de VSO, les enseignants de VSG voient une certaine convergence entre l'ECR et leurs pratiques évaluatives.

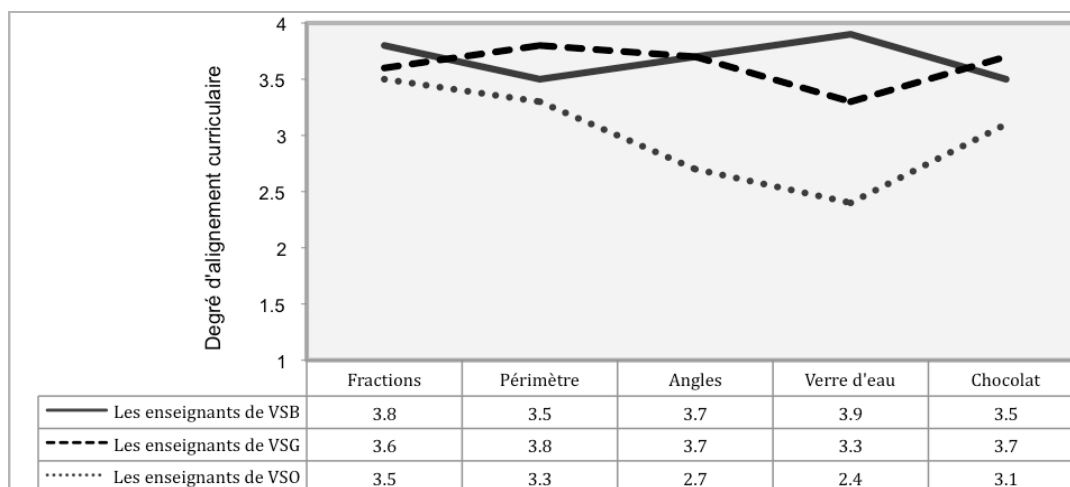
- 3 La majorité des enseignants de français ne voit pas l'utilité de l'ECR. Son indifférence à l'égard d'une épreuve externe à enjeux faibles n'étonne guère. Pourtant, le fait que près d'un tiers d'entre eux dit tirer profit d'une manière ou d'une autre de l'épreuve, et que plus du tiers lui reconnaît un rôle régulateur de sa pratique évaluative conforte l'image d'une épreuve externe plutôt bien perçue.
- 4 La majorité des enseignants, même ceux d'entre eux qui reconnaissent à l'ECR de français une validité curriculaire et une validité pédagogique satisfaisantes, est opposée au principe d'une épreuve unique commune aux trois voies. Ces enseignants préféreraient, soit trois épreuves différentes, une par voie, soit trois épreuves différentes mais avec une partie commune et des parties spécifiques aux voies. Mais à ce sujet également, d'importants désaccords existent entre enseignants des trois voies. Les maîtres de VSO qui dans leur majorité rejettent l'idée d'une épreuve unique pour les trois voies préfèrent qu'il y ait trois épreuves différentes avec une partie commune et des parties spécifiques. Les enseignants de VSB, majoritairement opposés à l'épreuve unique également, préfèrent quant à eux qu'il y ait trois épreuves entièrement différentes ou dans une moindre mesure trois épreuves partiellement différenciées, avec une partie commune et une partie spécifique à chaque voie. A la différence de ceux de VSO et de VSB, les enseignants de VSG ne rejettent pas l'idée d'une épreuve unique commune aux trois voies, mais préfèrent qu'il y ait trois épreuves partiellement différenciées, avec une partie commune et une partie spécifique à chaque voie. Ils rejettent en revanche la formule de trois épreuves entièrement différenciées.

5.5 VALIDITÉ ET UTILITÉ PERÇUES DE L'ECR DE MATHÉMATIQUES

5.5.1 DEGRÉ D'ALIGNEMENT CURRICULAIRE

Du point de vue de tous les enseignants sans distinction quant à leur voie d'affectation, l'ECR de mathématiques est bien alignée sur le plan d'études. Le degré moyen d'alignement curriculaire perçu des cinq problèmes varie de 3.1 à 3.6 sur 4.

Figure 5.13 : Degré moyen d'alignement curriculaire perçu des différentes tâches de l'ECR de mathématiques par les enseignants de VSO, VSG et VSB



L'appréciation du degré d'alignement curriculaire des cinq problèmes est toutefois différente si on analyse séparément les réponses des enseignants de chaque voie. Ce sont

surtout les enseignants de VSB et ceux de VSG qui perçoivent un degré élevé d'alignement curriculaire de tous les problèmes. Les enseignants de VSO ne reconnaissent un bon alignement curriculaire qu'à trois problèmes (*fractions*, *périmètre* et *chocolat*), estimant que les deux autres (*angles* et *verre d'eau*) ne sont pas adaptés au programme de leur voie.

D'après les résultats de l'analyse de variance multivariée, l'effet global de la voie sur le degré d'alignement curriculaire perçu est statistiquement significatif : lambda de Wilks = .49, $F(10,144) = 6.10$, $p < .001$. D'après les analyses a posteriori à l'aide du test de Tukey, cependant, les différences d'appréciation du degré d'alignement curriculaire sont statistiquement significatives seulement au niveau de trois problèmes : les *angles*, entre les enseignants de VSO et ceux de VSB ($p < .001$) ainsi que ceux de VSG ($p < .001$) ; le *verre d'eau*, entre les enseignants de VSO et ceux de VSB ($p < .001$) ainsi que ceux de VSG ($p < .01$) ; le *chocolat*, entre les enseignants de VSO et de VSG ($p < .05$). Seuls deux problèmes (*fractions* et *périmètre*) sont donc perçus comme bien alignés aux plans d'études suivis dans les trois voies.

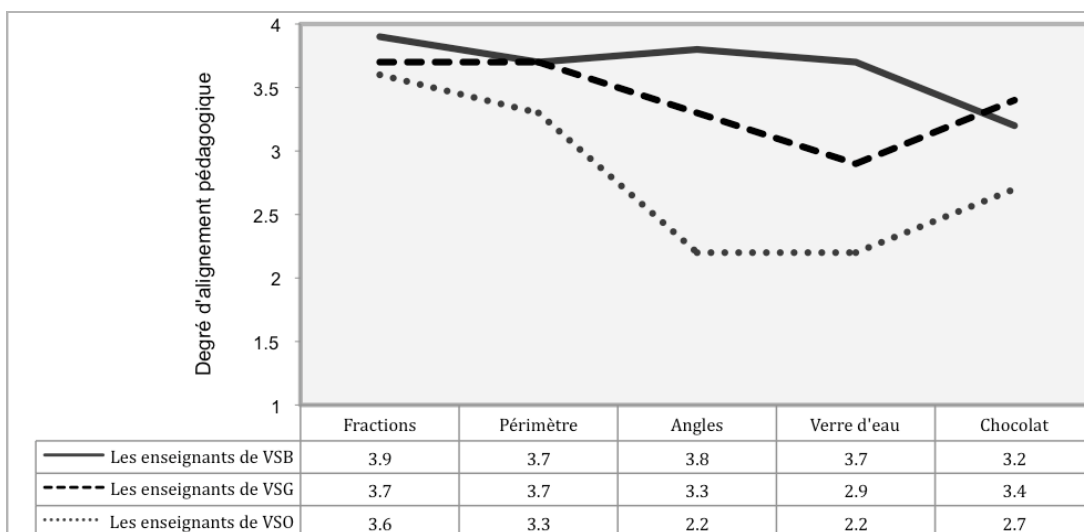
Dans leurs commentaires, beaucoup d'enseignants de VSO et dans une moindre mesure ceux de VSG pointent du doigt l'alignement curriculaire du problème *verre d'eau* : *verre d'eau difficilement accessible pour des élèves de VSO, cumul de difficultés* (un enseignant de VSO) ; *verre d'eau niveau VSB car pas d'aiguillage, de questions intermédiaires amenant à la réponse* (un enseignant de VSG).

5.5.2 DEGRÉ D'ALIGNEMENT PÉDAGOGIQUE

Tout comme leur alignement curriculaire, le degré moyen d'alignement pédagogique perçu par les enseignants de toutes les voies confondues est assez élevé pour les cinq problèmes de l'ECR de mathématiques, entre 2.9 et 3.7 sur 4.

De même, les différences d'appréciation du degré d'alignement pédagogique selon la voie dans laquelle on enseigne sont assez importantes (figure 5.14). En attestent les résultats de l'analyse de la variance multivariée ainsi que des analyses a posteriori à l'aide du test de Tukey.

Figure 5.14 : Degré moyen d'alignement pédagogique perçu des différentes tâches de l'ECR de mathématiques par les enseignants de VSO, VSG et VSB



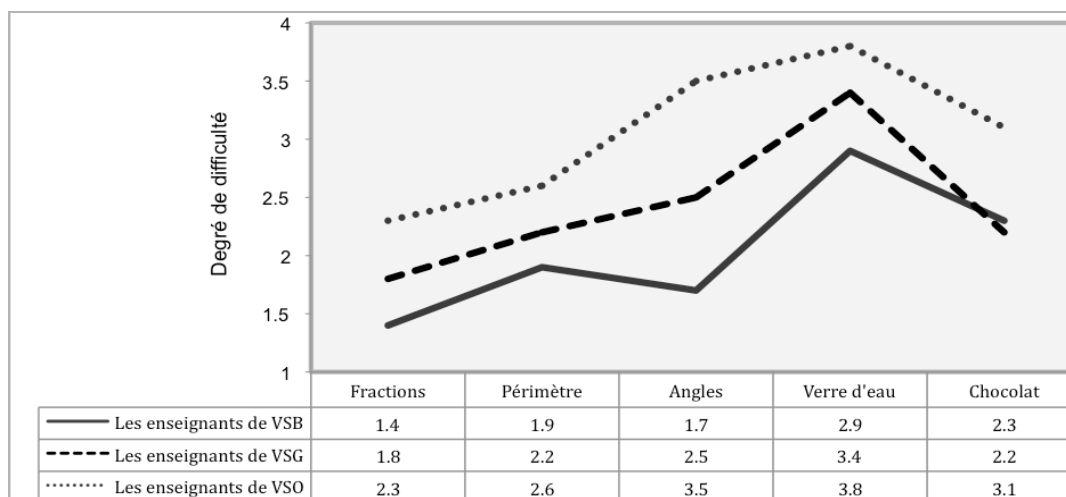
L'analyse de variance multivariée révèle un effet global de la voie à laquelle les enseignants sont rattachés sur leur perception du degré d'alignement pédagogique des cinq problèmes : λ de Wilks = .43, $F(10,148) = 7.72$, $p < .001$. Le test de Tukey met en évidence des différences d'appréciation du degré d'alignement curriculaire statistiquement significatives au niveau des mêmes problèmes qu'au sujet de l'alignement curriculaire : les *angles*, entre les enseignants de VSO et ceux de VSB ($p < .001$) ainsi que ceux de VSG ($p < .001$) ; le *verre d'eau*, entre les enseignants de VSO et ceux de VSB ($p < .001$) ainsi que ceux de VSG ($p < .01$) et entre ceux de VSG et ceux de VSB ($p < .01$) ; le *chocolat*, entre les enseignants de VSO et de VSG ($p < .05$). Seuls deux problèmes (*fractions* et *périmètre*) sont donc perçus comme bien alignés sur l'enseignement effectivement assuré dans les trois voies.

Le problème *verre d'eau* suscite les plus fortes oppositions des enseignants de VSO qui expliquent dans leurs commentaires que *Pythagore dans l'espace est prévu en 9^e*.

5.5.3 DEGRÉ DE DIFFICULTÉ

A l'exception du *verre d'eau*, tous les problèmes sont perçus comme relativement faciles par la majorité des enseignants dans les trois voies. Sur l'échelle de 1 à 4, le degré moyen de difficulté perçue est de 1.9, 2.2, 2.6, 3.4 et 2.6 respectivement pour les problèmes *fractions*, *périmètre*, *angles*, *verre d'eau* et *chocolat*. Néanmoins, comme en ce qui concerne l'alignement curriculaire ou l'alignement pédagogique, les différences d'appréciation du degré de difficulté perçue, selon la voie dans laquelle on enseigne, sont assez importantes. Les résultats de l'analyse de la variance multivariée ainsi que des analyses a posteriori à l'aide du test de Tukey, le montrent.

Figure 5.15 : Degré moyen de difficulté perçue des différentes parties de l'ECR de mathématiques par les enseignants de VSO, VSG et VSB



L'analyse de variance multivariée met en évidence un effet global de la voie dans laquelle les enseignants sont affectés sur leur perception du degré de difficulté des cinq problèmes : λ de Wilks = .28, $F(10,152) = 13.47$, $p < .001$.

D'après les résultats du test de Tukey, les différences d'appréciation du degré de difficulté selon la voie dans laquelle on enseigne sont statistiquement significatives au niveau de tous les problèmes : les *fractions*, entre les enseignants de VSO et ceux de VSB ($p < .001$) ainsi que ceux de VSG ($p < .01$) ; *périmètre*, entre les enseignants de VSO et ceux de VSB ($p < .001$) ainsi que ceux de VSG ($p < .05$) ; les *angles*, entre les enseignants de VSO et ceux

de VSB ($p < .001$) ainsi que ceux de VSG ($p < .001$) et entre ceux de VSG et ceux de VSB ($p < .001$) ; le *verre d'eau*, entre les enseignants de VSO et ceux de VSB ($p < .001$) et entre ceux de VSG et ceux de VSB ($p < .05$) ; le *chocolat*, entre les enseignants de VSO et ceux de VSG ($p < .001$) ainsi que ceux de VSB ($p < .001$).

Dans leurs commentaires, certains enseignants de VSO attribuent la difficulté des problèmes à l'aspect transversal de ceux-ci : *trop de notions à mobiliser simultanément ; difficulté de décodage de ce qui est demandé ; notions pas clairement exprimées ; l'élève ne sait pas dans quel domaine il se situe ; les élèves ne comprennent pas ce qu'on leur demande ; présentation des problèmes difficile à comprendre pour les élèves.*

Entre ceux de VSB et ceux de VSO, les enseignants de VSG perçoivent un certain équilibre entre les problèmes faciles et ceux qui leur paraissent difficiles pour leurs élèves. La majorité d'entre eux estime que les problèmes *fractions, périmètre et chocolat* sont très faciles ou plutôt faciles alors qu'à l'unanimité, le problème du *verre d'eau* est très difficile ou plutôt difficile pour ses élèves. Ces enseignants sont très partagés sur le degré de difficulté du problème des *angles* : 50% le trouvent facile, 50% sont d'un avis contraire.

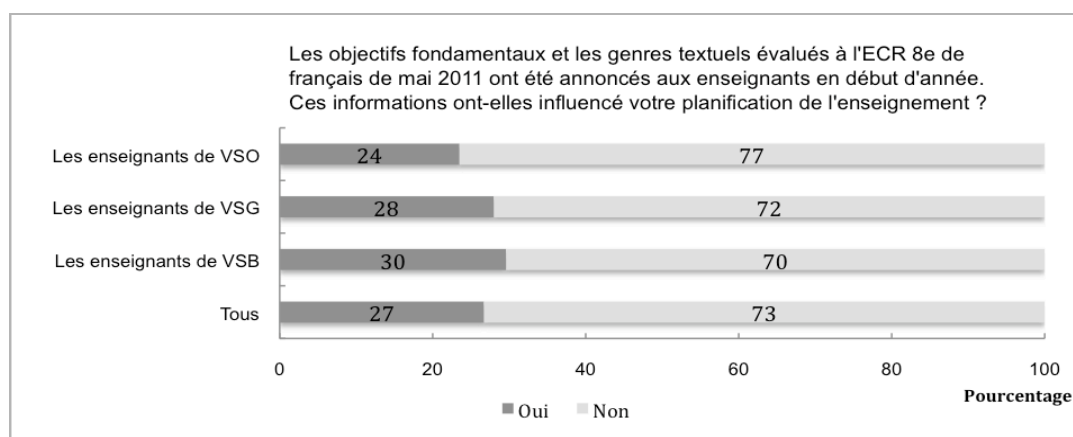
5.5.4 PRISE EN COMPTE DE L'ECR DANS LA PLANIFICATION DE L'ENSEIGNEMENT

Contrairement à ceux de français, la majorité des enseignants de mathématiques dit ne pas avoir tenu compte du courrier de la Direction générale de l'enseignement obligatoire annonçant les objectifs évalués à l'ECR.

Les écarts entre les voies sont faibles, statistiquement non significatifs : 77% en VSO, 72% en VSG et 70% en VSB.

Un enseignant de VSO explique ainsi sa position : *priorité est donnée aux élèves, à leur capacité, au programme réparti sur 3 ans afin d'être prêt pour les examens en fin de 9^e ; un certain nombre d'objectifs demandent une maturité qu'un élève de VSO n'atteint qu'en 9^e après avoir "digéré" des notions de base acquises en 7^e et en 8^e.*

Figure 5.16 : Pourcentage d'enseignants de mathématiques ayant adapté leur planification de l'enseignement en fonction de l'annonce des objectifs évalués dans l'ECR

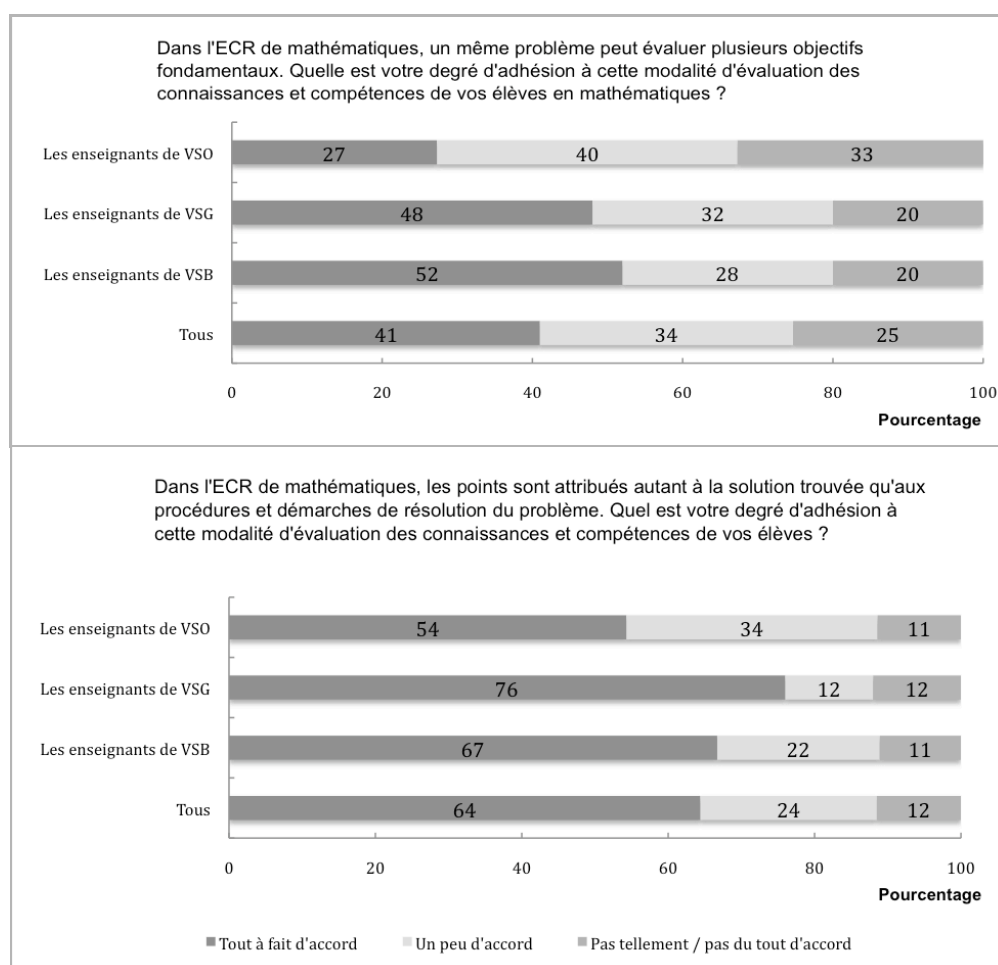


Certains enseignants trouvent même inapproprié le fait de connaître à l'avance les objectifs évalués, ce qu'ils expriment en ces termes : *ils ne devraient pas être connus des enseignants à l'avance* (un enseignants de VSG) ; *ils m'ont empêché d'user de ma liberté pédagogique fondamentale et ont induit un bachotage systématique de ces objectifs* (un enseignant de VSB).

5.5.5 DEGRÉ D'ADHÉSION AUX MODALITÉS D'ÉVALUATION CARACTÉRISTIQUES DE L'ECR

La majorité des enseignants des trois voies (75%) est *tout à fait* ou *un peu d'accord* avec le fait d'évaluer plusieurs compétences dans un même problème. Les différences d'appréciation de cette caractéristique des ECR de mathématiques ne sont pas significativement liées au fait d'enseigner dans telle ou telle autre voie. Dans leurs commentaires, plusieurs enseignants justifient ainsi cette façon de faire : *les élèves ont tendance à travailler par chapitres et oublier une fois le test terminé ; des exercices globaux permettent une meilleure maîtrise des compétences.*

Figure 5.17 : Pourcentage d'enseignants favorables aux modalités d'évaluation spécifiques à l'ECR 8^e mathématiques



Le fait d'attribuer des points aussi bien aux solutions trouvées qu'aux démarches de résolution des problèmes, une autre caractéristique des ECR de mathématiques, est également approuvée par la majorité des enseignants, soit 88% qui sont *tout à fait* ou *un peu d'accord* avec cette caractéristique des ECR. Les différences d'appréciation entre les voies ne sont pas significatives sur ce point non plus.

Dans leurs commentaires, plusieurs enseignants expriment cependant des réserves ou leurs opposition sur ce point : *d'accord sur le principe, mais dans les ECR le système d'évaluation des points d'un problème est si mauvais que l'on voit souvent un élève*

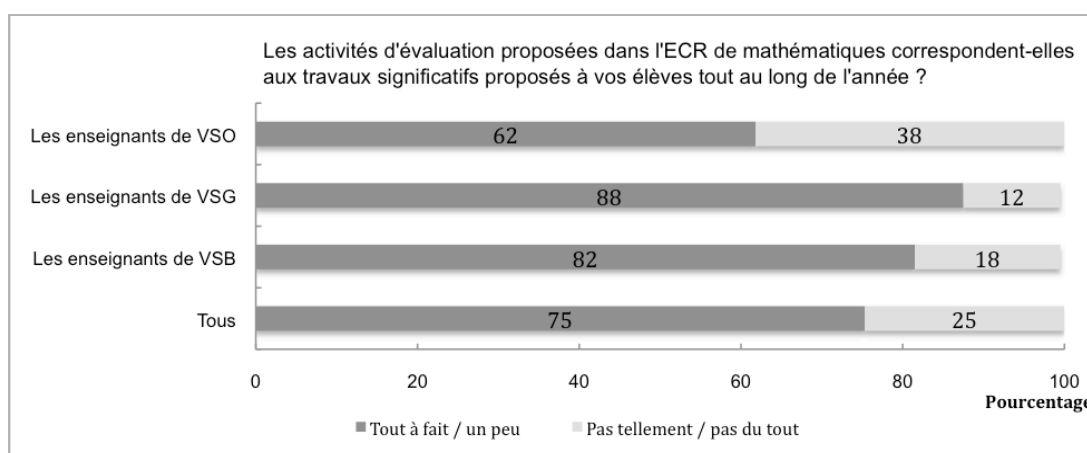
n'ayant pas compris l'exercice (et ne l'ayant pas réussi, bien entendu) obtenir plus de la moitié des points, voire les troisquarts (un enseignant de VSO) ; cette façon de faire, totalement inefficace, implique un temps de correction de plus de 3 heures pour une seule classe de 24 élèves, alors qu'un test significatif bien conçu ne prend jamais plus d'une heure (un enseignant de VSB).

5.5.6 DEGRÉ DE COHÉRENCE ENTRE L'ECR ET LES TRAVAUX SIGNIFICATIFS

La majorité des enseignants (75%) considère que les ECR correspondent *tout à fait* ou *un peu* à ses propres épreuves d'évaluation sommative. Les différences de perception entre les enseignants de VSB, VSG et VSO à ce sujet ne sont toutefois pas statistiquement significatives : $\chi^2(2) = 5.82, p > .05$.

D'après les commentaires de certains, les enseignants de VSO estiment que l'ECR de mathématiques est plus difficile que leurs propres épreuves d'évaluation sommative : *les consignes sont trop compliquées dans l'ECR* (un enseignant de VSO) ; *les tests significatifs en VSO portent sur des sujets vus durant les semaines qui précèdent, plus tard, les élèves ont beaucoup de peine à reproduire ce qu'ils ont vu il y a plus d'un mois* (un enseignant de VSO).

Figure 5.18 : Degré de correspondance perçu par les enseignants entre l'ECR de mathématiques et les travaux significatifs



Les enseignants de VSB estiment en revanche que leurs propres épreuves d'évaluation certificative sont plus difficiles que l'ECR : *je fais plus pointu*, déclare l'un d'eux.

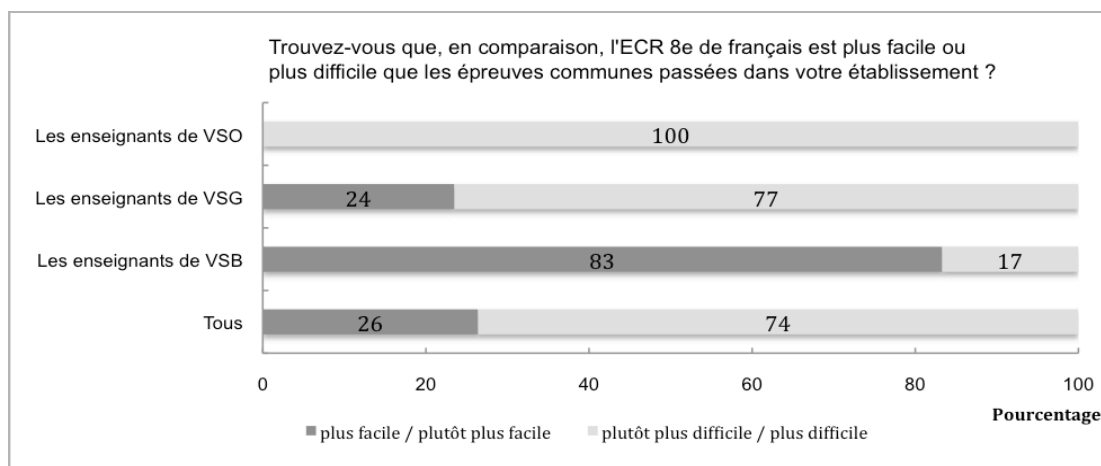
5.5.7 DEGRÉ DE COHÉRENCE ENTRE L'ECR ET LES ÉPREUVES COMMUNES D'ÉTABLISSEMENT

La majorité des enseignants dans toutes les voies, 77% en VSO, 68% en VSB et 59% en VSG dit avoir mis en place une ou plusieurs épreuves communes à plusieurs classes de son établissement au cours de l'année. Dans 95% des cas, ces épreuves ont été passées dans des classes de même voie, rarement de deux voies différentes, jamais de trois voies. En comparaison, l'ECR de mathématiques est perçue comme plus difficile que les épreuves communes d'établissement par tous les enseignants de VSO ainsi que par la majorité des enseignants de VSG (77%).

Dans leurs commentaires, ceux-ci expliquent en ces termes la plus grande difficulté des ECR : *nous évaluons 1 ou 2 thèmes alors que l'ECR mélange plusieurs thèmes, pas*

toujours étudiés (un enseignant de VSO) ; nos épreuves communes sont ciblées sur des thèmes précis. Cette année : fractions, calcul littéral (un enseignant de VSG).

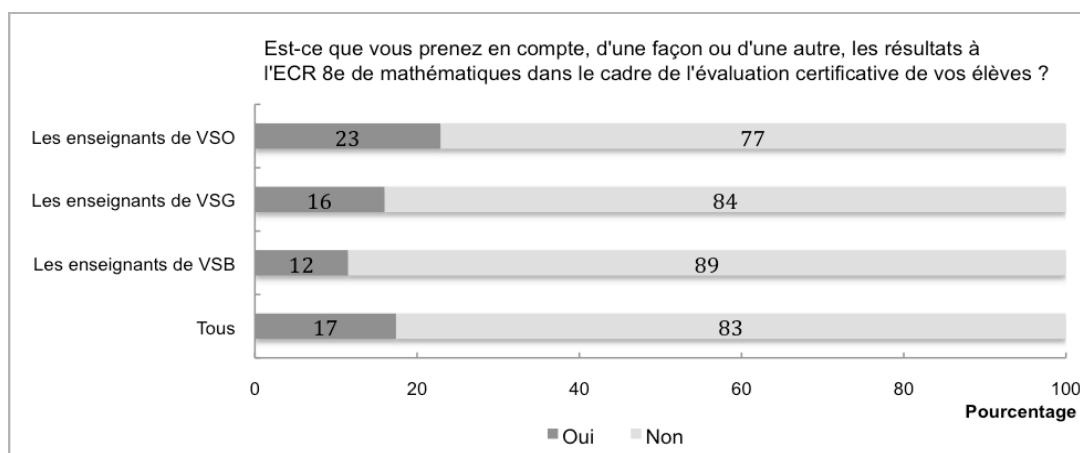
Figure 5.19 : Différence de niveau de difficulté perçue par les enseignants entre l'ECR de mathématiques et les épreuves communes d'établissement



A l'inverse, la majorité des enseignants de VSB (83%) considère que les ECR sont plus faciles. La différence de perception de la difficulté relative de l'ECR comparée aux épreuves communes d'établissement est statistiquement significative : $\chi^2(2) = 28.69$, $p < .001$.

5.5.8 UTILISATION DE L'ECR DANS LE CADRE DE L'ÉVALUATION CERTIFICATIVE

Figure 5.20 : Pourcentage d'enseignants de mathématiques utilisant les résultats de l'ECR dans le cadre de l'évaluation certificative de leurs élèves



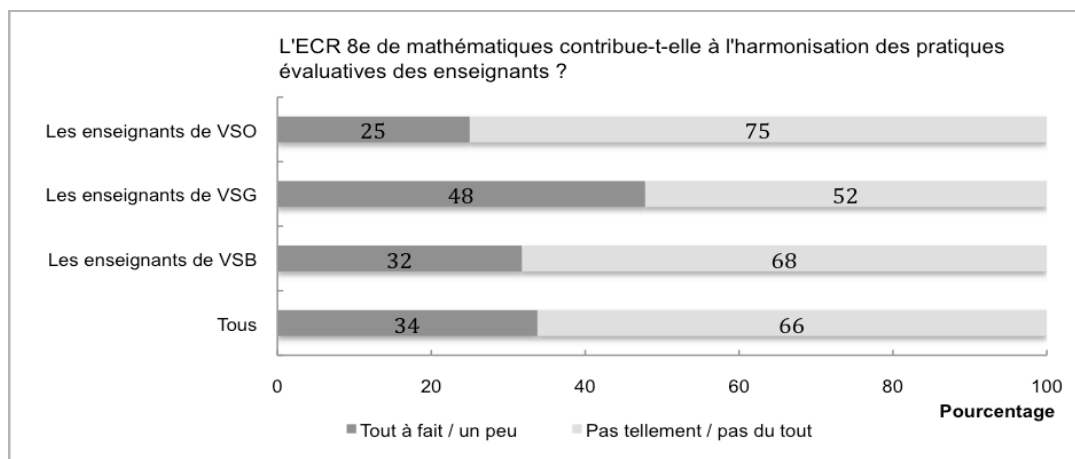
Un peu moins que leurs collègues enseignants de français, quelques enseignants de mathématiques disent utiliser les résultats de leurs élèves à l'ECR, dans le cadre de l'évaluation du travail des élèves : 23% en VSO, 16% en VSG et 12% en VSB. La différence entre les trois voies n'est pas significative.

Dans leurs commentaires, certains enseignants précisent quel usage ils font des résultats à l'ECR : *l'ECR est une bonne indication de maîtrise des objectifs mélangés ; je fais travailler davantage les compétences non maîtrisées ; en reprenant ce qui leur a procuré le plus de difficultés (ex : volume).*

5.5.9 RÔLE DE L'ECR DANS LA RÉGULATION DES PRATIQUES ÉVALUATIVES DES ENSEIGNANTS

Seuls 32% des enseignants en VSB, 48% en VSG et 25% en VSO estiment que les ECR contribuent à l'harmonisation de leurs pratiques évaluatives. La majorité des enseignants ne croit donc pas au rôle régulateur de l'évaluation scolaire que les autorités reconnaissent aux ECR. Les différences entre les trois voies ne sont pas statistiquement significatives.

Figure 5.21 : Pourcentage d'enseignants considérant que l'ECR contribue à l'harmonisation de leurs pratiques évaluatives



Dans leurs commentaires, quelques enseignants ironisent sur la visée harmonisatrice des ECR : *je continue à faire mes évaluations comme il y a 20 ans* (un enseignant de VSG) ; *si vous entendez par là les tableaux d'évaluation compliqués, par objectifs, etc., c'est sûr, là on est harmonisé ; on n'en fait pas, aucun collègue n'en fait* (un enseignant de VSB).

5.5.10 MODALITÉS DE DIFFÉRENCIATION DE L'ECR EN FONCTION DES VOIES

A la question de savoir s'ils préféreraient une épreuve commune unique pour les trois voies, une épreuve différente pour chaque voie ou une épreuve en partie commune et en partie spécifique à chaque voie, les enseignants de mathématiques ont répondu presque comme un seul groupe, sans différence significative liée aux voies d'enseignement.

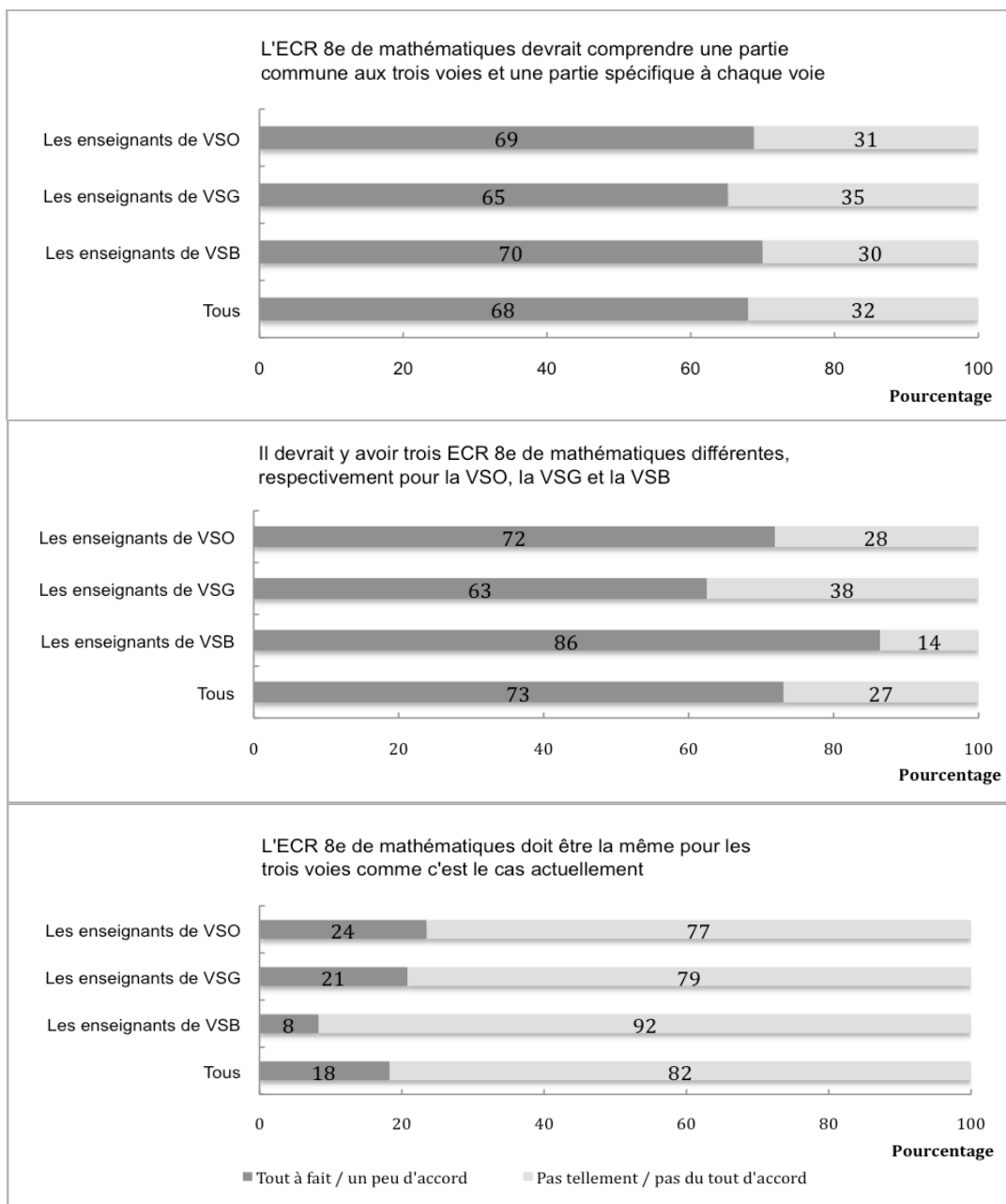
La formule actuelle d'une épreuve unique commune à toutes les voies est rejetée par la majorité des enseignants (92% en VSB, 79% en VSG et 77% en VSO) qui préféreraient lui substituer celle d'une épreuve différente pour chaque voie (86% en VSB, 63% en VSG et 72% en VSO) ou, tout au moins, celle d'une épreuve avec une partie commune et des parties spécifiques à chaque voie (70% en VSB, 65% en VSG et 69% en VSO).

Dans leurs commentaires, les enseignants soulignent les différences de niveau entre les voies : *c'est un leurre d'imaginer que des élèves de 8 VSO et 8 VSB ont les mêmes capacités* (un enseignant en VSB) ; *je suis un des rares enseignants vaudois à avoir enseigné dans les 3 voies, à mon humble avis, il n'est pas possible de faire une épreuve satisfaisante pour les 3 voies (il n'y a qu'à considérer les résultats des VSO = décourageant pour les élèves, voire les profs !!!* (un enseignant en VSB).

Il s'en trouve néanmoins qui soutiennent la formule d'une épreuve unique commune à toutes les voies : *afin d'harmoniser les pratiques il convient de garder une épreuve commune, des épreuves différentes ne feraient que renforcer les écarts "théoriques" entre*

les filières et ne feraient que renforcer l'avis d'enseignants pour lesquels "l'éducabilité" des VSO n'existe que très peu (un enseignant en VSG).

Figure 5.22 : Degré d'accord des enseignants de mathématiques avec trois modalités de différenciation de l'ECR de 8^e année



5.5.11 SYNTHÈSE

- 1 Le degré d'alignement curriculaire de l'ECR de mathématiques est apprécié différemment par les enseignants selon la voie à laquelle ils sont rattachés. La majorité des enseignants de VSB et de VSG estime que tous les problèmes sont bien alignés sur le plan d'études, alors que pour les enseignants de VSO, trois problèmes sur les cinq ne sont pas conformes au programme.

- 2 Le degré d'alignement pédagogique perçu de l'ECR de mathématiques est caractérisé par le même contraste. Pour les enseignants de VSB et ceux de VSG, tous les problèmes sont bien alignés sur l'enseignement effectivement assuré alors que pour ceux de VSO, trois problèmes seraient inappropriés.
- 3 Le degré de difficulté perçu suscite la même opposition entre les enseignants de VSO estimant que trois problèmes sont difficiles et ceux de VSB et de VSG qui les trouvent faciles. Les enseignants de VSG s'accordent avec ceux de VSO pour dire que l'ECR de mathématiques est plus difficile que leurs épreuves communes d'établissement, à l'opposé de ceux de VSB qui les trouvent plus faciles.
- 4 Les enseignants de mathématiques qui ne voient pas l'utilité de l'ECR sont plus nombreux que ceux de français. A peine 17% disent se référer d'une manière ou d'une autre aux résultats à l'ECR, dans le cadre de l'évaluation certificative. En revanche, un tiers d'entre eux reconnaissent à l'ECR un rôle régulateur de leurs pratiques évaluatives.
- 5 Malgré leurs divergences de vue sur la validité curriculaire et pédagogique de l'ECR, les enseignants des trois voies s'accordent pour revendiquer une différenciation de celle-ci en fonction des pratiques d'enseignement spécifiques à chaque voie. La formule actuelle d'une épreuve unique commune à toutes les voies est rejetée par la majorité des enseignants qui lui préférerait celle de trois épreuves différentes ou, tout au moins, celle d'une épreuve avec une partie commune et des parties spécifiques à chaque voie.

6 LES ÉPREUVES COMMUNES D'ÉTABLISSEMENT

6.1 UN CADRE LÉGISLATIF ET ADMINISTRATIF RESTREINT POUR UNE PRATIQUE COURANTE

Le fait d'administrer une même épreuve à plusieurs classes de même degré d'un établissement scolaire est une pratique relativement courante dans le Canton de Vaud. Beaucoup de directeurs d'établissement encouragent cette modalité d'évaluation susceptible de favoriser la collaboration entre enseignants, ainsi que l'harmonisation de leurs pratiques d'évaluation. Cette modalité est perçue comme plus équitable que celle d'une épreuve par classe, dans la mesure où les élèves de même degré, ayant suivi le même programme, devraient être évalués selon des critères communs. Beaucoup d'enseignants et de directeurs d'établissements opposent les épreuves communes d'établissement aux épreuves cantonales dont les standards de contenu et de performance paraissent parfois éloignés de la réalité des apprentissages dans les classes. Ils estiment que les autorités scolaires feraient mieux d'encourager les établissements à développer des épreuves communes locales.

Il nous a semblé important dans cette étude d'aborder quelques questions que cette pratique d'évaluation suscite. Dans quelle mesure les épreuves communes d'établissement sont-elles pratiquées ? Sont-elles communes aux classes de différentes voies comme les ECR ? Ouvriraient-elles une alternative ou seraient-elles un complément aux épreuves cantonales de référence ?

Avant d'expliquer notre démarche de prise d'informations, il n'est pas inutile de préciser que les dispositions législatives ou administratives concernant les épreuves communes d'établissement sont lacunaires. Alors que les procédures d'évaluation du travail des élèves par les enseignants et les dispositions relatives aux épreuves cantonales de référence figurent dans la loi et son règlement d'application, et dans le *Cadre général de l'évaluation*, les épreuves communes d'établissement sont seulement mentionnées dans le *Règlement d'application de la loi scolaire* du 12 juin 1984, adopté le 25 juin 1997. L'article 144 de ce règlement définit, en effet, les tâches du chef de file d'une discipline, dont l'une consiste à *présider à l'élaboration des épreuves communes et d'examens dans l'établissement*. Le *Cadre général de l'évaluation* ne fait aucune mention des épreuves communes d'établissement (DGEO, 2010).

6.2 FRÉQUENCE DES ÉPREUVES COMMUNES D'ÉTABLISSEMENT

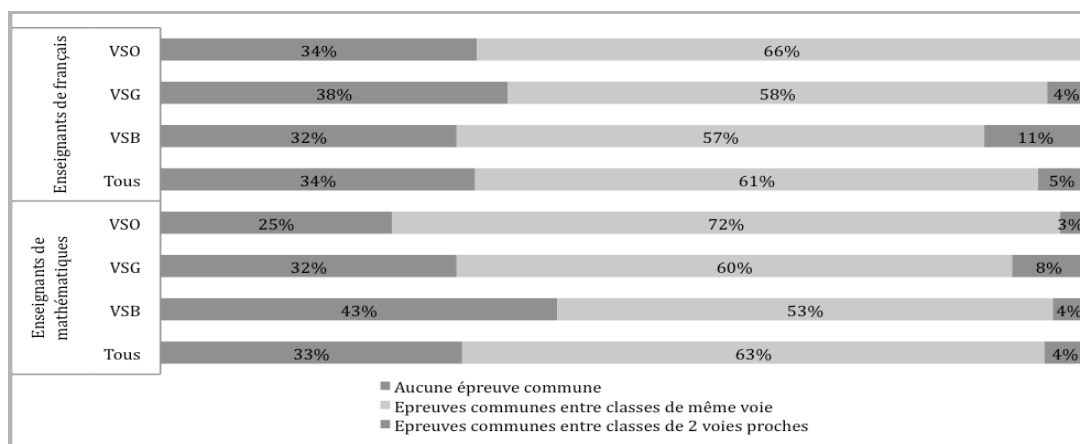
Dans un item du questionnaire aux enseignants présenté au chapitre 5, nous avons demandé à ceux-ci de dire :

- si oui ou non ils ont mis en place une ou plusieurs épreuves communes à plusieurs classes durant l'année scolaire 2010-2011 (les enseignants ont reçu le questionnaire vers la fin de l'année scolaire, au mois de mai) ;
- le cas échéant, si leurs épreuves communes ont été passées dans des classes d'une même voie ou de voies différentes (VSB + VSG, VSG + VSO ; VSB + VSO, VSB + VSG + VSO).

Les réponses des 98 enseignants de français (38 en VSO, 24 en VSG, 36 en VSB) et 89 enseignants de mathématiques (36 en VSO, 25 en VSG, 28 en VSB) à la question de savoir

s'ils ont mis en place des épreuves communes à plusieurs classes de même voie ou de voie différentes sont présentées dans la figure 6.1.

Figure 6.1 : Fréquence du recours des enseignants de français et de mathématiques aux épreuves communes d'établissement selon les voies



On constate que la pratique des épreuves communes est assez courante puisque environ deux tiers des enseignants de français comme ceux de mathématiques disent y avoir eu recours une ou plusieurs fois dans l'année³. Dans presque tous les cas, les épreuves communes sont organisées dans des classes de même voie, rarement dans des classes de deux voies proches (VSO + VSG ou VSG + VSB), jamais entre classes des trois voies (VSO + VSG + VSB).

Le pourcentage d'enseignants de français ayant eu recours aux épreuves communes ne varie pas significativement selon la voie dans laquelle on enseigne. Le pourcentage d'enseignants de mathématiques ayant eu recours aux épreuves communes est un peu plus élevé en VSO qu'en VSG et en VSG qu'en VSB, mais cette variation n'est pas statistiquement significative.

6.3 PRATIQUES D'ÉVALUATION COMMUNE EN FRANÇAIS

6.3.1 D'APRÈS LES CHEFS DE FILE INTERROGÉS

Nous avons réalisé des entretiens semi-dirigés avec des chefs de file de français choisis de manière volontaire dans des établissements pratiquant plus ou moins couramment les épreuves communes.

Au cours de ces entretiens, nous avons abordé quatre séries de questions sur les épreuves communes, en prenant comme année de référence 2010-2011.

- Le nombre d'épreuves communes passées dans l'établissement en 2010-2011 et leur caractérisation, selon qu'elles ont été passées dans des classes de même voie ou de voies différentes (VSB + VSG, VSG + VSO ; VSB + VSO, VSB + VSG + VSO).
- Les dispositions formelles ou informelles concernant les épreuves communes dans l'établissement. Y a-t-il un écrit du conseil de direction de l'établissement concernant l'évaluation commune ? Par qui les épreuves communes ont-elles été

³ Dans ce questionnaire anonyme, nous n'avons pas demandé aux enseignants d'indiquer le nom de leur établissement ; une information qui nous aurait permis d'estimer la proportion d'établissements ayant recours aux épreuves communes.

élaborées ? En cas d'élaboration commune, comment le travail est-il partagé, organisé ? Comment le choix du contenu à évaluer est-il négocié ?

- La pratique des épreuves communes en 2010-2011. Les épreuves communes d'établissement ont-elles été passées dans les mêmes conditions par toutes les classes concernées ? Les critères de correction des copies étaient-ils les mêmes ? Quelle utilisation des résultats à l'épreuve commune ? Le résultat à l'épreuve commune est-il assimilé à un travail significatif parmi d'autres ? Que représente la note obtenue à l'épreuve commune dans la moyenne semestrielle ou annuelle ?
- La perception des épreuves communes d'établissement par les enseignants et les élèves ? Les épreuves communes d'établissement sont-elles mieux ou moins bien perçues que les ECR ?

Les entretiens avec les chefs de file de français ont eu lieu en octobre et novembre 2011. Nous avons interrogé cinq chefs de file de français issus de cinq établissements. D'après les chefs de file interrogés, une ou plusieurs épreuves communes à quelques classes de 8^e année ont été organisées dans leur établissement au cours de l'année scolaire. Parfois, les directeurs les ont exigées mais, le plus souvent, l'initiative est venue des enseignants qui tentent ainsi de réduire les écarts que les parents d'élèves constatent entre les exigences des divers enseignants travaillant dans des classes parallèles.

Dans tous les cas, les pratiques d'évaluation communes diffèrent d'un établissement à l'autre, et pour un même établissement, d'un degré à l'autre et d'une voie à l'autre. Le terme de test commun ne signifie pas précisément qu'il soit le produit d'un travail collectif mené en équipe. Les épreuves communes sont élaborées soit par le chef de file, soit par une équipe de deux ou plusieurs enseignants concernés qui, d'une épreuve à l'autre, alternent les rôles de rédacteur principal et de relecteur. Plusieurs épreuves communes d'établissement échappent ainsi au contrôle du chef de file, ce qui contrevient au règlement scolaire qui veut que le chef de file *préside à l'élaboration des épreuves communes et d'examens* (article 144 du *Règlement d'application de la loi scolaire du 12 juin 1984*, article 35 *Règlement d'application de la loi du 7 juin 2011 sur l'enseignement obligatoire*).

L'attribution très variable selon les établissements de la fonction de chef de file explique en partie la situation. Dans un des cinq établissements, trois chefs de file, un par voie (VSO, VSG et VSB) coexistent sans concertation entre eux. Dans les quatre autres, un ou deux chef de file interviennent dans les trois voies mais en laissant une grande autonomie aux collègues de VSO qui sont invités aux rencontres, sans exigence formelle.

Le travail commun des équipes de maîtres, sous la supervision du chef de file, se concentre dans trois établissements sur une planification annuelle. Le plus souvent, cette dernière indique les contenus à travailler durant l'année scolaire répartie en quatre étapes principales. Les contenus à enseigner relèvent principalement du fonctionnement de la langue (orthographe, grammaire, conjugaison, vocabulaire). Dans les trois cas, les lectures littéraires sont, plusieurs fois dans l'année, effectuées de façon commune, des activités d'écriture en commun sont également proposées dans deux des cinq établissements.

Deux établissements ont institué des règles du jeu précises dans la collaboration. Des réunions fixées dans l'horaire sont organisées chaque semaine ou chaque mois. Il faut relever que dans ces deux établissements cette collaboration ne concerne qu'en partie les collègues de VSO qui travaillent de façon plus individuelle ou en équipe de deux ou trois. La planification, le suivi de la progression dans les classes et les travaux communs sont à l'ordre du jour des réunions d'équipe.

Dans tous les établissements, les chefs de file insistent sur la collaboration informelle, par exemple sous forme de dossiers recensant les activités de tous les collègues et qui sont organisés selon les grandes entrées du français (production écrite, lecture, orthographe, etc.). Les dossiers s'échangent et permettent de fait d'homogénéiser les pratiques en équipes réduites sans qu'un calendrier directif soit imposé. Le terme d'affinité entre les collègues est souvent mis en évidence. Selon les affinités, des équipes entament une collaboration suivie et prévoient quelques tests communs durant l'année par équipes de deux ou trois maîtres sans que les chefs de file ne soient au courant directement et sans qu'il y ait une communication à la direction des résultats de ces travaux communs.

Les réunions de file dans les deux établissements marquant une forte tradition de collaboration fonctionnent, dans un cas, par voie, dans l'autre en associant les VSG et les VSB. Ces réunions permettent de suivre la planification décidée en début d'année, de la réajuster si nécessaire, mais aussi d'échanger des séquences d'enseignement complètes et de faire le point à partir des tests communs. De plus, lorsqu'une épreuve commune est mise en œuvre, on prévoit une séance pour reprendre le corrigé suite aux corrections des travaux d'élèves. L'acceptation de certaines réponses est débattue et les décisions sont prises en toute collégialité. Les notes aux travaux communs permettent d'ajuster les échelles des travaux non communs afin d'éviter des écarts de notes trop importants entre les tests communs et les travaux notés dans le cadre de la classe seulement. Les résultats des diverses classes sont synthétisés sur des tableaux permettant la comparaison entre les classes. Les enseignants, selon les écarts constatés, peuvent ajuster les échelles de leurs évaluations personnelles pour mieux prendre en compte le niveau de leur classe face aux autres classes. Dans quatre établissements, les résultats des diverses classes sont transmis à la direction.

Les chefs de file constatent que les exigences provenant de la direction sont appuyées par la majorité des enseignants qui, dans les deux établissements où la tradition de la concertation est bien établie, relève le gain de temps et l'équité entre les classes comme points positifs de cette approche du travail en équipe. Dans ces deux établissements, les exigences de la direction se fondent sur un discours commun qu'il faut pouvoir tenir aux parents pour se présenter comme une école cohérente dans ses attentes. D'autres membres des équipes dans des proportions très variées selon les établissements (de quelques cas isolés pour un établissement à une évidente majorité dans un autre) préfèrent préserver leur marge de manœuvres et ne prévoient aucune collaboration.

Les résistances des enseignants sont parfois importantes quand le directeur leur demande formellement d'organiser les épreuves communes. Dans deux établissements sur cinq, aucune épreuve commune n'a été organisée dans une voie malgré la demande explicite de la direction. Plusieurs raisons sont avancées par les enseignants pour justifier le refus des enseignants. D'abord, ces derniers pointent la difficulté à mener des progressions similaires dans des classes parallèles. Quelques-uns relèvent le non-sens de calendriers fixés a priori et standardisant la progression des élèves quelles que soient les activités de la classe. Certains y voient un appauvrissement de leur métier d'enseignant qui requiert une marge de manœuvres importante pour être bien réalisé. Les opposants aux épreuves communes mettent également en évidence la difficulté à se mettre d'accord sur un corrigé réellement commun, beaucoup de temps et d'efforts pour, le plus souvent, ne faire plaisir qu'à la direction puisque les résultats des classes ne sont pas repris et analysés pour éventuellement prendre des mesures correctives. Dans un établissement, des négociations difficiles se sont déroulées simplement pour convenir si l'on devait compter des points ou des erreurs. Des conceptions de l'évaluation antagonistes se dévoilent et mettent parfois à l'épreuve les relations entre les collègues. Ainsi, dans cet établissement, les tests communs bousculent les habitudes et inquiètent. De l'avis de ces maîtres, les épreuves communes

représentent beaucoup de travail pour un objectif finalement assez flou. Un chef de file signale un autre point de tensions : le choix des thèmes à évaluer. Ce chef de file relève qu'au final le test commun est un collage d'activités issues de l'expérience des différents collègues et de thèmes hétérogènes. La note à la sortie ne signifie rien, si ce n'est l'addition de points délivrés dans des activités décrochées les unes des autres. *Finalement, les notes sont lissées et la plupart des élèves obtiennent des résultats proches de 4.* Il est vrai, beaucoup de remarques des chefs de file tournent autour des problèmes de calendriers : comment prévoir des tests communs au même moment alors que les enseignants organisent les activités d'apprentissage des élèves de manières fort différentes. Le défi se révèle particulièrement difficile à relever dans les grands établissements où, dans une même voie, sept ou huit classes travaillent en parallèle.

6.3.2 D'APRÈS QUELQUES ÉPREUVES COMMUNES COLLECTÉES

Nous avons récolté quelques épreuves communes de français (cinq épreuves, une par établissement, numérotées de 1 à 5 dans le tableau 6.1) auprès des chefs de file interrogés. Nous constatons que ces épreuves sont destinées, le plus souvent, aux classes de même voie. Une seule épreuve associe les élèves de VSB et de VSG, deux épreuves s'adressent aux classes de VSB, deux autres étant réservées aux classes de VSO.

Tableau 6.1 : Exemples d'objectifs visés et d'activités correspondantes des épreuves communes de français en 8^e

	Voies	Objectifs visés	Activités
1	VSB	FDL : le discours rapporté direct et indirect.	<ul style="list-style-type: none"> Exercices de reconnaissance et transformation de discours rapportés dans un texte littéraire. réécriture d'une BD en faisant alterner récit et dialogue.
2	VSB	FDL : identifier la catégorie et la fonction d'un mot ou d'un groupe de mots ; distinguer les catégories grammaticales des fonctions ; connaître les caractéristiques principales des catégories grammaticales ; orthographier correctement.	<ul style="list-style-type: none"> Exercices de restitution de concepts grammaticaux (tableaux à compléter) ; repérage de catégories grammaticales ; lacunes à compléter avec les pronoms relatifs qui conviennent ; identification des fonctions de pronoms relatifs ; construction d'adverbes ; repérage de fonctions dans la phrase ; repérage de verbes auxiliaires, semi-auxiliaires et à sens plein.
3	VSB VSG	FDL : orthographe.	<ul style="list-style-type: none"> Dictée d'un texte littéraire de 240 mots écrit à l'imparfait / passé simple : orthographe lexicale et grammaticale.
4	VSO	FDL : connaître et utiliser les indicateurs de lieu ; connaître les verbes attributifs ; distinguer les mots du lexique mélioratif, neutre ou péjoratif.	<ul style="list-style-type: none"> Identifier les adjectifs et les noms qu'ils qualifient ; identifier les verbes attributifs en contexte ; compléter un texte lacunaire ; classer des mots selon le niveau de la langue ; retrouver un adjectif qui convient au contexte dans une liste donnée. Ecrire un texte descriptif d'une maison au bord de la mer en respectant le temps de verbe indiqué, en évitant les répétitions et en utilisant les connecteurs de lieu.
5	VSO	Compréhension écrite. FDL.	<ul style="list-style-type: none"> Lecture d'un conte de Grimm et questions de compréhension portant sur le contexte d'écriture (auteur/narrateur), le genre textuel, l'organisation du récit, le rôle des personnages, la trame et la chronologie du récit. Exercices de structuration en lien avec le texte : pronominalisation du groupe nominal, synonymes, antonymes, homonymes, accords dans le GN, accord des adjectifs, choix d'un homonyme orthographiquement et sémantiquement correct.

Le fonctionnement de la langue (FDL) constitue le principal domaine de compétences évaluées à travers les cinq épreuves. Quatre épreuves portent exclusivement sur ce domaine, le cinquième lui associe un deuxième domaine, celui de la compréhension écrite.

Bien que les cinq épreuves communes soient loin d'être représentatives de ce qui se passe dans d'autres établissements, on peut tout de même relever qu'elles couvrent une petite partie des apprentissages réalisés dans une discipline à un moment donné. Alors que les épreuves cantonales de référence sont censées être représentatives des apprentissages réalisés toute l'année dans les principaux domaines, les épreuves communes d'établissement, comme les travaux significatifs d'évaluation certificative au niveau d'une classe, sont davantage ciblés sur une partie des apprentissages réalisés dans un ou deux domaines.

D'après les résultats de l'enquête présentée au chapitre 5, tous les enseignants de VSB et une majorité des enseignants de VSG estime que les épreuves cantonales de référence sont plus faciles que ses épreuves communes d'établissement. Trois constats de l'analyse de contenu des épreuves 1 et 2 destinées aux classes de VSB et de l'épreuve 3 destinée aux classes de VSB et de VSG, confortent cet avis.

- 1 Le niveau d'exigences de la première épreuve mentionnée dans le tableau 6.1 va au-delà de l'objectif fondamental du plan d'études sur lequel il est aligné : *identifier les caractéristiques du discours rapporté direct (DRD) et indirect (DRI)*, prévu dans le plan d'études. La partie *exercices de structuration* demandant aux élèves de repérer dans un texte des DRD et DRI ainsi que de transformer des DRD en DRI semble cohérente avec l'objectif fondamental. Dans la partie rédaction, en revanche, le texte demandé (400 mots) est long. En comparaison, les élèves devaient produire un texte de 200 mots à l'ECR de 2011.
- 2 La deuxième épreuve commune indiquée dans le tableau 6.1 pour les élèves de VSB se révèle également plus exigeante que les ECR en ce qui concerne la grammaire. D'une part, les connaissances d'un métalangage grammatical évaluées sont très précises. Une bonne partie de ce métalangage n'apparaît d'ailleurs pas dans le plan d'études vaudois (verbes auxiliaires, semi-auxiliaires, de sens plein, les types de déterminants – possessif, numéral cardinal, ...). D'autre part, dans le test, le terme de fonction est employé alors que dans le *Plan d'études vaudois*, il est fait mention de rôle syntaxique. On observe donc une exigence élevée dans la connaissance du métalangage grammatical et dans son application pour analyser des phrases. Autant de connaissances qui ne figurent que d'une manière plus simple dans le *Plan d'études vaudois*, concernant en particulier l'objectif fondamental *reconnaître les constituants de la phrase*. De plus, les activités demandées apparaissent de façon décrochée les unes des autres sans que des liens avec des textes réels soient mis en œuvre. En observant ce test commun, se profile en filigrane la question du sens des apprentissages grammaticaux et de la terminologie à employer qui, ici, s'éloigne des attentes du PEV. Grammaire pour la production et la compréhension de l'écrit, ou grammaire pour permettre une description précise de la phrase ou du texte ?
- 3 La dictée soumise aux classes de VSB et de VSG dans la troisième épreuve commune (tableau 6.1) est sans doute plus compliquée que celle faisant partie de l'ECR de français. Le vocabulaire, les temps de verbes, l'apparition de cas particuliers d'orthographe française, la longueur, la cotation, autant d'aspects sur lesquels les deux épreuves se différencient en termes d'exigences. Le texte dicté à l'ECR 8^e de 2011, extrait d'un article de journal écrit au présent, comprend 95 mots. A l'épreuve

commune 3, la dictée porte sur un texte littéraire de 240 mots, écrit à l'imparfait et au passé simple.

Les épreuves communes aux classes de VSO sont, elles, moins exigeantes que les ECR. Outre la prédominance de la structuration relevée ci-dessus, dans les contenus évalués (catégories de mots, fonctions, classe du nom, verbes attributifs, pluriel des adjectifs de couleur, reprises, dérivation, etc.), les quatrième et cinquième épreuves mentionnées dans le tableau 6.1 sont davantage ciblées sur quelques contenus précis, alors que l'ECR se montre davantage généraliste.

L'activité de production écrite prévue dans la quatrième épreuve demande un texte plus bref et plus simple (description d'un lieu au présent ou à l'imparfait avec des connecteurs de lieu) que ce que demande l'ECR (texte complet avec un destinataire précis, consignes plus nombreuses et exigence d'une construction argumentative).

A la cinquième épreuve, le texte à lire (un conte) est long et d'un niveau de langue élevé, mais les questions portent sur une compréhension littérale du texte. A l'ECR, le texte à lire est accessible au niveau du vocabulaire, mais les questions d'inférence ou d'interprétation alternent avec la compréhension littérale.

6.4 SYNTHÈSE

Les épreuves communes d'établissement sont conçues le plus souvent pour des classes de la même voie, parfois pour des classes de deux voies proches, rarement pour des classes des trois voies.

Dans de nombreux cas, les épreuves communes organisées dans l'établissement le sont à l'initiative d'enseignants qui s'accordent pour harmoniser ou au moins mutualiser leurs pratiques évaluatives.

A l'instar des travaux d'évaluation certificative conçus pour une seule classe, les épreuves communes d'établissement portent sur un ou deux domaines de contenu qui semblent être le dénominateur commun de l'enseignement assuré dans les différentes classes concernées. Leur degré d'alignement curriculaire semble limité.

7 CONCLUSION

Le problème de la validité différentielle des épreuves externes est au centre de cette étude. Suivant les théories de la mesure et de l'évaluation en éducation (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2004 ; Laveault et Grégoire, 2002), la validité d'un test standardisé doit être assurée non seulement pour l'ensemble de la population mais également pour les différents groupes d'appartenance de cette population. Le degré de validité d'une épreuve standardisée doit être le même pour les filles que pour les garçons, par exemple. Dans le cas de l'évaluation externe des élèves du secondaire I, les tests standardisés doivent présenter le même degré de validité pour les différentes filières.

L'étude avait pour objectif principal de vérifier l'hypothèse d'un biais de validité des épreuves cantonales de référence (ECR) passées en fin de 8^e année. Depuis leur mise en place en 2007, ces épreuves ont fait l'objet de nombreuses critiques dont la principale est qu'elles ne seraient pas adaptées au niveau des élèves de chacune des trois voies, particulièrement la moins exigeante, VSO, et la plus exigeante, VSB. Conçues comme une épreuve unique alignée sur les objectifs fondamentaux communs aux trois voies, les épreuves cantonales de référence de 8^e n'ont guère convaincu les enseignants qui en attendent des repères pour leurs propres évaluations.

L'étude a mis en évidence les limites des ECR de 8^e année, pour évaluer solidement et équitablement les élèves des trois voies de l'enseignement secondaire I. Afin de les améliorer, nous proposons une différenciation partielle ainsi qu'un réajustement du contenu de ces épreuves cantonales.

Avant la réforme scolaire mise en place après l'adoption de la *Loi sur l'enseignement obligatoire* (LEO), la différenciation partielle des ECR de 8^e année aurait permis d'évaluer plus équitablement les élèves des trois voies (VSO, VSG et VSB). Dans le cadre de l'organisation scolaire actuelle, la différenciation partielle des ECR de 10^e année (l'équivalent de la 8^e année avant la réforme induite par la LEO) permettrait d'évaluer sûrement et équitablement les élèves dans le respect du niveau d'exigences, propre à chaque voie, la voie pré-gymnastique (VP) et la voie générale (VG), ainsi qu'à chaque niveau (niveau 1 et niveau 2) de la voie générale (VG). Elle est d'autant plus pertinente que le statut des ECR change puisqu'elles comptent désormais comme des travaux significatifs d'évaluation certificative.

La validité différentielle limitée des ECR de huitième année

L'analyse du contenu des ECR passées en 2011 met en évidence un alignement curriculaire partiel de celles-ci. Une partie des objectifs pédagogiques évalués à l'ECR de français, par exemple, sont considérés comme fondamentaux en 7^e ou en 9^e mais non en 8^e année, d'autres ne sont reconnus comme objectifs fondamentaux dans aucune degré.

Les épreuves cantonales de référence de 8^e sont en outre inégalement alignées sur le plan d'études de chaque voie. A l'ECR de mathématiques, plus particulièrement, sur 11 objectifs évalués, seuls huit, neuf et sept d'entre eux sont reconnus comme objectifs fondamentaux respectivement en VSB, VSG et VSO.

Une partie des activités d'évaluation qui composent les ECR de 8^e ne sont que partiellement alignées sur le plan d'études, en partie liées à des objectifs fondamentaux et, en partie, à des objectifs non considérés comme fondamentaux. Certaines activités sont même entièrement non alignées au référentiel d'évaluation. L'activité de rédaction d'une

lettre argumentative, par exemple, est entièrement non alignée au référentiel d'évaluation dans la mesure où les objectifs correspondants ne sont considérés comme fondamentaux dans aucune des trois voies d'enseignement pour la 8^e année. Il en est de même à l'ECR de mathématiques où, à titre d'exemple, le problème des *Fractions* est aligné au plan d'études de la 8^e VSO et de la 8^e VSG mais pas à celui de la 8^e VSB, alors que le problème des *Angles* n'est cohérent avec les objectifs fondamentaux d'aucune des trois voies.

Une transgression aussi importante du référentiel conventionnel de l'évaluation certificative que constitue l'ensemble des objectifs fondamentaux de l'enseignement en 8^e année limite la validité curriculaire de l'épreuve. Elle est par ailleurs une source d'inégalité de traitement entre les voies, dans la mesure où certains objectifs évalués ont le statut d'objectifs fondamentaux dans une voie et pas dans une autre. Le plus souvent, à l'ECR de mathématiques surtout, la VSO est la plus nettement désavantagée.

Le fait que les enseignants soient informés, en début d'année, des objectifs évalués à l'ECR ne règle pas le problème. Au contraire, en demandant aux enseignants de préparer les élèves à passer l'ECR en avançant d'une année une partie du programme, ou en renforçant les apprentissages consacrés aux objectifs non fondamentaux, on encourage le *teaching to the test* aux inconvénients si souvent dénoncés. Il n'est du reste pas sûr que les enseignants soient en mesure ou acceptent de changer significativement leur planification de l'enseignement. D'après notre enquête, 42% des enseignants de français et 73% de mathématiques disent ne pas avoir revu leur planification de l'enseignement en fonction de l'annonce des objectifs évalués à l'ECR.

La validité différentielle limitée ressort également de l'analyse des résultats des élèves aux épreuves cantonales de référence passées en 8^e année en 2011. Conçues, semble-t-il, en fonction du niveau des élèves de VSG, elles donnent une mesure biaisée des compétences des élèves de VSO et de VSB qu'elles tendent à sous-estimer dans le premier cas, à surestimer dans le deuxième.

Vers une différenciation partielle des ECR de dixième année

La majorité des enseignants de français et de mathématiques est opposée à l'épreuve unique commune aux trois voies. Seuls les enseignants de français en VSG sont majoritairement favorables à cette conception des épreuves cantonales qui vise à donner l'image de l'égalité des élèves devant l'enseignement malgré leurs différences de parcours dans les trois voies.

Une différenciation totale des ECR (trois épreuves différentes, une par voie) serait le plus sûr moyen d'adapter l'évaluation au niveau d'exigences propre à chaque voie mais cette option ne convient pas aux autres fonctions des épreuves standardisées, celle de l'évaluation du système en particulier. Une comparaison des performances des élèves dans différentes voies et niveaux, afin de renseigner sur le degré d'équité du système d'enseignement, nécessite des tâches communes à tous les élèves.

La formule d'une différenciation partielle (trois épreuves avec une partie commune et des parties spécifiques) serait plus appropriée surtout dans le contexte actuel où les ECR représentent un enjeu significatif pour les élèves. A partir de l'année scolaire 2014-2015, en effet, les résultats aux ECR de 10^e année compteront dans la moyenne du deuxième semestre au même titre que les travaux significatifs d'évaluation certificative conçus et administrés par chaque enseignant. La partie commune, assez importante du point de vue du nombre d'items et de points, supportera la validité de l'épreuve pour toute la population en vue du monitoring du système. Les parties spécifiques, adaptées au niveau

d'exigences propre à chaque voie et à chaque niveau, renforceront la validité différentielle qui pourrait faire défaut à la partie commune de l'épreuve.

Rééquilibrage des domaines de compétences évalués à l'ECR de français

La pondération des points attribués aux divers domaines de l'enseignement du français reflète une certaine prépondérance du fonctionnement de la langue (50% des points attribués). Ce domaine est présent dans chaque partie de l'ECR. En revanche, on observe une sous-représentation de la compréhension écrite. Ce dernier domaine apparaît comme le parent pauvre de l'épreuve analysée en terme d'attribution de points, mais aussi en terme de variété des habiletés de compréhension testées. Le niveau des questions est le plus souvent élémentaire, puisque les tâches relatives à la compréhension sollicitent prioritairement la capacité de repérer des informations figurant explicitement dans le texte. Les processus de compréhension permettant aux élèves d'effectuer des inférences et des raisonnements se révèlent moins sollicités créant ainsi un déséquilibre dans l'évaluation du domaine de la compréhension de textes écrits.

L'épreuve gagnerait en validité de contenu en accordant plus d'importance à la compréhension écrite et particulièrement en questionnant davantage les capacités d'inférence et de raisonnement des élèves.

Davantage de problèmes spécifiques dans l'ECR de mathématiques

La conception des tâches de l'ECR de mathématiques explique en partie sa validité limitée. Les ECR de mathématiques sont généralement constituées de problèmes complexes, chacun comptant pour plusieurs domaines de compétences. Dans l'ECR passée en mai 2011, ce sont cinq problèmes complexes, alignés chacun sur deux à trois des cinq domaines de compétences. Cette ligne s'est un peu assouplie par la suite, en ajoutant aux problèmes complexes quelques exercices d'évaluation de connaissances conceptuelles ou procédurales, adaptés au niveau des élèves de VSO.

L'évaluation standardisée basée sur des tâches complexes pose un problème fondamental de validité que de nombreux chercheurs (Crahay & Detheux, 2005 ; De Ketele & Gerard, 2005 ; Dierendonck & Fagnant, 2010 ; Marc & Wirthner, 2013) ont mis en évidence. Il n'est pas envisageable de proposer plus de cinq problèmes complexes à résoudre dans les 90 minutes que dure une ECR, alors qu'il faudrait au moins deux problèmes par domaine de compétence évalué, soit au minimum une dizaine de problèmes pour composer une épreuve suffisamment représentative des apprentissages réalisés, durant l'année, dans les cinq axes thématiques du *Plan d'études romand*. Cette condition ne peut être remplie qu'en proposant davantage de problèmes spécifiques et moins de problèmes complexes.

Revisiter les épreuves communes d'établissement

Les épreuves communes d'établissement pourraient jouer le même rôle que les épreuves cantonales au niveau de l'établissement. Elles permettraient à la fois d'évaluer plus équitablement les élèves et d'harmoniser les pratiques évaluatives des enseignants. Bien conçues et mises en œuvre, elles contribueraient au monitoring local et au pilotage proximal de l'enseignement.

La pratique des épreuves communes d'établissement dans le Canton de Vaud est cependant loin de cette vision. Microsommatives à l'instar des épreuves certificatives conçues et administrées par chaque enseignant pour sa classe, les évaluations communes d'établissement dans le Canton de Vaud ne portent que sur un ou deux domaines de compétences ayant fait l'objet d'un à deux mois d'enseignement quand les épreuves

cantoniales doivent s'aligner sur le programme annuel. Elles sont communes aux classes de même voie, parfois de deux voies proches, rarement de trois voies.

L'analyse de quelques épreuves communes de français fait apparaître que les standards de contenu et de performance de ces épreuves sont très variables. Un meilleur alignement curriculaire de ces épreuves est indispensable pour qu'elles puissent concourir à l'harmonisation des pratiques d'évaluation ainsi qu'à un traitement plus équitable des élèves. Des dispositions administratives spécifiques au niveau des directions d'établissement contribueraient sans doute à relancer cette piste de régulation des pratiques d'enseignement et d'évaluation.

BIBLIOGRAPHIE

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2004). *Standards for educational and psychological testing*. Washington : American Educational Research Association (2nd impression).
- Antonietti, J.-P., Guignard, N., Mudry, A., Ntamakiliro, L., Riesen, W., Tièche Christinat, C. & Van Der Klink, A.-C. (2005). *Évaluation des compétences en mathématiques en fin de 4^e primaire : résultats de la seconde phase de l'enquête Mathéval*. Neuchâtel : Institut de recherche et de documentation pédagogique.
- Bain, D. (2010). Pour évaluer les qualités docimologiques des tests de maîtrise : l'intérêt de recourir à la généralisabilité. *Mesure et évaluation en éducation*, 33(2), 35-64.
- Bertrand, R. & Blais, J.-G. (2004). *Modèles de mesure : l'apport de la théorie des réponses aux items*. Montréal : Presses de l'Université du Québec.
- Blais, J.-G. (2005). L'obligation de résultats à la lumière de l'interaction entre le quantitatif et le social. In C. Lessard & P. Meirieu (Eds.), *L'obligation de résultats en éducation. Evolution, perspectives et enjeux internationaux* (pp. 123-144). Bruxelles : De Boeck.
- Bloom, B. S. & al. (1969). *Taxonomie des objectifs pédagogiques. I. Domaine cognitif*. Montréal : Education Nouvelle.
- Bloom, B. S., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto : Longmans, Green.
- Carver, R. P. (1974). Two dimensions of tests : psychometric and edumetric. *American Psychologist*, 29, 512-518.
- Crahay, M. & Detheux, M. (2005). L'évaluation des compétences, une entreprise impossible ? (Résolution de problèmes complexes et maîtrise de procédures mathématiques). *Mesure et évaluation en éducation*, 28(1), 57-76.
- Danesi, M. (2011). Des épreuves qui embelliraient les résultats. In *Le Temps*, 31.08. 2011.
- De Ketele, J.-M. & Gerard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et évaluation en éducation*, 28(3), 1-26.
- De Landsheere, V. & De Landsheere, G. (1980). *Définir les objectifs de l'éducation*. Paris : Presses Universitaires de France.
- Dierendonck, C & Fagnant, A. (2010). Quelques réflexions autour des épreuves d'évaluation développées dans le cadre de l'approche par compétences. *Le Bulletin de l'ADMEE*, 2010/1, 5-20.
- Direction générale de l'enseignement obligatoire, DGEO (2010). *Cadre général de l'évaluation. Procédures et dispositions légales et réglementaires en relation avec l'évaluation*. 3^e édition. Lausanne : Etat de Vaud, Département de la formation, de la jeunesse et de la culture.
- Eurydice (2009). *Les évaluations standardisées des élèves en Europe : objectifs, organisation et utilisation des résultats*. Bruxelles : Eurydice.
- Hardy, R. A. (1984). Measuring instructional validity : a report of an instructional validity study for the Alabama high school graduation examination. *Journal of Educational Measurement*, 21 (3), 291-301.
- Klieme, E. (2004). *Le développement de standards nationaux de formation : Une expertise*. Bonn : Ministère fédéral de l'Education et de la Recherche (BMBF).
- Laveault, D. & Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation*. Bruxelles : De Boeck.

- Madaus, G. F. & Stufflebeam, D. L. (1988). *Educational Evaluation : Classic Works of Ralph W. Tyler*. Boston : Kluwer Academic Publishers.
- Marc, V. & Wirthner, M. (2013). *Développement d'un modèle d'évaluation adapté au PER. Rapport scientifique du projet d'épreuves romandes communes*. Neuchâtel : IRDP.
- Martone, A. & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79 (4), 1332-1361.
- Moreau, J. (2004). *Compétences et facteurs de réussite au terme de la scolarité. Analyse des données vaudoises de PISA 2000*. Lausanne : URSP.
- Ntamakiliro, L. & Tessaro, W. (2002). La perception de l'évaluation externe des élèves par les enseignants primaires genevois. *Actes du 15^e colloque de l'ADMEE-Europe et congrès annuel de la SSRE 2002*.
- Ntamakiliro, L. & Tessaro, W. (2010). Plus d'épreuves externes : vers une standardisation de l'évaluation scolaire ? In P. Gilliéron Giroud & L. Ntamakiliro (Eds.). *Réformer l'évaluation scolaire : mission impossible ?* (pp. 165-185). Berne : Peter Lang.
- OCDE (2004). *Compétences en sciences, lecture et mathématiques : le cadre d'évaluation de PISA 2006*. Paris : OCDE.
- Raïche, G. (2004). Le testing adaptatif. In R. Bertrand & J.-G. Blais (Eds). *Modèles de mesure : l'apport de la théorie des réponses aux items*. Montréal : Presses de l'Université du Québec.
- Resnick, L. B., Rothman, R., Slattey, J. B. & Vranck, J. L. (2003). Benchmarking and alignment of standards and testing. *Educational Assessment*, 9 (1 & 2), 1-27.
- Roegiers, X. (2004). *L'école et l'évaluation. Des situations pour évaluer les compétences des élèves*. Bruxelles : De Boeck.
- Shiel, G., Kellaghan, T. & Moran, G. (2010). *Standardised Testing In Lower Secondary Education*. Dublin : National Council for Curriculum and Assessment (NCCA).
- Stufflebeam, D. L., Madaus, G. F. & Kellaghan, T. (2000). *Evaluation models : viewpoints on educational and human services evaluation* (2nd edition). Boston : Kluwer Academic Publishers.
- Trahan, M. et Dassa, C. (1978). La courbe normale en psychométrie et en éduométrie : postulat, constante naturelle ou erreur scientifique ? *Revue des sciences de l'éducation*, 4(1), 73-80.